GNSS pre-correlation waveform feature extraction methods for multipath-afflicted signal classification

Moustakas, Panagiotis

https://doi.org/10.34737/wzw9x

UNIVERSITY OF WESTMINSTER

GNSS pre-correlation waveform feature extraction methods for multipath-afflicted signal classification

written by

Panagiotis D. Moustakas

A THESIS

SUBMITTED TO THE COLLEGE OF DESIGN, CREATIVE AND DIGITAL INDUSTRIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF PHILOSOPHY IN ELECTRONICS ENGINEERING

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

MARCH 2025

Contents

Abstract

Prime goal of this research is to propose and test novel algorithms for GNSS multipath environment classification on the receiver Digital Signal Processing (DSP) stage, but at an earlier processing point than it is used usually; at the generation of the digitized samples of the RF signal. Towards this direction, a detailed study behind the theory and modelling of multipath has been conducted.

Multipath interference is the result of a signal's reception via two or more paths due to reflection or diffraction of the transmitted signal. Physically, the path distance travelled by non-line-of-sight (NLoS) signals is larger than the line-of-sight (LoS) one, therefore all multipath components incident to the antenna arrive with a delay with respect to the corresponding LoS signal. Essentially, the composite signal is a superposition of the LoS and NLoS components which sum constructively and destructively; consequently, segments of the composite waveform are either amplified or attenuated.

Different feature extraction methods were studied and assessed according to their suitability to characterize multipath-contaminated waveforms. Namely, the methods of Generalized Hurst Exponent, Detrended Fluctuation Analysis, Correlation Dimension, Fuzzy Entropy and Recurrence Period Density Entropy were proposed and tested in numerical simulations implemented in MATLAB.

For testing and simulation purposes, the GPS L1 C/A signal structure was selected, as it represents the most fundamental GNSS signal, being the simplest in structure, and potentially allows for further extensions considering more complex-structured signals. The test data are numerically generated by custom MATLAB scripts representing different multipath-afflicted signals. The validity of the selection of feature extraction algorithms was performed by further simulations using off-the-shelf classifier estimators such as LDA and SVM. Finally, a combination of different features is tested to specify the optimal solution to the classification problem.

Acknowledgements

In this section I would like to thank all those who contributed in various ways to the successful completion of this study.

I wish to thank my supervisor Dr. Vasileios Kontogiannis for his continuous support and guidance throughout the course of this research work. Prof. Izzet Kale and Dr. Djuradj Budimir are gratefully acknowledged for their constructive criticism and comments, but also Dr. Dimitris Dracopoulos for his participation in the supervision of this research work.

It would be certainly impossible to finish this research without the support and encouragement of my wife Polyxeni to whom I would like to express my deep gratitude for her patience throughout the whole duration of this study.

List of Figures

List of Tables

List of Symbols

General Symbols

$\sum_{i=1}^{N} x_i$	Summation
$\sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j}$	Double summation
$\sum_{\vec{y}\in \boldsymbol{Y}} \vec{y}$	Summation over vectors $\vec{y}$ belonging to set $\boldsymbol{Y}$
$\int dx$	Integration
$\| \|$	Norm
$\| \|$	Magnitude of complex quantity or absolute value of a real quantity
$\forall$	For every
$\propto$	Proportional to
$\in$	Belongs to
$\ln( )$	Natural logarithm
$\exp( )$	Exponential function
$\Pr( )$	Probability
i	Satellite index
n	Multipath component index
e,j,k,p	Generic indices
$\mathrm{Re}\{ \}$	Real value of a complex quantity
$\mathbb{R}$	Set of Real numbers

Special functions

$E( )$	Electric filed
$\delta(t)$	Dirac delta function
τ	Time delay
$P(\tau)$	Power delay profile
$H( )$	Hurst Exponent
$\langle X(t+\tau)X(t)\rangle$	Auto-correlation function
$F[*]$	Fourier transform of *
$h( )$	Channel impulse response
h_{TF}	Channel transfer function
$HE( )$	Heaviside function
m	Embedding dimension
u_p	State in state-space representation of point p
$C( )$	Correlation integral
$\mathcal{R}_{p,e}$	Recurrent matrix between points p and e
$O( )$	Big O notation

Operators

$\cdot$	Inner product between two vectors
$\times$	Scalar multiplication

List of Abbreviations

ADC	Analog-to-Digital Converter
AGC	Automatic Gain Control
AR	Auto-regressive
ATX	Antenna
BOC	Binary Offset Carrier
BPSK	Binary Phase-Shift Keying
BSS	Blind Source Separation
C/A	Coarse/Acquisition
CD	Correlation Dimension
CDMA	Code Division Multiple Access
CS	Commercial Service
CIR	Channel Impulse Response
DFA	Detrended Fluctuation Analysis
DOA	Direction Of Arrival
DS	Direct Sequence
DSP	Digital Signal Processing
DSSS	Direct-Sequence Spread Spectrum
EC	European Commission
EEG	Electroencephalography
EMD	Empirical Mode Decomposition
E/M	Electromagnetic
ESA	European Space Agency
FE	Fuzzy Entropy
FOC	Full Operational Capability
GEO	Geosynchronous Equatorial Orbit
GHE	Generalized Hurst Exponent
GLONASS	Global'naya Navigatsionnaya Sputnikovaya Sistema
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HE	Hurst Exponent
ICA	Independent Component Analysis
IF	Intermediate Frequency
i.i.d.	Independent and identically distributed random variables
IGSO	Inclined Geosynchronous Orbit
IRNSS/NavIC	Indian Regional Navigation Satellite System/ Navigation with Indian Constellation
ISI	Inter-symbol interference
ITS	Intelligent Transportation System
KL	Kullback-Leibler
LDA	Linear Discriminant Analysis
LNA	Low Noise Amplifier
LoS	Line-of-sight
MEO	Medium Earth Orbit
MIMO	Multiple-Input Multiple-Output
MP	Multipath
NLoS	Non-Line-of-Sight
NMF	Non-negative Matrix Factorization
OFDM	Orthogonal Frequency-Division Multiplexing
OS	Open Service
PCA	Principal Component Analysis
PDF	Probability Density Function
PLL	Phase Lock Loop
PPS	Precise Positioning Service
PRN	Pseudorandom Noise

PRS	Public Regulated Service
PSD	Power Spectral Density
QZSS	Quasi-Zenith Satellite System
RF	Radio Frequency
RFI	Radio Frequency Interference
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RSA	Rescaled Range Statistical Analysis
SPS	Standard Positioning Service
SDR	Software-Defined Radio
SIR	Signal-to-Interference Ratio
SIS	Signal-In-Space
SMR	Signal-to-Multipath Ratio
SSA	Singular Spectrum Analysis
SSS	Small Sample Problem
STD	Standard Deviation
SVD	Singular Value Decomposition
SVM	Support Vector Machine
S-V	Saleh and Valenzuela
TXCO	Temperature Compensated Crystal Oscillator
WGN	White Gaussian Noise
WSS	Wide Sense Stationary

1 Introduction

The purpose of this research is to propose an alternative to the existing approaches for classifying the multipath propagation environment of an independent receiver, based on the use of the RF samples available at the end of the RF-frontend. The proposed approach uses these samples as inputs into one or several feature extraction algorithms, and then the extracted feature vector is passed on to an off-the-self classifier. The main difference with respect to conventional/traditional GNSS receiver processing is the addition of a feature extraction module instead of the common matched filter (usually combined with a discriminator scheme). A classifier can be added in both schemes at the end of each processing chain. The proposed approach offers the advantage that it is capable of classifying very challenging environments where there are very few (or none) and weak direct satellite signals available, and correlators receivers fail to produce a correlation result. The main attribute is that there is no dependency on post-correlation observables. Based on [1], another advantage of adopting such an approach is that interference information is maintained since it avoids the de-spreading operation which inevitably causes losses and decreases the precision of interference detection.

Based on extensive research survey, works on pre-correlation digital signal processing (DSP) stage for multipath detection and characterization are highly limited and currently, they are implemented only in experimental software-defined-radio (SDR) receivers. In addition, a thorough research on components provided by commercial GNSS-receiver manufacturers also indicates that GNSS sampler components are very few, and the corresponding outputs are not used for multipath detection. Therefore, the topic aims to investigate a new approach exploiting observables from a stage of the receiver processing chain that is not yet studied in detail.

It is speculated that this lack of relevant hardware is due to the high-end requirements of such components which translates to increased manufacturing cost, an area where the SDR technology can provide a solution. Powerful processors and large memory units are nowadays available in many every-day-life devices suggesting that the technology under study might be already possible to be implemented in some applications such as transportation in cities where the urban multipath channel could be compromised by various interferences.

Another aim of this research is to propose alternative approaches that are capable of capturing the multipath and radio frequency interference (RFI) patterns in feature space. For this purpose, a research in the suitability of candidate metrics in other technical and scientific fields was conducted. Since the problem of adaptive channel characterization and context awareness is not specific to GNSS applications, works on other RF applications such as radar or Multiple-Input Multiple-Output (MIMO) where reviewed, as well as from the field of biomedical and speech signal processing, among several others.

The review of the state-of-the-art, analysis and results of this research work are presented in the next five chapters. In Chapter 2 the description of the basic physics, and the different types of GNSS multipath phenomenon is described, presenting the necessary mathematical background to model its behaviour. Chapter 3 provides a review of the current state-of-the-art techniques used in detection, characterization, mitigation and estimation of generic GNSS interference, and further assessing their applicability to GNSS multipath. In Chapter 4, the feature-based classification approach in introduced along with a review of candidate feature extraction techniques. Then, according to the theoretical criteria that indicate the suitability of the techniques, five feature extraction are presented and analytically described. Numerical simulations assess the capability to detect different multipath propagation environments for each of the five feature extraction methods. Chapter 5 demonstrates the performance of multipath environment classification by employing the feature extraction results from the previous

chapter, and by using off-the-shelf classifiers such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). Finally, Chapter 6 includes conclusions and for future work recommendations.

1.1 Global Navigation Satellite Systems and main characteristics

Global Navigation Satellite Systems (GNSSs) constitute constellations of satellites providing radio-frequency signals from space, that disseminate positioning and timing data to user receivers on a global scale. All GNSSs provide signals on at least two different frequencies to cancel out the effect of ionosphere in their measurements, in the range of 1.2 - 1.6 GHz (L-band) corresponding to wavelengths of 19−25 cm. The user receiver can distinguish between satellites that transmit on the same frequency by a unique pseudo-random noise sequence modulated on the corresponding carrier wave of each satellite. Each satellite signal carries also information on the orbital position and its offset to system time, to facilitate time (or distance) measurements on signal propagation from a satellite to the receiver. By considering the system time, the local receiver time, and the signal propagation time, the receiver can estimate the distance from receiver to satellite. For positioning, it essentially requires at least four distances (four satellites) to solve for the four unknown quantities: the three spatial coordinates and the time offset from GNSS time.

There are four GNSSs available at the time of writing of this Thesis: GPS, GLONASS, Galileo and Beidou. The Quasi-Zenith Satellite System (QZSS) and the Indian Regional Navigation Satellite System (IRNSS)/NavIC are excluded as they are considered regional navigation satellite systems. A summary of the main features of each GNSS is given in the table below:

Table 1 - Main GNSS characteristics [1]

System	GPS	GLONASS	BeiDou	Galileo
Orbit	MEO	MEO	MEO, IGSO, GEO	MEO
Number of Satellites	24	24	27,3,5	30
Constellation	6 planes 56º inclination	Walker (24/3/1) 64.8º inclination	Walker (24/3/1) 55º inclination	Walker (24/3/1) 56º inclination
Nominal orbital altitude (Km)	20180	19100	MEOs 21530 IGSOs 36000 GEOs 36000	29600.318
Frequencies (MHz)	L1 1575.42 L2 1227.60 L5 1176.45	L1 1602.00 L2 1246.00 L3 1202.025	B1 1561.098 B2 1207.14 B3 1268.52	E1 1575.42 E5a 1176.45 E5b 1207.14 E6 1278.75

The Global Positioning System (GPS) is the first GNSS achieving Full Operational Capability in June 1995 while the launch of the first Navstar satellite dates back to 1978. GPS is operated by the US government and initially provided two services: Precise Positioning Service (PPS) on the GPS L1 (1575.42 MHz) and L2 (1227.6 MHz) frequencies and the Standard Positioning Service (SPS) on the GPS L1 frequency containing a coarse/acquisition (C/A) code. PPS L1 and L2 also include an encrypted precision (P) code ranging signal (also known as the Y-code) with a navigation data message for authorized users. Later three new signals were added L2C and L5 (1176.45MHz) for civilian users, and M-code at L1 & L2 for military/restricted use.

GLONASS stands for Global'naya Navigatsionnaya Sputnikovaya Sistema, it is developed and operated by Russia achieving a fully operational constellation of 24 satellites in orbit by 1996. However, shortly after that date, the number of operational satellites started to drop due to financial problems and the next time it achieved full operational capability was by the end of 2011. GLONASS signals are designed based on the frequency division multiple access (FDMA) scheme, and in the original design, the system transmitted the signals within bands Ll (1602−1615.5 MHz) and L2 (1246−1256.5 MHz), at frequencies spaced by 0.5625 MHz at L1 and by 0.4375 MHz at L2. The first block of satellites that deploys code

division multiple access (CDMA) is GLONASS-K satellites (also transmitting FDMA signals). Finally, GLONASS-K1 as well as GLONASS-M satellites transmit a CDMA signal on a new L3 frequency (1202.025 MHz).

The Galileo programme is owned by the EU, with the European Commission bearing the overall responsibility of managing and overseeing the implementation of all activities on behalf of the EU. The first two satellites were launched on October 2011, and for its Full Operational Capability (FOC) consists of 24 operational satellites. Galileo offers several types of services including Open Service (OS), Public Regulated Service (PRS), Navigation Message Authentication, and High Accuracy Service. The Open Service (OS) and the Public Regulated Service (PRS) are transmitted in the E1 frequency band centered on 1575.46MHz, the same as the GPS L1 frequency. The PRN ranging codes are modulated onto the carrier using binary offset carrier (BOC) techniques. The Commercial Service (CS) signal and the PRS are transmitted in the E6 frequency band centered on 1278.75 MHz using binary phase-shift keying (BPSK) and BOC modulation, respectively. Data and data-less (pilot) signals are transmitted in the E5 frequency band centered on 1191.795 MHz, again using BOC modulation. Data and pilot signals are also available on E1 and E6. The signals are separated into an E5a and an E5b component and either can be tracked separately or together.

Finally, China begun the development of a satellite-based navigation system known as BeiDou (Chinese for the Big Dipper asterism) in 1980. The initial constellation of three geosynchronous equatorial orbit (GEO) satellites was completed in 2003. A fourth GEO satellite was launched in 2007. The initial regional BeiDou system (BeiDou-1) has been replaced by a global system known as BeiDou-2 (or simply BeiDou and formerly known as Compass). The BeiDou Navigation Satellite System (BDS) as it is officially now known will eventually include five GEO satellites, 27 MEO satellites, and five Inclined Geosynchronous Orbit (IGSO) satellites. BeiDou-2 provides global coverage since 2020. The BeiDou-3 satellites have been launched since March 2015, which transmit two levels of service, an open service and an authorized service primarily for the Chinese government and military using three frequency bands. The bands and the central frequencies for the satellites now in use, the BeiDou-2 satellites, are B1 at 1561.098 MHz, B2 at 1207.14 MHz, and B3 at 1268.52MHz. BeiDou-3 transmits signals in L1/E1 and L5/E5 bands as well as the BeiDou B3 band.

2 Multipath interference in GNSS receivers

2.1 Introduction

This chapter aims at reviewing the fundamental physics and mathematical tools that describe analytically the multipath (MP) interference phenomenon in GNSS receivers.

In radio frequency (RF) systems such as GNSSs (L-band), the transmitter sends a signal to a receiver using a physical channel which - in almost all realistic scenarios - alters its physical properties. As a result, a modified waveform is incident to the receiver antenna compared to the one originally transmitted from the corresponding satellite. Multipath interference (experienced on the receiver side), is the result of electromagnetic interaction of the emitted signal with its reflected, diffracted or refracted copies due to the propagation-channel properties. The path distance travelled by the multipath echoes is expected to be larger than that of the corresponding direct line-of-sight (LoS) signal, and naturally, the echoes arrive to the receiver with a delay relative to the LoS signal. The multipath components are usually referred in literature as non-line-of-sight (NLoS) signals.

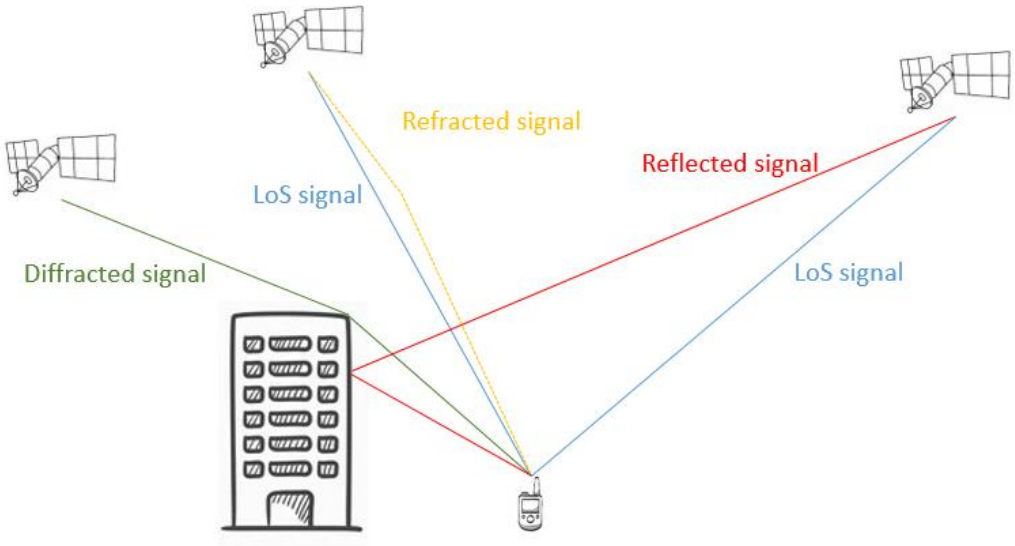


Figure 1 - MP propagation in GNSS systems

In the GNSS context, most user-receivers are located near the surface of the earth and the signals are transmitted from satellites orbiting at a distance of over 20,000 Km. A schematic of generic multipath propagation is shown in Figure 1, where all the above-mentioned physical phenomena are depicted. Causes of multipath propagation include atmospheric phenomena such as tropospheric refraction, ionospheric reflection and refraction, as well as reflections from terrestrial objects located in the vicinity of the receiver. Scattering can also be considered as another physical phenomenon that leads to multipath, as GNSS signals can propagate through street signs, foliage, or walls, causing the reflected energy to spread out (scatter) in all directions.

The result of reception of a variety of these signals from a GNSS receiver is the formation of a composite signal including both LoS and NLoS components. Specifically, the NLoS signals can either sum up constructively or destructively with the LoS one (but also between them), causing amplification or attenuation of the composite signal. From the user perspective, the communication might be affected when destructive interference occurs causing the signal power to fade or, in severe cases, complete loss of communication. Scattering of the LoS signal could also take place especially in indoor propagation

environment where no LoS component is present. This extreme scattering environment generally characterized in literature as NLoS conditions.

In general, the physical model for GNSS multipath propagation is built by considering the properties of physical channel that the RF signal propagates. Channel properties are determined by the number, geometry and material properties of different media included in the signal propagation environment and the relative transmitter-receiver geometry. Mathematically, the multipath interference phenomenon is due to the superposition of signals arriving at the same location (antenna) at one specific time(-instance). Aside from the shift in time with respect to the LoS signal due to the delay introduced from the longer propagation path, each of these signals will have altered signal properties, such as different phase-shifts, after passing over transmission paths of different lengths [2], and/or different frequencies due to the interaction of the radio waves with the objects of different material properties.

A key assumption used for modelling the resulting time-domain superposition of the different signal amplitudes is that RF waves are planar, which has proved to be an accurate approximation. At transmission from a single-point satellite source, the signal propagates outwards in all directions, and a reasonable choice to mathematically describe it would suggest to adopt the spherical waves model. However, for travelling distances of (approx.) 20,000 Km where satellite transmitters fly above the Earth's surface, the received signal wave-front (surfaces of constant phase) will appear as having the same amplitude everywhere on the plane perpendicular to its direction of travel in the (near) vicinity of the receiver. The plane wave simplification does not imply that the wave is traveling in a single direction and, as shown in Figure 1, due to the fact that the signal is spreading out in all directions, signals in red and yellow are present (on top of the blue ones, i.e. LoS signals) and contribute to the composite signal received. A simplistic model to mathematically describe an E/M wave is by considering an unmodulated carrier wave - that corresponds to a sinusoidal function of constant period/frequency - that oscillates around zero at the maximal amplitude and minimal amplitude A [3]:

$$E(\vec{r},t) = \frac{A}{\vec{r}} e^{j\omega t - j\vec{k}\cdot\vec{r}} \quad (1)$$

with E the electric filed of the signal, $\vec{k}$ the wave-vector in 3 dimensions, $\vec{r}$ the position vector (usually including Cartesian coordinates X, Y, Z), and $\omega = 2\pi f$ the angular frequency (f stands for the frequency). In GNSS systems a geocentric reference system is used, and if the source of the wave is located at distance $-X_0$ from the receiver, under the assumption that $X_0 \gg X, Y, Z$, the in-between (Euclidean) distance $|\vec{r}| = \sqrt{(X+X_0)^2 + Y^2 + Z^2}$ can be approximated as $|\vec{r}| = X + X_0$ after a Taylor expansion (and neglecting higher order terms for X, Y, Z). Intuitively, this means that a displacement along the X dimension causes a proportional change in the radius of the spherical wave, while any displacement along the other two dimensions causes a negligible first-order effect. Therefore, the previous equation could be replaced by:

$$E(\vec{r},t) = \frac{A}{X+X_0} e^{j\omega t - jk(X+X_0)} \quad (2)$$

Here, the wavenumber k is used instead of the wave vector $\vec{k}$, since it represents the magnitude of the latter and, quantifies the phase variation for the plane wave $k = 2\pi/\lambda$, where λ stands for the wavelength. By factoring out the X_0-dependent term the equation becomes:

$$E(X_0,t) = \frac{A}{X_0} e^{-jkX_0} e^{j\omega t - jkX} \quad (3)$$

with constant amplitude $A_0 = A/X_0$ and phase shift $\phi = e^{-ikX_0}$. Here, the influence of X is neglected as $X_0 \gg X$. This result is consistent with the direct derivation of the planar wave solutions from the Helmhotz equations [4]. Assuming that the antenna/receiver measures the signal at fixed position or at

time-intervals where the radial distance can be considered approximately constant, the X-related term can be omitted, yielding the model for the received unmodulated signal:

$$E(t) = A_0 e^{j\omega t} \quad (4)$$

The plane wave approximation result of (4), which is a simplified mathematical formulation of signals propagating under the underlying electromagnetic filed, can serve as the basis of a (more straightforward) multipath interference model.

A starting point to mathematically describe the resulting superposition of with the generated copies of the LoS signal is through the superposition principle. The superposition principle holds only if a linear response is assumed on the propagation channel. Thus implying that only reflective phenomena take place (also implying that the frequency of all signals remains the same) due to the objects in the vicinity of the antenna/receiver, or "mild" atmospheric conditions. In this case, the amplitude of the received wave in each time-instant shall be equal to the sum of the wave amplitudes of the individual waves. A common approach to express the multipath superposition of the received signal is by employing the formulation of impulse responses on linear systems, and, as a starting point, introducing an ideal Dirac pulse of electromagnetic power transmitted at time 0 (represented by $\delta(t)$). Replacing the angular frequency ω in (4) with the phase term ϕ yields the impulse response:

$$h(t) = \sum_{n=0}^{N} A_n \, e^{j\phi_n} \delta(t - \tau_n) \quad (5)$$

where N is the number of received impulses (or number of electromagnetic paths), τ_n is the excess delay due to multipath propagation of the generic nth impulse, A_n is the signal amplitude and ϕ_n the phase of the nth received pulse respectively. The analytical expression for the phase is $\phi_n = 2\pi f + \varphi_n$, with f the frequency (assumed to remain intact for all components) of the signal and φ_n the phase offset of each NLoS signal. It is noted that the LoS signal represented by index with $n = 0$ and $\tau_0 = 0$ and it is assumed that the channel has infinite bandwidth. Essentially, t is the time of observation from the receiver, and $t - \tau_n$ the time the impulse was introduced to the channel, i.e. τ_n in the past. By $s(t)$ we denote the signal displacement which is equal to the Channel Impulse Response (CIR) function $h(t)$ of the equivalent multipath model.

The summation of amplitudes A_n, or the CIR magnitude $h(t)$, as described by (5) results in constructive and destructive interference of the LoS signal, and causes its phase to change. This phase shift is due to the combination of out-of-phase signals. For the scenario where the antenna is moving, the frequencies of the different replicas might slightly differ, and apart from the amplitude modulation phenomenon described, phase modulation is also expected to occur [5]. Assuming that all signals have the same frequency and the phase difference between them is constant i.e. they are coherent, a fixed (stationary) interference pattern is expected to emerge manifesting the result of linear superposition. However, linear superposition almost never occurs in real conditions but there are certain static scenarios that coherence approximately holds such as in specular multipath propagation which is described in the next paragraph.

An alternative way to formulate the received signal $s(t)$ in (5) that is very common in literature is by showing explicitly the multipath relative amplitude a_n:

$$s(t) = \mathrm{Re}\left\{A_0 \, e^{j\phi}\delta(t) + A_0 \, \sum_{n=1}^{N} a_n \, e^{j\phi_n}\delta(t - \tau_n)\right\} \quad (6)$$

Here the first term represents the LoS signal which is separated from the multipath components showing the difference of the emitted (LoS) signal from the received signal $s(t)$. Also, A_0 is complex amplitude and in realistic conditions, a_n is the product of the free space loss, the signal attenuation due to reflection, the depolarization loss and the signal attenuation due to the antenna gain pattern. Possible attenuation due to shadowing can also be present.

In general, geometrical reflection conditions might be dynamic, and parameters A_n, ϕ_n, and τ_n shall be considered as time-varying. Respectively, at each time t each replica $n \neq 0$ arrives to the antenna/receiver system with a different delay $\tau_n(t)$, a different amplitude $A_n\big(t, \tau_n(t)\big)$, and goes through a different phase shift $\phi_n\big(t, \tau_n(t)\big)$. In the CIR given next, the LoS signal and the time-dependence of the excess delay in the arguments of relative amplitude and phase is omitted:

$$h(t,\tau) = \sum_{n=1}^{N} a_n(t,\tau)\, e^{j\phi_n(t,\tau)} \delta\big(\tau - \tau_n(t)\big) \tag{7}$$

In the next figure, τ is split into delay bins τ_i, with $i \in [1, N]$ each representing a multipath component, while also demonstrating the different excess delay profiles in discrete times t_0, t_1, t_2, t_3. Here, the multipath delay $\tau(t)$ denotes the parameter that takes excess delay τ_i.

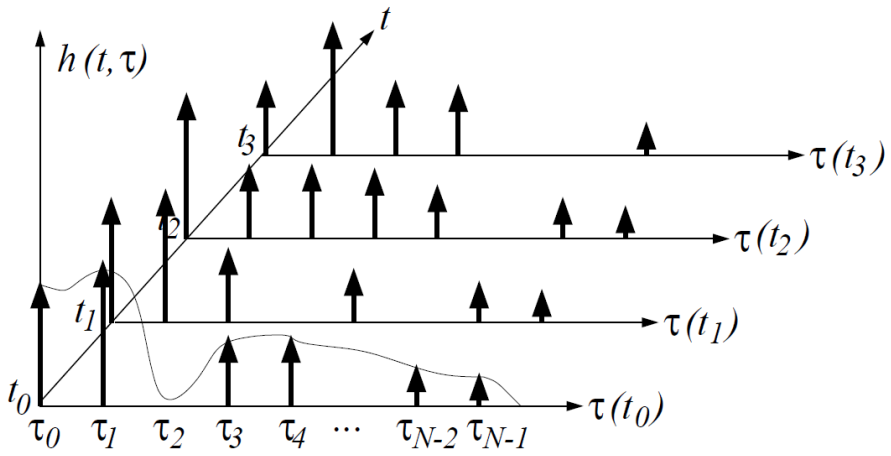


Figure 2 - CIR and delay bins

In noisy channels, where increased multipath and shadowing are present, such as urban canyons, signal amplitude fluctuations in time are pronounced due to the highly dynamic nature of the channel parameters.

2.2 Multipath interference characteristics of GNSS channels

In this paragraph, the metrics that quantify the most fundamental properties that characterize multipath interference are briefly described. In formulas $(5)-(7)$ it is assumed that N resolvable multipath components exist. Two multipath components with excess delays τ_1 and τ_2 are resolvable when their delay difference is very large compared to the inverse signal bandwidth $\tau_1 - \tau_2 \gg 1/B_{signal}$ and non-resolvable when $\tau_1 \approx \tau_2$ which, as a consequence, are observed as a single multipath component. Therefore, given that this assumption is valid, a first fundamental concept related to the excess delay parameter τ, is the multipath time T_M, which corresponds to the difference between the first and last received (resolvable) impulses:

$$T_M = \tau_N - \tau_1 \tag{8}$$

In real conditions the multipath time is computed by considering as last impulse the first one which allows to receive a determined amount of the total transmitted power (scaled by the atmospheric and propagation losses), for instance 99%. The coherence bandwidth can therefore be defined as:

$$B_C \approx \frac{1}{T_M} \tag{9}$$

Keeping the assumption of a linear time invariant channel, another metric that characterizes multipath is the channel transfer function h_{TF}, which is the continuous-time Fourier transform of the impulse response $h(t)$ given in (5):

$$h_{TF} = F[h(t)] = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt = \sum_{n=0}^{N-1} \rho_n e^{j\phi_n} e^{-j2\pi ft} \tag{10}$$

For this result the property of Fourier-transforming a Dirac-delta to a complex exponential function was used, and the resulting function represents an eigen-function of every linear system. A practical rule derived from the obtained channel transfer measure states that the distance (in frequency) between two consecutive local maxima or minima is approximately inversely proportional to the multipath time.

Another useful analysis tool used to analyse multipath interference is the squared CIR of (7) which generates the power delay profile $P(\tau)$. It is described by the power spectral density as function of delay, quantifying how the channel power is distributed along the different excess delays. An alternative way to describe it is its correspondence to the Fourier transform of the auto-correlation function of the CIR (7). A more practical quantity however, is the mean (excess) delay which is computed as the first moment of the power delay profile:

$$\overline{\tau} = \frac{\sum_{n=1}^{N} P(\tau_i)\tau_i}{P_T} \tag{11}$$

The denominator is corresponding to the total power in the channel $P_T = \sum_{n=1}^{N} P(\tau_i)$. A more detailed indicator of multipath system performance is the second moment of the power delay profile, usually referred to as root mean square (RMS) delay spread [6]. The latter takes into account the relative powers of the multipath components as well as their delays, and it is described mathematically as:

$$\tau_{RMS} = \sqrt{\frac{\sum_{n=1}^{N} P(\tau_i){\tau_i}^2 - \overline{\tau}^2}{P_T}} \tag{12}$$

The reciprocal of the RMS delay spread $B_c \approx 1/\tau_{RMS}$, is the coherence bandwidth, which was also defined in an alternative definition in (9). It is used as a measure of the range of frequencies over which the channel is "flat", i.e. specific spectral components (corresponding to a power delay profile) that have approximately equal gain and linear phase. The implication of a variable carrier frequency is that the

magnitude of the change in amplitude will also vary and therefore B_c quantifies the minimum separation in frequency after which two signals will experience uncorrelated fading. Specifically, in flat fading conditions the coherence bandwidth of the channel is larger than the bandwidth of the signal and all frequency components of the signal will experience the same magnitude of fading. On the other hand, in frequency-selective fading the coherence bandwidth of the channel is smaller than the bandwidth of the signal and different frequency components of the signal experience de-correlated fading.

Aside from time dispersion caused on the LoS signal, multipath propagation in time-varying channels also introduces frequency dispersion caused by Doppler frequency f_D. When the receiver - or the reflector objects in its local environment are moving - a shift in the frequency of the signal transmitted along each signal path is introduced, known as the Doppler shift. Signals travelling along different paths can have different Doppler shifts corresponding to different rates of change in phase. In fact, in GNSS systems the movement of the satellites always introduces Doppler shifts approximately +/-5kHz per satellite [51], [52], in a static antenna/receiver scenario.

The spectral broadening due to Doppler effect can be quantified using the Doppler spread D_s, which is essentially the bandwidth of the corresponding Doppler spectrum. Channels with a large Doppler spread have signal components that are each changing independently in phase over time. The reciprocal of the Doppler spread is a very useful measure of characterizing the time variability of the channel in the time domain. The coherence time is defined as:

$$T_{coherence} = \frac{\ell}{D_s} \tag{13}$$

where ℓ is a constant taking on values in the range of 0.25 to 0.5. It is a measure of the minimum time required for the magnitude change of the channel to become de-correlated from its previous value. The coherence time of the channel is practically used to determine if the channel is fading at a slow rate, where the effects of the Doppler spread are negligible, or at a fast rate otherwise. Since fading depends on whether signal components add constructively or destructively such channels are expected to have a very short coherence time.

One of the benefits of a frequency-selective channel is that it is rarely possible for all frequency components to be simultaneously affected by a deep fade. The frequency diversity introduced by the channel causes independent fading conditions between the different frequency components of the signal. This robustness against fading is one of the key features of the Code Division Multiple Access (CDMA) modulation [7], under which the majority of the GNSS signals are designed. However, frequency-selective fading channels are also energy-dispersive, causing the signal energy associated with each symbol to spread out in time and therefore resulting in interference between adjacent symbols. Inter-symbol interference (ISI) can be avoided by introducing a guard interval between the symbols, while for CDMA, the Rake receiver was designed to mitigate these effects by employing several correlators, each assigned to a different multipath component [8].

In conclusion, common measures that quantify MP interference are time dispersion and delay spread [9], Doppler Power Spectral Densities [10], but other statistical measures have also been used such as the channel capacity and mutual information [11].

2.3 Types of RF multipath interference

Multipath interference can appear as periodic signal variations to the received signal, characterized mainly from their frequency, which in its turn depends on the satellite-antenna-reflector geometry. A distinction between two types of multipath interference is usually found in the literature: the low-frequency (slow) or high-frequency (fast) multipath depending on the rate of change of the magnitude and phase introduced by the channel, given that the channel is not time-varying (static surrounding objects and antenna). Another terminology that essentially leads to the same separation is related to the refection surface that generates the observed multipath interference: the diffuse (or dense) multipath caused by rough surfaces which is manifested by fast variations in the received signal, and the specular multipath caused by smooth surfaces which appear as slow variations [12] (Figure 3). The roughness of a surface depends on the incident angle θ_i, and it is characterized by the Rayleigh criterion [13]:

$$\Delta h < \frac{\lambda}{8 \cos \theta_i} \tag{14}$$

Here, Δh represents the height difference of two points on the incident surface, and λ corresponds to the E/M wave wavelength [6].

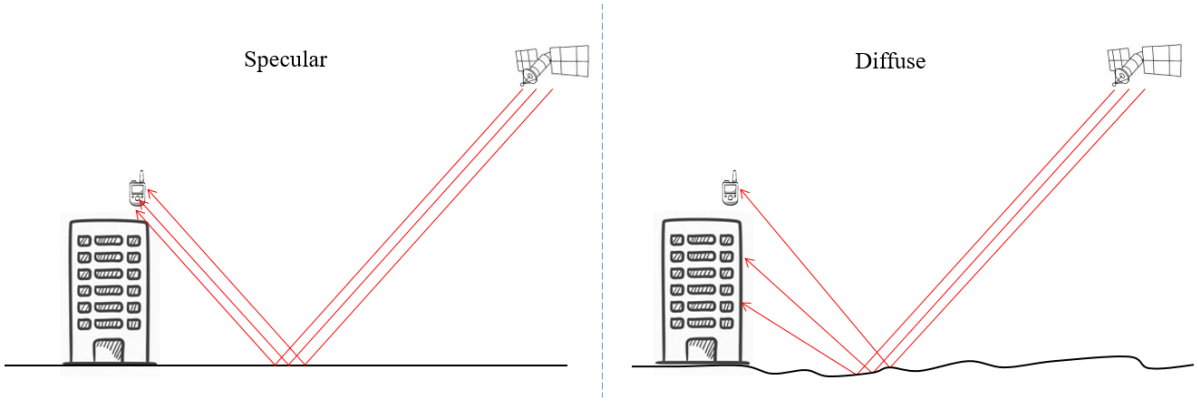


Figure 3 - Specular and diffuse multipath

In RF systems mirror-like smooth surfaces can be manifested as a single point source due to the fact that several coherent and directional propagation paths are added over a surface area, which introduce a dominant pattern in the signal. For a stationary transmitter-receiver-reflecting surface geometry the specular reflection is expected to show a constant phase difference with respect to the LoS signal. On the other hand, the randomness of the phase of the diffuse propagation paths (non-coherent) and the scattering of signal power in various directions make this type of multipath appear weaker than the specular case. The diffuse reflection then consists of a collection of signals of varying amplitudes and phases [14]. In addition, diffuse multipath might appear to originate from a sizable region and it is spread in path delay. Although in some cases the maximum absolute amplitude of the diffuse multipath may be relatively weak, the total strength of the diffuse multipath can become significant when it is non-coherently combined after reflecting from a large area surface [14].

According to this separation, the statistical nature of specular paths could be characterized as deterministic and in reality each path (excluding the LoS) is usually a cluster of paths with similar angles and delays. Diffuse multipath tends towards a stochastic behaviour and might appear as increased random noise in the time-domain. In this study, the generic case of a combination between diffuse and specular components is considered, where the assumption that multipath propagation in wireless propagation channels can be modelled merely by deterministic paths is considered inadequate.

The channel's coherence time (13) constitutes the definitive measure for slow or fast multipath. For slow fading, the coherence time of the channel is large relative to the symbol period of the system. In this regime, the amplitude and phase variation introduced by the channel can be considered approximately constant. Slow fading is usually caused by shadowing where a large obstruction such as a large building diffracts the LoS signal or the signal is scattered by tree foliage. Therefore, specular components and shadowing could individually generate similarly structured waveforms at signal reception. In fast fading channels the opposite is expected where the coherence time of the channel is small relative to the symbol period of the system. It is noted that under deep fading circumstances the time-diversity of a fast-fading channel can prove as an advantage to recover parts of the information of the signal within the transmission time, while for a slow-fading channel it is not the case, and the signal information will be unrecoverable.

Finally, in GNSS, multipath caused by either diffuse or specular reflections can manifest in both carrier phase and code measurements [15]. The distinction between carrier phase and code multipath is useful mostly due to the difference in error magnitude. Maximum carrier phase error is frequency dependent and can reach 4.8 and 6.1 cm for GPS L1 and L2 respectively [16]. Code multipath on the other hand is in the order of magnitude of several meters. Moreover, the ionospheric channels are characterized by multiple propagation paths, resulting in echoes for pulse transmission and in selective fading for narrowband waveforms [2].

2.4 Multipath modelling in DS-CDMA systems

Typically, all wireless systems experience signal fading due to the propagation media and the induced multipath phenomena. Most communication receivers assume that they track accurately the time-varying channel parameters, and according to the derived relevant information they proceed to perform the decoding of the transmitted signal. Channel parameter estimates are afterwards updated in the receiver given the transmitted and received signal pair. If the receiver considers the channel characteristics known in the inherent signal processing it also assumes an additional white Gaussian noise. Then the estimation of the transmitted signal is also based on the white Gaussian noise model. The problem with this approach is that channel characteristics might not be reliably estimated given these input signals. Another questionable aspect is, if an accurate channel estimation is necessary to reliable communication [11]. In particular, wideband Code Division Multiple Access (CDMA) systems, might be prone to such problems due to their large bandwidth feature where information is spread across a "large" range of frequencies.

In digital radio communications multipath propagation introduces time dispersion, which can cause errors due to inter-symbol interference (ISI). ISI is defined as the interference between adjacent bits of the same transmitter, and in a full description of the Direct Sequence CDMA (DS-CDMA) model, multiple access interference (MAI) shall also be considered. MAI regards the interference between bits of different transmitters [17], [18] and it is usually minimal under system design. Next, the mathematical representation of the transmitted signal $\varkappa_{sat}$ for a generic transmitter sat is given [17]:

$$\varkappa_{sat}(t) = A_{sat} d_{sat} \grave{c}_{sat}(t - \epsilon T - \zeta_{sat}) \tag{15}$$

- ✓ A_{sat} is the signal power of satellite transmitter sat
- ✓ $d_{sat}(n_{bit})$ is the n[th] bit of the data transmitted from satellite sat
- ✓ ζ_{sat} is a constant hardware delay
  Note: $0 \leq t - (n_{bit} - 1)N_c - \tau_{sat} - l_{sat} \leq (T_c - 1)$
- ✓ $\grave{c}_{sat}$ is the spreading-code sequence (PRN code) transmitted from satellite sat with number of chips N_c per bit. In detail, $\grave{c}_{sat}(t) = c_{sat}\psi(t - \epsilon T_c)$, with c_{sat} the chip value, ϵ represents an integer number, and ψ is a spreading code waveform with duration T_c.
  - Timing variables are all multiples of the chip period:
    - τ_{sat} is the arbitrary delay of the signal sat which represents the bit-asynchronous nature of the system and it takes integer values $\{0, \dots, N_c - 1\}$, assuming the transmission is chip synchronous
    - l_{sat} is the path delay expressed as chip multiples $\{0, \dots, L_{sat}\}$

    Time variables usually correspond to chip rate sampling, although extension to arbitrary sampling can be straightforward to derive.

Received signal sampled at chip rate [19]:

$$r(\xi) = \sum_{i}^{N_{sat}} \sum_{l_{sat}=0}^{L_{sat}} A_{sat} \gamma_{sat l_{sat}} d_{sat}(n_{bit}) \grave{c}_{sat}(\xi - (n_{bit} - 1)N_c - \tau_{sat} - l_{sat}) \tag{16}$$

- ✓ $\gamma_{sat l_{sat}}$ are the $L_{sat} + 1$ path coefficients for the channel sat (+1 for the LoS signal)
- ✓ ξ is an integer number

2.5 GNSS pre-correlation signal model

By design, GNSS signals have complicated waveforms since they reflect spreading code and data bit modulations embedded in their structure (and in some cases, sub-carriers are also used). The useful component for the user is the one that carries system-related information and it is encoded through the data bits whereas the spreading code modulation serves for channel access utilization where in the majority of GNSS systems the DS-CDMA is adopted.

By employing different Direct-Sequence Spread Spectrum (DSSS) sequences (Figure 4) GNSS systems utilize multiple satellites to transmit signals at the same frequency at the same time. The user receiver is capable of separating these signals by exploiting the known copies for these DSSS PRN-modulated codes (with chip duration T_c). The method of several DSSS signals sharing a common carrier frequency, is referred to as Code Division Multiple Access (CDMA).

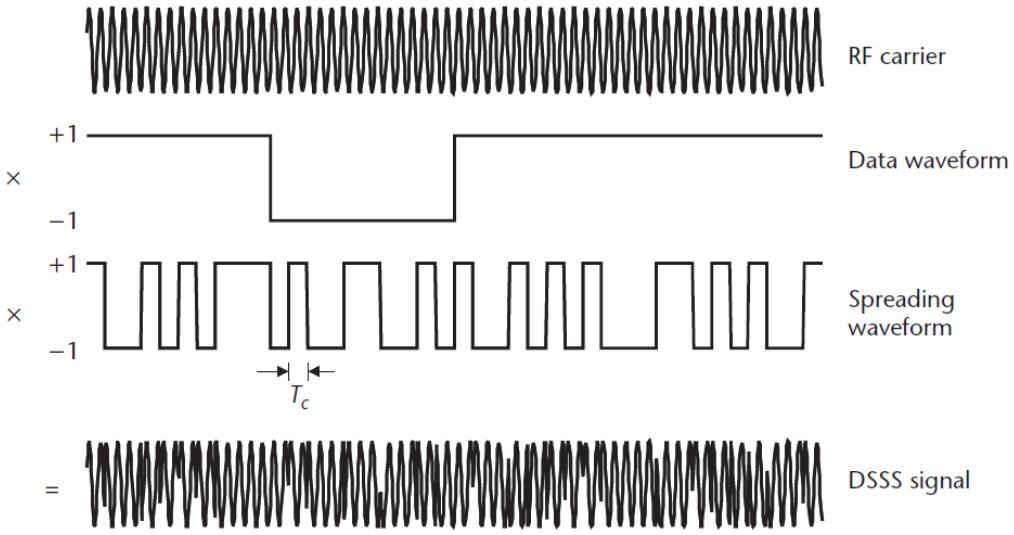


Figure 4 - DSSS signal modulation [20]

The signal carrier is a sinusoid generated by an oscillator and serves as the basis where all the above modulations are embedded. These signals are modulated in the time-domain to form an individual GNSS signal. A straightforward way to visualize the underlying frequencies is transforming the time-domain signal into the Fourier spectrum, i.e. the power spectral density (PSD) function, as shown in Figure 5.

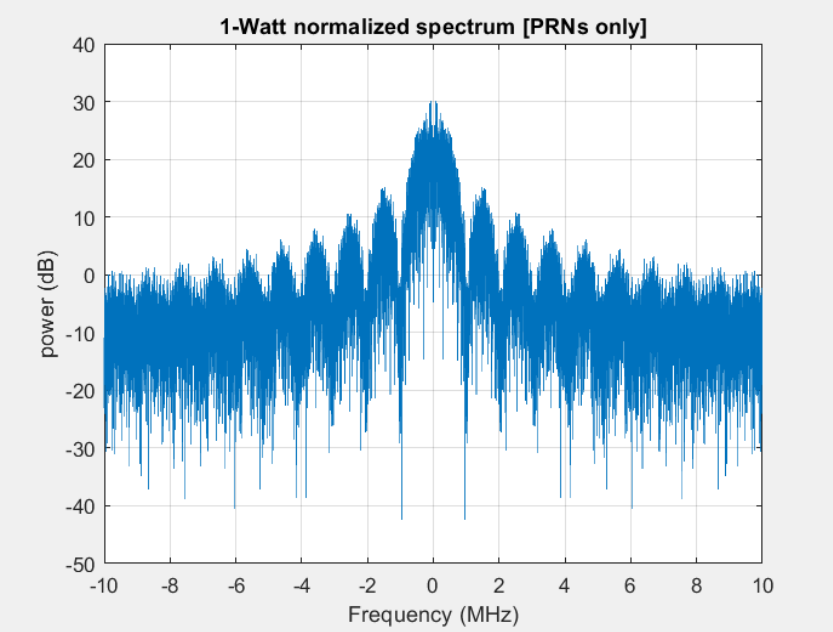


Figure 5 - Example PSD of an L1 C/A signal incident to a GNSS antenna from 7 different satellites sampled at 20 Mhz

The PSD shown in Figure 5 corresponds to a GPS L1 C/A simulated composite (RF) signal that includes PRNs 2, 4, 9, 11, 23, 27, and 31, sampled at 20 MHz (assuming IF of 4.13 MHz)

Moreover, the user antenna usually receives multiple GNSS signals simultaneously from different satellite transmitters, and the receiver requires a matched filter approach (cross correlation at spreading code level) in order to separate between them and proceed to the data decoding per satellite transmitter. Before getting to the latter stage of processing, the composite signal (of the different transmitted signals) is passing through the RF-frontend where filtering, amplification and down-conversion is applied (Figure 6). The down-conversion is not a necessary operation but it is included in most commercial GNSS receivers and it is usually implemented by heterodyne mixers. Heterodyning is a two-stage approach performing consecutive down-conversions.

A generic design and data flow of a GNSS antenna-receiver system is shown in Figure 6 where the antenna (ATX), Low-noise Amplifier (LNA), and down converter are still based on analog technology (RF chain). After the down-conversion step the digitization via the Analog-to-Digital Converter (ADC) takes place, and in some receiver designs an Automatic Gain Control (AGC) module is preceding the latter. In most receivers the addition of an AGC can benefit the processing of the ADC as it ensures that the signal-plus-noise magnitude is maintained within the required limits of the ADC detector. After the digitization of the signal from the ADC a bank of digital channels is built up performing acquisition and tracking.

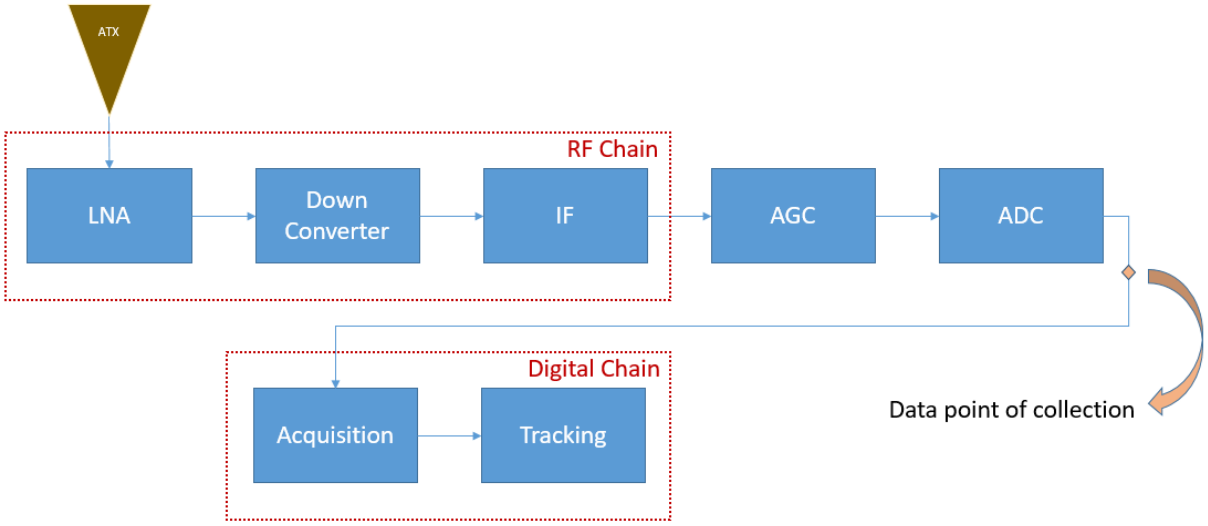


Figure 6 - GNSS receiver data flow showing the data-point of collection for the studied algorithms

As denoted in Figure 6 the data point of collection of the algorithms studied is essentially affected by the RF chain and ADC specifications. In particular, the RF chain presenting the relevant electronic analog components of a typical heterodyne GNSS frontend is shown in Figure 7.

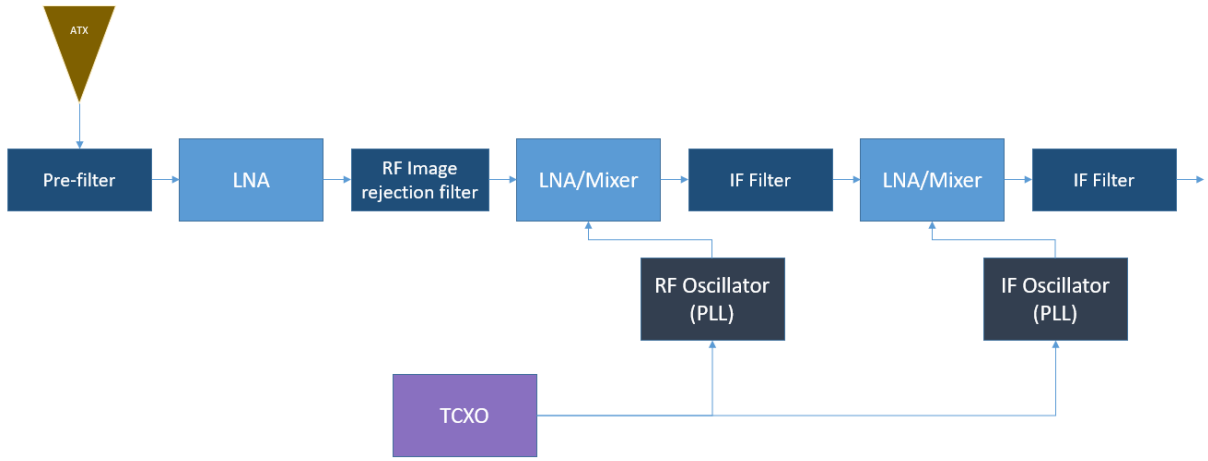


Figure 7 – A typical heterodyne RF frontend architecture of GNSS receivers [1]

The pre-filter after the signal reception from the antenna facilitates as a protection from possible spurious emissions and prepares a "cleaner" signal to be input to the LNA. The conditioning imposed by the pre-filter but also the amount of noise introduced by the LNA determine the quality of the observables in a definitive manner. After amplification the signal is usually filtered from image response and it is down-converted into hundreds of MHz. This intermediate frequency (IF) output is filtered by the following narrow-band filter to reject potential nuisance spectral components and another down-conversion takes place next. The IF signal then, is in the frequency range of a few MHz. As shown in Figure 7, the down conversion (mixing and amplification) process depends on the quality of the underlying oscillator which, for a generic frontend, a Temperature Compensated Crystal Oscillator (TCXO) might be used in combination with the corresponding Phase-Locked Loops (PLLs) in each down-conversion stage.

Mathematically, the model of the signal at the end of the receiver antenna is given as [1]:

$$x_{antenna}(t) = \sum_{i=1}^{N_{sat}} a_i e^{j\varphi_i} r_i(t-\tau_i) + v(t) \quad (17)$$

with a_i being the signal power attenuation factor for satellite i from N_{sat} visible satellites, ϕ_i is the phase delay and τ_i the signal propagation time from each satellite i to the receiver. Also, $r_i(t)$ represents the signal from satellite i and $v(t)$ is a generic noise component. Without considering any effects, the signal $s_i(t)$ is expected to follow the signal-in-space design for each GNSS system. After down conversion and then digitization by the ADC, a signal component of the digitized intermediate frequency (IF) received signal at the end of the RF-frontend can be rewritten in the discrete time domain as:

$$r_{IF,i}[t] = A_i \tilde{d}_i[t - \tau_i]\tilde{c}_i[t - \tau_i]\cos\left[2\pi\left(f_{IF} + f_{D_i}\right)t + \varphi_i\right] \tag{18}$$

Here, $\tilde{d}[t]$ represents the data sequence, $\tilde{c}_i[t]$ the spreading code sequence, f_{D_i} the carrier Doppler frequency shift due to the relative movement between receiver and satellite i, and A_i is the amplitude received and modified by the RF-frontend. The tilde above the data and spreading code parameters represents possible modifications introduced from the RF-frontend electronics.

In a more realistic scenario the reception of multiple reflected signals (also diffracted and deflected ones) from one GNSS transmitter are also included in this superposed signal due to the channel's geometry and material properties. In GNSS literature these effects are referred to as multipath propagation effects. It is noted that a wide variety of other influences might affect this waveform including hardware effects and interference from artificial sources, among others. The composite signal in discrete time-domain can be written as:

$$x[t] = s[t] + v[t] \tag{19}$$

with

$$s[t] = \sum_{i=1}^{N_{sat}} \check{r}_{IF,i}[t] \tag{20}$$

and $\check{r}_{IF,i}[t]$ being the multipath-afflicted version of the digitized IF received signal:

$$\check{r}_{IF,i}[t] = r_{IF,i}[t] + MP_i \tag{21}$$

Here the LoS signal $r_{IF,i}[t]$ shall have an associated multipath contribution that can be mathematically described from:

$$MP_i = A_i \sum_{n=1}^{N_{echoes,i}} a_{i_j}[t]\tilde{d}_i\left[t - \tau_i - \tau_j\right]\tilde{c}_i\left[t - \tau_i - \tau_j\right] \times \cos\left(2\pi f_{IF}t + \varphi_i + \Delta\varphi_{i,j} + \left(\Delta\omega_j - \Delta\omega_{IF}\right)t\right) \tag{22}$$

Therefore, the signal without the random noise is given next, and a schematic of the system models follows.

$$s[t] = \sum_{i=1}^{N_{sat}} r_{IF,i}[t] + \sum_{i=1}^{N_{sat}} MP_i \tag{23}$$

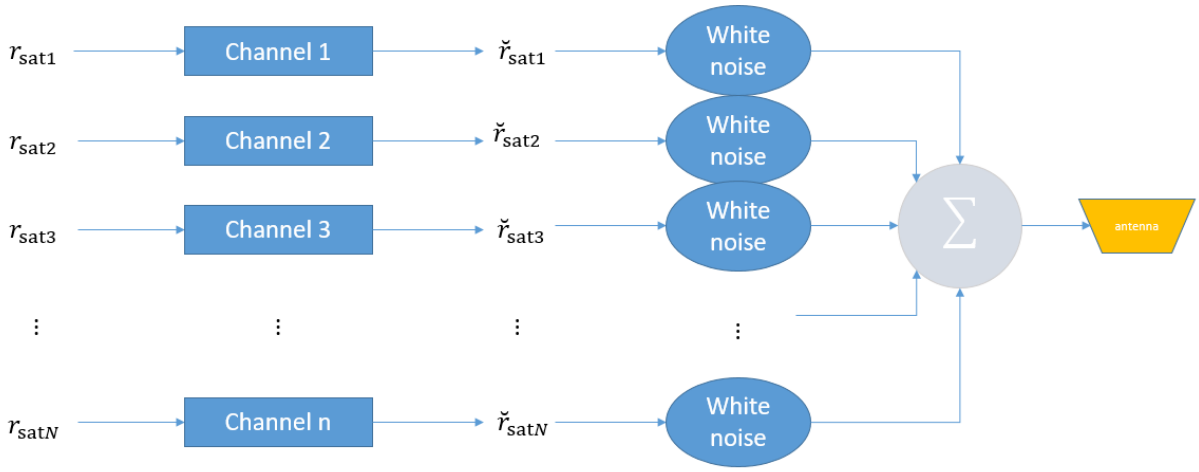


Figure 8 - Combination of signals upon reception in simulations

Therefore, the pre-correlation composite signal can be described from (Figure 8):

$$x[t] = \sum_{i=1}^{N_{sat}} r_{IF,i}[t] + \sum_{i=1}^{N_{sat}} MP_i + v[t] \quad (24)$$

with the random noise process v representing the thermal noise of the receiver.

An example of how the first (summation) term is given next. Here we simulate a waveform for 7 GPS L1 signals (PRNs 3,8,10,12,20,22,29) where the carrier amplitudes have been selected to be equal and having value 1 for simplicity. However, each carrier has different phase offsets as well as different Doppler frequency shifts to each other from the intermediate frequency. The intermediate frequency is selected as 4.1304 MHz and the sampling frequency is 20 MHz. Also, waveform duration is chosen to be 1 ms which is equal to the spreading code duration while the codephase of these spreading codes is randomly defined.

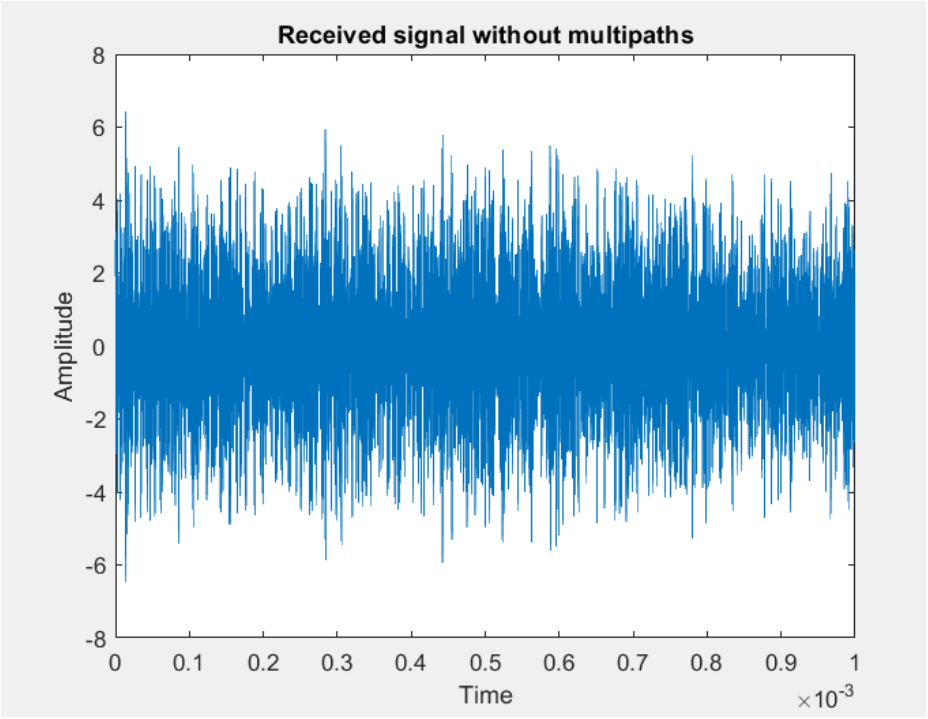


Figure 9 - An example of a GNSS signal waveform for 1ms duration received from an antenna/receiver system

Simulating one multipath echo for each transmitter and adding them on top of the previous term, the total signal $s[t]$ will look like the orange graph:

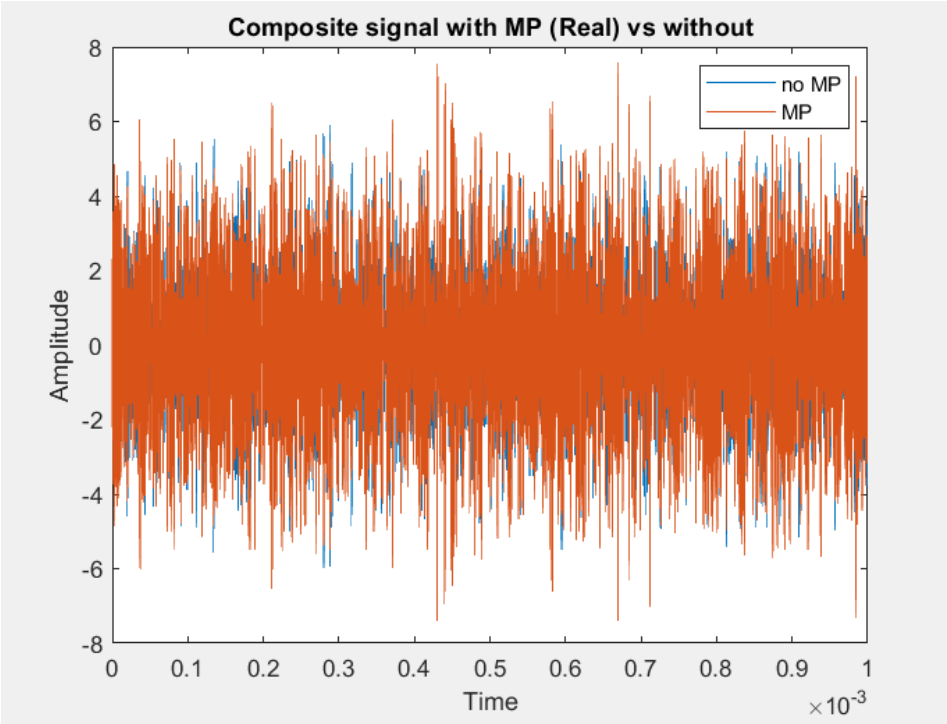


Figure 10 - An example of GNSS signal waveform for 1 ms duration both for the case of considering only SIS structure (no MP) and the same signal including multipath components

Then,

$$\sum_{i=1}^{N_{sat}} MP_i = s[t] - \sum_{i=1}^{N_{sat}} r_{IF,i}[t] \quad (25)$$

The "residual" MP signal would be:

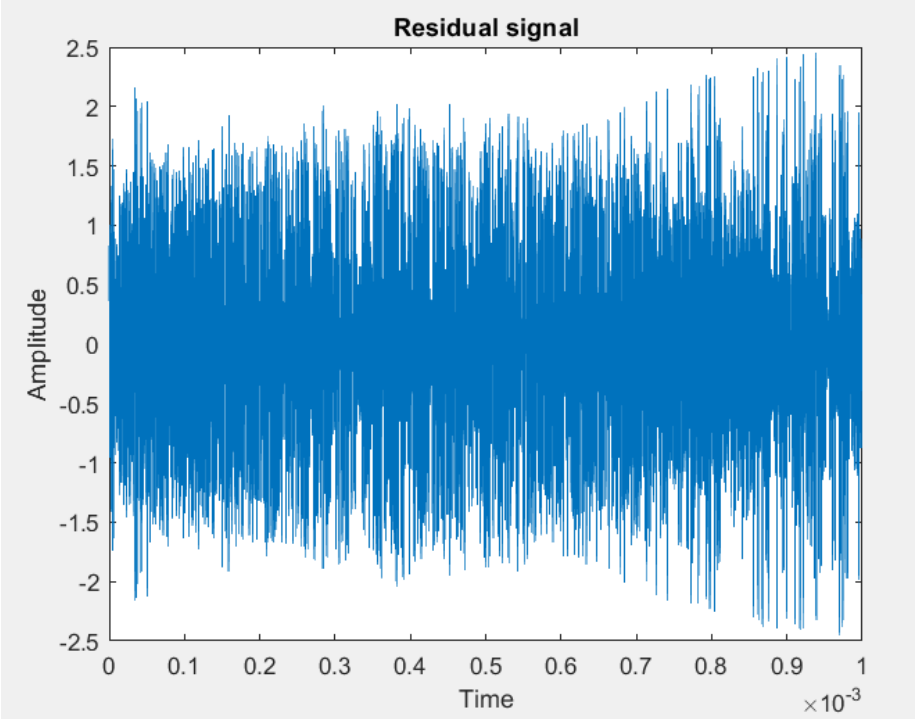


Figure 11 - The signal that is produced by considering only the MP contributions of Figure 6

The main goal is to exploit this “residual” MP signal from $s[t]$ to infer the multipath propagation environment that generated the final signal.

2.6 Empirical/Statistical models representing GNSS multipath

Fading channel models are used to describe the statistical behaviour of electromagnetic transmission of signals in wireless communication. Mathematically, fading is modelled as a time-varying random change in the amplitude $A_n\big(t, \tau_n(t)\big)$ and phase $\phi_n\big(t, \tau_n(t)\big)$ of the transmitted signal, given the delay $\tau_n(t)$, therefore, these parameters can be represented by stochastic processes. Depending on the signal propagation conditions a variety of models have been developed over the past years. Next, the most widely used statistical models are briefly reviewed, each sufficiently describing many possible scenarios under specific conditions.

Multipath propagation causes Rayleigh fading [21] and the standard statistical model of this gives a distribution known as the Rayleigh distribution. Usually, Rayleigh fading occurs in dispersive environments with several echoes each exposed to different delay, gain, and phase shift, and often constant. If a high amplitude (usually LoS) component is included in the mixture, the resulting distribution becomes Rician corresponding to so-called Rician fading conditions [22], [23], [24]. In fact, Rayleigh fading is a special case of Rician fading. Both Rayleigh and Rician distributions describe the statistical behaviour of the signal gains when propagating through dense environments that include numerous reflectors. Usually, either type of fading implies a frequency selective channel and the presence of inter-symbol interference. The random process of flat Rayleigh fading with a number of multipaths can be simulated with the sum-of-sinusoid method resulting in to Clarke‘s reference model [25], which in turn gives rise to the Jakes PSD (see [26]).

Although Rayleigh and Rice distributions are the most commonly applicable into indoors and urban telecommunication channels, several other statistical models exist (specific to the mobile satellite channel) such as the Norton [27], Nakagami [28], Loo [29], Lutz [30] and $\kappa - \mu$ [31] model. It is worth mentioning that a spoofing or meaconing attack can in principle cause multipath interference as well.

A thorough study covering all the related work on the statistical model for the indoor channel (NLoS conditions) is found in [32]. The current state-of-the-art of the indoor model was pioneered by Saleh and Valenzuela (S-V) [33] which showed some similarities to earlier work conducted to describe the urban canyon multipath behaviour [34]. In the latter reference authors showed that rays arrive at the receiver as Poisson-distributed events with independent Rayleigh-distributed amplitudes and uniform distributed phases. In the S-V model the major finding was that rays arrive at the receiver grouped in clusters and a second Poisson process was introduced to describe clusters’ times of arrival. Up to present, the S-V model remains the fundamental indoor channel model and it is generic to RF propagation not just for the GNSS bands.

One restriction of the S-V model is it describes a simple path for the multipath rays, meaning that the signal is not assumed to penetrate a number of walls [32]. Authors also mention an idea taken from the Lutz model (originally developed for outdoor multipath) where a classification of direct path, near echoes, and far echoes comes with a statistical characterization given for each part. Most work related to indoor channel characterization is pioneered mostly by the joint groups of IEEE 802 which have established several standards. The formulation of these standards is found on a comprehensive consideration of all the relevant physical aspects comprising the propagation channel. It appears that according to their studies, a model similar to the one propose by Lutz [30] covers most practical scenarios. Finally, the S-V model is still considered the baseline for all indoor related environments, and modifications could be used to improve accuracy. In particular, the physics of the indoor propagation channel are discussed in [35] where authors provide a detailed study that could aid in the development of more accurate models.

Depending on the surrounding environment of the receiver, models that accurately represent each (generalized) scenario have been proposed in the literature. For open sky conditions, the Weibull distribution [36] has been proposed to describe corresponding amplitude fading [37], [38]. For environments that shadowing takes place, amplitude change is often modelled using a log-normal distribution with a standard deviation according to the Log Distance Path Loss Model. Experimental results have been collected and characterized in [39] for no reflectors, giving rise to Log-Normal and Gamma distributions. The best models to represent fast multipath-fading are the Rician, Nakagami-m or Rayleigh, and in [39] an experimental verification and comparison according to corresponding detailed scenarios is included. The Nakagami-m distribution has also been proven to fit ionosphecic scintillation induced fading, a phenomenon causing the alteration of the signal amplitude and the velocity when the ionospheric refractive index changes. Parameters that affect scintillations include the strength, scale and location of the ionospheric irregularities, but also the signal frequency or the transmitter and receiver relative geometries. An excellent review on ionospheric scintillation models can be found in [40] with the most "preferred" ones being a bivariate distribution that would reduce to the product of a Nakagami-m distribution for the intensity [41], and a normal distribution for the phase in the limit of zero correlation between the two variates [42]. Performance results on employing these distributions can also be found in [43].

The suitability of each model has been assessed in [44] where a comparison of the 'goodness-of-fit' of Rayleigh, Nakagami, Weibull, Rician and Beta distributions based on the Kullback-Leibler (KL) divergence criteria. Another view on stochastic process modelling is provided by [45] which showed that the dynamics of the instantaneous power associated with each path can be modelled using mean-reverting Ornstein-Uhlenbeck processes, and higher order models. The authors introduce the concept of the state of the channel which corresponds to the solution of stochastic differential equations driven by white-noise (Brownian motion).

All models mentioned so far regard a single source signal. Considering the case of the RF GNSS signal where several signals from different transmitters might be superposed along with their corresponding echoes, the statistical modeling becomes more complex. An excellent introduction to parametric models that describe the sum of multiple fading signal components is given in [46]. Essentially, the resulting model is a combination of several random variables following Rice, Nakagami or Rayleigh stochastic models with deterministic parameters. An even more complete treatment on the subject that incorporates the Gaussian noise is discussed in [47]. In general, notable research has been performed in this area [48], [49], [50], [51], [52], [53], but the main drawbacks of estimating a parametric model such as the ones described in the references is the complexity and the lack of closed form solution.

3 State-of-the-art in detection, characterization and mitigation of GNSS interference

3.1 Introduction

An up-to-date and complete study of the vulnerabilities of GNSS systems is presented in [54], and a recent review of the relevant literature is compiled in [55]. Given these surveys a few receiver-autonomous architectures have been presented that combine detectors and countermeasures for most prominent GNSS threats, as for instance [56] and [57]. The two latter studies follow two different paths to address "multi-threat" detection and mitigation. In [56] several receiver architecture stages of a vector tracking loop (and the corresponding outputs) are adapted by a dedicated threat detection-mitigation algorithm each mitigating multipath, RFI and scintillation events. In [57] the algorithm exploit outputs from two stages: the post-correlation signal monitoring quality metrics and signal power measurements demonstrating high performance in distinguishing low-power spoofing from nominal multipath. It is also capable of providing information on the presence of any other type of RFI source.

Another interesting direction (and the one followed in this study) for detecting GNSS threats can be viewed from a context detection perspective [58]. For instance, in [59] authors presented a machine learning algorithm that classifies different types of multipath environments using direct correlator outputs (thus employing post-correlation observables). An almost exhaustive survey of the methods used to detect and characterize GNSS channel threats is [60]. As pre-correlation detectors are the focus of this research, several pre-despreading detection methods can be found in [61] and [62] and the references therein. The latter two references include algorithms that address mostly RFI-related threats and it is worth mentioning that pre-correlation multipath detectors (and/or classifiers) are rare in literature. A few notable references on multipath detectors are [63], [64] and [65], [66] for multipath channel estimation (but not restricted to GNSS applications).

RFI is one of the most feared events compromising the functionalities of a GNSS receiver and the corresponding user-level applications dependent on it [67], [68]. RFI can be coarsely categorized to unintentional and intentional with the latter further separated to jamming and spoofing. A study on different types of interference characteristics can be found [69]. Jammer signal characteristics have been studied in detail in [70], [71] and [72]. Different approaches to jammer's detection have been investigated such as the detection of the start of jamming event/ transient signal characteristics [38], Carrier-to-Noise Density Ratio [69] tests and Structural Power Content Analysis in [62]. Also, a review on specific types of GNSS jammers and their impact on low-cost GNSS receivers is presented in [73]. These types of interference show remote similarity to the auto-correlation behaviour produced my multipath and therefore these approaches are not studied further. However, there is a particular type of RFI that resembles the characteristics of multipath interference, that is a spoofing attack.

The main similarity between spoofing and multipath interference is that the received spoofing interference is statistically correlated with the authentic signal. However, multipath-induced structured interference is not intentional while spoofing usually involves an attacker who can adjust signal parameters such as signal power, code phase, carrier phase, and signal structure for deceiving the signal processing of the user receiver.

Detailed surveys on spoofing attack characteristics can be found in [74] and [75], as well as in [76] but also in [77] where authors are focusing more on GNSS vulnerabilities specific to spoofing. The impact of spoofing on various applications is discussed in [78]. A relatively recent study on spoofing impact on smartphones [79] showed that a simplistic spoofing attack is not fully successful in open-sky

conditions. All these works can prove a valuable resource in defining signature features for specific threat detection. A survey on spoofing detection SDR solutions [80] but also the proposal and test of an SVM-based detector can be found in [81] and [82].

From the available literature on pre-correlation threat detection and characterization it is evident that relevant research is limited. Most threat detectors involve binary characterization on whether the threat is present or not, for instance for an RFI source or an ionospheric scintillation event. On the other hand, the vast majority of excessive multipath detectors use as inputs post-correlation observables, i.e. sequences per satellite signal. The main reason for avoiding to use the superposition of GNSS signals in pre-correlation level (RF signal) is feasibility on real-time applications and the reliability of the result. This is because the estimation of the channel parameters such as amplitudes and code delays of multipath or non-line-of-sight signals requires very high sampling rates (increased memory requirement) and in consequence, a powerful processor to provide a fast solution; but also a very complicated model.

3.2 Multipath detection methods

DSP methods are usually distinguished in GNSS literature according to the applicable processing stage in the receiver. As described previously, the great majority of GNSS transmitter signals are modulated according to the DSSS technique and follow the CDMA property where multiple transmitters share the same channel. The receiver employing its saved copy of the spreading sequence can then distinguish each transmitter signal through the cross-correlation operation (and specify if the signal is received). Observations available before or after correlation give rise to pre-correlation or post-correlation DSP algorithms respectively. Here, we are interested in studying methods that operate to the former category.

In the GNSS-RFI literature detection techniques are usually separated into time-domain, frequency domain, and statistical-domain. The spectrogram method can also be considered as another distinct class, providing a representation of the frequency spectrum of a signal as it varies with time (a time-frequency method of analysis). A big percentage of pre-correlation GNSS threat signal detection methods involve statistical hypothesis testing and require the construction of a test statistic that employs specific observables [39]. In time domain the observable is usually the power content, in the frequency domain the power spectral density, and in the statistical domain the observed time-series distribution. Binary or multiple hypothesis testing is a powerful tool that enables detection and classification of signals in two or several classes, respectively. Prime requirement in most cases is the prior knowledge of the functional form of the distribution of the data, free from other influences.

One of the very few tools available - not tested on pre-correlation GNSS data - and specifically designed to analyse periodic structures in frequency spectra is referred to as Cepstrum. Its name is an anagram of the spectrum and it is computed as the inverse Fourier transform of the logarithm of the estimated signal spectrum (Fourier transform) [83]. It is a time domain representation and it is the result of three steps:

1. Transformation of a signal from the time domain to the Fourier domain,

2. Computation of the logarithm of the spectral amplitude,

3. Transformation to “quefrency” domain, where the final independent variable, the quefrency has a time scale.

Depending on whether the initial signal is in complex form or it is a power time-series it computes the complex or power cepstrum, respectively.

Frequency domain techniques are generally based on spectral estimation [84], [85], [86] of the incoming signal and their performance relies on time-frequency resolution which is essential in processing time-varying signals. For instance in [87] it is shown that long observation windows enables detection of low power RFI (structured interference). Frequency domain techniques typically compare the spectrum (PSD) of the received signal with a theoretical threshold which is usually determined according to a statistical model representing the received signal. Possibly the lack of related works using the frequency domain techniques for the multipath is that it does not demonstrated a clear pattern according to a strong frequency component present (in the spectral density it would be expected as a spike).

Statistical techniques monitor the input signal statistical distribution. The main assumption in these techniques is that in the absence of any interference the incoming signal is approximately Gaussian because both the thermal noise and the set of received signals are well described by Gaussian distributions. In the presence of sufficiently strong interference the distribution of the received signal is often non-Gaussian. Gaussianity tests in time-series collected by the receiver can be applied both in the time-domain and frequency-domain [88]. Although these goodness-of-fit tests are frequently used, they were developed for independent and identically distributed time-series and have been shown to perform

poorly for dependent variables. This is mainly the reason that these detection methods are not suitable for multipath interference.

One of the first references of statistical methods dealing with correlated/colored data can be found in [89] and later in [90]. The main drawback of such Gaussianity test methods is that they require a sufficient period where interference is present to reliably estimate the Probability Density Function (PDF). It has been shown that for highly dynamic RFI sources as well as for wideband interferers such as spoofers, such methods are likely to fail [60]. An example of statistical techniques in GNSS interference context is [91] where information from second order statistics is deployed to perform a goodness-of-fit test. Higher order statistics have also been used to detect non/Gaussianities such as the kurtosis method using fourth-order cumulants in [92], as well as references [93], [94] which are not restrictive to GNSS.

In cases where the a priori probability distribution function is not known - for instance for the reception of a weak signal - further statistical concepts must be introduced such as locally optimal tests [95]. GNSS signal changes greatly from time to time but it is also uncertain, which means it is difficult to build exact models or prescribe an exact threshold. Numerous pre-correlation interference detection methods relying on hypothesis testing and/or setting decision threshold that do not require non-interference GNSS signal model have been proposed, for instance [96], [97], [98], [99], [100], [101], [102], [103], [104]. These methods are strongly linked to the non-parametric measures of the Receiver Operating Characteristic (ROC) curve. The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

An evaluation of the detection performance of several commercial pre-correlation interference detectors can be found in [105]. The evaluations are based on actual measurements of GPS signals and different types of jamming signals. In this survey a method of detection on automatic gain control (AGC) level is also included. In the presence of spoofing, received signals show an increase on their power content, leading to changes in the AGC level. However, the AGC gain can be disrupted by various interfering signals without necessarily distinguishing between them [62].

Furthermore, in [106] the performance of four algorithms was compared (pulse, peak-picking, single channel kurtosis, and spectral kurtosis methods) for detecting sinusoidal pulsed interference in microwave radiometer systems. Another work related to pre-correlation interference detection is presented in [107] where authors built an algorithm based on four distinct steps: 1) Setting Short-time Fourier-transform parameters to obtain the spectrograms of the radiometer output signal; 2) Insert the spectrograms of the radiometer output signal into a power-law detection algorithm [108]; 3) Determine the detection threshold given by the false alarm rate; 4) Obtain the detection result by comparing the power-law detector output with the threshold. Again, unless there is a very fine-tuning of detection thresholds, involved multipath effects might be too subtle to be detected. Other examples include the more sophisticated Bayesian M-ary hypothesis testing [95] where the major drawback - as analysed by [109] - regards the fact that as the number of tests increases the number of ‘false findings’ tends to increase as well. Another alternative of detections’ fusion approach is presented in [110] which is based on Dempster-Shafer theory of belief functions [111].

Finally, there are very few published works that make use of statistical methods in pre-correlator GNSS signal processing for filtering purposes. A direct application of singular value decomposition (SVD) on a data observation matrix composed of a number of realizations of the signal vector is presented in [112]. As the authors state this type of SVD-based signal detection is widely used in cognitive radio networks, but here, it is employed for RFI signal detection. Also, in [113] a PCA-based interference detection scheme (Karhunen–Loeve Transform) is presented where the statistical representation of the signals makes the discrimination of unwanted RFI easier.

3.3 Multipath mitigation and filtering techniques overview

In real conditions and depending on the signal path there is a significant probability that the channel will experience a deep fade. Multipath propagation is the main cause of signal fading and poor performance in traditional communication systems, therefore its mitigation is of great importance to any wireless receiver (GNSS or not). There have been several different approaches toward multipath reduction and/or filtering by introducing a different type of domain-diversity. For instance, multipath effects can be mitigated by using diversity: transmitting the signal over multiple channels (composite channel). This way each channel experiences independent fading and signals are coherently combined at the receiver, yielding a low probability of fading since it is proportional to the probability of a simultaneous fade of all component channels. Diversity can be achieved in time, frequency, or space domains and most common techniques include diversity reception and transmission, the use of orthogonal frequency-division multiplexing (OFDM) in transmission, the use of so-called Rake receivers (i.e. multiple correlators and analysis per visible signal), the standard matched filtering [2] technique, but also the use of space–time codes, and multiple-input and multiple-output (MIMO) method.

Specifically for GNSS, Betz [114] designed the binary offset carrier (BOC) modulation to enhance robustness to multipath interference with respect to the binary phase-shift keying (BPSK), which was widespread in GNSS systems. The main idea behind BOC modulation is to reduce the interference with BPSK-modulated signal, which has a *sinc* function shaped spectrum. Variations to this approach exist, see for instance [115]. Near-far interference in multiple-access communications can also be reduced by multiuser detection methods [116]. However, implementation of such a detection scheme requires that the amplitudes and delays of the incoming waveforms to be known a priori, or jointly estimated [117]. Finally, a comprehensive survey of the dense/diffuse multipath filtering methods is given in [118].

Before closing this paragraph, an alternative case [119] is included that has been applied on narrowband interference suppression for GPS navigation using a neural-network-based filter. This is an example of work that relies on machine learning algorithms to perform interference detection and shows a potential for improved detection sensitivity that could potentially be applicable to multipath as well. Machine learning algorithms have been increasingly showing their capacity to perform more accurate detection tasks and they could prove very useful tools in mitigating the complex effect of multipath propagation.

3.4 Multipath estimation techniques overview

This paragraph is focused on established techniques used to perform multipath channel estimation. Channel estimation and channel characterization are two very closely related terms, since the estimation of the parameters could in principle also lead to a characterization when compared with templates. The overview is performed in spite of the fact that in a GNSS receiver the multipath features of the propagation channel might be very subtle or could possibly disappear in the noise floor [120]. This is the reason that "mapping of the local environment" or "context awareness" or channel characterization, is a very demanding task, and ways to exploit the maximum information available are the main goal of this research.

Multipath parameters such as the direction of arrival (DOA) are considered out of scope as the focus here is on time-diversity for a single antenna application. The latter restriction also implies that spatial-diversity techniques are also excluded from this review as well. In particular, the estimation methods used for the (complex) amplitude and delay multipath channel parameters would only be considered, and there have been few research publications considered relevant to the objectives of the thesis. Notwithstanding this general direction followed, it is worth referring briefly to the SAGE (Space Alternating Generalized Expectation Maximization) algorithm that has been applied for channel estimation in DS-CDMA communication systems. SAGE has been used in estimating the propagation time-delay of the LoS signal in a GNSS receiver under the presence of severe multipath (SMR = 5dB) [121] but also tested under a Kalman filter approach through simulations with 2 GPS C/A signals [120]. It assumes the use of a sensor array and the estimation parameters comprise of complex amplitude, delay, Doppler frequency, incident azimuth and incident elevation of all multipath signals. The power of SAGE lies in estimating the parameters of each multipath component sequentially, in contrast to typical Expectation Maximization that computes those in parallel in one iteration step.

An example of estimating the complex amplitude and delay by using OFDM signals is described in [122] where authors propose the exploitation a large number of known pilot symbols. The possibility to adapt this concept to GNSS signals could theoretically be feasible by taking into account the trade-off between the data rate and the multipath channel estimation performance, but it would involve re-designing system parameters. A parametric multipath estimation method in the frequency domain is found in [123] but only for BOC signals. A different example of deploying the multipath for mapping purposes regards 5G communication signals and mmWave propagation [124].

A different approach to signal cross-correlation performed by all GNSS receivers could be regarded as a pre-correlation method from the perspective that it does not require a cross-correlation operation. A data-driven method is presented in [125] where the standard correlator scheme is replaced with a deep neural network architecture, aimed to provide enhanced robustness to the receiver in challenging environments. The method shows that it can learn the complex characteristics of the GNSS satellite-to-land channel, however, it is observed that the investigated Neural Network model requires a multi-correlation scheme thus involving an increased computational cost.

A potential route to perform channel estimation for each different satellite signal visible from the user antenna on the same time could be given by Blind Source Separation (BSS) techniques. Exploiting the orthogonality feature of the signals originating from different transmitters due to the spreading code sequence, CDMA signals can be theoretically separated when the sampling rate of the signal is equal to the chipping rate. The study of the use of BSS approaches to maximize output signal-to-interference ratio (SIR) for recovering source signals from observed signals (without prior knowledge of a mixing matrix or the source signals, i.e. unsupervised), is investigated. In the simplest case, received signals in a CDMA receiver can be considered as signals generated by the linear convolutive model of statistically

independent components of independent users [126]. However, there are several challenges before any BSS algorithm can be used for the GNSS RF signal.

First challenge in the pre-correlator stage of the receiver is that the number of satellite signals visible is unknown, and the great majority of BSS methods rely on the a-priori knowledge of the number of source signals in order to perform the separation. Second problem is that they expect a number of synchronous signals, not a single stream of data. One of rare examples in literature of direct decomposition of the single channel data to independent components is studied in [127] where Non-negative Matrix Factorization (NMF) methods are used. Another difficulty regards the fact that they work with statistically independent source signals, and although the satellite signals are quasi-orthogonal, each one (or a subset) of them in "accompanied" by one or many delayed replica, therefore the performance of the algorithms might be sub-optimal. Also, since BSS algorithms rely on a known number of source signals they are distinguished in overdetermined (if the number of sensors exceeds the number of signals source) or underdetermined optimization which limits their capacity to be adaptive.

Solutions to each of the above-mentioned challenges have been proposed. The determination of the number of source signals can be resolved by using a blind multiuser detector [128]. Another approach is to initialize the Independent Component Analysis (ICA) algorithm by first performing a clustering step to derive an initial number of sources estimate [129]. Alternative methods for estimating the number of sources in instantaneous and anechoic mixing systems without the knowledge of the system parameters is proposed in [130] and [131]. The term 'anechoic mixing' is often used in acoustics' literature and it refers to observations of different sources made from different sensors with but with no multipath echoes, only direct signals.

The problem of separating a mixture of signals from a single sensor/channel is rather common in many different fields such as in communication and biomedical signal processing applications [132], [133], [134], [135], [136], [137], [135][138], [139], [140] as well as speech and audio processing [141], [127], [142], [143], [144], [145], bioinformatics-chemometrics-environmetrics-image processing (see [127] and references therein), and telecommunications [146]. The common practice in all these works is the use of a signal decomposition step which results to a multi-dimensional expansion of the single-channel data, and then the derivative mode functions are used to feed a standard BSS algorithm. A superior performance in detecting and filtering RFI threats in the GNSS context has also been reported [60] by using this approach. Another advantage of employing decomposition methods is that it could resolve the problem of over- or under-determined optimization, given that the decomposition result always provides more mode functions than the source/satellite signals.

Finally, as highlighted in [145], most ICA algorithms are proven to be consistent even when considering statistically dependent data. In fact, blind separation of convolved mixtures [147](Ch.19.2) is particularly challenging considering that observations are the mixtures of latent components as well as their delayed versions, thus the unknown parameters increase significantly. However, there are research works which propose algorithms for convolutive BSS such as [148], [149], [150], [151], [152], [153], [154]. Other methods include Principal Components Combining for dense correlated multipath fading environments [155], [156], and the use of a deep Recurrent Neural Network [157], [158].

3.5 Summary and discussion

In this chapter an overview of the techniques used to analyse and mitigate the multipath effect in GNSS signals is presented. Most methods appear to avoid operating on the raw data directly, but as a first step they transform the data into feature space to isolate multipath effects from the useful data. In the pattern recognition terminology, this step is usually referred to as feature extraction, and it commonly takes place before the detection/filtering/estimation algorithm. The different methods performing feature extraction is GNSS signals per domain are summarized:

- Statistical-domain methods; for instance, based on optimization criteria of mutual information, entropy and kurtosis.
- Frequency-domain methods; based on spectral transformations such as Fourier, Wavelet, or adaptive basis transformations and the corresponding spectral representation.
- Time-Domain methods
- A combination of time and frequency-domain methods named time-frequency representations such as 3D representations (power-time-frequency). This family of approaches could be seen as an image of contours (2D) but since there is a grid (defined number of points with certain time-frequency coordinates), the resolution plays an important role.

In conclusion, the main objective of the feature extraction step is the transformation of the data to a domain that the properties of underlying processes are manifested as distinctly as possible. An accurate and reliable feature vector can significantly improve the performance of the estimator that follows. In the next chapter several feature extraction methods with the potential to capture the essential properties of the multipath effect in the time domain GNSS signals are presented. The goal of these algorithms is to provide the input to the multipath environment classifier.

4 Feature extraction techniques for GNSS multipath

4.1 Introduction

This chapter includes the theoretical investigation and numerical simulations on existing feature extraction methods and/or analysis tools. A selection of methods suitable for capturing essential properties of multipath patterns in the pre-correlation waveform of a GNSS receiver was performed according to numerical simulation tests relying on GPS L1 C/A signal characteristics. Feature extraction is a critical step in the methodology followed as the algorithmic steps that precede classification depend on the quality and the "separability" of the features generated during this stage.

In challenging single-channel signal analysis, it is custom to use a multi-dimensional expansion of the signal into several time-domain components to create features in multiple scales and/or frequencies by using signal decomposition methods. This process is usually characterized as a feature creation step that generates features on each component, and could be seen as a pre-processing and/or filtering step before the final feature extraction. Examples of multi-dimensional expansions are Wavelet Decompositions, Empirical Mode Decomposition (EMD) [159], [160] and its variants [161], and Singular Spectrum Analysis (SSA) [162], [163]. Then, feature extraction methods in the statistical and/or time-domain are usually applied to each component/mode generated.

A special mention to the importance of non-linear behaviour of the multipath phenomenon in wireless communication channels [164] shall be made at this point as an extra consideration in the selection of the appropriate analysis tool/s. Without delving deeper to the detail of each method, time-frequency representations and signal decomposition methods such as the Windowed Fourier-based method and the Wavelet Transform are built on the additive principle, and theoretically, might not be suitable to analyse multipath behaviour. Decomposition methods such as EMD and SSA might be more appropriate or more advanced tools such as Holo-Hilbert Spectral Analysis [165] that is reported to be able to extract not only linear additive information. In conclusion, the selection of the feature extraction approach shall not only be able to accommodate for the non-linearities of GNSS multipath but also maintain a low algorithmic complexity. Further details on the criteria considered for the methodology followed in this research are given in the next paragraph.

4.2 Feature extraction methods

This section reviews and assesses methods potentially suitable to model multipath afflicted power-fluctuation sequences generated at the end of the ADC in the receiver GNSS RF-frontend. In the simplest terms, and only by considering the GNSS signal design and the thermal noise of the receiver, the waveform under study (GNSS signal) is a Wide Sense Stationary (WSS) process [102]. However, the main assumption made for the rest of this paragraph is that generic pre-correlation data sequences - given that multipath interference is present - can be modelled as autoregressive process AR(k) of order k and the WSS model does not hold.

The reason of selecting an AR(k) model is because the signal under study could be described as a superposition of self-similar processes due to the repetitive structure of spreading code sequences when received from several paths with different delays. A model could be built using a linear combination of an arbitrary number of Markov (autoregressive) processes if parametric estimation was the objective. More in detail, the RF sequence contains one to several transmitter signals and due to the quasi-orthogonality of the spreading code sequences, interference signatures (constructive-destructive interference) might not be evident since they will "cancel-out" within a spreading code duration. However, if at least one multipath echo (delayed copy) is also "added" in the RF sequence for each transmitter then some constructive-destructive interference pattern might arise.

Given the autocorrelated random processes assumption holds, the criteria for adopting feature creation methods are (to be):

- Suitable for capturing the properties of crossover phenomena and in particular to be able to distinguish between autocorrelated random processes and (uniform) random noise
- One-value (scalar) metric describing the characteristics of the total time-series/waveform (for instance the PSD is a power-frequency function, not a scalar value)

A list and the corresponding description of candidate methods for the feature extraction block of the feature-based classification approach (Figure 12), is presented. Candidates presented next include the Hurst Exponent, Detrended Fluctuation Analysis, Correlation Dimension, Fuzzy Entropy and Recurrence Period Density Entropy. It is noted that multi-scale features could theoretically enhance the performance of the feature extraction by adding more feature observables per measurement. Representative examples of decompositions in individual bands using one sample in feature-based classification but regarding different applications can be found in [166], [167] and [168]. However, this direction will not be explored further in this work.

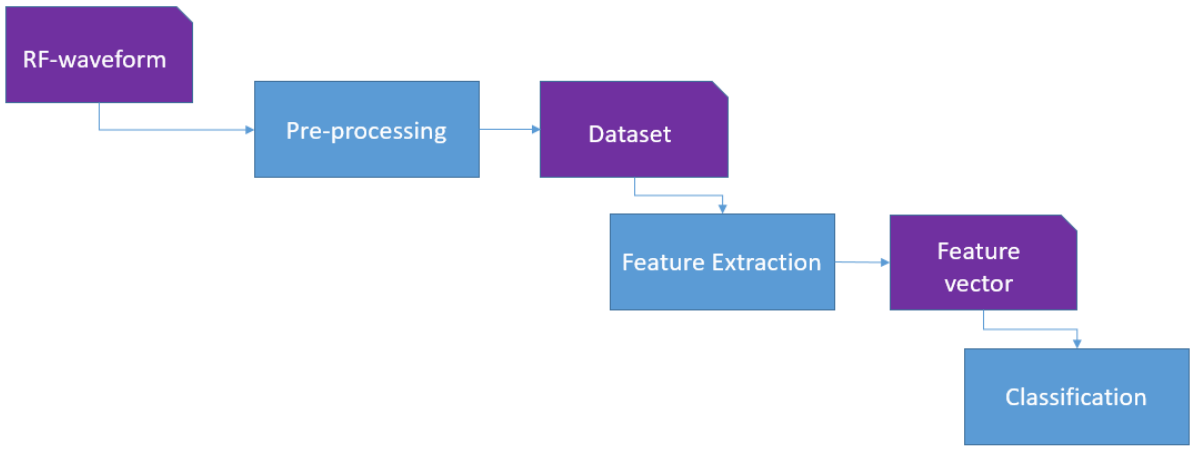


Figure 12 - Feature-based classification building blocks and data flow

4.2.1 Hurst Exponent

The Hurst exponent (HE) was first introduced in statistical hydrology by Harold Edwin Hurst [169], [170], and it quantifies the autocorrelations of time series and corresponding scaling properties. Originally it was developed as a metric to describe linear combination of (independent) autoregressive processes of order one. Mandelbrot [171] established its direct relation to the fractal dimension in fractal geometry, usually conceptualized as a measure of the "intensity" of randomness. In fractal geometry a fractal is a geometric shape that is self-similar and has fractional dimensions.

The Hurst method yields an exponent H, which results in a fractal dimension $D_F = 2 - H$ and it is defined as:

$$H = \frac{\log(R/S)}{\log T_{interval}} \tag{26}$$

with $T_{interval}$ the duration of the sample data and R/S the rescaled range which quantifies the amount of divergence of time series, defined as "the range of the mean-centered values for a given duration $T_{interval}$ divided by the standard deviation for that duration".

The Hurst exponent $H = 0.5$ corresponds to a completely random system such the Brownian motion (i.e. a random walk process), and $R = {T_{interval}}^{1/2}$, where R is the distance covered by the particle in time $T_{interval}$. For $H = 1$ the time series indicate a completely deterministic system where autocorrelations dominate.

The original algorithm [170] to estimate Hurst exponent H and the fractal (fractional) dimension of a time series was the rescaled range statistical analysis (RSA). RSA is based on differences between the maximum and the minimum cumulative values and the standard deviation from the observed values. Considering a time-series $\beta_1, \dots, \beta_Z$ the steps for RSA are:

1. *Compute the mean $\hat{\beta}$ and the standard deviation S of the series*
2. *Compute rescaled range time series by subtracting the sample mean from the original series $S_k = \beta_k - \hat{\beta}$ for $1 \leq k \leq Z$, S_k is zero-mean.*
3. *Compute the cumulative time series $\Gamma_k = \sum_{i=1}^{k} S_k$ for $1 \leq k \leq Z$*
4. *Calculate the adjusted range R_k, i.e. the difference between the maximum value and the minimum value of Γ_k: $R = \max_{1 \leq k \leq Z} \Gamma_k - \min_{1 \leq k \leq N} \Gamma_k$ (over all Γ_k)*
5. *Then compute the rescaled range R/S*
6. *Next, split the Z sample time series into two $Z/2$ samples, repeat steps 1-5 for each sample and compute the average rescaled range R/S*
7. *Repeat for successively smaller intervals over the data set, dividing each segment obtained in each step in two and calculating R/S for each segment and finally computing the average R/S.*
8. *Plot the values of $log(R/S)$ against $log(Z)$ for all the different sample sizes.*
9. *Compute the slope of the best fitting line (for instance using least squares regression) which gives the estimate of the Hurst exponent H.*

Several methods for estimating H exist [172], [173] based on various estimators of variance, Fourier Spectra (Whittle, Periodogram), Zero-Level-Crossings of paths, quadratic variations, convex rearrangements, and also wavelet-based estimation [174].

The Hurst exponent can only take values between 0 and 1, and a value with $H > 0.5$ indicates persistent long memory (or long-term positive autocorrelation) in the time series, while $H < 0.5$ is interpreted as evidence of anti-correlated time series, and $H = 0.5$ reflects a random white-noise (uncorrelated) time

series [175]. For the values $0.5 < H < 1$ and $0 < H < 0.5$ long-range correlations typically decrease as a power law.

An application regarding channel estimation in mobile radio communication has been reported [176] where the proposed algorithm can completely characterize the channel by estimating solely the Hurst exponent. It has also been employed in functional magnetic resonance imaging (fMRI) time series [175], [177]. Several alternative approaches/extension of the Hurst exponent have been proposed to address some of the drawbacks of the Hurst RSA approach for its poor performance in the presence of short memory, heteroskedasticity, and multiple scale behaviours, as well as vulnerability to outliers [178].

The Generalized Hurst Exponent (GHE) method was introduced in [179], [178] while applied to financial time series data. This method involves the computation of the q-moments of the distribution of the increments. The q-moments provide a direct measure for the scaling properties of the time series and add robustness when outliers are present compared to the adjusted range R_k (step 4 of the original R/S algorithm) used from the RSA method. The basic element for the computation of the GHE is the quantity $K_q(\tau_{interval})$:

$$K_q(\tau_{interval}) = \frac{\langle|\mathrm{B}(t-\tau_{interval})-\mathrm{B}(t)|^q\rangle}{\langle \mathrm{B}(t)\rangle} \quad (27)$$

Here $\mathrm{B} = \beta_1, \dots, \beta_Z$, and $\tau_{interval}$ is the time-interval that can vary between the time resolution/sampling rate ϱ $(t = \varrho, 2\varrho, \dots, k\varrho, \dots T_{interval})$ and a maximum value $\tau_{interval\_max}$. Different exponents q represent the different multi-scale characteristics of the series and, in particular for $q = 2$, the $K_q(\tau_{interval})$ is proportional to the autocorrelation function.

The GHE measure denoted as $H(q)$ can be defined from the scaling behaviour of $K_q(\tau_{interval})$ [180], as :

$$K_q(\tau_{interval}) \sim \left(\frac{\tau_{interval}}{\varrho}\right)^{qH(q)} \quad (28)$$

The Hurst exponent (25) is closely related to (27) when $q = 1$, i.e. $H(1)$ as both metrics quantify the scaling of the absolute spread in the increments. The exponent at $q = 2$ is describing the scaling of the autocorrelation function $a(\tau_{interval})$ as $K_q(\tau_{interval}) \propto a(\tau_{interval})$, $a(\tau_{interval}) = \langle \mathrm{B}(t + \tau_{interval})\mathrm{B}(t)\rangle$) which is related to the power spectrum [181]. Here, $H(q) = 0$ corresponds to $1/f$ pink noise.

4.2.2 Detrended Fluctuation Analysis

Detrended fluctuation analysis (DFA) is technique used to quantify long-range power-law correlations in signals and it is considered as a generalization of the Hurst exponent that is more robust. It was originally introduced to be applied to signals whose underlying statistics (such as mean and variance) or dynamics are non-stationary [182]. Both RSA and DFA constitute analysis tools for estimating the scaling of the variance (or energy) of the time series at different time scales and it has been shown [183] that they are analytically connected through an integral transform. Among the numerous extensions proposed in the past to improve the original RSA method for the computation of the HE, DFA is one of the most successful due to capability to mitigate local non-stationarities [182] by employing a ''detrending'' operation. The basic algorithm for a time-series $\beta_1, \dots, \beta_N$ is:

1. *Computation of the mean* $\hat{\beta}$
2. *Compute rescaled range time series by subtracting the sample mean from the original series* $S_k = \beta_k - \hat{\beta}$ *for* $1 \leq k \leq Z$ *(*S_K *becomes zero-mean)*
3. *Compute the cumulative time series* $\Gamma_k = \sum_{i=1}^{k} S_k$ *for* $1 \leq k \leq Z$
4. *Split the N sample time series into windows of various sizes*

For all the different window sizes:

5. *Perform a polynomial fit in each integrated (cumulative) time series (n-th order polynomial regression denoted as DFAn)*
6. *Compute the mean squared residuals*
7. *Plot the values of log mean squared residuals against* $log(N)$*, also referred to as fluctuation plot.*

The above algorithm detects self-similarity (or fractal properties) by quantifying the dispersion of the residual of integrated fluctuations from the regression, at different resolutions (or window sizes). The power law scaling is derived from the fluctuation plot by performing a least-squares fit. The scaling exponent (from the DFA) value provides information about the series self-correlations, and - as with the original HE - if the value is < 0.5 it indicates anti-correlated behavior. A value of $\simeq 0.5$ is for uncorrelated white noise and for $> 1/2$ it corresponds to correlated time series. Also, if the value is $\simeq 1$ then the series shows 1/f-noise (pink noise).

Trends of higher order can be removed by higher order DFAn (n here stands for the order) where for instance, a linear fit can be replaced by a polynomial fit. The Hurst RSA analysis removes constant trends in the original sequence and thus in its detrending, it is equivalent to DFA1. Robust Detrended Fluctuation Analysis (r-DFA) is an extension of the standard method [184] that is capable to determine statistically significant scaling exponents and optimum crossover locations. An example of recent application where the DFA is uses regards Electroencephalography (EEG) signals [185], [186].

It is noted that this method is computationally more expensive than RSA and according to [177], DFA can introduce an uncontrolled bias and might not be as efficient as designed to be when non-stationarities are present.

4.2.3 Correlation Dimension

The correlation dimension (CD) measure is used in chaotic dynamical systems to estimate the dimension of attractors [187]. An attractor is a set of states that a dynamical system tends to evolve regardless of the initial conditions, and when state values get close to attractor values they are sustained close even if the system is perturbed. Mathematically the CD is based in the concept of the correlation integral C, and for a set of Z points $\{\beta_1, \beta_2, \cdots, \beta_N\}$ it is expressed as [188]:

$$C(Z, r_{sim}) = \frac{2}{Z(Z-1)} \sum_{p,j} HS\left(r_{sim} - \left\|\beta_p - \beta_j\right\|\right) \tag{29}$$

which is a function of the number of data points Z and the similarity distance r_{sim}, $HS(x)$ is the Heaviside function, and the summation goes over number of pairs for which the between-point distance is less than r_{sim}. For large Z and small r_{sim} the correlation integral is expected to scale according to the relation [189], [190]:

$$C(Z, r_{sim}) \approx \left(\frac{r_{sim}}{R_{eff}}\right)^{\nu} \tag{30}$$

with R_{eff} a constant of proportionality (effective radius) and ν the dimension of the attractor or the correlation dimension. There are several algorithms available that can estimate u such as the construction of a log-log plot of $C(Z, r_{sim})$ against r_{sim}, and by computing the slope of the curve at a specific range [188], or by computing the logarithm of every distance within the specified range and then estimating an average value a which then is used to find $\nu = -1/a$ [191]. Also, box assisted algorithms have been widely used [187].

The correlation dimension is measure of the dimensionality of the space occupied by a set of points and it is a type of the fractal dimension. As Falconer states [192], "many different definitions of 'fractal dimension' are scattered throughout the mathematical literature", however in Mandelbrot's paper [193] fractional dimensions where based on self-similarity characteristics. The value of correlation dimension is directly proportional to the level of chaos in the system, that is: a higher value of represents a high level of chaotic complexity in the system.

4.2.4 Fuzzy Entropy

The Fuzzy Entropy (FE) measure was introduced in 1972 [194], and many modifications on its definition were proposed over time, see for instance in [195]. Entropy is a value which increases as the level of irregularity (or the chaos degree of a system) increases. FE was designed to measure the regularity/complexity of time series, but the work of [196] demonstrated that it measures randomness instead. Its superior performance on pattern recognition problems has been demonstrated by a wide range of applications. It is capable of capturing the non-linear characteristics of EEG signals by quantifying the complexity [167], but also used successfully for an electrocardiography application in [197] and motor bearing fault [198].

The foundations of FE lie in fuzzy set theory, a framework that has been successfully used in modelling complex and imprecise systems. Fuzzy sets contain elements with varying membership degrees [199] or equivalently, elements are mapped to membership values through a membership function. A membership function essentially maps elements of a fuzzy set into a real value within the interval $[0,1]$, and it is the counterpart of the indicator function in classical (crisp) sets. The membership function [200] is key to fuzzy logic as it represents the degree of truth, an easily misunderstood concept with probabilities. In particular, a fuzzy set is defined as abstract and it is not related to the probability of occurrence on an event, therefore the two concepts are clearly distinct. For the formulation of the FE metric the probability employed by Shannon entropy is replaced by the membership function. The derivation of the mathematical representation is given in [201]:

$$H_{FE} = -\eta \sum_{e=1}^{n_{member}} [\mu_e \log \mu_e + (1-\mu_e)\log(1-\mu_e)] \tag{31}$$

where μ_e is the membership function for $e = 1, \dots, n_{member}$ and $\eta = 1/n_{member}$ is a constant. From (31) it is assumed that for the computation of the FE the computation of the memberships is required. One of the widely used membership functions is the Gaussian where original values are transformed into a point of the normal distribution. The mean of the Gaussian determines the set definition and it is assigned a membership value equal to 1. Membership values decrease depending on how close the transformed values lie to the mean in both negative and positive directions.

Several algorithms have been proposed to compute the FE, but for this study the method presented in [202] is used. Essentially the FE measure represents a more accurate variant of the Sample Entropy by adding a modification on the Heaviside function replaced with a fuzzy membership function. The algorithm at first constructs a state-space representation is by forming $Z - m\tau_t$ vectors $\vec{b}_m(p) = \{u_p, u_{p+\tau_t}, u_{p+2\tau_t}, \cdots, u_{p+(m-1)\tau_t}\}$ with $1 \le p \le Z - m\tau_t$ and denoting with τ_t the time-delay, and m is the embedding dimension. Then distances $distance(p,e)$ between $\vec{b}_m(p)$ and $\vec{b}_m(e)$ for $e \neq p$ are calculated, and the determination of $\Theta_e^{m+1}(distance, radius)$ follows next, a quantity that ranks the similarity between vectors with the next point $m+1$. The two vectors are compared given a $radius$ threshold. More details especially considering multi-scale analysis can be found in [203] and the time delay parameter is provided in [197]. The computation of the H_{FE} is [197]:

$$H_{FE}(m, \tau_t, radius) = -\ln \frac{\sum_{e=1}^{Z-m\tau_t} \Theta_i^{m+1}(distance, radius)}{\sum_{e=1}^{Z-m\tau_t} \exp(-\ln(2)(distance(p,e)/radius)^2)} \tag{32}$$

with $\Theta_e^m(distance, radius)$ denoting the average number of vectors $\vec{b}_m(p)$ within the $radius$ of $\vec{b}_m(e)$ or $distance(p,e) \le radius$ for all $e = 1,2, \dots, Z - m\tau_t$ and $e \neq p$. Also, according to the algorithm of [202] the local mean of each vector is removed before the distance $distance(p,e)$ calculation. The computation formula (32) was used in an electromyography application by [204] to evaluate the similarity between vectors based mainly on their shape.

The FE can quantify the irregularity of time-series by quantifying how unpredictable are the fluctuation patterns. According to [196] values close to 0 suggest a periodic time series, values between 1 and 2 indicate mixed processes, and values from 2 and above correspond to Gaussian noise.

4.2.5 Recurrence Period Density Entropy

The measure of Recurrence Period Density Entropy (RPDE) is developed from the idea of recurrent time between the recurrent points [205], [206] and it constitutes an average entropic measure [207]. It is based on the recurrence analysis technique (recurrence plot) which is conceptually simpler than spectral analysis tools like Fourier or Wavelet transform, and easier to be implemented as it does not require signal decomposition within a basis. It has been successfully used to detect abnormalities in biomedical contexts such as speech signal. The RPDE computation is based on the construction of the recurrent matrix:

$$\mathcal{R}_{p,e} = HE\big(radius - \|\beta_p - \beta_e\|\big) \tag{33}$$

for a set of Z points $\{\beta_1, \beta_2, \cdots, \beta_Z\}$ which correspond to the dynamical state at times p and e, while $HE(\beta)$ represents the Heaviside function. In this context recurrence is used to describe the closeness of any two points, which are considered close if their state vectors inside a $radius$. The recurrence matrix is essentially a symmetric matrix of ones signifying recurrent points and zeros for non-recurrent points in phase space using segments of the signal in repetition [208].

RPDE requires the embedding of the time series in phase space, an operation that can be carried out by forming time-delayed vectors. A phase space sequence is reconstructed by selecting both a suitable time-delay τ_t and a proper embedding dimension m as $\{u_p, u_{p+\tau_t}, u_{p+2\tau_t}, \cdots, u_{p+(m-1)\tau_t}\} = \{\beta_1, \beta_2, \cdots, \beta_Z\}$. Optimal values from delay τ_t and m can be estimated. Essentially these parameters are estimated by systematic search for the optimal set as practical embedding parameter techniques for stochastic systems are not available [209]. For each point β_p recurrence times (time-difference between one recurrence) within the $radius$ are reported in a histogram that is normalized to sum to unity. It is noted that dedicated algorithms for optimizing τ_t and m with respect to the time series available can be found in literature (see [207] and references therein). This normalized histogram corresponds to the estimate of the recurrence period density function, and the normalized density of this function gives the (normalized) RPDE [206]:

$$H_{norm} = -\frac{\sum_{p=1}^{T_{maxR}} \Pr(p) \ln \Pr(p)}{\ln T_{maxR}} \tag{34}$$

T_{maxR} is the maximum recurrence time found in the embedded state-space and H_{norm} is unit-less with real values in the range [0,1]. For purely periodic signals the RPDE equals 0 whereas for purely i.i.d. uniform white noise it shall equal to 1. The RPDE is designed to distinguish between autocorrelated random processes and (at the extreme case of) uniform random noise, and in characterising the extent to which a time series repeats the same sequence. Therefore, it is very similar to linear autocorrelation or time delayed mutual information, except that it measures repetitiveness in the phase space of the system, and that it is a more reliable measure based upon the underlying dynamics that generated the signal. The RPDE also offers the advantage that it does not require assumptions on linearity, Gaussianity or dynamical determinism.

4.3 Numerical simulations

The verification of the suitability of the selected feature algorithms under study is performed using simulated data implemented in MATLAB. The input pre-correlation RF signal sequence is generated from a custom GPS L1 C/A simulator which is implemented from scratch, generating signal samples at millisecond scales. The RF signal structure is deterministically defined by the time-dependent convolution of the spreading codes of all satellites in view. On top of the SIS (LoS) signal structure the corresponding multipath echoes are superposed. The simulator also applies the fading channel influences and potential interferences to produce different data samples for each scenario.

The development of the algorithms that generate the synthetic data for the simulation is based on the assumptions listed next:

- Fast fading variations are usually modelled as the sum of a limited number of sinusoids with random phases and different frequencies [39].
- The above operation leads to an extremely time correlated amplitude profile that can also be modelled as a non-white random process [210].
- Receiver dynamics are considered; therefore, amplitude variations are also affected from the receiver movement (or the movement of the surrounding objects). Indicative values are adopted according to [6].
- In a single reflector scenario located in the vicinity of the receiver, dominant periodic patterns in the carrier power variations might be expected. In any other case, this assumption shall not hold and simulations regard only the latter case.
- Power attenuation and fluctuations due to path loss and shadowing are neglected (considered constant for the 1 ms duration).
- Data bit transitions are not considered in the sampled waveforms.
- One term representing the total white noise is considered corresponding to a sum of the individual white noise contributions per channel.

The model of the synthetic signals considering $(18), (19), (22)$, is simulated from:

$$x_{RF}[t] = \sum_{i=1}^{N_{sat}} \left(r_{IF,i}[t] + MP_i[t]\right) + v[t] \tag{35}$$

and $|x_{RF}[t]|^2$ corresponds to the power time-series that are used as input to the feature extraction algorithms. It is noted that complex signal carriers are employed (and the real part is extracted for $x_{RF}[t]$) and the entire usable bandwidth of the ADC $(0 - f_s/2)$ is used. The simulation on the different environments is based on the Monte Carlo method, and the parameters are drawn randomly per execution. The number of executions depends on the runtime of the feature extraction algorithms. In particular, the most computationally efficient method is the GHE, then DFA, CD and FE which are more demanding, and the most computationally expensive method is RPDE.

The random parameters, changing per execution are given in the table below:

Table 2 - L1 C/A signal/s simulation parameters

LoS-related: • $i \in [1,32]$ representing the satellite PRN, is randomly chosen (using uniformly distributed pseudorandom integer generator from MATLAB) • $\tilde{c}_i$ sequence is introduced with a random codephase (uniformly distributed pseudorandom integer generator from MATLAB) • $A_i \in [0.5, 1]$ is random (uniformly distributed) per LoS signal • $\varphi_i \in [0,2\pi)$ is random (uniformly distributed) • $\Delta f_{D,i}$ is random (uniformly distributed) in the interval +/-5kHz in the IF carrier freq.
Multipath-related (NLoS): • Number of NLoS components J is randomly chosen (using uniformly distributed pseudorandom integer generator from MATLAB with a predefined min value J_{min} and max value J_{max}) for each LoS signal • Path amplitude $A_i a_{i_j} \in [0, 0.85]$ for $j \in [1, J]$ is random (uniformly distributed) for each NLoS echo • Δf_j is random (uniformly distributed) in the interval +/-300Hz in the IF LoS carrier freq. for each NLoS echo • $\varphi_j \in [0,2\pi)$ is random (uniformly distributed) for each NLoS echo • $\tilde{c}_{i_j}$ is introduced with a random codephase (uniformly distributed pseudorandom integer generator from MATLAB) for each NLoS echo

In addition, the RF-frontend IF is assumed to be 4.13 MHz and the ADC sampling rate equal to 20 MHz. A White-Gaussian-noise process is also added, generated from the "wgn" function of MATLAB following the predefined sampling rate. A receiver's front-end bandwidth determines the SNR that is output of the RF-frontend. Typical values for the GPS L1 C/A signal for a receiver frontend bandwidth of 4 MHz shall be between -29 dB to -21 dB [211]. However, since the amplitudes of the signals are not selected according to real physical conditions rather relying on indicative/placeholder numbers, the resulting average SNR ratio in all simulations is ~11 dB.

Several assumptions were adopted in order to build a simulation environment that is not very complex and heavily time-consuming for a typical PC to execute, but in parallel to have some relevance to realistic conditions. First consideration is that in the majority of realistic scenarios few LoS signals (direct satellite signals) implies that the receiver is located in a relatively cluttered environment with obstructions and consequently, the presence of one to several MP components. In open-sky conditions it is assumed that 6-11 satellites should be able to provide direct LoS signal to any user on the Earth's surface simultaneously at any time. Therefore, any scenario with less than 6 satellite signals received implies that there might be at least one obstruction in the environment of the receiver. On the other hand, many LoS signals available would suggest an open-sky, free-from-many obstructions environment.

Following the same logic, the more numerous obstructions may suggest a larger number of MP echoes received, and it is considered here that maximum number of possible MP components increase as the number of received satellite signals becomes smaller. The quantification of these simplistic assumptions is shown in the table below:

Table 3 - Generic simulation parameters

Number of LoS	**J_{min} to J_{max} (num NLoS per LoS signal)**	**Signal-to-Multipath ratio (SMR) [dB]**
0 (1-3 differ. signals)	20-50	NA
1	20-40	1.2 - 20
2	10-25	
3	7-15	
4	3-10	2-25
5	0-5	
6	0-2	
7	0-2	
8	0-1	3-25
9	0-1	
10	0-1	
11	0-1	

A summary of key points regarding simulations:

- Simulations rely on indicative numerical values that serve as inputs to the models presented, rather on synthetic data from a multipath simulator
- Several assumptions have been made on the correspondence between the values and MP conditions and in consequence (see Table 2), on how the resulting waveform structure might be caused in a receiver
- The main scope is to provide evidence on the suitability of use of these techniques for this application

Finally, for verification purposes, feature values are provided for four possible cases:

1. the multipath- and noise-free signal
2. the noise-free signal under multipath
3. the multipath-free signal under (white) noise
4. and the multipath-contaminated signal in noise.

Signals 1. correspond to a PRN-modulated carrier from which - in principle - pseudo-randomness ("whiteness") shall be confirmed through the feature extraction method results. In case 2. the signal sequence is coloured by correlated multipath effects and their characteristics shall become evident through simulations. Signals 3. and 4. take into consideration the white noise which is expected in all realistic scenarios. In 3. the uncorrelated noise property is expected to hold since two uncorrelated random processes are superposed. On the other hand, 4. is a noisier counterpart of case 2. which is considered closer to real conditions. Although measure distributions are treated as unknown, the Gaussian mean estimator was used to determine the mode of each distribution.

4.3.1 Generalized Hurst Exponent

The generalized Hurst exponent [178] constitutes a tool for analysis of q-order moments of the distribution of the increments. For $q = 1$ it is equivalent to the original Hurst exponent definition, thus quantifying the scaling behaviour (of the absolute values) of the increments. For $q = 2$ the scaling behaviour proportional to the autocorrelation function is analysed. The latter is of greater relevance given the nature of the signal under analysis, although, both measures where tested. The overall time complexity of the algorithm is $O(\tau_{interval\_max} \times Z)$ with $\tau_{interval\_max}$ the maximum value of the scaling window and Z is the length of the data series.

In the first two figures, the results for $H(1) = 1$ are shown for 5000 iterations:

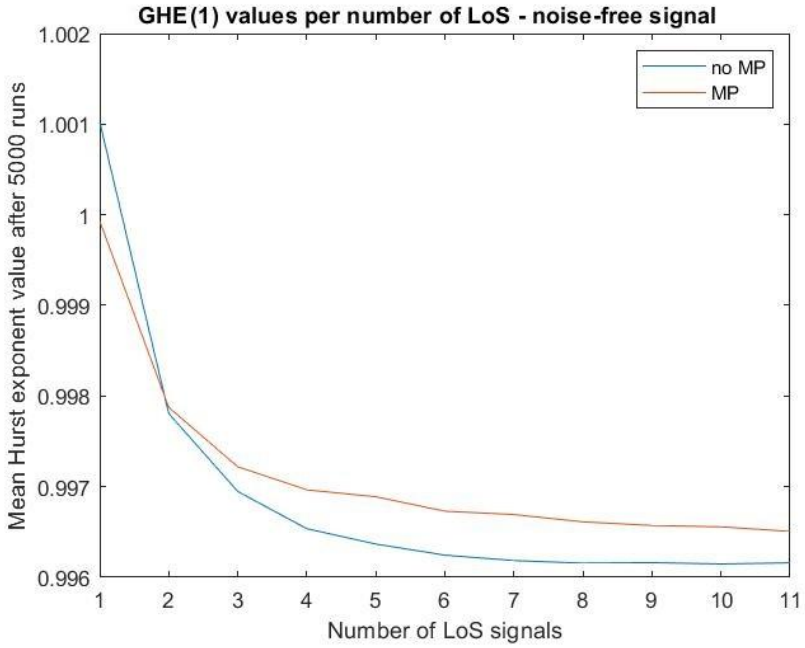


Figure 13 - GHE (q=1) simulation results without additional white noise after 5000 iterations

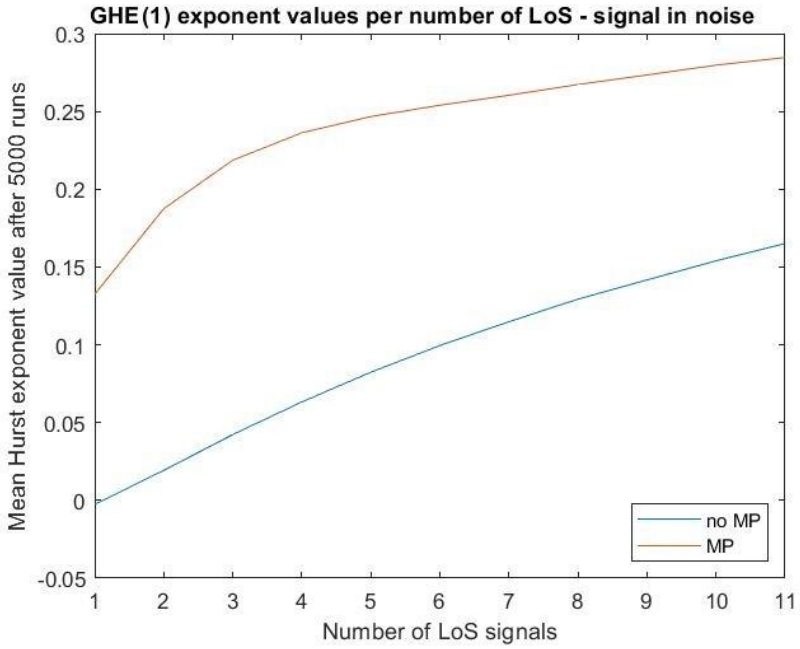


Figure 14 - GHE (q=1) simulation results with added white noise after 5000 iterations

Theoretically, for values close to $H(1) = 1$ the time series indicate a strongly deterministic system where autocorrelations dominate. While this assumption agrees very well with the MP-afflicted noise-free case, it appears to hold also for to the superposition of pure PRNs (noise-free case). Moreover, both noise-free simulations show monotonically non-increasing characteristics where less number of LoS signals shows a higher value (although numerically not significant), a result that is slightly contradictory with the theoretical prediction. Overall, it can be said that values are very close to each other in these simulations.

In the signal-in-noise case, both functions appear to be monotonically non-decreasing. Here, there is strong evidence of anti-correlation as $H(1) < 0.5$ and the result agrees with the expectation that the more PRN codes and/or MP echoes are superposed the closer to the $H(1) = 0.5$ (that corresponds to Brownian motion) the exponent values should approach. It is however denoted that in the 'no MP' case the initial point of the function (1 LoS) has $H(1) \cong 0$ that corresponds to $1 / f$ pink noise, rather than white noise according to the expectation for the pseudo-random nature of one spreading code sequence.

Next, the corresponding results for $q = 2$ are shown:

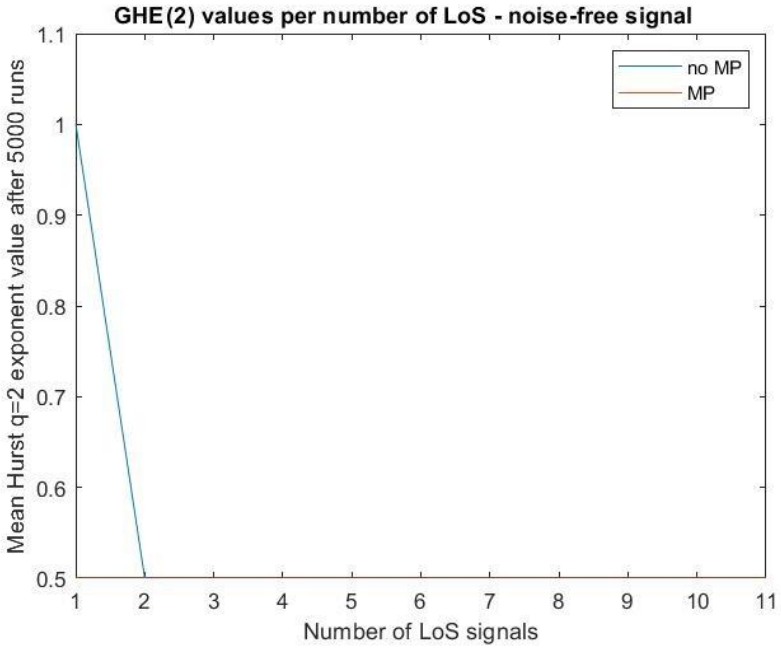


Figure 15 - GHE (q=2) simulation results without additional white noise after 5000 iterations

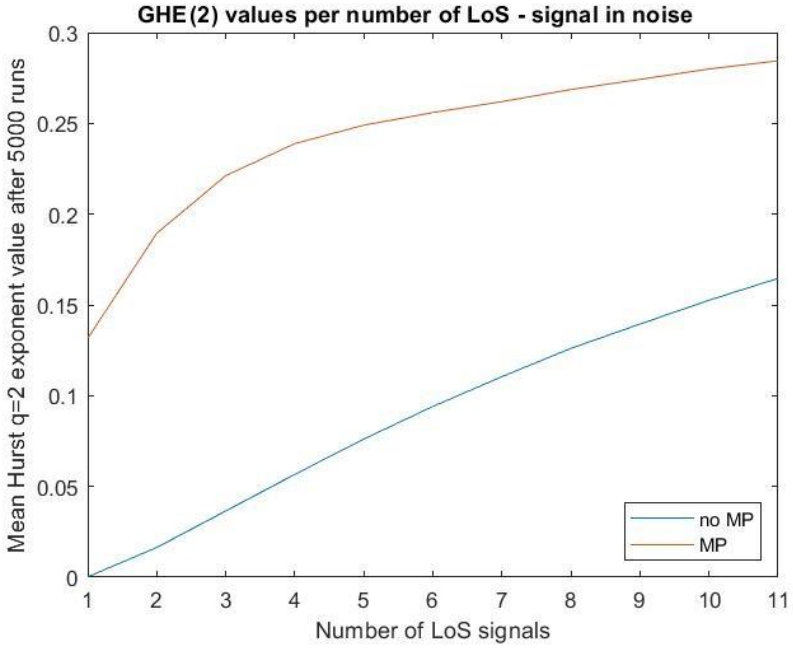


Figure 16 - GHE (q=2) simulation results with added white noise after 5000 iterations

A first remark after the inspection of both $q = 1{,}2$ results, is that functions look very similar. For the simulations regarding signal-in-noise, evidence shows that the exponent is a monotonic function of the number of LoS signals (considering the indicative MP-parameter values used). Also, the result is consistent for low SMR with a Rayleigh random walk where the mean value of the displacement is $std.\times \sqrt{N_{samples}}$. It is noted that the histograms of GHE estimates after 5000 runs indicate non-Gaussian structure, therefore the mean estimator might not provide accurate results (see Appendix A). Finally, in high-SNR conditions (see Appendix B) the measure tends to become constant as a function of LoS signals and GHE exponent values are very close to each other.

4.3.2 Detrended Fluctuation Analysis

The DFA simulations were performed using the standard function from the MATLAB library, while the polynomial order for detrending was configured to be equal to 2 (quadratic). The overall time complexity of the algorithm is $O(Z + M)$ where Z is the size of the input data and M is the number of scales generated. Due to computation resource restrictions the iteration number here is 1000. The results for the "noiseless" and "noisy" case are presented in Figures 17 & 18:

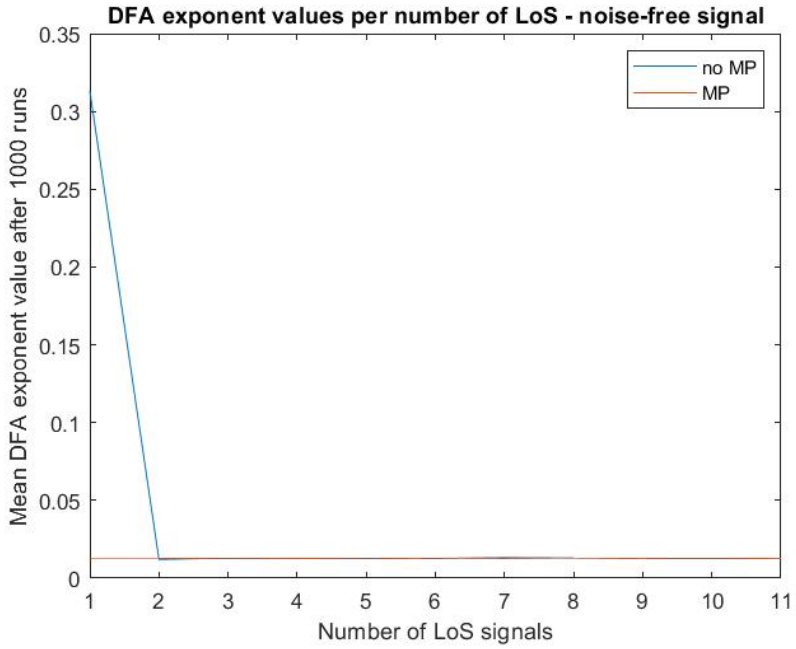


Figure 17 - DFA simulation results without additional white noise after 2000 iterations

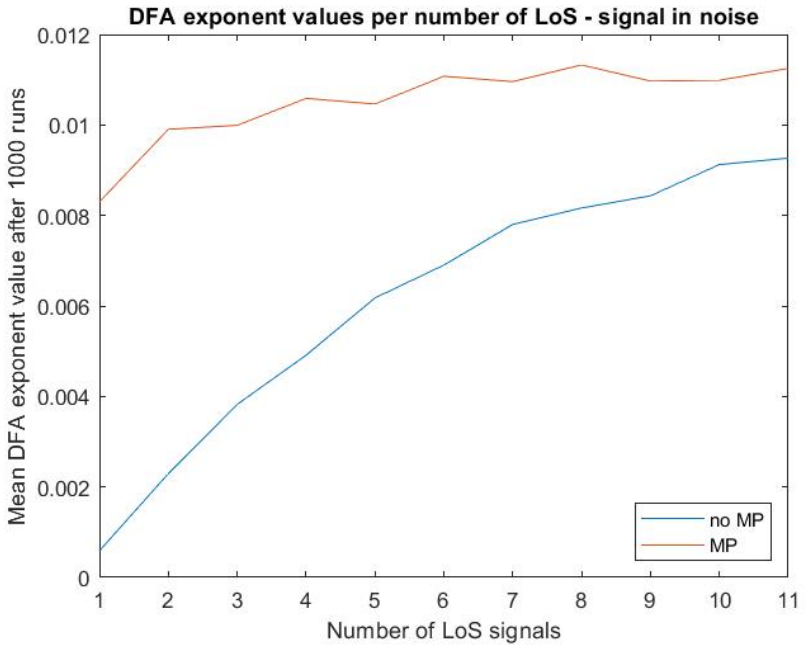


Figure 18 - DFA simulation results with added white noise after 2000 iterations

A generic remark on all cases is that a strong anti-correlated scaling behaviour is evident according to the DFA exponent scale definition (< 0.5). Similar to GHE, randomness appears to increase when the number of LoS signals increases. What is notable once again (after the GHE $q = 1$) is that without adding WGN the DFA as a function of number of LoS appears constant except for the 1 LoS in the 'No MP' case. Even by adding MP echoes to the SIS signals, makes no difference. In the "noisy" case DFA functions show similar properties to the GHE. However, in contrast to the GHE exponent, the signal-in-noise simulations show that the exponent function appears to break the monotonic property for the MP-afflicted case.

4.3.3 Correlation Dimension

For the implementation of the CD algorithm, the corresponding MATLAB function was used, i.e. "correlationDimension()" that estimates the correlation dimension of the uniformly sampled time-domain signal. It has a time complexity of $O(Z^2)$ with Z being the length of data. The function's runtime was rather long and only 200 Monte Carlo runs were performed (lasting approx. 24h). The results are given in the following two figures:

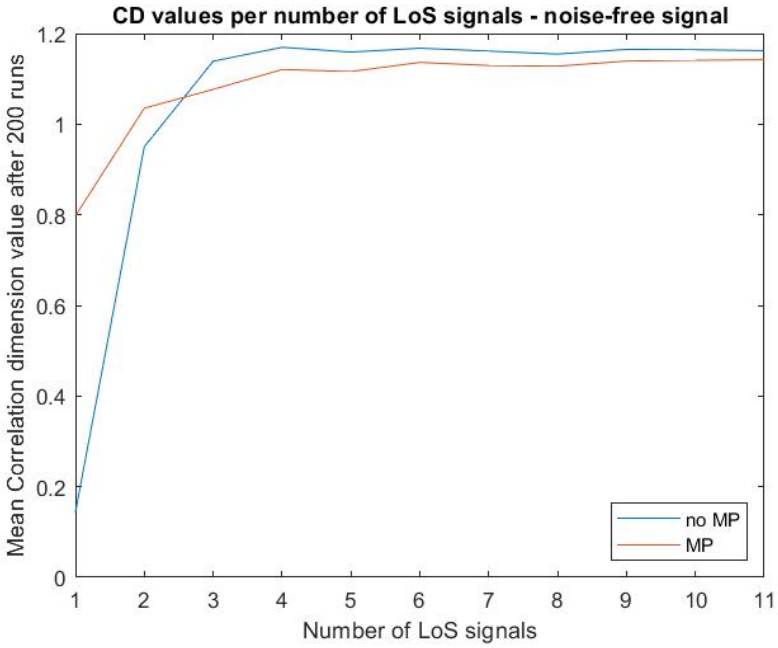


Figure 19 - CD simulation results without additional white noise after 200 iterations

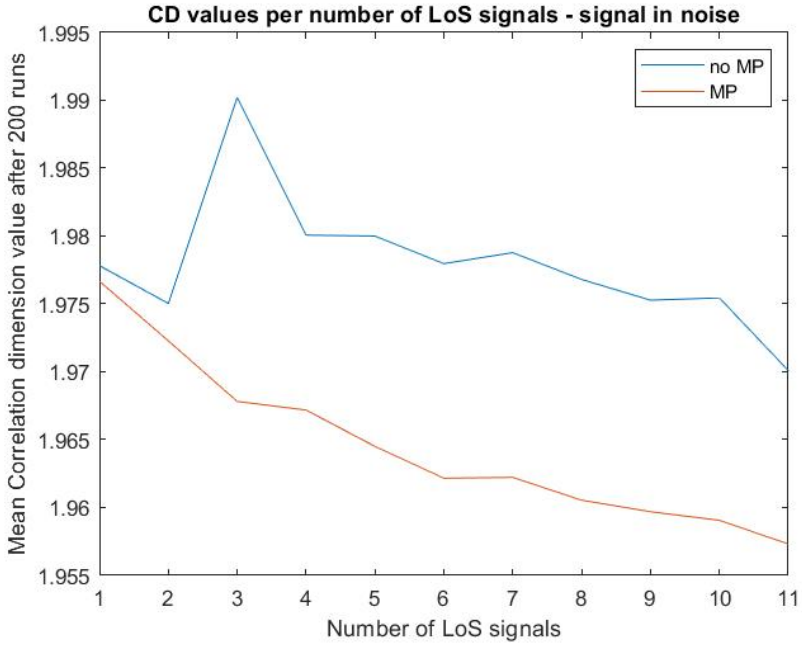


Figure 20 - CD simulation results with added white noise after 200 iterations

According to Figure 19 for one LoS and no added noise the majority of CD values are close to 0. This finding confirms that an individual sampled pseudo-random (Gold code) sequence is indeed approximating a random noise process. For 2 LoS the majority of CD values are concentrated around values 0.9-1 confirming that the combination of two sampled pseudo-random (Gold code) sequences shall result into a higher complexity indicator (when compared to 1 LoS). For 3 LoS and above the majority of values are between 1 and 1.3, again, suggesting higher complexity of the composite signal (from 1 or 2 LoS). However, it seems that the measure is not sensitive to the number of LoS in this regime.

In the MP-afflicted (Figure 20) case with noise, the CD value decreases while the number of LoS increases (and consequently the number of NLoS decreases), indicating that the level of chaotic complexity in the system decreases. This implies that randomness increases as the superposed PRN modulations increase in number and it appears as a plausible justification as one PRN sequence indicates a higher level of deterministic chaos than two or more (superposed). However, in the MP case the opposite result is produced (closer to the random end). Also, the time-correlated components (i.e. waveforms with NLoS signals) seem to produce systematically lower correlation exponent values than when these were absent, except for 1 or 2 LoS components in the noise-free runs. Finally, the CD function does not appear to be monotonic (MP-afflicted case with noise) in contradiction to the GHE exponent. It is noted that results rely only on the 4 % of the runs compared to the latter.

Given the above points, conclusions regarding the validity of the use of this metric cannot be made according to the current simulations. One possible explanation is that the number of executions (200) were not sufficient to provide a reliable result. For this reason, histograms for the correlation dimension estimates of the multipath- and noise-free signal where generated (see Appendix A).

4.3.4 Fuzzy Entropy

The FE bears approximately the same computational cost as the CD algorithm. Implementation and configuration values were set according to the work of [197], [202] with the time-delay $\tau_t = 1$, $m = 2$, and $radius = 2 \times std$. The $std.$ corresponds to the standard deviation of the power time series. The approximate time complexity of the implementation is $O(Z^2/(m^2 \times {\tau_t}^2))$, with Z representing the length of the input series. The result is presented in the next two figures:

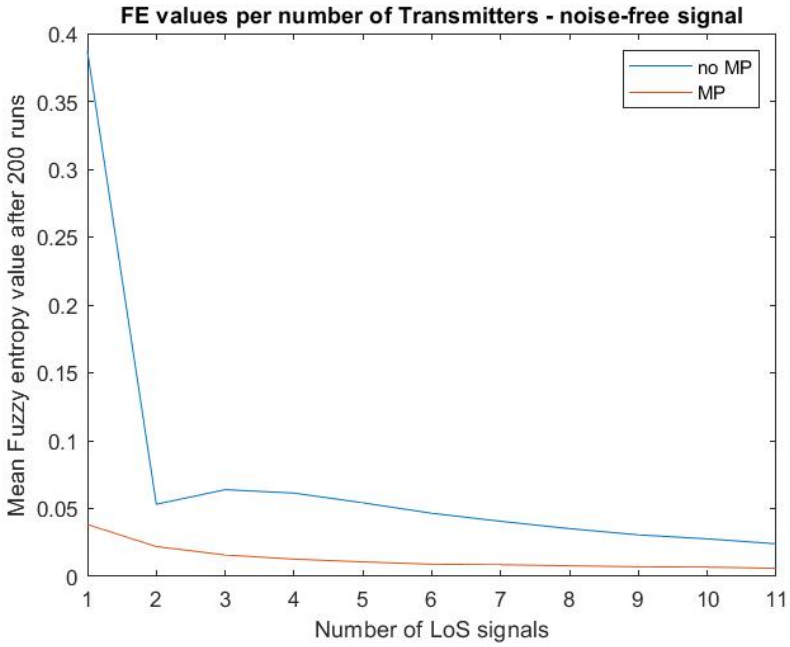


Figure 21 - FE simulation results without additional white noise after 200 iterations

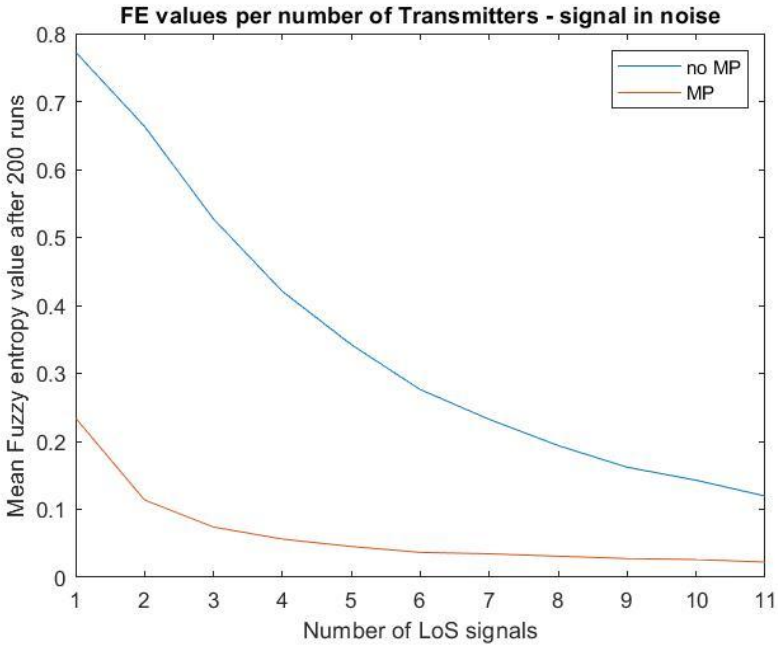


Figure 22 - FE simulation results with added white noise after 200 iterations

Given that a Gaussian noise component is added to the series in Figure 22, the FE values are higher than the ones for the noiseless case (Figure 21), which is consistent with the theoretical definition of the measure scale. It is also evident that MP-afflicted signals (Figure 21) report very low FE values close to zero, an indication of structured periodic signals according to the definition of the metric. Same applies as the LoS components increase in the case where no MP is present, as several PRN-modulated carriers combine.

In both figures, although the MP case appears to be monotonically non-increasing function, they also appear to converge after 4-5 LoS indicating that any number of LoS above this region might be difficult to distinguish as values become numerically very close. This is not the case with the corresponding 'No MP' functions. Another remark is related to the highest value denoted on 1 LoS in the noiseless simulation, which, although the FE value is "closer" to the Gaussian noise end, it is still very far from the theoretical value of (<)2. Same applies for the rest number-of-LoS FE values that appear very low (and close to zero). It is also noted that since values close to 0 suggest a periodic time series in the FE, these results suggest that in higher numbers of LoS the composite signals appear as more structured and periodic.

4.3.5 Recurrence Period Density Entropy

Due to the very long runtime of the implementation of the RPDE algorithm employed from [212] only 100 Monte Carlo iterations were performed (lasting more than 24 hours. The time complexity can be approximated as $O(Z + m^2)$, m being the embedding dimension and Z the length of the time series. Certainly, the results shown in the Figures 23 & 24 serve as a first indication of performance, as with the other simulations.

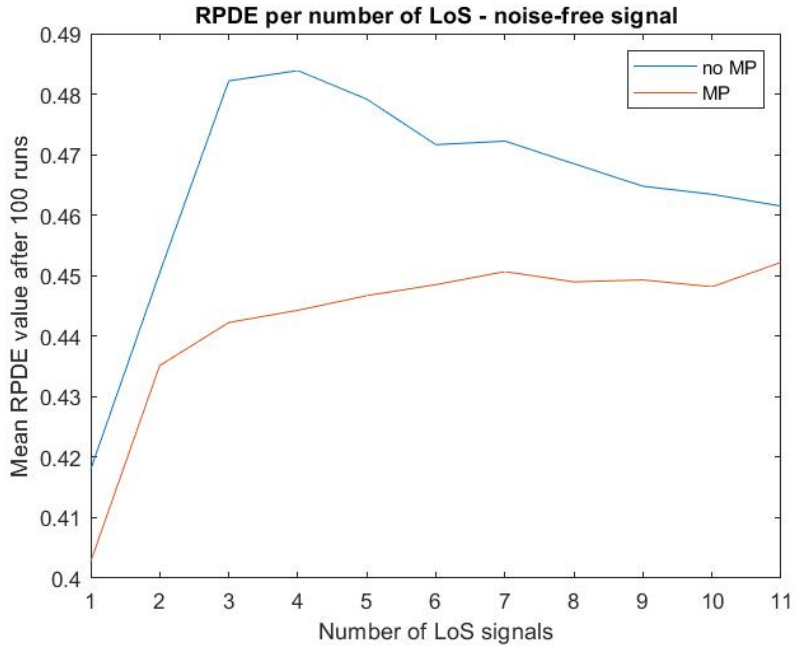


Figure 23 – RPDE simulation results without additional white noise after 100 iterations

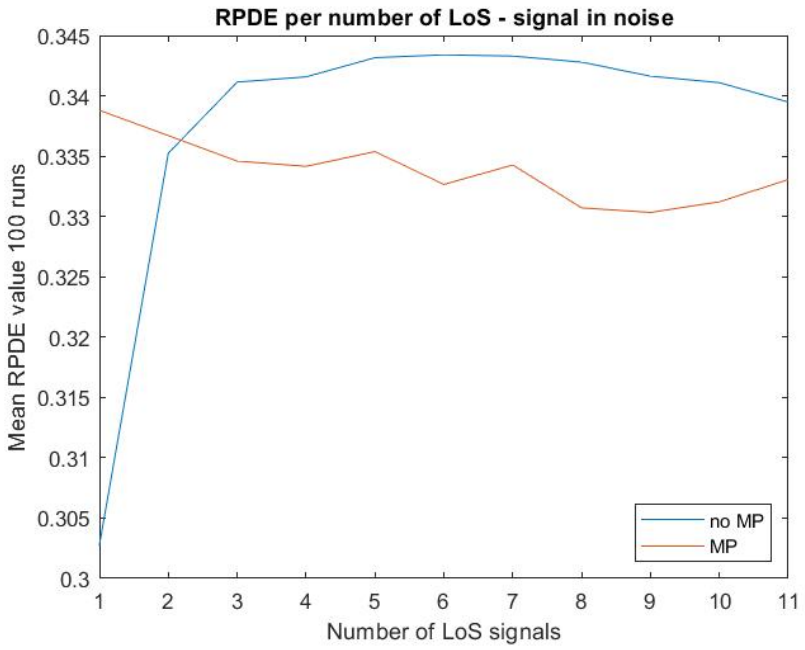


Figure 24 - RPDE simulation results with added white noise after 100 iterations

For the noisy simulation scenario, it is observed that the differences in the RPDE values between the 'No MP' and 'MP' case are subtle. For number of LoS above 2 the multipath-afflicted function shows a slightly lower value that the SIS one and it is not monotonic as it shows certain fluctuations. In contrast, the noise-free case shows that the 'No MP' function is systematically lower than the 'MP' one and the magnitude of their difference is more pronounced. Numerically the noisy results show a lower value than the noise-free case, a fact that suggests that the additional noise influences the results to lower values approaching the 0 end of the RDPE scale which corresponds to periodic/structured signals. This is contradicting to the expectation as additive Gaussian noise should lead to a value increase approaching 1 which indicates pure WGN and therefore, appears as counterintuitive. In general, the interpretation of RPDE results is not intuitive and straightforward. In any case some valuable information could be contained in this measure to determine a classification outcome but its inclusion might come with the risk of misclassification and performance degradation.

4.4 NLoS conditions

The non-line of sight scenario is a special case of signal reception which is usually equivalent to a receiver located indoors where no direct LoS is available but the antenna receives reflected, scattered or/and diffracted signals. In this setting it is common to assume that there is large number of (resolvable) echoes of each signal due to a congested environment. The numerical simulation parameters selected are presented in the table below:

Table 4 - Simulation parameters for indoor L1 C/A signal reception

NLoS conditions (Indoors) simulation parameters:

- Number of unique (transmitted from a single satellite) received signals present is defined to be within $[1,\ 3]$
- Number of (NLoS) echoes J per received signal is randomly chosen (using uniformly distributed pseudorandom integer generator from MATLAB with a predefined min value J_{min} and max value J_{max}) per unique signal (see Table 2).
- Path amplitude $A_i a_{i_j} \in [0,\ 1]$ for $j \in [1,\ J]$ is random (uniformly distributed) for each NLoS echo
- Δf_j is random (uniformly distributed) in the interval +/-300Hz in the IF LoS carrier freq. for each echo
- $\varphi_j \in [0,2\pi)$ is random (uniformly distributed) for each NLoS echo
- $\tilde{c}_{i_j}$ is introduced with a random codephase (uniformly distributed pseudorandom integer generator from MATLAB) for each NLoS echo

For each measure the results on noiseless and signal-in-noise cases are presented next:

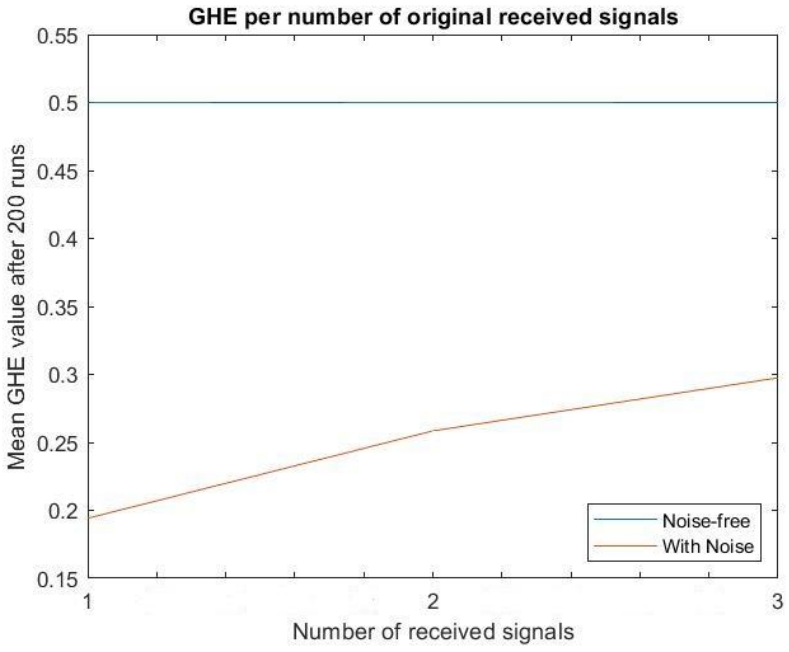


Figure 25 - Mean GHE value for simulated NLoS conditions

For the GHE ($H(2)$), the signal that is composed only of echoes (no WGN) appears as an approximately constant function with respect to the different PRNs included and it is close to 0.5, i.e. “Brownian” noise. When noise is added the value appears to be roughly between 0.2-0.3. The latter result is very similar to the case where LoS signals are present.

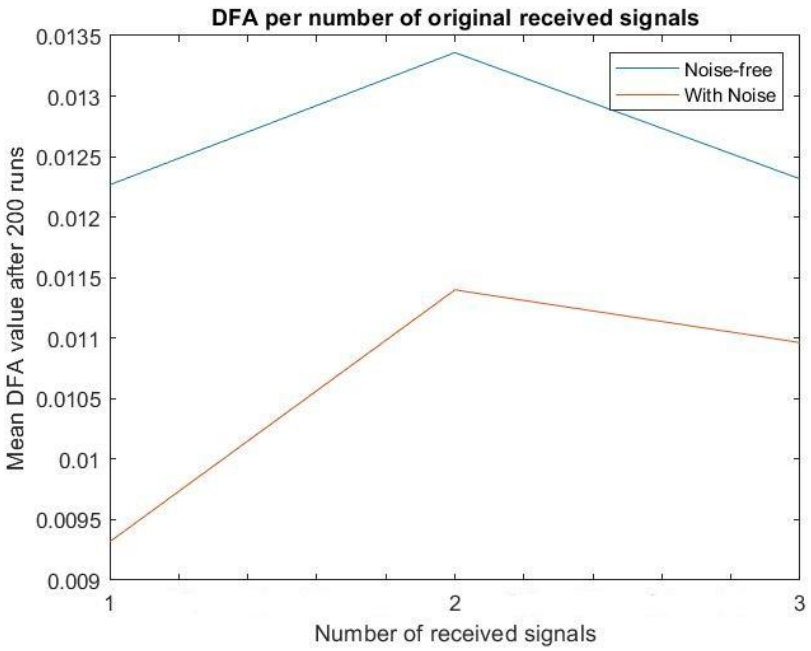


Figure 26 - Mean DFA value for simulated NLoS conditions

DFA values are again, very similar to the scenarios with LoS signals present. Also, the two depicted functions (Figure 26) appear to have a similar shape. It is reminded here that values so low close to zero indicate a strong anti-correlated behavior.

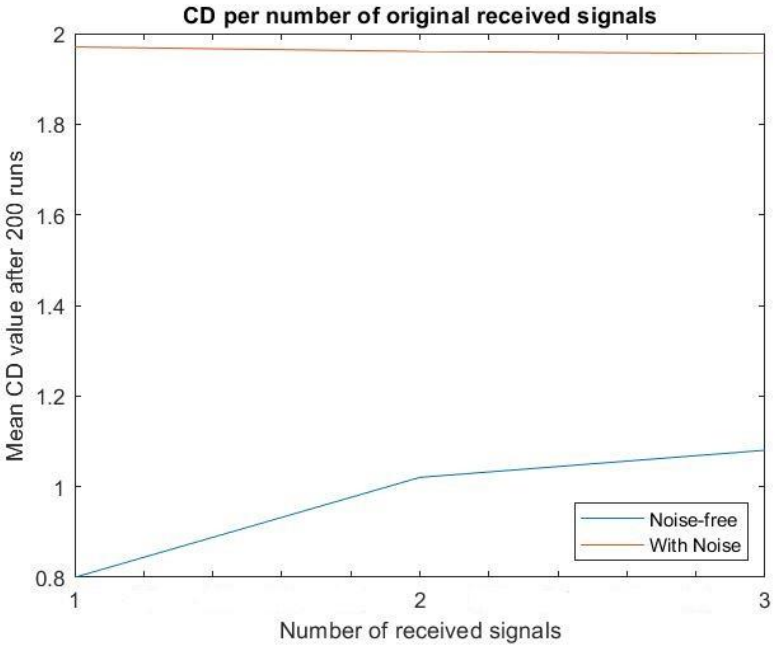


Figure 27 - Mean CD value for simulated NLoS conditions

The CD functions in simulated NLoS conditions (Figure 27) have shown results that are in agreement to the corresponding LoS conditions, suggesting that subtle differences if any, could distinguish the different environments.

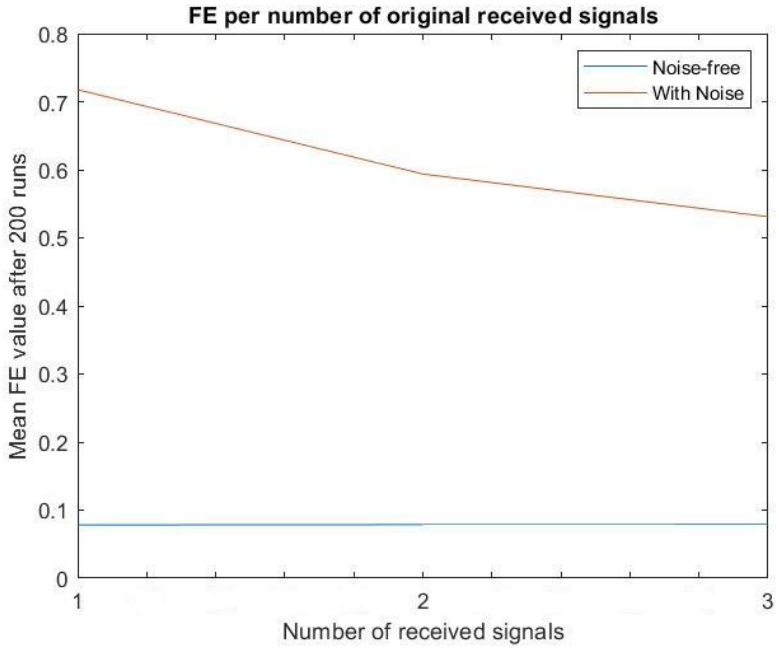


Figure 28 - Mean FE value for simulated NLoS conditions

The FE seems to provide a distinctly different result that the corresponding LoS simulations. Here in the noisy case, values are higher ranging roughly between 0.53 - 0.72 while in the latter all values where below 0.23. Both of these functions are monotonically non-increasing. The function corresponding to the noiseless case is approximately constant and values < 0.1 indicate strongly structured periodic signals which is consistent to the theory. Accordingly, the "noisy" function regards higher values.

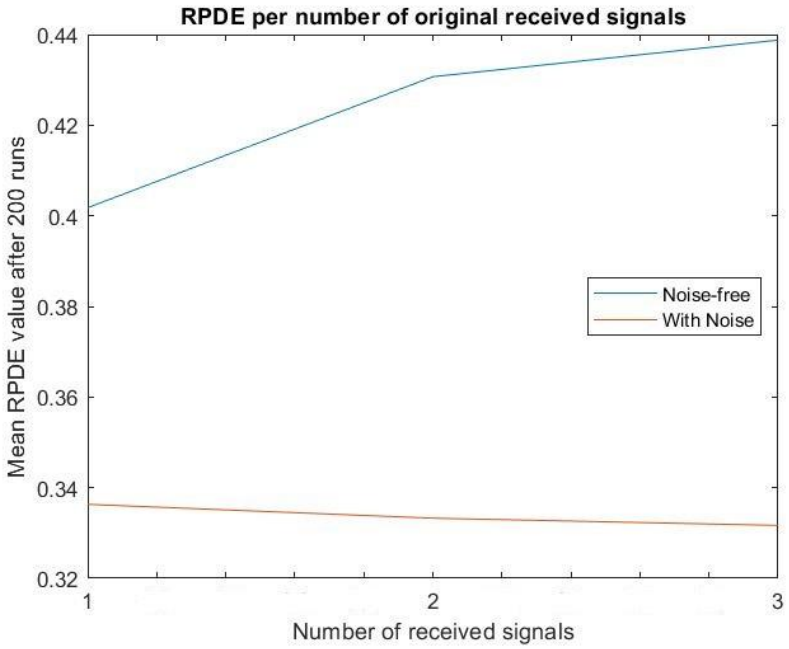


Figure 29 - Mean RPDE value for simulated NLoS conditions

Finally, the RPDE functions look very similar to the corresponding LoS simulations, both in numbers and in shape.

4.5 Remarks and discussion

Based on the results presented in the previous section:

- The GHE is favoured over the DFA algorithm to compute the scaling behaviour of the sampled waveform as it is more computationally efficient and provides a result that is easily interpretable (monotonic function). The latter statement relates to the fact that for every different number of LoS signals (and corresponding multipaths) as defined in Table 2 appears to correspond to a unique GHE value.
- CD provides a different type of information, it is a measure of deterministic complexity which could add diversity and redundancy in the feature extraction stage. It is denoted however, that it is increasing significantly the computational cost compared to GHE and DFA.
- FE but also RPDE are costly and the resulting functions manifest some difficulties. The major problem with the former is that measures are numerically very close after 4-5 LoS signals, while the latter is not monotonic. However, the FE has shown the best performance in the NLoS simulations.
- Overall, all metrics may provide valuable feature information and their contribution to a classification problem will be assessed in the next chapter.

Regarding the computational complexity, the GHE algorithm seems to be the less demanding while the rest methods follow in increasing complexity order: DFA, CD, FE and RPDE.

5 Multipath environment classification

5.1 Introduction

The purpose of this chapter is to demonstrate and validate that the studied feature extraction methods can provide an accurate classification result when used in conjunction to an off-the-self classifier. Towards this objective, the techniques of Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) were selected for simulation testing due to their simplicity and high performance when applied to supervised problems. The simplest test case would be to distinguish between two extreme classes of nominal (SIS) GNSS signal and NLoS/Indoors condition. However, a three-class classification simulation is also presented afterwards.

In the first paragraph the key principles and fundamental mathematics of the two main classification techniques are presented while the simulation results are given in the second paragraph of this chapter. In order to keep the computational cost of the proposed approach as low as possible no more than two features are used as inputs to the classifiers. Essentially this analysis could specify the highest performing feature couples and provide a first indication of a configuration that is capable of capturing and distinguishing accurately the different multipath-afflicted waveforms. Features combined in triplets are also presented next in order to test if the classification performance is improved for the (more challenging) three-class problem.

The methodology for the calculation of the predicted classes of measurement data followed is:

1. Generation of simulated training and test/measurement (features) datasets:
    - ✓ In the two-class classification between 'Indoors/NLoS' and 'No MP' signals the training sets are composed of 300 and 1200 simulated samples, respectively. For testing: two sets of 600 data points were generated from a different simulation than the one producing the training samples, both for 'Indoors/NLoS' and 'No MP' labels.
    - ✓ In the three-class classification between 'Indoors/NLoS', 'MP-afflicted' and 'No MP' signals the same number of training and testing data points were used for the 'Indoors/NLoS' and 'No MP' signals as above. For training the 'MP-afflicted' class: 1200 training points and 600 testing points were produced.
2. Computation of GHE, DFA, CD, FE and RPDE features using the training datasets. These features facilitate as inputs to the LDA and SVM classifiers to train the corresponding classification models.
3. Computation of different classification models for LDA and SVM:
    - ✓ Using only pairs of features for all possible combinations (10).
    - ✓ Using all possible combinations (10) of three features for the three-class problem.
4. Prediction of labels for all different combinations using the test data sets as inputs to the classifiers.
5. Performance assessment using the confusion matrix graph and the derived accuracy, sensitivity and specificity.

5.2 Classification Methods

5.2.1 Linear Discriminant Analysis

The method of Linear Discriminant Analysis (LDA) is considered one of the most well-used data reduction techniques, used in several application such as automatic speech recognition [213]. Gaussian linear discriminant analysis allows linear separation by considering joint Gaussian probability densities and modelling the data conditional upon each class-label [214].

There are two types of LDA technique to deal with classes [215]:

- class-dependent
- class-independent

In the class-dependent LDA, the input data are projected onto one separate lower-dimensional space per class, while in the class-independent method each class is considered as separate from others. In this study the class-dependent type is applied to the simulations.

A vector of samples $\boldsymbol{\mathcal{B}} = \left\{\vec{b}_1, \dots, \vec{b}_{N_{feat}}\right\}$ is considered that consists of N_{feat} feature-vectors $\vec{b}_i$ in a D-dimensional space with class labels $\Omega = \{\omega_1, \dots, \omega_{N_{class}}\}$ with $\omega \in \{1, \dots, N_{class}\}$ and N_{class} being the number of classes. Then the set $\boldsymbol{\mathcal{B}}$ can be subdivided into C subsets $\boldsymbol{\mathcal{B}}_1, \boldsymbol{\mathcal{B}}_2, \dots, \boldsymbol{\mathcal{B}}_{\mathsf{C}}$ denoted as $\boldsymbol{\mathcal{B}}_j \subset \boldsymbol{\mathcal{B}}$ which corresponds to class j. Also, μ is denoted as the centroid of set $\boldsymbol{\mathcal{B}}$ and μ_j the centroid of $\boldsymbol{\mathcal{B}}_j$.

The total scatter matrix $S_{Tot} \in \mathbb{R}^{D \times D}$ is then defined as [216]:

$$S_{Tot} = \sum_{\vec{b} \in \boldsymbol{\mathcal{B}}} (\vec{b} - \mu)(\vec{b} - \mu)^T \tag{36}$$

The within-class scatter matrix $S_{WC} \in \mathbb{R}^{D \times D}$:

$$S_{WC} = \sum_{j=1}^{\mathsf{C}} \sum_{\vec{b} \in \boldsymbol{\mathcal{B}}_j} (\vec{b} - \mu_j)(\vec{b} - \mu_j)^T \tag{37}$$

The between-class scatter matrix $S_{BC} \in \mathbb{R}^{D \times D}$:

$$S_{BC} = \sum_{j=1}^{\mathsf{C}} N_j^{samples} (\mu_j - \mu)(\mu_j - \mu)^T \tag{38}$$

with $N_j^{samples}$ representing the number of samples per class j. The transformation matrix W is given from [217]:

$$S_{BC} w_i = \lambda_i S_{WC} w_i \tag{39}$$

with $W = \{w_k\}$ for $k = 1, \dots, K$ where w_k are eigenvectors, and the columns of matrix W correspond to eigenvalues λ_k.

There are two major issues related to the use of the LDA method. First, if the dimensions are much higher than the number of samples in the data matrix, it is incapable of finding the lower dimensional space resulting to a singular S_{WC}. This is known as the small sample problem (SSS); a detailed analysis is given in [218]. The other drawback is that LDA is a linear discriminator, therefore, if different classes are non-linearly separable it cannot discriminate between them. The first problem does not apply here while for the second it might prove to be an issue if features are not linearly separated in feature space.

5.2.2 Support Vector Machines

A support vector machine (SVM) is a supervised learning algorithm originally developed to tackle binary classification problems. Methods have been proposed to expanded the SVM algorithm to multiclass classification problems by reducing the single multiclass problem into multiple binary classification problems [219].

As a maximum margin classifier [220], the criterion that defines the optimal decision boundary between two classes is the hyperplane that offers the largest possible separation between data points. Therefore, the SVM algorithm is estimating a hyperplane that is parallel to two (again parallel) support vectors that define each class boundary, given no data points are included in-between (Figure 30). The original SVM algorithm however was intended to separate linearly separable data points. In the MATLAB implementation the algorithm maximizes a soft margin allowing a small number of misclassifications.

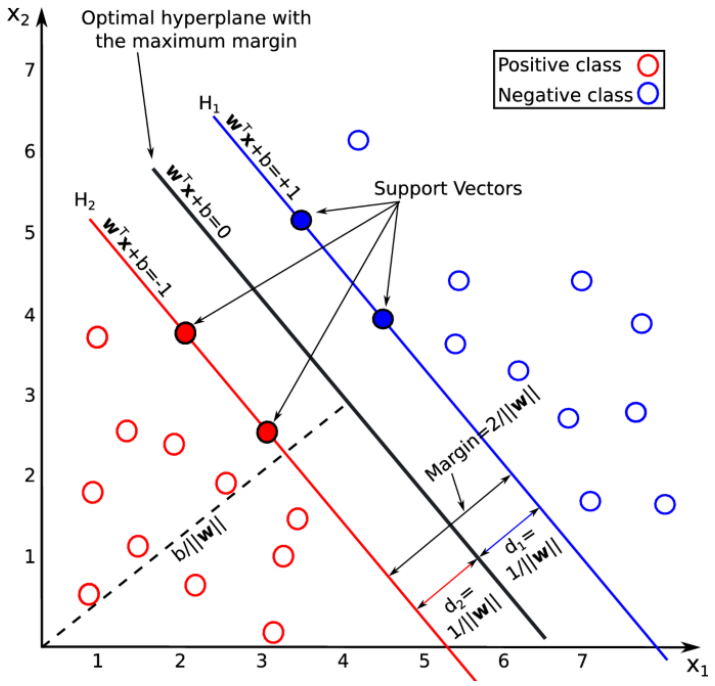


Figure 30 – The SVM concept and key parameters of the method [221]

The SVM algorithm is based on generic linear models of the form [220]:

$$y(\vec{x}) = w^T \phi(\vec{x}) + b \tag{40}$$

Here $\vec{x}$ corresponds to the training data set vector, w is an M-dimensional weight vector, $\phi(\vec{x})$ represents a fixed feature-space transformation, and b is a bias term.

The distance of a point x_n to the decision hyperplane is:

$$\frac{t_n y(x_n)}{\|w\|} = \frac{t_n(w^T \phi(\vec{x})+b)}{\|w\|} \tag{41}$$

where the training data set vector x_n gets target value $t_n \in \{-1, 1\}$, and new data points x are classified according to the sign of $y(\vec{x})$. The perpendicular distance to the closest point x_n from the data set is called the margin, and the goal of the algorithm is to optimize parameters w and b for the maximum distance. The maximum margin optimization is mathematically given by:

$$arg \max_{w,b} \left\{ \frac{1}{\|w\|} \min_n [t_n(w^T \phi(\vec{x}) + b)] \right\} \tag{41}$$

Further details on the solution of this optimization problem are provided in [220].

5.3 Simulation parameters and assumptions

In supervised pattern classification methods such as LDA and SVM, the pattern classifier is trained using a set of training feature vectors. Its performance is evaluated by classifying the feature vectors from the test dataset (which is normally different from the training dataset). First, the assessment of the capability of LDA and SVM in separating between two classes is performed in the simplest characterization problem between a signal without multipath echoes and an indoors reception. Afterwards, a simulation that involves three classes is presented.

To determine the optimal combination of features all possible combinations of feature pairs (10 in total) are produced. Therefore, 10 datasets are "fitted" to generate 10 different models per pair. The most accurate classification result should then determine the optimal combination of features. In the multi-class experimentation triplets of features are also combined and tested for all possible combinations (10).

The 'No-MP' – 'Indoors/NLoS' classification requires the generation of the corresponding training datasets. Again, training datasets were numerically simulated under very simple numerical assumptions about the physical conditions in which the antenna receives the signals and the propagation of the signal through the RF-frontend components. Starting with the 'No-MP' label it is assumed that any user in open-sky conditions shall be able to receive 6-11 signals from different satellites, and no MP echoes are affecting the composite signal. The signal however includes WGN to better approximate realistic conditions, but with average $SNR \approx 10\ dB$. On the other hand, the 'Indoors/NLoS' training signals are generated with the simulation parameters shown below:

Table 5 - Training set simulation parameters for indoor/NLoS L1 C/A signal reception

NLoS conditions (Indoors) simulation parameters:

- Number of unique (transmitted from a single satellite) received signals present is defined to be 1
- Number of (NLoS) echoes J per received signal is randomly chosen using uniformly distributed pseudorandom integer generator from MATLAB with min value $J_{min} = 20$ and max value $J_{max} = 50$ per unique signal.
- Path amplitude $A_i a_{i_j} \in [0, 2]$ for $j \in [1, J]$ is random (uniformly distributed) for each NLoS echo (Here is it assumed that one signal is generated with path amplitude of 1 and all the echoes are added to it).
- Δf_j is random (uniformly distributed) in the interval +/-300Hz in the IF LoS carrier freq. for each echo
- $\varphi_j \in [0,2\pi)$ is random (uniformly distributed) for each NLoS echo
- $\tilde{c}_{i_j}$ is introduced with a random codephase (uniformly distributed pseudorandom integer generator from MATLAB) for each NLoS echo

Feature samples are computed for every one of the studied features to train the LDA and SVM classification models. The 'Indoors/NLoS' training set is composed of 300 samples while for the 'No MP' class 1200 samples were used (200 per number of LoS). Both training sets (in feature pairs or triplets) were pre-processed before being input to classifiers by removing the outliers using the "rmoutliers" function of MATLAB and maintaining values that are above the 10th and below the 90th percentile. The data flow is shown in the next figure:

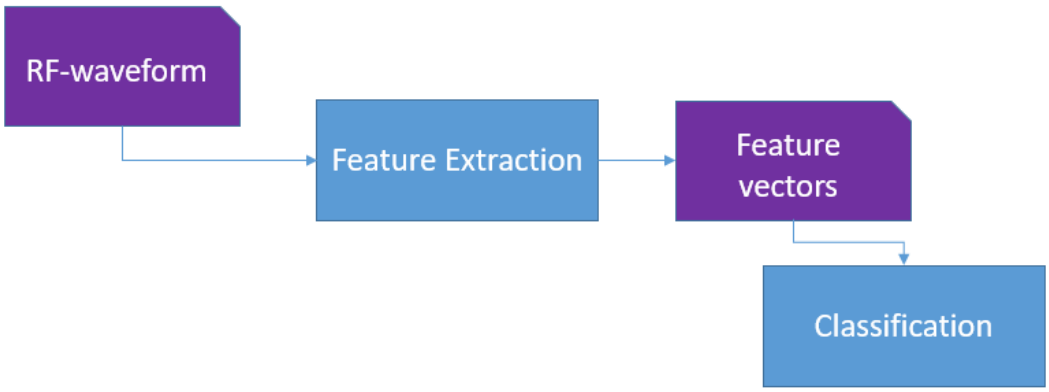


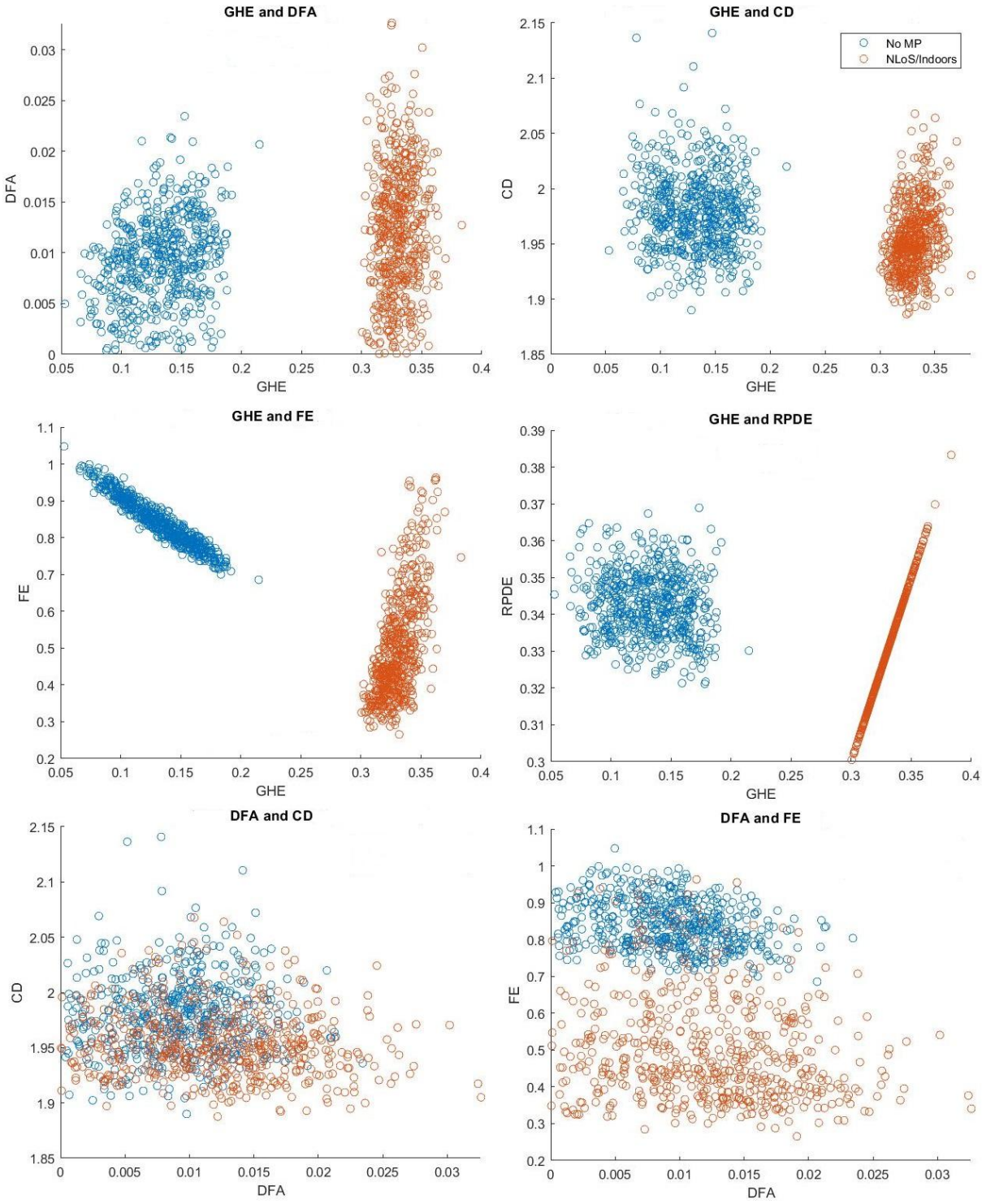
Figure 31 - Feature-based classification implementation data flow

As for the measurement samples, 1200 data points were simulated from a different implementation than the one for producing the training samples from which 600 samples were generated from the 'Indoors/NLoS' simulation parameters (shown in Table 5) and the 600 for the corresponding 'No MP' configuration.

Table 6 - Measurement set simulation parameters for indoor/NLoS L1 C/A signal reception

NLoS conditions (Indoors) simulation parameters:
• Number of unique (transmitted from a single satellite) received signals present is defined to be within $[1,\ 3]$
• Number of (NLoS) echoes J per received signal is randomly chosen using uniformly distributed pseudorandom integer generator from MATLAB with min value $J_{min} = 7$ and max value $J_{max} = 30$ per unique signal.
• Path amplitude $A_i a_{i_j} \in [0,\ 2]$ for $j \in [1,\ J]$ is random (uniformly distributed) for each NLoS echo (Here is it assumed that one signal is generated with path amplitude of 1 and all the echoes are added to it).
• Δf_j is random (uniformly distributed) in the interval +/-300Hz in the IF LoS carrier freq. for each echo
• $\varphi_j \in [0,2\pi)$ is random (uniformly distributed) for each NLoS echo
• $\tilde{c}_{i_j}$ is introduced with a random codephase (uniformly distributed pseudorandom integer generator from MATLAB) for each NLoS echo

Each feature is calculated using as input the measurement samples. For every possible sampled feature pair, a scatter plot was generated to show the distribution of the two simulated environments in feature space:


GHE and DFA
DFA
GHE
GHE and CD
CD
GHE
No MP
NLoS/Indoors
GHE and FE
FE
GHE
GHE and RPDE
RPDE
GHE
DFA and CD
CD
DFA
DFA and FE
FE
DFA

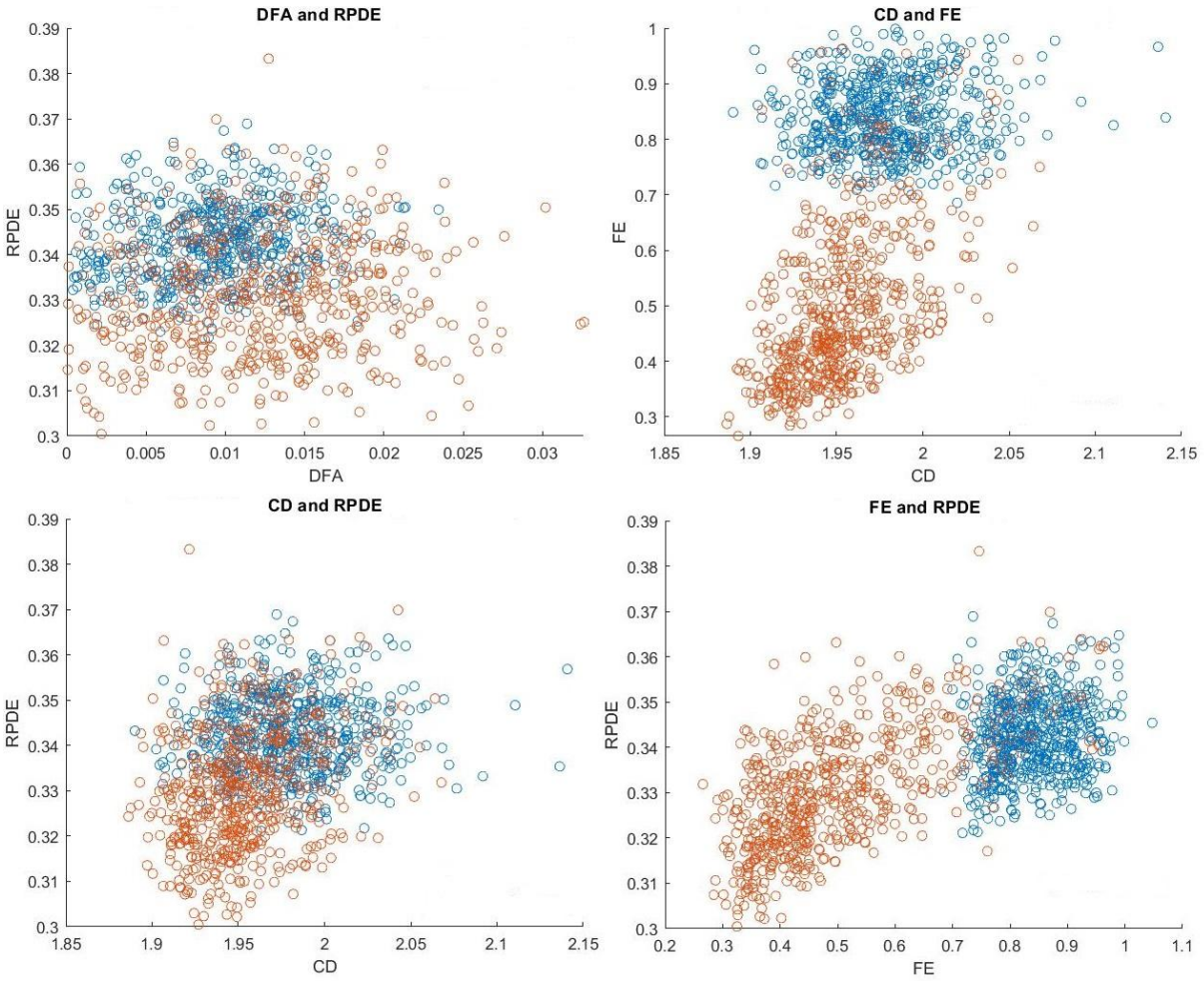


Figure 32 - Measurement samples in 2D feature space

5.4 Classification results and discussion

Performance is assessed through the confusion charts that are produced from MATLAB. The confusion matrix (or chart) is a visual tool that assesses the performance of the prediction model in a tabular manner. The generic structure of a confusion matrix for a binary classification problem is shown in the next table:

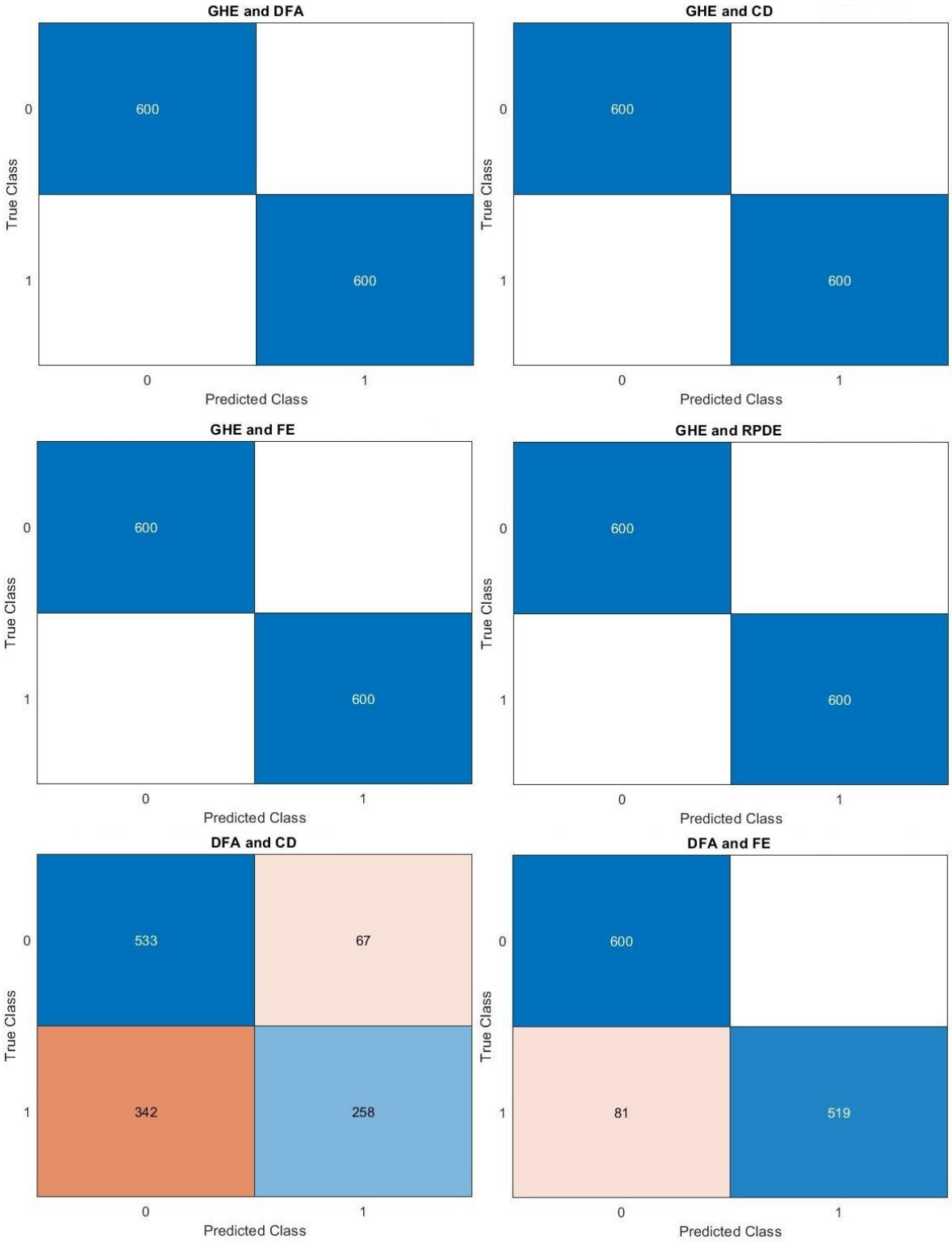
Table 7 - Confusion matrix structure

	Actual (True/False)	
Predicted (Positive/Negative)	True Positive [TP]	False Positive (Type I error) [FP]
	False Negative (Type II error) [FN]	True Negative [TN]

Each entry in a confusion matrix denotes the number of predictions made by the model depending on correct or incorrect classification outcome. For the predicted result of a classifier the Positive/Negative characterization holds while the True/False refers to the actual class of the data. The False Positive (FP) and False Negative (FN) represent the Type I and Type II statistical errors respectively.

In the multi-class case the confusion matrix is converted into a one against all type matrix and essentially reduces to the binary-class confusion matrix, but it requires the construction of one binary confusion matrix per class.

First, by using the standard LDA method the classification predictions between the options of using all possible feature pairs are given next:



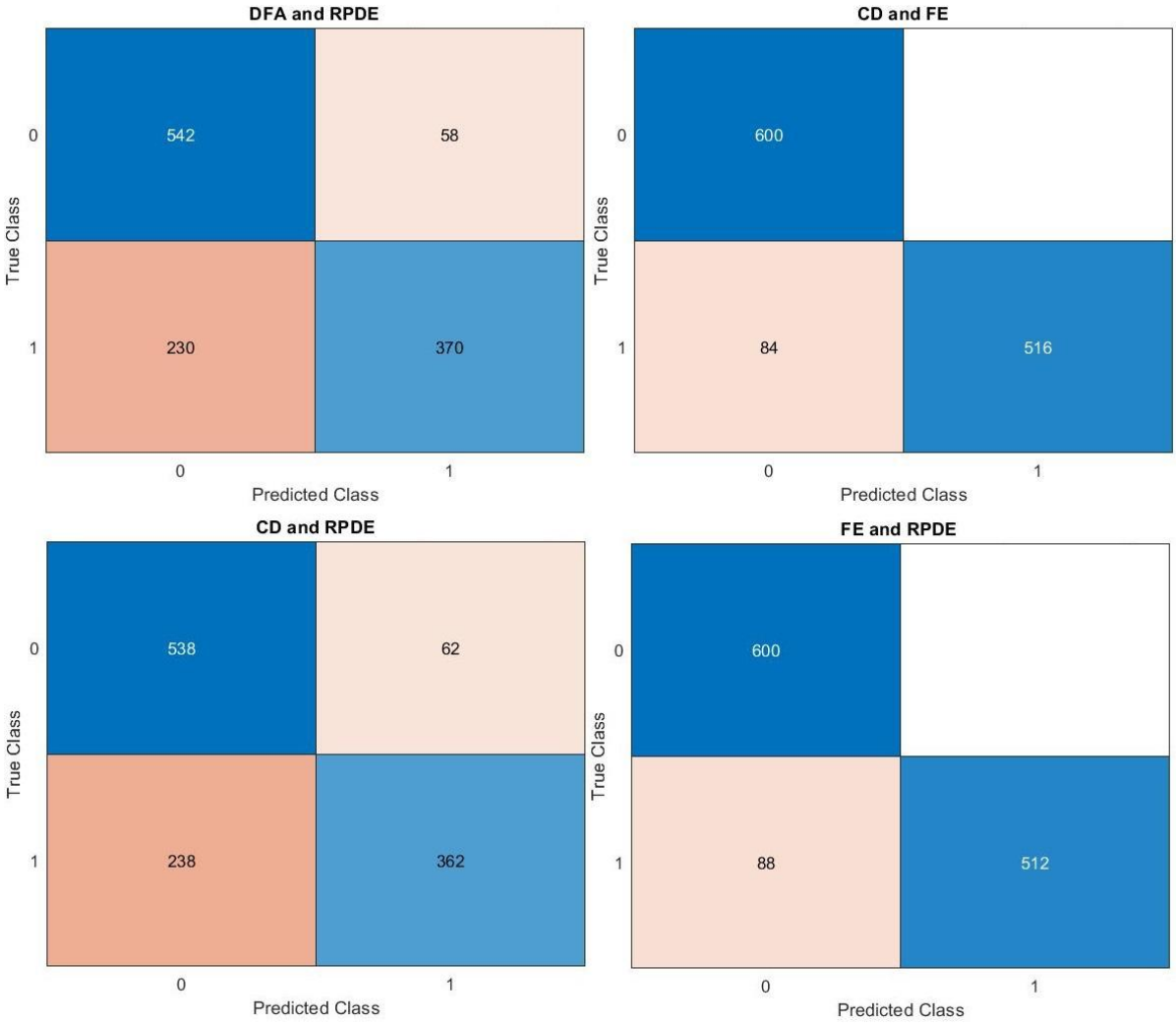
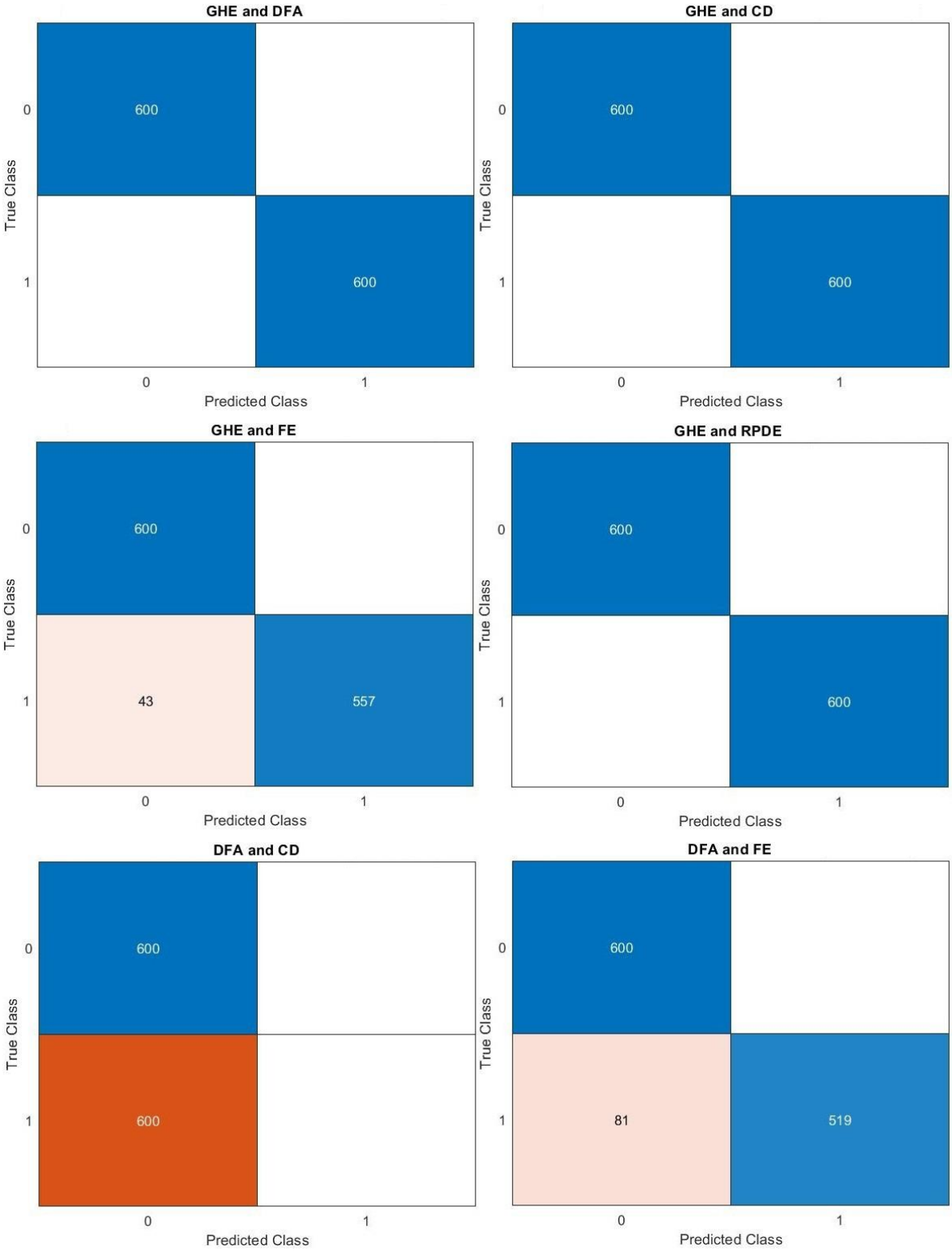


Figure 33 - Confusion charts for binary classification using the LDA method

It is evident that all feature pairs that involve the GHE measurement show excellent performance in this particular binary classification simulation. It is also noticed that DFA-CD, DFA-RPDE and CD-RPDE test cases show a degraded performance.

Next, the corresponding results using the standard SVM algorithm for binary classification are presented:


GHE and DFA
600
600
True Class
Predicted Class
0
1
GHE and CD
600
600
True Class
Predicted Class
0
1
GHE and FE
600
43
557
True Class
Predicted Class
0
1
GHE and RPDE
600
600
True Class
Predicted Class
0
1
DFA and CD
600
600
True Class
Predicted Class
0
1
DFA and FE
600
81
519
True Class
Predicted Class
0
1

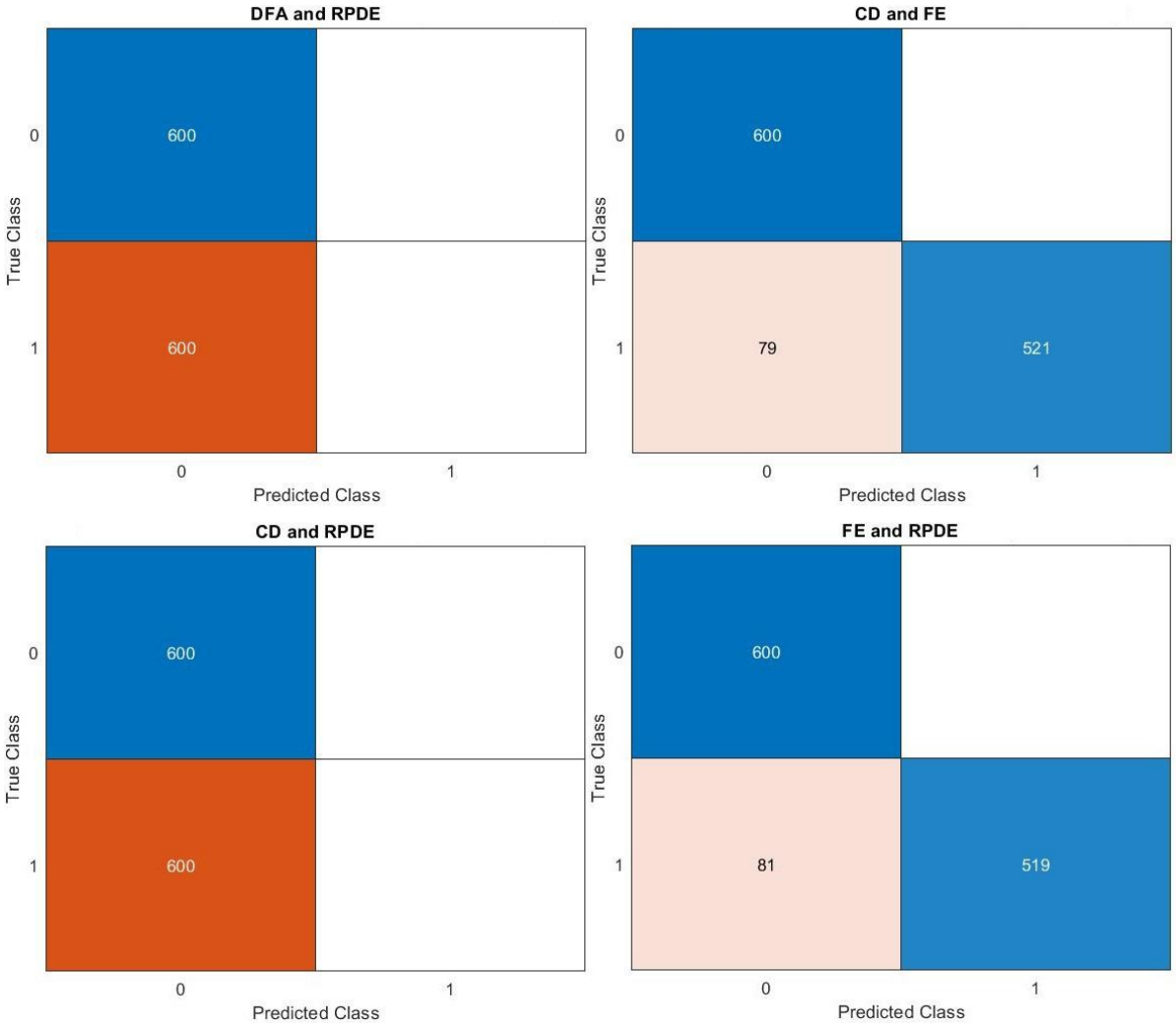


Figure 34 - Confusion charts for binary classification using the SVM method

The SVM classification showed excellent results for all cases that include the GHE measure, except when combined with the FE. Poor performance when using DFA-CD, DFA-RPDE and CD-RPDE feature pairs in the SVM classifier is, again, noticeable and even more pronounced. The numerical simulations have verified that an accurate classification result is possible using two off-the-shelf classifiers for the extreme cases of pure multipath signals and "clean" SIS signals.

Next, a third class label is added to test the performance of the algorithms in a more challenging setup. The third class corresponds to the presence of several MP components on top of the LoS signal. In order to simulate such an intermediate state between the 'No MP' and 'NLoS' class labels the training parameters shown in the next table were selected:

Table 8 - Training set simulation parameters for the "MP" label

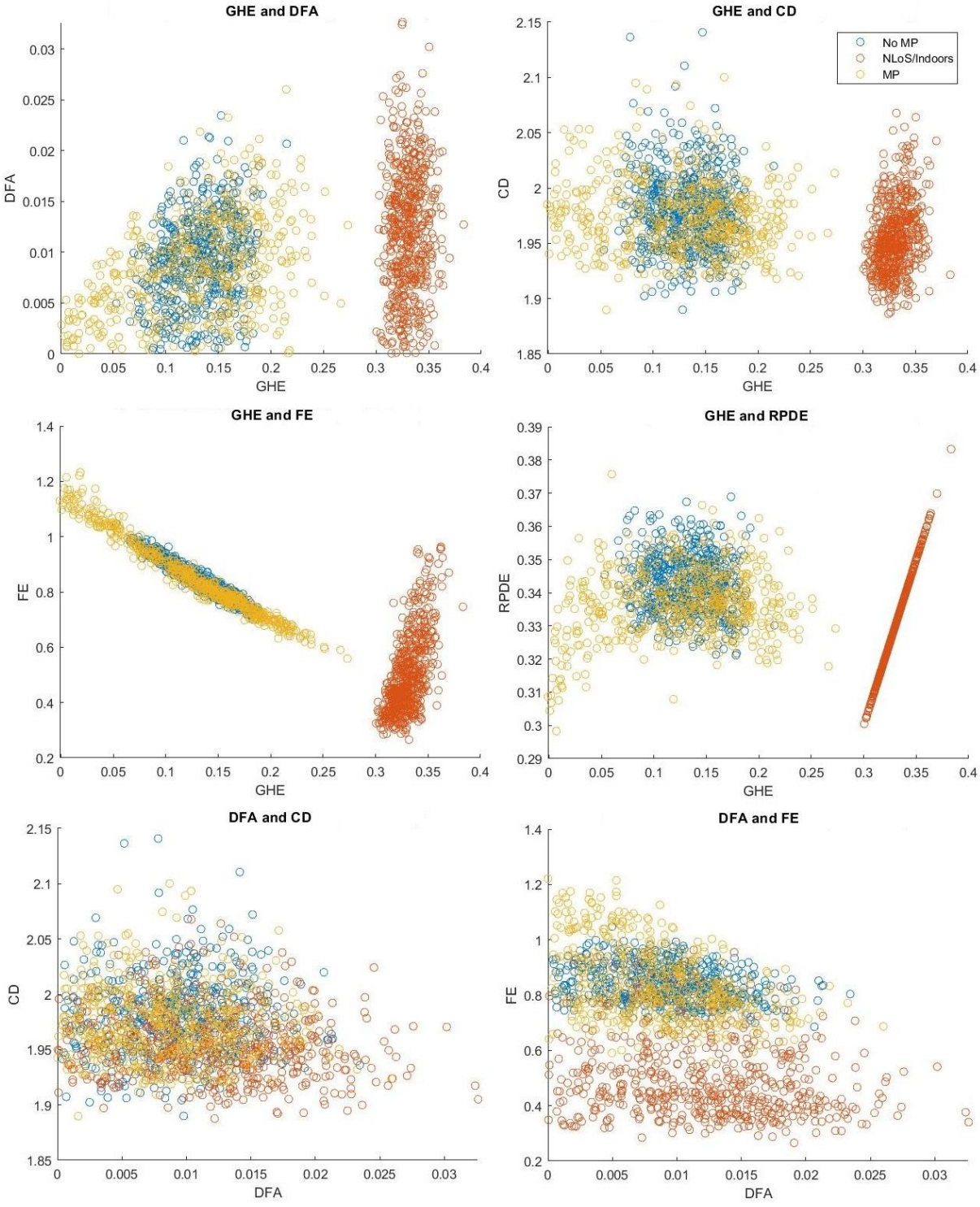
Number of LoS	J_{min} to J_{max} (num NLoS per LoS signal)
1	5-30
2	3-20
3	2-15
4	1-10
5	1-5

The ‘MP’ class was added to the training models without modifying any of the training sets corresponding to the other two classes. For training the ‘MP’ class 1200 training points and 600 testing points were produced. The total samples for the three-class classification test set is 1800, 600 points per each class-label generated from three different signal simulation implementations. Specifically, the implementation employed to produce the corresponding measurement set for the ‘MP’ class considered the following parameters:

Table 9 - Measurement set simulation parameters for the "MP" label

Number of LoS	J_{min} to J_{max} (num NLoS per LoS signal)
1	1-25
2	1-15
3	1-10
4	1-7
5	1-5

The corresponding scatter plots in 2D feature space are compiled next:



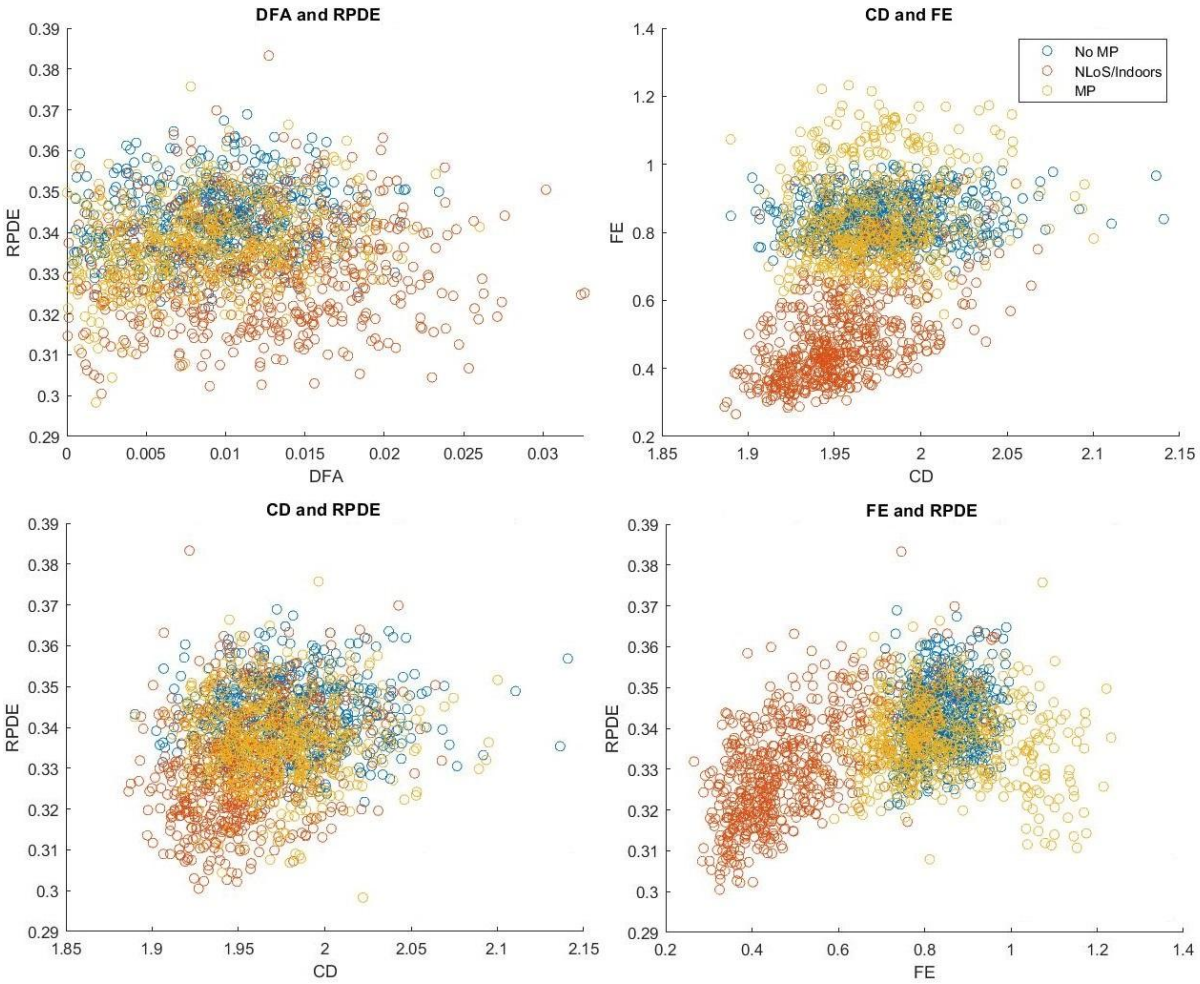
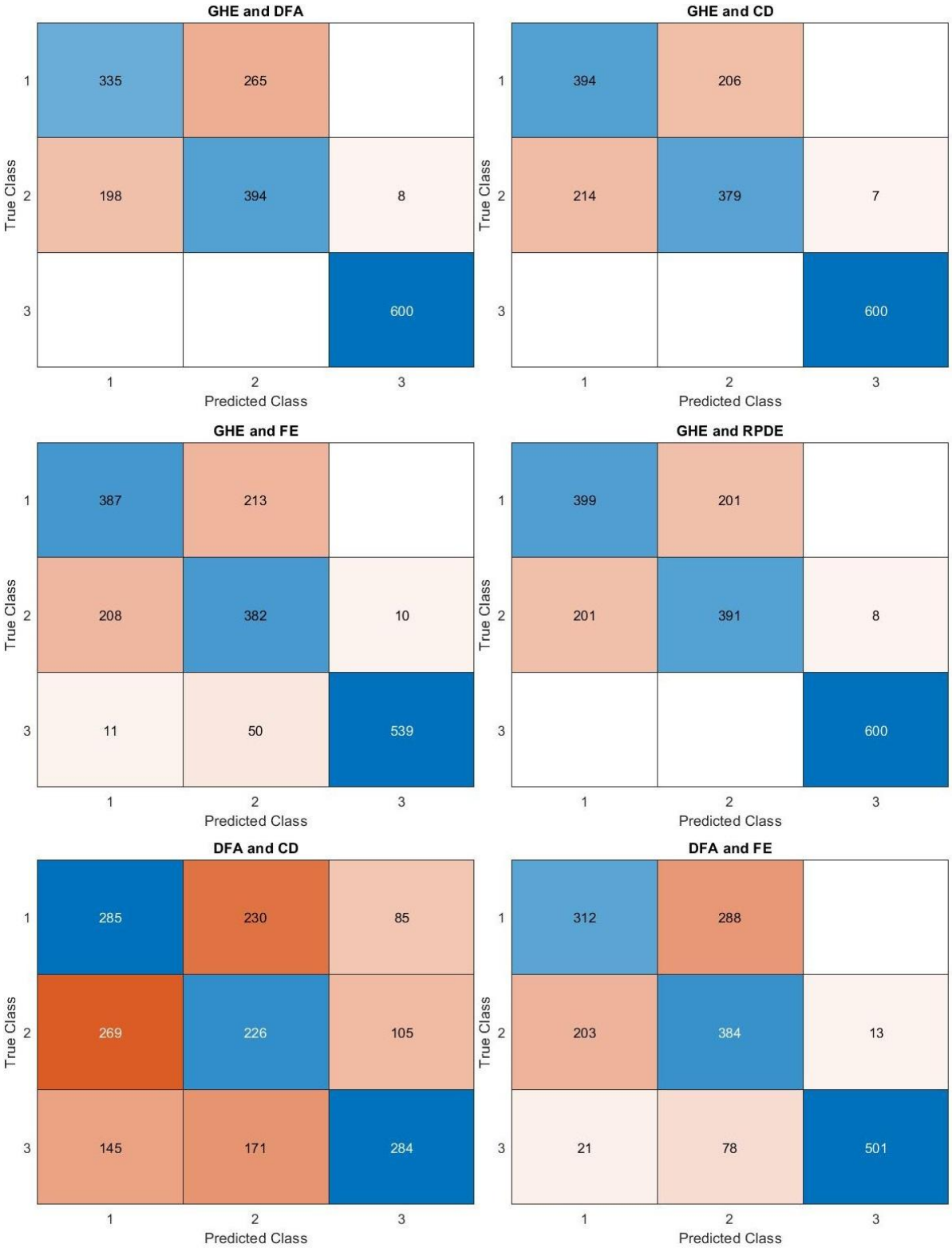


Figure 35 - Measurement samples in 2D feature space for 3 classes

For this three-class classification problem the LDA and SVM methods were not providing a good classification performance. This is mainly because the 'MP' and 'No MP' samples are entangled in almost all possible test cases (see Figure 35).

For this reason, a random forest algorithm was deployed which demonstrated a better performance. In particular, the MATLAB function "fitcensemble(_, _, 'Method', 'Bag')" was used as the random forest implementation. As with the previous simulations, features were combined in pairs to determine the best performing combination. The results are shown next, where class label 1 corresponds to 'No MP', label 2 to 'MP', and 3 to 'Indoors/NLoS'.


GHE and DFA
335
265
198
394
8
600
True Class
Predicted Class
GHE and CD
394
206
214
379
7
600
True Class
Predicted Class
GHE and FE
387
213
208
382
10
11
50
539
True Class
Predicted Class
GHE and RPDE
399
201
201
391
8
600
True Class
Predicted Class
DFA and CD
285
230
85
269
226
105
145
171
284
True Class
Predicted Class
DFA and FE
312
288
203
384
13
21
78
501
True Class
Predicted Class

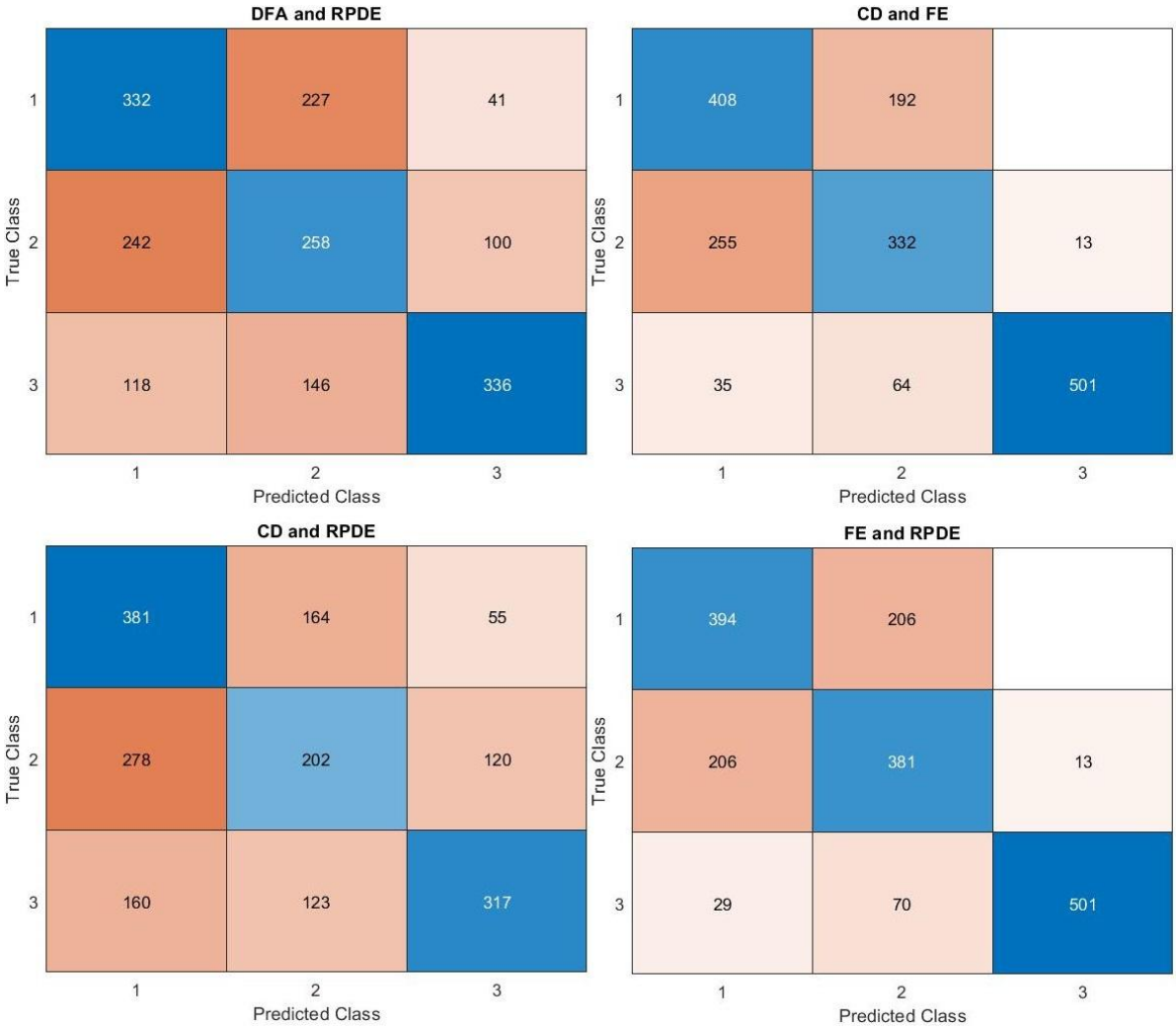
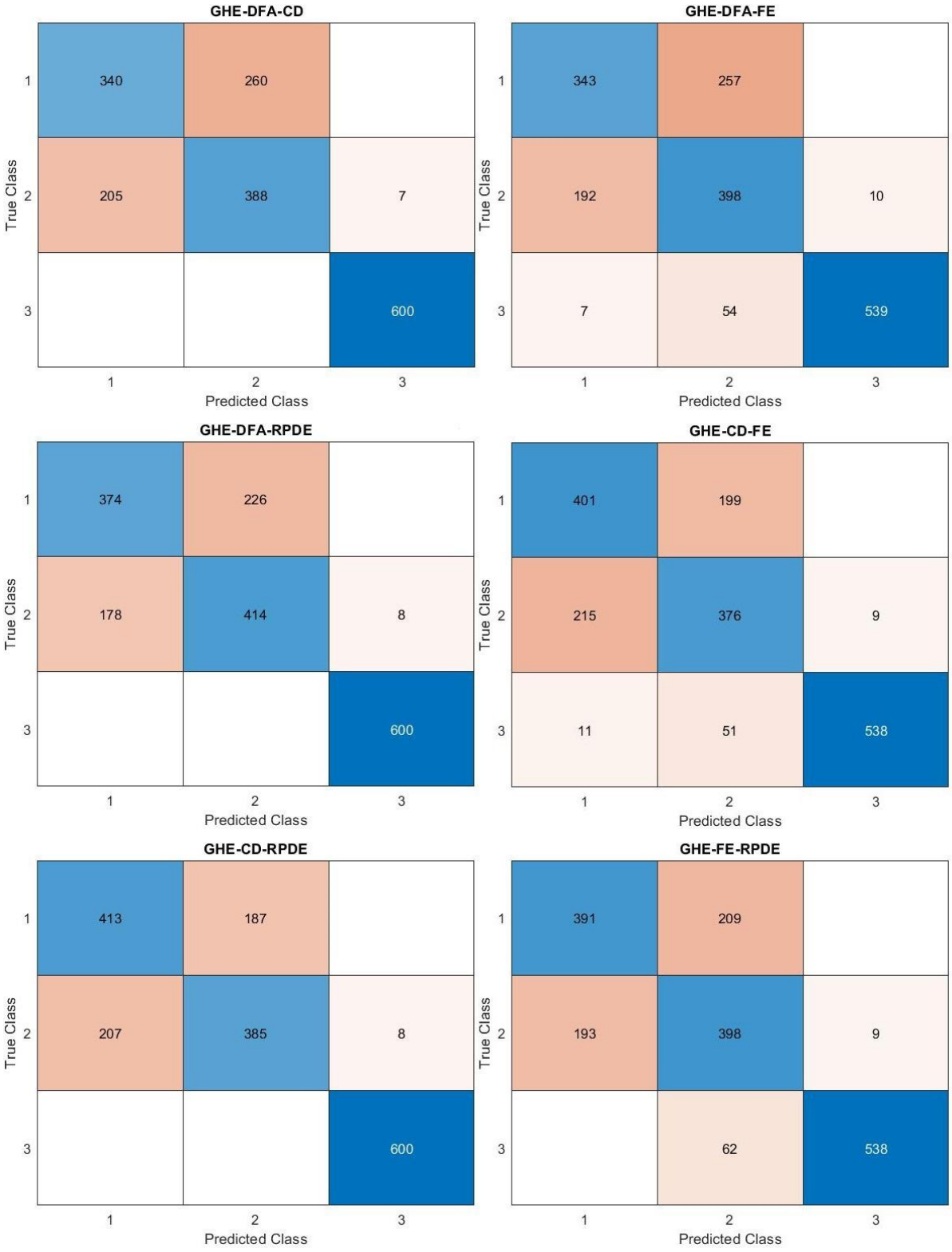


Figure 36 - Confusion charts for 3-label classification using the Random Forest method

The results depicted in Figure 36 and Table 12 demonstrate a rather poor overall performance when a third class is added to the total sample. In particular, the performance appears to be good in the majority of the feature pairs only for the prediction of the simulated NLoS (class 3) signal mixture, but not predicting the 'No MP' or 'MP' class points. For this reason the case where the employment of a third feature was tested, again for all possible triplet combinations (i.e. 10 cases):


GHE-DFA-CD
340
260
205
388
7
600
True Class
Predicted Class
GHE-DFA-FE
343
257
192
398
10
7
54
539
GHE-DFA-RPDE
374
226
178
414
8
600
GHE-CD-FE
401
199
215
376
9
11
51
538
GHE-CD-RPDE
413
187
207
385
8
600
GHE-FE-RPDE
391
209
193
398
9
62
538

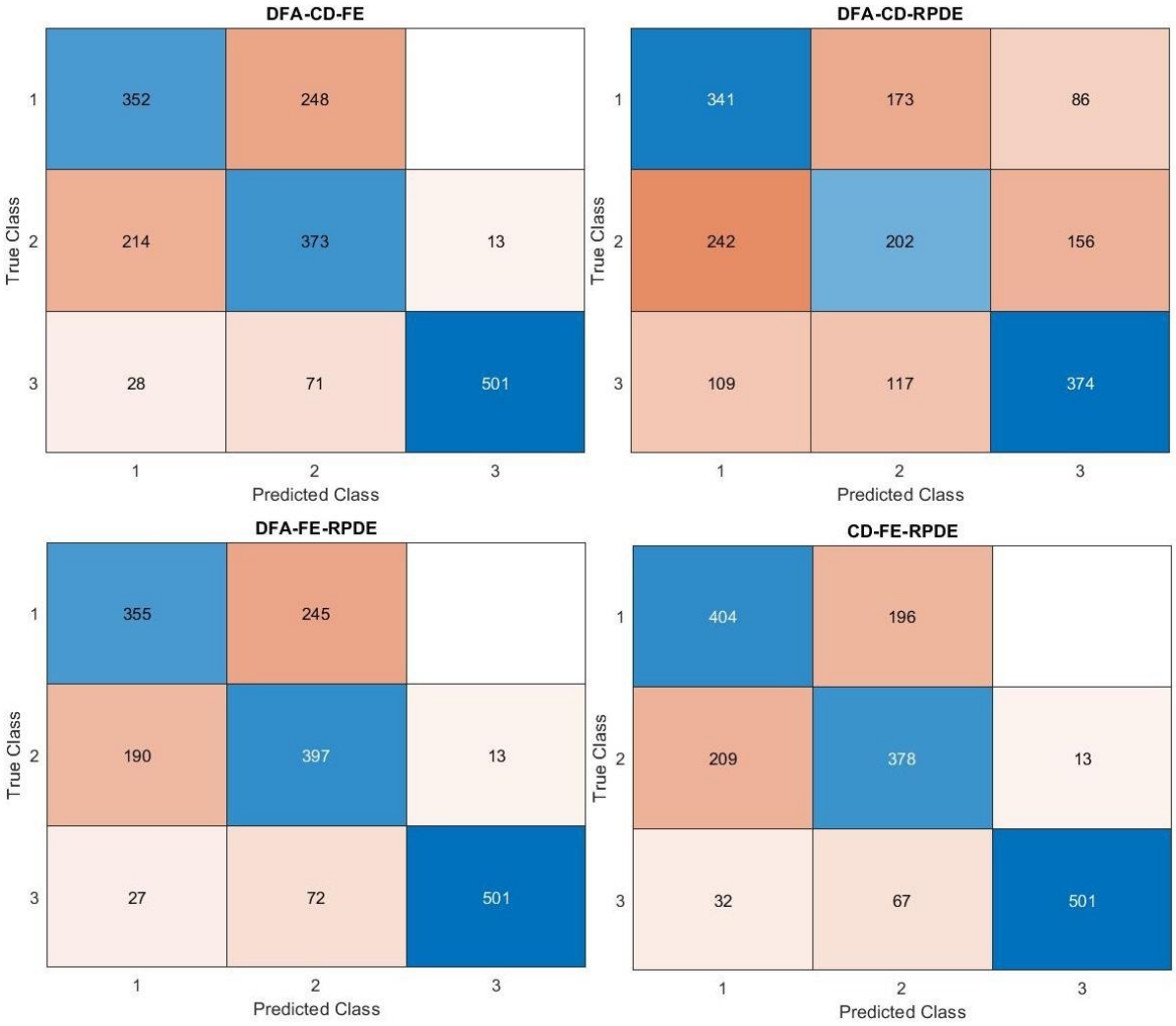


Figure 37 - Confusion charts for 3-label classification and a combination of 3 features using the Random Forest method

Even after adding a third feature the final performance (according to Figure 37 and Table 13) did not improve significantly but it is worth noticing that combinations involving the GHE and RPDE metrics showed overall a slightly better performance than the rest, especially in predicting classes 1 and 2, while also being very efficient in detecting/predicting NLoS measurements. Possible improvements on the methodology presented to increase performance are summarized in the next chapter.

5.5 Validation of classification methods

In the final section of this chapter, three metrics are deployed to assess further the validity of the classification models. These metrics deploy quantities defined in the confusion matrix:

- Accuracy (all correct) which is computed as (TP + TN) / (TP + TN + FP + FN)
- Sensitivity (true positives / all actual positives) which is computed as TP / (TP + FN)
- Specificity (true negatives / all actual negatives) which is computed as TN / (TN + FP)

Figure 33 values of the confusion matrix metrics follow in the next table:

Table 10 - Confusion matrix metric values for binary LDA classification

	Accuracy	Sensitivity	Specificity
GHE-DFA	100%	100%	100%
GHE-CD	100%	100%	100%
GHE-FE	100%	100%	100%
GHE-RPDE	100%	100%	100%
DFA-CD	66%	61%	79%
DFA-FE	93%	88%	100%
DFA-RPDE	76%	70%	86%
CD-FE	93%	87%	100%
CD-RPDE	75%	69%	85%
FE-RPDE	92%	87%	100%

Next, based on the results depicted in Figure 34, the corresponding values of the confusion matrix metrics are given:

Table 11 - Confusion matrix metric values for binary SVM classification

	Accuracy	Sensitivity	Specificity
GHE-DFA	100%	100%	100%
GHE-CD	100%	100%	100%
GHE-FE	96%	93%	100%
GHE-RPDE	100%	100%	100%
DFA-CD	50%	50%	NA
DFA-FE	93%	88%	100%
DFA-RPDE	50%	70%	NA
CD-FE	93%	88%	100%
CD-RPDE	50%	50%	NA
FE-RPDE	93%	88%	100%

Figure 36 values for accuracy, sensitivity and specificity follow:

Table 12 - Confusion matrix metric values for three-class Random Forest classification

	Class	Accuracy	Sensitivity	Specificity
GHE-DFA	1	74%	63%	79%
	2	74%	60%	82%
	3	99%	98%	100%
GHE-CD	1	76%	65%	83%
	2	76%	65%	82%
	3	99%	99%	100%
GHE-FE	1	76%	64%	82%
	2	72%	59%	81%
	3	96%	98%	95%
GHE-RPDE	1	77%	66%	83%
	2	77%	66%	82%
	3	99%	98%	100%
DFA-CD	1	59%	41%	71%
	2	57%	36%	68%
	3	72%	60%	76%
DFA-FE	1	71%	58%	77%
	2	67%	51%	79%
	3	93%	97%	92%
DFA-RPDE	1	65%	48%	76%
	2	60%	41%	71%
	3	77%	70%	80%
CD-FE	1	73%	58%	82%
	2	71%	56%	78%
	3	94%	97%	92%
CD-RPDE	1	63%	46%	77%
	2	62%	41%	69%
	3	74%	64%	78%
FE-RPDE	1	75%	62%	82%
	2	72%	58%	81%
	3	93%	97%	92%

Figure 37 result table:

Table 13 - Confusion matrix metric values for three-class Random Forest classification (combinations of 3 features)

	Class	Accuracy	Sensitivity	Specificity
GHE-DFA-CD	1	74%	62%	79%
	2	73%	60%	81%
	3	99%	99%	100%
GHE-DFA-FE	1	74%	63%	79%
	2	71%	56%	81%
	3	96%	98%	95%
GHE-DFA-RPDE	1	76%	67%	82%
	2	77%	64%	84%
	3	99%	98%	100%
GHE-CD-FE	1	76%	64%	83%
	2	73%	60%	81%
	3	96%	98%	95%
GHE-CD-RPDE	1	78%	66%	84%
	2	77%	67%	82%
	3	99%	98%	100%
GHE-FE-RPDE	1	77%	67%	83%
	2	73%	59%	82%
	3	96%	98%	95%
DFA-CD-FE	1	72%	59%	79%
	2	69%	54%	79%
	3	93%	97%	92%
DFA-CD-RPDE	1	66%	49%	76%
	2	61%	41%	69%
	3	74%	60%	72%
DFA-FE-RPDE	1	74%	62%	80%
	2	71%	55%	81%
	3	93%	97%	92%
CD-FE-RPDE	1	75%	62%	83%
	2	73%	59%	81%
	3	93%	97%	92%

6 Conclusions and future work

The performance of an independent (in standalone mode) GNSS receiver is predominantly driven by the signal processing strategy - usually pre-selected by the manufacturer - and its capabilities to mitigate unwanted effects that signals experience from transmission to reception. The scope of this study is to research signal processing algorithms that detect and characterize the underlying multipath propagation environment by studying the classification performance of different multipath-afflicted signal waveforms. Context awareness plays a key role in GNSS integrity applications [222] such as in intelligent transportation systems (ITS)s [223], and spoofing detection [62] by providing information on the reliability of GNSS measurements. Integrity related information could activate further mitigation actions in the receiver side such as adaptation of signal processing (tracking bandwidth etc.). A similarly designed component could be utilized in snapshot receivers and emerging cloud integrated systems [224].

Several feature extraction methods were presented and tested through numerical simulations. The objective of selecting and studying these methods was based on two factors; their suitability to capture multipath patterns on a basic RF GNSS signal (pre-correlator stage) such as the GPS L1 C/A, and their combined employment to be utilised in effective classification of the multipath-afflicted waveform presumably caused by the corresponding multipath propagation environment. The presented numerical simulations showed excellent performance in classifying between a waveform that is composed from pure LI C/A spreading codes and a waveform with tens of multipath echoes (and no 'strong' LoS component present). In particular, all possible feature pairs that include the GHE metric could classify the two waveform types with 100% accuracy, sensitivity and specificity (see Figures 33, 34 and Tables 7 and 8).

When a third class of generic (intermediate type of) multipath-afflicted signals is added to the problem, the classification accuracy between 'No MP' or 'MP' labels was at best at 77% for the GHE-RPDE using a Random Forest classifier (Table 11) as LDA and SVM showed a poor performance. Towards the goal of improving performance, all possible combinations of triplets of features were employed as inputs to the selected classifiers (Table 12). However, showing no improvement compared to input feature-pairs.

The most accurate and computationally efficient method for feature extraction is shown to be the GHE (Hurst exponent). Evidence shows that in certain scenarios GHE could be employed to characterize the multipath environment from the RF signal even without requiring an additional classifier. The majority of the rest of feature extraction methods are rather computationally expensive so their use might be limited in certain applications.

Given the numerical nature of the simulations the outcome of the research is a preliminary analysis of the multipath-afflicted waveforms and the detection performance using the selected feature extraction methods. Although iterations (in a Monte Carlo simulation setting) are relatively low to cover for all different realistic scenarios, evidence shows that the studied feature techniques could potentially accommodate the context awareness capacity under certain SNR conditions.

To improve classification performance, one proposal is to further investigate the optimal configuration of the feature extraction methods for the specific signal. Another case could be the employment of batch measurements combining several 1 ms samples. Without a matched-filter (cross-correlation) using local PRN replicas and therefore distinguishing one signal per satellite for analysis, it is unavoidable that ambiguous and challenging settings are anticipated. Depending on the number and properties of satellite signals in comparison to the number and properties of their resolvable echoes multiple physical

conditions might cause the same signal waveform. Subtle differences could possibly give extra information through the feature algorithms, and quite probably more information might be required.

One of the most noticeable limiting factors in real conditions is that the central frequencies of different GNSSs might coincide such as GPS L1 band and Galileo E1 therefore affecting the results. Also the RF processing and quality of the components plays an important role as it defines the SNR factor. Several points can be listed on potential improvements:

- Experimentation using real data for L1/CA which implies a careful selection of the antenna-frontend pair (and a GNSS sampler/baseband unit).
- Optimal tuning of feature extraction algorithms to improve performance and reduce computational cost/runtime with the prospect of employing it into a real GNSS system.
- Assessing performance using different sampling rates for the ADC output and/or different step/increment lengths. The data-window could also be adapted.
- The use of signal decompositions as a pre-processing step on one or several bands/signals and study potential improvement in feature selection.

Moreover, the general aims of this work is to establish a framework on the methodology to detect/classify MP propagation with potential extensions to:

- Extension to several classes, according to the different multipath propagation environments such as the ones widely used: Open-air, Rural, Urban, Indoors etc.
- Extension of the considered classes by including Ionospheric Scintillation (which manifests itself in a similar manner as a multipath effect) and/or structural interference (RFI).
- RFI and Scintillation could be accommodated and the classifier shall be designed to perform multi-label multiclass classification.
- Generalization to other systems and signals (modern GNSS signals).

References

[1] Teunissen and Montenbruck, *Handbook of Global Navigation Systems*. 2018.

[2] S. M. Sussman, "A Matched Filter Communication System for Multipath Channels," *IRE Trans. Inf. Theory*, vol. 6, no. 3, pp. 367–373, 1960, doi: 10.1109/TIT.1960.1057565.

[3] I. S. Grant and W. R. Phillips, *Electromagnetism*, 2nd editio. Wiley, 1991.

[4] J. D. Jackson, *Classical Electrodynamics*, 3rd editio. John Wiley & Sons, Inc., 1999.

[5] L. M. Aŕevalo Aguilar, C. Robledo-Śanchez, M. L. Arroyo Carrasco, and M. M. Ḿendez Otero, "The principle of superposition for waves: The amplitude and phase modulation phenomena," *Appl. Math. Inf. Sci.*, vol. 6, no. 2, pp. 307–315, 2012.

[6] S. R. Saunders and A. A. Aragón-Zavala, *Antennas and Propagation for Wireless Communication Systems*, 2nd Editio. John Wiley & Sons, Ltd, 2007.

[7] J. S. Lee and L. E. Miller, *CDMA Systems Engineering Handbook*. Artech House, 1998.

[8] D. Torrieri, *Principles of Spread-Spectrum Communication Systems*. Cham: Springer International Publishing, 2018.

[9] T. Yücek and H. Arslan, "Time dispersion and delay spread estimation for adaptive OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1715–1722, 2008, doi: 10.1109/TVT.2007.909247.

[10] S. M. Rana, "A Study of Multipath Propagation and Doppler Effect at 24GHz ISM band," 2020.

[11] I. Emre Telatar and D. Tse, "Capacity and mutual information of broadband multipath fading channels," *IEEE Int. Symp. Inf. Theory - Proc.*, vol. 46, no. 4, p. 395, 1998, doi: 10.1109/ISIT.1998.709000.

[12] J. M. Tranquilla and J. P. Carr, "GPS Multipath Field Observations at Land and Water Sites," *Navigation*, vol. 37, no. 4, pp. 393–414, Dec. 1990, doi: 10.1002/j.2161-4296.1990.tb01564.x.

[13] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces (Artech House Radar Library)*. Artech House, 1987.

[14] K. Haspert and M. Tuley, "Comparison of predicted and measured multipath impulse responses," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1696–1709, 2011, doi: 10.1109/TAES.2011.5937259.

[15] C. Rost and L. Wanninger, "Carrier phase multipath mitigation based on GNSS signal quality measurements Christian Rost and Lambert Wanninger," *J. Appl. Geod.*, vol. 3, no. 2, pp. 81–87, 2009, doi: 10.1515/JAG.2009.009.

[16] P. Y. Georgiadou and A. Kleusberg, "No Title," *Manuscripta Geod.*, vol. 13, no. 3, pp. 172–179, 1988.

[17] X. Wang and A. Høst-Madsen, "Group-blind multiuser detection for uplink CDMA," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1971–1984, 1999, doi: 10.1109/49.806826.

[18] K. Mahmood, S. M. Asad, M. Moinuddin, A. Zerguine, and L. Cheded, "Multiple access interference in MIMO-CDMA systems under Rayleigh fading: statistical characterization and applications," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–18, 2016, doi: 10.1186/s13634-016-0338-y.

[19] R. H. Mahadevappa and J. G. Proakis, "Mitigating multiple access interference and intersymbol interference in uncoded CDMA systems with chip-level interleaving," *IEEE Trans. Wirel. Commun.*, vol. 1, no. 4, pp. 781–791, 2002, doi: 10.1109/TWC.2002.804163.

[20] E. D. Kaplan, *Understanding GPS: Principles and Applications*. Artech House Publishers, 1996.

[21] B. Sklar, “Rayleigh fading channels in mobile digital communication systems .I. Characterization,” *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 90–100, Jul. 1997, doi: 10.1109/35.601747.

[22] C. Xiao, Y. R. Zheng, and N. C. Beaulieu, “Statistical simulation models for Rayleigh and Rician fading,” *IEEE Int. Conf. Commun.*, vol. 5, pp. 3524–3529, 2003, doi: 10.1109/icc.2003.1204109.

[23] C. Pimentel, T. H. Falk, and L. Lisbôa, “Finite-state Markov modeling of correlated rician-fading channels,” *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1491–1501, 2004, doi: 10.1109/TVT.2004.832413.

[24] T. Eltoft, “The Rician inverse Gaussian distribution: A new model for non-Rayleigh signal amplitude statistics,” *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1722–1735, 2005, doi: 10.1109/TIP.2005.857281.

[25] R. H. Clarke, “A Statistical Theory of Mobile-Radio reception,” 1968, doi: 10.1002/j.1538-7305.1968.tb00069.x.

[26] M. A. Skima, H. Ghariani, and M. Lahiani, “A multi-criteria comparative analysis of different Rayleigh fading channel simulators,” *AEU - Int. J. Electron. Commun.*, vol. 68, no. 6, pp. 550–560, 2014, doi: 10.1016/j.aeue.2014.01.001.

[27] K. A. Norton, L. E. Voglert, W. V Mansfieldt, and P. J. Shortt, “The Probability Distribution of the Amplitude of a Constant Vector Plus a Rayleigh- Distributed Vector,” in *IRE Proceedings*, 1955, pp. 1354–1361, doi: 10.1109/JRPROC.1955.277948.

[28] M. Nakagami, “The m-Distribution—A General Formula of Intensity Distribution of Rapid Fading,” 1960, pp. 3–6, doi: 10.1016/B978-0-08-009306-2.50005-4.

[29] C. Loo, “A Statistical Model for a Land Mobile Satellite,” *IEEE Trans. Veh. Technol.*, vol. V, no. 3, pp. 122–127, 1985.

[30] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, “The Land Mobile Satellite Communication and Channel Model,” *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, 1991, doi: 10.1109/25.289418.

[31] V. C. Lima Filho and A. Moraes, “Modeling multifrequency GPS multipath fading in land vehicle environments,” *GPS Solut.*, vol. 25, no. 1, pp. 1–14, 2021, doi: 10.1007/s10291-020-01040-8.

[32] G. W. Hein, M. Paonni, V. Kropp, and A. Teuber, “GNSS Indoors - Fighting the Fading part I,” *GNSS Solut.*, pp. 43–52, 2008.

[33] A. A. M. Saleh and R. Valenzuela, “A Statistical . Model for Indoor Multipath Propagation,” *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, 1987, doi: 10.1109/JSAC.1987.1146527.

[34] G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, “A statistical model of urban multipath propagation,” *IEEE Trans. Veh. Technol.*, vol. 21, no. 1, pp. 1–9, Feb. 1972, doi: 10.1109/T-VT.1972.23492.

[35] G. W. Hein, A. Teuber, H. J. Thierfelder, and A. Wolfe, “GNSS Indoors, Fighting the Fading, Part 2,” *InsideGNSS*, no. june, pp. 47–53, 2008.

[36] G. Tzeremes and C. G. Christodoulou, “Use of Weibull distribution for describing outdoor multipath fading,” *IEEE Antennas Propag. Soc. AP-S Int. Symp.*, vol. 1, pp. 232–235, 2002, doi: 10.1109/aps.2002.1016291.

[37] F. Van Graas and A. Soloviev, “Precise Velocity Estimation Using a Stand-Alone GPS Receiver,” *Navigation*, vol. 51, no. 4, pp. 283–292, Dec. 2004, doi: 10.1002/j.2161-4296.2004.tb00359.x.

[38] B. Aminian, V. Renaudin, D. Borio, and G. Lachapelle, “Indoor doppler measurement and velocity characterization using a reference-rover receiver,” *23rd Int. Tech. Meet. Satell. Div. Inst. Navig. 2010, ION GNSS 2010*, vol. 4, no. September, pp. 3069–3079, 2010.

[39] S. Satyanarayana, “GNSS Channel Characterization and Enhanced Weak Signal Processing,” University of Calgary, 2011.

[40] S. Priyadarshi, “A Review of Ionospheric Scintillation Models,” *Surv. Geophys.*, vol. 36, no. 2, pp. 295–324, 2015, doi: 10.1007/s10712-015-9319-1.

[41] M. Nakagami, “Statistical Methods in Radio Wave Propagation,” in *Statistical Methods in Radio Wave Propagation*, 1960, p. 348.

[42] E. J. Fremouw, R. C. Livingston, and D. A. Miller, “On the statistics of scintillating signals,” *J. Atmos. Terr. Phys.*, vol. 42, no. 8, pp. 717–731, Aug. 1980, doi: 10.1016/0021-9169(80)90055-0.

[43] N. Linty and F. Dovis, “An open-loop receiver architecture for monitoring of ionospheric scintillations by means of GNSS signals,” *Appl. Sci.*, vol. 9, no. 12, 2019, doi: 10.3390/app9122482.

[44] H. Kulhandjian and T. Melodia, “Modeling underwater acoustic channels in short-range shallow water environments,” *Proc. 9th ACM Int. Conf. Underw. Networks Syst. WUWNET 2014*, 2014, doi: 10.1145/2671490.2674560.

[45] C. D. Charalambous and N. Menemenlis, “A state-space approach in modeling multipath fading channels via stochastic differential equations,” *IEEE Int. Conf. Commun.*, vol. 7, pp. 2251–2255, 2001, doi: 10.1109/ICC.2001.937056.

[46] J. C. S. Santos Filho and M. D. Yacoub, “Highly accurate -μ approximation to sum of M independent non-identical Ricean variates,” *Electron. Lett.*, vol. 41, no. 6, p. 338, 2005, doi: 10.1049/el:20057727.

[47] Z. Yu, T. T. Tjhung, and C. C. Chai, “PDF of the sum of Rician signals and Gaussian noise,” *IEEE Veh. Technol. Conf.*, vol. 58, no. 1, pp. 21–25, 2003, doi: 10.1109/vetecf.2003.1284970.

[48] G. K. Karagiannidis and S. A. Kotsopoulos, “On the distribution of the weighted sum of L Independent Rician and Nakagami envelopes in the presence of AWGN,” *J. Commun. Networks*, vol. 3, no. 2, pp. 112–119, 2001, doi: 10.1109/jcn.2001.6596855.

[49] J. Hu and N. C. Beaulieu, “Accurate closed-form approximations to Ricean sum distributions and densities,” *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 133–135, 2005, doi: 10.1109/LCOMM.2005.02029.

[50] A. Alimohammad, S. F. Fard, B. F. Cockburn, and C. Schlegel, “Compact Rayleigh and Rician fading simulator based on random walk processes,” *IET Commun.*, vol. 3, no. 8, pp. 1333–1342, 2009, doi: 10.1049/iet-com.2008.0297.

[51] N. C. Beaulieu and X. Jiandong, “A novel fading model for channels with multiple dominant specular components,” *IEEE Wirel. Commun. Lett.*, vol. 4, no. 1, pp. 54–57, 2015, doi: 10.1109/LWC.2014.2367501.

[52] M. R. Bhatnagar, “On the sum of correlated squared κ-μ Shadowed random variables and its application to performance analysis of MRC,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2678–2684, 2015, doi: 10.1109/TVT.2014.2343453.

[53] S. Aja-Fernández and G. Vegas-Sánchez-Ferrero, “Statistical analysis of noise in MRI:

Modeling, filtering and estimation," in *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*, 2016, pp. 1–327.

[54] R. T. Ioannides, T. Pany, and G. Gibbons, "Known Vulnerabilities of Global Navigation Satellite Systems, Status, and Potential Mitigation Techniques," *Proc. IEEE*, vol. 104, no. 6, pp. 1174–1194, 2016, doi: 10.1109/JPROC.2016.2535898.

[55] J. Zidan, E. I. Adegoke, E. Kampert, S. A. Birrell, C. R. Ford, and M. D. Higgins, "GNSS Vulnerabilities and Existing Solutions: A Review of the Literature," *IEEE Access*, vol. 4, pp. 1–1, 2020, doi: 10.1109/access.2020.2973759.

[56] N. G. Ferrara *et al.*, "Combined architecture enhancing multi-dimensional signal quality in GNSS receivers," *InsideGNSS*, vol. 2, no. 1, pp. 54–62, 2016.

[57] K. D. Wesson, J. N. Gross, T. E. Humphreys, and B. L. Evans, "GNSS Signal Authentication Via Power and Distortion Monitoring," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 2, pp. 739–754, 2018, doi: 10.1109/TAES.2017.2765258.

[58] P. D. Groves, H. Martin, K. Voutsis, D. Walter, and L. Wang, "Context Detection , Categorization and Connectivity for Advanced Adaptive Integrated Navigation," no. September, pp. 16–20, 2013.

[59] N. Sokhandan, N. Ziedan, A. Broumandan, and G. Lachapelle, "Context-Aware Adaptive Multipath Compensation Based on Channel Pattern Recognition for GNSS Receivers," *J. Navig.*, vol. 70, no. 5, pp. 944–962, 2017, doi: 10.1017/S0373463317000121.

[60] R. Morales-Ferre, P. Richter, E. Falletti, A. de la Fuente, and E. S. Lohan, "A survey on coping with intentional interference in satellite navigation for manned and unmanned aircraft," *IEEE Commun. Surv. Tutorials*, vol. PP, pp. 1–1, 2019, doi: 10.1109/comst.2019.2949178.

[61] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, "Pre-despreading authenticity verification for GPS L1 C/A signals," *Navig. J. Inst. Navig.*, vol. 61, no. 1, pp. 1–11, 2014, doi: 10.1002/navi.50.

[62] A. Broumandan, R. Siddakatte, and G. Lachapelle, "An approach to detect GNSS spoofing," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 8, pp. 64–75, Aug. 2017, doi: 10.1109/MAES.2017.160190.

[63] A. H. Dodson, X. Meng, and G. W. Roberts, "Adaptive method for multipath mitigation and its application for structural deflection monitoring," *Proc. Int. Symp. Kinematic Syst. Geod. Geomatics Navig.*, no. June, pp. 5–8, 2001.

[64] Y. Xia, S. Pan, X. Meng, W. Gao, and H. Wen, "Robust statistical detection of gnss multipath using inter-frequency C/N0 differences," *Remote Sens.*, vol. 12, no. 20, pp. 1–25, 2020, doi: 10.3390/rs12203388.

[65] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "Blind multipath MIMO channel parameter estimation using the Parafac decomposition," *IEEE Int. Conf. Commun.*, pp. 0–4, 2009, doi: 10.1109/ICC.2009.5198793.

[66] X. Jiang, W. J. Zeng, E. Cheng, and C. R. Lin, "Multipath channel estimation using fast least-squares algorithm," *Proc. - 2011 3rd Int. Conf. Commun. Mob. Comput. C. 2011*, no. 1, pp. 433–436, 2011, doi: 10.1109/CMC.2011.12.

[67] F. Dovis, *GNSS Interference Threats & Countermeasures*. Artech House Publishers, 2015.

[68] M. L. Psiaki and T. E. Humphreys, "GNSS Spoofing and Detection," *Proc. IEEE*, vol. 104, no. 6, pp. 1258–1270, 2016, doi: 10.1109/JPROC.2016.2526658.

[69] S. Bartl, P. Berglez, and B. Hofmann-Wellenhof, "GNSS interference detection, classification and localization using Software-Defined Radio," *2017 Eur. Navig. Conf. ENC 2017*, pp. 159–

169, 2017, doi: 10.1109/EURONAV.2017.7954205.

[70] R. H. Mitch *et al.*, "Signal characteristics of civil GPS jammers," *24th Int. Tech. Meet. Satell. Div. Inst. Navig. 2011, ION GNSS 2011*, vol. 3, pp. 1907–1919, 2011.

[71] T. Kraus, R. Bauemfeind, and B. Eissfeller, "Survey of In-Car Jammers - Analysis and modeling of the RF signals and IF samples (suitable for active signal cancellation)," *24th Int. Tech. Meet. Satell. Div. Inst. Navig. 2011, ION GNSS 2011*, vol. 1, no. September, pp. 430–435, 2011.

[72] L. Zhang, J. Ren, and T. Li, "Time-varying jamming modeling and classification," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3902–3907, 2012, doi: 10.1109/TSP.2012.2193574.

[73] D. Borio, F. Dovis, H. Kuusniemi, and L. Lo Presti, "Impact and Detection of GNSS Jammers on Consumer Grade Satellite Navigation Receivers," *Proc. IEEE*, vol. 104, no. 6, pp. 1233–1245, 2016, doi: 10.1109/JPROC.2016.2543266.

[74] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Čapkun, "On the requirements for successful GPS spoofing attacks," *Proc. ACM Conf. Comput. Commun. Secur.*, pp. 75–85, 2011, doi: 10.1145/2046707.2046719.

[75] C. Günther, "A Survey of Spoofing and Counter-Measures," *Navig. J. Inst. Navig.*, vol. 61, no. 3, pp. 159–177, 2014, doi: 10.1002/navi.65.

[76] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, "GPS vulnerability to spoofing threats and a review of antispoofing techniques," *Int. J. Navig. Obs.*, 2012, doi: 10.1155/2012/127072.

[77] S. F. Bian, Y. F. Hu, C. Chen, Z. M. Li, and B. Ji, "Research on GNSS repeater spoofing technique for fake Position, fake Time & fake Velocity," *IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM*, pp. 1430–1434, 2017, doi: 10.1109/AIM.2017.8014219.

[78] D. Schmidt, K. Radke, S. Camtepe, E. Foo, and M. Ren, "A survey and analysis of the GNSS spoofing threat and countermeasures," *ACM Comput. Surv.*, vol. 48, no. 4, 2016, doi: 10.1145/2897166.

[79] A. Rustamov, N. Gogoi, A. Minetto, and F. Dovis, "Assessment of the Vulnerability to Spoofing Attacks of GNSS Receivers Integrated in Consumer Devices," *2020 Int. Conf. Localization GNSS, ICL-GNSS 2020 - Proc.*, pp. 1–6, 2020, doi: 10.1109/ICL-GNSS49876.2020.9115489.

[80] E. Schmidt, Z. A. Ruble, D. Akopian, and D. J. Pack, "Software-Defined Radio GNSS Instrumentation for Spoofing Mitigation : A Review and a Case Study," *IEEE Trans. Instrum. Measuerment*, 2019.

[81] S. Semanjski, I. Semanjski, W. De Wilde, and A. Muls, "Use and Validation of Supervised Machine Learning Approach for Detection of GNSS Signal Spoofing," in *2019 International Conference on Localization and GNSS (ICL-GNSS)*, 2019, pp. 1–6.

[82] S. Semanjski, I. Semanjski, W. De Wilde, and A. Muls, "Use of supervised machine learning for GNSS signal spoofing detection with validation on real- world meaconing and spoofing data—Part I," *Sensors (Switzerland)*, vol. 20, no. 4, 2020, doi: 10.3390/s20041171.

[83] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The Quefrency Alanysis of Time Series for Echoes," in *Proceedings of the Symposium on Time Series Analysis*, 1963, pp. 209–243.

[84] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.

[85] H. K. Liang, J. W. Pierre, and R. F. Kubichek, "Asymptotic statistics for an uncommon

averaged DFT magnitude spectral estimator used in test and measurement equipment," *Proc. 2002 IEEE 10th Digit. Signal Process. Work. DSP 2002 2nd Signal Process. Educ. Work. SPE 2002*, pp. 84–87, 2002, doi: 10.1109/DSPWS.2002.1231081.

[86] Wackerly, Mendenhall, and Scheaffer, *Mathematical Statistics with Applications*, 7th editio. 2017.

[87] L. Marti and F. van Graas, "Interference Detection by Means of the Software Defined Radio," in *Proceedings of the 17th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2004)*, 2004.

[88] E. Moulines, J. W. Dalle Molle, K. Choukri, and M. Charbit, "Testing that a stationary time-series is Gaussian: Time-domain vs. frequency-domain approaches," *Proc. - IEEE Signal Process. Work. Higher-Order Stat. HOST 1993*, pp. 336–340, 1993, doi: 10.1109/HOST.1993.264540.

[89] T. Gasser, "Goodness-of-Fit Tests for Correlated Data," *Biometrika*, vol. 62, no. 3, p. 563, Dec. 1975, doi: 10.2307/2335511.

[90] T. W. Epps, "Testing That a Stationary Time Series is Gaussian," *Ann. Stat.*, vol. 15, no. 4, pp. 1683–1698, Dec. 1987, doi: 10.1214/aos/1176350618.

[91] B. Motella and L. Lo Presti, "Methods of goodness of fit for GNSS interference detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 1690–1700, 2014, doi: 10.1109/TAES.2014.120368.

[92] F. D. Nunes and F. M. G. Sousa, "Gnss blind interference detection based on fourth-order autocumulants," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2574–2586, 2016, doi: 10.1109/TAES.2016.150499.

[93] R. D. De Roo, S. Misra, and C. S. Ruf, "Sensitivity of the kurtosis statistic as a detector of pulsed sinusoidal radio frequency interference in a microwave radiometer receiver," *Int. Geosci. Remote Sens. Symp.*, vol. 45, no. 7, pp. 2706–2709, 2007, doi: 10.1109/IGARSS.2007.4423401.

[94] A. J. Schoenwald, D. C. Bradley, P. N. Mohammed, J. R. Piepmeier, and M. Wong, "Performance analysis of a hardware implemented complex signal kurtosis radio-frequency interference detector," *14th Spec. Meet. Microw. Radiom. Remote Sens. Environ. MicroRad 2016 - Proc.*, vol. 1, no. 1, pp. 71–75, 2016, doi: 10.1109/MICRORAD.2016.7530507.

[95] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd editio. New York, NY: Springer New York, 1994.

[96] P. T. Capozza *et al.*, "Measured Effects of a Narrowband Interference Suppressor on GPS Receivers," no. June, pp. 28–30, 1999.

[97] F. Bastide *et al.*, "Automatic gain control (AGC) as an interference assessment tool," in *Proceedings of the 16th Interna-tional Technical Meeting of the Satellite Division of The Institute of Nav- igation (ION GPS/GNSS 2003)*, 2003, pp. 2042–2053.

[98] L. Marti and F. Van Graas, "Bias detection and its confidence assessment in global positioning system signals," *IEEE Aerosp. Conf. Proc.*, vol. 3, pp. 1608–1617, 2004, doi: 10.1109/aero.2004.1367935.

[99] J. Lehtomäki, "Analysis of energy based signal detection," 2005.

[100] A. T. Balaei, "Statistical inference technique in pre-correlation interference detection in GPS receivers," *Proc. Inst. Navig. - 19th Int. Tech. Meet. Satell. Div. ION GNSS 2006*, vol. 4, pp. 2232–2240, 2006.

[101] D. Borio, L. Camoriano, S. Savasta, and L. Lo Presti, "Time-Frequency Excision for GNSS

Applications," *IEEE Syst. J.*, vol. 2, no. 1, pp. 27–37, 2008, doi: 10.1109/JSYST.2007.914914.

[102] A. Tani and R. Fantacci, "Performance Evaluation of a Precorrelation Interference Detection Algorithm for the GNSS Based on Nonparametrical Spectral Estimation," *IEEE Syst. J.*, vol. 2, no. 1, pp. 20–26, 2008, doi: 10.1109/JSYST.2007.914772.

[103] A. T. Balaei and A. G. Dempster, "A statistical inference technique for gps interference detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 4, p. 1499, 2009, doi: 10.1109/TAES.2009.5310313.

[104] F. Faurie and A. Giremus, "Bayesian detection of interference in satellite navigation systems," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4348–4351, doi: 10.1109/ICASSP.2011.5947316.

[105] E. Axell, F. M. Eklöf, P. Johansson, M. Alexandersson, and D. M. Akos, "Jamming detection in GNSS receivers: Performance evaluation of field trials," *Navig. J. Inst. Navig.*, vol. 62, no. 1, pp. 73–82, 2015, doi: 10.1002/navi.74.

[106] J. T. Johnson and L. C. Potter, "A study of algorithms for detecting pulsed sinusoidal interference in microwave radiometry," *Int. Geosci. Remote Sens. Symp.*, vol. 2, no. 1, pp. 628–636, 2008, doi: 10.1109/IGARSS.2008.4778952.

[107] W. Li, Z. Jiaqi, L. Yan, and P. Shusheng, "Detection of radiometer radio frequency interference with power-law detector," *2017 Int. Appl. Comput. Electromagn. Soc. Symp. China, ACES-China 2017*, no. 4, pp. 16–17, 2017.

[108] Z. Wang and P. K. Willett, "All-purpose and plug-in power-law detectors for transient signals," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2454–2466, 2001, doi: 10.1109/78.960393.

[109] A. Sjölander and S. Vansteelandt, "Frequentist versus Bayesian approaches to multiple testing," *Eur. J. Epidemiol.*, vol. 34, no. 9, pp. 809–821, 2019, doi: 10.1007/s10654-019-00517-2.

[110] H. Tao, H. Li, and M. Lu, "A method of detections' fusion for GNSS anti-spoofing," *Sensors (Switzerland)*, vol. 16, no. 12, 2016, doi: 10.3390/s16122187.

[111] R. R. Yager and L. Liu, Eds., *Classic Works of the Dempster-Shafer Theory of Belief Functions*, vol. 219. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[112] M. Sgammini, F. Antreich, and M. Meurer, "SVD-based RF interference detection and mitigation for GNSS," *27th Int. Tech. Meet. Satell. Div. Inst. Navig. ION GNSS 2014*, vol. 4, no. October, pp. 3475–3483, 2014.

[113] F. Dovis and L. Musumeci, "Use of the Karhunen – Loeve Transform for interference detection and mitigation in GNSS," *ICT Express*, vol. 2, no. 1, pp. 33–36, 2016, doi: 10.1016/j.icte.2016.02.008.

[114] J. W. Betz, "Binary Offset Carrier Modulations for Radionavigation," *Navigation*, vol. 48, no. 4, pp. 227–246, Dec. 2001, doi: 10.1002/j.2161-4296.2001.tb00247.x.

[115] J. Wu and A. G. Dempster, "'BOC-gated-PRN' a multipath mitigation technique for BOC(n,n) waveforms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1136–1153, 2011, doi: 10.1109/TAES.2011.5751248.

[116] S. Verdú, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 85–96, 1986, doi: 10.1109/TIT.1986.1057121.

[117] R. A. Litis and L. Mailaender, "A symbol-by-symbol multiuser detector with joint amplitude and delay estimation," *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, vol. 12, no. 5, pp. 108–112, 1992, doi: 10.1109/ACSSC.1992.269248.

[118] S. Jiang, W. Wang, Y. Miao, W. Fan, and A. Molisch, “A Survey of Dense Multipath and Its Impact on Wireless Systems,” *IEEE Open J. Antennas Propag.*, 2022.

[119] M. R. Mosavi and F. Shafiee, “Narrowband interference suppression for GPS navigation using neural networks,” *GPS Solut.*, vol. 20, no. 3, pp. 341–351, 2016, doi: 10.1007/s10291-015-0442-8.

[120] T. Jost, W. Wang, F. Schubert, F. Antreich, and U. C. Fiebig, “Channel sounding using GNSS signals,” *Proc. 5th Eur. Conf. Antennas Propagation, EUCAP 2011*, pp. 3724–3728, 2011.

[121] F. Antreich, J. A. Nossek, and W. Utschick, “Maximum likelihood delay estimation in a navigation receiver for aeronautical applications,” *Aerosp. Sci. Technol.*, vol. 12, no. 3, pp. 256–267, 2008, doi: 10.1016/j.ast.2007.06.005.

[122] W. Wang, T. Jost, C. Gentner, S. Zhang, and A. Dammann, “A Semiblind Tracking Algorithm for Joint Communication and Ranging With OFDM Signals,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5237–5250, 2016, doi: 10.1109/TVT.2015.2468079.

[123] C. Sun, H. Zhao, W. Feng, and S. Du, “A frequency-domain multipath parameter estimation and mitigation method for boc-modulated gnss signals,” *Sensors (Switzerland)*, vol. 18, no. 3, 2018, doi: 10.3390/s18030721.

[124] F. Wen, J. Kulmer, K. Witrisal, and H. Wymeersch, “5G positioning and mapping with diffuse multipath,” *arXiv*, no. 700044, pp. 1–10, 2019, doi: 10.1109/twc.2020.3031180.

[125] H. Li, P. Borhani-Darian, P. Wu, and P. Closas, “Deep learning of GNSS signal correlation,” *Proc. 33rd Int. Tech. Meet. Satell. Div. Inst. Navig. ION GNSS+ 2020*, no. September, pp. 2836–2847, 2020, doi: 10.33012/2020.17598.

[126] Z. Albataineh and F. M. Salem, “Adaptive Blind CDMA Receivers Based on ICA Filtered Structures,” *Circuits, Syst. Signal Process.*, vol. 36, no. 8, pp. 3320–3348, 2017, doi: 10.1007/s00034-016-0459-4.

[127] M. N. Schmidt, “Single-channel source separation using complex matrix factorization,” Technical University of Denmark, 2009.

[128] D. Avitzour, “Detection of asynchronous CDMA with unknown user waveforms,” *IEEE Signal Process. Lett.*, vol. 11, no. 2 PART II, pp. 209–211, 2004, doi: 10.1109/LSP.2003.821678.

[129] B. Loesch and B. Yang, “Source Number Estimation and Clustering for Undeterminded Blind Source Separation,” *Proc. Int. Work. Acoust. Echo Noise Control*, no. 5, 2008.

[130] H. C. Bing, S. Kadambe, V. Adams, and W. Dewley, “Estimation of the number of sources present in instantaneous and anechoic mixtures,” *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, pp. 1242–1246, 2006, doi: 10.1109/ACSSC.2006.354953.

[131] D. B. Haddad, L. Lovisolo, M. R. Petraglia, P. B. Batalheiro, and J. C. P. Filho, “Blind and Semi-blind Anechoic Mixing System Identification Using Multichannel Matching Pursuit,” *Circuits, Syst. Signal Process.*, vol. 40, no. 9, pp. 4546–4575, 2021, doi: 10.1007/s00034-021-01681-1.

[132] C. Tapia, O. Daud, and J. Ruiz-del-Solar, “EMG Signal Filtering Based on Independent Component Analysis and Empirical Mode Decomposition for Estimation of Motor Activation Patterns,” *J. Med. Biol. Eng.*, vol. 37, no. 1, pp. 140–155, 2017, doi: 10.1007/s40846-016-0201-5.

[133] Y. Guo, G. R. Naik, and H. Nguyen, “Single channel blind source separation based local mean decomposition for Biomedical applications,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6812–6815, 2013, doi: 10.1109/EMBC.2013.6611121.

[134] Y. Guo, S. Huang, Y. Li, and G. R. Naik, “Edge effect elimination in single-mixture blind

source separation," *Circuits, Syst. Signal Process.*, vol. 32, no. 5, pp. 2317–2334, 2013, doi: 10.1007/s00034-013-9556-9.

[135] A. K. Maddirala and R. A. Shaik, "Separation of sources from single-channel EEG signals using independent component analysis," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 382–393, 2018, doi: 10.1109/TIM.2017.2775358.

[136] S. Kouchaki and S. Sanei, "Supervised single channel source separation of EEG signals," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013, doi: 10.1109/MLSP.2013.6661895.

[137] A. K. Maddirala and R. A. Shaik, "Removal of EOG Artifacts from Single Channel EEG Signals Using Combined Singular Spectrum Analysis and Adaptive Noise Canceler," *IEEE Sens. J.*, vol. 16, no. 23, pp. 8279–8287, 2016, doi: 10.1109/JSEN.2016.2560219.

[138] B. Mijović, M. De Vos, I. Gligorijević, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, 2010, doi: 10.1109/TBME.2010.2051440.

[139] Z. Zhang, H. Li, and D. Mandic, "Blind source separation and artefact cancellation for single channel bioelectrical signal," *BSN 2016 - 13th Annu. Body Sens. Networks Conf.*, pp. 177–182, 2016, doi: 10.1109/BSN.2016.7516255.

[140] B. Mijović, M. De Vos, I. Gligorijević, and S. Van Huffel, "Combining EMD with ICA for extracting independent sources from single channel and two-channel data," *2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC'10*, pp. 5387–5390, 2010, doi: 10.1109/IEMBS.2010.5626482.

[141] G. J. Jang and T. W. Lee, "A maximum likelihood approach to single-channel source separation," *J. Mach. Learn. Res.*, vol. 4, no. 7–8, pp. 1365–1392, 2004.

[142] F. Weninger, J. R. Hershey, J. Le Roux, and B. Schuller, "Discriminatively trained recurrent neural networks for single-channel speech separation," *2014 IEEE Glob. Conf. Signal Inf. Process. Glob. 2014*, pp. 577–581, 2014, doi: 10.1109/GlobalSIP.2014.7032183.

[143] R. Aihara, G. Wichern, and J. Le Roux, "Deep clustering-based single-channel speech separation and recent advances," *Acoust. Sci. Technol.*, vol. 41, no. 2, pp. 465–471, 2020, doi: 10.1250/ast.41.465.

[144] E. M. Grais, G. Roma, A. J. R. Simpson, and M. D. Plumbley, "Two-Stage Single-Channel Audio Source Separation Using Deep Neural Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 9, pp. 1773–1783, 2017, doi: 10.1109/TASLP.2017.2716443.

[145] E. M. Grais, M. U. Sen, and H. Erdogan, "Deep neural networks for single channel source separation," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 3734–3738, 2014, doi: 10.1109/ICASSP.2014.6854299.

[146] S. Kouchaki and S. Sanei, "Tensor factorisation for narrowband single channel source decomposition," *J. Franklin Inst.*, vol. 354, no. 7, pp. 3152–3169, 2017, doi: 10.1016/j.jfranklin.2017.01.018.

[147] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, USA: John Wiley & Sons, Inc., 2001.

[148] T. Routtenberg and J. Tabrikian, "MIMO-AR system identification and blind source separation for GMM-distributed sources," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1717–1730, 2009, doi: 10.1109/TSP.2008.2011825.

[149] I. Russell, J. Xi, A. Mertins, and J. Chicharo, "Blind source separation of nonstationary convolutively mixed signals in the time domain," *ICASSP, IEEE Int. Conf. Acoust. Speech*

Signal Process. - Proc., vol. 5, no. 2, pp. 93–98, 2004, doi: 10.1109/icassp.2004.1327152.

[150] K. Xie, G. Zhou, J. Yang, Z. He, and S. Xie, "Eliminating the Permutation Ambiguity of Convolutive Blind Source Separation by Using Coupled Frequency Bins," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, no. 2, pp. 589–599, 2020, doi: 10.1109/TNNLS.2019.2906833.

[151] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," in *Proc. IEEE Int. Symposium on Independent Component Analysis and Blind Source Separation (ICA)*, 2003, no. 2, pp. 1–6.

[152] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, 2005, doi: 10.1109/TSA.2004.838775.

[153] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, *A Survey of Convolutive Blind Source Separation Methods*. 2008.

[154] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 1, no. 4, pp. 423–426, 1997, doi: 10.1109/icassp.1997.599665.

[155] M. S. Alouini, A. Scaglione, and G. B. Giannakis, "PCC: Principal components combining for dense correlated multipath fading environments," *IEEE Veh. Technol. Conf.*, vol. 5, no. 52 ND, pp. 2510–2517, 2000, doi: 10.1109/vetecf.2000.883312.

[156] R. H. Lambert, M. Joho, and H. Mathis, "Polynomial Singular Values for Number of Wideband Sources Estimation and Principal Component Analysis," pp. 379–383, 2001.

[157] J. He, W. Chen, and Y. Song, "Single Channel Blind Source Separation Under Deep Recurrent Neural Network," *Wirel. Pers. Commun.*, vol. 115, no. 2, pp. 1277–1289, 2020, doi: 10.1007/s11277-020-07624-4.

[158] S. P. Chatzis and D. Kosmopoulos, "A nonparametric Bayesian approach toward stacked convolutional independent component analysis," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 2803–2811, 2015, doi: 10.1109/ICCV.2015.321.

[159] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

[160] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: The Hilbert spectrum," *Annu. Rev. Fluid Mech.*, vol. 31, pp. 417–457, 1999, doi: 10.1146/annurev.fluid.31.1.417.

[161] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, 2014, doi: 10.1109/TSP.2013.2288675.

[162] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Phys. D Nonlinear Phenom.*, vol. 20, no. 2–3, pp. 217–236, Jun. 1986, doi: 10.1016/0167-2789(86)90031-X.

[163] J. B. Elsner and A. A. Tsonis, *Singular Spectrum Analysis*. Boston, MA: Springer US, 1996.

[164] C. Fu and F. F. Zhang, "On the nonlinear properties analysis of multipath fading channel," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, 2007, doi: 10.1109/TENCON.2007.4429129.

[165] N. E. Huang *et al.*, "On holo-hilbert spectral analysis: A full informational spectral representation for nonlinear and non-stationary data," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0206.

[166] S. Madan, K. Srivastava, A. Sharmila, and P. Mahalakshmi, "A case study on Discrete Wavelet

Transform based Hurst exponent for epilepsy detection," *J. Med. Eng. Technol.*, vol. 42, no. 1, pp. 9–17, 2018, doi: 10.1080/03091902.2017.1394390.

[167] M. Li, W. Chen, and T. Zhang, "Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM," *Biocybern. Biomed. Eng.*, vol. 36, no. 4, pp. 708–718, 2016, doi: 10.1016/j.bbe.2016.07.004.

[168] F. P. Karegar, A. Fallah, and S. Rashidi, "Using Recurrence quantification analysis and Generalized Hurst Exponents of ECG for human authentication," *2nd Conf. Swarm Intell. Evol. Comput. CSIEC 2017 - Proc.*, pp. 66–71, 2017, doi: 10.1109/CSIEC.2017.7940172.

[169] H. E. Hurst, "Long-Term Storage Capacity of Reservoirs," *Trans. Am. Soc. Civ. Eng.*, vol. 116, no. 1, pp. 770–799, Jan. 1951, doi: 10.1061/TACEAT.0006518.

[170] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-term Storage: An Experimental Study*. Constable, 1965.

[171] B. B. Mandelbrot and J. R. Wallis, "Noah, Joseph, and Operational Hydrology," *Water Resour. Res.*, vol. 4, no. 5, pp. 909–918, Oct. 1968, doi: 10.1029/WR004i005p00909.

[172] D. Koutsoyiannis, "The Hurst phenomenon and fractional Gaussian noise made easy," *Hydrol. Sci. J.*, vol. 47, no. 4, 2002.

[173] T. Graves, R. Gramacy, N. Watkins, and C. Franzke, "A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951-1980," *Entropy*, vol. 19, no. 9, pp. 0–21, 2017, doi: 10.3390/e19090437.

[174] I. Simonsen, A. Hansen, and O. M. Nes, "Determination of the Hurst exponent by use of wavelet transforms," *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 58, no. 3, pp. 2779–2787, 1998, doi: 10.1103/PhysRevE.58.2779.

[175] C. Gentili, N. Vanello, I. Cristea, D. David, E. Ricciardi, and P. Pietrini, "Proneness to social anxiety modulates neural complexity in the absence of exposure: A resting state fMRI study using Hurst exponent," *Psychiatry Res. - Neuroimaging*, vol. 232, no. 2, pp. 135–144, 2015, doi: 10.1016/j.pscychresns.2015.03.005.

[176] A. Gupta and S. Joshi, "Estimation of multipath fading channel using fractal based VSLMS algorithm," *WSEAS Trans. Signal Process.*, vol. 10, no. 1, pp. 230–241, 2014.

[177] J. Dong, B. Jing, X. Ma, H. Liu, X. Mo, and H. Li, "Hurst exponent analysis of resting-state fMRI signal complexity across the adult lifespan," *Front. Neurosci.*, vol. 12, no. FEB, pp. 1–10, 2018, doi: 10.3389/fnins.2018.00034.

[178] T. Di Matteo, T. Aste, and M. M. Dacorogna, "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development," *J. Bank. Financ.*, vol. 29, no. 4, pp. 827–851, 2005, doi: 10.1016/j.jbankfin.2004.08.004.

[179] T. Di Matteo, T. Aste, and M. M. Dacorogna, "Scaling behaviors in differently developed markets," *Phys. A Stat. Mech. its Appl.*, vol. 324, no. 1–2, pp. 183–188, 2003, doi: 10.1016/S0378-4371(02)01996-9.

[180] A.-L. Barabási and T. Vicsek, "Multifractality of self-affine fractals," *Phys. Rev. A*, vol. 44, no. 4, pp. 2730–2733, Aug. 1991, doi: 10.1103/PhysRevA.44.2730.

[181] P. Flandrin, "On the spectrum of fractional Brownian motions," *Information Theory, IEEE Transactions on*, vol. 35, no. 1. pp. 197–199, 1989, [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=42195%0Apapers3://publication/doi/10.1109/18.42195.

[182] R. M. Bryce and K. B. Sprague, "Revisiting detrended fluctuation analysis," *Sci. Rep.*, vol. 2, 2012, doi: 10.1038/srep00315.

[183] C. Heneghan and G. McDarby, “Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes,” *Phys. Rev. E*, vol. 62, no. 5, pp. 6103–6110, Nov. 2000, doi: 10.1103/PhysRevE.62.6103.

[184] A. Habib, J. P. R. Sorensen, J. P. Bloomfield, K. Muchan, A. J. Newell, and A. P. Butler, “Temporal scaling phenomena in groundwater-floodplain systems using robust detrended fluctuation analysis,” *J. Hydrol.*, vol. 549, pp. 715–730, 2017, doi: 10.1016/j.jhydrol.2017.04.034.

[185] A. K. Maity *et al.*, “Multifractal Detrended Fluctuation Analysis of alpha and theta EEG rhythms with musical stimuli,” *Chaos, Solitons and Fractals*, vol. 81, pp. 52–67, 2015, doi: 10.1016/j.chaos.2015.08.016.

[186] R. Hardstone *et al.*, “Detrended fluctuation analysis: A scale-free view on neuronal oscillations,” *Front. Physiol.*, vol. 3 NOV, no. November, pp. 1–13, 2012, doi: 10.3389/fphys.2012.00450.

[187] J. Theiler, “Efficient algorithm for estimating the correlation dimension from a set of discrete points,” *Physical Review A*, vol. 36, no. 9. pp. 4456–4462, 1987, doi: 10.1103/PhysRevA.36.4456.

[188] P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” *Phys. D Nonlinear Phenom.*, vol. 9, no. 1–2, pp. 189–208, Oct. 1983, doi: 10.1016/0167-2789(83)90298-1.

[189] G. L. Baker and J. P. Gollub, *Chaotic Dynamics*. Cambridge University Press, 1996.

[190] A. H. Nayfeh and B. Balachandran, *Applied Nonlinear Dynamics*. Wiley, 1995.

[191] F. Takens, “On the numerical determination of the dimension of an attractor,” 1985, pp. 99–106.

[192] K. Falconer, *Fractal Geometry*. Chichester, UK: John Wiley & Sons, Ltd, 2003.

[193] B. Mandelbrot, “How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension,” *Science (80-.).*, vol. 156, no. 3775, pp. 636–638, May 1967, doi: 10.1126/science.156.3775.636.

[194] A. De Luca and S. Termini, “A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory,” *Inf. Control*, vol. 20, no. 4, pp. 301–312, 1972, doi: 10.1016/S0019-9958(72)90199-4.

[195] N. R. Pal and J. C. Brzdek, “Measuring Fuzzy Uncertainty,” *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 2, pp. 107–118, 1994, doi: 10.1109/91.277960.

[196] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, and Y. Hou, “Assessing the complexity of short-term heartbeat interval series by distribution entropy,” *Med. Biol. Eng. Comput.*, vol. 53, no. 1, pp. 77–87, 2015, doi: 10.1007/s11517-014-1216-0.

[197] B. Shi, Y. Zhang, C. Yuan, S. Wang, and P. Li, “Entropy Analysis of Short-Term Heartbeat Interval Time Series during Regular Walking,” *Entropy*, vol. 19, no. 10, pp. 1–14, 2017, doi: 10.3390/e19100568.

[198] Meng Hu and Hualou Liang, “Adaptive Multiscale Entropy Analysis of Multivariate Neural Data,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 12–15, Jan. 2012, doi: 10.1109/TBME.2011.2162511.

[199] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Boston, 1980.

[200] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.

[201] S. Al-Sharhan, F. Karray, W. Gueaieb, and O. Basir, “Fuzzy entropy: A brief survey,” *IEEE Int. Conf. Fuzzy Syst.*, vol. 3, pp. 1135–1139, 2001, doi: 10.1109/fuzz.2001.1008855.

[202] W. Chen, J. Zhuang, W. Yu, and Z. Wang, “Measuring complexity using FuzzyEn, ApEn, and SampEn,” *Med. Eng. Phys.*, vol. 31, no. 1, pp. 61–68, Jan. 2009, doi: 10.1016/j.medengphy.2008.04.005.

[203] Y. Li, M. Xu, H. Zhao, and W. Huang, “Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis,” vol. 98, pp. 114–132, 2016, doi: 10.1016/j.mechmachtheory.2015.11.010.

[204] W. Chen, Z. Wang, H. Xie, and W. Yu, “Characterization of surface EMG signal based on fuzzy entropy,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, 2007, doi: 10.1109/TNSRE.2007.897025.

[205] M. Little, P. McSharry, I. Moroz, and S. Roberts, “Nonlinear, biophysically-informed speech pathology detection,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2, pp. 1080–1083, 2006, doi: 10.1109/icassp.2006.1660534.

[206] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *Biomed. Eng. Online*, vol. 6, pp. 1–19, 2007, doi: 10.1186/1475-925X-6-23.

[207] S. Mukherjee, S. K. Palit, S. Banerjee, M. R. K. Ariffin, L. Rondoni, and D. K. Bhattacharya, “Can complexity decrease in congestive heart failure?,” *Phys. A Stat. Mech. its Appl.*, vol. 439, pp. 93–102, 2015, doi: 10.1016/j.physa.2015.07.030.

[208] G. Corso, T. L. Prado, G. Z. dos S. Lima, and S. R. Lopes, “A novel entropy recurrence quantification analysis,” pp. 1–10, 2017, [Online]. Available: http://arxiv.org/abs/1707.00944.

[209] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, “Delay embeddings for forced systems. II. stochastic forcing,” *J. Nonlinear Sci.*, vol. 13, no. 6, pp. 519–577, 2003, doi: 10.1007/s00332-003-0534-4.

[210] M. Pätzold, *Mobile Fading Channels*. Wiley, 2002.

[211] A. Joseph, “Measuring GNSS Signal Strength,” *InsideGNSS*, 2010.

[212] H. Yang, “Cross recurrence plot and quantification analysis -CRP, CRQA (https://www.mathworks.com/matlabcentral/fileexchange/100164-cross-recurrence-plot-and-quantification-analysis-crp-crqa), MATLAB Central File Exchange.” 2023.

[213] P. Viszlay, M. Lojka, and J. Juhar, “Class-dependent two-dimensional linear discriminant analysis using two-pass recognition strategy,” in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 1796–1800.

[214] C. M. Bishop, “Neural Networks for Pattern Recognition.”

[215] S. Balakrishnama and A. Ganapathiraju, “Linear discriminant analysis-a brief tutorial,” Institute for Signal and information Processing, 1998.

[216] A. Sharma, K. K. Paliwal, and G. C. Onwubolu, “Class-dependent PCA, MDC and LDA: A combined classifier for pattern classification,” *Pattern Recognit.*, vol. 39, no. 7, pp. 1215–1229, 2006, doi: 10.1016/j.patcog.2006.02.001.

[217] M. Thompson, R. O. Duda, and P. E. Hart, “Pattern Classification and Scene Analysis,” *Leonardo*, vol. 7, no. 4, p. 370, 1974, doi: 10.2307/1573081.

[218] A. Sharma and K. K. Paliwal, “Linear discriminant analysis for the small sample size problem: an overview,” *Int. J. Mach. Learn. Cybern.*, vol. 6, no. 3, pp. 443–454, 2015, doi: 10.1007/s13042-013-0226-9.

[219] K. B. Duan and S. S. Keerthi, “Which is the best multiclass SVM method? An empirical study,” *Lect. Notes Comput. Sci.*, vol. 3541, pp. 278–285, 2005, doi: 10.1007/11494683_28.

[220] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2006.

[221] A. Tharwat, “Parameter investigation of support vector machine classifier with kernel functions,” *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1269–1302, Dec. 2019, doi: 10.1007/s10115-019-01335-4.

[222] N. Zhu, J. Marais, D. Betaille, and M. Berbineau, “GNSS Position Integrity in Urban Environments: A Review of Literature,” *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 2762–2778, 2018, doi: 10.1109/TITS.2017.2766768.

[223] J. Breßler, P. Reisdorf, M. Obst, and G. Wanielik, “GNSS positioning in non-line-of-sight context-A survey,” *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 1147–1154, 2016, doi: 10.1109/ITSC.2016.7795701.

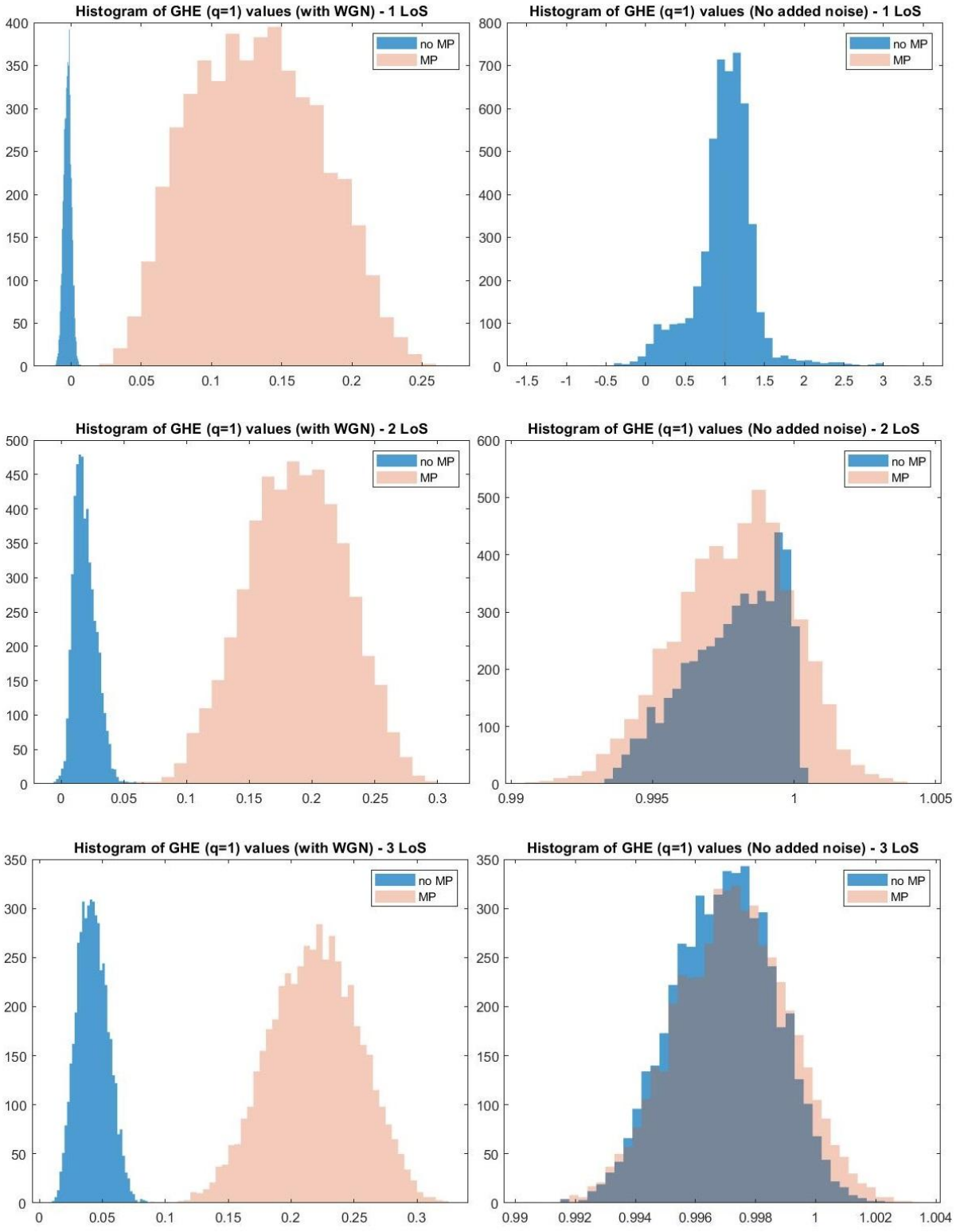
[224] S. V. Shafran, E. A. Gizatulova, and I. A. Kudryavtsev, “Snapshot technology in GNSS receivers,” in *2018 25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*, May 2018, pp. 1–3, doi: 10.23919/ICINS.2018.8405873.

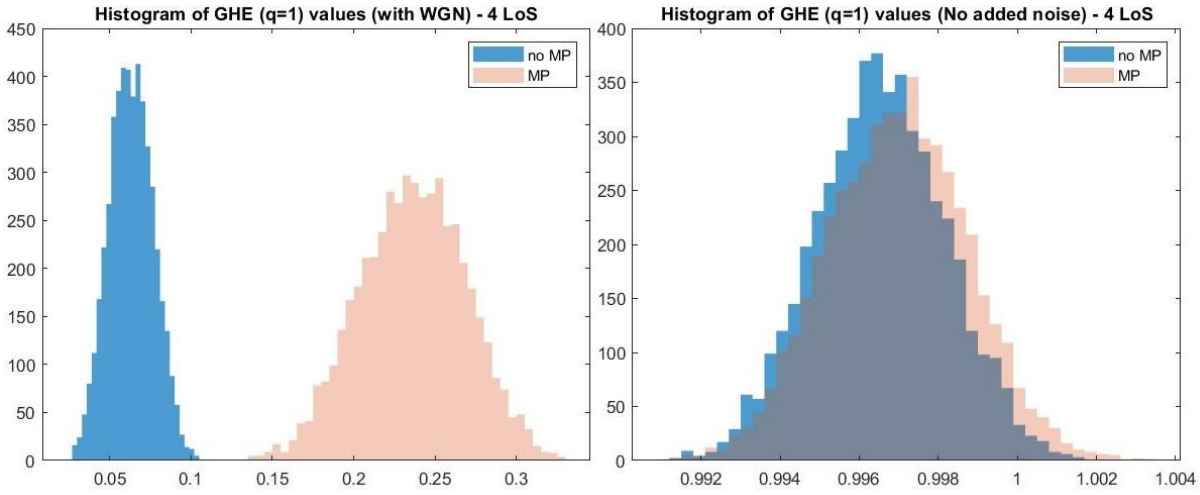
Appendix A – Histograms

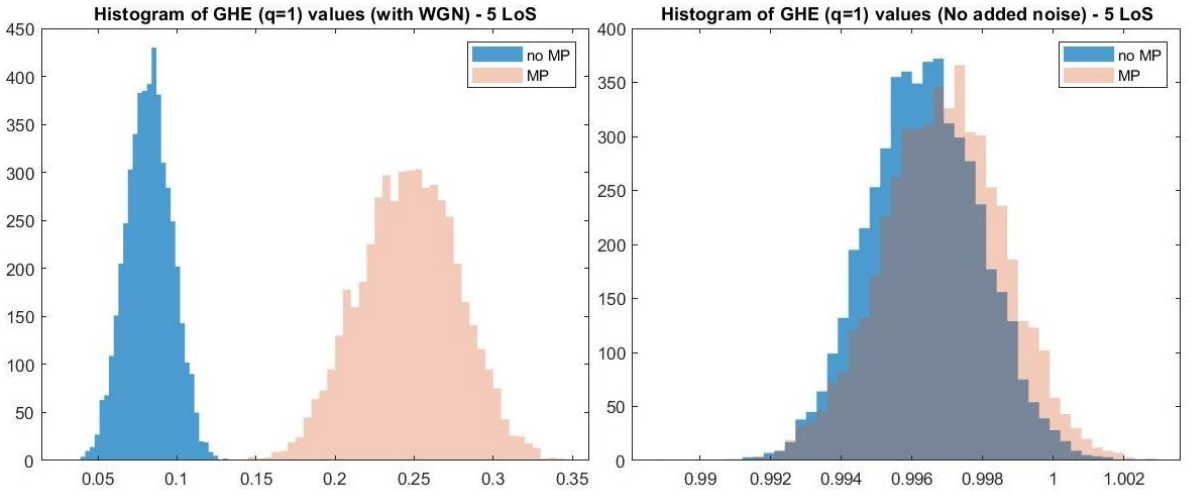
The purpose of presenting the histograms corresponding to method is to visualize the underlying statistics of the numerical simulations, and not to infer any statistical quantities as density functions remain unknown.

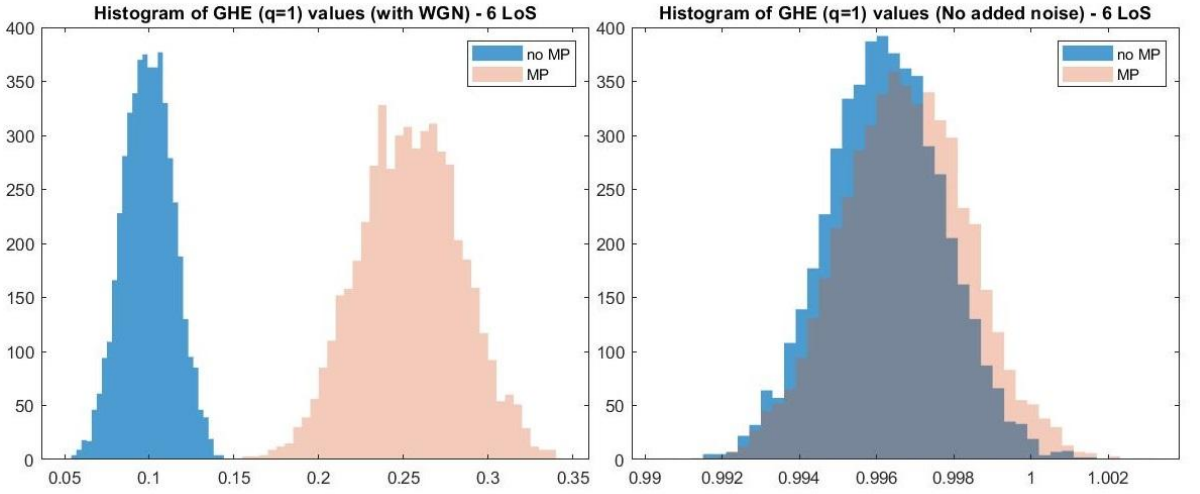
A.1 Generalized Hurst Exponent

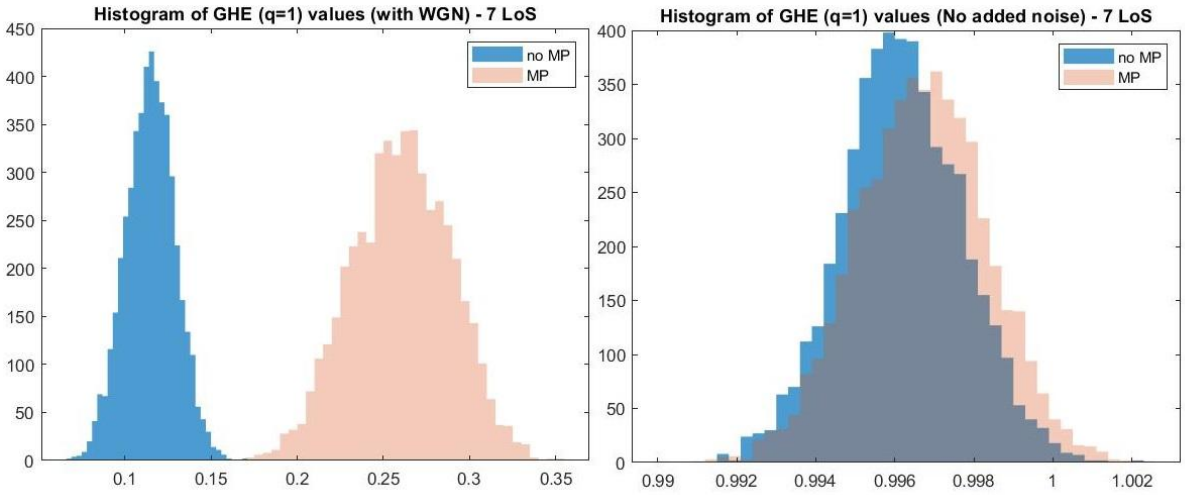
The GHE implementation is the most computationally efficient from the ones tested in this research, thus it allowed for experimentation using 5000 Monte Carlo Iterations. Next, the $H(1)$ result histograms are presented:

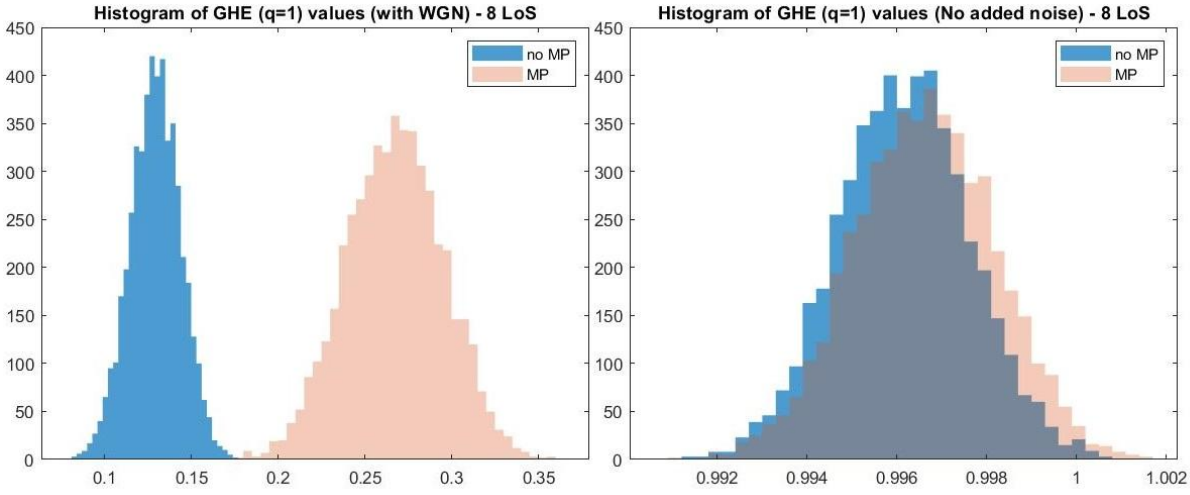


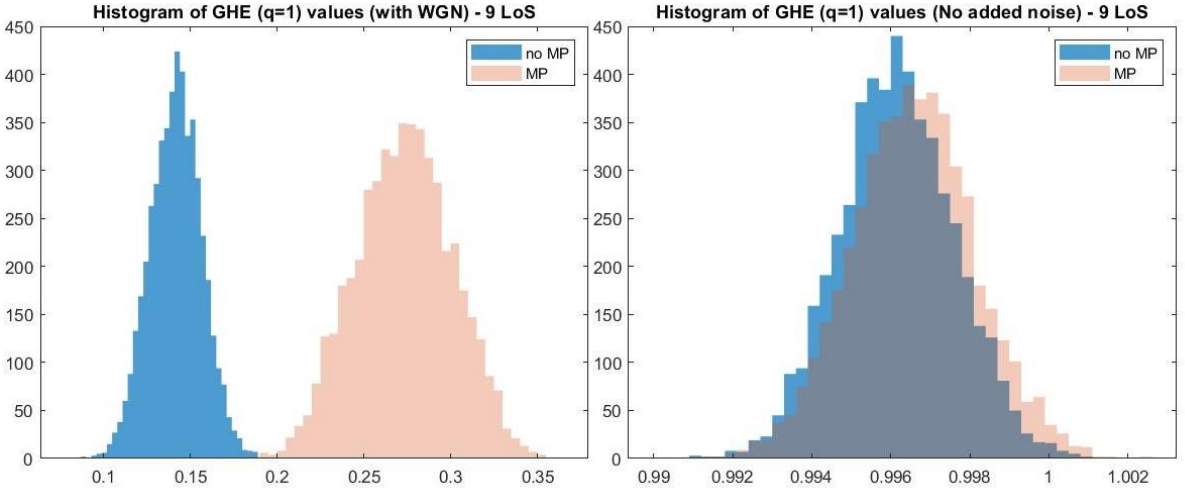

Histogram of GHE (q=1) values (with WGN) - 4 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 4 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 5 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 5 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 6 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 6 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 7 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 7 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 8 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 8 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 9 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 9 LoS
no MP
MP

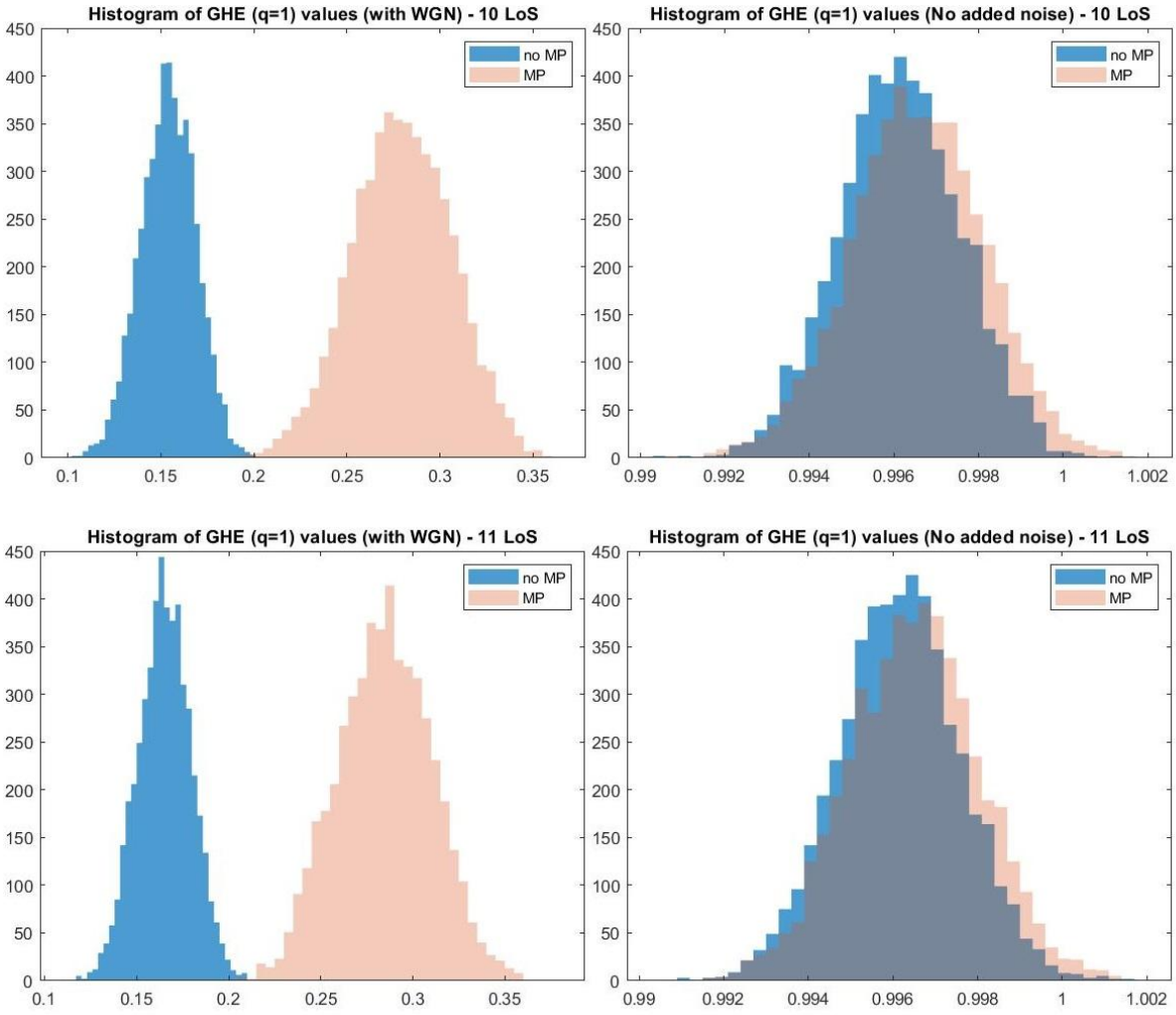
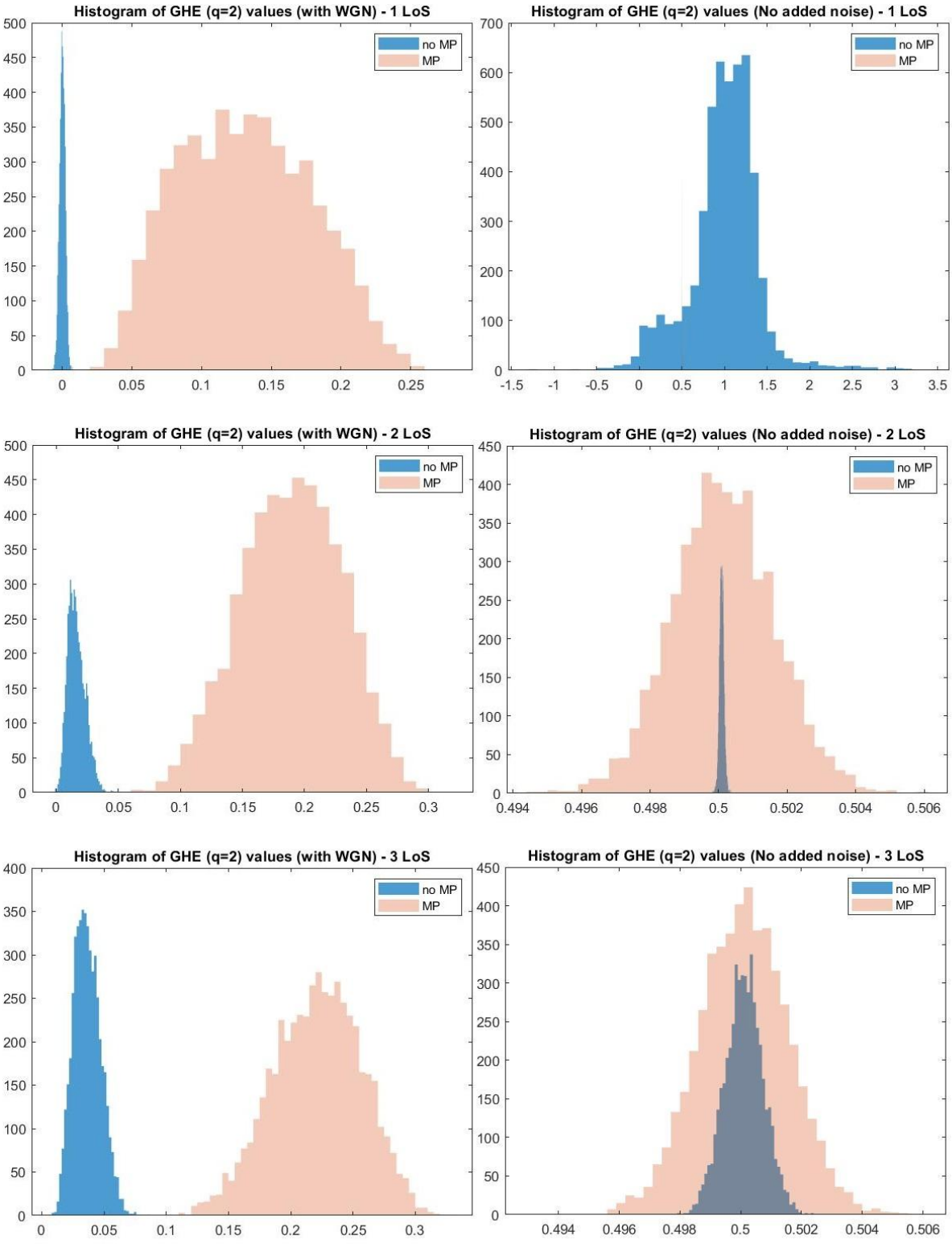
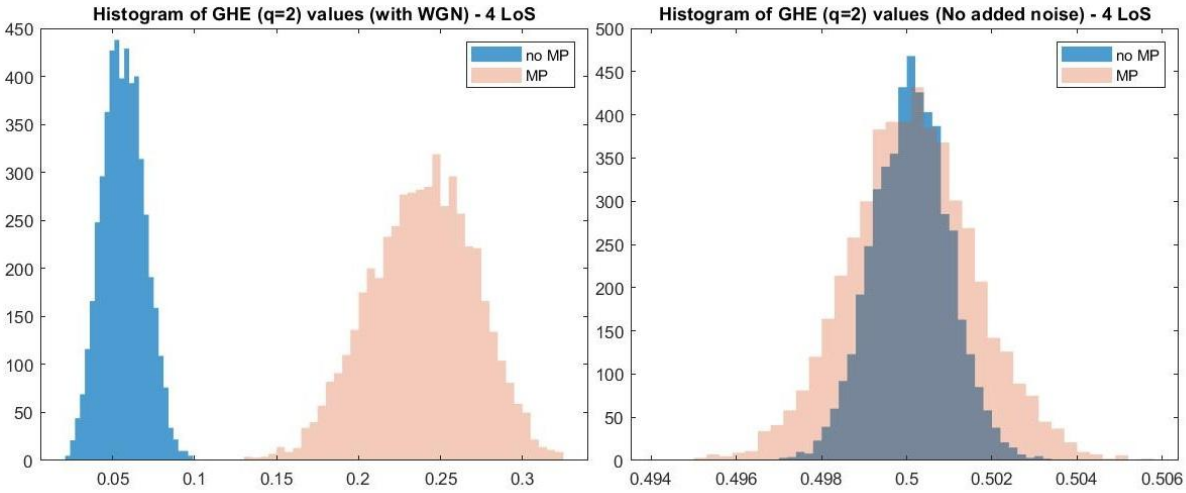
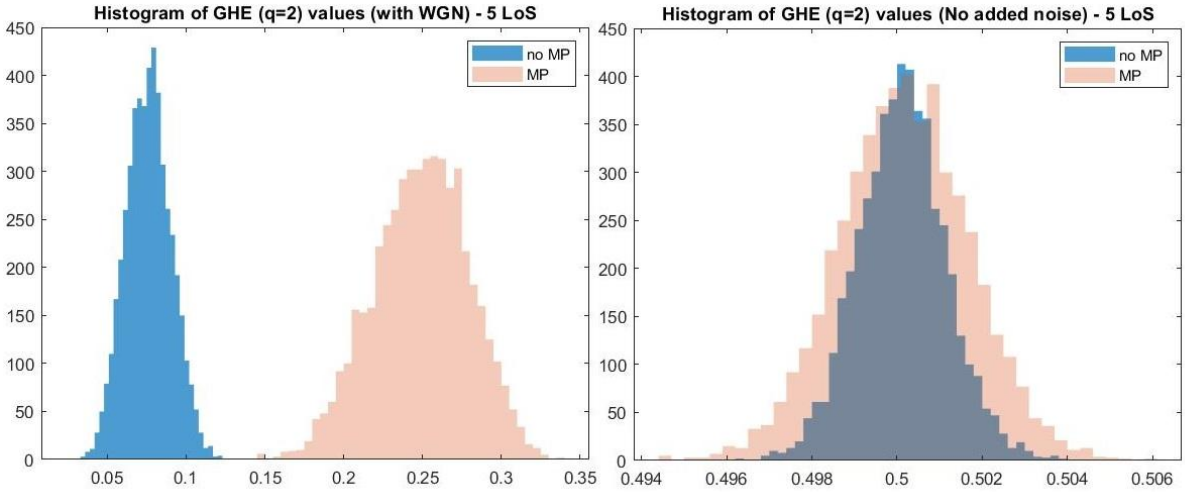


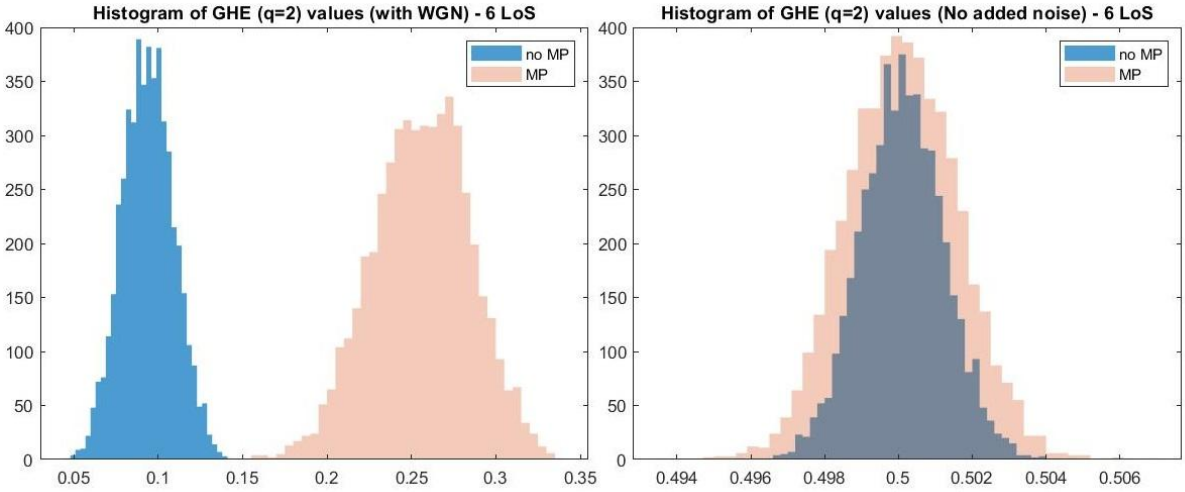
Figure 38 – GHE (q=1) histograms after 5000 iterations

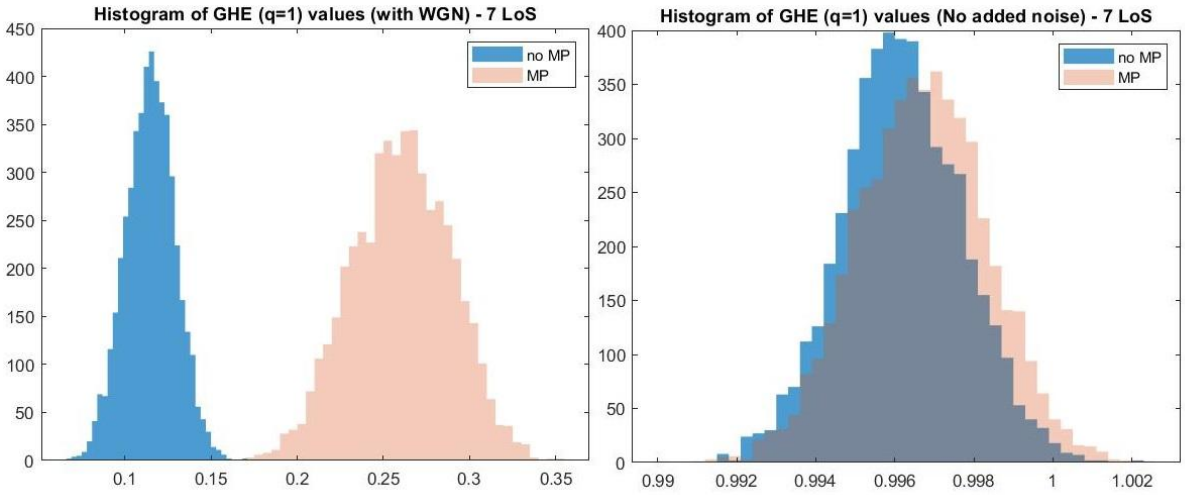
For $H(2)$ respectively:

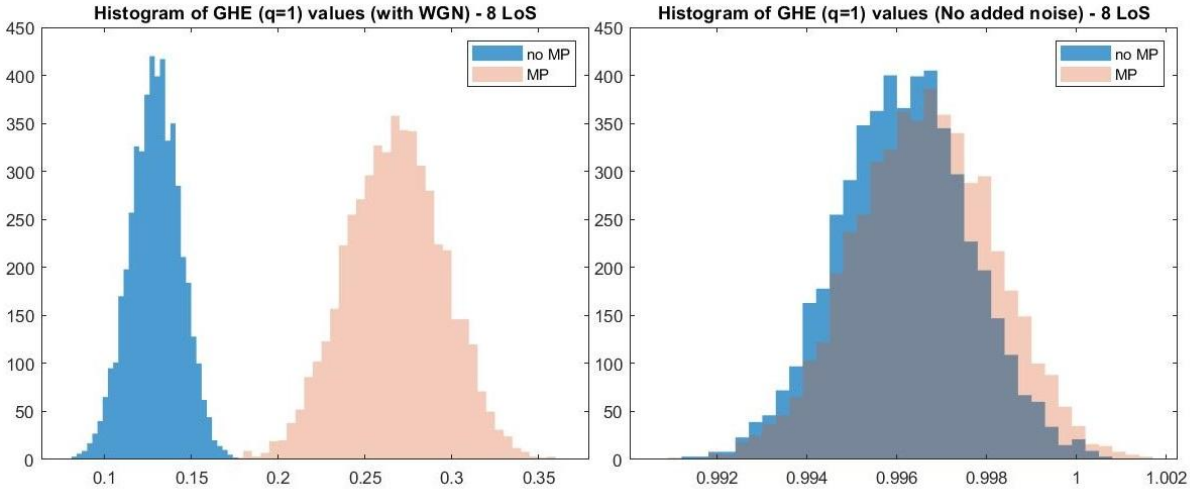


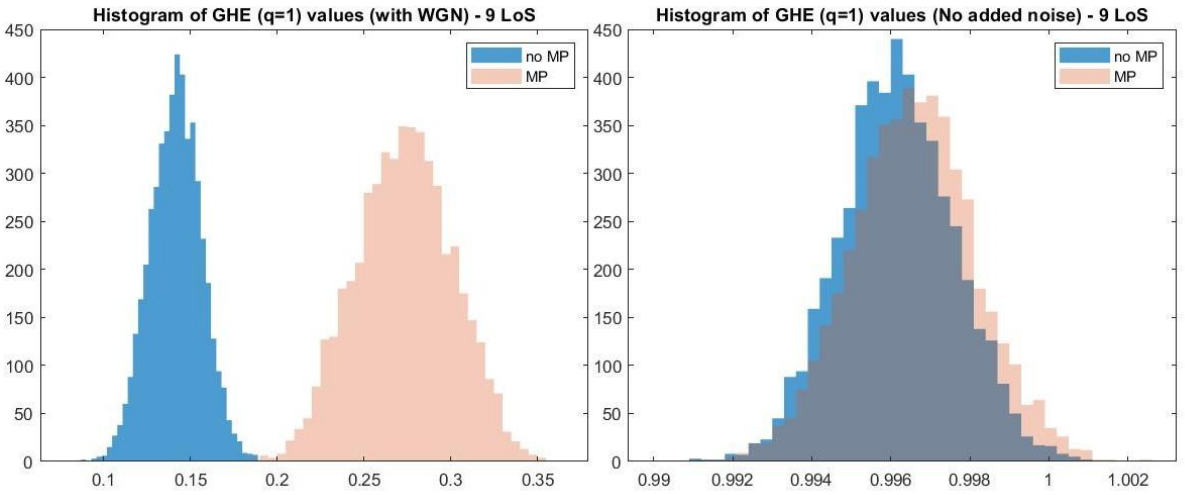

Histogram of GHE (q=2) values (with WGN) - 4 LoS
no MP
MP
Histogram of GHE (q=2) values (No added noise) - 4 LoS
no MP
MP


Histogram of GHE (q=2) values (with WGN) - 5 LoS
no MP
MP
Histogram of GHE (q=2) values (No added noise) - 5 LoS
no MP
MP


Histogram of GHE (q=2) values (with WGN) - 6 LoS
no MP
MP
Histogram of GHE (q=2) values (No added noise) - 6 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 7 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 7 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 8 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 8 LoS
no MP
MP


Histogram of GHE (q=1) values (with WGN) - 9 LoS
no MP
MP
Histogram of GHE (q=1) values (No added noise) - 9 LoS
no MP
MP

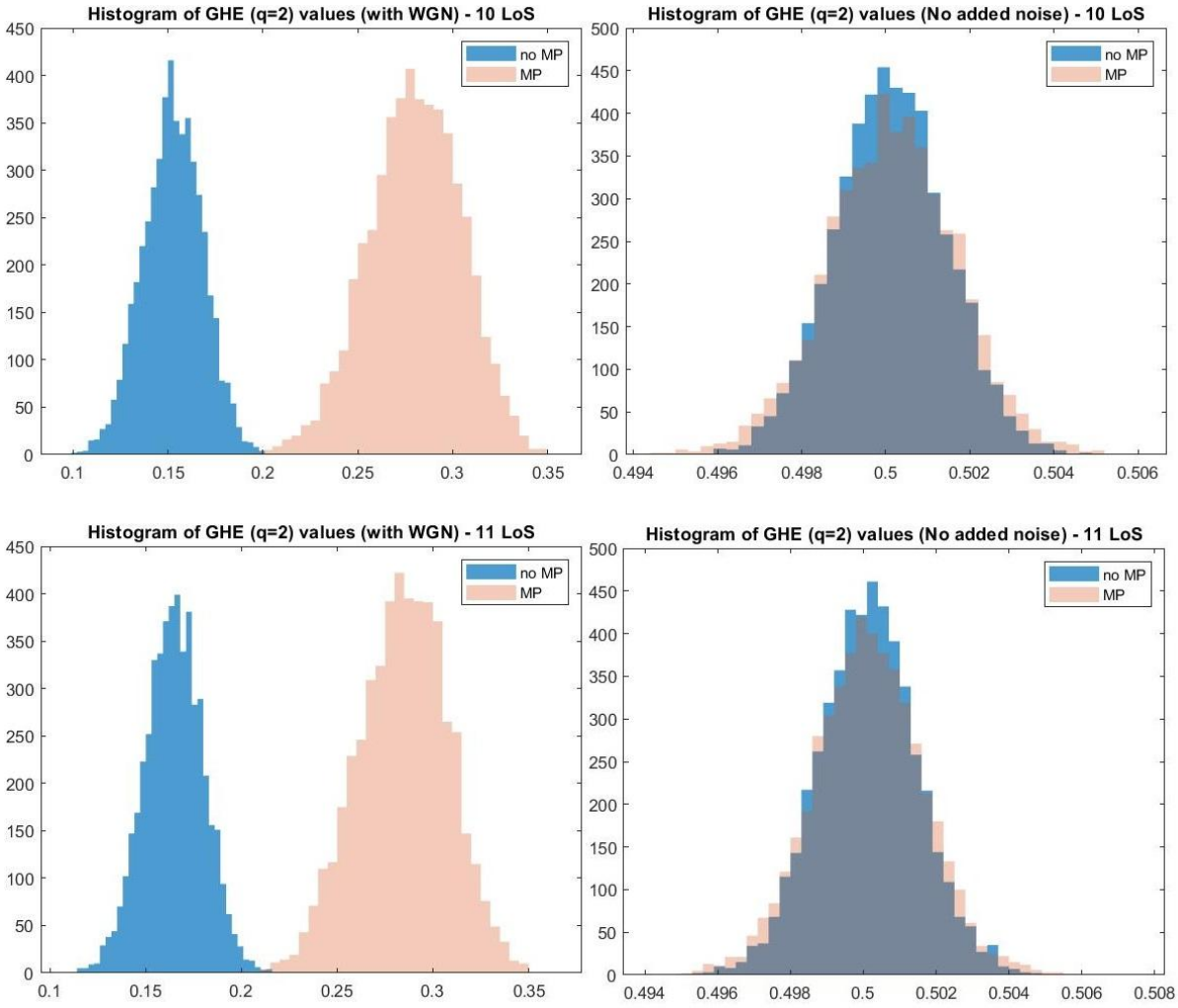
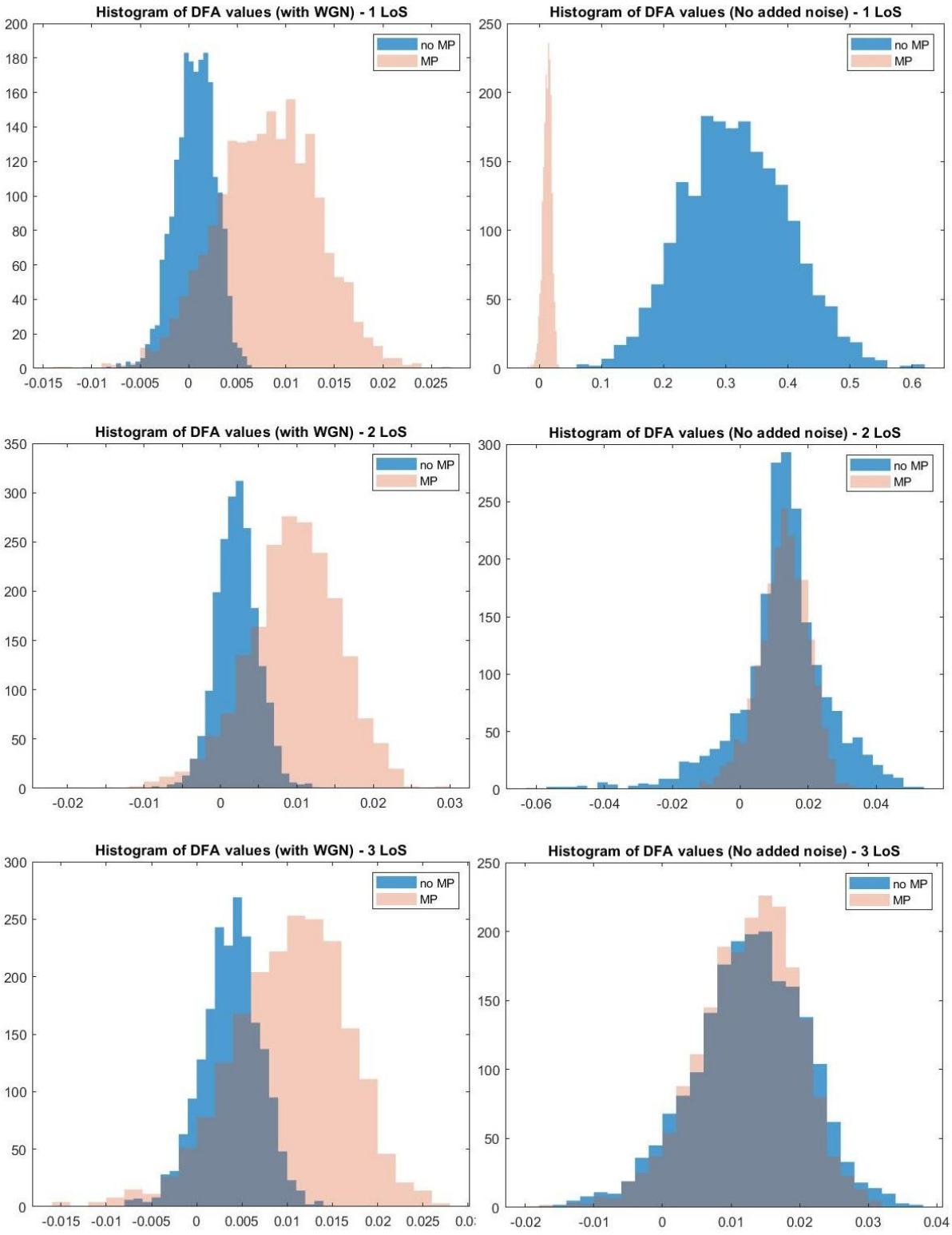
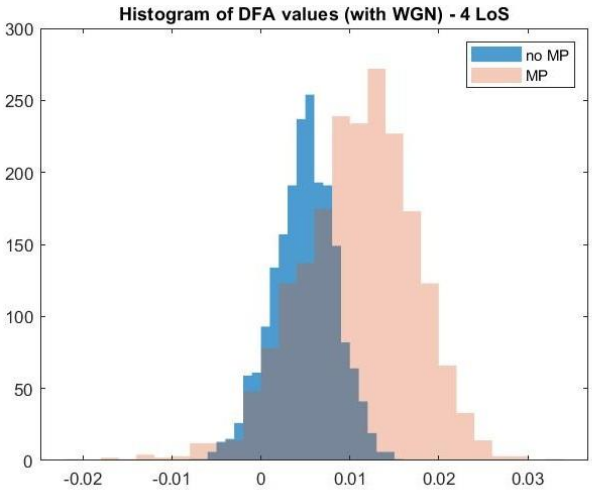


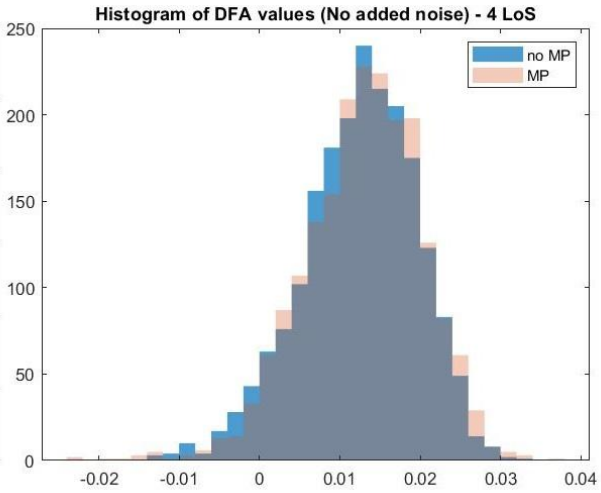
Figure 39 - GHE (q=2) histograms after 5000 iterations

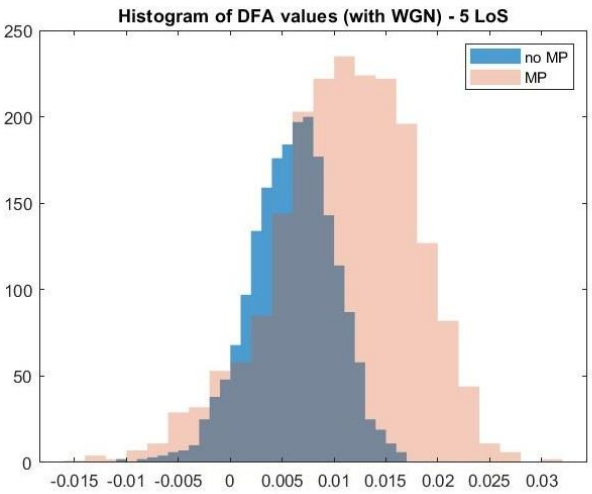
A.2 Detrended fluctuation analysis

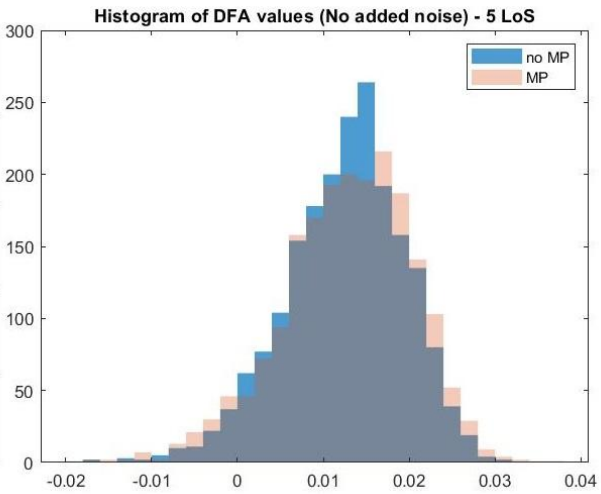
For DFA 2000 iterations were feasible yielding the histograms shown below:

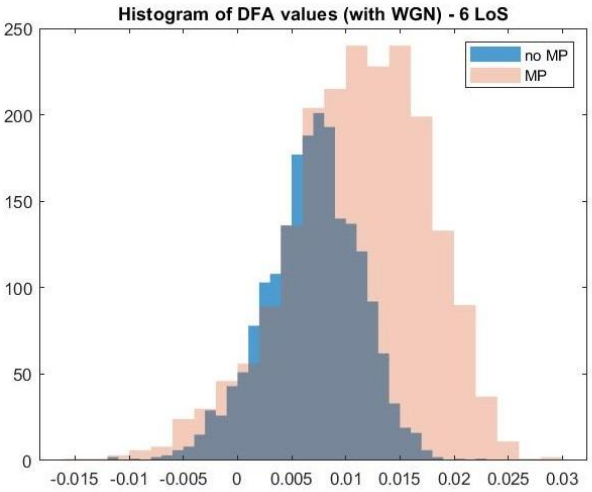


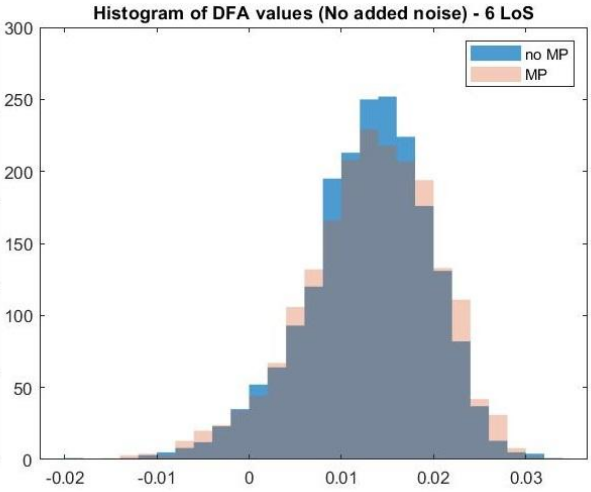

Histogram of DFA values (with WGN) - 4 LoS
no MP
MP
300
250
200
150
100
50
0
-0.02
-0.01
0
0.01
0.02
0.03

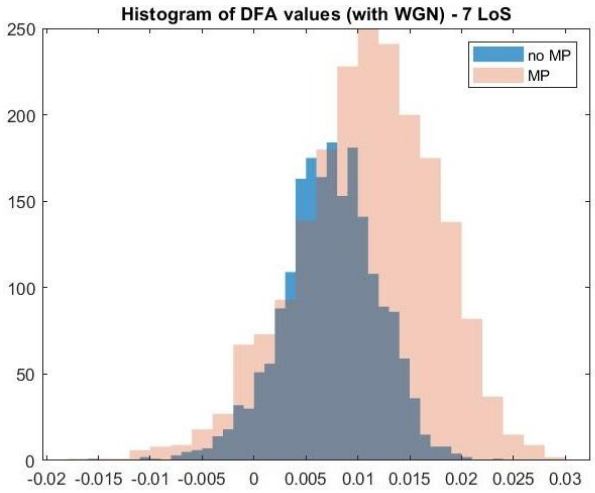

Histogram of DFA values (No added noise) - 4 LoS
no MP
MP
250
200
150
100
50
0
-0.02
-0.01
0
0.01
0.02
0.03
0.04

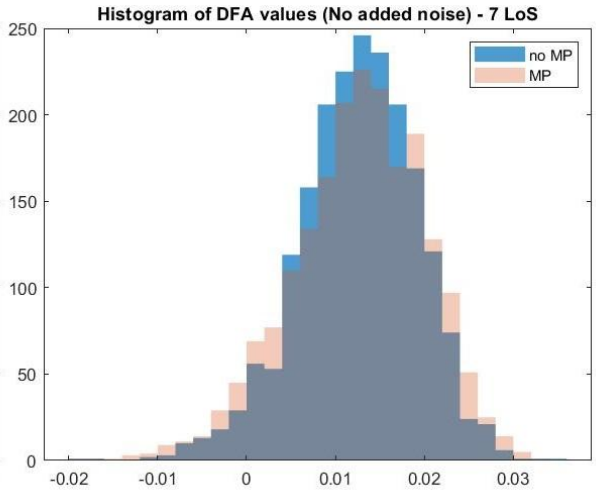

Histogram of DFA values (with WGN) - 5 LoS
no MP
MP
250
200
150
100
50
0
-0.015
-0.01
-0.005
0
0.005
0.01
0.015
0.02
0.025
0.03

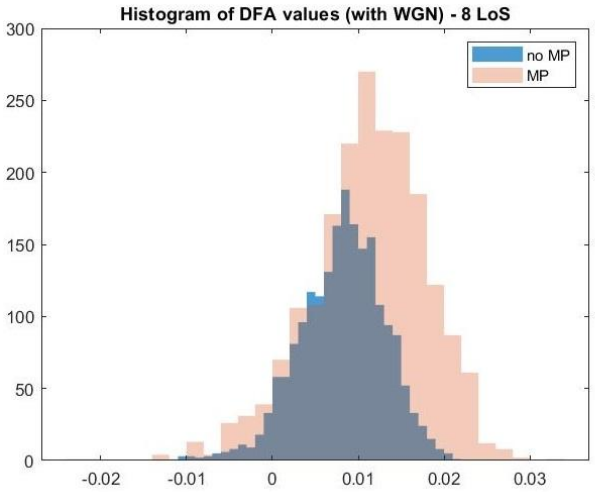

Histogram of DFA values (No added noise) - 5 LoS
no MP
MP
300
250
200
150
100
50
0
-0.02
-0.01
0
0.01
0.02
0.03
0.04

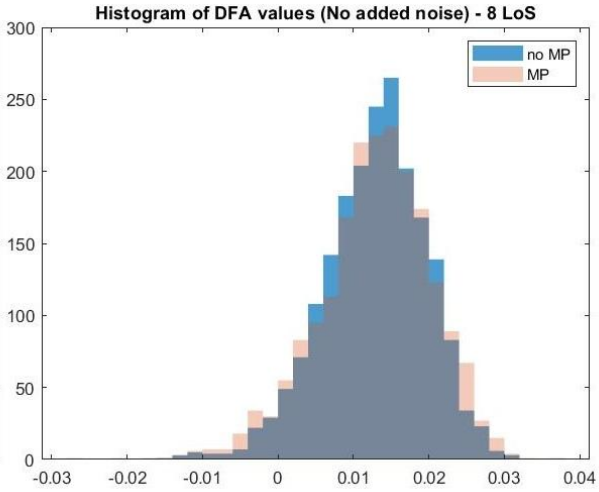

Histogram of DFA values (with WGN) - 6 LoS
no MP
MP
250
200
150
100
50
0
-0.015
-0.01
-0.005
0
0.005
0.01
0.015
0.02
0.025
0.03

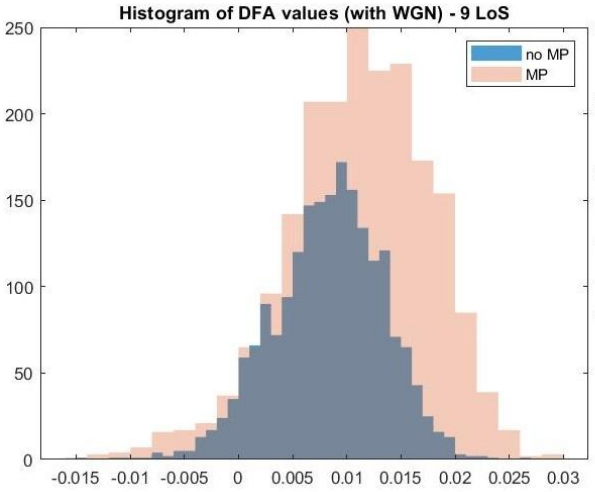

Histogram of DFA values (No added noise) - 6 LoS
no MP
MP
300
250
200
150
100
50
0
-0.02
-0.01
0
0.01
0.02
0.03

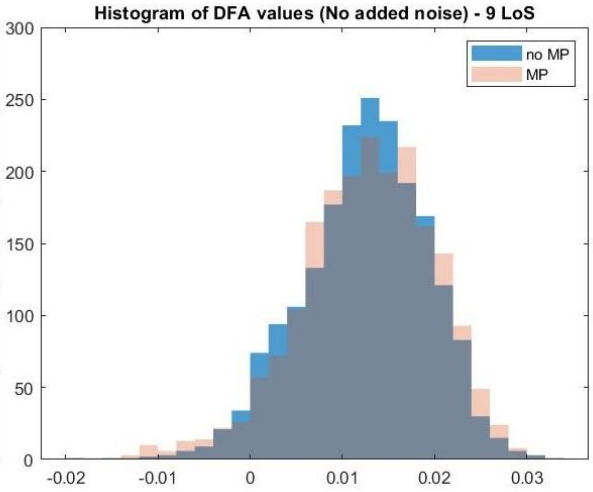

Histogram of DFA values (with WGN) - 7 LoS
no MP
MP
250
200
150
100
50
0
-0.02 -0.015 -0.01 -0.005 0 0.005 0.01 0.015 0.02 0.025 0.03


Histogram of DFA values (No added noise) - 7 LoS
no MP
MP
250
200
150
100
50
0
-0.02 -0.01 0 0.01 0.02 0.03


Histogram of DFA values (with WGN) - 8 LoS
no MP
MP
300
250
200
150
100
50
0
-0.02 -0.01 0 0.01 0.02 0.03


Histogram of DFA values (No added noise) - 8 LoS
no MP
MP
300
250
200
150
100
50
0
-0.03 -0.02 -0.01 0 0.01 0.02 0.03 0.04


Histogram of DFA values (with WGN) - 9 LoS
no MP
MP
250
200
150
100
50
0
-0.015 -0.01 -0.005 0 0.005 0.01 0.015 0.02 0.025 0.03


Histogram of DFA values (No added noise) - 9 LoS
no MP
MP
300
250
200
150
100
50
0
-0.02 -0.01 0 0.01 0.02 0.03

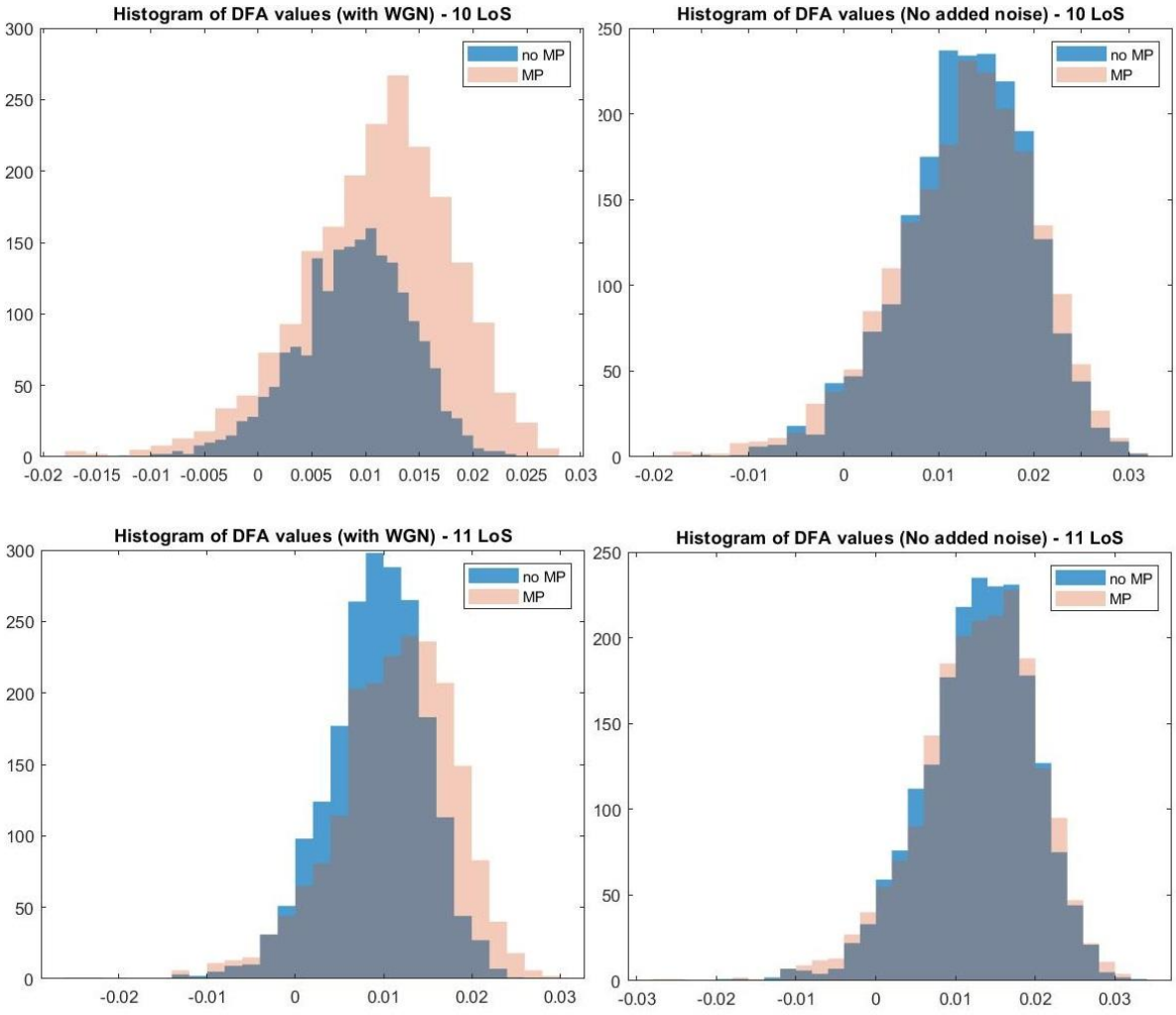
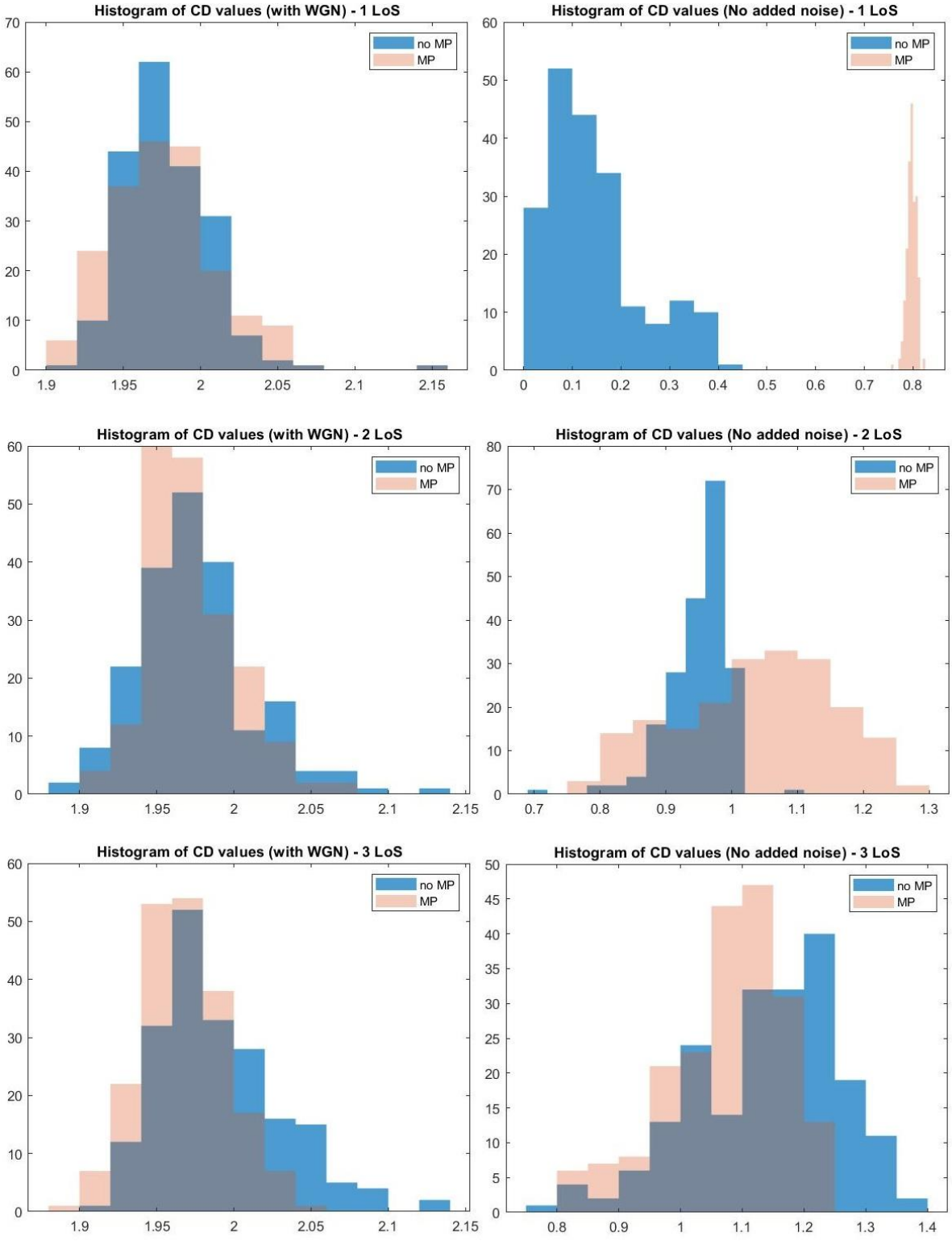
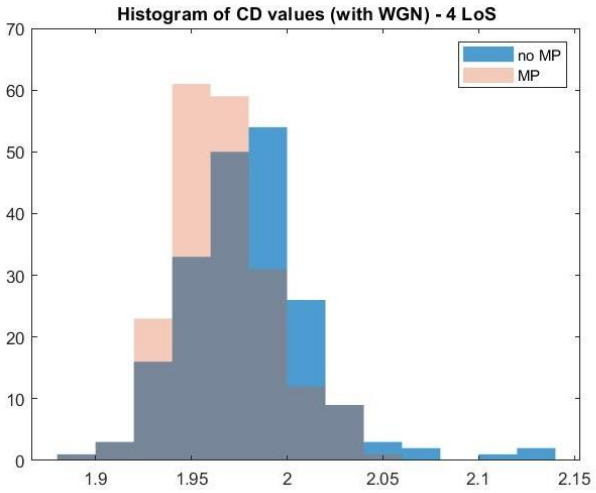


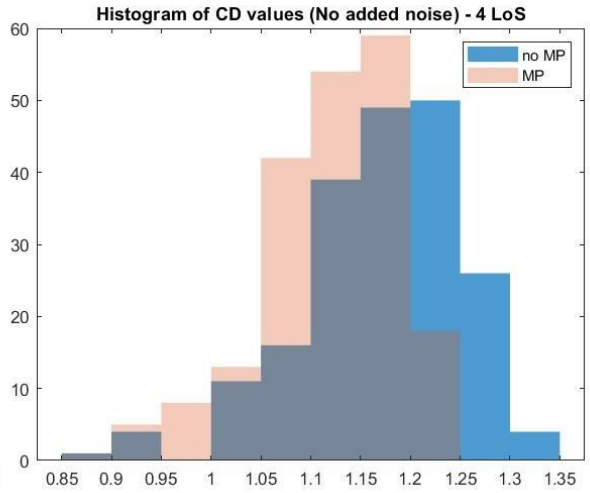
Figure 40 - DFA histograms after 2000 iterations

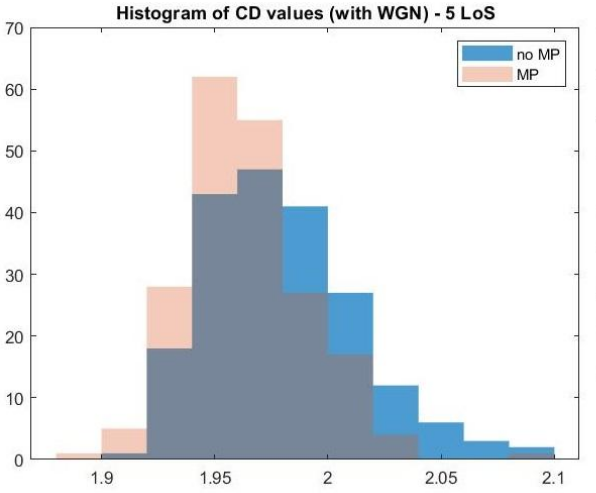
A.3 Correlation dimension

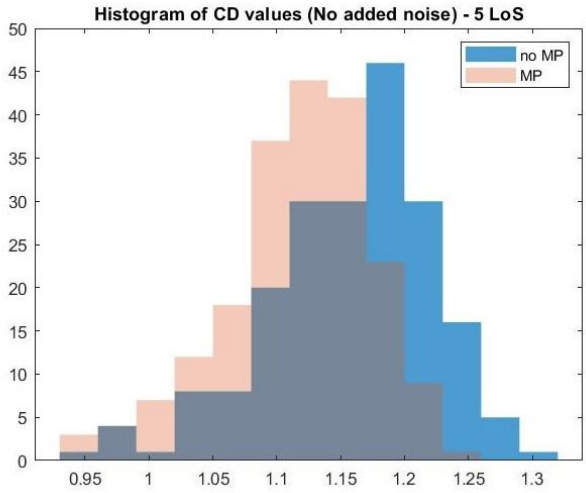
Due to the lengthy runtime of the CD implementation within the numerical simulator the number of iterations performed were 200 and the statistics are shown below:

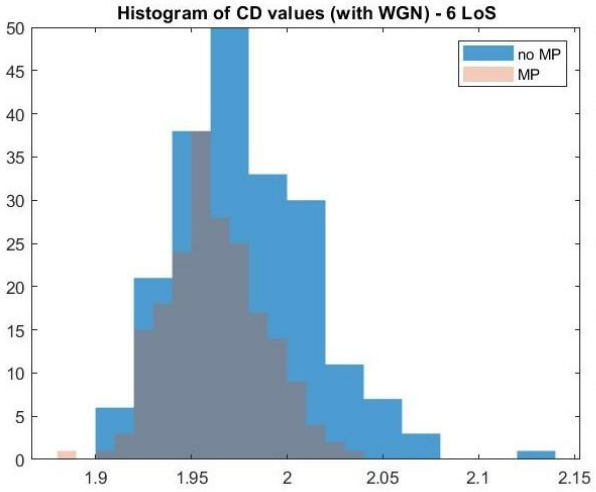


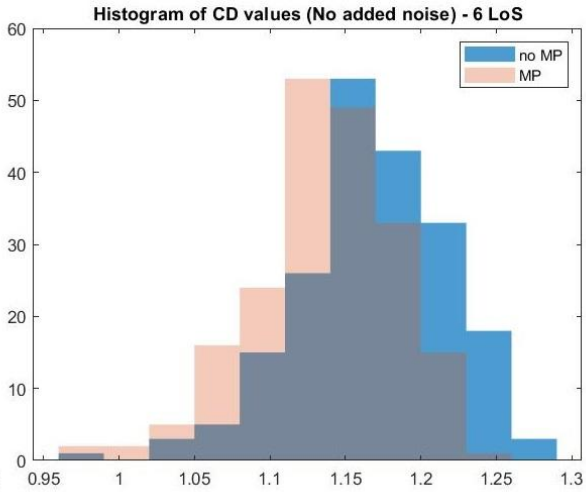

Histogram of CD values (with WGN) - 4 LoS
no MP
MP
70
60
50
40
30
20
10
0
1.9
1.95
2
2.05
2.1
2.15

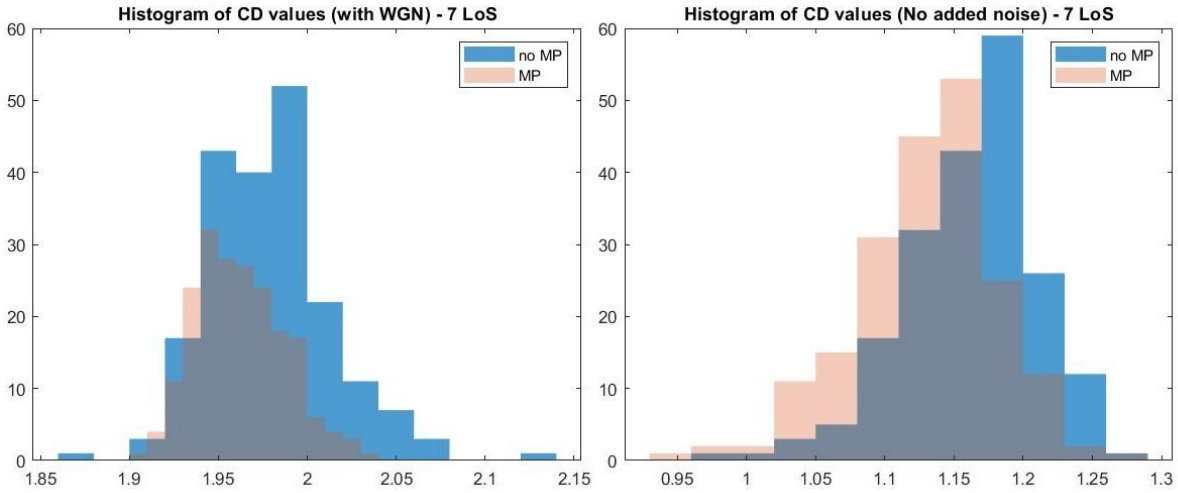

Histogram of CD values (No added noise) - 4 LoS
no MP
MP
60
50
40
30
20
10
0
0.85
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3
1.35


Histogram of CD values (with WGN) - 5 LoS
no MP
MP
70
60
50
40
30
20
10
0
1.9
1.95
2
2.05
2.1

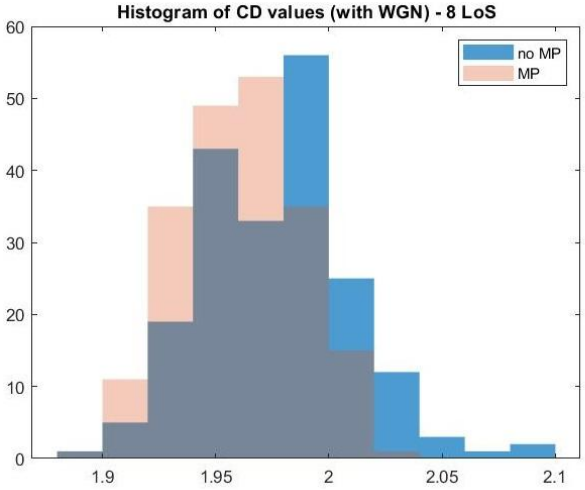

Histogram of CD values (No added noise) - 5 LoS
no MP
MP
50
45
40
35
30
25
20
15
10
5
0
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3

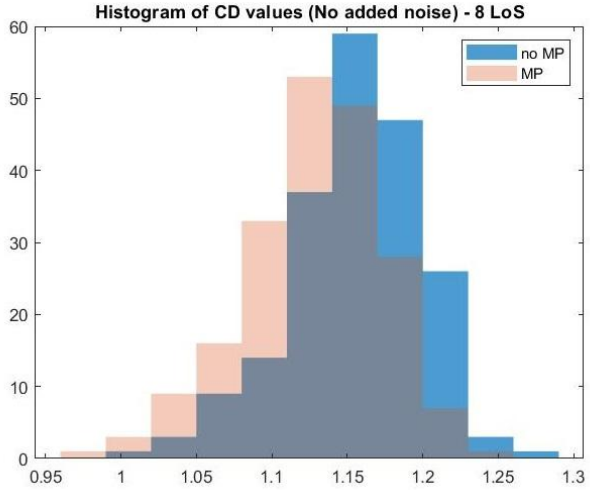

Histogram of CD values (with WGN) - 6 LoS
no MP
MP
50
45
40
35
30
25
20
15
10
5
0
1.9
1.95
2
2.05
2.1
2.15

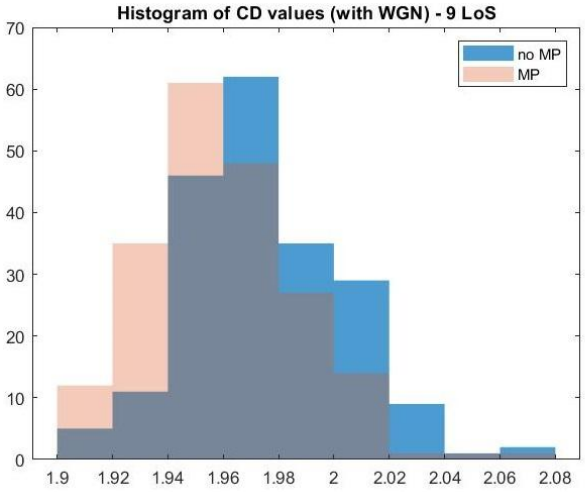

Histogram of CD values (No added noise) - 6 LoS
no MP
MP
60
50
40
30
20
10
0
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3

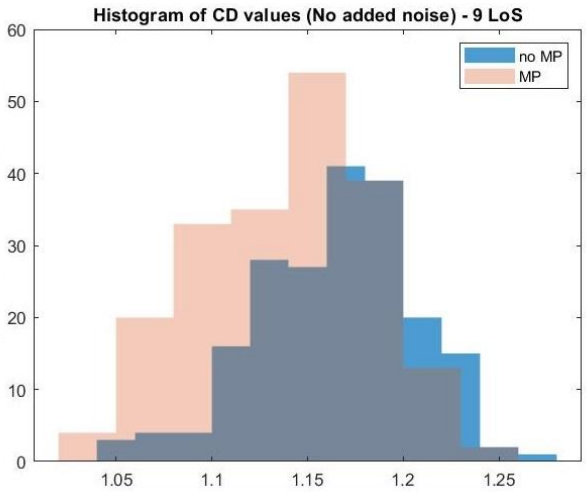

Histogram of CD values (with WGN) - 7 LoS
no MP
MP
0
10
20
30
40
50
60
1.85
1.9
1.95
2
2.05
2.1
2.15

Histogram of CD values (No added noise) - 7 LoS
no MP
MP
0
10
20
30
40
50
60
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3


Histogram of CD values (with WGN) - 8 LoS
no MP
MP
0
10
20
30
40
50
60
1.9
1.95
2
2.05
2.1


Histogram of CD values (No added noise) - 8 LoS
no MP
MP
0
10
20
30
40
50
60
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3


Histogram of CD values (with WGN) - 9 LoS
no MP
MP
0
10
20
30
40
50
60
70
1.9
1.92
1.94
1.96
1.98
2
2.02
2.04
2.06
2.08


Histogram of CD values (No added noise) - 9 LoS
no MP
MP
0
10
20
30
40
50
60
1.05
1.1
1.15
1.2
1.25

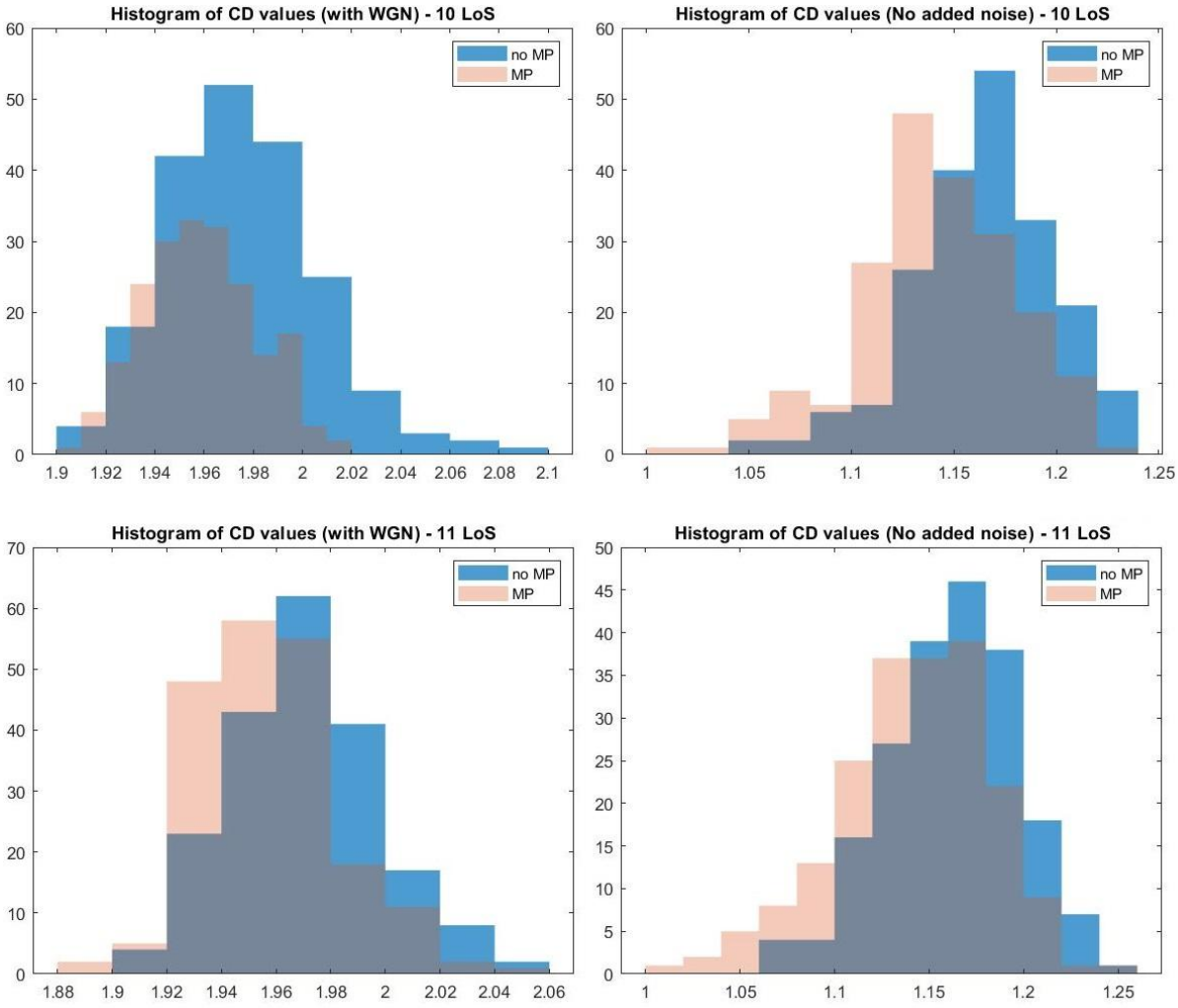
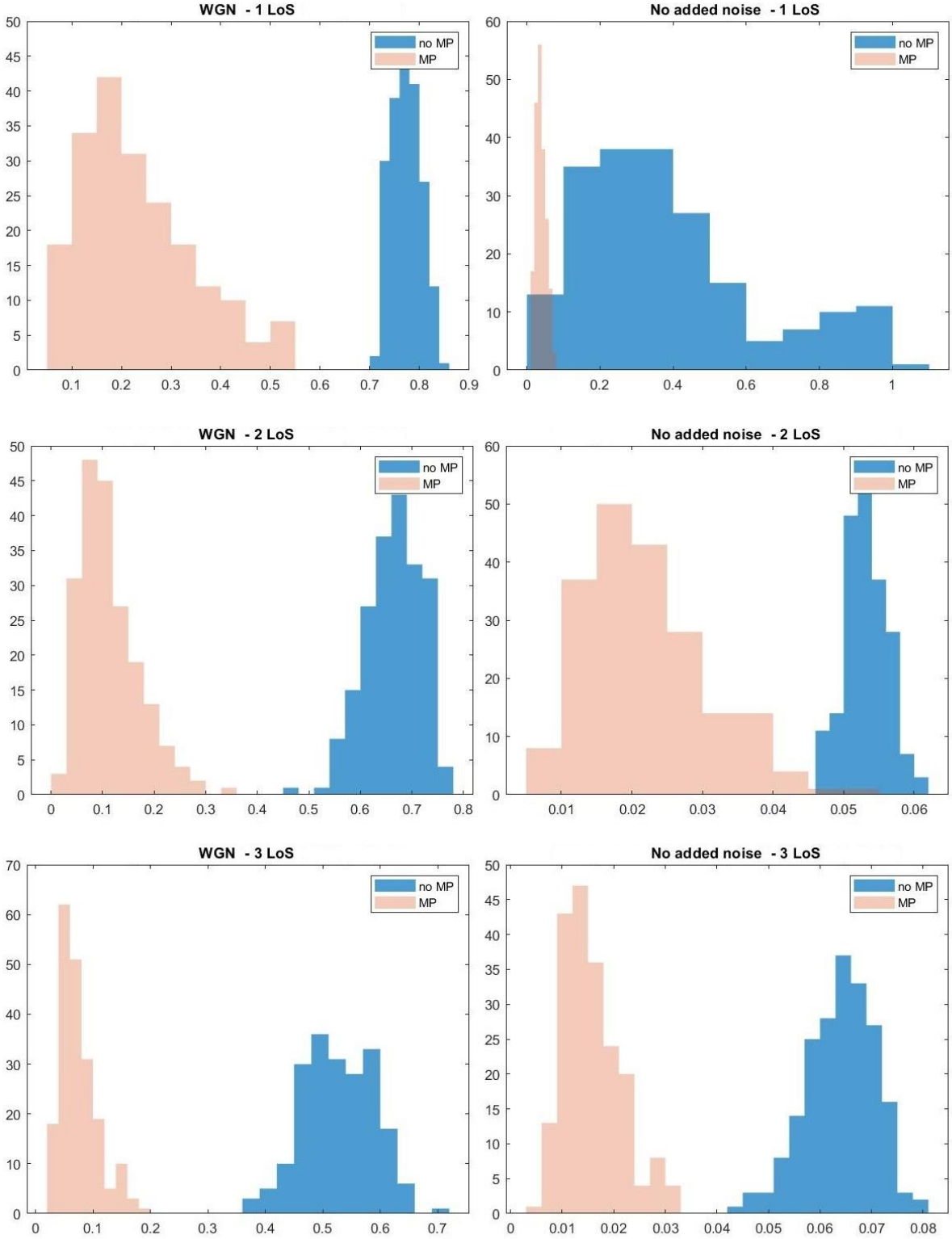


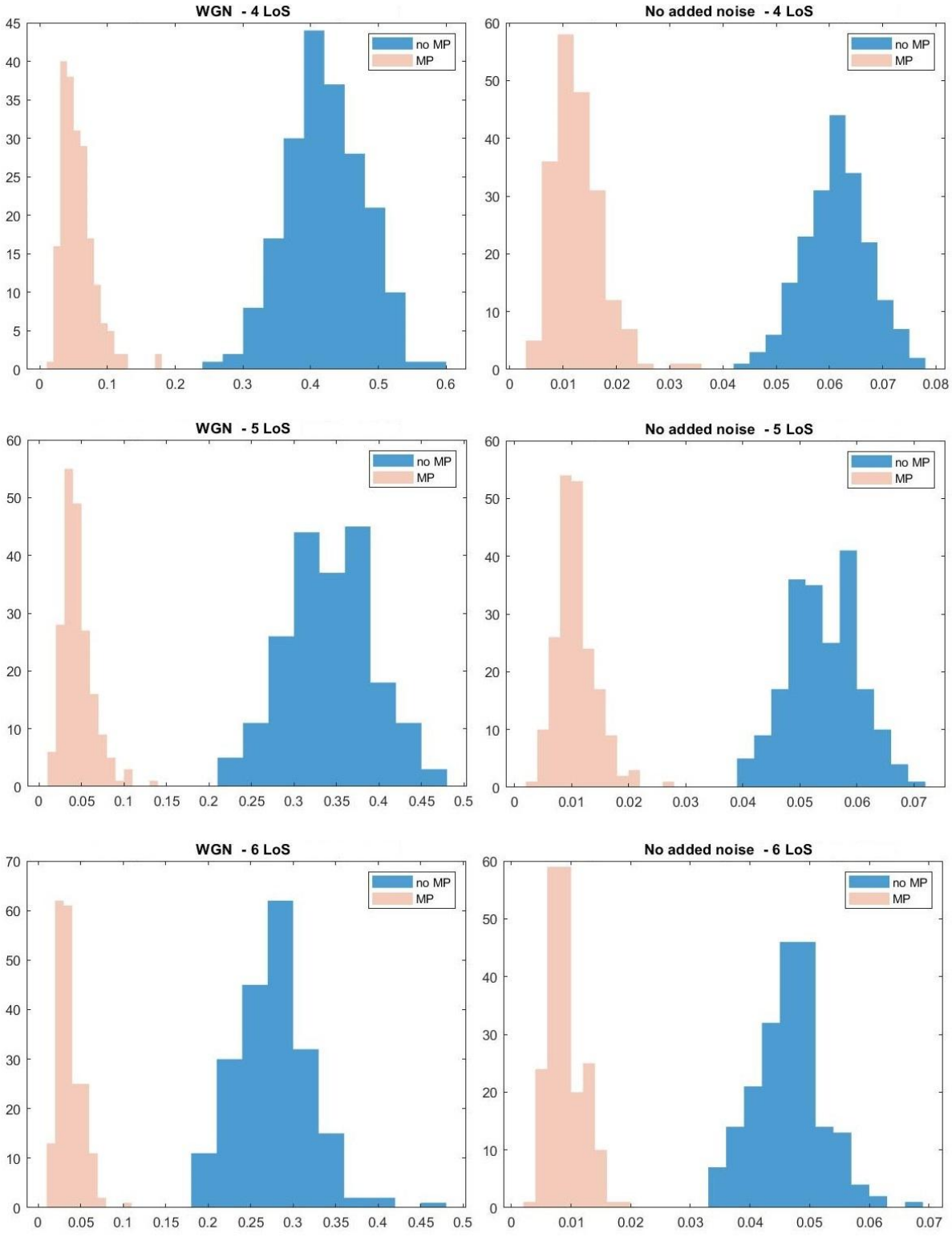
Figure 41 - CD histograms after 200 iterations

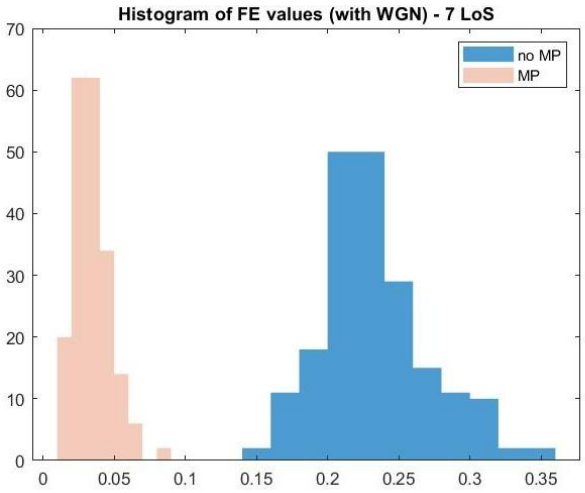
Inspection of histograms generated from the noisy versions show skewed distributions and long tails (kurtosis), and the mode appears to be flattened. Same effect was observed to the MP-afflicted results and their histogram representation, without the long tails. The combined effects of random noise and MP showed a relatively more pronounced mode.

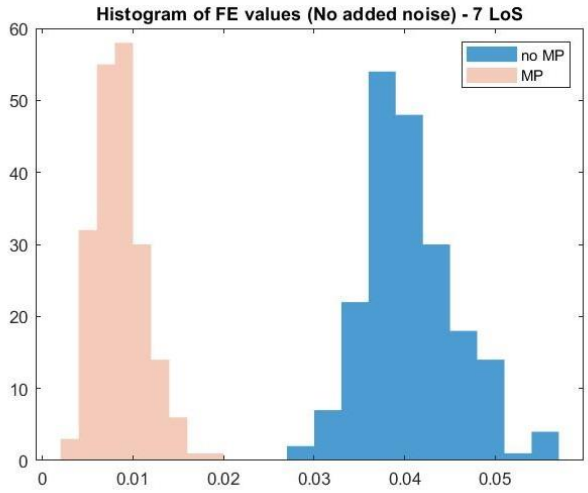
A.4 Fuzzy entropy

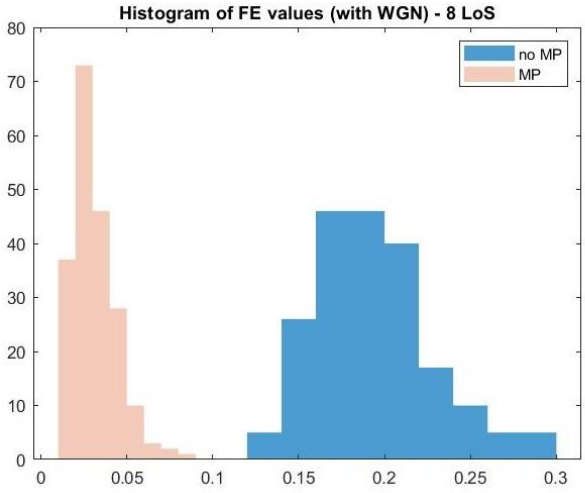
The following histograms are generated from 200 iterations/particles for the FE measure simulation:

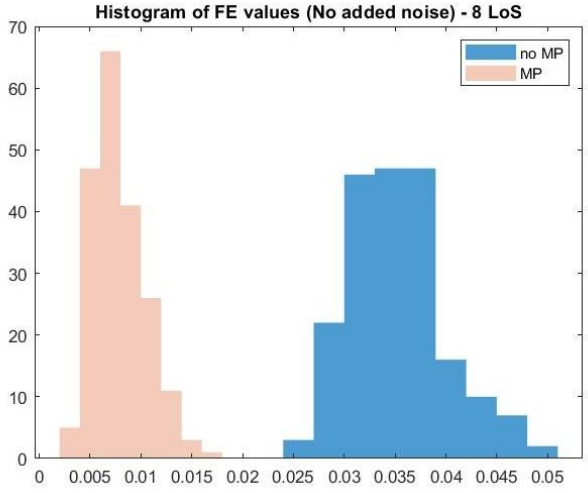


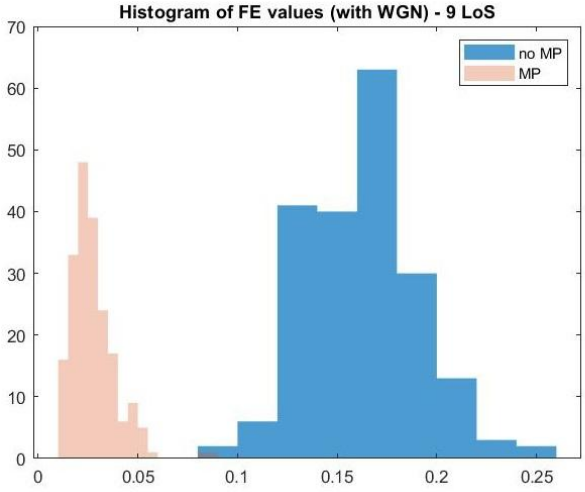

WGN - 4 LoS
no MP
MP
No added noise - 4 LoS
no MP
MP
WGN - 5 LoS
no MP
MP
No added noise - 5 LoS
no MP
MP
WGN - 6 LoS
no MP
MP
No added noise - 6 LoS
no MP
MP

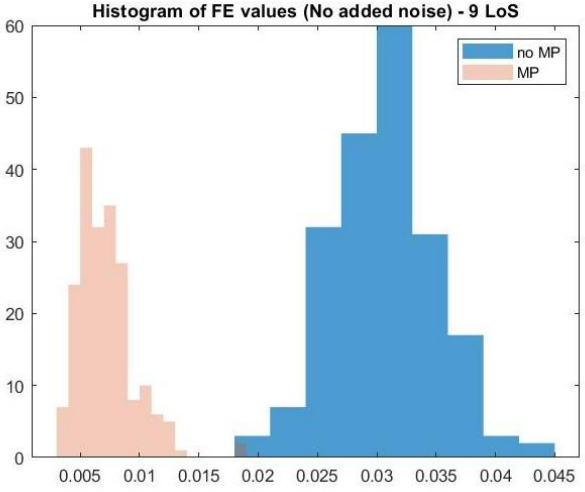

Histogram of FE values (with WGN) - 7 LoS
no MP
MP


Histogram of FE values (No added noise) - 7 LoS
no MP
MP


Histogram of FE values (with WGN) - 8 LoS
no MP
MP


Histogram of FE values (No added noise) - 8 LoS
no MP
MP


Histogram of FE values (with WGN) - 9 LoS
no MP
MP


Histogram of FE values (No added noise) - 9 LoS
no MP
MP

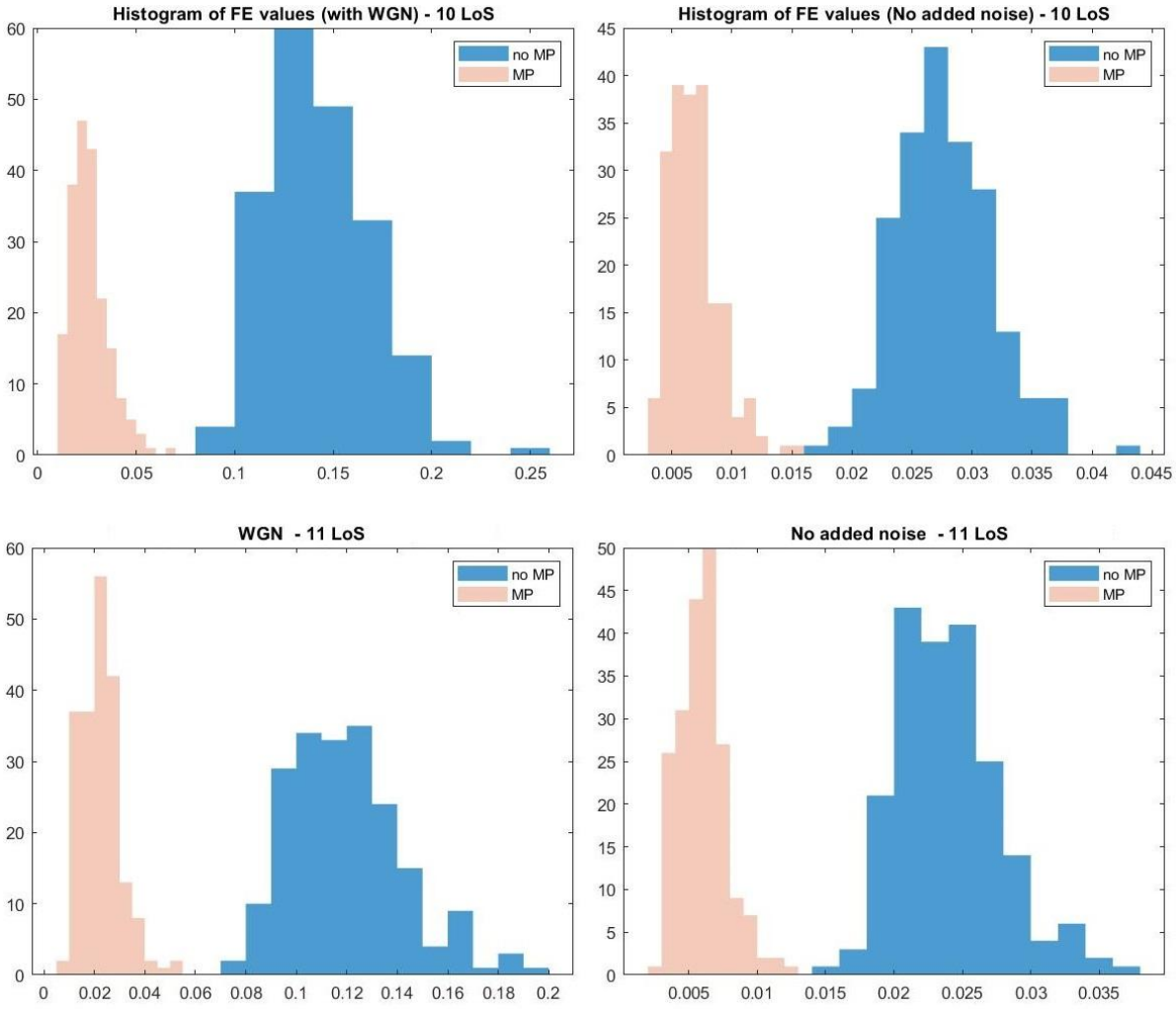
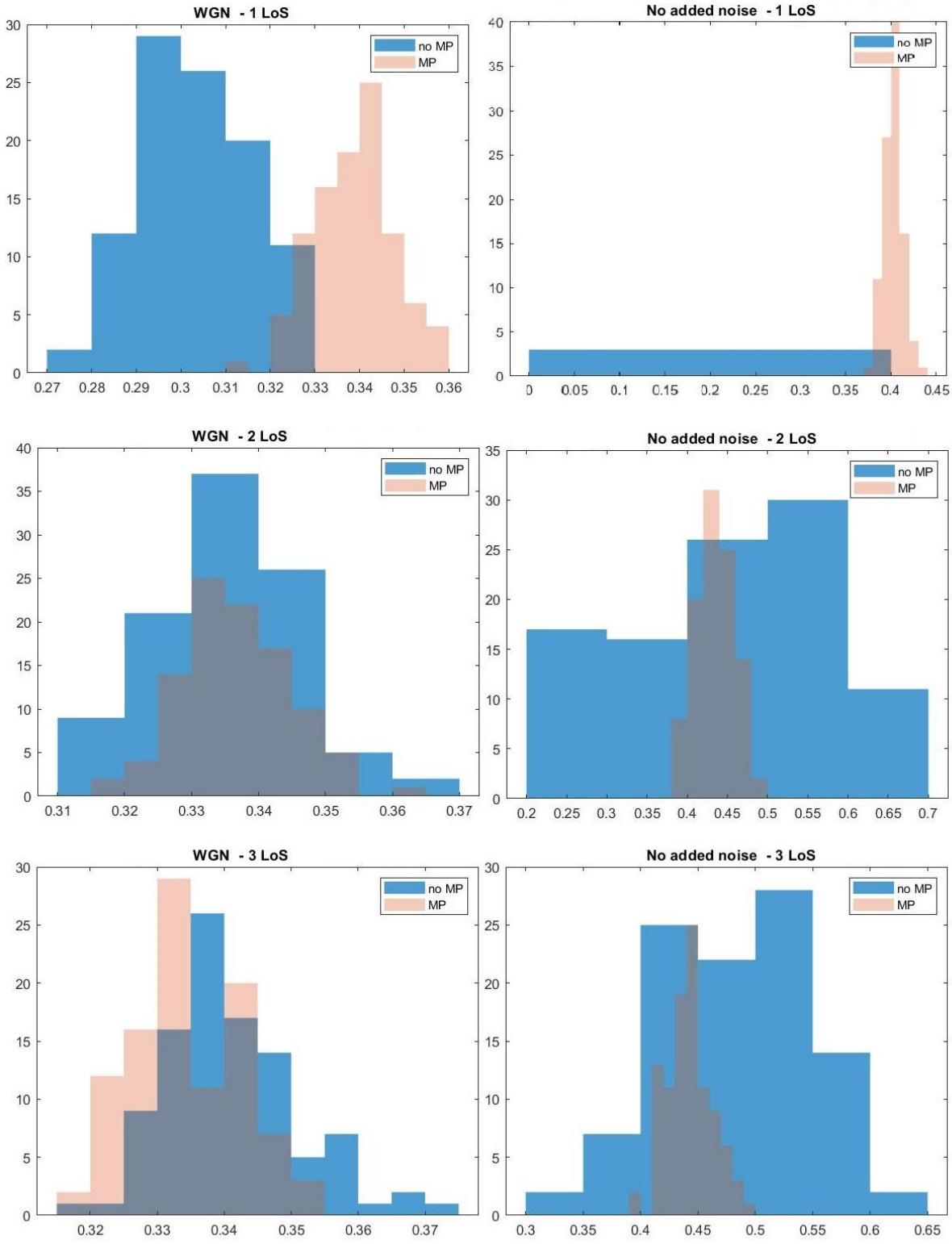
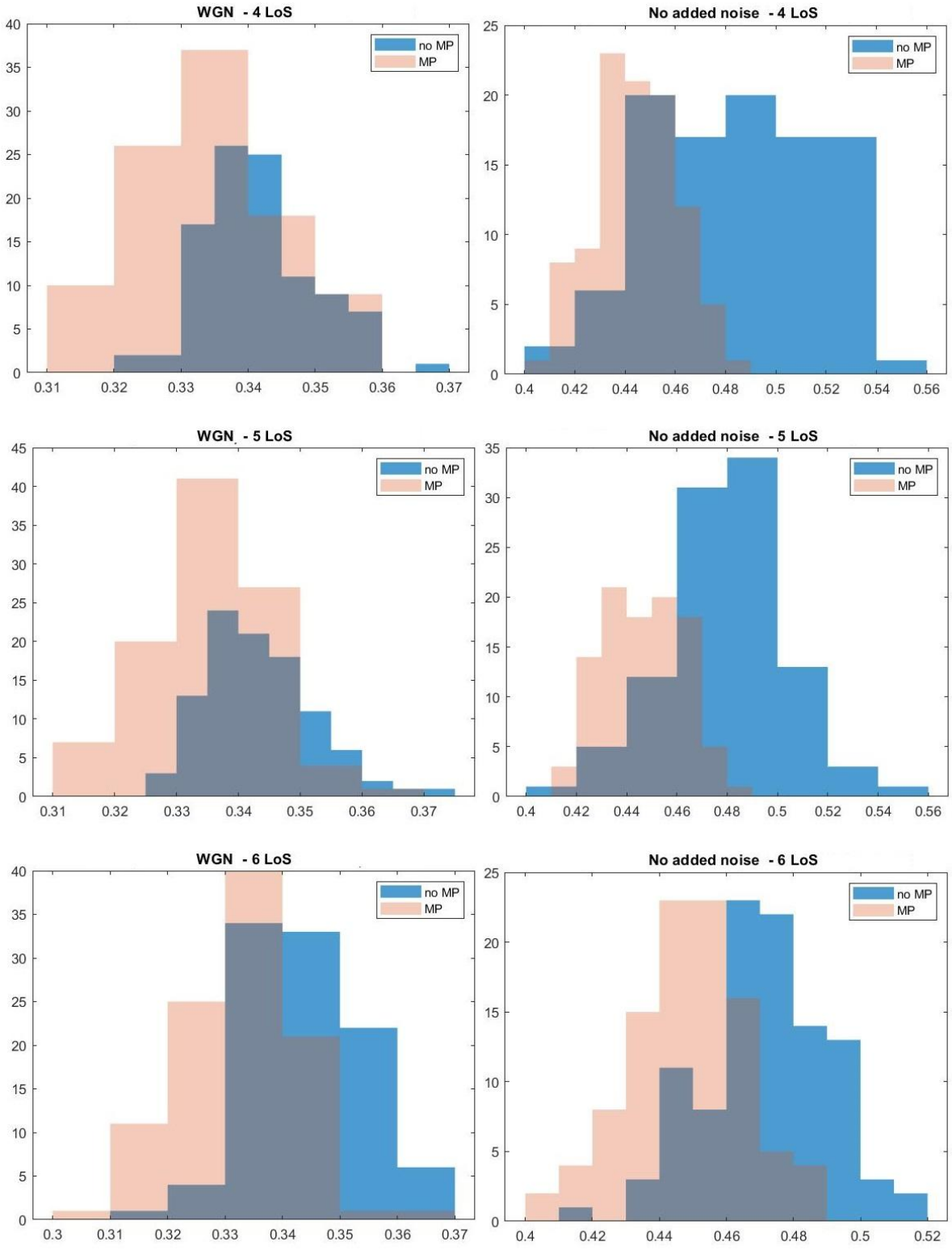


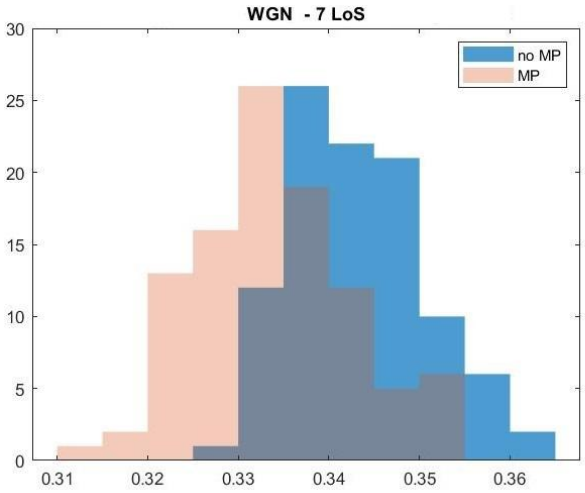
Figure 42 - FE histograms after 200 iterations

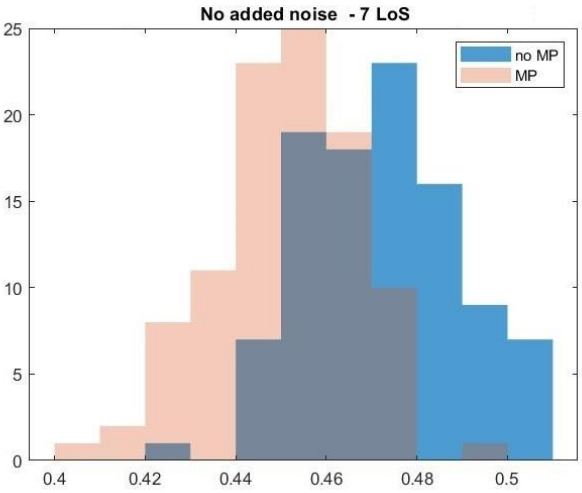
A.5 Recurrence period density entropy

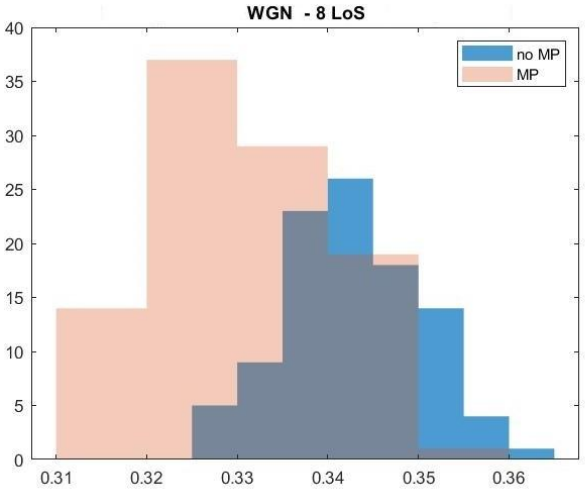
The following histograms are generated from 100 iterations/particles:

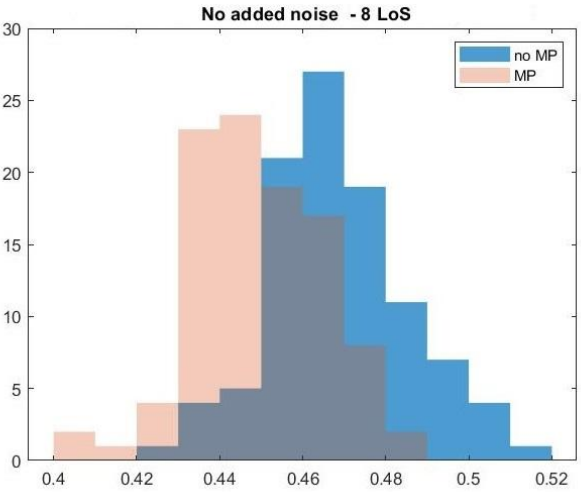


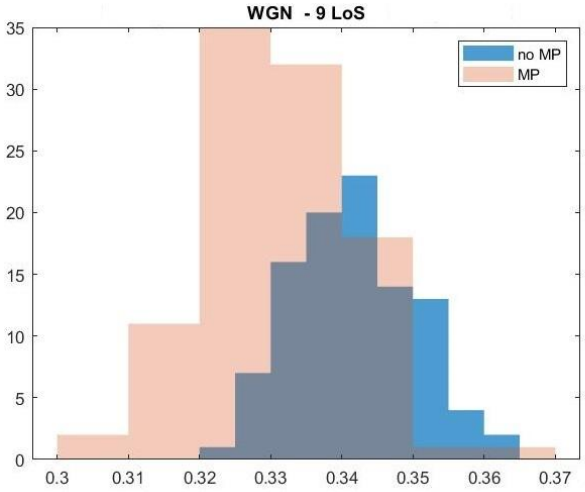

WGN - 4 LoS
no MP
MP
No added noise - 4 LoS
no MP
MP
WGN - 5 LoS
no MP
MP
No added noise - 5 LoS
no MP
MP
WGN - 6 LoS
no MP
MP
No added noise - 6 LoS
no MP
MP

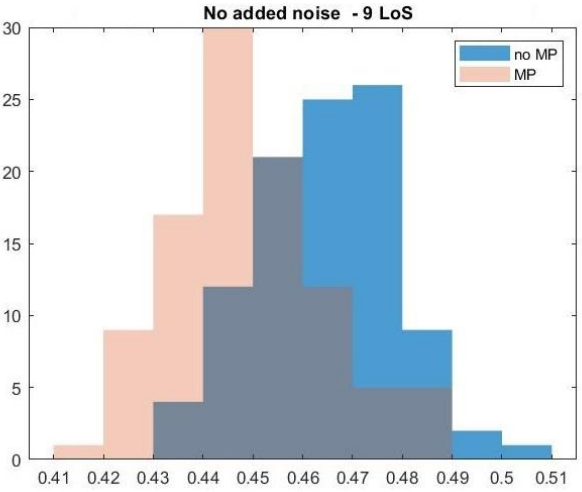

WGN - 7 LoS
no MP
MP


No added noise - 7 LoS
no MP
MP


WGN - 8 LoS
no MP
MP


No added noise - 8 LoS
no MP
MP


WGN - 9 LoS
no MP
MP


No added noise - 9 LoS
no MP
MP

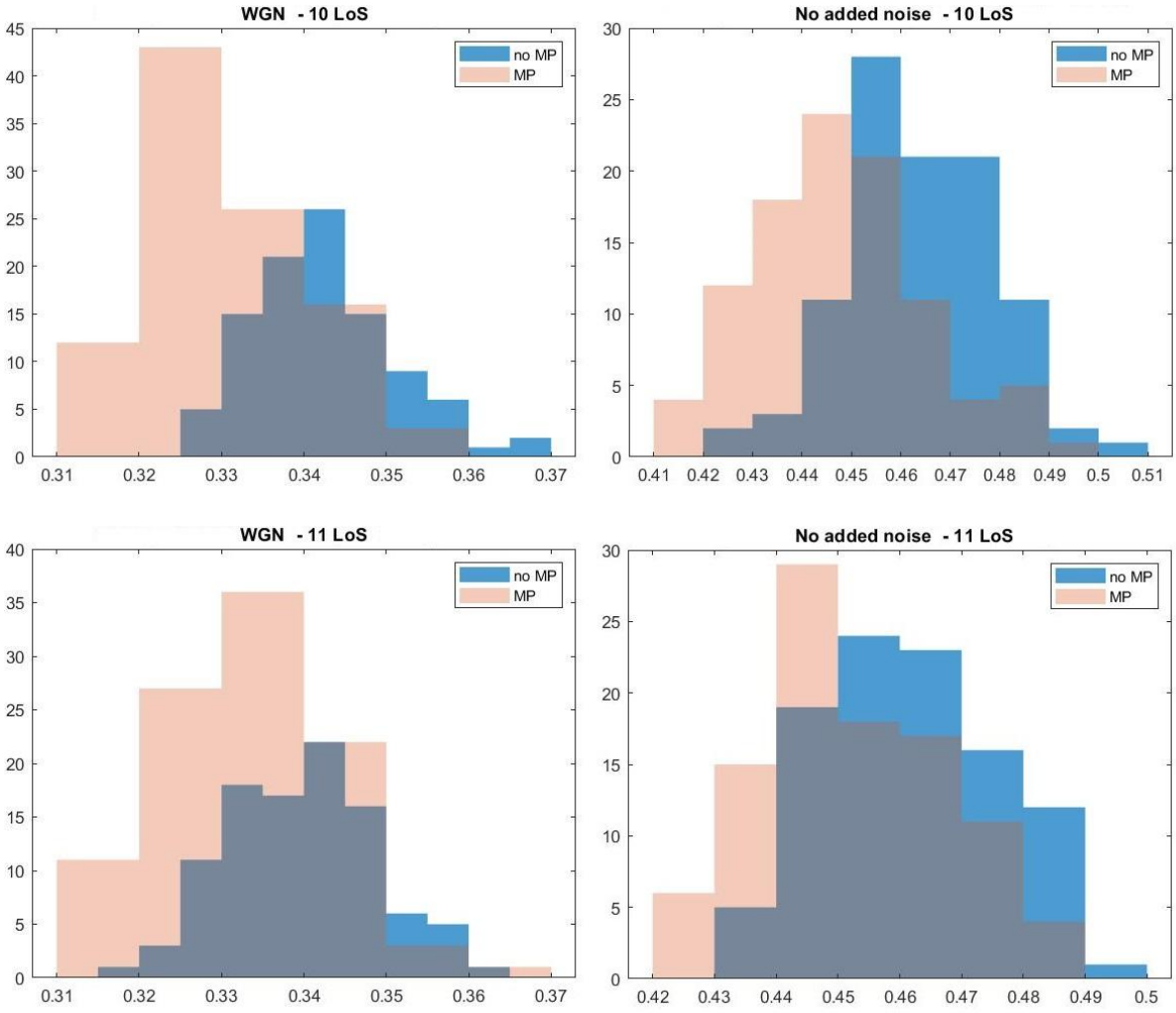


Figure 43 - RPDE histograms after 100 iterations

Appendix B – Simulation GHE values under different SNRs

In this Appendix, the results from extended simulations on the GHE computation are presented. 2000 iterations were configured for four different average SNRs and the GHE was computed for both $q = 1,2$.

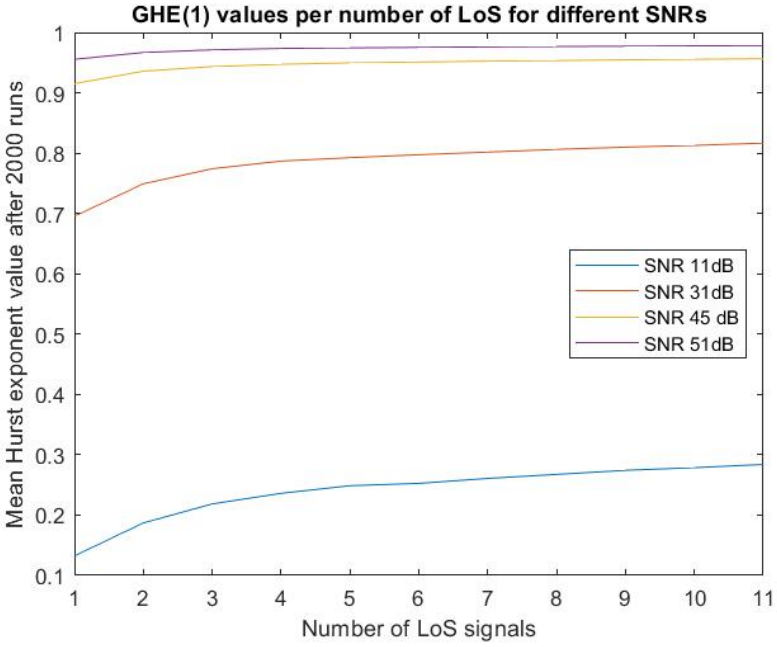


Figure 44 - GHE (q=1) simulated functions for different SNRs

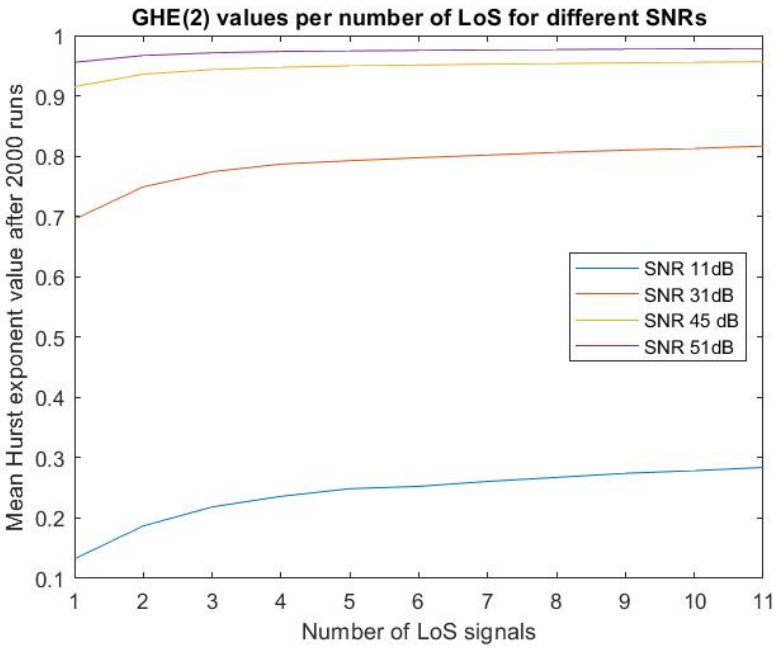


Figure 45 - GHE (q=2) simulated functions for different SNRs

As shown, for either value of q the results are very similar, and the higher the SNR, the more "flat" the resulting functions they get. This result suggests that the GHE metric might perform better in "characterizing" the different signals (simulation scenarios) in lower SNRs or "weaker" signals.