Thèse de Doctorat

Mention : Sciences et technologies de l'information et de la communication
Spécialité : Traitement du signal et des images

présentée à *l'École Doctorale en Sciences Technologie et Santé (ED 585)*

de l'Université du Littoral Côte d'Opale

par

Sarah El Hajj Chehade

pour obtenir le grade de Docteur de l'Université du Littoral Côte d'Opale

Statistical Modeling of Amplitude and Phase in GNSS Reflectometry for Surface Classification

Soutenue le 23 Janvier 2026, après avis des rapporteurs, devant le jury d'examen :

M. Guillaume Ginolhac, Professeur, Université Savoie Mont Blanc	**Président**
M. Adriano José Camps, Professeur, Universitat Politècnica de Catalunya	**Rapporteur**
M. Jocelyn Chanussot, Professeur, Université Grenoble Alpes	**Examinateur**
M. Manuel Martin-Neira, Ingenieur, ESA	**Examinateur**
M. Maximilian Semmling, Chercheur, DLR	**Examinateur**
M. Oussama Bazzi, Professeur, Université Libanaise	**Examinateur**
M. Serge Reboul, Professeur, Université du Littoral Côte d'Opale	**Directeur de thèse**
M. Georges Stienne, MCF-HDR, Université du Littoral Côte d'Opale	**Co-directeur**

Thèse de Doctorat

Mention: Information and Communication Sciences and Technologies
Specialty : Signal and Image Processing

presented to *the Doctoral School in Science, Technology and Health (ED 585)*

of the Université du Littoral Côte d'Opale

by

Sarah EL HAJJ CHEHADE

to obtain the degree of Doctor of Philosophy from the Université du Littoral Côte d'Opale

Statistical Modeling of Amplitude and Phase in GNSS Reflectometry for Surface Classification

Defended on January 23, 2026, after the approval of the reviewers, in front of the examining board:

M. Guillaume Ginolhac, Professor, Université Savoie Mont Blanc	**President**
M. Adriano José Camps, Professor, Universitat Politècnica de Catalunya	**Reviewer**
M. Jocelyn Chanussot, Professor, Université Grenoble Alpes	**Examiner**
M. Manuel Martin-Neira, Engineer, ESA	**Examiner**
M. Maximilian Semmling, Researcher, DLR	**Examiner**
M. Oussama Bazzi, Professor, Université Libanaise	**Examiner**
M. Serge Reboul, Professor, Université du Littoral Côte d'Opale	**Thesis Director**
M. Georges Stienne, Associate Professor, Université du Littoral Côte d'Opale	**Thesis co-director**

To my parents and the memory of my grandparents.

Acknowledgements

The completion of this doctoral work marks the end of an intense scientific and personal journey, and it is an opportunity to express my sincere gratitude to all those who, at one time or another, made it possible or contributed to its fulfillment. First and foremost, I would like to thank my thesis supervisor, Professor Serge Reboul, for the three years of continuous guidance, trust, and collaboration we shared. I am deeply grateful for his unwavering support, scientific rigor, and constant encouragement throughout this work. His clear vision and high standards played an essential role in the progress and completion of this thesis.

I would also like to express my sincere and warm thanks to my co-supervisor, Dr. Georges Stienne, whose support, availability, and guidance were invaluable throughout this doctoral work. His expertise, thoughtful feedback, and constructive discussions greatly contributed to shaping this research. I am truly grateful for both his mentorship and friendship.

I would like to thank all the members of my thesis jury for accepting to evaluate this work and for the time and effort devoted to its assessment. I am honored by their presence and grateful for their valuable comments, remarks, and suggestions, which contributed to improving the quality and clarity of this manuscript.

I gratefully acknowledge Université du Littoral Côte d'Opale (ULCO) for providing me with the opportunity and institutional framework to carry out this doctoral research. I would also like to thank Pôle Métropolitain de la Côte d'Opale (PMCO) for the financial support that made this work possible.

My sincere thanks go to all the members of the LISIC laboratory for their support and for the enriching exchanges, related to both research and everyday life, that accompanied my PhD journey. I also thank the administrative and technical staff for their constant availability and assistance.

I would like to express my special thanks to Dr. Hamza Issa and Dr. Mohamad Choaib, with whom I shared the doctoral journey, for their constant support, valuable scientific discussions, and practical help throughout these years. Beyond their scientific involvement, their friendship and presence made this period more motivating and easier to go through.

I would like to extend my deepest gratitude to Safae Abouelouafa and Rita Arfoul, who have been far more than colleagues met at the laboratory. Their constant support, both professionally and personally, played a crucial role in helping me move forward during this PhD. Their presence, encouragement, and generosity, whether through scientific discussions, daily support, or their involvement in the preparation of this defense, from organization to rehearsals, were invaluable. I am profoundly thankful for everything we shared throughout together.

I would also like to thank my friends who were there for me and helped me maintain balance throughout this demanding period, particularly Othmane El Attiq, Maxime Lernould, Dr. Dharmendra Singh, and Sharanya Dattamandal, for their friendship, encouragement, and support. My gratitude also extends to all my friends in France and in Lebanon who, in different ways, contributed to making this journey lighter and more meaningful. I would also like to thank Dr. Williams Kouassi, my laboratory colleague during my first year, with whom I remain in contact, for his support and the valuable exchanges we shared.

Finally, and most importantly, I would like to express my deepest and most heartfelt gratitude to my family. They have been my greatest source of strength, support, and motivation throughout this journey. Their unconditional love, constant encouragement, and unwavering belief in me never faltered, even during the most challenging moments.

Contents

Acknowledgements v

List of Figures xi

List of Tables xv

List of Abbreviations xvii

List of Symbols xix

General Introduction 1

1 The Evolution of GNSS-R Signal Processing: Towards Phase-Aware Remote Sensing 9

1.1 Introduction 9

1.2 GNSS-R Signal Processing Evolution 10

1.2.1 Traditional Power-Based Approaches and Their Limitations 10

1.2.2 Evolution of Correlation and Demodulation Techniques 13

1.2.3 Coherent vs Incoherent Integration Strategies 14

1.2.4 Multi-Constellation Processing Developments 16

1.2.5 GNSS-R Fundamentals: Geometry, Signal Paths, and Coherence Conditions 17

1.3 Phase Information in GNSS-R: The Underexploited Domain . 18

1.3.1 Evolution of Phase-Based Approaches 18

1.3.2 Theoretical Foundations 20

1.3.3 Advanced Coherence Detection Algorithms 21

1.3.4 Current State and Research Gaps 23

1.3.5 Critical Research Gaps Requiring Investigation 24

1.4 Experimental Framework and Data Foundation 25

1.4.1 Airborne GNSS-R Experimental Campaign 25

1.4.2 Platform and Instrumentation 25

1.4.3 GNSS-R Signal Processing Architecture 27

1.4.4 Data Quality and Validation Framework 29
1.5 Conclusion 31

2 Literature Review on Change Point Detection in Reflectometry Signals 33
2.1 Introduction 33
2.2 Different Statistical Models for Speckle Noise 34
2.2.1 Empirical models 34
2.2.2 Physically Based Models 37
2.2.3 Discussion 41
2.3 Single Change Point Detection: Methods and Algorithms . . . 42
2.3.1 Unsupervised 43
2.3.2 Implementation 55
2.3.3 Tuning 58
2.4 CPD for Multiplicative Noise 60
2.4.1 Sequential Detection under Gamma Noise 60
2.4.2 Discussion 63
2.5 Conclusion 64

3 Segmentation of the GNSS-R Reflectivity Signals affected by Speckle Noise 67
3.1 Introduction 67
3.2 Statistical model of the GNSS-R reflectivity 69
3.2.1 Direct signal tracking: Phase Locked Loop 69
3.2.2 Reflected signal tracking: Assisted open loop phase tracking 70
3.2.3 Reflectivity of the GNSS-R signal 72
3.2.4 Homomorphic transformation of the signal 73
3.3 Change point detection 75
3.3.1 Proposed change point detector 76
3.3.2 Proposed change point localization approach 78
3.3.3 Parameter definitions 80
3.3.4 Post-Localization Segment Merging Process 81
3.4 Experimentation: 84
3.4.1 Assessment on Synthetic Data 84
3.4.2 Application to Real Airborne GNSS-R data 87
3.5 Conclusion 97

4 GNSS-R Phase Modeling and Excess Doppler Analysis 101
4.1 Introduction . . . 101
4.2 Surface Geometry and Signal Representation . . . 102
4.2.1 From Fresnel Zone to Glistening Zone: The Impact of Surface Roughness . . . 103
4.2.2 Signal Composition and Scattering Components . . . 104
4.3 Signal Processing and Phase Estimation . . . 105
4.3.1 Linear Phase Assumption . . . 106
4.3.2 Numerical Implementation for Reflected Signal . . . 106
4.3.3 Phase Estimation and Coherence Analysis . . . 108
4.4 Excess Doppler Analysis . . . 109
4.4.1 Doppler Spectrum Computation . . . 109
4.4.2 Phase Coherence Across Surfaces . . . 112
4.4.3 Phase Model Verification . . . 113
4.5 Experimentation . . . 114
4.5.1 Synthetic Data Assessment . . . 114
4.5.2 Evaluation of Doppler-Based Criteria on Synthetic GNSS-R Signals . . . 119
4.5.3 Application of Excess Doppler Criteria on Real data . . . 121
4.6 Conclusion . . . 125

5 Machine Learning Applications for GNSS-R Surface Classification 127
5.1 Introduction . . . 127
5.2 Literature Review: Machine Learning in GNSS-R . . . 129
5.2.1 Machine Learning in Spaceborne GNSS-R . . . 129
5.2.2 ML Applications in Ocean Remote Sensing . . . 130
5.2.3 Land Surface Applications and Soil Moisture Retrieval 130
5.2.4 Cryosphere Applications and Sea Ice Classification . . . 132
5.2.5 Inland Water Body Detection and Hydrological Applications . . . 133
5.2.6 Forest Fire Disturbance Detection . . . 134
5.2.7 Methodological Advances and Algorithm Development 134
5.2.8 Performance Evaluation and Comparative Analysis . . . 136
5.2.9 Challenges and Future Directions . . . 136
5.3 Data Collection and Generation . . . 138
5.3.1 Real Airborne GNSS-R Dataset . . . 138
5.3.2 Synthetic Signal Generation . . . 139
5.3.3 Feature Engineering . . . 139

5.4 Data Preprocessing . . . 142
5.5 Model Selection . . . 142
5.6 Model Training and Evaluation . . . 143
5.6.1 Experimental Setup and Training Procedure . . . 143
5.6.2 Evaluation Strategy and Metrics . . . 144
5.7 Results . . . 145
5.7.1 Model Performance on Synthetic and Real Data . . . 145
5.7.2 Visual Comparison of Generalization . . . 146
5.8 Conclusion . . . 146

General Conclusion **149**

A Signal Processing Derivations for Multi-Path GNSS-R Phase Modeling **155**
A.1 Continuous-Time Reflected Signal Processing Model . . . 156
A.2 Statistical Phase Model . . . 156
A.2.1 Linear Phase Assumption . . . 156
A.2.2 In-phase and Quadrature Signal Modeling . . . 157

Résumé Étendu de la Thèse **175**
1 Modèle statistique de la réflectivité GNSS-R . . . 178
2 Segmentation de la réflectivité GNSS-R . . . 180
2.1 Modèle de transition . . . 180
2.2 Système de détection de ruptures en ligne et hors ligne 181
2.3 Application aux données GNSS-R aéroportées . . . 183
3 Modélisation de la phase GNSS-R et analyse d'excès Doppler 185
3.1 Traitement du signal et estimation de la phase . . . 185
3.2 Analyse du spectre d'excès Doppler . . . 187
3.3 Résultats de classification des surfaces . . . 189
4 Apprentissage automatique pour la classification des surfaces 190
4.1 Approche bi-dataset et ingénierie des caractéristiques . 190
4.2 Prétraitement des données et sélection des modèles . . 191
4.3 Performance et analyse de classification . . . 192

Résumé **195**

Abstract **197**

List of Figures

1.1 Timeline of spaceborne GNSS-R signal processing evolution. . 11
1.2 The flight trajectory [13]. 26
1.3 The gyrocopter used during the flights with the sensors embedded on it. 27
1.4 Flow chart of the base-band receiver processing. 28
1.5 GNSS-R processing chain overview: (a) Satellite footprint traces (PRN 7 & 5) over inland water bodies, (b) reflected signal amplitude from PRN 5, (c) corresponding phase showing coherence variations, and (d) automated segmentation and surface classification results. 31

2.1 Typology of the online change point detection algorithms. . . 43

3.1 Signal model in a working window. 76
3.2 Change detection system. 77
3.3 Variance estimate of the Homomorphic log-transform W . . . 88
3.4 Detection, Localization and Segmentation Process of an airborne GNSS signal reflectivity at (50.876064°, 1.948393°) 90
3.5 Segmentation of the GNSS measurements at (50.877390°, 1.933011°) – **Flight 1** 91
3.6 Segmentation of the GNSS-R reflectivity measurements at (50.878126°, 1.925893°) – **Flight 2**. 92
3.7 Reflectivity Classification for PRN 5, 7 & 30 – **Flight 1**. 94
3.8 Reflectivity Classification for PRN 8 & 21 – **Flight 2**. 95
3.9 Reflectivity reference areas 96

4.1 GNSS-R observation system. 103
4.2 Surface Roughness Impact on Reflection Characteristics. . . . 104
4.3 Geometric representation of coherent and incoherent GNSS-R scattering over calm and wavy sea surfaces, illustrating the Fresnel and glistening zones. 105

4.4 FFT spectrum of the reflected GNSS-R signal segment illustrating the computation of C^1, C^2 and C^3 (spectral dispersion). Data segment at $(50.876064^\circ, 1.948393^\circ)$ between 1770–1780 s. 112
4.5 Specular reflection point trajectory of GPS satellite PRN 5 over a 10-second segment, showing a transition from rough land (blue) to a smooth lake surface (cyan). 113
4.6 Corresponding phase of the reflected GNSS signal from GPS satellite PRN 5 over the same 10-second interval. 113
4.7 Comparison between the real phase observation of the reflected GNSS-R signal (blue) and the estimated phase computed from the I/Q components (red dashed) over a transition from land to lake segment. 114
4.8 Synthetic GNSS-R phase signal with surface type classification labels. 117
4.9 Phase signal over sand surface demonstrating coherent scattering with smooth phase evolution. 118
4.10 Phase signal over land surface showing incoherent scattering behavior with rapid phase fluctuations. 119
4.11 Global confusion matrices obtained on synthetic GNSS-R signals using (a) reflectivity-based parameters and (b) excess Doppler spectrum criteria. 121
4.12 Comparative FFT spectra and phase evolution extracted from Google Earth data for different surface types: (a) Land region at $(50.886062^\circ, 1.889681^\circ)$ (b) Lake region at $(50.888349^\circ, 1.872098^\circ)$ 122
4.13 Excess Doppler spectrum representation of significant peaks (in blue) and maximum peaks (in red), showing peak variations across different surface types. 124
4.14 Classification of the Max Peaks Values. 125

5.1 Coefficient of variation (CV) of the amplitude versus CV of the spectral peaks for different surface types. 141
5.2 Amplitude skewness versus kurtosis across surface classes. . . 141
5.3 Model accuracy comparison between synthetic and real testing data. 147
5.4 Model F1-score comparison between synthetic and real testing data. 147

A.1 Modèle du signal dans une fenêtre de travail. 181
A.2 Système de détection de ruptures. 182

A.3 Segmentation des mesures GNSS à (50.877390°, 1.933011°). . . 184
A.4 Classification de la réflectivité pour les satellites PRN 5, 7 et 30. 185
A.5 Classification des valeurs de pics maximaux. 189
A.6 Comparaison de la précision des modèles entre les données de test synthétiques et réelles. 194
A.7 Comparaison du F1-score des modèles entre les données de test synthétiques et réelles. 194

List of Tables

3.1 Statistical parameters of different change point detectors as a function of the reflectivity mean of first and second segments (λ_0 & λ_1). The parameters are: the mean, the standard deviation and the number of detected changes (Nbr). 86
3.2 Overlapping Percentages $P(i, PRN)$ of GNSS-R Reflectivity Signals 96

4.1 Synthetic GNSS-R signal generation parameters. 115
4.2 Classification performance on synthetic GNSS-R signals using reflectivity-based parameters (mean and variance). 120
4.3 Classification performance on synthetic GNSS-R signals using excess Doppler spectrum criteria. 120

5.1 Performance comparison of machine learning and deep learning models on synthetic and real GNSS-R data. 145

A.1 Comparaison des performances des modèles d'apprentissage automatique et profond sur les données GNSS-R synthétiques et réelles. 193

List of Abbreviations

1D-CNN	**1D C**onvolutional **N**eural **N**etwork
AML	**A**pproximate **M**aximum **L**og-likelihood
ANN	**A**rtificial **N**eural **N**etwork
ARL	**A**verage **R**un **L**ength
Bi-LSTM	**Bi**directional **L**ong **S**hort-**T**erm **M**emory
BOCPD	**B**ayesian **O**nline **C**hange**p**oint **D**etection
BRCS	**B**istatic **R**adar **C**ross-**S**ection
CART	**C**lassification **A**nd **R**egression **T**rees
CDMA	**C**ode **D**ivision **M**ultiple **A**ccess
cGNSS-R	**c**onventional **GNSS-R**
CNN	**C**onvolutional **N**eural **N**etwork
CPD	**C**hange **P**oint **D**etection
CUSUM	**Cu**mulative **Sum**
CV	**C**oefficient of **V**ariation
CYGNSS	**Cy**clone **G**lobal **N**avigation **S**atellite **S**ystem
DBSCAN	**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise
DCNN	**D**eep **C**onvolutional **N**eural **N**etwork
DDM	**D**elay-**D**oppler **M**ap
DL	**D**eep **L**earning
DLL	**D**elay-**L**ocked **L**oop
DOC	**D**egree **o**f **C**oherence
DP	**D**ynamic **P**rogramming
EM	Expectation-**M**aximization
FDMA	**F**requency **D**ivision **M**ultiple **A**ccess
FFT	**F**ast **F**ourier **T**ransform
FLL	**F**requency-**L**ocked **L**oop
FMM	**F**inite **M**ixture **M**odel
GBDT	**G**radient **B**oosting **D**ecision **T**ree
GBI	**G**eneral **B**ayesian **I**nference
GGD	**G**eneralized **G**aussian **D**istribution
GGR	**G**eneralized **G**aussian **R**ayleigh
GLR	**G**eneralized **L**ikelihood **R**atio
GMLE	**G**amma **M**aximum **L**ikelihood **E**stimate
GMM	**G**aussian **M**ixture **M**odel
GNSS	**G**lobal **N**avigation **S**atellite **S**ystem
GNSS-R	**G**lobal **N**avigation **S**atellite **S**ystem **R**eflectometry
GPS	**G**lobal **P**ositioning **S**ystem
HMM	**H**idden **M**arkov **M**odel
I/Q	**I**n-phase/**Q**uadrature

ICESat	**I**ce, **C**loud and land **E**levation **Sat**ellite
ID3	**I**terative **D**ichotomiser **3**
IGN	**I**nstitut **G**éographique **N**ational
KLD	**K**ullback-**L**eibler **D**ivergence
KNN	**K**-**N**earest **N**eighbor
LHCP	**L**eft-**H**and **C**ircular **P**olarization
LSTM	**L**ong **S**hort-**T**erm **M**emory
MAE	**M**ean **A**bsolute **E**rror
ML	**M**achine **L**earning
MLE	**M**aximum **L**ikelihood **E**stimation
MLP	**M**ulti**l**ayer **P**erceptron
OLS	**O**rdinary **L**east **S**quares
PARIS	**P**assive **R**eflectometry **a**nd **I**nterferometry **S**ystem
PCA	**P**rincipal **C**omponent **A**nalysis
PDF	**P**robability **D**ensity **F**unction
PLL	**P**hase-**L**ocked **L**oop
PRN	**P**seudo-**R**andom **N**oise
RCS	**R**adar **C**ross-**S**ection
RF	**R**andom **F**orest
RHCP	**R**ight-**H**and **C**ircular **P**olarization
RMSE	**R**oot **M**ean **S**quare **E**rror
RNN	**R**ecurrent **N**eural **N**etwork
SAM	**S**pectral **A**ngle **M**apper
SAR	**S**ynthetic **A**perture **R**adar
SGR-ReSI	**S**pace **GNSS** **R**eceiver **Re**mote **S**ensing **I**nstrument
SHAP	**SH**apley **A**dditive ex**P**lanations
SLR	**S**tepwise **L**inear **R**egression
SMAP	**S**oil **M**oisture **A**ctive **P**assive
SNR	**S**ignal-to-**N**oise **R**atio
SPRT	**S**equential **P**robability **R**atio **T**est
SVI	**S**tructural **V**ariational **I**nference
SVM	**S**upport **V**ector **M**achine
SVR	**S**upport **V**ector **R**egressor
TDS-1	**T**ech**D**emo**S**at-**1**
TEC	**T**otal **E**lectron **C**ontent
UK-DMC	**UK** **D**isaster **M**onitoring **C**onstellation
XAI	e**X**plainable **A**rtificial **I**ntelligence
XGBoost	e**X**treme **G**radient **Boost**ing

List of Symbols

Signal and Correlation Parameters

A	Amplitude of the received signal
$C_A(t)$	CDMA spreading code
$c(t)$	Local code replica
I_{kT_c}	In-phase component
Q_{kT_c}	Quadrature component
$s(t)$	Received signal
s_D	Demultiplexed signal
$s_l(t)$	Local sine replica for demodulation
$c_l(t)$	Local cosine replica for demodulation
M	Number of reflections of reflected signal
S	Spectral characteristics
T_c	Coherent integration time
T_l	Longer integration duration
T_d	Direct signal intensity
T_r	Reflected signal intensity
$Y(\tau, f_D)$	Correlation output vs delay and Doppler

Geometric and Physical Quantities

$d_1(r)$	Distance from transmitter to surface point r
$d_2(r)$	Distance from surface point r to receiver
d_R	Distance from specular point to receiver
d_T	Distance from transmitter to specular point
$\Delta\hat{d}$	Geometric path difference estimate
$\Delta p(t)$	Path difference (reflected vs direct)
d_{atm}	Neutral-atmosphere delay
γ	Incidence angle
$h(t)$	Receiver height above surface
λ	Signal wavelength
r	Surface point (vector)
r_0	Footprint radius
$R(\gamma)$	Reflectivity as a function of γ
$\sigma_0(q)$	Bistatic scattering coefficient
θ_e	Satellite elevation angle
v	Platform speed
ω_D	Angular Doppler frequency
ζ	Circular length

Noise and Distribution Parameters

a	Scale parameter
b	Shape parameter
c	Shape parameter
μ	Mean value
σ	Standard deviation
$\Gamma(\cdot)$	Gamma function
$\Psi(N)$	Digamma function
$\Psi^{(1)}(N)$	Trigamma function
N	Shape parameter of Reflectivity
λ	Mean parameter of the reflectivity
λ_0, λ_1	Pre– and post–change mean reflectivities
$W(N, \lambda)$	Log–Gamma transform of reflectivity
$R(N, \lambda)$	Gamma–distributed reflectivity model
$\bar{w}_t$	Recursive mean of log–intensity w_t
w_t	Log–intensity sample ($\log r_t$)
$\eta(t)$	Additive white Gaussian noise
L_s	Number of scatterers
NF	Thermal noise floor

Phase and Coherence Quantities

α_i	Linear phase rate as function of f_D
ϕ_i	Individual phase measurement
$\phi_{i,0}$	Initial phase
K	Circular kurtosis
P_{coh}	Coherent power
ψ_σ	Rayleigh parameter
ρ	Phase ambiguity constant

Change Point Detection (CPD) Parameters

g_t^-	Negative CUSUM accumulator
g_t^+	Positive CUSUM accumulator
h	Detection threshold
C_s	CUSUM decision statistic
WT	Wald test statistic
$\text{ARL}(0)$	Average run length under null hypothesis
$\text{ARL}(1)$	Average run length under hypothesis 1
α	Test significance level
β	Probability of detection
P_{FA}	Probability of false alarm
P_D	Probability of detection

τ'	Detected change instant
τ_1, τ_2	Start/end of gradual transition interval
λ_t	Time–varying mean during transition
m_i	Sample mean of segment i
n_i	Number of samples in segment i
i_t	Innovation term in recursive mean filter
$\tilde{i}_t$	Normalized innovation
K	Window length for recursive averaging
Q	Process parameter in adaptive filter
P_t	Variance of recursive mean estimate $\bar{w}_t$
L	Number of independent looks

Constants and System Parameters

c	Speed of light
f_D	Doppler frequency offset
$f_{1,2}$	Carrier frequencies
f_{L1}	GPS L1 carrier frequency
$\hat{f}_1$	Doppler estimate from FLL (direct signal)
f_e	Sampling frequency
G_r	Receiver antenna gain
G_t	Transmitter antenna gain
P_t	Transmitted power
T_e	Sampling period ($1/f_e$)
Δf	Hardware delays / frequency bias term
R_c	Correlation coefficient
$E[\cdot]$	Expectation operator
$V[\cdot]$	Variance operator

General Introduction

Global Navigation Satellite System Reflectometry (GNSS-R) has emerged as a revolutionary paradigm in the science of Earth observation using the reflected navigation signals as opportunistic resources for bistatic radar remote sensing. Continuous global coverage, passive reception architecture, and the ability to exploit existing satellite infrastructure designed for positioning and navigation purposes are among the fundamental advantages of GNSS-R. These attributes make GNSS-R an economically viable complement to the established active and passive microwave remote sensing techniques, offering exclusive functionalities for Earth surface properties in diverse environmental contexts [1].

The development of GNSS-R technology over the last three decades demonstrates a progressive deepening of scientific understanding of the content of information inherent in the reflected navigation signal. Initial demonstrations focused primarily on proof-of-concept applications, thus establishing the potential for deriving geophysical parameters from signals reflected off ocean surfaces. In 1993, Martin-Neira proposed the concept of the Passive Reflectometry and Interferometry System (PARIS) to provide an underlying theoretical foundation for leveraging the use of GPS signals as bistatic radar sources [2]. Subsequent developments broadened the application field beyond the marine environment to land [3], cryosphere [4, 5], and inland water body surfaces [6], and shifted from ground-based experimentation to spaceborne missions such as the UK-DMC satellite [7], TechDemoSat-1 [8, 9], and the Cyclone Global Navigation Satellite System (CYGNSS) constellation [10–12].

Despite the advancements in GNSS-R capabilities, most of the research efforts and operational frameworks have focused primarily on amplitude-based observables, particularly in the analysis of signal power contributions through Delay-Doppler Maps (DDMs) and reflectivity measurements. The traditional methods used for GNSS-R applications show high efficiency in various applications, including ocean wind speed retrieval, sea surface

height estimation, and soil moisture monitoring, yet they fail to extract essential information from reflected signals. The phase component of GNSS reflections is an understudied aspect that illustrates the temporal evolution of signal interference patterns resulting from complex scattering mechanisms, providing opportunities for improved surface characterization and measurement accuracy.

Coherent reflections, occurring predominantly over smooth surfaces such as calm water bodies, sea ice, and certain terrestrial areas, maintain stable phase relationships enabling interferometric processing techniques and accurate altimetric measurements. The statistical properties of incoherent reflection phases from rough surfaces directly reveal information about scattering mechanisms and surface characteristics. The distinction between coherent and incoherent reflection patterns depends on electromagnetic wavelength and surface roughness and observation angle which makes phase behavior an effective method for automated surface classification.

This dissertation addresses the critical gap between the theoretical potential of phase-based GNSS-R and its practical implementation for Earth observation applications. We develop an integrated framework of statistical signal modeling, processing algorithms, and machine learning algorithms for enhanced surface classification performance. Airborne GNSS-R observations acquired in an experimental campaign over northern France are utilized as the research data set. It consists of high-resolution observations over different surface types ranging from sandy beaches and open ocean to inland water bodies and mixed land covers. This rich dataset enables both the development of physically-motivated signal processing methodologies and their rigorous validation against ground-truth observations.

This work tackles the challenge of noise in GNSS-R data as that of inherent multiplicative speckle noise, which fundamentally transforms the statistical properties of reflective signals compared to traditional additive noise scenarios. While previous work on GNSS-R change point detection assumed additive Gaussian noise models [13], this simplification proves inadequate when confronted with the complex scatter environments characteristic of Earth surface monitoring. We therefore derive specialized statistical models based on Log-Gamma distribution theory. This approach considers that multi-look averaged GNSS-R reflectivity follows probability distributions

based on the underlying principles of electromagnetic scattering theory. This physical basis enables detection algorithms with statistically derived thresholds. These algorithms can perform consistently without requiring empirical parameter tuning.

The transition from amplitude to phase analysis for classification applications requires the use of sophisticated signal segmentation techniques. The continuous nature of airborne GNSS-R data streams presents a significant challenge. As the platform traverses heterogeneous terrain, frequent surface type transitions occur. This requires robust change point detection algorithms that can identify transitions while minimizing false alarms. We address this using an online-offline detection scheme combining the efficiency of Cumulative Sum (CUSUM) methods with Maximum Likelihood Estimation techniques adapted for Log Gamma-distributed reflectivity data.

The general aims and contributions of this work can be summarized as follows:

- Development of an automated GNSS-R signal segmentation system using Gamma distribution modeling and homomorphic transform, allowing for effective detection of surface changes in airborne measurements.

- Development of phase-aware statistical models for GNSS-R processing, establishing the relationship between phase coherence and surface scattering properties for better surface discrimination.

- Introduction of excess Doppler spectrum analysis with quantitative measurements that provide complementary criteria along with conventional reflectivity-based techniques.

- Generation of synthetic data, based on the physical mechanisms, due to the limited availability of labeled real GNSS-R datasets suitable for supervised learning training.

- Implementation and validation of machine learning algorithms for surface classification, with effective application of models trained on synthetic data to real airborne observations.

The developed techniques are presented through five chapters. The first two chapters are mainly bibliographical, while the last three chapters are dedicated to the presentation of the research work. Chapter 1 covers the evolution of GNSS-R signal processing from power-based to phase-aware approaches. This chapter provides an overview of the bistatic radar theory, signal processing architectures, and phase-based GNSS-R, indicating research gaps. The comprehensive outline of the experimental framework and the airborne data collection campaign is presented, establishing a solid basis for validation.

The second chapter first establishes the statistical modeling framework for speckle noise, comparing empirical and physically-based distribution models with particular emphasis on the Gamma distribution, which is suitable distribution for multi-look GNSS-R applications. Different change point detection algorithms are reviewed, including both unsupervised methods (including GLR, Bayesian, and CUSUM approaches) and supervised classification techniques. The chapter concludes by examining existing methodologies for change detection under Gamma noise assumptions, identifying critical limitations such as parameter sensitivity and computational complexity, motivating the contributions of Chapter 3.

Chapter 3 proposes a segmentation model that divides the GNSS signals into stationary parts associated with different areas of reflection. The power of the reflectivity signals changes significantly with the displacement of the satellite footprints from one area to another. The chapter begins with the derivation of the Gamma distribution model for multi-look reflectivity measurements, establishing the theoretical justification for a homomorphic transformation that converts multiplicative noise into constant-variance additive noise. The integrated online-offline change detection system is presented in detail, including CUSUM-based detection with statistically determined thresholds. Then, Log Gamma Maximum Likelihood Estimation is proposed as an algorithm that localizes these changes in a working window. A Wald test-based segment merging method is implemented to eliminate over-segmentation and to decrease the number of false change detections. This chapter concludes by assessing the performance on both synthetic and real GNSS-R airborne data, demonstrating the feasibility of the proposed segmentation methodology.

Chapter 4 outlines both theoretical and practical foundations of phase-aware GNSS-R processing. In this chapter, details are given on the overall signal processing chain, deriving implementations in discrete time for direct and reflected signal analysis. The von Mises statistical phase model demonstrates relationships between concentration parameters and signal-to-noise ratios and coherence properties. Excess Doppler spectrum analysis is introduced, and it provides three quantitative measures that determine peak amplitude, spectral complexity, and energy dispersion features. A comprehensive synthetic data generation methodology is developed, based on the multi-path reflection model, which extracts surface-related parameters from real airborne measurements. K-means clustering is applied to the three excess Doppler metrics computed for each segment, demonstrating improved surface classification in an unsupervised mode compared to reflectivity alone.

Chapter 5 utilizes machine learning algorithms and phase models, as well as features defined in the previous chapter. This chapter establishes a dual-dataset approach, describing the generation of synthetic data from physical models with feature engineering approaches for classical classifiers and deep learning algorithms. A total of six machine learning models are presented with a systematic study, from classical ones like Logistic Regression, Random Forest, XGBoost, and Multilayer Perceptron, to deep learning ones like 1D-CNN and Bi-LSTM. Detailed experimental results compute classification performance on both synthetic and real test datasets, demonstrating successful application of models trained on physics-based synthetic data to real airborne surface classification.

Finally, a general conclusion is presented, which synthesizes the contributions described throughout the dissertation and proposes future perspectives to further advance the topic.

Publications

This thesis work is valorized in 1 journal, 1 international conference with proceedings, 2 international conferences without proceedings, 3 national communications, in addition to 1 journal awaiting submission.

Publications in a journal

1. **S. E. H. Chehade.**, H. Issa, G. Stienne and S. Reboul. Change Point Detection in Radar Reflectivity Measurements Contaminated by Speckle Noise, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 11208-11218, 2024, doi: 10.1109/JSTARS.2024.3410039.

International conferences with proceedings

1. **S. E. H. Chehade**, H. Issa, G. Stienne and S. Reboul. Segmentation and Classification of Airborne GNSS-R Reflectivity Signals with Speckle Noise Mitigation, International Conference on Localization and GNSS (ICL-GNSS) 2024, Antwerp, Belgium, 2024, pp. 1-7, https://doi.org/10.1109/ICL-GNSS60721.2024.10578377.

International conferences without proceedings

1. **S. E. H. Chehade.**, G. Stienne and S. Reboul. Excess Doppler Spectrum Analysis for Surface Characterization in GNSS-R Reflections, IEEE GNSS+R 2025 - Specialist Meeting on Reflectometry using GNSS and other Signals of Opportunity.

2. **S. E. H. Chehade.**, H. Issa, G. Stienne and S. Reboul. An Improved Change Detection Approach for Airborne GNSS Signals Segmentation in the Presence of Speckle Noise, IEEE GNSS+R 2023 - Specialist Meeting on Reflectometry using GNSS and other Signals of Opportunity.

National communications

1. Journée des Doctorants du Pôle MTE (Doctoral Students Day of the MTE Pole), held in ULCO Calais on June 20,2023. I had the opportunity to present a poster about my thesis work.

2. 9ème Journée Régionale des Doctorants en Automatique (JRDA) on June 21, 2023, held in Roubaix, Lille, France. I had the opportunity to present a 60-second talk about my thesis, in addition to presenting a poster.

3. Journée Régionale des Doctorants en Sciences de la Mer et Automatique (JRDMA) on Jul y 8, 2024, held in Centre universitaire du Musée à Boulogne-sur-Mer. I presented a poster about my research.

4. Journée Doctorale 2024 SFR CAMPUS DE LA MER, held in Centre Universitaire du Musée de Boulogne-sur-Mer on November, 2024. Presented a talk on my research related to detecting inland water bodies with satellite reflectometry.

Chapter 1

The Evolution of GNSS-R Signal Processing: Towards Phase-Aware Remote Sensing

1.1 Introduction

The development of spaceborne Global Navigation Satellite System Reflectometry (GNSS-R) is a remarkable transformation from initial proof-of-concept demonstrations to sophisticated spaceborne remote sensing applications. Although early spaceborne GNSS-R techniques made use of only amplitude based signal processing, the domain has undergone a fundamental shift with the understanding that phase information conveys dense, complementary observables far beyond conventional power measurements.

This chapter describes the overall evolution of GNSS-R signal processing methods and records the evolution from traditional power-based methodologies to sophisticated phase-aware methods. This is more than a refinement of technologies; it is a paradigmatic change in concepts of using reflected GNSS signals for Earth observation applications from satellite platforms.

The chapter begins with a discussion of early power-based processing schemes that were the foundation of the underlying GNSS-R principles. While those early schemes were effective in demonstrating basic capabilities, they showed that they had inherent limitations that prompted more sophisticated processing structures. Developments of correlation and demodulation approaches, integration schemes, and multi-constellation processing innovations have increasingly expanded GNSS-R capabilities.

Key to this account is the development of phase-aware processing as a revolutionary method of GNSS-R signal processing. Systematic exploration of the phase information has uncovered previously considerably underutilized areas of surface-sensitive measurement and has made possible applications from precise altimetry to improved surface characterization. Bistatic radar operation theory, coherence requirements, and probabilistic modeling infrastructures are formulated and developed to offer the analytical foundation of phase-based measurement.

1.2 GNSS-R Signal Processing Evolution

The development of spaceborne GNSS-R has been characterized by successive improvements in signal processing methods, as each subsequent generation as shown in figure 1.1 has targeted the limitations of its predecessor and expanded the potential effectiveness of Earth observation methods. This evolutionary timeline illustrates the key milestones in this evolution from traditional power-based approaches to sophisticated phase-aware processing techniques. This evolutionary process represents not just a technological refinement but, in itself, reveals a correspondingly deeper understanding of the rich informational content inherent in the reflected GNSS signals, from simple power-based measurements to sophisticated multi-dimensional analytical methods.

1.2.1 Traditional Power-Based Approaches and Their Limitations

The early implementation of GNSS-R was defined by conventional processing methods that focused on power property extraction, despite Martin Neira's original PARIS concept envisioning interferometric phase-based approaches [2]. These conventional methods, known as conventional GNSS-R (cGNSS-R), laid the groundwork for basic signal processing by utilizing correlation-based methods that involved the cross-correlation of incoming reflected signals with locally generated clean replicas of transmitted GNSS codes.

The matched filtering operation forms the mathematical basis of conventional processing. It can be written as [14]:

$$Y(\tau, f_D) = \frac{1}{T_c} \int_0^{T_c} s(t) \cdot c^*(t-\tau) \cdot e^{-j2\pi f_D t} dt \tag{1.1}$$

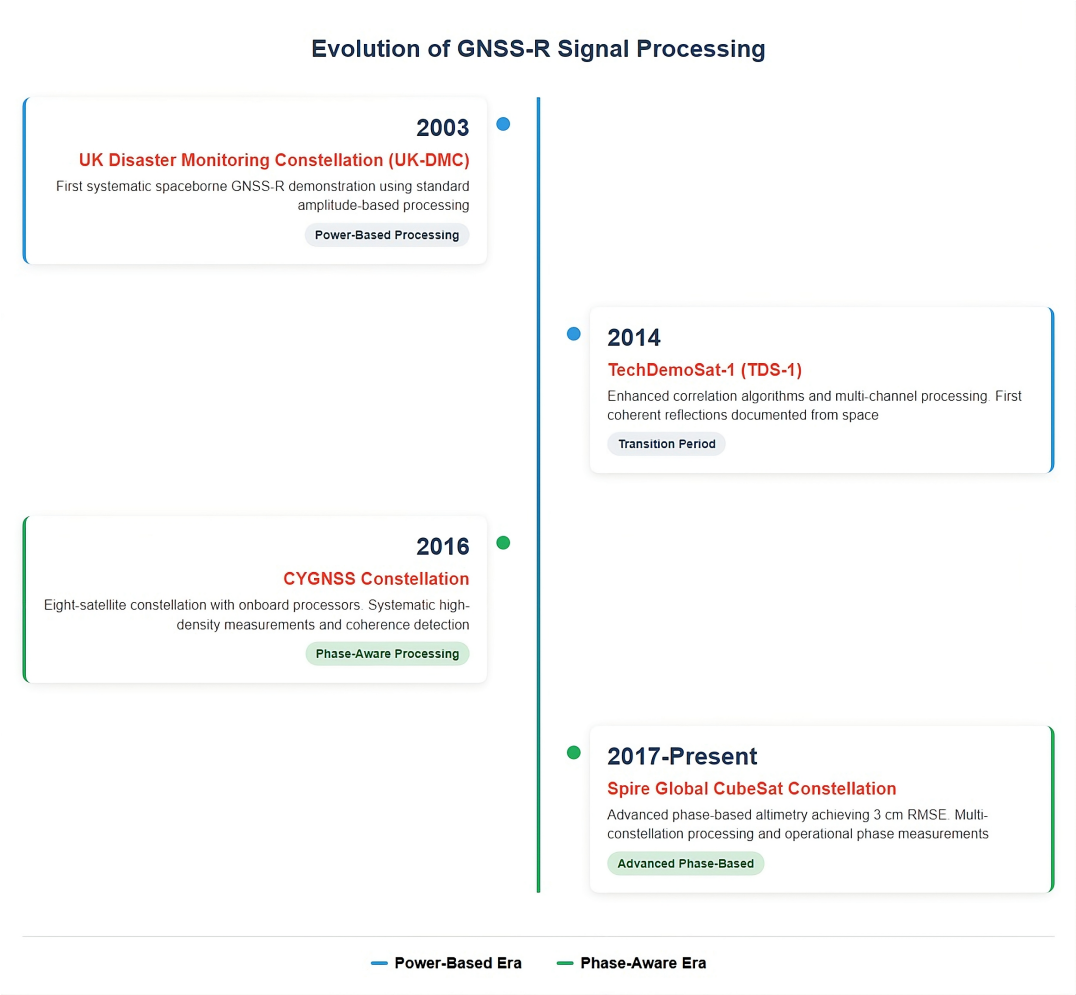


FIGURE 1.1: Timeline of spaceborne GNSS-R signal processing evolution.

where $s(t)$ is the received reflected signal, $c(t)$ is the local code replica, T_c is the coherent integration time, τ is the code delay, and f_D is the Doppler frequency offset. This correlation approach exploits the auto-correlation properties of pseudo-random noise codes to maximize the signal-to-noise ratio and enable range discrimination.

The UK Disaster Monitoring Constellation satellite, launched in 2003, provided the first systematic demonstration of spaceborne GNSS-R using standard amplitude-based processing techniques [15–17]. The mission employed straightforward yet effective correlation processing, demonstrating the potential of space-based GNSS-R and identifying key differences in reflection characteristics over ocean and ice surfaces.

The extension to two-dimensional delay-Doppler mapping represented

a substantial advancement in traditional processing. Delay-Doppler Maps (DDMs) provide an extensive representation of scattered signal power as functions of time delay and Doppler frequency shift. The theoretical foundation for DDM generation is established by the bistatic radar equation formulated by Zavorotny and Voronovich [1], which relates expected scattered power to surface and geometric parameters:

$$\langle |Y(\tau, f_D)|^2 \rangle = \frac{P_t G_t G_r \lambda^2}{(4\pi)^3} \int\int_A \frac{\sigma^0(q)}{d_1^2(r) d_2^2(r)} |\Lambda(\tau - \tau(r))|^2 |S(f_D - f_D(r))|^2 dr \tag{1.2}$$

where P_t is the transmitted power, G_t and G_r are transmitter and receiver antenna gains, λ is the signal wavelength, $\sigma^0(q)$ is the bistatic scattering coefficient dependent on the scattering geometry parameter q (which characterizes the local incidence and scattering angles), $d_1(r)$ and $d_2(r)$ are the distances from the transmitter to surface point r and from surface point r to the receiver respectively. Λ and S represent the signal's autocorrelation and spectral characteristics respectively. The double integral is performed over the scattering area A (the glistening zone) representing the differential surface area element, accounting for contributions from all surface points that scatter the transmitted signal toward the receiver.

The power-based methods, on the other hand, had a few fundamental limitations of their own which grew in significance with field development [18]. While successful in revealing basic GNSS-R potential, applications for retrieving ocean wind speed relied heavily on empirical relationships between observable characteristics and the geophysical parameters. These relationships showed significant variability across different environmental conditions and limited transferability between surface types and observation geometries.

Land applications were especially difficult to conventional methods due to the increased diversity of scattering mechanisms and surfaces. Early attempts had tried utilizing peak signal-to-noise ratio by proxy of moisture content of surfaces, but accuracy only had a narrow range, and generalizability across a variety of surfaces was in doubt. The main limitation was that amplitude measurements were not sufficient to differentiate surfaces of similar reflectivity but having different scattering mechanisms.

1.2.2 Evolution of Correlation and Demodulation Techniques

Identifying some limitations within the underlying correlation processing of reflected signals triggered the development of more sophisticated correlation and demodulation techniques intended for deriving more information from backscattering GNSS signals. The UK TechDemoSat-1 (TDS-1) mission, launched in 2014, marked an important step forward in space-based GNSS-R processing by using better correlation algorithms and multi-channel processing capabilities [16].

TDS-1 used advanced signal processing algorithms that improved the fidelity and quality of correlation outputs through enhanced calibration methods and superior noise floor estimation. The Space GNSS Receiver Remote Sensing Instrument (SGR-ReSI) of TDS-1 had multiple correlation channels operating in parallel, thus allowing for the simultaneous tracking of multiple GNSS satellites and the formation of different types of DDMs simultaneously.

TDS-1 captured the first systematic record of coherent GNSS reflections from outer space. Observations made within regions of sea ice cover revealed that the DDMs recovered were of characteristics significantly different than those measured over rough ocean surfaces [19]. In contrast to the horseshoe pattern characteristic of diffuse scattering, the coherent reflections provided responses that were peak-shaped and narrow, with a resemblance to the ambiguity function of the transmitted signal.

The Cyclone Global Navigation Satellite System (CYGNSS), launched in 2016, is a milestone in systematic and high-density spaceborne GNSS-R observations [20]. It uses eight microsatellites that have been specially designed to estimate wind speeds over the oceans, improving the capabilities to predict and understand hurricanes. The eight-satellite constellation features onboard processors that successfully address some of the limitations of previous missions. Each satellite utilizes a multi-channel architecture that enables simultaneous processing of signals from four GPS satellites combined with advanced satellite selection algorithms to maximize both quality and data coverage.

The CYGNSS processing combines several important advances in the correlation and demodulation techniques. Signal-to-noise ratio (SNR) computation

for GNSS-R observations is given by [16]:

$$SNR = \frac{P_{inc,peak} - NF}{NF} \tag{1.3}$$

Here, $P_{inc,peak}$ refers to the peak of the incoherently averaged power waveforms and NF refers to the thermal noise floor, which is estimated by taking the average of the first 100 delay bins of the power waveforms.

1.2.3 Coherent vs Incoherent Integration Strategies

An important aspect in the evolution of GNSS-R signal processing has been the development of increasingly sophisticated integration methods that balance the requirement for noise removal with information preservation. The inherently weak nature of the reflected GNSS signal, typically 20–30 dB below the level of direct signals, requires extensive averaging for useful surface data; nevertheless, different integration schemes have profound implications regarding what information can be extracted.

In conventional GNSS-R systems, incoherent integration is used, which means that the results of the squared correlation are averaged over long periods [20]:

$$P_{avg}(\tau, f_D) = \frac{1}{N} \sum_{i=1}^{N} |Y_i(\tau, f_D)|^2 \tag{1.4}$$

where N represents the number of individual correlation products averaged, typically corresponding to integration times of 1 second or longer. This approach effectively reduces thermal noise and provides stable power estimates, but fundamentally discards phase information contained in individual correlation products.

Conversely, coherent integration preserves both amplitude and phase information by vector averaging of complex correlation outcomes before calculation of signal power [21]. For coherent integration, the processing gain G_p of a fully coherent signal with a longer integration duration T_l compared to the usual integration duration T_c is given by the following [22]:

$$G_p = 10 \log_{10} \left(\frac{T_l}{T_c} \right) \tag{1.5}$$

This connection demonstrates the possibility of coherent signals achieving significant processing gains through extended integration, which depends on maintaining phase coherence during the integration period.

Semi-coherent integration is a hybrid technique that attempts to combine the benefits inherent in noise reduction of extended integration while, in part, maintaining phase information. Semi-coherent integration uses short coherent integration periods followed by incoherent averaging of the following correlation values, thus providing improved performance over purely incoherent integration while maintaining some robustness to phase variations.

The practical implementation of these integration strategies becomes particularly complex when considering multiple independent observations, known as multi-look processing. Multi-look processing is a key concept in GNSS-R signal processing that has a direct effect on noise reduction versus information preservation. For GNSS-R systems, multi-look processing generally consists of taking averages over several time instances or spatial points of the outputs from correlation, with significant impacts on statistical quantities as well as the capacity to retrieve surface information.

The choice between coherent and incoherent approaches becomes critical in multi-look processing. For coherent scattering, this case is complicated due to the nature of the underlying signal being deterministic as opposed to random. Averaging multi-looks in coherent signals can have phase information preservation when averaging is performed as vector averaging in the complex outputs from correlation:

$$\langle Y \rangle = \frac{1}{N} \sum_{i=1}^{N} Y_i \tag{1.6}$$

where Y_i represents complex correlation output per pixel. If multi-look processing involves averaging over magnitudes, phase is lost irretrievably even when the underlying scattering is coherent.

The statistical nature of multi-look processed signals is heavily reliant on the main mechanism of scattering. For incoherent scattering, individual correlation outputs are provided by exponential distributions, while the multi-look mean will tend towards Gaussian statistics. For L independent looks, the coefficient variation in the multi-look mean will decrease as [10, 23–26]:

$$CV = \frac{\sigma}{\mu} = \frac{1}{\sqrt{L}} \tag{1.7}$$

This is the basic trade-off between reducing noise and having independent samples available to average.

The temporal and spatial characteristics of the observations determine the effectiveness of multi-look processing strategies. The characteristic time for the surface scatter affects the efficiency in multi-look processing as well as the independence between separate looks. The scatter signal's correlation time can be estimated as:

$$t_{\text{cor}} = \frac{d_r \lambda}{r_0 v} \tag{1.8}$$

where d_r is the range from receiver to effective surface footprint, r_0 is the radius of the footprint and v is the platform speed. This equation provides the minimum time separation to have independent looks.

To quantify the degree of coherence in the case of coherent scatterings, we can use [27]:

$$DOC = \frac{|E[Y]|^2}{E\left[|Y|^2\right]} \tag{1.9}$$

where a DOC close to 1 indicates highly coherent scattering, while values approaching 0 suggest incoherent scattering dominated by speckle noise [27].

1.2.4 Multi-Constellation Processing Developments

The modernization of GNSS constellations and the availability of multi-frequency signals have opened new opportunities for advanced GNSS-R processing. Extending from GPS to Galileo, BeiDou, and GLONASS signals has significantly increased reflection possibilities and improved global coverage, especially at high latitudes where GPS coverage is limited [21].

Galileo and BeiDou reflected signals have been successfully tested for GNSS-R applications, with studies showing that their signal-to-noise ratios are generally 1-2 dB lower than those of GPS signals [21]. The reason for this lies mainly in the differences in transmitted power and receiver bandwidth characteristics, but overall performance is still adequate for the majority of

GNSS-R applications.

GLONASS signals, with their frequency division multiple access (FDMA) structure, have specific advantages for certain GNSS-R applications [28]. The frequency split between the different GLONASS satellites enables straightforward signal separation and processing, allowing individual satellite selection based on their distinct frequency channels.

Multi-frequency processing offers improved capabilities with the ionospheric correction and better surface penetration qualities. Dual-frequency observations allow the separation of surface and ionospheric effects [29], with the geometric difference in reflected and direct signal paths given by:

$$\Delta\hat{d} = \Delta\phi_{1,2} - d_{atm} + \frac{k_p\,\Delta TEC}{f_{1,2}^2} - \Delta\rho_{1,2} \tag{1.10}$$

In this equation, $\phi_{1,2}$ are phase measurements, respectively, on each frequency (labeled by subscripts 1 and 2). d_{atm} is the delay induced by the neutral atmosphere for the reflected signal path, the Total Electron Content TEC is the number of electrons integrated along a signal path, k_p is a proportionality constant, f is a carrier frequency of a signal, and ρ is the phase ambiguity constant. Values preceded by Δ denote the difference between the reflected and direct signal paths.

1.2.5 GNSS-R Fundamentals: Geometry, Signal Paths, and Coherence Conditions

Understanding the fundamental concepts of bistatic radar operation is essential for recognizing the unique characteristics and challenges of GNSS-R systems. Unlike conventional monostatic radar systems that involve co-located transmitters and receivers, GNSS-R operates in bistatic mode, utilizing navigation satellites as opportunistic sources of illumination and dedicated receivers that receive the signals reflected from the Earth surface.

In GNSS-R, coherent scattering occurs when smooth surfaces maintain the phase relationship between the incident and reflected signals, enabling phase-based measurements and geometric retrievals. On the other hand, incoherent scattering from rough surfaces disperses the phase over the scattering area, resulting in diffuse reflections.

The bistatic radar equation serves as the theoretical foundation for understanding signal power relationships in GNSS-R systems, and work conducted by Voronovich and Zavorotny has recently enhanced it to incorporate the transition from coherent to incoherent scattering regimes [1].

The bistatic coherence conditions depend on the interaction between electromagnetic wavelength, surface roughness, and observation geometry. The Rayleigh parameter is a dimensionless measure of the effect due to surface roughness [22]:

$$\psi_\sigma = \frac{2\pi\sqrt{\langle\Delta h^2\rangle}\cos\gamma}{\lambda} \tag{1.11}$$

where $\sqrt{\langle\Delta h^2\rangle}$ is the Root Mean Square (RMS) surface height, λ is the wavelength, and γ is the incidence angle. Surfaces appear smooth when this parameter is much less than unity and rough when it approaches or exceeds unity. For bistatic scattering from smooth surfaces, the coherent power can be expressed as:

$$P_{coh} = \frac{P_t G_t \lambda^2 G_r}{(4\pi)^2 (d_T + d_R)^2} R(\gamma)\varepsilon^2 \exp(-(2k\sigma\cos(\gamma))^2) \tag{1.12}$$

where, d_T and d_R are the distances from the transmitter to the specular point and from the specular point to the receiver, respectively, R is the reflectivity and ε the transmissivity including attenuation due to vegetation. The exponential term describes the signal attenuation due to surface roughness.

1.3 Phase Information in GNSS-R: The Underexploited Domain

1.3.1 Evolution of Phase-Based Approaches

The realization that phase information potential is available in GNSS-R has evolved since the initial concept of Passive Reflectometry and Interferometry System (PARIS) was proposed by Martin-Neira in 1993 [2]. While the original concept of PARIS was based on interferometric processing for altimetry with the oceans, the broad phase-based measurement possibilities that exist were scarcely exploited in early implementations.

The pioneering Zavorotny-Voronovich work laid the theory for GNSS signal backscattering from ocean surfaces, although mainly dealing with

amplitude-based observables [1]. This pioneering contribution introduced the bistatic radar equation that formed the prototype for GNSS-R applications, though the phase part of the scatter-received signals was given marginal consideration in the early formulation.

An early systematic study on phase coherence in GNSS-R originated from the phase measurements by the UK-DMC satellite, where researchers in [30] observed potential phase coherence in GPS reflections from the sea ice. This early spaceborne evidence demonstrated that backscattered GNSS waves can conserve phase links under specific surface conditions, disproving the assumption that all reflected signals are incoherent.

Experimental Breakthroughs in Phase Altimetry

A significant breakthrough was reached in Fabra et al.'s research [30] when they introduced systematic dual-polarization phase altimetry over sea ice from ground GNSS-R observations. From their seven-month campaign over the Greenland region, they discovered that phase samples that are coherent can be retained over sea ice surfaces while enabling ellipsoidal height estimations that were consistent with Arctic tide models together with ICESat laser altimetry products. This pioneering study entrenched the fundamental rule that phase coherence is optimized for smooth reflective surfaces.

Based on these preliminary works, Semmling et al. then applied phase-based measurement to ocean tides with interferometric GNSS-R reflections to show the detection potential for tides in the Arctic Ocean [31]. What their study demonstrated was that with adequate signal-to-noise ratios and suitable surface states, coherent processing was capable of retrieving geometric information from reflected waves.

Spaceborne Phase Implementations

The transition from airborne phase-based to spaceborne phase-based GNSS-R was another breakthrough. Li et al. [32] initially demonstrated spaceborne phase altimetry over sea ice utilizing TechDemoSat-1 GNSS-R echoes to demonstrate that coherent reflections can be utilized from Low Earth Orbit satellites. Decimeter-level height retrieval accuracy from their

experiments validated the feasibility of phase measurement from space.

Later, Nguyen et al. [33] demonstrated performance from the Spire Global CubeSat constellation to within 3 cm RMSE agreement with sea surface height models after removal of biases. This provided the best accuracy altimetric estimates to date from spaceborne GNSS-R measurements, the best reported performance for spaceborne GNSS-R altimetry.

Multi-Constellation and Advanced Processing

The latest advancements have expanded from GPS to multi-constellation phase processing. The implementation of spaceborne BeiDou-3 signals for use in GNSS-R has been demonstrated in [34], exhibiting similar performance to GPS while offering additional measurements.

The advancement in coherence detection algorithms has paralleled these expansions in constellations. Al-Khaldi et al. in [35] proposed power spread ratio techniques for flagging coherence in CYGNSS measurements, whereas Roesler et al. in [36] provided coherence tests from circular statistics of excess-phase measurements. These algorithm advances have made uniform detection from large datasets possible, enabling statistical study of the patterns of coherence in various kinds of surfaces.

1.3.2 Theoretical Foundations

While the previous section laid the basic signal processing structure and bistatic radar fundamentals of GNSS-R, the true potential of this method lies far beyond that of standard amplitude-based observables. The phase content embedded in GNSS reflected signals is a vastly underutilized area that provides unprecedented potential for improved surface description and high-accuracy measurement.

1.3.2.1 The Phase Paradigm Shift

The conventional GNSS-R processing has mainly employed incoherent integration methodologies that intentionally suppress phase information due to the instability in amplitudes. However, latest breakthroughs in signal processing as well as receiver technologies have unveiled that the phase domain provides abundant information sensitive to ground properties,

mechanisms of scatterings, as well as geometrical relationships that are unattainable from single amplitude measurements.

The main limitation in the amplitude-only methodology is revealed when looking into the information content provided in the complex correlation output. Whereas, standard processing compresses the complex correlation to:

$$P(\tau, f_D) = |Y(\tau, f_D)|^2 \tag{1.13}$$

the complete complex observable has both magnitude and phase components:

$$Y(\tau, f_D) = |Y(\tau, f_D)| \cdot e^{j\phi(\tau, f_D)} \tag{1.14}$$

where $\phi(\tau, f_D)$ is the interferometric phase that carries geometrical as well as surface information lost in standard processing.

1.3.2.2 Phase Information Extraction Strategies

The critical phase-based GNSS-R property is not in the scatter mechanisms themselves (Section 1.2), but in how phase information can be extracted and utilized depending on the dominant scattering regime.

For Coherent-Dominated Signals: Phase relationships are maintained, which allows interferometric processing methods as well as precise geometrical measurements.

For Incoherent-Dominated Signals: Since single-phase measurement is no longer available, statistical phase properties can still provide information in the surface characterization.

For Mixed Scattering: Higher order processing methodologies can separate coherent from incoherent parts, extracting different types of information from each.

1.3.3 Advanced Coherence Detection Algorithms

Recent developments in coherence detection have introduced multiple complementary approaches for identifying coherent reflections from GNSS-R datasets.

1.3.3.1 Circular Statistics for Phase Analysis

The phase measurement from GNSS-R is inherently circular data that must be handled statistically in specialized ways. The statistical treatment of circular data is impossible in the same manner as linear statistical data because of the periodic nature of the phase angles. The classical phase analysis in GNSS-R has utilized simple coherence thresholds or power-based estimators. The phase nature, being, however, circular, means that advanced statistical techniques from circular statistics can provide enhanced discrimination among surface types.

Studies have introduced complementary methods to identify coherent reflections from GNSS-R observations using these circular statistics approaches. Coherence tests have been developed from circular statistics of phase excess measurements through the use of circular length and kurtosis parameters in phase-coherence measurement [36, 37]. The circular length is provided as: **[roe]**

$$\zeta = \frac{1}{M}\left|\sum_{i=1}^{M}\cos(\phi_i) + j\sum_{i=1}^{M}\sin(\phi_i)\right| \tag{1.15}$$

where M is the number of phase samples and ϕ_i represents the individual phase measurements. The circular kurtosis is another indicator of the degree of the "peakedness" in an angular distribution:

$$K = \frac{1}{M}\sum_{i=1}^{M}\cos\left(2\left(\phi_i - \bar{\phi}\right)\right) \tag{1.16}$$

where $\bar{\phi}$ is the mean of the dataset. If phases are uniformly distributed, then $\zeta = 0$ and $K \leq 0$, while completely coherent signals yield $\zeta = 1$ and $K = 1$.

These approaches have been applied with success to the Spire Global CubeSat dataset, which showed that about 1% and 44% of the GPS reflections over the ocean and sea ice, respectively, are viable for precise altimetry [29].

1.3.3.2 Entropy-Based Detection Methods

Von Neumann entropy provides a coherence detection framework based on the eigenvalue distribution of complex correlation matrices [22, 38]. This method involves the eigendecomposition of the correlation matrix, formed from complex zero-Doppler delay waveforms acquired over multiple

coherent integration periods. The resulting entropy value quantifies the degree of coherence, where entropy equal to zero corresponds to a single dominant coherent component and entropy approaching unity indicates fully incoherent scattering.

Unlike phase-based metrics that examine temporal variations, the entropy metric analyzes the correlation structure rather than direct temporal phase variations, even though it operates on complex-valued samples. The entropy computes how the received signal is spread among dimensions of the signal basis. Clearly, the thermal noise is perfectly diffused along all dimensions, whereas the signal with partial coherence spans all values of entropy from zero to one, with perfectly deterministic waveform having entropy equal to 0.

1.3.3.3 Machine Learning Approaches

Advanced coherence detection algorithms have transitioned from simple statistical thresholds to explorations of machine learning methodologies. Studies constructed their algorithms from estimates of power spread ratios in delay-Doppler maps to identify coherence [22], whereas newer algorithms utilize support vector machines and random forest classifiers that are trained on several indicators of coherence as features [35].

These machine learning methodologies mark a significant progression in automatic detection of coherence from broad GNSS-R datasets, attaining detection performance levels in excess of 95% when implemented with multi-constellation datasets [35].

1.3.4 Current State and Research Gaps

Phase-based GNSS-R has demonstrated extraordinary ability in three large application domains and is currently the most advanced implementation of coherent signal processing for satellite reflectometry. Its most advanced application has been that of altimetry, with spaceborne platforms achieving remarkable accuracy levels [29].

Atmospheric sounding is one of these new application areas for which coherent measurements are sensitive to tropospheric parameters. Phase recording for atmospheric parameter retrieval allows for better accuracy for atmospheric parameter derivation, especially in coastal areas for obtaining precise

surface height with proper tropospheric correction. The coherence time of GPS signals reflected off surfaces and statistical properties of these signals still greatly affect remote sensing applications of sea surface topography and wind determinations.

1.3.5 Critical Research Gaps Requiring Investigation

Limitation in statistical modeling is an important gap in existing GNSS-R approaches. CYGNSS land observation analysis demonstrates that 89.6% of observations are controlled by the coherent part, but present classification techniques assume empirically defined thresholds instead of stable statistical models [11]. The problem lies in developing techniques that utilize scattering power-spreading shape and magnitude features to effectively separate coherent and incoherent DDMs, as different classification estimators show varying performance quality across different surfaces [11].

Theoretical knowledge deficiencies also limit the development of phase-based procedures beyond the present uses. The existing bistatic radar equation used for the simulation of waveforms reflected by the surfaces was obtained in the past with the assumption of a strong diffuse scattering characteristic for very rough surfaces by using the Kirchhoff approximation [29]. This method results in an erroneous solution for weak diffuse scattering or in the presence of coherent reflection, demonstrating the requirement for generalized bistatic radar equations for the whole range of surfaces.

The integration of phase information with other GNSS-R observations has not yet been explored extensively, with potential for greatly increased ability for the characterization of the surface. Coherence detection techniques with circular statistics of the phase have promise but are not yet developed rigorously for the separation of coherent and incoherent scattering components [27]. Extraction with coherent component analysis has potential for improved scatterometry measurements with greater spatial resolution, as the coherent components will maintain the reflection at the specular point and hence the first Fresnel zone [27].

These comprehensive gaps indicate the urgent need for advanced statistical frameworks that can bridge the gap between theoretical understanding and

practical implementation, and establish robust phase-based GNSS-R methods for diverse environmental conditions and terrains.

1.4 Experimental Framework and Data Foundation

The signal processing and theoretical advances presented in this thesis are validated and proven through an extensive dataset of airborne GNSS-R data collected during experimental campaigns in Northern France [13, 39]. This section provides the experimental basis upon which research findings in later chapters rest.

1.4.1 Airborne GNSS-R Experimental Campaign

These experimental observations were carried out in airborne GNSS-R missions [13, 39] over the coastal area of Northern France. The first flight took place on October 19, 2020, starting at 14h45 UTC, and the second flight on June 16, 2021, starting at 13:56 UTC. Both flights departed from Calais–Dunkerque Airport and followed an ellipse-shaped trajectory of around 71 km, including diverse terrestrial and water surfaces between the cities of Calais, Escalles, and Ardres as shown in Figure 1.2. The 45-minute flights were dedicated, in particular, to recording GNSS reflections over various surface types, to offer a comprehensive dataset to support advanced signal processing and surface classification studies.

The experimental location was intentionally selected for its extensive heterogeneity of reflecting surfaces, i.e., the region between Guînes and Ardres, which contains over 50 different water body surfaces of lakes, ponds, rivers, wetlands, and swamps of different characteristics and sizes. This heterogeneous environment also contains other terrain surface types, such as plain land, groves, woodland, sandy beaches, and coasts. This offers an excellent experimental field to investigate potential GNSS-R surface differentiation under diverse scattering conditions.

1.4.2 Platform and Instrumentation

The airborne platform utilized in this study was selected because of its stability and capacity to operate in low-altitude flight regimes conducive to high-resolution observations from GNSS-R. The lightweight aircraft shown in Figure 1.3 presents unique advantages for GNSS-R missions including high

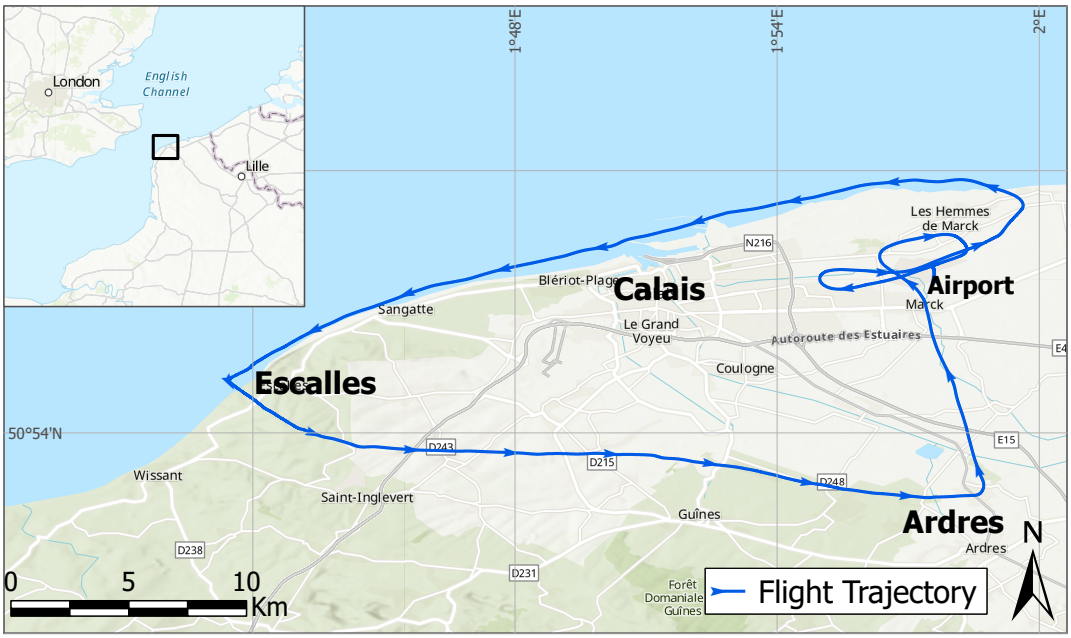


FIGURE 1.2: The flight trajectory [13].

maneuverability to enable precise trajectory control over target areas. Flying at an average altitude of 315 meters above sea level at speeds of about 90 km/h, the platform provided the necessary geometric configurations for both coherent and incoherent reception of GNSS reflections with very high spatial resolution. Considering the average speed of the gyrocopter, the distance between two consecutive specular points is approximately 0.5 m. The spatial resolution of the measurements is primarily determined by the size of the first Fresnel zone. For the high-elevation satellites (greater than 50) analyzed in this work, the major axis size of the footprints was fixed to a maximum of 23 m.

This experimental configuration was specifically designed for this GNSS-R instrumentation system, using a dual-channel receiver structure to allow for simultaneous observation of both direct and reflected signals. The system comprised one right-hand circular polarization (RHCP) antenna mounted at the nose of the gyrocopter in order to receive direct signals. The second antenna, optimized for left-hand circular polarization (LHCP), was mounted on the bottom of the gyrocopter, for receiving reflected signals after specular scattering by the surface of the Earth below.

An essential new feature in the design of the receiver was to apply a delay line system based on fiber optical elements, which made it possible to synchronously process both direct and reflected signals using a single

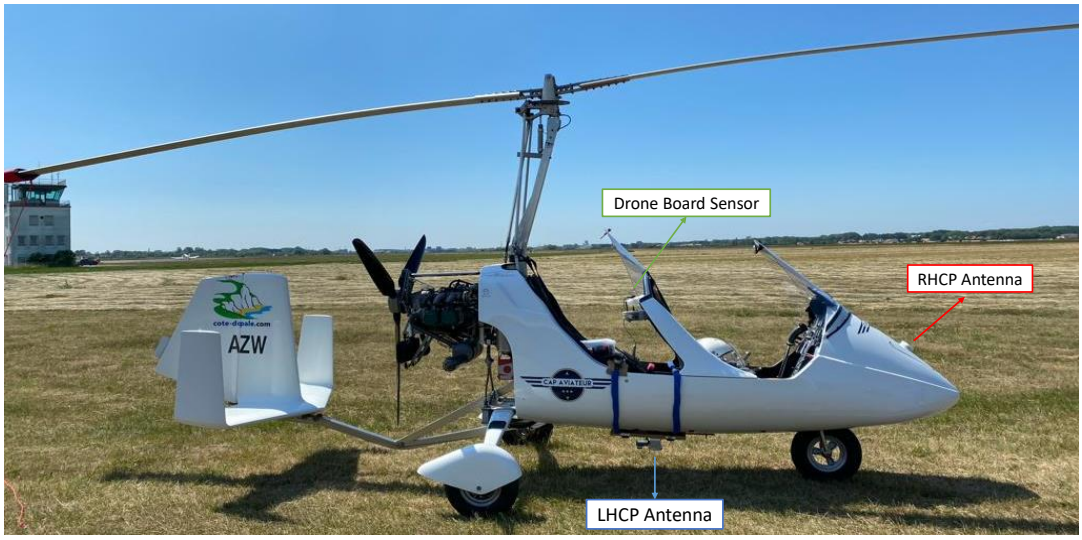


FIGURE 1.3: The gyrocopter used during the flights with the sensors embedded on it.

mono-channel digitizer. The usage of the delay line, which added an imposed time offset of approximately 1.5 *km* between the RHCP and the LHCP channels, along with the mono-channel bit grabber allows synchronous digitization of the GNSS signals from the two channels. The raw data were digitized and recorded using a Syntony bit grabber configured for GPS L1 signals at 25 MHz sampling frequency with 1-bit quantization, providing stable and reliable signal correlation functions essential for high-precision GNSS-R measurements.

Additional instrumentation comprised highly accurate attitude and navigation sensors that could provide the reliable geo-referencing of all of the GNSS-R measurements. A drone board sensor recorded the gyrocopter's position, altitude, and attitude with respect to GPS time throughout the flights' duration, providing the geometric information necessary for specular point localization and surface characterization. This comprehensive instrumentation suite enabled the collection of high-quality GNSS-R data and also the precise association of reflectivity measurements with specific surface locations and geometric configurations.

1.4.3 GNSS-R Signal Processing Architecture

The GNSS-R receiver implemented for this airborne campaign features a dual-channel design that comprises separate tracking loops optimized for processing direct and reflected GNSS signals from all visible satellites. As

shown in Figure 1.4, the flowchart describes the process of baseband signal processing after down-conversion and sampling of the direct signal (s_d) and the reflected signal (s_r) from the GNSS satellites.

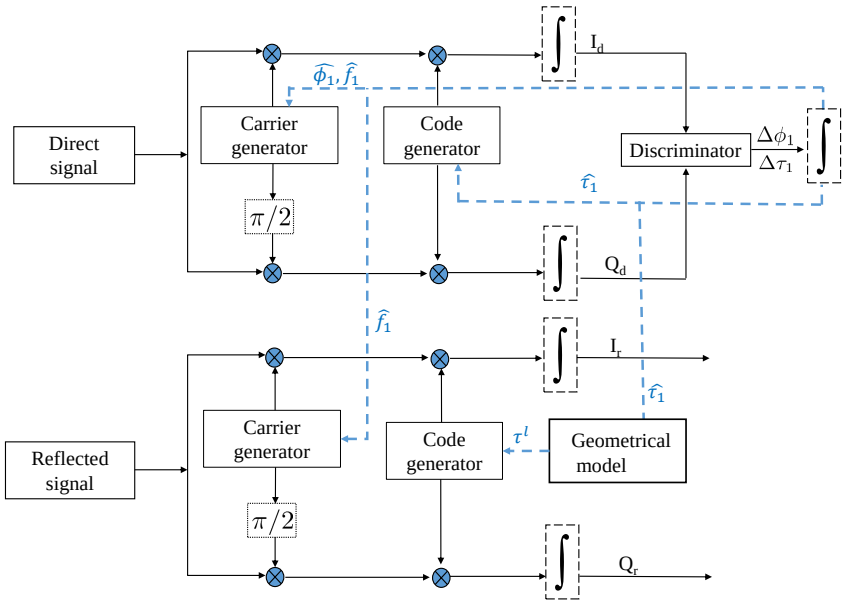


FIGURE 1.4: Flow chart of the base-band receiver processing.

The direct signal processing follows conventional GNSS receiver architecture. For GPS C/A signals, the received signal for satellite l can be expressed as:

$$s(t) = \sum_{l \in V} A^l CA^l(t + \tau^l) n^l(t + \tau^l) \sin[2\pi(f_{L1} + f_D^l)t + \phi^l] + \eta(t) \qquad (1.17)$$

where V represents the set of visible satellites, A^l is the amplitude of the signal, $CA^l(t)$ is the CDMA code, τ^l is the code delay, f_{L1} is the GPS L1 carrier frequency (1.57542 GHz), f_D^l is the Doppler frequency, ϕ^l is the carrier phase and $\eta(t)$ is the additive white Gaussian noise.

Tracking the direct signal employs traditional Phase-Locked Loop (PLL) and Delay-Locked Loop (DLL) techniques operating within integration time-scales of milliseconds [13]. The DLL considers both in-phase and quadrature-phase components via processing of early, prompt, and late iterations of pseudo-random noise (PRN) code replicas. Due to the dynamic nature of airborne GNSS-R operation, a Frequency-Locked Loop (FLL) is also implemented to accurately track the Doppler frequency along the flight

trajectory, primarily at coherent integration timescales.

In processing the reflected signals, we employ a new assisted tracking approach that uses the information contained in the direct signal channel [13]. Employing a master/slave configuration successfully addresses the leading challenge faced by reflected GNSS signals.

The assisted DLL uses pseudo ranges obtained from the direct signal and a model of the geometrical path difference to predict the delay experienced by the reflected signal as it traverses in specular reflection. The path difference, $\Delta p(t)$, between the reflected and direct signal is given as:

$$\Delta p(t) = 2h(t) \sin \theta_e(t)/c + \Delta_f \tag{1.18}$$

Here, $h(t)$ refers to the height of the receiver above the surface of reflection, $\theta_e(t)$ is the elevation angle of the satellite, c represents the speed of light, and Δ_f covers the hardware delays.

The hardware receiver configuration of the GNSS-R utilizes a delay line approach, which enables synchronous processing of both direct and reflected signals using a single-channel architecture. In addition, the phase-assisted open-loop tracking of the reflected signal leverages Doppler frequency estimates from tracking the direct signal, thus compensating for the geometric path differences.

This processing method facilitates coherent integration interval of 20 ms, enhancing signal-to-noise ratios while preserving the phase coherence essential for sophisticated GNSS-R applications such carrier phase altimetry.

1.4.4 Data Quality and Validation Framework

Development of an extensive data quality framework proved essential in ensuring reliability and scientific validity of the resulting GNSS-R observations throughout the airborne mission. The framework consists of multilevel quality control measures and validation processes, which together ensure the integrity of the experimental dataset used in this research work.

The primary quality control procedures established during the data acquisition process focus on signal strength and geometric aspects. A

minimum satellite elevation angle of 50° was established to increase the spatial resolution of the application, since higher elevation satellite signals tend to provide better gain for the LHCP antenna. The elevation criterion also ensures that the largest dimension of the principal axis of the first Fresnel zone remains less than 23 meters, thus providing adequate spatial resolution for surface characterization.

Specular points' coordinates and reflectivity observations were linked using GPS time provided by on-board sensors in conjunction with GPS time inferred from digitized GNSS observations. This time synchronization enables an accurate correspondence between GNSS observations and in-flight surface conditions witnessed by the flight trajectory.

The data processing quality framework incorporates statistical validation measures based on the theoretical foundation that GNSS-R reflectivity follows a Gamma distribution. The processing chain includes proper signal dating and synchronization procedures to ensure temporal consistency between GNSS observations and reference positioning data. The system architecture incorporates dedicated hardware components designed specifically for airborne GNSS-R applications, including appropriate antenna configurations and signal processing chains optimized for the simultaneous tracking of direct and reflected signals under dynamic flight conditions.

These quality control procedures collectively ensure the reliability and consistency of the experimental dataset for further analysis.

Figure 1.5 presents GNSS-R observations from Flight 1 to demonstrate the nature of the data acquired and motivate this research. This area was intentionally selected for its variety of land surfaces and more than 50 diverse water bodies, as shown in subfigure 1.5a with specular point traces of two high-elevation GPS satellites (PRN 7 and 5). As shown in subfigure 1.5b, the signal amplitude is significantly higher over water surfaces in both reflected signals. This variation forms the basis for inland water detection applications. The corresponding phase signals illustrated in subfigure 1.5c provide additional surface information, particularly demonstrating coherent behavior over smooth surfaces. The main objective of this dissertation is illustrated conceptually in subfigure 1.5d, which shows the reflected signal from PRN 5 to visualize the segmentation and classification of GNSS-R signals

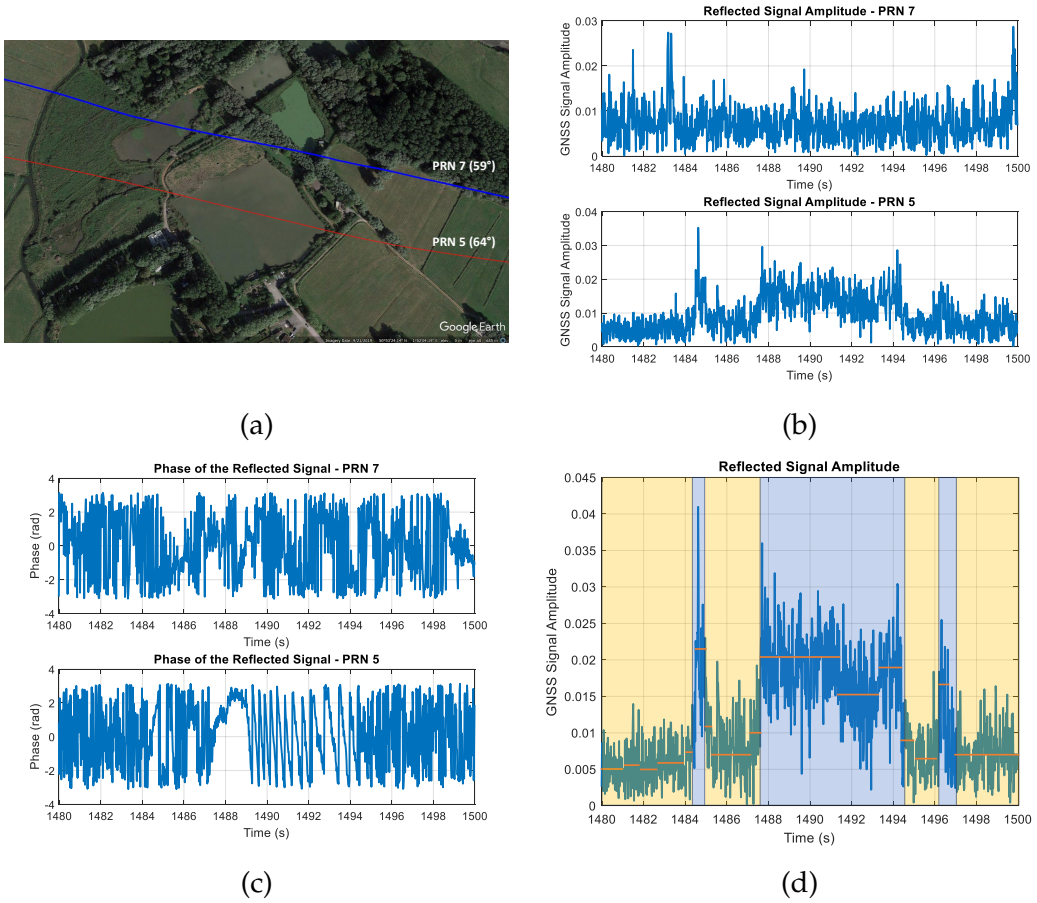


FIGURE 1.5: GNSS-R processing chain overview: (a) Satellite footprint traces (PRN 7 & 5) over inland water bodies, (b) reflected signal amplitude from PRN 5, (c) corresponding phase showing coherence variations, and (d) automated segmentation and surface classification results.

for distinguishing among different surface types.

1.5 Conclusion

The literature review has laid the theoretical groundwork and uncovered the key research gaps that drive the enhanced phase analysis of the advanced GNSS-R phase analysis in Chapter 4 of the thesis. The development of the GNSS-R signal processing from conventional power-based methods to phase-sensitive methods reflects an inherent limitation: conventional processing of the phase information is lost via incoherent integration, yet the phase properties of the signal on reflection have abundant information on the scattering mechanisms of the surface and coherence properties.

The detailed examination of bistatic radar fundamentals, integration techniques, and multi-constellation processing developments gives evidence of the growing acknowledgment that phase information provides proxy observables inaccessible by the use of amplitude observations alone. However, as the above assessment of current developments and gaps highlights, meaningful research gaps persist that constrain the operational utilization of phase-based GNSS-R systems.

The importance of proper statistical models that surpass empirical thresholding is one crucial research area to further enhance the abilities of GNSS-R to describe the surface.

These overall gaps and the pressing requirements of more advanced statistical paradigms that can translate theoretical knowledge to practical application provide the direct motivation for the research presented in Chapter 4, which presents new analytical methods address these fundamental limitations through systematic phase analysis of airborne GNSS-R measurements.

Chapter 2

Literature Review on Change Point Detection in Reflectometry Signals

2.1 Introduction

Change point detection (CPD) in reflectometry signals is a fundamental challenge due to the multiplicative speckle noise caused by coherent electromagnetic scattering phenomena. In GNSS-R, change point detection in reflectivity measurements is important for the detection of surface transitions and environmental monitoring, but speckle noise intrinsically alters signal statistics and makes traditional additive noise-based detection techniques insufficient.

Theoretical foundations to address this issue include both statistical models that are good descriptors of speckle-contaminated signals and algorithmic frameworks that can detect changes in such environments. While there is extensive literature on change point detection in additive noise environments, the multiplicative noise environment demands theoretical development and methodological innovation.

This comprehensive analysis provides a systematic review of the theoretical and methodological background required for change point detection in reflectometry signals. Section 2.2 builds the statistical modeling background based on speckle noise models, considering the Gamma distribution when dealing with multi-look GNSS-R signals. Section 2.3 classifies change point detection techniques, including unsupervised and supervised techniques, implementation schemes, and performance tuning. Section 2.4 offers an overview of current CPD techniques for multiplicative noise considering their theoretical backgrounds and limitations.

The theoretical foundation required to understand the limitations and potential of change point detection for reflectometry applications is provided by this comprehensive analysis. The review highlights the fundamental gaps in the current approaches, particularly in the areas of computational efficiency, robustness against parameter variations, and adaptability under varying surface conditions.

2.2 Different Statistical Models for Speckle Noise

Multiplicative noise, dominated principally by speckle noise [23, 40], is the principal challenge for signal analysis in coherent remote sensing systems. This inherent characteristic, in contrast to the fundamental additive noise, affects reflectivity signals which constitute the subject of this dissertation (as discussed in Chapter 3).

Speckle noise appears as a textured grain pattern resulting from coherent interference between waves scattered from a large number of elements over a resolution cell [40, 41]. Although this phenomenon is enlightening when it comes to sub-resolution surface topography, speckle patterns obscure larger-scale features of interest, and this makes it extremely difficult to perform activities such as change point detection [23, 42]. Consequently, there has been significant effort to statistically model speckle behavior, and as a result, several probabilistic models have been developed, predominantly in radar and SAR disciplines, as thoroughly reviewed in the literature [42, 43].

This part now briefly summarizes major speckle statistical models. Empirical models, selected for their ability to fit data, are introduced first, followed by physically based models, which connect speckle statistics to the underlying scattering physics. This provides an overview of the speckle modeling within this dissertation.

2.2.1 Empirical models

Empirical statistical models characterize speckle noise through mathematical distributions designed to optimally fit observed data patterns without explicit regard to the underlying electromagnetic scattering physics. The focus is on characterizing statistical properties of pixel intensities or amplitudes, thereby providing good tools for SAR image processing applications where

computational efficiency and mathematical tractability are of greater concern than physical interpretability.

The strength of empirical approaches is that they can be used in a flexible manner to represent diverse statistical behavior observed in real SAR data, particularly in complex circumstances where physical scattering mechanisms are difficult to model analytically. However, this flexibility comes at the cost of reduced physical insight, because the model parameters lack direct analogy to measurable surface parameters.

2.2.1.1 Log-Normal Distribution

The Log-normal distribution is among the most widely used empirical models of SAR imagery, particularly apt for high-resolution radar imaging and sea clutter modeling [44–47]. The distribution is particularly useful for modeling intensity variation in urban areas with spatially heterogeneous clutter in which more extreme values are observed than would be predicted under Gaussian models [23, 48].

The Log-normal distribution is selected based on its empirical ability to fit observed intensity distributions in heterogeneous scenes, rather than derivation from first-principles scattering theory. The probability density function is expressed as:

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \tag{2.1}$$

where x represents the intensity or amplitude, μ denotes the mean of the logarithm of x, and σ represents the standard deviation of the logarithm of x.

2.2.1.2 Weibull Distribution

Weibull distribution provides a general empirical model for the simulation of various kinds of clutter in SAR images, terrestrial, meteorological, and sea ice environments [23, 49–51]. The flexibility of the distribution is due to the shape parameter, which can be changed to match observed statistical behavior in various instances without explicit modeling of the scattering mechanisms. Probability density function is given by:

$$p(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b} \tag{2.2}$$

where a denotes the scale parameter and b represents the shape parameter that controls the distribution's form.

2.2.1.3 Fisher Distribution

The Fisher distribution is an empirical reparametrization of the physically-based G^0 distribution, widely applied to the description of highly spatially-resolved non-uniform SAR imagery [52]. It is, despite its physical model origin, empirically focused on fitting statistically rather than explicit consideration of the scattering mechanism.

The probability density function is expressed as:

$$p(x) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} \frac{x^{b-1}}{(1+x)^{b+c}} \tag{2.3}$$

where $\Gamma(..)$ denotes the Gamma function and c, like b, is a shape parameter that determines the distribution's characteristics.

2.2.1.4 Generalized Gamma Distribution

The Generalized Gamma Distribution (GΓD) is highly flexible in modeling as it can reduce to different specialist distributions under certain parameter conditions [53, 54]. This kind of ability facilitates the successful characterization of various types of terrain and scattering conditions, but the model parameters do not have an easy-to-interpret physical meaning.

Two canonical parameterizations are most commonly employed in the literature. The GΓD-1 formulation is as follows:

$$p(x) = \frac{b}{a^b \Gamma(a)} x^{ab-1} e^{-(\frac{x}{a})^b} \tag{2.4}$$

The alternative GΓD-2 parameterization is expressed as:

$$p(x) = \frac{b x^{bc-1}}{a^{bc} \Gamma(c)} e^{-(\frac{x}{a})^b} \tag{2.5}$$

2.2.1.5 Finite Mixture Models (FMMs)

Finite Mixture Models (FMMs) give high-level statistical models by formulating the joint intensity distribution as a weighted linear combination of simpler component distributions [48, 55]. It is a useful approach for Synthetic

Aperture Radar (SAR) scenes that image multiple types of terrain or multiple scattering mechanisms within one image.

It involves more parameters than the single distribution model, but has a wider range of applications. FMMs are considered empirical when the component distributions are themselves empirical and the mixing proportions are determined solely by data fitting, without reference to a physical model of scene composition. The PDF of an FMM is given by:

$$f_x(x|M) = \sum_{i=1}^{k} \alpha_i f_i(x|m_i) \tag{2.6}$$

where k denotes the number of mixture components, α_i are the mixing proportions such that $\sum_{i=1}^{k} \alpha_i = 1$ and $0 \leq \alpha_i \leq 1$, $f_i(x|m_i)$ represents the PDF of the i-th mixture component with parameters m_i and M is the set of all parameters.

2.2.1.6 Non-Parametric Modeling

Non-parametric estimation infers probability distributions from data without specifying functional forms, the most general empirical approach [56, 57]. Non-parametric techniques make as few assumptions as possible about the data generation mechanism underlying, relying solely on statistical inference from data.

Widely employed non-parametric techniques include kernel density estimation techniques, for example, the Parzen window method, machine learning techniques that encompass neural networks [43, 58, 59] and sophisticated classification techniques like support vector machines.

2.2.2 Physically Based Models

Compared with empirical approaches, physically based statistical models establish explicit relationships between quantified speckle features and underlying electromagnetic scattering physics. The physically based models derive their mathematics from fundamental principles of wave propagation and scattering theory, providing parameters in terms of measurable physical features of the surfaces being observed.

The major advantage of physically-based models is that they can be interpreted and theoretically founded, with explicit statistical parameters to surface properties such as roughness, dielectric constants and density of scatterers. This physical basis enables more effective extrapolation to novel conditions and the development of inversion algorithms for estimating surface parameters.

2.2.2.1 Rayleigh Distribution

The Rayleigh distribution is a fundamental model for speckle noise, particularly in homogeneous regions of SAR images [23, 60–62]. It arises from the central limit theorem applied to the coherent scatterer model, where the backscattered signal is considered the sum of numerous independent scatterers within a resolution cell. This scenario is typical of surfaces with a large number of randomly positioned, statistically independent scatterers of similar scattering amplitudes, such as short grass fields or calm water surfaces.

The Rayleigh model is established when the coherent scatterer model satisfies the following assumptions [62–64]:

- **Statistical Independence:** Scatterer amplitudes (a_i) and phases (ϕ_i) are mutually independent.
- **Spatial Randomness:** Scatterer amplitudes and positions are uncorrelated.
- **Identical Distributions:** Scattered amplitudes ($a_1, a_2, ...$) are independently and identically distributed with finite first and second moments, ensuring no dominant scatterers within the resolution cell.
- **Uniform Phase Distribution:** Random scatterer positions, with the maximum distance significantly greater than the wavelength, yield uniformly distributed phases in $(-\pi, \pi]$.
- **Large Population:** The number of scatterers (L_s) approaches infinity within a resolution cell.

Under these conditions, the real ($\Re$) and imaginary ($\Im$) components of the scattered field are expressed as:

$$\Re = \sum_{i=1}^{L_s} a_i \cos \phi_i \tag{2.7}$$

$$\Im = \sum_{i=1}^{L_s} a_i \sin \phi_i \tag{2.8}$$

Applying the central limit theorem, both $\Re$ and $\Im$ are Gaussian distributed random variables with zero mean and equal variance ($\sigma^2/2$). The joint probability density function (PDF) of $\Re$ and $\Im$ is given by [40]:

$$p_{\Re,\Im}(\Re, \Im) = \frac{1}{\pi\sigma^2} e^{-\frac{\Re^2+\Im^2}{\sigma^2}} \tag{2.9}$$

2.2.2.2 Gamma Distribution

The Gamma distribution emerges as a natural extension of the Rayleigh model for multi-look SAR processing [23, 42], where multiple independent observations of the same scene are averaged to reduce speckle noise. From a physical perspective, this distribution also models scenarios involving incoherent summation of multiple Rayleigh-distributed contributions. This may occur in scenarios where the resolution cell contains multiple independent homogeneous scattering areas, such as agricultural fields with varying vegetation density or forests with different canopy layers.

If n pixels are averaged, the resulting image is called an n-look image and the corresponding n-look intensity, I_n, is given by:

$$I_n = \frac{1}{n} \sum_{\ell=1}^{n} I(\ell) \tag{2.10}$$

where $I(\ell)$ represents the intensity of the ℓ-th look. The PDF of the Gamma distribution for n-look intensity is:

$$p(I_n) = \frac{1}{\Gamma(n)} \left(\frac{n}{\mu}\right)^n I_n^{n-1} e^{-\frac{nI_n}{\mu}} \tag{2.11}$$

where μ represents the mean intensity and $\Gamma(n)$ denotes the Gamma function.

The Gamma distribution has a mean and variance of μ and $\frac{\mu^2}{n}$, respectively. The coefficient of variation (CV) of the Gamma distribution is $\frac{1}{\sqrt{n}}$, which implies that the speckle noise is reduced as the number of looks is increased.

The gamma distribution can also result under the incoherent scatterer sum model when the number of scatterers within a resolution cell follows the negative binomial distribution with an infinite mean [47, 61].

2.2.2.3 G^0 Distribution

The G^0 distribution addresses the limitations of homogeneous scattering models by incorporating the spatial variability of the radar cross-section (RCS) through the product model framework, which considers the SAR image intensity as the product of the RCS and speckle. This statistical model is particularly relevant for high-resolution SAR imagery [52, 65–67], where the common assumption of numerous scatterers per resolution cell—required by the Rayleigh distribution—may be violated. The G^0 distribution effectively describes inhomogeneous regions with spatially varying RCS, such as forests, urban areas and mixed terrain regions.

$$p(x;a,b,n) = \frac{n^n \Gamma(n-b) a^{-b} x^{n-1}}{\Gamma(n)\Gamma(-b)(a+nx)^{n-b}} \tag{2.12}$$

The distribution arises when the RCS intensity follows an inverse gamma distribution and the speckle follows a gamma distribution under the product model framework [67]. The shape parameter b, where b<0, indicates the surface roughness and can characterize different land cover types: $b \leq -15$ represents homogeneous area such as grassland, $b \in (-15, -5]$ represents inhomogeneous areas such as forests and $b \in (-5, 0]$ represents extremely inhomogeneous area.

Mathematically, the G^0 distribution is equivalent to the Fisher distribution, also called "Fisher–Snedecor" [68].

2.2.2.4 Rice Distribution

The Rice distribution models the amplitude of a circularly symmetric Gaussian random variable with a non-zero mean [69]. In SAR imagery, the Rice distribution arises when the scattered field is the summation of a constant amplitude A_0 and an infinite number of random scatterers of amplitude a_i [40, 70]:

$$E = A_0 + \sum_i a_i e^{j\phi_i} = A_0 + A_n e^{j\theta_n} \tag{2.13}$$

The scattered amplitude $r = |E|$ then follows the Rice distribution:

$$p(r) = \frac{r}{\sigma_n^2} \exp\left(-\frac{r^2 + A_0^2}{2\sigma_n^2}\right) I_0\left(\frac{rA_0}{\sigma_n^2}\right) \tag{2.14}$$

where σ_n denotes the standard deviation of the zero-mean Gaussian noise component and $I_0(\cdot)$ is the modified zero-order Bessel function of the first kind [71]. When $A_0 = 0$, the Rice distribution reduces to the Rayleigh distribution.

For the intensity variable $x = r^2$, the corresponding PDF becomes:

$$p(x) = \frac{1}{2\sigma_n^2} \exp\left(-\frac{x + A_0^2}{2\sigma_n^2}\right) I_0\left(\frac{A_0\sqrt{x}}{\sigma_n^2}\right) \tag{2.15}$$

2.2.2.5 Generalized Gaussian Distribution

The Generalized Gaussian Distribution (GGD) effectively models the real and imaginary components of scattered fields [52]. It generalizes the standard Gaussian model, allowing for heavier or lighter tails through a shape parameter that adjusts kurtosis. When the real and imaginary components follow the GGD, the amplitude follows the Generalized Gaussian Rayleigh (GGR) distribution [42, 72]. This makes the model suitable for heterogeneous scattering regions that deviate from Rayleigh assumptions.

The probability density function (PDF) of the GGD is given by:

$$p(u; a, c, \mu) = \frac{a\, c}{2\,\Gamma\left(\frac{1}{c}\right)} \exp[-|a(u-\mu)|^c], \tag{2.16}$$

When the shape parameter $c = 2$, this reduces to the standard Gaussian distribution, and when $c = 1$, it becomes the Laplace distribution.

2.2.3 Discussion

The selection of appropriate statistical models for speckle characterization is a balance between mathematical flexibility, computational efficiency, and physical interpretability. Empirical models provide superior data-fitting accuracy and mathematical flexibility, particularly for cases involving heterogeneous environments, but with the disadvantage of limited physical insight into surface characteristics. Physically-based models, alternatively, establish direct connections between statistical parameters and observable surface characteristics, enabling reasonable algorithm development with

theoretical foundation.

In coherent remote sensing applications, particular properties of bistatic geometry and intrinsic correlation processing multi-look averaging inherently require the use of physically-based models, with the Gamma distribution standing out among them. This choice aligns with the multi-look processing nature of coherent signals and also provides the theoretical framework that is suitable for efficient change point detection algorithms.

The Gamma distribution is found to be appropriate for this purpose on physical basis in multi-look processing, mathematical tractability for algorithm implementation, and sufficient flexibility due to its two-parameter nature. This rationale has been documented along with experimental validation in coherent remote sensing literature.

This review establishes that although empirical models offer mathematical flexibility, physically-based models provide the interpretability and theoretical basis that is needed for coherent remote sensing applications. For the purposes of this dissertation, the Gamma distribution will be employed for our GNSS-R application.

2.3 Single Change Point Detection: Methods and Algorithms

Having established the statistical framework for speckle modeling, the focus shifts to change detection techniques for statistically modeled signals.

Online Change Point Detection (CPD) refers to the set of algorithms to identify abrupt changes or shifts in time series data as the data arrives sequentially. The algorithms are critical for real-time applications where rapid detection of changes is crucial for monitoring, decision-making and adaptive control of systems. The overall goal of online CPD is to identify a change point at the smallest possible delay after the change point has been reached.

Formally, single change point detection presents the statistical inference problem of whether or not there exists an unobserved time point τ in a sequence of observations $y_1, y_2, \ldots, y_T$ such that the underlying probability

distribution that is generating the observations has switched regimes [73].

The methods can be broadly categorized as shown in Figure 2.1 in terms of the nature of the change that they are designed to identify, their implementation strategy and the methods used for tuning their performance. This review focuses exclusively on unsupervised change point detection methods. The following sections therefore examine unsupervised techniques for detecting both abrupt and gradual changes.

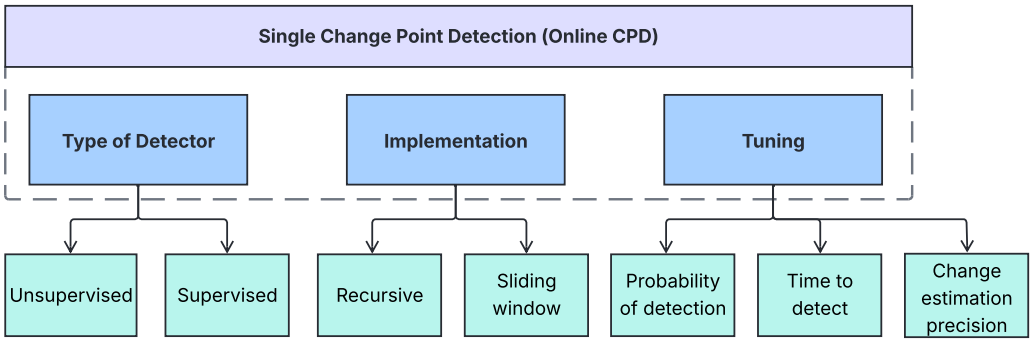


FIGURE 2.1: Typology of the online change point detection algorithms.

2.3.1 Unsupervised

Unsupervised CPD methods operate without prior knowledge of system states or labeled examples of changes. They detect changes by identifying deviations from established statistical properties of the data stream itself.

2.3.1.1 Abrupt Change

Abrupt change algorithms detect sudden, instantaneous shifts in the underlying distribution or parameters. These changes are modeled as step functions, reflecting instantaneous or near-instantaneous transitions in statistical characteristics that occur on time scales much shorter than the sampling interval.

Abrupt changes appear as sudden discontinuities that produce discrete statistical regimes with well-defined temporal boundaries. Mathematically, it appears as a piecewise constant parameter model where the system remains stable in each regime but suddenly changes from one regime to another.

For a series of observations, an abrupt change at time τ is defined as:

$$y(t) = \begin{cases} y_0, & \text{for } t < \tau \\ y_1, & \text{for } t \geq \tau \end{cases} \tag{2.17}$$

where $t \in \{1, 2, \ldots, T\}$.

The applicability of abrupt change detection prevails in numerous application domains with the requirement of real-time monitoring of systems and fault detection. In process industries, abrupt changes signal equipment, sensor, or process upset faults requiring immediate attention [73]. In biomedical systems, abrupt changes may represent pathological occurrences or therapy effects requiring urgent clinical intervention [73].

A. Generalized Likelihood Ratio (GLR) Algorithm

The Generalized Likelihood Ratio (GLR) algorithm represents one of the most effective and universally applicable techniques for the detection of abrupt changes in signals and linear systems [74, 75]. Originally developed for situations where post-change parameters are unknown, the GLR method generalizes the classic likelihood ratio procedure to realistic scenarios lacking a priori knowledge.

The theoretical foundation of GLR derives from the Neyman-Pearson lemma, which establishes optimality of likelihood ratio tests for hypothesis testing with known parameters [74, 76, 77]. While optimality does not directly extend to the unknown parameter case, GLR provides a theoretically motivated extension that maintains the hypothesis testing framework while adapting to unknown parameter scenarios.

Consider a linear state-space system described by:

$$x_{t+1} = F_t x_t + G_t u_t + \zeta_t + \delta(t - \tau)v \tag{2.18}$$

$$y_t = H_t x_t + \xi_t \tag{2.19}$$

where x_t represents the state vector at time t, u_t denotes the known input vector, y_t is the observation vector, ζ_t and ξ_t are zero-mean Gaussian noise processes, $\delta(t - \tau)$ is the Dirac delta function, v represents the unknown jump magnitude occurring at unknown time τ and F_t, G_t, H_t are known system

matrices [75].

The fundamental hypothesis testing problem becomes:

$\boldsymbol{H_0}$: No jump occurs (($v = 0$)).

$\boldsymbol{H_1(t, v)}$: A jump of magnitude v occurs at time t.

The likelihood functions under each hypothesis are:

$$p\left(y_{1:t} \mid H_0\right) = p\left(y_{1:t} \mid t = T\right) \tag{2.20}$$

$$p\left(y_{1:t} \mid H_1(t, v)\right) = p\left(y_{1:t} \mid t, v\right) \tag{2.21}$$

The log-likelihood ratio for testing $H_1(t, v)$ against H_0 is:

$$\ell(t, v) = 2 \ln \left(\frac{p(y_{1:T} \mid t, v)}{p(y_{1:T} \mid t = T)} \right) \tag{2.22}$$

This ratio quantifies the evidence supporting the jump hypothesis relative to the no-jump hypothesis [75].

B. Bayesian Online Change Point Detection Algorithm

Bayesian Online Change Point Detection (BOCPD) takes a fundamentally different approach by using prior knowledge about where changes might occur, what parameter values to expect, and how models should be structured [73, 78]. This framework provides principled uncertainty quantification, computing posterior probabilities over change point locations rather than point estimates, particularly valuable for applications where comprehensive risk analysis is required [79].

The Bayesian approach simultaneously accommodates multiple sources of uncertainty: uncertainty about change occurrence, timing and characterizing parameters for the pre-change and post-change regimes. This comprehensive treatment of uncertainty distinguishes Bayesian methodology from frequentist approaches and enables more informed decision-making in complex scenarios [78, 79].

The Bayesian change point detection problem begins with the assumption that observations can be segmented into independent and identically

distributed sections, each characterized by different parameter values.

The mathematical framework employs a product partition model, assuming time-series data can be segmented into contiguous sections with change points, where all data points within each section are sampled from identical parameters:

$$\theta_i \sim p(\theta_i|\eta) \tag{2.23}$$

where θ_i represents the parameters for the run starting at time i and η represents hyperparameters shared between all priors [79].

The product partition model can be equivalently formulated as a Hidden Markov Model (HMM) with a potentially infinite number of hidden states [79, 80]. Let $\rho_t \in \{0,1\}^t$ represent a latent variable specifying which parameter setting θ_i was responsible for generating observation y_t, with the constraint that $\sum_{i=1}^{t} \rho_{t,i} = 1$ [73]. This generative process can be described as follows:

$$\rho_t|\rho_{t-1} \sim Q(\rho_t, \rho_{t-1}) \tag{2.24}$$

$$y_t \mid \theta_i,\ \rho_{t,i} = 1 \sim f(y_t \mid \theta_i) \tag{2.25}$$

where Q represents the conditional prior governing the transitions between possible segmentations. In this formulation, the variable ρ_t identifies which run (or segment) is currently active. If $\rho_{t,i} = 1$, it means that the observation y_t was generated under the parameter θ_i, corresponding to the regime that began at time i.

The run-length is then defined as the number of observations since the most recent change point, and it can be expressed as:

$$\omega_t = t - i \ \textbf{ if } \rho_{t,i} = 1 \tag{2.26}$$

Hence, ρ_t identifies the current run and ω_t quantifies its length. The pair (ρ_t, ω_t) thus jointly describe the segmentation structure of the time series.

The transition between regimes is regulated by a hazard function, $q(\cdot)$, which specifies the probability that a new change point occurs after a given run length. The probability that a change occurs at time t is represented by $q(\omega_{t-1}+1)$, given that the current regime has lasted for $\omega_{t-1}+1$ observations.

This function defines the prior over the run-length process and is often modeled using a geometric distribution, assuming a constant probability of change over time (i.e., a memoryless process) [80, 81]. Accordingly, the run-length prior can be written as:

$$p(\omega_t = 0 \mid \omega_{t-1} = \rho) = q(\rho + 1), \tag{2.27}$$

$$p(\omega_t = \omega_{t-1} + 1 \mid \omega_{t-1} = \rho) = 1 - q(\rho + 1), \tag{2.28}$$

This probabilistic formulation enables recursive Bayesian updating of the posterior over possible run-lengths as new data arrive, forming the basis of the Bayesian online change point detection framework [79, 81].

The central objective in Bayesian change point detection is to estimate the joint posterior distribution over possible run-lengths and model parameters [80]. This posterior characterizes the full uncertainty about both the timing of changes and the statistical properties of each segment. Given the sequence of observations $y_{1:t}$, the joint posterior can be expressed as:

$$p(\rho_t, \theta \mid y_{1:t}) \propto p(y_{1:t} \mid \rho_t, \theta)\, p(\rho_t, \theta) \tag{2.29}$$

where the prior factorizes as:

$$p(\rho_t, \theta) = p(\rho_t)\, p(\theta \mid \rho_t) \tag{2.30}$$

This formulation enables the recursive calculation of posterior probabilities over change point positions, providing a comprehensive probabilistic specification of the detection problem [78, 79]. To ensure computational tractability, the model usually uses conjugate exponential family models, in which the likelihood is in an exponential family and the prior is set to be the respective conjugate distribution, allowing for analytic integration over parameters and closed-form posterior updates [73, 80, 82].

C. β-Divergence Algorithm for Robust Bayesian Inference

The β-divergence algorithm is a significant innovation in robust Bayesian change point detection that addresses major limitations of traditional Kullback-Leibler divergence (KLD) based methods under the influence of outliers and model misspecification [83]. Traditional Bayesian inference aims to minimize the KLD between the estimated model and data generation

process, but is not robust as the KLD/log-score has a strictly increasing influence function that cannot accommodate outlying values [83].

The β-divergence algorithm remedies this limitation by employing β-divergences with an influence function with a well-defined peak, which enables effective outlier handling without losing the theoretical foundations of Bayesian inference. The approach is founded on the assumptions of General Bayesian Inference (GBI), extending standard Bayesian updating from the KLD to a broader family of divergences for enabling stronger robustness [83].

The theoretical foundation of β-divergence inference is based on the interaction between losses ℓ and divergences D such that for any given divergence D, there exists an associated loss ℓ_D. Under a model, the best posterior update for divergence D results in the following distribution:

$$p(\theta \mid y_{(t-\omega_t+1):t}) \propto p(\theta) \exp\left\{ - \sum_{i=t-\omega_t+1}^{t} \ell_D(\theta \mid y_i) \right\} \tag{2.31}$$

For parameter inference with the KLD and β-divergence, the corresponding losses are the log score and the Tsallis score respectively [83]:

$$\ell_{\text{KLD}}(\theta|y_t) = -\log f(y_t|\theta) \tag{2.32}$$

$$\ell_\beta(\theta|y_t) = -\frac{1}{\beta_p}\left[f(y_t|\theta)^{\beta_p} \left(\int f(z|\theta)^{1+\beta_p} dz \right)^{-1/(1+\beta_p)} - 1 \right] \tag{2.33}$$

where $\beta_p \in [0,1]$ controls the degree of robustness.

The β-divergence algorithm generalizes the BOCPD by introducing robustness into the run-length and parameter posteriors. The algorithm uses β-divergence posteriors $p_\beta^{rlm}(\omega_t \mid y_{1:t})$ for the run-length and $p_\beta(\theta \mid y_{1:t})$ for the parameters, with $\beta = (\beta_{rlm}, \beta_p) > 0$, where β_{rlm} guards against abrupt changes in the run-length posterior due to few observations [83].

The robust BOCPD implementation substitutes the usual predictive densities $f(y_t \mid y_0)$ and $f(y_t \mid y_{1:t-1})$ with their β-divergence versions:

$$\exp\left(\ell_{\beta_{rlm}}(\theta \mid y_t)\right) \tag{2.34}$$

where:

$$\ell_{\beta_{rlm}}(\theta \mid y_t) = -\frac{1}{\beta_{rlm}}\Bigg[f(y_t \mid y_{1:t-1})^{\beta_{rlm}} \tag{2.35}$$

$$\cdot \left(\int f(z \mid y_{1:t-1})^{1+\beta_{rlm}}\, dz\right)^{-1/(1+\beta_{rlm})} - 1\Bigg] \tag{2.36}$$

One of the basic theoretical contributions of the β-divergence algorithm is the formal odds ratio characterization of robustness. In the case of Bayesian Linear Regression models, the algorithm offers conditions that guarantee a single outlying observation y_{t+1} cannot force a false change point declaration [83].

The odds of $w_{t+1} \in \theta, w+1$ conditional on $\omega_t = w$ are given by:

$$\frac{p\left(\omega_{t+1} = 0 \mid y_{1:t+1}, \omega_t = w\right)}{p\left(\omega_{t+1} = w+1 \mid y_{1:t+1}, \omega_t = w\right)} = \frac{q(w+1)\, f_\beta\left(y_{t+1} \mid y_{1:t}\right)}{\left(1 - q(w+1)\right) f_\beta\left(y_{t+1} \mid y_0\right)} \tag{2.37}$$

where f_β represents the β-divergence predictive density and $q(\cdot)$ is the hazard function defined in equation 2.28.

Because β-divergence posteriors are intractable, the algorithm uses Structural Variational Inference *(SVI)* that retains parameter dependence and becomes exact as $\beta_p \to 0$. In contrast to mean-field approximations, SVI maintains parameter correlations while allowing computational tractability.

The variational approximation is optimized via variance-reduced Stochastic Gradient Descent with anchored gradients:

$$\theta \leftarrow \theta + \eta \cdot \left[\nabla^{\text{new}} - \nabla^{\text{old}} + \nabla^{\text{anchor}}\right] \tag{2.38}$$

where the anchor gradients supply variance reduction and better convergence for small run-lengths ω_t.

A key component of the algorithm is the principled choice of the divergence parameter β. The algorithm initializes β_p by maximizing the β-divergence impact for a fixed Mahalanobis Distance under the prior $p(\theta)$ and then optimizes β online by minimizing predictive losses:

$$\beta_t = \beta_{t-1} - \eta \cdot \begin{bmatrix} \nabla_{\beta_{rlm,t}} J\left(\varepsilon_t\left(\beta_{1:(t-1)}\right)\right) \\ \nabla_{\beta_{p,t}} J\left(\varepsilon_t\left(\beta_{1:(t-1)}\right)\right) \end{bmatrix} \tag{2.39}$$

where $\varepsilon_t(\beta) = y_t - \hat{y}_t(\beta)$ is the prediction error and J is a suitable bounded loss function.

This principled strategy for β selection allows automatic adaptation to the level of contamination and model misspecification in the data, yielding a self-tuning robust inference procedure that outperforms standard methods by a considerable margin in practical problems [83].

D. Spectral Distance Algorithm

The spectral distance approach is a fundamental method for remote sensing and signal processing change detection, where change manifests itself as a change in spectral characteristics between temporal instances [84]. Unlike parametric methods that assume specific distributional forms, spectral distance techniques measure dissimilarities between corresponding pixels or signal segments using various distance metrics, providing a flexible framework for change detection across diverse application domains [84].

Spectral distances possess certain advantages in cases when changes' character is unknown, or in the case of high-dimensional data when usual parametric assumptions may be inappropriate. The approach is fast computationally, simple to apply and does not require ground-truth training data and is thus especially interesting for unsupervised change detection applications.

Consider multidimensional observations that can be represented as vectors. In different applications, these may correspond to spectral bands in imaging systems, delay-Doppler bins in radar systems, or feature channels in multi-sensor configurations. At each measurement location (i, j) let $y^{(1)}_{(i,j)}$ and $y^{(2)}_{(i,j)}$ represent observation vectors at two temporal instances, where each vector contains B measurements. The spectral distance framework computes dissimilarity measures $d(y^{(1)}, y^{(2)})$ to quantify changes between observations

[84].

The spectral distance framework incorporates multiple distance measures, each with distinct mathematical properties and sensitivity characteristics. Key formulations include:

- Euclidean Distance:

$$d_{\text{Euclidean}} = \sqrt{\sum_{k=1}^{B} \left(y_{i,j,k}^{(1)} - y_{i,j,k}^{(2)} \right)^2} \tag{2.40}$$

- City-Block (Manhattan) Distance:

$$d_{\text{City-Block}} = \sum_{k=1}^{B} \left| y_{i,j,k}^{(1)} - y_{i,j,k}^{(2)} \right| \tag{2.41}$$

- Spectral Angle Mapper (SAM):

$$d_{\text{SAM}} = \cos^{-1} \left(\frac{\sum_{k=1}^{B} y_{i,j,k}^{(1)} y_{i,j,k}^{(2)}}{\sqrt{\sum_{k=1}^{B} \left(y_{i,j,k}^{(1)} \right)^2} \sqrt{\sum_{k=1}^{B} \left(y_{i,j,k}^{(2)} \right)^2}} \right) \tag{2.42}$$

where B represents the total number of measurement dimensions.

One highly significant class of spectral distances relies on metrics of correlation measures, which have demonstrated superior performance in comparative studies. The Pearson correlation distance is defined as:

$$d_{\text{Pearson}} = 1 - \frac{\sum_{k=1}^{B} \left(y_{i,j,k}^{(1)} - \bar{y}_{i,j}^{(1)} \right) \left(y_{i,j,k}^{(2)} - \bar{y}_{i,j}^{(2)} \right)}{\sqrt{\sum_{k=1}^{B} \left(y_{i,j,k}^{(1)} - \bar{y}_{i,j}^{(1)} \right)^2} \sqrt{\sum_{k=1}^{B} \left(y_{i,j,k}^{(2)} - \bar{y}_{i,j}^{(2)} \right)^2}} \tag{2.43}$$

where $\bar{y}^{(1)}$ and $\bar{y}^{(2)}$ represent the mean values across dimensions for the respective observations.

The Spearman correlation employs rank-based correlation, where observations are first converted to their ordinal ranks based on sorted positions, making this distance measure robust to outliers and non-linear monotonic

relationships:

$$d_{\text{Spearman}} = 1 - \frac{6\sum_{k=1}^{B} d_k^2}{B(B^2-1)} \tag{2.44}$$

where d_k represents the difference between ranks of corresponding dimensions.

Spectral distance algorithms convert computed distance measures into binary change decisions in accordance with thresholding operations. Selecting the best threshold becomes a critical factor as it directly influences the detection accuracy and false alarm rates.

For a computed distance measure d, the binary change decision is given by:

$$\begin{cases} 1 & \text{if } d \geq h \\ 0 & \text{otherwise} \end{cases} \tag{2.45}$$

where h represents the detection threshold.

The algorithm commonly employs the automatic threshold approach by Otsu for optimal thresholding [85]. This approach computes the threshold by reducing intra-class variance, or equivalently by maximizing inter-class variance:

$$h^* = \arg\min_h \left[\omega_1(h)\sigma_1^2(h) + \omega_2(h)\sigma_2^2(h)\right] \tag{2.46}$$

where $\omega_1(h)$ and $\omega_2(h)$ represent the probabilities of the two classes and $\sigma_1^2(h)$ and $\sigma_2^2(h)$ denote their respective variances.

In order to obtain more precision, the algorithm can use multilevel threshold to find multiple optimal thresholds simultaneously:

$$h_1^*, h_2^*, \ldots, h_K^* = \arg\min_{h_1,\ldots,h_K} \sum_{i=1}^{K+1} \omega_i \sigma_i^2 \tag{2.47}$$

where K is the number of thresholds. This allows more sophisticated change classifications than simple binary alternatives.

2.3.1.2 Slow or Gradual Change

Previous approaches detected abrupt changes in system parameters, but most real processes instead degrade or drift slowly over time. Slow changes, containing continuous deviations from nominal operating regimes, are

challenging to detect with algorithms designed for abrupt shifts [73]. Unlike abrupt changes showing up as discrete jumps in parameter values, gradual changes are continuous evolutionary tendencies where parameters slowly deviate from their initial levels.

Slow change detection employs techniques to accumulate evidence rather than depending on single observations. Adaptive estimation monitors changes [73], but sequential detection is necessary for immediate action on cumulative changes. Moving from single-event to trend analysis is necessary to concentrate on processes rather than on events.

A. The Cumulative Sum (CUSUM) Algorithm

Cumulative Sum, or CUSUM, is probably the simplest but most widely applied algorithm for deciding sequential changes. The CUSUM method was first suggested by Page [86] in 1954, with subsequent formal development by Lorden [87] in 1971. CUSUM relies on the idea of accumulating deviations from some optimum or target value and thus is highly sensitive to slow, consistent changes.

The basic CUSUM statistic for an upward shift in mean from λ_0 to λ_1 is given by:

$$g_t^+ = \max(0, g_{t-1}^+ + y_t - \lambda_0 - \Lambda) \tag{2.48}$$

where $\Lambda = \frac{\lambda_1 - \lambda_0}{2}$ represents the reference value and $g_0^+ = 0$. For bidirectional detection, the algorithm employs parallel statistics:

$$g_t^- = \max(0, g_{t-1}^- - y_t - \lambda_0 - \Lambda) \tag{2.49}$$

The stopping rule declares a change at time:

$$\tau_a = min(t : g_t^+ \geq h \text{ or } g_t^- \geq h) \tag{2.50}$$

where h represents the decision threshold.

The Sequential Probability Ratio Test (SPRT) is the basis for the CUSUM algorithm. The log-likelihood ratio for identifying a transition from λ_0 to λ_1 at unknown time τ is [88]:

$$\log \ell_t = \sum_{i=\tau}^{t} \log \frac{f(y_i|\lambda_1)}{f(y_i|\lambda_0)} \tag{2.51}$$

The CUSUM statistic maximizes this over all possible change times:

$$g_t = \max_{1 \le \tau \le t} \log \ell_t \tag{2.52}$$

The V-mask gives a geometric representation of the CUSUM algorithm [86]. The cumulative sum is graphed versus time and a V-shaped mask is overlaid. The mask parameters are defined as:

- Lead distance: $d_L = \frac{h}{\Lambda}$
- Half-angle: $\Omega = \arctan\left(\frac{\Lambda}{\sigma}\right)$

where σ denotes the standard deviation of the process in the in-control state (i.e., before any change occurs). A change is indicated when the cumulative sum trajectory intersects either arm of the V-mask.

CUSUM uses score values for continuous inspection systems [86]. Consider a scoring system in which every defective item is assigned a score of +1 and every non-defective is assigned a score of -1. The cumulative score is:

$$g_t = g_{t-1} + \psi_t \tag{2.53}$$

where ψ_t is the score assigned to the $t - th$ inspected item. If the cumulative score g_t becomes large and positive, it indicates a possible deterioration in the process (increasing number of defectives). Conversely, if g_t becomes large and negative, it suggests improvement (decreasing number of defectives). The CUSUM test signals a change when g_t exceeds a predefined threshold h.

Traditional CUSUM can be sensitive to outliers due to its unbounded quadratic loss function [89]. Robust variants address this limitation by employing bounded loss functions that reduce the influence of extreme deviations. The *Huber loss*-based CUSUM replaces the standard quadratic loss with:

$$Loss_H(y_t - \lambda_0) = \begin{cases} \frac{1}{2}(y_t - \lambda_0)^2, & \text{if } |y_t - \lambda_0| \le \eta_H \\ \eta_H |y_t - \lambda_0| - \frac{1}{2}\eta_H^2, & \text{if } |y_t - \lambda_0| > \eta_H \end{cases} \tag{2.54}$$

The *biweight loss* provides even stronger robustness:

$$Loss_B(y_t - \lambda_0) = \begin{cases} \frac{\eta_B^2}{6}\left[1 - \left(1 - \left(\frac{y_t - \lambda_0}{\eta_B}\right)^2\right)^3\right], & \text{if } |y_t - \lambda_0| \leq \eta_B \\ \frac{\eta_B^2}{6}, & \text{if } |y_t - \lambda_0| > \eta_B \end{cases} \tag{2.55}$$

where η_H and η_B are the respective tuning constants controlling the robustness of the Huber and biweight loss functions. Smaller values of these constants make the CUSUM statistic less sensitive to outliers but slower to detect small changes.

2.3.2 Implementation

Successful implementation of the change point detection algorithm requires careful consideration of computational structure and memory management strategies. Two main paradigms of implementation have emerged as the dominant methodologies: recursive methods that maintain concise representations of state and sliding window strategies that operate via fixed-size data buffers. Paradigm selection directly affects computational complexity, memory demand and viability for real-time applications [90].

2.3.2.1 Recursive Implementation Methods

Recursive implementations maintain algorithmic state through efficient summary statistics that are incrementally updated with the addition of new observations, thus avoiding detailed data histories by passing adequate statistics over time [90]. The global structure of recursive updates maintains a state vector s_t that contains all relevant information derived from the observations:

$$s_t = f(s_{t-1}, y_t) \tag{2.56}$$

where $f(\cdot)$ represents the recursive update function. The change detection decision is based solely on the current state:

$$\text{Decision}_t = g(s_t) \tag{2.57}$$

For **CUSUM** algorithms, the recursive state consists of the cumulative sum statistic mentioned in equations 2.48 and 2.49. The algorithm maintains only two scalar values regardless of sequence length, achieving $O(1)$ space

complexity.

BOCPD is another example of a recursive architecture. It maintains a posterior distribution over the run length variable ω_t, representing the number of samples since the last change point. The update equations for $P(\omega_t|y_{1:t})$ follow directly from the recursive Bayesian formulation previously introduced, and are derived in detail there [82, 90]. BOCPD thus fits naturally within the general recursive framework by maintaining and updating sufficient statistics through time.

GLR methods can also be formulated recursively through incremental updates of sufficient statistics. At each time step, the likelihood ratio is updated incrementally as new data arrive, enabling efficient online implementation without storing the full observation history. This recursive structure makes GLR-based approaches suitable for real-time applications and computationally constrained systems, where parameter estimates and likelihood values are propagated as new data become available [80].

2.3.2.2 Sliding-Window Implementation Methods

Sliding-window implementations process fixed-size windows of recent data, offering computational efficiency and bounded memory requirements [90]. Consider a sliding window of size w_s containing the most recent samples, the detection statistic is computed over this window:

$$S_t = g(y_t) = g\big([\, y_{t-w_s+1}, \ldots, y_t \,]\big) \tag{2.58}$$

Typical sliding-window statistics include moving averages and variances:

$$\bar{y}_t = \frac{1}{w_s} \sum_{i=t-w_s+1}^{t} y_i \tag{2.59}$$

$$s_t^2 = \frac{1}{w_s - 1} \sum_{i=t-w_s+1}^{t} \big(y_i - \bar{y}_t\big)^2 \tag{2.60}$$

For efficient computation, the mean can be updated recursively as:

$$\bar{y}_t = \bar{y}_{t-1} + \frac{y_t - y_{t-w_s}}{w_s} \tag{2.61}$$

The Page–Hinkley test maintains cumulative sums within the sliding window:

$$m_t = \sum_{i=t-w_s+1}^{t} \left(y_i - \bar{y}_t - \nu\right) \qquad M_t = \min_{t-w_s+1 \le i \le t} m_i \tag{2.62}$$

A change is declared when $m_t - M_t > h_{\mathrm{PH}}$.

Multi-window schemes employ several windows of different sizes to capture dynamics at multiple time scales:

$$S_t^{(w_s)} = g\left(y_t^{(w_s)}\right) \tag{2.63}$$

for window sizes $w_s \in \{w_1, w_2, \ldots, w_k\}$. The final decision combines evidence across all window sizes:

$$\text{Decision}_t = \bigvee_{i=1}^{k} \left(S_t^{(w_i)} > \theta_i\right) \tag{2.64}$$

Performance Trade-offs

Recursive techniques provide static storage allocation ($O(1)$ to $O(\log t)$), optimal usage of all available information and theoretical optimality characteristics, making them suitable for resource-constrained environments [80, 90]. However, they require complex implementation for sophisticated algorithms and may accumulate numerical errors over time.

Sliding window methods are characterized by simplicity, small incremental updates and space-bounded computation requirements (space complexity $O(w_s)$) [90]. Drawback is information loss in ignored data and the challenge of window size specification. Computational complexity on each update is $O(1)$ with efficient incremental updates and $O(w_s)$ re-computation of all.

Most recent implementations combine these approaches into hybrid approaches:

$$S_t = \alpha_h S_t^{recursive} + (1 - \alpha_h) S_t^{sliding}$$

where α_h balances between recursive precision and sliding window adaptability, leveraging the advantages of both paradigms while mitigating their respective limitations.

2.3.3 Tuning

Algorithm tuning is a critical phase of change point detection system design that requires careful optimization of detection metrics to achieve the desired performance characteristics. Algorithm tuning involves trading off three fundamental performance metrics: probability of detection, detection time and dynamic change estimation precision [73].

2.3.3.1 Probability of Detection

Tuning is selecting detection thresholds with maximum probability of detection p_d at achievable false alarm rates.

The probability of detection quantifies the algorithm's ability to identify true change points when they occur. It is mostly determined by the relationship between the minimal detectable change size and the detection threshold [73]. In the context of CUSUM-based algorithms, the detection probability is also related to the SNR and the threshold h, as:

$$p_d = p(\tau < \infty | \text{change occurred}) \tag{2.65}$$

The detectability condition requires that the Kullback-Leibler divergence between pre- and post-change distributions exceeds a minimum threshold:

$$D_{KL}(p_1||p_0) = \int p_1(x) \log \frac{p_1(x)}{p_0(x)} dx > h_{min} \tag{2.66}$$

2.3.3.2 Time to Detection

The detection time is the duration that elapses between the occurrence of a change point and the subsequent alarm activation. It is quantified by the expected detection delay $E[t_d - \tau]$, where t_d denotes the alarm instant and τ the true change point. The Average Run Length (ARL) function is the primary tool for analyzing the behavior of the detection delay [73]:

$$ARL_0 = E_{y_t}[\tau]. \tag{2.67}$$

The ARL function encapsulates the key performance trade-off of optimal algorithms that aim to minimize the mean detection delay while maintaining a fixed mean time between false alarms. The mean time between false alarms constrains the selection of the detection threshold:

$$E[t_d \mid H_0] \geq ARL_{\min}. \tag{2.68}$$

Proper tuning requires balancing detection speed against false alarm frequency. The detection threshold must be adjusted so that the average time between false alarms does not fall significantly below the expected time between successive changes in piecewise-constant processes.

2.3.3.3 Change Estimation Precision

Dynamic change estimation refers to an algorithm's capability to adapt to varying magnitudes of change while providing accurate estimates of both the timing and the parameters of those changes. The accuracy of these estimates depends on the choice of the minimum jump amplitude as well as on the adaptive threshold selection [73].

For algorithms such as GLR and CUSUM that yield change-time estimates, the distribution of the estimation error guides parameter selection:

$$P(\hat{\tau} = \tau \pm n), \tag{2.69}$$

where $\hat{\tau}$ denotes the estimated change time and τ the true change point.

2.3.3.4 Tuning Process

The tuning process follows in a systematic manner. It begins with an extensive examination of the problem at hand, including the identification of expected types of changes, relevant time constants and prevailing noise conditions. Based on this analysis, the most suitable detection algorithm is chosen by matching the required detection capabilities to the algorithm's strengths, taking estimation requirements and any available prior information into account. This is followed by the refinement of parameters through, for instance, average run length (ARL) analysis or using detectability indices to determine optimum thresholds, window lengths and minimum detectable change magnitudes. The detectability index provides a weaker performance measure when ARL computation becomes intractable for complex problems. Tuning parameters must account for the sensitivity of algorithms to nuisance parameters and modeling errors, prioritizing robust performance over

theoretical optimality.

Tuning requires continuous refining obtained through performance measurement through sets, which involves a recursive process that cycles back from accuracy verification to parameter adjustment, algorithm selection, or problem reformulation. Successful tuning requires maintaining a subtle balance between detecting actual changes and ignoring false alarms to ensure its applicability in real-world situations.

2.4 CPD for Multiplicative Noise

CPD of signals contaminated by multiplicative noise, such as speckle noise found in many radar and reflectometry measurements, presents unique challenges best addressed by parametric methods specifically designed. Unlike additive noise, the amplitude of multiplicative noise depends on the signal amplitude, which can vary with changes in the reflecting surface. The Gamma model, identified as a physically realistic distribution for homogeneous reflection surfaces in radar applications, forms the basis for most CPD methods addressing situations with multiplicative noise.

2.4.1 Sequential Detection under Gamma Noise

The CPD analysis under multiplicative noise is commonly addressed through parametric methods that are based on known distributions, with the Gamma distribution being the predominant choice due to its derivation from electromagnetic scattering theory. Two primary sequential detection approaches have emerged: MLE and CUSUM control charts adapted for Gamma-distributed signals.

2.4.1.1 MLE for Gamma Distribution

Maximum Likelihood Estimation is a fundamental method used in determining change points in time series data that follows a Gamma distribution [91, 92]. The key approach involves the testing of a null hypothesis that no change exists, in opposition to an alternative one that asserts a change occurs at an unknown time, τ. In independent sequences of Gamma random variables, estimators for the shape parameter (k) and scale (θ) are derived from the log-likelihood function [91, 92].

The position of the change point, $\hat{\tau}$, is estimated through optimization of the sum of log-likelihoods of segments separated by potential change points. In combined testing and estimation, this process combines the evaluation of a change with the estimation of its location. Even though Gamma's log-likelihood function is concave with a unique maximum, identifying MLE usually requires iterative optimization methods due to the non-existence of closed-form expressions [91].

In solving computational problems, efficient methods use closed-form estimators to estimate gamma log-likelihood maximization value. The Approximate Maximum Log-likelihood (AML) achieves an acceptable approximation of the actual maximum, allowing faster detection without sacrificing standards in performance of exact likelihood methods. The procedure reduces time complexity from $O(t)$ to $O(t^2)$ and, in certain cases, achieves linear complexity [91].

For multiple change points, the problem is resolved using Dynamic Programming (DP) techniques [91]. The Generalized Log-Likelihood Ratio (GLR) technique handles the case in which process characteristics are unknown, requiring parameter estimation preceding and following each potential change point.

Several research studies have examined the distribution of this maximum likelihood estimate of change points, without taking into account changes in shape parameter, only changes in scale parameter [92]. In the long term, this normalized change point is uniformly distributed, meaning that the probability of uncovering a correct position of this change point is independent of its true location [92].

2.4.1.2 CUSUM Control Charts for Gamma-Distributed Signals

CUSUM control charts are the second significant method for sequential detection in the presence of Gamma noise [93, 94]. These charts are effective in identifying moderate changes in process parameters and provide improved sensitivity compared with conventional Shewhart charts, especially for gradual parameter deviations [94].

The CUSUM algorithm operates by recursively calculating a tracking statistic that accumulates evidence for or against change. The alarm is raised

when the statistic exceeds a pre-set threshold, chosen to control the false alarm rate through the ARL under the null hypothesis.

In the case of Gamma distributions, the optimal CUSUM step is the log-likelihood ratio of the alternative to the null hypothesis [95]. The CUSUM sample path can be quantized into an approximating discrete-state Markov chain to overcome computational challenges [94].

Several adaptations of the CUSUM chart have been proposed to address different types of parameter changes in the Gamma distribution. For detecting variations in the scale parameter, specific test statistics have been developed that are asymptotically optimal in efficiency for two-sided alternatives [93].

Their performance is usually assessed in terms of ARL calculations derived via integral equation approaches, Markov chain approximations, or Monte Carlo simulations.

2.4.1.3 Specialized Sequential Methods

In addition to general MLE and CUSUM methods, there are specialized techniques for Gamma distributed signals, targeting changes in particular parameters. To detect changes in scale parameters when change points are unspecified, test statistics give asymptotically optimal likelihood ratio tests [93]. These are especially useful for detecting continuous change in the scale parameter.

For changes in the shape parameter, CUSUM-based estimators that employ logarithmic transformations have been suggested [96]. The transformation produces a sequence whose statistical characteristics are connected with the shape parameter via mathematical functions.

Sequential Probability Ratio Test (SPRT) procedures have also been derived for Gamma failure distributions in reliability applications [97]. These procedures make use of the monotone likelihood ratio property of the Gamma distribution to test hypotheses in reliability.

2.4.2 Discussion

Despite their theoretical underpinnings and application potentials, current CPD approaches for multiplicative Gamma noise have some serious drawbacks that impact their real-world performance and reliability.

2.4.2.1 Parameter Sensitivity and Estimation Challenges

The most significant limitation is high sensitivity to the precise specification of the Gamma shape and scale parameters [91, 92]. Detection performance is dependent on these parameters and in practice variations can dramatically undermine reproducibility and false alarm control. Classic approaches tend to require shape parameters to be fixed and known, restricting applicability when both parameters can change in real data streams.

The exact MLE computation for Gamma parameters involves iterative numerical optimization, which can be computationally costly for large samples or multiple change point models. Although approximate approaches offer computational ease, they come at the expense of introducing estimation errors, which can compromise precision, especially for small sample sizes.

2.4.2.2 Distribution and Model Assumptions

Analytical tractability becomes extremely limited when both shape and scale parameters vary together. Analytical solutions remain mostly limited to single parameter changes and large empirical simulations are needed for double parameter cases [92]. The assumption that underlying data follows a Gamma distribution exactly can be problematic. Methods struggle with model misspecification and even simple dependence structures in data streams. When the true distribution deviates from Gamma assumptions, detection performance can deteriorate significantly.

2.4.2.3 Operational and Performance Limitations

Threshold decisions in CUSUM applications are based solely on the assumed Gamma parameters and desired ARL performance. Threshold selection is an issue in the presence of varying true Gamma parameters in reality, as typically faced in Earth remote sensing applications whose surface conditions vary.

The false alarm probabilities can increase with the transition time and can produce premature signaling before any change occurs [98]. Chart sensitivity varies with shift magnitude and sample size and increasing the sample size typically produces capacity of more efficient detection but is computationally more costly.

Detection delay versus false alarm rate is an issue of operation. Increased sensitivity results in reduced detection delay while generating additional false alarms and capping sensitivity risks missing subtle surface transitions essential to the applications of remote sensing.

2.5 Conclusion

This chapter has presented a critical outline of the theoretical and methodological requirements of detecting change points in the case of multiplicative noise interference. The examination identifies the Gamma distribution as an appropriate statistical model for Earth remote sensing applications, based on electromagnetic scattering theory. The two-parameter Gamma model offers a proper compromise between flexibility and the mathematical tractability required for efficient implementation of the algorithm.

A systematic analysis of change point detection methodologies uncovers an extensive variety of approaches while highlighting notable gaps in multiplicative noise environments. Most available change point detection (CPD) approaches rely on assumptions about additive noise, which leads to their failure in coherent remote sensing scenarios where multiplicative speckle overpowers signal characteristics.

Maximum Likelihood Estimation and CUSUM-based methods offer theoretically sound frameworks for Gamma-distributed signals; however, they exhibit significant parameter sensitivity, computational complexity and limited robustness to model misspecification. These methods usually need very accurate information about Gamma parameters and have trouble when both shape and scale parameters change at the same time, which is the case in real Earth remote sensing applications.

These comprehensive gaps indicate the need for advanced statistical frameworks that can bridge the gap between theoretical understanding and practical

implementation. The limitations in existing methodologies, especially the lack of flexible, parameter-adaptive algorithms for multiplicative noise, directly motivate the research contributions presented in Chapter 3. This chapter addresses these fundamental challenges by developing a novel segmentation framework specifically designed for our GNSS-R reflectivity signals in the presence of speckle noise.

Chapter 3

Segmentation of the GNSS-R Reflectivity Signals affected by Speckle Noise

3.1 Introduction

The primary objective of this chapter is to introduce an automatic segmentation model specifically designed to divide airborne GNSS-R reflectivity measurements, assumed to follow a Gamma distribution, into statistically stationary segments. These segments are associated with different mean signal levels, which arise from changes in the amplitude and characteristics of the reflected GNSS signals as the sensor carrier passes over diverse terrestrial and water surfaces. To achieve robust and accurate segmentation, we propose a methodology founded on a specific statistical model of the signal, leading to a mixture of an on-line change point detection algorithm and an off-line change point localization and refinement process.

Given the continuous and often extensive nature of data collected during airborne GNSS-R campaigns, an on-line change point detection process is crucial for timely data processing. In this regard, we employ the CUSUM algorithm to detect changes in the signal characteristics as they occur. A key innovation in our application of the CUSUM algorithm stems directly from the statistical model developed for GNSS-R reflectivity, which is a central theme of this chapter. In accordance with Chapter 2 and the groundwork laid out concerning noise characteristics, we assume that the raw GNSS-R reflectivity is affected by multiplicative speckle noise. Our model assumes that the GNSS-R reflectivity follows a Gamma distribution. By applying a homomorphic log-transformation to the reflectivity data, we convert this multiplicative noise into an additive noise component with a constant

variance. This critical property allows the CUSUM detector to operate with a fixed detection threshold. This threshold is not empirically set but is statistically derived based on a desired Average Run Length (ARL(0)), which defines the expected number of observations between false alarms under the null hypothesis (no change). This fixed statistical threshold ensures a consistent and reproducible detection performance across varied datasets and environmental conditions, a significant advantage over approaches where CUSUM sensitivity might depend on fluctuating, difficult-to-estimate local noise parameters.

Once the on-line CUSUM algorithm detects a change, an off-line localization approach is implemented to determine the precise instant of the change with greater accuracy. For this, we propose a Log Gamma Maximum Likelihood Estimate (LGMLE) specifically designed for change localization within the Log-Gamma distributed, transformed reflectivity data. The MLE process identifies the most probable time of transition by evaluating the likelihood function across a working window centered around the initial CUSUM detection, thereby optimizing the accuracy of the change point localization.

In any real-world signal processing application, particularly with noisy data, the initial detection stage may yield false alarms or over-segmentation. Hence, a further merging process is included. Changes detected with a dynamic less than a minimum value are eliminated. This practical step serves to decrease the number of less significant detections, especially in regions where there is slight fluctuations that don't represent actual transitions among different land-form types.

Finally, the GNSS-R signals are segmented into stationary parts, each characterized by its unique statistical properties. A further step in our methodology involves using these properties (specifically, the mean and standard deviation of the reflectivity within each segment) for automatic classification. K-means clustering is then applied to these features to group segments of similar characteristics and facilitate differentiation and labeling of various land-forms along the flight path.

This chapter formulates the full framework for GNSS-R signal segmentation, starting with statistical modeling of reflectivity measurements and

moving through change point detection to surface differentiation. The treatment begins with the theory, deriving the Gamma distribution model for speckle noise and adding the homomorphic transform to the Log-Gamma domain. The change point detection algorithm follows, detailing both the CUSUM detection strategy and Maximum Likelihood localization techniques. The proposed methodology is tested using synthetic as well as real data, and validation with airborne GNSS-R datasets demonstrates the effectiveness of the system in differentiating diverse surface properties. The conclusion is provided in the final section.

3.2 Statistical model of the GNSS-R reflectivity

In a GNSS-R radar observation system, the reflectivity is processed as the ratio between the intensity of the reflected signal and the intensity of the direct signal from the GNSS satellite.

3.2.1 Direct signal tracking: Phase Locked Loop

After the demultiplexing stage, the direct signal processed by the PLL for each GNSS satellite can be represented as follows:

$$s_d(t) = a \sin(2\pi f_d t + \phi) + \eta \tag{3.1}$$

where a denotes the amplitude of the signal, f_d represents its Doppler and η represents an additive noise component following a centred normal distribution. To process the correlation components of the signal, the following local replicas are generated:

$$s_l(t) = \sin\left(2\pi \hat{f}_1 t + \hat{\phi}\right) \tag{3.2}$$

$$c_l(t) = \cos\left(2\pi \hat{f}_1 t + \hat{\phi}\right) \tag{3.3}$$

where $\hat{f}_1$ and $\hat{\phi}$ respectively represent the Doppler frequency and phase delay estimated by the FLL and the PLL. In its implementation, the DLL processes the following two correlation components:

$$I_d(\tau) = \int_0^{Tc} s_d(t) s_l(t+\tau) dt \tag{3.4}$$

$$Q_d(\tau) = \int_0^{Tc} s_d(t) c_l(t+\tau) dt \tag{3.5}$$

where $I_d(0)$ and $Q_d(0)$ represent the two quadrature components of correlation. Using these two correlation components, we calculate the signal intensity as:

$$T_d = I_d^2(0) + Q_d^2(0) \tag{3.6}$$

$I_d(0)$ and $Q_d(0)$ are assumed to be random variables distributed according to the following Gaussian distributions:

$$I_d(0) \sim N(a\cos(\phi_e), \sigma^2) \tag{3.7}$$
$$Q_d(0) \sim N(a\sin(\phi_e), \sigma^2) \tag{3.8}$$

where $\phi_e = \phi - \hat{\phi}$ which is close to zero. The standard deviation, σ, of the Gaussian distribution is a function of the noise power in the direct signal. The intensity, $\frac{Td}{\sigma^2}$, follows a non-central chi-squared distribution with two degrees of freedom and non-centrality parameter a^2. This leads to the expected value:

$$E(T_d) = a^2 + 2\sigma^2 \tag{3.9}$$

In our implementation, the direct signal is assumed to be stationary for large period of time, where its mean value is estimated.

3.2.2 Reflected signal tracking: Assisted open loop phase tracking

For the reflected signal, the demultiplexing stage is processed using the code delay of the direct signal, assuming the receiver's location is known. This operation is facilitated by an assisted DLL, as described in [39]. After demultiplexing, the expression for the reflected signal is given by:

$$s_r(t) = \sum_{i=1}^{L} a_i \sin(2\pi f_i t + \phi_i) + \eta \tag{3.10}$$

where a_i, f_i, and ϕ_i represent the parameters of the multiple reflections of the GNSS signal on the ground reaching the receiver antenna. The number of scatterers L is typically large for rough surfaces, ensuring that the central limit theorem conditions are satisfied for the subsequent statistical analysis.

The assisted open loop tracking of the received signal processes the two quadrature components using the following replicas:

$$s_l(t) = \sin\left(2\pi\hat{f}_1 t\right) \tag{3.11}$$

$$c_l(t) = \cos\left(2\pi\hat{f}_1 t\right) \tag{3.12}$$

where $\hat{f}_1$ is the Doppler frequency estimate by the FLL of the direct signal tracking. The two quadrature components of correlation are defined as follows:

$$I_r(\tau) = \int_0^{Tc} s_r(t)s_l(t+\tau)dt \tag{3.13}$$

$$Q_r(\tau) = \int_0^{Tc} s_r(t)c_l(t+\tau)dt \tag{3.14}$$

To process the intensity of the reflected signal, we use the component of correlation obtained for $\tau = 0$. In this context, the signal intensity, as a function of the two quadrature components, can be expressed as follows:

$$T_r = I_r^2(0) + Q_r^2(0) \tag{3.15}$$

When the GNSS signal is reflected off a rough surface, the phases ϕ_i are assumed to be random following a uniform distribution between 0 and 2π according to the Goodman model [99]. This assumption is physically justified by the random nature of surface height variations that are much larger than the wavelength, causing the reflected signals from different scatterers to have random phase relationships.

Assuming that the central limit theorem is satisfied and if the reflecting surface is rough, the components of correlation are considered as random variables and are assumed to follow a centered normal distribution:

$$I_r(0) \sim N(0, \sigma_r^2) \tag{3.16}$$

$$Q_r(0) \sim N(0, \sigma_r^2) \tag{3.17}$$

where the standard deviation σ_r is a function of the surface roughness. Physically, σ_r^2 represents the average scattered power and is directly related to the radar cross-section of the illuminated surface area. For rougher surfaces, σ_r^2 increases due to enhanced scattering, whereas for smooth surfaces, the Goodman's model no longer applies.

In our implementation, we process the mean intensity of the received signal within a working window of size N.

$$T_r = \frac{1}{N}\sum_{i=1}^{N} T = \frac{1}{N}\sum_{i=1}^{N}\left(I_r^2(0) + Q_r^2(0)\right) \tag{3.18}$$

In this implementation, as we increase N, we enhance the SNR of the intensity. However, increasing N also reduces the temporal resolution of change detection, creating a fundamental trade-off between noise reduction and detection sensitivity. In the proposed framework, N denotes the number of reflectivity samples aggregated to compute each intensity estimate and can be interpreted as an effective number of looks in the sense of multi-look processing. In our case for the GPS L1 signal, the components of correlation we processed with $Tc = 1ms$ data and a typical value of $N = 20$.

From the equations above, each individual intensity measurement T follows a chi-squared distribution with 2 degrees of freedom and scale parameter σ_r^2. When we average N independent chi-squared random variables, the sum follows a Gamma distribution. Specifically, if $X \sim X^2\left(2, \sigma_r^2\right)$, then $\sum_{i=1}^{N} X_i \sim Gamma(N, 2\sigma_r^2)$. Since T_r is the average of N such variables, we have:

$$T_r \sim Gamma\left(N, 2\sigma_r^2/N\right) \tag{3.19}$$

where $2\sigma_r^2$ is the mean value of the intensity and $4\sigma_r^4/N$ represents its variance.

3.2.3 Reflectivity of the GNSS-R signal

Previous analytical work derived an electromagnetic step-response for GNSS-R across sharp reflectivity transitions and validated its spatial-response predictions at L1 and L5 [100]. In this work, we develop a statistical model that considers a linear variation of reflectivity between two surfaces under multiplicative noise conditions.

In a GNSS-R radar system, both the direct signal from the satellite and the reflected signal obtained after the reflection of the direct signal on ground are observed. The power of the reflected signal is a function of the power of the

direct signal. To normalize the observation of the reflected signal, we use the expression of reflectivity that is given by:

$$R = \frac{T_r}{E(T_d)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(I_r^2(0) + Q_r^2(0)\right)}{(a^2 + 2\sigma^2)} \tag{3.20}$$

where, $E(T_d)$, the mean intensity of the direct signal is considered constant during the duration of the experimentation. In practice, the intensity of the direct signal changes slowly over time due to satellite and receiver motion and atmospheric effects, but these variations occur on time scales much longer than the observation window, allowing for reliable mean estimation.

In order to understand the meaning of R, the ideal coherent reflection is defined where the surface is flat and perfectly conductive, and the reflection is purely specular. The intensity of the reflected signal is maximum and its mean value $E(T_r)$ tends toward $E(T_d)$. Then we have:

$$E(R) \quad = \quad \frac{E(T_r)}{E(T_d)} \approx 1 \tag{3.21}$$

For the incoherent reflections, where we have multiple reflections of the direct signal due to the roughness of the surface, the mean value of the reflectivity is defined as:

$$E(R) = \frac{E(T_r)}{E(T_d)} = \frac{2\sigma_r^2}{(a^2 + 2\sigma^2)} \tag{3.22}$$

In this case, the intensity of the reflected signal is dispersed and $E(T_r) < E(T_d)$. Finally, we have:

$$0 < E(R) < 1 \tag{3.23}$$

3.2.4 Homomorphic transformation of the signal

From the mathematical derivation above, in the case of non-coherent reflectivity R follows a Gamma distribution as a direct consequence of the signal processing chain and the statistical properties of the correlation components [23, 99]. In this case, the noise on the observations is a multiplicative speckle noise. Let:

$$R(N, \lambda) \quad \sim \quad Gamma\,(N, \lambda/N) \tag{3.24}$$

with:

$$E(R(N,\lambda)) = \lambda = \frac{2\sigma_r^2}{(a^2 + 2\sigma^2)} \tag{3.25}$$

The multiplicative nature of speckle noise presents fundamental challenges for change point detection algorithms designed for additive noise scenarios. In multiplicative noise environments, the noise power scales with the signal amplitude, violating the constant variance assumption underlying most classical detection methods. This signal-dependent noise characteristic necessitates specific processing approaches.

In the case of a multiplicative model, the noise power is a function of the signal amplitude that varies in our application with the surface of reflection. The homomorphic log-operation transforms the multiplicative noise into additive noise, with constant noise power. The offset log-transform is defined by:

$$W(N,\lambda) \quad = \quad \log\left(R(N,\lambda)\right) \tag{3.26}$$

with:

$$W(N,\lambda) \quad \sim \quad Log - Gamma\left(N, \lambda/N\right) \tag{3.27}$$

where $V(W(N,\lambda))$, the variance of $W(N,\lambda)$, is constant. We have the following expressions:

$$E(W(N,\lambda)) \quad = \quad \Psi(N) + \log(\lambda/N) \tag{3.28}$$
$$V(W(N,\lambda)) \quad = \quad \Psi^{(1)}(N) \tag{3.29}$$

where $\Psi(\dots)$ and $\Psi^{(1)}(\dots)$ are respectively the digamma function and the trigamma function [101, 102].

The logarithmic transformation is important as it transforms the multiplicative noise model into an additive one, while maintaining the relative variations in the signal. This transformation facilitates the application of established additive noise processing techniques while preserving the essential statistical features needed for precise change detection.

3.3 Change point detection

Building on the statistical model outlined in the previous section, this section presents the change point detection approach specifically designed for GNSS-R reflectivity. One of the main features of this approach is that it is based on the Log-Gamma distribution, which has a constant variance, making it easy to develop optimal detection algorithms.

In our radar application, a large amount of data is available, and we need to process these data sequentially as a chronological series. This series is considered to be in control when the statistical parameters of the process are stationary. In contrast, it is considered out of control in the non-stationary case when there is a change in the process.

The problem of change point detection in GNSS-R applications is particularly difficult given the possibility for continuous transitions from one type of surface to another. In contrast to instantaneously occurring abrupt changes, real-world boundary transitions tend to be characterized by smooth variations as the satellite footprint traverses boundary areas. This calls for a higher-level hypothesis testing framework that can handle both abrupt and continuous change possibilities. In this context, the hypothesis test for the change point detection can be defined as follows:

$$\begin{aligned} H_0 &: R_t \sim Gamma\left(N, \frac{\lambda_0}{N}\right), \quad \forall t \in \{1, \ldots, n\} \\ H_1 &: R_t \sim \begin{cases} Gamma\left(N, \frac{\lambda_0}{N}\right), & 1 \leq t \leq \tau_1 \\ Gamma\left(N, \frac{\lambda_t}{N}\right), & \tau_1 < t < \tau_2 \\ Gamma\left(N, \frac{\lambda_1}{N}\right), & \tau_2 \leq t \leq n \end{cases} \end{aligned} \tag{3.30}$$

where λ_t denotes the mean of the reflectivity process during the transition, defined as:

$$\lambda_t = \lambda_0 + (\lambda_1 - \lambda_0) \cdot \frac{t - \tau_1}{\tau_2 - \tau_1}, \quad \tau_1 < t < \tau_2 \tag{3.31}$$

This expression accounts for the real situation in which surface transitions happen across several observation samples and not instantaneously. The parameters τ_1 and τ_2 define the start and end of the transition interval, and λ_t is the time-dependent mean reflectivity within this duration. This model would be especially applicable to airborne GNSS-R scenarios in which the

size of the footprint and platform speed determine the boundary crossing durations.

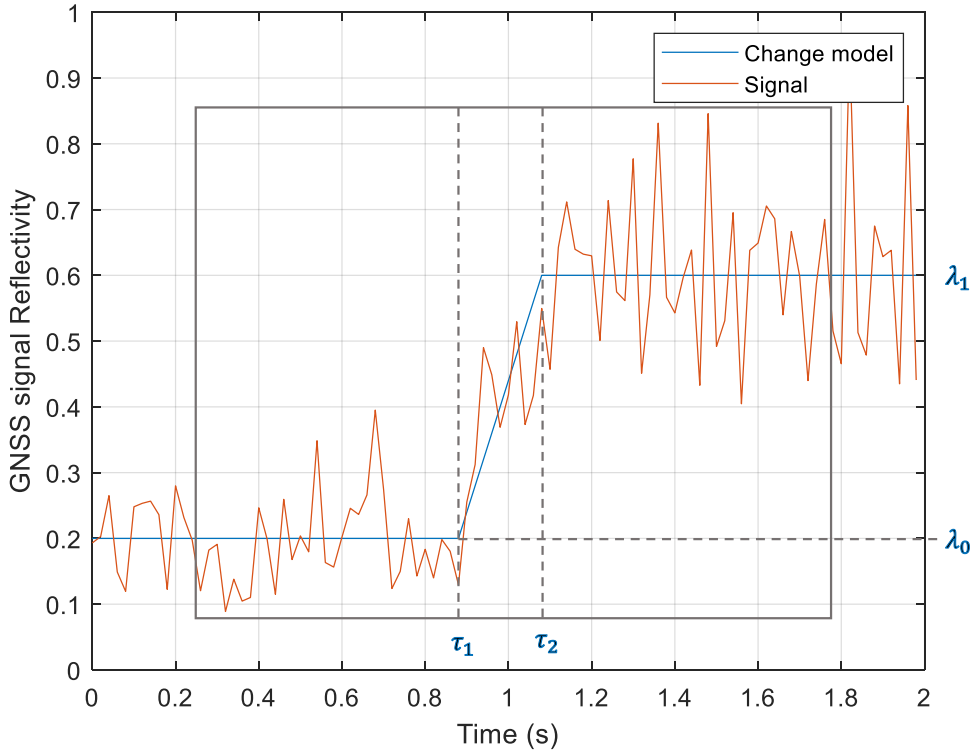


FIGURE 3.1: Signal model in a working window.

Figure A.1 represents the signal and its parameters under H_τ in the working window N. We can observe the increasing evolution of the signal model between the mean values λ_0 and λ_1 associated to the satellite footprint displacement from one area to another.

3.3.1 Proposed change point detector

The main objective of the proposed detector is to detect a change in the observed process. Figure 3.2 illustrates the flow chart of the proposed online/off-line change point detector. As shown in this figure, the localization of the change is processed in a second step within a working window centered on τ', the detected instant of change.

The filter can be defined as a recursive mean estimate derived from the log-transformation of the reflectivity observations:

$$\bar{w}_t = \bar{w}_{t-1} + \frac{1}{K}\left(\log(r_t) - \bar{w}_{t-1}\right) \tag{3.32}$$

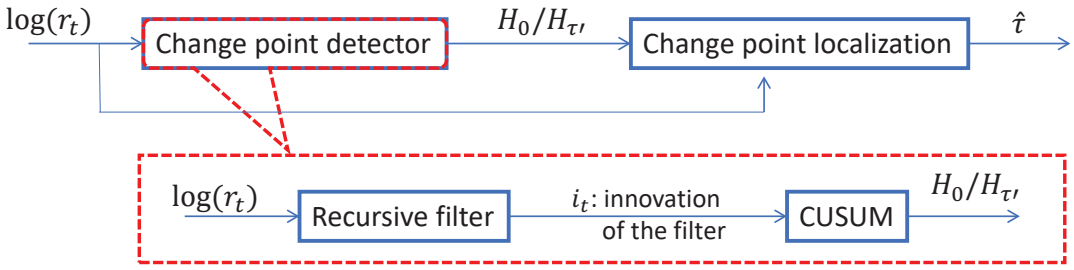


FIGURE 3.2: Change detection system.

where K is the size of the window used to estimate the mean $\bar{w}_t$. In practice, K is a parameter that increases with time and removes information provided by the observation r_t. This estimate can be expressed as a weighted sum:

$$\begin{aligned} &\bar{w}_t = \alpha_{1,t}\, \bar{w}_{t-1} + \alpha_{2,t}\, \log(r_t) \\ &\text{with } \alpha_{1,t} + \alpha_{2,t} = 1 \\ &\text{and } \alpha_{1,t} = 1 - \frac{1}{K},\ \ \alpha_{2,t} = \frac{1}{K} \end{aligned} \tag{3.33}$$

This can be rewritten as:

$$\bar{w}_t = \bar{w}_{t-1} + \alpha_{2,t}\, i_t \tag{3.34}$$

$$\text{with } i_t = \log(r_t) - \bar{w}_{t-1} \tag{3.35}$$

where i_t is the innovation of the filter. In the framework of optimal adaptive filtering [103], the optimal expression of $\alpha_{2,t}$ can be derived in the following recursive form :

$$P_t = \frac{(P_{t-1} + Q)\, \Psi^{(1)}(N)}{P_{t-1} + Q + \Psi^{(1)}(N)} \tag{3.36}$$

$$\alpha_{2,t} = \frac{P_t}{\Psi^{(1)}(N)} \tag{3.37}$$

where Q is a parameter defined by the user, and P_t the variance of the estimate $\bar{w}_t$. In this recursive form, $\alpha_{2,t}$ decreases with time toward a constant value, and the smoothing of the mean estimate increases until a constant smoothing level.

We can define the normalized innovation as:

$$\tilde{i}_t = \frac{i_t}{P_{t-1} + \Psi^{(1)}(N)} \tag{3.38}$$

and the CUSUM detector in its simple form:

$$g_t^+ = \left(g_{t-}^+ + \tilde{i}_t\right)^+ \tag{3.39}$$
$$g_t^- = \left(g_{t-}^- - \tilde{i}_t\right)^+ \tag{3.40}$$
$$\tau' = \min\left\{t : (g_t^+ \geq h) \cup (g_t^- \geq h)\right\} \tag{3.41}$$

where $(X)^+ = max(X, 0)$. In the case of an innovation process, g_t^+ and g_t^-, the integration of the innovation process, evolves as a random walk before the change. After the change instant $\tau + t$, g_t^+ and g_t^- are monotonic increasing function.

The threshold h, fixed by the user, defines the probability of false alarm under H_0. However, the threshold also determines the delay between the instant where the change occurred and the instant an alarm is raised under H_τ. In our approach, this delay aims to be corrected by the proposed change point localization process described in the next section.

3.3.2 Proposed change point localization approach

Let the observed samples be $w_t = \log(r_t)$, where r_t are the reflectivity values following a Gamma distribution. After applying the homomorphic transformation, we assume that the samples follow a Log-Gamma distribution:

$$f(w_t; N, \lambda_t/N) = \frac{\left(e^{w_t}\right)^N \cdot e^{-\frac{Ne^{w_t}}{\lambda_t}}}{\left(\frac{\lambda_t}{N}\right)^N \Gamma(N)} \tag{3.42}$$

where λ_t denotes the mean of the underlying Gamma-distributed reflectivity and is varying with time.

In this gradual change scenario, the interval $[1, n]$ is divided into three segments:

- For $t \in [1, \tau_1]$, the process is stationary with mean λ_0.
- For $t \in (\tau_1, \tau_2)$, the mean evolves linearly from λ_0 to λ_1.

- For $t \in [\tau_2, n]$, the process stabilizes again with mean λ_1.

During the transition interval (τ_1, τ_2), we define the time-varying mean λ_t as a linear interpolation in eq. 3.31.

The total log-likelihood of the window $w_{1:n} = (w_1, \ldots, w_n)$ is expressed by summing the contributions of the three segments:

Before the change ($t \in [1, \tau_1]$):

$$\log L_1 = \sum_{t=1}^{\tau_1} \left[N w_t - \frac{N e^{w_t}}{\lambda_0} - N \log \left(\frac{\lambda_0}{N} \right) - \log \Gamma(N) \right] \tag{3.43}$$

During the change ($t \in (\tau_1, \tau_2)$):

$$\log L_2 = \sum_{t=\tau_1+1}^{\tau_2-1} \left[N w_t - \frac{N e^{w_t}}{\lambda_t} - N \log \left(\frac{\lambda_t}{N} \right) - \log \Gamma(N) \right] \tag{3.44}$$

After the change ($t \in [\tau_2, n]$):

$$\log L_3 = \sum_{t=\tau_2}^{n} \left[N w_t - \frac{N e^{w_t}}{\lambda_1} - N \log \left(\frac{\lambda_1}{N} \right) - \log \Gamma(N) \right] \tag{3.45}$$

The total log-likelihood is then given by:

$$\log L_{\text{total}}(\tau_1, \tau_2, \lambda_0, \lambda_1) = \log L_1 + \log L_2 + \log L_3 \tag{3.46}$$

To estimate the gradual change location, we search for the optimal transition boundaries $(\hat{\tau}_1, \hat{\tau}_2)$ and corresponding mean values $(\hat{\lambda}_0, \hat{\lambda}_1)$ that maximize the total likelihood:

$$(\hat{\tau}_1, \hat{\tau}_2) = \arg\max_{\tau_1, \tau_2} \log L_{\text{total}}(\tau_1, \tau_2, \hat{\lambda}_0, \hat{\lambda}_1) \tag{3.47}$$

We estimate λ_0 and λ_1 independently using the segments $t \in [1, \tau_1]$ and $t \in [\tau_2, n]$, respectively, according to the Log-Gamma distribution. The log-likelihood for m samples $w_1, \ldots, w_m$ is:

$$\mathcal{L}(\lambda) = \sum_{t=1}^{m} \log f(w_t; N, \lambda/N) \tag{3.48}$$

$$= N \sum_{t=1}^{m} w_t - \frac{N}{\lambda} \sum_{t=1}^{m} e^{w_t} - mN \log\left(\frac{\lambda}{N}\right) - m \log \Gamma(N) \tag{3.49}$$

Taking the derivative with respect to λ and setting it to zero:

$$\frac{d\mathcal{L}}{d\lambda} = \frac{N}{\lambda^2} \sum_{t=1}^{m} e^{w_t} - \frac{mN}{\lambda} = 0 \tag{3.50}$$

Multiplying both sides by λ^2/N:

$$\sum_{t=1}^{m} e^{w_t} = m\lambda \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{m} \sum_{t=1}^{m} e^{w_t} \tag{3.51}$$

A more precise estimate leverages the known expectation of w_t under the Log-Gamma model:

$$\mathbb{E}[w_t] = \Psi(N) + \log\left(\frac{\lambda}{N}\right) \tag{3.52}$$

Approximating the expectation using the sample mean $\bar{w}$:

$$\log\left(\frac{\lambda}{N}\right) = \bar{w} - \Psi(N) \tag{3.53}$$

$$\Rightarrow \quad \hat{\lambda} = N \cdot e^{\bar{w} - \Psi(N)} \tag{3.54}$$

Hence, the LGMLE are:

$$\bar{w}_0 = \frac{1}{\tau_1} \sum_{t=1}^{\tau_1} w_t, \quad \hat{\lambda}_0 = N \cdot e^{\bar{w}_0 - \Psi(N)} \tag{3.55}$$

$$\bar{w}_1 = \frac{1}{n - \tau_2 + 1} \sum_{t=\tau_2}^{n} w_t, \quad \hat{\lambda}_1 = N \cdot e^{\bar{w}_1 - \Psi(N)} \tag{3.56}$$

3.3.3 Parameter definitions

To ensure the reproducibility of the signals segmentation, the parameters of both the change point detector and change point estimate must be independent of the parameters and of the direct signal intensity (a and σ) and the reflected signal intensity (σ_r).

For the change point detector, we suggest fixing the threshold h as a function of the probability of false alarm rate. In change point detection, the false alarm rate is characterized by the Average Run Length (ARL) under H_0, denoted as ARL(0). ARL(0) is the expected number of observations between false alarms in a sequence of observations that represents noise without a change point. The value of ARL(0) is determined by simulation with a noise power $\sqrt{\Psi^{(1)}(N)}$ under H_0. The noise power defined by N is independent of the direct and reflected signal intensity.

The accuracy of the change point estimate relies only on the precision of estimating $\overline{W}$. This precision is determined by the noise power $\sqrt{\Psi^{(1)}(N)}$ and remains independent of the direct and reflected signal intensity.

3.3.4 Post-Localization Segment Merging Process

Upon change point detection and localization, the reflectivity signal after segmenting is typically over-segmented, particularly when the detection threshold is chosen such that the process is kept very sensitive to detecting changes. This over-segmentation occurs as adjacent segments that are statistically indistinguishable have been separated due to minor fluctuations or noise-induced false detections. To address this issue, we propose a systematic merging approach built on statistical hypothesis testing, specifically designed for Gamma-distributed reflectivity data.

3.3.4.1 Motivation and Statistical Framework

The merging process attempts to combine adjacent segments originating from the same underlying statistical population and preserve segments that represent true transitions between different surface types. Because the GNSS-R reflectivity is Gamma distributed as established in Section 3.2, the merging problem is framed as a statistical hypothesis test between the mean parameters of adjacent segments.

For two consecutive segments S_i and S_{i+1} with n_i and n_{i+1} observations respectively, the following hypothesis test is considered:

$$H_0 : \lambda_i = \lambda_{i+1} \quad \text{(segments originate from the same population)} \tag{3.57}$$

$$H_1 : \lambda_i \neq \lambda_{i+1} \quad \text{(segments represent different surface types)} \tag{3.58}$$

where λ_i and λ_{i+1} are the mean parameters of the Gamma distributions $\Gamma(N, \lambda_i/N)$ and $\Gamma(N, \lambda_{i+1}/N)$ for the respective segments.

3.3.4.2 Wald Test for Gamma Distribution Parameters

We apply the Wald test, a classical statistical approach specifically designed for testing parameters in exponential family distributions of which the Gamma distribution is one. The Wald test takes advantage of the asymptotic normality of MLE in order to construct a test statistic under the null hypothesis.

Parameter Estimation For each segment with reflectivity observations $R_{i,1}$, $R_{i,2}, \ldots, R_{i,n_i}$ following Gamma$(N, \lambda_i/N)$, the MLE for the mean parameter is:

$$\hat{\lambda}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{i,j} = \bar{R}_i \tag{3.59}$$

This estimator has the asymptotic property:

$$\hat{\lambda}_i \sim \mathcal{N}\left(\lambda_i, \frac{\lambda_i^2}{N \cdot n_i}\right) \quad \text{for large } n_i \tag{3.60}$$

Test Statistic Construction The Wald test statistic is formulated according to the standardized difference between the estimates of the parameters:

$$W^T = \frac{(\hat{\lambda}_i - \hat{\lambda}_{i+1})^2}{\text{Var}(\hat{\lambda}_i) + \text{Var}(\hat{\lambda}_{i+1})} \tag{3.61}$$

Substituting the asymptotic variances and the MLEs:

$$W^T = \frac{(\bar{R}_i - \bar{R}_{i+1})^2}{\frac{\bar{R}_i^2}{N \cdot n_i} + \frac{\bar{R}_{i+1}^2}{N \cdot n_{i+1}}} \tag{3.62}$$

Under the null hypothesis $H_0 : \lambda_i = \lambda_{i+1}$, the test statistic W^T has asymptotically a chi-squared distribution with one degree of freedom:

$$W^T \sim \chi^2(1) \quad \text{under } H_0 \tag{3.63}$$

3.3.4.3 Decision Rule and Merging Criterion

The statistical decision between combining adjacent segments is based on comparing the computed Wald statistic with the critical value from the chi-squared distribution. For a significance level α (usually $\alpha = 0.05$), the decision criterion is:

- **Merge segments** if $W^T < \chi^2_{1,\alpha}$ (fail to reject H_0)
- **Keep segments separated** if $W^T \geq \chi^2_{1,\alpha}$ (reject H_0)

For $\alpha = 0.05$, the critical value is $\chi^2_{1,0.05} = 3.841$, this means that segments are merged when the Wald statistic is below this threshold value.

The physical interpretation of this criterion is that segments are combined when their average reflectivity levels are not statistically significantly different, taking into account the natural variability expected from the Gamma distribution model and the segment lengths, respectively.

3.3.4.4 Minimum Segment Length Constraint

In order to guarantee the validity of the asymptotic approximation assumed by the Wald test, we enforce a minimal segment length requirement. Segments with a length less than a predefined cutoff l_{min} are merged with their neighboring segment automatically, independent of the result of the statistical test. This constraint serves multiple purposes:

1. **Statistical Validity**: Provides a sufficient sample size for the asymptotic normality assumption
2. **Computational Stability**: Avoids numerical instabilities in variance calculations

The minimum segment length is given as:

$$l_{\min} = \max(2N, 10) \tag{3.64}$$

where N is the number of samples used in the averaging process during reflectivity computation. This selection provides for at least $2N$ observations to achieve reliable parameter estimation and yet keeps a practical minimum of 10 samples.

3.3.4.5 Sequential Merging Algorithm

The Algorithm 1 of merging is performed as a sequential process that works with adjacent segment pairs from left to right across the time series. The process is designed to accommodate cascading merges, where merging two segments can establish conditions for the next merges with neighboring segments.

3.3.4.6 Statistical Properties and Performance Characteristics

The Wald test-based merging approach possesses some appealing statistical properties:

Type I Error Control: α determines directly how often segments mistakenly are merged that actually reflect truly distinct types of surfaces. By setting $\alpha = 0.05$, such errors occur at most 5% of the time under optimal assumptions.

Power Characteristics: Statistical power of the test to distinguish segments that should be separated correctly is a function of actual difference in means of reflectivity $|\lambda_i - \lambda_{i+1}|$, lengths of segments n_i and n_{i+1}, and N shape parameter. Larger differences, longer segments, and larger values of N all increase greater statistical power.

Asymptotic Efficiency: With long segment lengths, the Wald test is asymptotically efficient, i.e., it requires small sample sizes to achieve desired detection performance compared to other possible tests.

3.4 Experimentation:

3.4.1 Assessment on Synthetic Data

To assess the proposed LGMLE method, a comparative simulation with the Maximum Log likelihood Localization Estimate (MLLE) [13] is performed under a multiplicative noise scenario. The evaluation is based on 1000 Monte Carlo realizations per test configuration, using simulated GNSS-R reflectivity signals generated according to the Gamma distribution model described in Section 3.2.4.

Data: reflectivity observations R_t, change points localized $\hat{\tau}^e$
Result: change points after merging $\tilde{\tau}^e$, segmentation m
Algorithm

———-*Initialization*———-

set values for : $\alpha = 0.05, l_{min} = \max(2N, 10), \chi^2_{1,\alpha} = 3.841; i = 1; j = 1;$
$merged = true$

———-*Implementation*———-

while $merged = true$ **do**
  $merged = false;$
  $m_i = mean(R_{\hat{\tau}^e_{i-1}} \dots R_{\hat{\tau}^e_i});$
  $m_{i+1} = mean(R_{\hat{\tau}^e_i} \dots R_{\hat{\tau}^e_{i+1}});$
  $n^f_i = length(R_{\hat{\tau}^e_{i-1}} \dots R_{\hat{\tau}^e_i});$
  $n^f_{i+1} = length(R_{\hat{\tau}^e_i} \dots R_{\hat{\tau}^e_{i+1}});$
  if $n^f_i < l_{min}$ ***or*** $n^f_{i+1} < l_{min}$ **then**
    ————————*Automatic Merging (Short Segments)*————————
    $m_i = m_{i+1} = mean(R_{\hat{\tau}^e_{i-1}} \dots R_{\hat{\tau}^e_{i+1}});;$ `/* Merge */`
    $merged = true;$
  else
    ————————*Wald Test Statistical Decision*————————
    $\hat{\lambda}_i = m_i; \quad \hat{\lambda}_{i+1} = m_{i+1};$
    $Var_i = \frac{\hat{\lambda}^2_i}{N \cdot n^f_i}; \quad Var_{i+1} = \frac{\hat{\lambda}^2_{i+1}}{N \cdot n^f_{i+1}};$
    $WT = \frac{(\hat{\lambda}_i - \hat{\lambda}_{i+1})^2}{Var_i + Var_{i+1}};;$ `/* Wald Test Statistic */`
    if $WT < \chi^2_{1,\alpha}$ **then**
      ————————*Statistical Merging (H_0 not rejected)*————————
      $m_i = m_{i+1} = mean(R_{\hat{\tau}^e_{i-1}} \dots R_{\hat{\tau}^e_{i+1}});;$ `/* Merge */`
      $merged = true;$
    else
      ————————*Keep Segments Separate (H_0 rejected)*————————
      $\tilde{\tau}^e_j = \hat{\tau}^e_i;$
      $j = j + 1;$
    end
  end
  $i = i + 1;$
end

Algorithm 1: Wald Test-Based Segment Merging Algorithm

The test signals are divided into two segments: the first segment has a fixed low reflectivity mean $\lambda_0 = 0.1038$, while the second segment has varying reflectivity mean λ_1 ranging from 0.1226 to 0.4722. A smooth transition of 20 samples in duration (corresponding to approximately 0.4 s) is implemented between the segments to model realistic boundary-crossing scenarios. The true change point is positioned at t = 1.0 s.

A critical characteristic of multiplicative speckle noise is that its variance is proportional to the signal power. When transitioning from low- to high-reflectivity surfaces, the noise variance in the second segment increases proportionally. This signal-dependent noise presents a fundamental challenge for the MLLE approach, which assumes constant noise variance across the entire signal.

TABLE 3.1: Statistical parameters of different change point detectors as a function of the reflectivity mean of first and second segments (λ_0 & λ_1). The parameters are: the mean, the standard deviation and the number of detected changes (*Nbr*).

Reflectivity		**LGMLE**			**MLLE**		
λ_0	λ_1	**Mean**	**Std**	**Nbr**	**Mean**	**Std**	**Nbr**
0.1038	0.1226	1.142	0.322	1358	1.176	0.280	1364
0.1038	0.1543	1.070	0.203	1188	1.105	0.127	1192
0.1038	0.1883	1.039	0.124	1084	1.069	0.080	1102
0.1038	0.2277	1.024	0.085	1036	1.057	0.060	1045
0.1038	0.2685	1.020	0.045	1003	1.053	0.046	1004
0.1038	0.3141	1.015	0.043	1003	1.050	0.039	1002
0.1038	0.3625	1.012	0.037	1000	1.042	0.037	1001
0.1038	0.4143	1.009	0.031	1000	1.043	0.033	1000
0.1038	0.4722	1.003	0.031	1000	1.047	0.032	1000

The results in table 3.1 reveal two distinct performance regimes. When the jump is low, there were many false alarms. Thresholds are fixed in terms of false alarm rate.

For low-change dynamics, where the reflectivity ratio between segments is relatively low, MLLE achieves a lower standard deviation (0.280 s to 0.060 s) compared to LGMLE (0.322 s to 0.085 s). However, MLLE consistently shows a higher detection delay, with mean estimates ranging from 1.176 s to 1.057 s, compared with the actual change at 1.0 s.

For high-change dynamics, both estimators achieve similar standard deviations (on the order of 0.03–0.04 s). LGMLE demonstrates superior detection timing, with mean estimates very close to the actual change point (1.003 s), while MLLE exhibits a consistent bias of around 40–50 ms.

The performance difference is attributed to the fundamental noise model employed. The proposed LGMLE approach operates on log-transformed reflectivity, where the multiplicative noise becomes additive with constant variance $\Psi^{(1)}(N)$, regardless of signal power. Conversely, MLLE directly processes the Gamma-distributed data, encountering increasing noise variance as reflectivity increases. This variance instability primarily affects MLLE's performance in high-reflectivity cases.

These results demonstrate that proper modeling of multiplicative noise through the Log-Gamma framework provides more consistent change-point localization across varying reflectivity conditions, especially in scenarios where surface reflectivity changes from one surface to another.

3.4.2 Application to Real Airborne GNSS-R data

A detailed description of the airborne experiment is provided in Chapter 1. In this section, figures present the reflectivity measurements acquired using our GNSS-R receiver mounted on a gyrocopter. Reflectivity tends to increase when signals are reflected off water content.

3.4.2.1 Discussion of the log Transform Noise Model

One of the main contributions of this chapter is to define a detection threshold that can be applied seamlessly to different sets of data. In our approach we assume that the log-transform of the reflectivity is distributed according to a Log-Gamma distribution. We can then define a detection threshold that is completely independent of the SNR because the standard deviation of the additive noise on the signal is fixed and known. In order to verify this assumption, we estimate the noise power on a set of real airborne GNSS-R reflectivity data as depicted in Figure 3.3.

Figure 3.3 shows the variance of the log-transformed reflectivity estimated over N samples processed for the trace of satellite PRN 5. We can observe that the estimated variance is noisy. In this data, there are two distinct reflecting

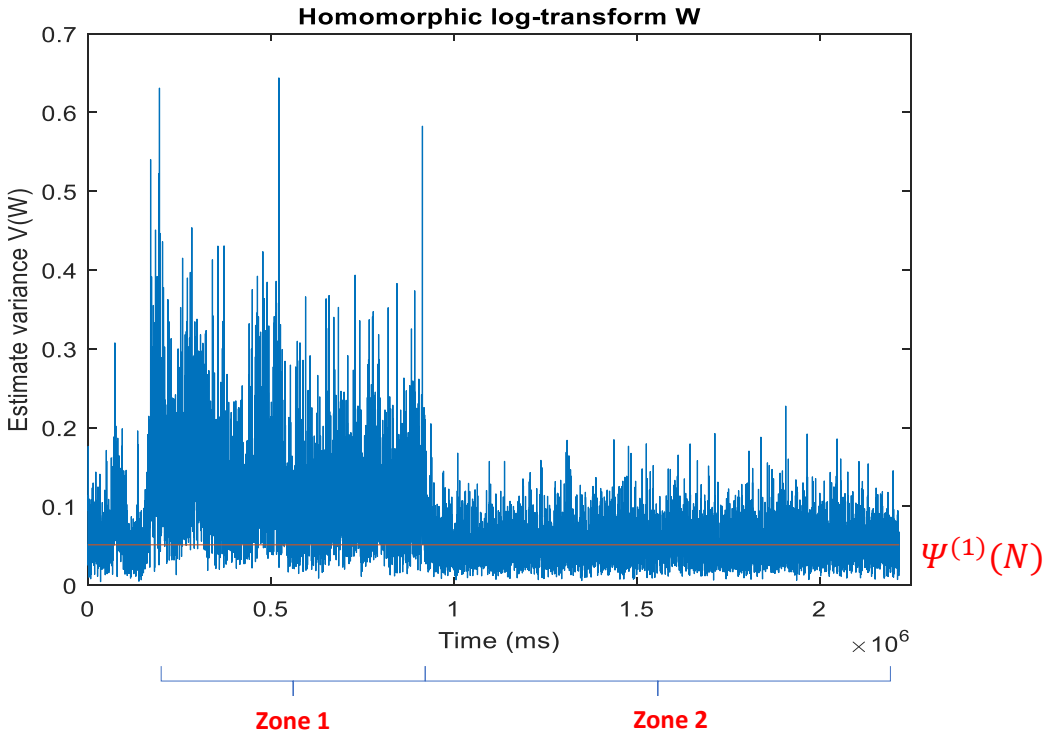


FIGURE 3.3: Variance estimate of the Homomorphic log-transform W

regimes. Zone 1, which represents the sand and sea regions in the map (figure 1.2), exhibits dominant-path scattering where one or a few reflection components have significantly higher power than multiple weaker multipath contributions; in this case, the Goodman model does not apply. Zone 2 (plain land) exhibits distributed multipath scattering, where many reflection paths contribute with destructive phase interference.

We show as a red line the theoretical variance of the log-transform of the reflectivity. This figure shows that this theoretical value is in accordance with the mean variance in the Zone 2. However, in Zone 1, the theoretical value underestimates the true value of the mean variance. In this case, the statistical model of the 1st Zone's dominant-path scattering regime differs from the Gamma distribution.

The Gamma distribution model, which is obtained in the assumption of numerous equally powered scatterers, properly describes Zone 2, in which distributed multipath scattering is prevalent. Under these conditions, the central limit theorem requirements are met: many independent scatterers

having similar amplitudes and phases uniformly distributed generate Rayleigh-distributed in-phase and quadrature components, which result in Gamma-distributed intensity measurements when a multi-look average is applied.

The threshold is fixed for the land region where diffuse reflections exist. However, in areas characterized by coherent reflections from water surfaces, false alarm detection may increase due to the increase in the standard deviation of that specific area.

However, since the main interest is detecting transitions between different land surface types, where distributed multipath scattering dominates, this limitation does not compromise the segmentation performance. The approach successfully identifies major surface transitions, and the merging process (Section 3.3.4) removes false detections.

3.4.2.2 Change Point estimation

In Figure 3.4, we present the airborne reflectivity measurements acquired during our flight experimentation for two water bodies. The GNSS signal reflectivity variations, as shown in the following graphs, are detected and processed by a series of change detection algorithms.

Initially, the CUSUM algorithm detects changes in the reflectivity of the GNSS signal. In the CUSUM implementation, ARL(0) is fixed to 3000 for N=20 and the smoothing parameter $Q = 0.001$. In order to decrease the number of false alarms in the area of specular reflection the changes with a dynamic below 0.01 are removed. Following the CUSUM detection, our proposed LGMLE approach is applied to localize the changes as shown in Figure 3.4. Finally, the signal is segmented into different sections associated to the change in the mean reflectivity levels based on the localized change point positions in Figure 3.4.

3.4.2.3 Reflectivity Signal Segmentation

We apply our radar signal segmentation technique for in-land water body detection. Figure 3.5 represent the radar signal segmentation for airborne GNSS measurements obtained along the trajectory of the flight. The Google Earth images of these figures illustrate the traces of the satellite footprints

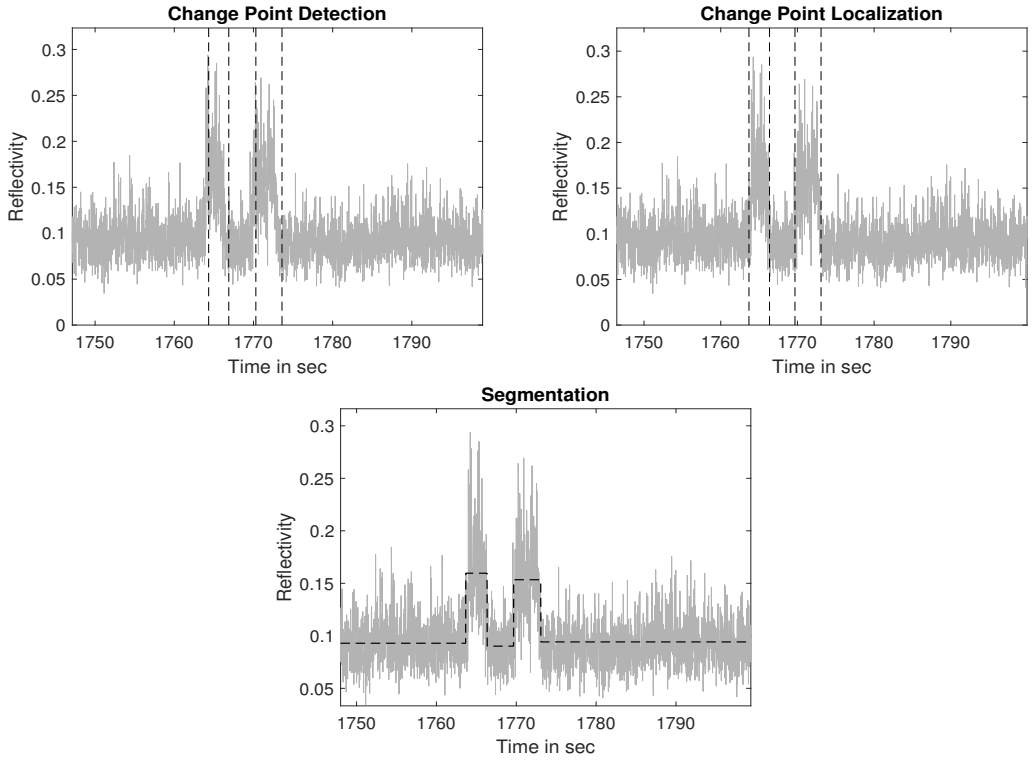


FIGURE 3.4: Detection, Localization and Segmentation Process of an airborne GNSS signal reflectivity at (50.876064°, 1.948393°)

represented by 20 *ms* localization of the specular points of reflection. It is color-coded —yellow denoting land and —blue indicating water bodies. These colors are associated to different mean reflectivity measurements. We also present the GNSS reflectivity measurements associated to each satellite signal and the corresponding signal segmentation with our proposed radar technique.

We analyze the GNSS measurements obtained from three different satellites (PRN 5, 7, and 30) of flight 1 and two satellites (PRN 8 and 21) of Flight 2, each characterized by a different elevation angle. The surface reflectivity exhibits a direct correlation with the water content in land. The reflectivity measurements increase when the signals are reflected from a water body. In addition, the variance of the noise increases with the mean reflectivity as it is supposed to be in multiplicative noise.

We observe that the radar technique detects in-land water bodies with various sizes and shapes under different environments. We also show our system's capacity to use the whole data set collected from the flight to automatically differentiate surfaces.

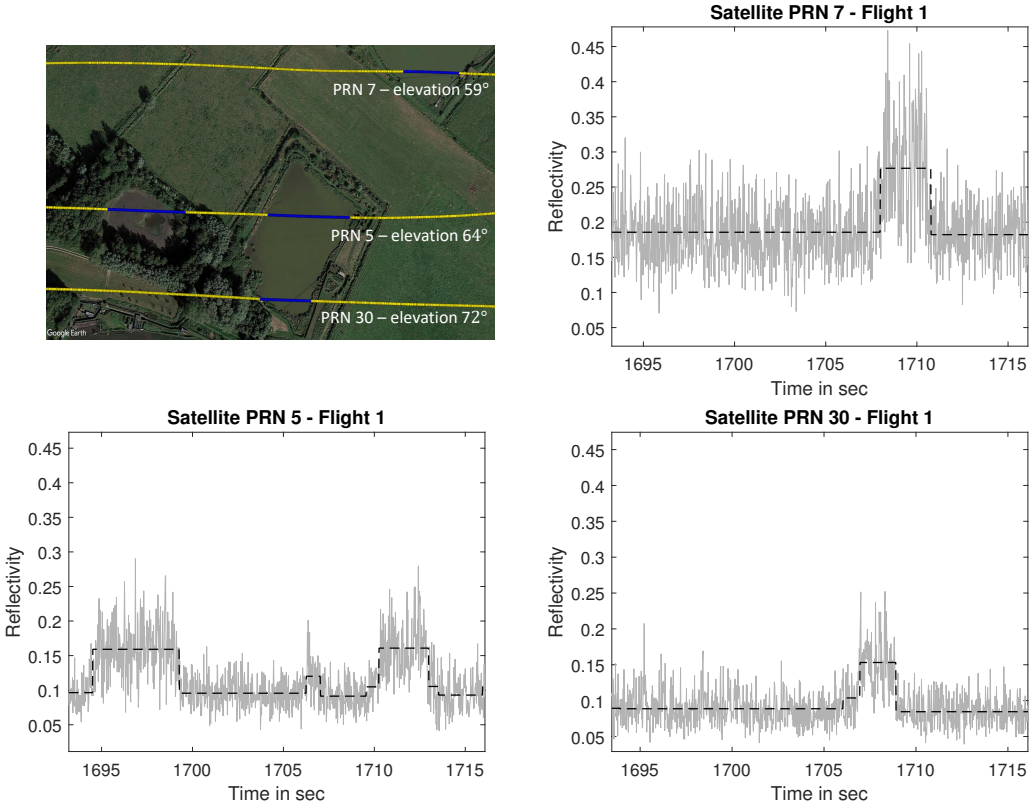


FIGURE 3.5: Segmentation of the GNSS measurements at (50.877390°, 1.933011°) – **Flight 1**

3.4.2.4 Surface Classification of Segmented Reflectivity Data

Following the change point detection and localization procedures, the segmented reflectivity signals require classification to associate each segment with meaningful surface types. The K-means clustering approach provides an unsupervised framework for automatic surface classification based on the statistical characteristics of the detected segments.

The statistical properties, i.e. the mean and standard deviation, of each segment are used to obtain a general differentiation of the signal using the K-means clustering technique. The choice of these two features (mean and standard deviation) is motivated by their direct relationship to the underlying surface scattering physics: the mean reflectivity indicates the average scattering strength of the surface, while the standard deviation captures the variability associated with surface roughness and coherence properties.

The K-means approach is an effective unsupervised machine learning method that clusters data points based on the similarity of their features.

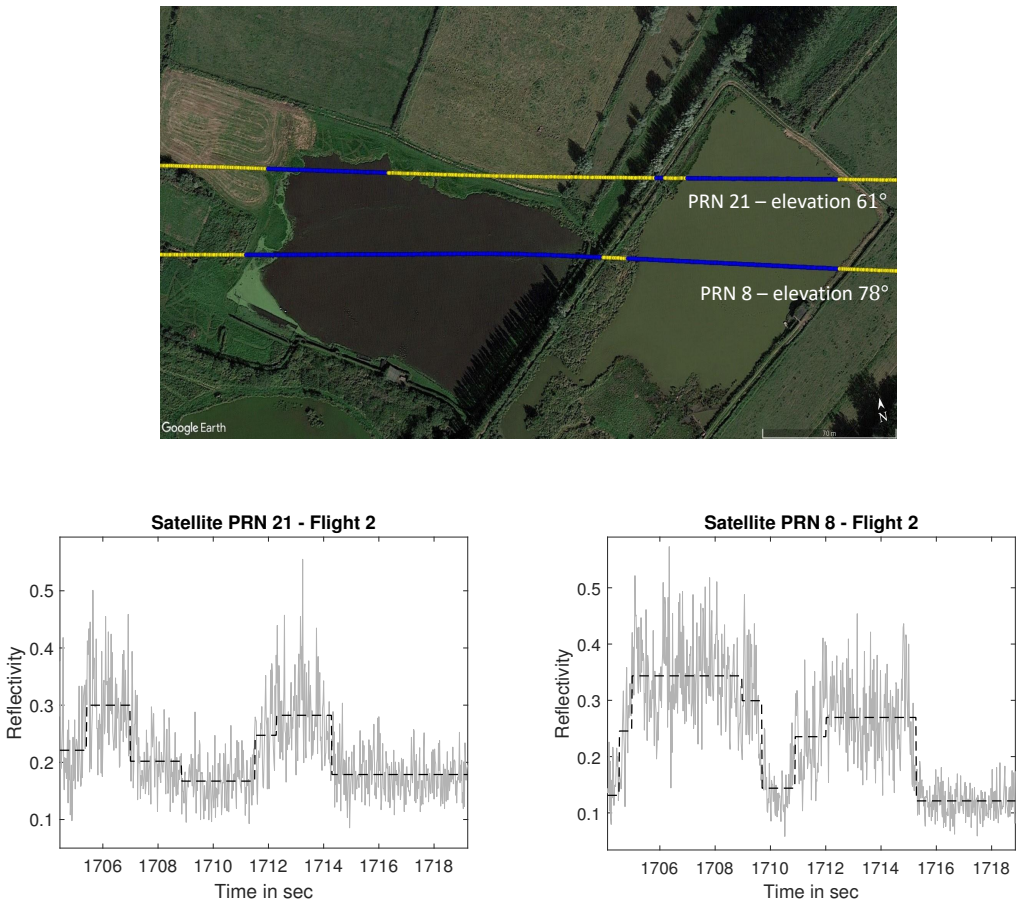


FIGURE 3.6: Segmentation of the GNSS-R reflectivity measurements at (50.878126°, 1.925893°) – **Flight 2**.

Unlike supervised classification methods that require labeled training data, K-means operates without prior knowledge of surface types, making it particularly suitable for exploratory analysis of GNSS-R data where ground truth may be limited or unavailable.

In our study, it is important to note that considered data point corresponds to the detected segment of the reflectivity signal. This segment-based approach, as opposed to pixel-based classification, leverages the temporal coherence of GNSS-R observations and reduces the impact of instantaneous noise fluctuations on the classification results. Each segment represents a statistically homogeneous region identified by the change point detection algorithm, ensuring that the clustering operates on meaningful physical units rather than arbitrary data windows.

The K-means method involves three iterative phases: initialization, assignment, and update. The algorithm begins with K_c cluster centers, $K_c = 3$

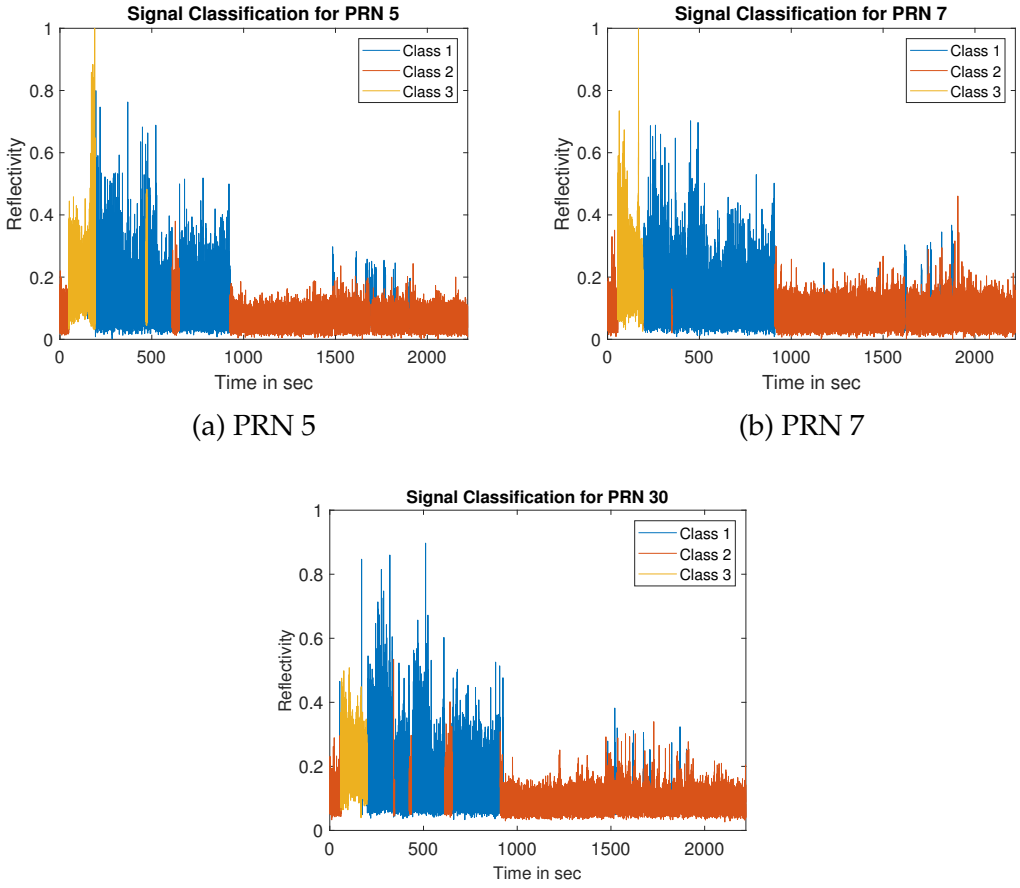
in our case, that are initialized randomly in the feature space defined by segment mean and standard deviation. This initialization strategy, while simple, has been shown to be effective for the relatively low-dimensional feature space used in this application.

Segments are then assigned to the nearest cluster center based on their feature similarity, mean and standard deviation in our case. The distance metric used is typically the Euclidean distance in the two-dimensional feature space, though other metrics could be employed if the features exhibit different scales or distributions. The assignment process ensures that each segment belongs to exactly one cluster, providing a hard classification suitable for discrete surface type identification.

Then, the cluster centers are updated by calculating the mean values of the points inside each cluster. This update step minimizes the within-cluster sum of squares, which is the objective function optimized by the K-means algorithm. The iterative nature of this process ensures convergence to a local optimum, though the global optimum is not guaranteed due to the non-convex nature of the clustering problem.

The iterative process continues until convergence is attained, usually when the distribution of segments to clusters achieves a state of stability. Convergence is typically determined by monitoring the change in cluster assignments between iterations or by tracking the decrease in the objective function. In practice, convergence usually occurs within 10-20 iterations for our GNSS-R clustering problem, making the algorithm computationally efficient for real-time applications.

In Figure 3.7, our K-means segment clusters are depicted, illustrating three main classes corresponding to sandy beach, sea, and land when overlaid on visible (Google Earth) and IGN classification maps. We observe a significant level of stability in the segmentation results of the flight's three satellite signals under consideration. This demonstration showcases our segmentation approach ability to identify three distinct classes, with an automatic categorization of the signal's characteristics with a general perspective on the nature of the GNSS-R data.



(a) PRN 5

(b) PRN 7

(c) PRN 30

FIGURE 3.7: Reflectivity Classification for PRN 5, 7 & 30 – **Flight 1**.

The same classification algorithm was applied to the reflectivity measurements obtained from PRNs 8 and 21 of Flight 2, and the corresponding results are presented in Figure 3.8. The results are in alignment with those obtained from previous GNSS-R data, hence highlighting the reliability and consistency of the automatic classification approach across different datasets.

It is worth noting that there are differences in sea and sandy beach reflectivity levels between the two flights, attributed to a variety of factors such as the difference in precipitation and wind speed between both flight dates. During the second flight, which was done on a day with clear weather conditions and relatively low wind speed, higher reflectivity measurements were obtained from the sea due to the reflections being mostly coherent, in contrast to Flight 1 where the presence of wind caused a decrease in the reflectivity levels due to a higher number of incoherent reflections.

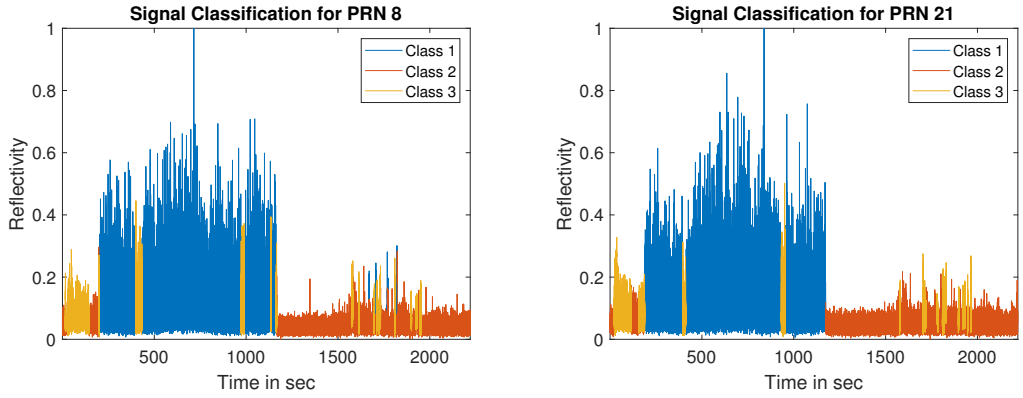


FIGURE 3.8: Reflectivity Classification for PRN 8 & 21 – **Flight 2**.

3.4.2.5 Percentage of Overlapping of GNSS-R Reflectivity Signals

The assessment of overlapping percentages for each class provides valuable insights into the performance and consistency of the GNSS-R classification methodology across diverse signals. These percentages serve as indicators of the level of agreement between the classified classes and reference signals.

The overlapping percentage is calculated using the equation:

$$\text{P(i,PRN)} = \frac{\text{card}(\text{class(i,ref)} \cap \text{area(i)} \cap \text{class(i,PRN)})}{\text{card}(\text{class(i,ref)} \cap \text{area(i)})} \times 100\% \tag{3.65}$$

where $i = [1,2,3]$ and ref denotes the PRN 5 as the reference for Flight 1 and PRN 8 for Flight 2). $class(i, PRN)$ is the set of samples classified as class i for PRN and $area(i)$ is the set of samples in area i defined in Figure 3.9 which was determined using Google Earth imagery and accurate IGN (Institut Géographique National) maps. These areas are carefully chosen to ensure visibility by all satellites. For the plain land, we took the part before $t = 1500$ sec as seen in Figure 3.9, as beyond this point, we enter a region where many lakes are present but not observed by all satellites.

In table 3.2, PRN 7 and PRN 30 of the flight demonstrated high level of consistency obtaining overlapping percentages of 97.37% and 94.42% for class 1, 99.67% (PRN 7) and 99.84% (PRN 30) for class 2 and 95.58% and 87.78% for class 3 with PRN 5 taken as the reference. As previously discussed, the traces followed by the three satellites are close but not exactly the same, and thus the reflectivity measurements vary depending on the reflecting surface characteristics and the detection length by the satellites. The

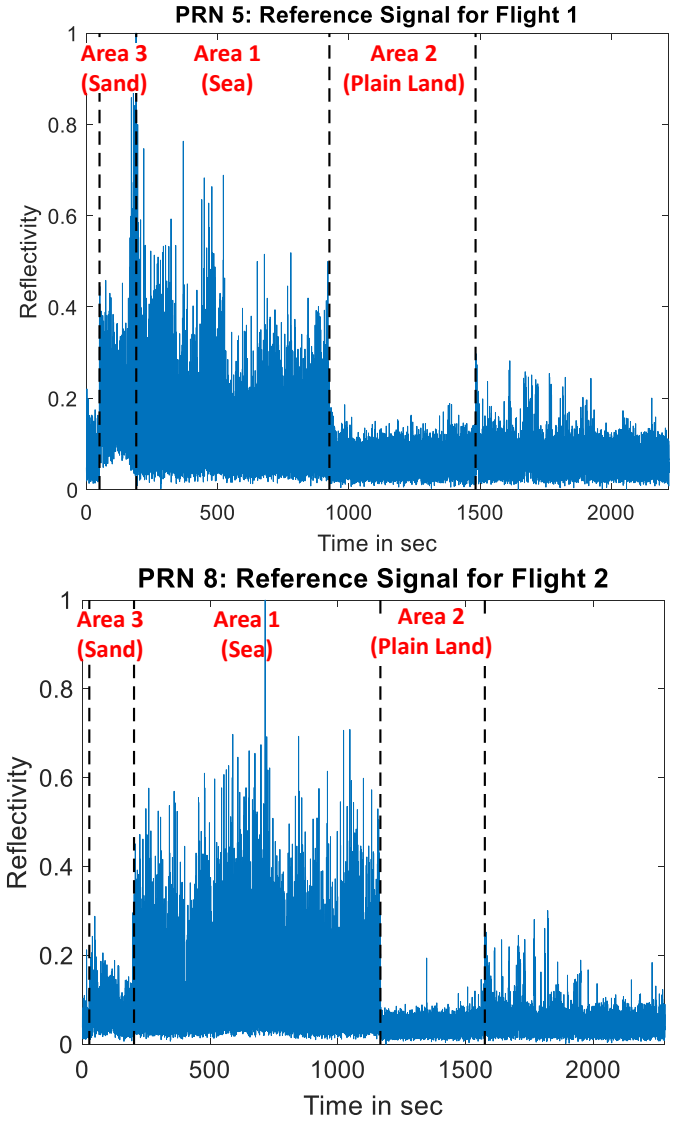


FIGURE 3.9: Reflectivity reference areas

TABLE 3.2: Overlapping Percentages $P(i, PRN)$ of GNSS-R Reflectivity Signals

	PRN \ i	1	2	3
FLIGHT 1	PRN 5	100	100	100
	PRN 7	97.37	99.67	95.58
	PRN 30	94.42	99.84	87.78
FLIGHT 2	PRN 8	100	100	100
	PRN 21	97.20	98.11	73.49

same conclusion can be made for Flight 2 when PRN 8 is taken as the reference.

These overlapping percentages show a good agreement between the reflectivity segmentation obtained for different PRNs for both flights. The overlapping differences are mainly due to the difference of satellite traces and the miss detection. However, the automatic classification that uses the mean and the variance estimate in each segment is not sufficient to fully differentiate between sea and sand as shown for the different flights. In perspective of this work, environmental conditions as well as the measurement of coherence of the reflectivity should be able to improve the classification.

During Flight 1, the wind speed was 24 km/h while it was 12 km/h in Flight 2. This difference explains why the reflectivity on the sea in Flight 1 is lower than that in Flight 2. Flight 1 occurred in October, a rainy month in Northern France with expected high humidity. In contrast, Flight 2 took place in June, a dry and warm month, explaining the lower reflectivity levels of sand and land compared to Flight 1.

3.5 Conclusion

This chapter proposes an integrated methodology for segmenting GNSS-R reflectivity signals affected by multiplicative speckle noise while addressing the core issues highlighted in Chapter 2 of detecting change points under non-additive noise conditions. The proposed methodology represents a breakthrough in GNSS-R signal processing techniques, offering statistically robust methods for the automatic detection and classification of surface transition.

The fundamental theoretical contribution entails statistical modeling of the reflectivity of GNSS-R using the Gamma distribution as an extension of electromagnetic scattering theory with extensive real airborne data analysis. The homomorphic log-transformation becomes a key innovation as a technique of transforming multiplicative speckle noise into constant variance additive noise and thus permitting the use of classical change detection algorithms with statistically derived parameter-independent thresholding.

The integrated online-offline detection system effectively merges the computational efficiency of CUSUM-based online detection with Gamma Maximum Likelihood Estimation of offline localization accuracy. The method overcomes the limitations of parameter sensitivity and computational complexity found in current methods discussed in Chapter 2, while maintaining consistent performance across various environmental conditions without needing prior knowledge of surface-specific parameters.

The post-localization merging process conducted with Wald test statistical hypothesis testing efficiently eliminates the tendency of over-segmentation of sensitive change detection systems. The approach offers criteria for differentiation of real surface transitions and false detections triggered by noise, significantly improving the practical use of the segmentation results.

We demonstrated on synthetic data that the proposed LGMLE change-point localization consistently outperforms the classical MLLE framework under multiplicative speckle noise across a range of reflectivity mean. The empirical distribution of the LGMLE estimates remains tightly centered on the true change time, whereas MLLE exhibits larger bias and dispersion.

Experimental verification with real airborne GNSS-R data of Northern France campaign proves the effectiveness of the framework with a wide variety of surface coverings ranging from sandy areas to open ocean and inland waterbodies, as well as other land coverings. Automatic classification through K-means clustering reveals overlapping percentages greater than or equal to 87% of the major surface coverings, supporting the reliability of the segmentation method.

The developed segmentation framework offers the fundamental prerequisite for advanced GNSS-R analysis through the provision of statistically uniform signal segments with well-defined properties. These are used as the underlying building blocks of phase-based analysis that follows and from which coherence information and surface-specific scattering properties studied in Chapter 4 are extracted. Robust segmentation ensures that the analysis of the phase takes place with statistically significant data regions and as a natural prerequisite enables the derivation of advanced coherence detection and surface classification algorithms.

The chapter's contributions at the methodological level can be used beyond GNSS-R applications and the overall framework of radar signal processing with multiplicative noise scenarios. It can provide a template to similar change detection challenges in other remote sensing and weather radar applications.

Chapter 4

GNSS-R Phase Modeling and Excess Doppler Analysis

4.1 Introduction

Although traditional GNSS-R methods have primarily used amplitude-based observables, the phase content of backscattered signals carries a significant amount of information on scattering mechanisms and surface parameters that has been largely underexplored.

This chapter develops a comprehensive framework for modeling GNSS-R phase and excess Doppler effects, thereby establishing the theoretical foundations and practical procedures necessary for phase-based surface classification. The approach is based on the geometry of bistatic radars and the principles of signal propagation from earlier chapters, and it extends the analysis for simulating the statistical properties of the phase estimator derived from in-phase and quadrature components of the received signal.

The basic concept is that the coherence properties of the reflected signal phase serve as a direct indicator of the underlying scattering regime. For smooth surfaces, specular scattering produces coherent phase variation, but over rough surfaces, multiple scattering routes lead to complex interference patterns that display incoherent phase changes. This fundamental difference provides a powerful discriminant for automated surface type identification.

We develop a statistical phase model based on the von Mises distribution framework, which captures the transition between coherent and incoherent regimes as a function of signal-to-noise ratio and reflection component magnitudes. The model is derived from the complete signal processing chain including demultiplexing, demodulation, and correlation operations, with

both continuous-time and discrete-time implementations presented.

We introduce excess Doppler spectrum analysis using three quantitative criteria: peak amplitude of the dominant Doppler component, count of statistically significant spectral peaks, and degree of spectral dispersion. The proposed criteria are evaluated using synthetic GNSS-R data generated from the statistical I/Q model, which enables the assessment of their discriminative capability. The theoretical developments are validated through a comprehensive analysis of real airborne GNSS-R data, demonstrating clear distinctions between different surface types.

The chapter is organized as: Section 4.2 establishes the surface geometry and signal representation model; Section 4.3 outlines the statistical phase model derivation with signal processing techniques and coherence analysis; Section 4.4 gives the excess Doppler spectrum analysis with results of validation; Section 4.5 displays experimentation results with classification performance on both synthetic and real data. The contributions provide the foundation for more advanced GNSS-R signal processing techniques that take advantage of phase and spectral data to enable enhanced surface classification capability.

4.2 Surface Geometry and Signal Representation

In a GNSS-R radar observation system, the reflection mechanism is heavily dependent upon surface roughness in terms of signal wavelength. Analyzing this relationship involves reviewing two primary geometric concepts: the specular point and the effective scattering area. The specular point is the spot wherein, at this location, the reflection angle and incident angle are equal so that it obeys Snell's law of reflection [104]. In a perfectly smooth surface, then, there is, at this single point, all of the coherent reflection [19]. Figure 4.1 shows the bistatic geometry of GNSS-R, in which both the direct and reflected paths of the signal converge at the specular point. The geometry is defined by the transmitter-specular point-receiver path, which determines the fundamental delay and Doppler characteristics of the reflected signal.

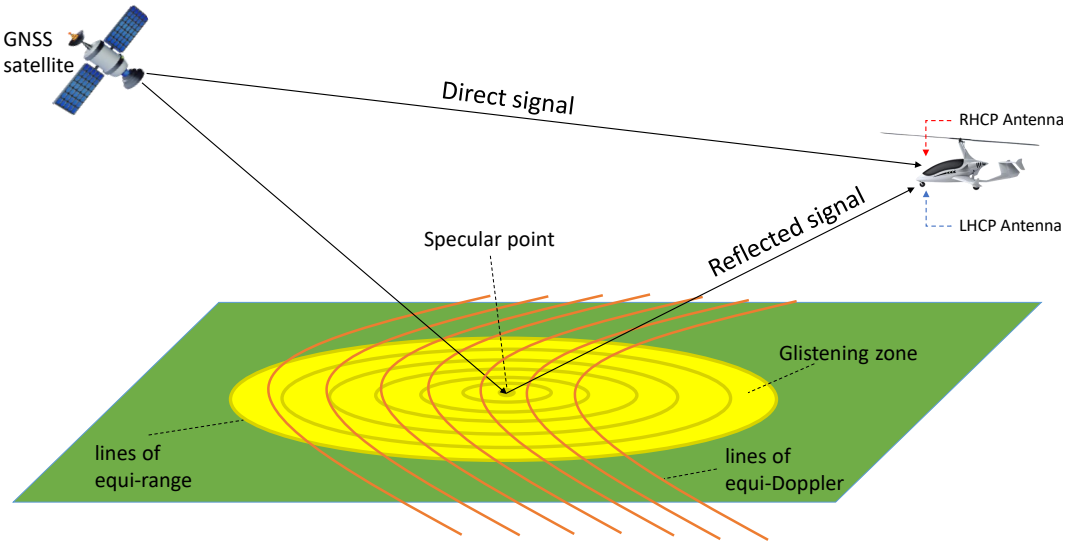


FIGURE 4.1: GNSS-R observation system.

4.2.1 From Fresnel Zone to Glistening Zone: The Impact of Surface Roughness

For smooth surfaces, in which variations in surface elevation are much smaller compared to the wavelength of incoming waves, this reflection phenomenon is primarily limited to the first Fresnel zone. In this elliptically defined region, variations in path length that the respective waves encounter are much less than $\lambda/4$ (for GPS L1: $\lambda = 19.042\ cm$). The size of this elliptical zone is defined by two axes: semi-major axis, termed as $a = \frac{\sqrt{\lambda h sin(\theta_e)}}{sin^2(\theta_e)}$, and semi-minor axis, termed as $b = \sqrt{\frac{\lambda h}{sin(\theta_e)}}$. The variables that affect these parameters are h, representing receiver height, θ_e, representing elevation of incoming wave, and λ, representing the GPS signal wavelength.

As surface roughness increases, there is a significant increase in the glistening zone, which is attributed to the entire region that contributes to providing power in reflected signals. This increase extends far beyond what is normally seen within the first Fresnel zone, describing a prominent signal reflection phenomenon. The cause behind this increase lies in that there are several facets of a surface that are of varying orientation; these facets are in a position to direct incoming signals towards a receiver in places that are not necessarily limited to the nominal specular point, which is the perfect reflection point. Each correctly oriented facet creates a new reflection path, adding distinctive properties in reflection delay as well

as Doppler effects attributed to signals. This fundamentally impacts the received signal, showing how surface roughness can influence signal behavior.

Figure 4.2 illustrates, in great detail, the transformation of confined zone scattering to more extensive glistening zone scattering. This graphical representation clearly indicates that as the roughness of the surface is higher, the effective reflection area increases substantially.

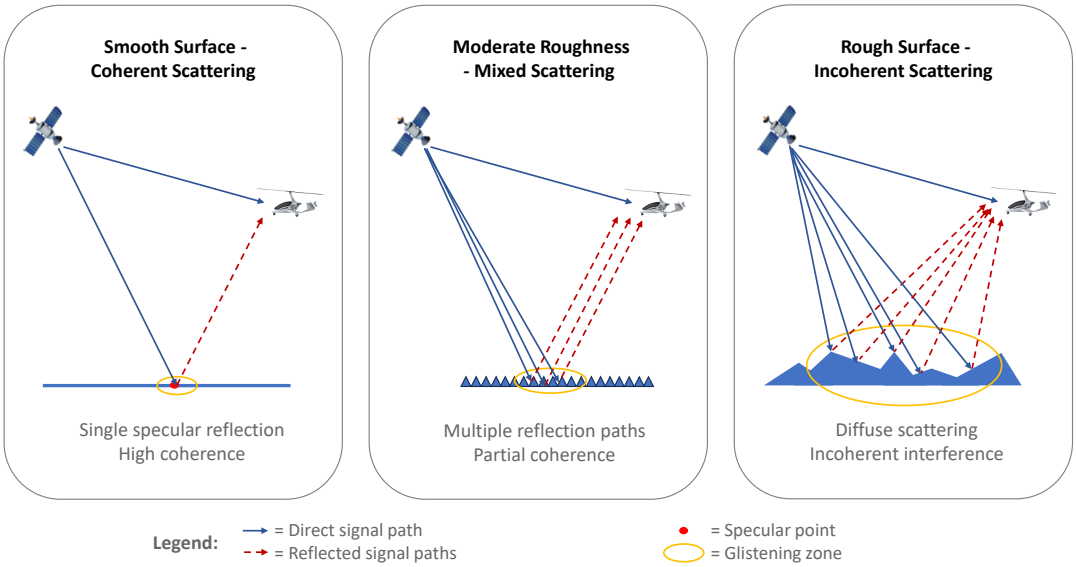


FIGURE 4.2: Surface Roughness Impact on Reflection Characteristics.

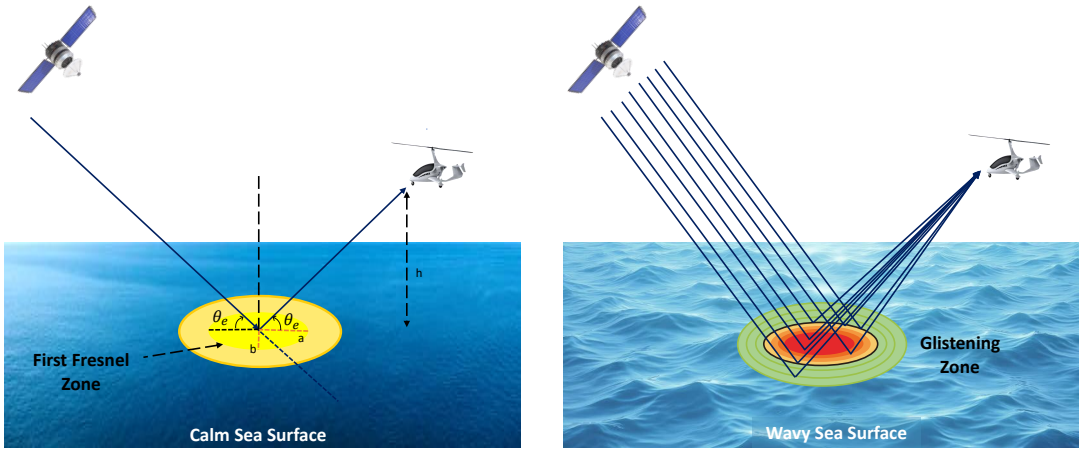
4.2.2 Signal Composition and Scattering Components

The overall scattering power of rough surfaces consists of two parts having essentially different natures. The coherent part is derived from regions where the correlation length of the surface is significantly greater than the first Fresnel zone size. The coherent part stays synchronized over the integration time, making it suitable for interferometric processing and precise measurements. It is usually predominant over quiet water surfaces, wetlands, and ice surfaces, for which the surface is effectively smooth compared to the GNSS wavelength.

The incoherent part is due to surface roughness, which induces several scattering paths with virtually random phase connections. Varied delays are representative of these several paths, and Doppler shifts occur at multiple points of reflection located within the extended glistening zone. This part prevails on vegetation, dry, rough soils, and roughened wind surfaces in lakes

and oceans where surface variations are approaching or exceeding the signal wavelength scale.

Figure 4.3 demonstrates how the relative contributions of these components change with surface roughness conditions, showing the transition from coherent-dominated to incoherent-dominated scattering regimes.



(a) Coherent Scattering from Calm Sea Surface

(b) Diffuse Scattering from Wavy Sea Surface

FIGURE 4.3: Geometric representation of coherent and incoherent GNSS-R scattering over calm and wavy sea surfaces, illustrating the Fresnel and glistening zones.

4.3 Signal Processing and Phase Estimation

This section derives the whole chain for recovering phase information from direct and reflected GNSS signals. The theory is formulated in terms of continuous-time signal processing, and explicit mathematical derivations are provided in Appendix A. The following are the discrete-time numerical schemes that are fundamental to our experimental phase estimation methodology.

The signal processing is systematic: demultiplexing to obtain the desired components of the signal, demodulating to obtain baseband information, and correlation to obtain the in-phase and quadrature measurements required for estimation of phase. Implementations for direct and backscattered signals are provided mainly to demonstrate the entire framework for bistatic processing.

4.3.1 Linear Phase Assumption

Before proceeding with the numerical implementations, we establish the linear phase model that underlies our processing approach.

For the reflected signal with multiple reflection paths, the phase of each reflection is assumed to be a linear function of time during the coherent integration interval T_c:

$$\phi_i(t) = \phi_{0_i} + \alpha_i t \tag{4.1}$$

where $\alpha_i = 2\pi f_{d_i}$ and f_{d_i} represents the Doppler shift of the ith reflection path within the glistening zone.

4.3.2 Numerical Implementation for Reflected Signal

For the numerical implementation of the reflected signal processing, we discretize the multi-path reflection model following the same systematic approach as the direct signal but accounting for the summation over multiple reflection paths. Here, u is an integer defined between 1 and Nf_eT_c, with M the number of reflections and N the number of milliseconds of the GNSS signal.

$$s_r(uT_e) = \sum_{i=1}^{M} A_i CA(uT_e + \tau_{r,i}) \sin(2\pi f_i uT_e + \phi_i(uT_e)) + \eta(uT_e) \tag{4.2}$$

Demultiplexing After demultiplexing, the output signal becomes:

$$\begin{aligned} s_D(uT_e) &= CA(uT_e + \tau_l) s_r(uT_e) \\ &= \sum_{i=1}^{M} A_i \sin(2\pi f_i uT_e + \phi_i(uT_e))|_{u\in S_1} \\ &\quad - \sum_{i=1}^{M} A_i \sin(2\pi f_i uT_e + \phi_i(uT_e))|_{u\in S_2} \\ &\quad + CA(uT_e + \tau_l)\eta(uT_e) \end{aligned} \tag{4.3}$$

Correlation If we process the correlation for a local delay τ_l, the correlation output is expressed as:

$$\begin{aligned} C_{kT_c} \approx & \sum_{i=1}^{M} A_i R(\tau_i - \tau_l) \\ & \times \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \sin(2\pi f_i u T_e + \phi_i(uT_e)) \\ & + \sum_{u=(k-1)f_eT_c}^{kf_eT_c} CA(uT_e + \tau_l)\eta(uT_e) \end{aligned} \tag{4.4}$$

where $R(\tau_i - \tau_l)$ is the cross-correlation function and we state $a_i = A_i R(\tau_i - \tau_l)$.

For noise, the summation simplifies to:

$$\sum_{u=(k-1)f_eT_c}^{kf_eT_c} CA(uT_e + \tau_l)\eta(uT_e) = \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \eta(uT_e) \tag{4.5}$$

Correlation after demodulation The in-phase (I) and quadrature (Q) correlation components are:

$$I_{kT_c} = \sum_{i=1}^{M} \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \frac{a_i}{2} \cos(\phi_i(uT_e)) + \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \sin(2\pi f_i u T_e)\eta(uT_e) \tag{4.6}$$

$$Q_{kT_c} = \sum_{i=1}^{M} \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \frac{a_i}{2} \sin(\phi_i(uT_e)) + \sum_{u=(k-1)f_eT_c}^{kf_eT_c} \cos(2\pi f_i u T_e)\eta(uT_e) \tag{4.7}$$

Applying the linear phase assumption established in section 4.3.1, as it is shown in Appendix A:

$$\sum_{u=(k-1)f_eT_c}^{kf_eT_c} \cos(\phi_{i,0} + \alpha_i u T_e) = \sin\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) \tag{4.8}$$

$$\sum_{u=(k-1)f_eT_c}^{kf_eT_c} \sin(\phi_{i,0} + \alpha_i u T_e) = \cos\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) \tag{4.9}$$

Open-Loop Approximation Finally, in an open-loop configuration, we have:

$$I_{kT_c} = \sum_{i=1}^{M} a_i \cos\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_I \tag{4.10}$$

$$Q_{kT_c} = \sum_{i=1}^{M} a_i \sin\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_Q \tag{4.11}$$

4.3.3 Phase Estimation and Coherence Analysis

The estimated phase of the signal is given by:

$$\hat{\phi}_{kT_c} = arctan\left(\frac{Q_{kT_c}}{I_{kT_c}}\right) \tag{4.12}$$

The weighted sum of the phase in the circular domain is defined as [105]:

$$\hat{\phi}_{kT_c} = \arctan 2\left(\frac{\sum_{i=1}^{M} a_i \sin\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_Q}{\sum_{i=1}^{M} a_i \cos\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_I}\right) \tag{4.13}$$

where $arctan2$ is the four-quadrant inverse tangent.

We show in the appendix A that η_I and η_Q are gaussian noise, in this case the underlying distribution of the noise $\eta_\phi(kTc)$ is a Von Mises distribution. When the concentration parameter of the noise $\eta_\phi(kT_c)$ is high, the von Mises distribution approaches a normal distribution, indicating that the phase ϕ_{kT_c} maintains coherence across successive integration periods. Conversely, when the concentration parameter is low, the distribution approaches uniform behavior, indicating that the phase becomes incoherent. According to the expressions derived for I_{kT_c} and Q_{kT_c}, this transition occurs based on the signal-to-noise ratio when the amplitudes a_i are small compared to the noise power.

The ratio of signal power to noise power can be directly related to the transition between coherent and incoherent surfaces. For the direct signal case, coherence is maintained when $|a_1|^2/2\sigma_\eta^2 >> 1$, where σ_η^2 represents the noise variance. The summation over several reflection components with varying phases and amplitudes makes this condition more complicated for the reflected signal with multiple paths.

The underlying scattering mechanisms directly influence these statistical phase properties through the interference patterns created by multiple reflection paths. This phenomenon is mathematically captured in the multi-path model using I_{kT_c} and Q_{kT_c}. When the amplitudes A_i are comparable and the phases $\phi_{i,0} + \alpha_i T_c(k-1)$ are distributed across the full range $[0, 2\pi]$, the vector sum results in a low-magnitude composite signal with highly variable phase.

In contrast, over smooth surfaces such as calm water bodies, reflection is dominated by the specular component where the coherent nature means that most reflected power comes from a single, well-defined reflection path. This results in stable phase relationships that vary smoothly with observation geometry changes. For specular-dominated reflections, the single-path approximation becomes valid: $I_{kT_c} \approx A_s \cos(\phi_{s,0} + \alpha_s T_c(k-1) + \frac{\alpha_s T_c}{2})$ and $Q_{kT_c} \approx A_s \sin(\phi_{s,0} + \alpha_s T_c(k-1) + \frac{\alpha_s T_c}{2})$, where the subscript s denotes the specular component. The phase $\hat{\phi}_{kT_c}$ follows predictable temporal evolution determined primarily by geometric Doppler shift, resulting in smooth, coherent phase behavior.

4.4 Excess Doppler Analysis

4.4.1 Doppler Spectrum Computation

To study the Doppler behavior of the reflected GNSS-R signals, we apply a Fast Fourier Transform (FFT) to the complex baseband signal $I + jQ$. While a sliding window approach is considered, each window is ensured that it remains entirely within a homogeneous surface segment, as defined by the surface transition indices from the segmentation obtained in Chapter 3. Windows that overlap segment boundaries are excluded to avoid spectral artifacts caused by mixed segments and to maintain interpretability. This approach enables a more precise correlation between surface types and spectral features.

The spectral representation of the excess Doppler norm, $|FFT\left(I_{kT_c} + jQ_{kT_c}\right)|$, is expressed as:

$$\hat{\Gamma}(f)_{kT_c} = \frac{f_e T_c}{2} \sum_{i=1}^{w} a_i \delta\left(f - f_{d_i}\right) + \Psi \tag{4.14}$$

Where $\delta(f)$ is the Dirac delta function, Ψ is a multiplicative noise term and f_{d_i} are the peak frequencies of the spectral representation where these peaks are obtained with local maxima search algorithm. To estimate the spectrum, we use the discrete Fourier transform with sampling period T_c and a maximum frequency of $1/2T_c$. For a spectral resolution of at least 1 Hz, the minimum number of data points required in the time domain is $1/T_c$. The spectral function $\hat{\Gamma}(f)_{kT_c}$ is processed using a sliding window approach.

The resulting spectrum illustrates the distribution of the reflected signal energy across Doppler frequencies serving as a foundation for the extraction of different features. We provide three quantitative criteria that capture different aspects of the Doppler energy distribution in order to carefully examine the spectral features and derive significant information about the target surface:

Peak Amplitude of Dominant Doppler Component: It quantifies the power of the primary Doppler component present in the spectrum. This parameter is defined mathematically as:

$$C^1_{kT_c} = \underbrace{\max}_{f\in[-1/2T_c,1/2T_c]} \hat{\Gamma}(f)_{kT_c} \tag{4.15}$$

with *Tc* denoting the coherent integration time used in radar processing. The maximum amplitude acts as a measure of the coherent scattering effect from the surface, where higher values are generally linked to more specular reflection behaviors. This criterion is notably sensitive to the influence of dominant scattering mechanisms and offers valuable understanding of the overall signal-to-noise ratio in the measurement.

Count of Significant Doppler Peaks: It is an estimate of the number of Doppler spectrum spectral peaks that are statistically significant. A spectral peak is considered significant when its amplitude exceeds a threshold set at half the maximum peak value:

$$C^2_{kT_c} = \sum_{f=-1/2T_c}^{1/2T_c} \sum_{i=1}^{p} \gamma(i)\delta\left(f - f_{d_i}\right) \tag{4.16}$$

$$\gamma(i) = \begin{cases} 1, & \text{if } a_i \geq C^1_{kT_c}/2 \\ 0, & \text{otherwise} \end{cases} \tag{4.17}$$

where $\gamma(i)$ is a function defined to differentiate between peaks originating from the signal and those arising from noise, while a_i represents the amplitude of each peak.

By ensuring that only peaks with a significant energy content are taken into account in the analysis, this threshold-based method successfully eliminates variations coming from the noise. This criterion acts as an indicator of spectral complexity and is closely associated with the diversity of the scattering environment. Surfaces characterized by greater roughness or various scattering mechanisms generally display higher peak counts, as they produce a wider range of Doppler frequency components resulting from each reflection scattering element.

Spectral Dispersion Across Peaks: It quantifies the degree to which the total spectral power is distributed among multiple significant peaks, rather than being concentrated in a single dominant component. It is determined by subtracting the maximum peak amplitude (C^1) from the cumulative amplitude of all significant peaks. Using $\gamma(i)$ from eq. 25, it is expressed as:

$$C^3_{kT_c} = \frac{f_e T_c}{2} \sum_{i=1}^{p} \gamma(i) a_i - \underbrace{\max_{f \in [-1/2T_c, 1/2T_c]}}_{} \hat{\Gamma}(f)_{kT_c} \tag{4.18}$$

The spectrum dispersion quantifies the spread of energy across the doppler spectrum range. Greater values of C^3 indicate that the reflected energy is more widely dispersed across several Doppler frequencies, which usually happens when the surface is significantly rough or when the backscattered signal is the result of various scattering mechanisms. This measure is very useful for determining the difference between surfaces that have similar peak amplitudes but different scattering properties.

The three spectral criteria are illustrated in Figure 4.4, which shows the FFT spectrum of the reflected GNSS-R signal segment for a time interval between 1770 and 1780 s. This segment includes a coherent scattering behaviors, providing a representative example for evaluating the criteria. The figure highlights the computation of the three quantitative parameters: the maximum Doppler peak amplitude C^1, the count of significant peaks C^2, and the spectral dispersion C^3.

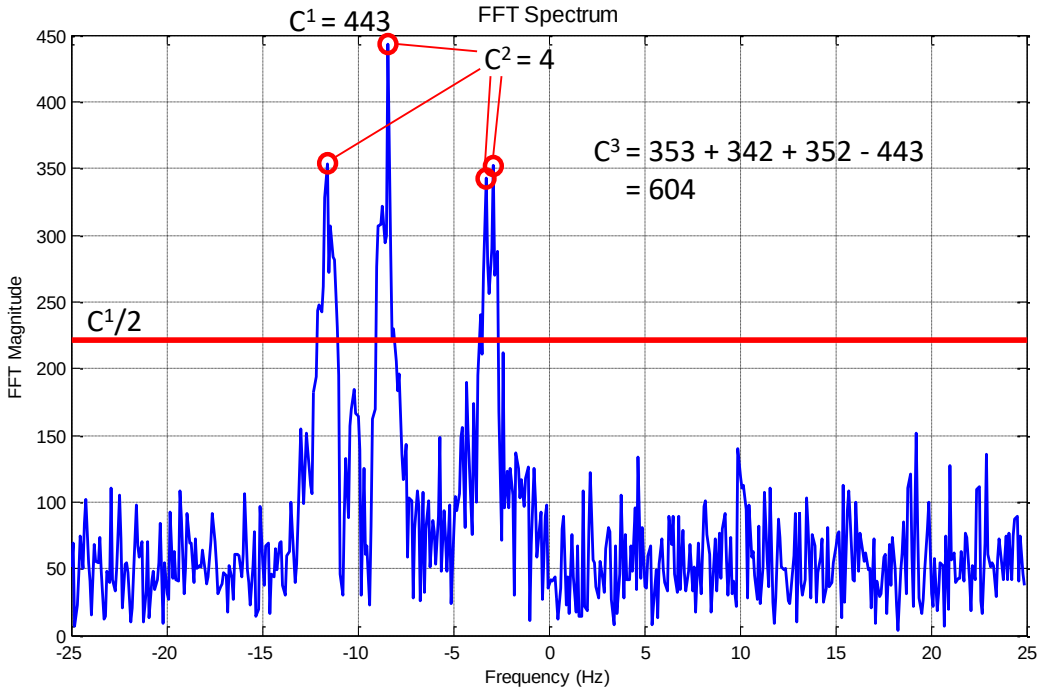


FIGURE 4.4: FFT spectrum of the reflected GNSS-R signal segment illustrating the computation of C^1, C^2 and C^3 (spectral dispersion). Data segment at $(50.876064°, 1.948393°)$ between 1770–1780 s.

4.4.2 Phase Coherence Across Surfaces

To interpret the phase behavior derived in 4.12, we analyze a real airborne GNSS-R example showing a transition from a rough land surface to a smooth lake.

Figure 4.5 displays the trajectory of the specular reflection points on a Google Earth image. The trajectory is color-coded to indicate the transition: the blue segment corresponds to signal reflections from land, while the cyan segment indicates reflections from a lake surface.

The phase of the reflected signal is presented in Figure 4.6. A clear difference in phase behavior is observed: over land, the phase exhibits rapid, irregular fluctuations, characteristic of incoherent reflections. Over the lake, the phase becomes smoother indicating increased coherence. This observation is consistent with the theory: as the amplitudes a_i decrease relative to the noise power, the signal-to-noise ratio decreases, and the phase tends toward a uniform distribution, resulting in incoherence. In contrast, over smoother surfaces like water, higher coherence is preserved.

These figures illustrate the phase coherence dependency on surface roughness and SNR, as formulated through the von Mises phase model introduced earlier.

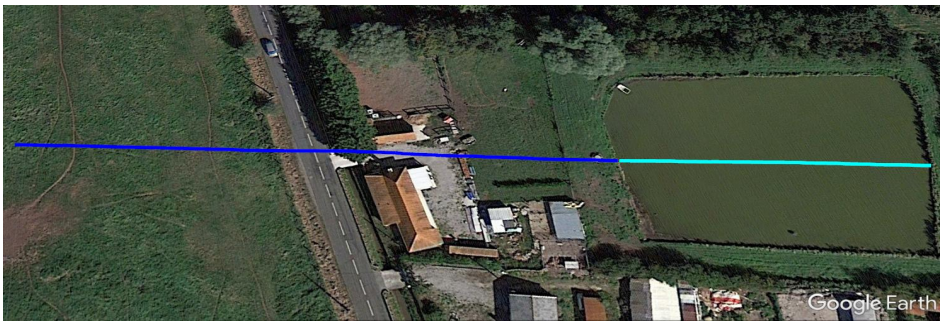


FIGURE 4.5: Specular reflection point trajectory of GPS satellite PRN 5 over a 10-second segment, showing a transition from rough land (blue) to a smooth lake surface (cyan).

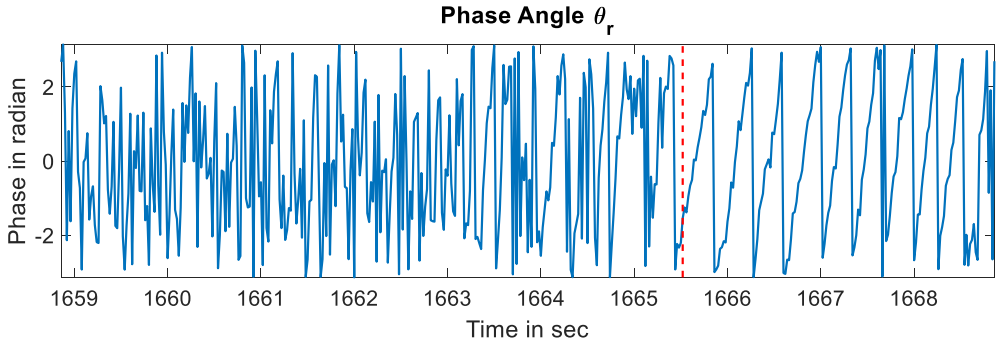


FIGURE 4.6: Corresponding phase of the reflected GNSS signal from GPS satellite PRN 5 over the same 10-second interval.

4.4.3 Phase Model Verification

In order to evaluate the accuracy of the estimated phase expression in 4.12, the observed carrier phase obtained from the signal reflected over the same region shown in Figure 4.5 is compared with the estimated phase derived using the previously discussed spectral analysis. The comparison is presented in Figure 4.7, where the blue curve denotes the unwrapped observed phase, and the red dashed line represents the estimated phase.

Even in the non-coherent part of the reflected phase signal, where the phase evolution is no longer linear, the estimated phase, obtained using

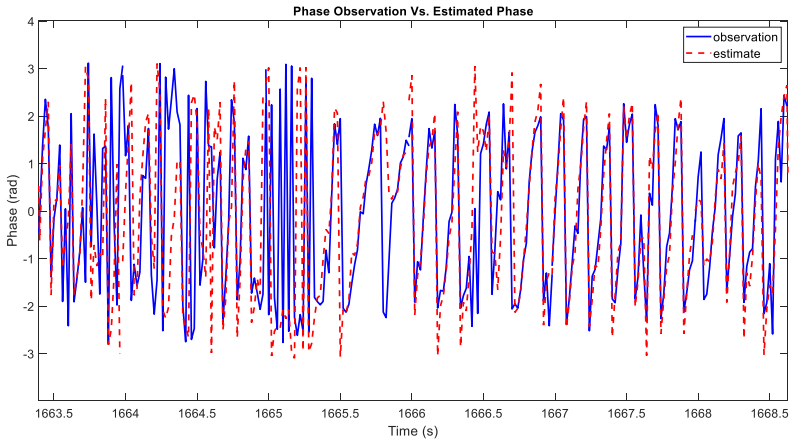


FIGURE 4.7: Comparison between the real phase observation of the reflected GNSS-R signal (blue) and the estimated phase computed from the I/Q components (red dashed) over a transition from land to lake segment.

parameters derived from the significant-peak criteria, shows good agreement with the observations. This agreement becomes even stronger in the coherent region over the lake.

This comparison offers further validation for the phase model employed in this study and strengthens its relevance for surface-type classification and subsequent Doppler-domain analysis in GNSS-R remote sensing.

4.5 Experimentation

4.5.1 Synthetic Data Assessment

4.5.1.1 Synthetic Data Generation Methodology

Synthetic generation of GNSS-R signals requires detailed knowledge of the physical processes that take place during signal reflection and scattering from surfaces. The synthetic signal generation approach employed in the current study is based on the statistical modeling of real airborne GNSS-R data. Large-scale parameter extraction was performed to retain the fundamental characteristics of all surfaces during the experimental flight campaign. This allows the synthetic signals to maintain properties of real GNSS-R measurements. This synthetic dataset is used to evaluate the discriminative capability of the proposed Doppler-based criteria under controlled signal

conditions.

The synthetic signal generation relies on the multi-path reflection model built in Section 4.2, where the received signal is a coherent superposition of multi-path reflection components within the glistening zone. Each multi-path reflection component is characterized by its own amplitude, Doppler frequency shift, and phase characteristics, jointly determining the received signal's spectral and temporal features. The characteristics of these parameters for different surfaces were derived by analysis of real data segments, ensuring that the synthetic signals exhibit physical relationships as observed in actual measurements.

Surface-Specific Parameter Characterization: The parameter extraction process required a detailed statistical analysis of experimental data to obtain feasible parameter ranges associated with each surface type. There are four surface classes: land, river, sea, and sand, each exhibiting distinct reflective characteristics that lead to different parameter distributions. The ranges of the generated parameters for each surface type are summarized in Table 4.1.

TABLE 4.1: Synthetic GNSS-R signal generation parameters.

Parameter	**Sand**	**Sea**	**Land**	**River**
Number of reflections	1–2	5–8	8–10	1–3
Doppler frequency (Hz)	$[-10, 10]$	$[-22, 22]$	$[-25, 25]$	$[-15, 15]$
Dominant peak amplitude	1200	400	200	600
Secondary peak amplitude	600–1000	200–375	125–225	400–650

The number of reflections is an indicator of the level of complexity of the scattering environment. Sea and land surfaces usually exhibit multiple reflection rays due to surface roughness, while river and sand surfaces experience more focused reflections.

Land surface parameters were generated using several reflections M, ranging from 8 to 10, revealing the inherently complex scattering environment produced by vegetation, terrain variability, and mixed surface covers. The excess Doppler frequency interval spans from -25 Hz to 25 Hz, providing sufficient coverage for the diverse scattering geometries observed over terrestrial surfaces. The dominant peak amplitude is relatively low (around 200), while multiple secondary peaks are present within the range 125 to 225,

reflecting the distributed and incoherent nature of the scattering.

River surface parameters represent the specular nature of water body reflection, commonly showing only one to three dominant paths ($M = [1, 3]$) and a restricted excess Doppler frequency range (-15 Hz to 15 Hz) that indicates coherent reflection with limited dispersion. The dominant peak amplitude is higher (around 600), with secondary peaks ranging from 400 to 650, consistent with strong specular reflection from smooth water surfaces.

Sea surface parameters are characterized by multiple reflection paths (M ranging from 5 to 8), reflecting the increased surface roughness induced by wind-driven waves. The excess Doppler frequency range extends from -22 Hz to 22 Hz, accounting for the increased spectral dispersion associated with dynamic sea states. The dominant peak amplitude is around 400, while several secondary peaks are present within the range 200 to 375, reflecting the contribution of multiple scattering components.

Sand surface parameters correspond to more homogeneous and relatively smooth terrains, generally characterized by a limited number of reflection components (M ranging from 1 to 2). The Doppler frequency range spans from -10 Hz to 10 Hz. The dominant peak amplitude is high (around 1200), with secondary peaks ranging from 600 to 1000, consistent with more coherent scattering behavior.

These parameter ranges enable reproducible generation of synthetic GNSS-R signals representative of the considered surface types.

Mathematical Framework for Signal Synthesis: The signal generation process utilizes the theoretical model with discrete time synthesis of the in-phase and quadrature components. For each time step k of a segment, the signal components are determined with the multi-path reflection model, where the total received signal is represented as the coherent superposition of multiple reflection contributions. The generation of the in-phase and quadrature components follows from the basic relationships given in eq. 18 and 19.

The time evolution of the phase incorporates both the linear variation caused by the Doppler shift and the geometric phase relations that arise from

the bistatic nature of the GNSS-R.

The noise components of I and Q are represented as additive white Gaussian noise, characterized by surface-type-specific variance values that correspond to the varying signal-to-noise ratios of different surface types. The noise power is calibrated to align with the statistical behaviors observed in real data.

4.5.1.2 Generated Synthetic Phase Signal

The resulting synthetic phase signal of GNSS-R has been generated by first computing the in-phase (I) and quadrature (Q) components from the multipath model of Section 4.2 and obtaining the phase as arctan2(Q/I). Figure 4.8 shows an example of a complete synthetic phase signal at a few surface areas with their respective class labels.

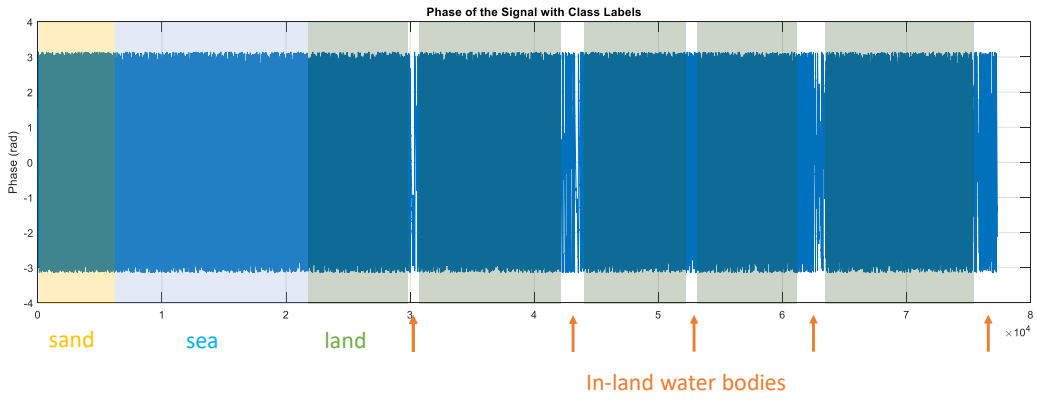


FIGURE 4.8: Synthetic GNSS-R phase signal with surface type classification labels.

The synthetic signal is made up of sequential regions of surfaces encountered in the flight operation with distinctly different coherence features that mirror their physical scattering behavior. The phase signal models the behaviors of transition from one surface type to another, with clear jumps at inter-regional boundaries, reflecting the real-world situation of the GNSS receiver footprint transitioning from one surface type to another.

Phase Coherence Analysis by Surface Type To gain insight into the coherence features relevant to each type of surface, a careful observation of specific regions of the synthetic signal reveals the specific phase behavior

responsible for the surface type classification. The following analysis examines illustrative sections of each type of surface as a demonstration of the coherent and incoherent scattering regimes captured in the synthetic data generation approach.

The synthetic signal exhibits clearly different phase coherence properties that relate directly to the physical scattering mechanisms on each type of surface. On sand surfaces (Figure 4.9), the phase is smooth and slowly varying phase behavior consistent with coherent scattering from homogeneous surfaces. The periodic phase progression is indicative of the dominance of the specular contribution of the reflected rays, where phase relationships remain stable across successive integration periods.

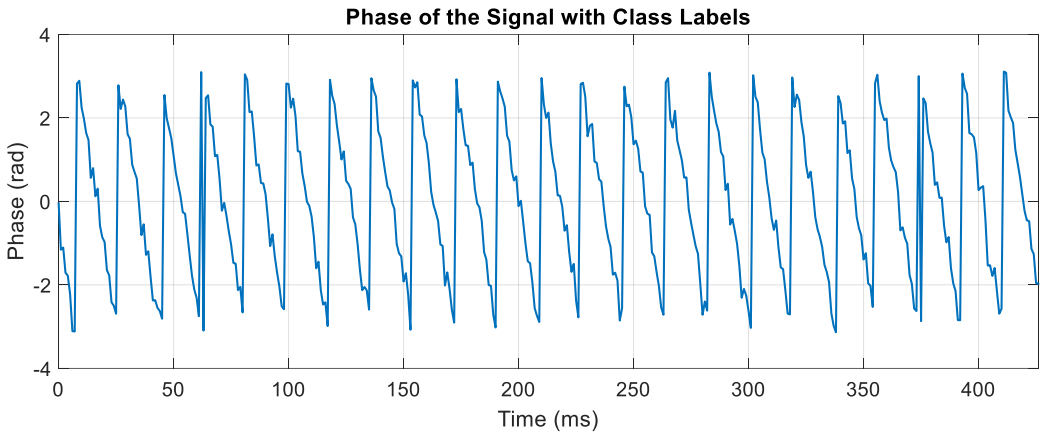


FIGURE 4.9: Phase signal over sand surface demonstrating coherent scattering with smooth phase evolution.

Conversely, land surfaces (referred to in Figure 4.10) produce rapid, random phase changes that indicate incoherent scattering. The extreme variation and irregular nature of phase variations arise due to the multiple scattering paths provided by the heterogeneity and changeability of the surface. This occurrence aligns with the theoretical expectation that when amplitudes are comparable and phases $\phi_{i,0} + \alpha_i T_c(k-1)$ are distributed across the full range $[0, 2\pi]$, the vector sum results in a low-magnitude composite signal with highly variable phase.

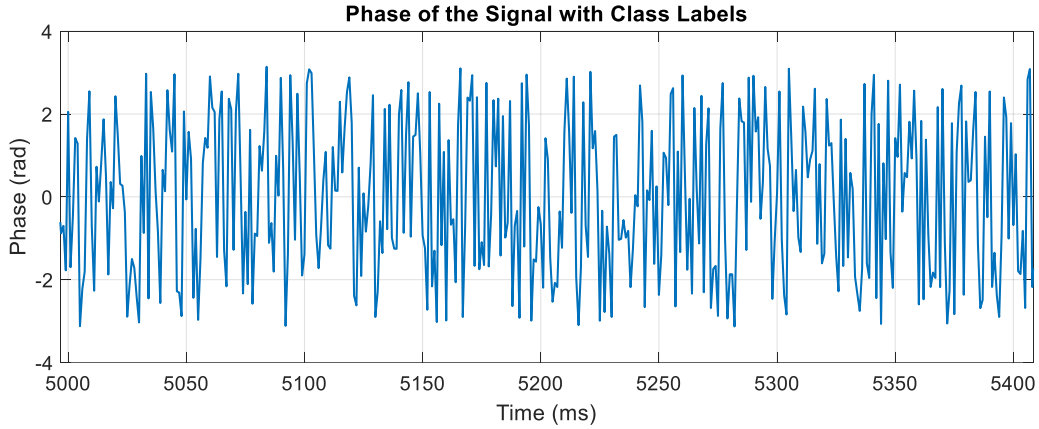


FIGURE 4.10: Phase signal over land surface showing incoherent scattering behavior with rapid phase fluctuations.

4.5.2 Evaluation of Doppler-Based Criteria on Synthetic GNSS-R Signals

To evaluate the discriminative capability of the excess Doppler criteria, K-means clustering is applied to the synthetic GNSS-R signals generated using the statistical I/Q model described previously. This evaluation is done under a fixed signal-to-noise ratio of 10 dB, which is an intermediate regime between coherent and incoherent scattering observed in real GNSS-R measurements.

Two sets of parameters are considered. The first set consists of reflectivity-based parameters, namely the mean and the variance of the reflectivity, introduced in Chapter 3. The second set is composed of the excess Doppler spectrum criteria (C_1, C_2, C_3) introduced in this chapter.

Classification performance is evaluated using standard detection and classification metrics. The probability of correct detection, P_d, measures the proportion of samples from a given class that are correctly detected, while the false alarm probability, P_{fa}, quantifies the proportion of samples incorrectly assigned to that class. The overall accuracy reflects the fraction of correctly classified samples across all classes. The F1-score provides a balanced measure by jointly accounting for false alarms and missed detections.

$$P_d = \frac{TP}{TP + FN} \tag{4.19}$$

$$P_{fa} = \frac{FP}{FP + TN} \tag{4.20}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.21}$$

$$F1 \text{ score} = \frac{2TP}{2TP + FP + FN} \tag{4.22}$$

where TP denotes the number of correctly detected segments of a given surface, FP the number of segments incorrectly assigned to that surface (false alarms), FN the number of missed segments of that surface, and TN the number of correctly rejected segments.

TABLE 4.2: Classification performance on synthetic GNSS-R signals using reflectivity-based parameters (mean and variance).

Class	P_d	P_{fa}	Accuracy	F1-score
Sand	0.88	0.06	0.9250	0.8543
Sea	0.63	0.24	0.7275	0.5361
Land	0.85	0.09	0.8925	0.7981
River	0.30	0.05	0.7850	0.4109

TABLE 4.3: Classification performance on synthetic GNSS-R signals using excess Doppler spectrum criteria.

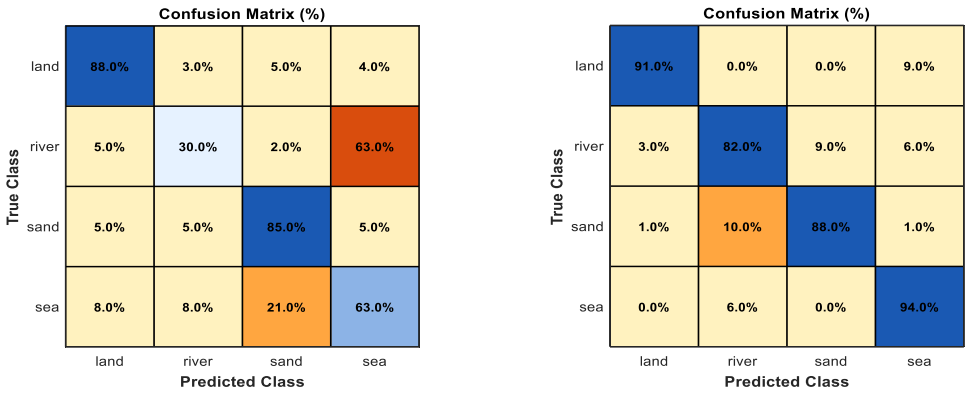
Class	P_d	P_{fa}	Accuracy	F1-score
Sand	0.94	0.0533	0.9450	0.8952
Sea	0.82	0.0500	0.9150	0.8280
Land	0.88	0.0300	0.9475	0.8934
River	0.91	0.0130	0.9675	0.9333

Table 4.2 shows the classification performance obtained using reflectivity-based parameters. While sand and land surfaces show better classification performance, sea and river surfaces exhibit lower probabilities of detection and reduced F1-scores. This indicates limited separability when relying on reflectivity information alone, particularly for water-related surfaces.

The results obtained with the excess Doppler spectrum criteria are presented in Table 4.3. It is shown that there is a consistent improvement in performance for all types of surfaces with an increase in probability of

detection, a decrease in false alarm probability, and an increase in accuracy. The most significant performance improvements are observed on sea and river surfaces, with a marked increase in F1-score.

These results demonstrate the effectiveness of the excess Doppler-based criteria in providing supplementary information with respect to reflectivity-based parameters.



(a) Reflectivity-based parameters (b) Excess Doppler spectrum criteria

FIGURE 4.11: Global confusion matrices obtained on synthetic GNSS-R signals using (a) reflectivity-based parameters and (b) excess Doppler spectrum criteria.

When using reflectivity-based parameters, the confusion matrix shows significant misclassification between several surface types, particularly for sea and river classes. In contrast, the confusion matrix obtained using excess Doppler spectrum criteria demonstrates a clearer diagonal dominance and reduced confusion across all classes. These findings further illustrate the improved class separability achieved through the Doppler-related information.

4.5.3 Application of Excess Doppler Criteria on Real data

The processing sequence involved applying FFT operations on the complex baseband signal of the airborne GNSS-R dataset $I + jQ$ at each stage of coherent integration, such that the processing windows stayed entirely within homogeneous surface parts as characterized by the segmentation indices from Chapter 3. This method enables precise correlation of types of surfaces and respective spectral characteristics, avoiding the artifacts due to the mixed surfaces. The three criteria were calculated according to the mathematical operations laid out in Section 4.4.1.

A key element of this analysis involves comparing the spectral characteristics over distinct surface types, such as the transition from a lake to a land surface, as illustrated in Figure 4.12. The spectral results clearly distinguish between the reflecting surfaces based on scattering mechanisms and energy distribution. For the lake region, the FFT magnitude spectrum is characterized by a dominant frequency component, indicative of coherent reflection. The corresponding phase evolution is smooth, reflecting the preservation of phase coherence over time. In contrast, the land region exhibits multiple distinct frequency components (or peaks), associated with incoherent scattering from rough terrestrial surfaces. The phase evolution in this case is highly variable, indicating loss of coherence due to multiple scattering contributions. These differences in spectral structure and phase behavior highlight the relevance of Doppler-based characteristics for surface discrimination in GNSS-R applications.

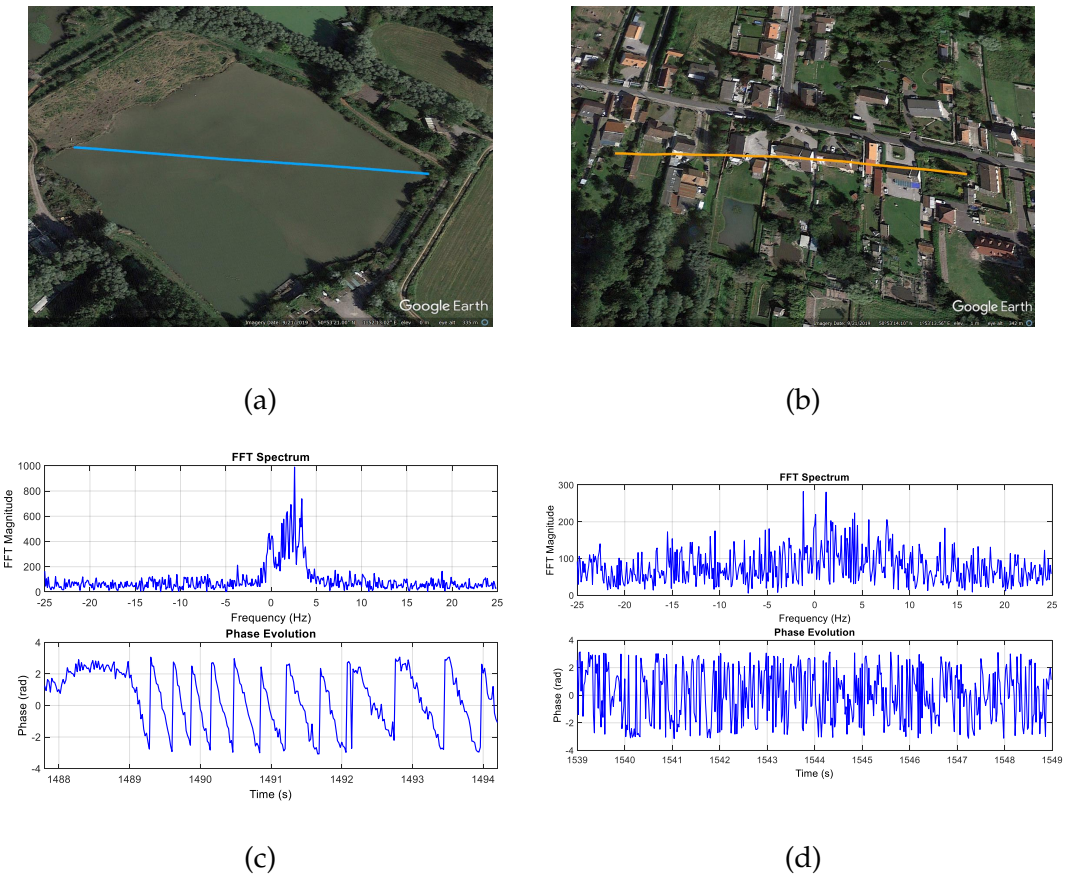


FIGURE 4.12: Comparative FFT spectra and phase evolution extracted from Google Earth data for different surface types: (a) Land region at (50.886062°, 1.889681°) (b) Lake region at (50.888349°, 1.872098°)

4.5.3.1 Criteria Performance Analysis

The performance evaluation of the developed criteria revealed distinct behavioral patterns across different surface types encountered during the airborne experiment. The analysis demonstrated that the excess Doppler characteristics provide complementary information to traditional reflectivity measurements, offering enhanced discrimination capabilities for surface classification applications. The maximum peak amplitude criterion showed particularly strong sensitivity to surface coherence properties, with calm water surfaces exhibiting significantly higher values compared to rough terrestrial surfaces due to the preservation of specular reflection characteristics.

The relationship between surface roughness and spectral characteristics became evident through the analysis of the significant peaks criterion, which effectively captured the transition from coherent to incoherent scattering regimes. Smooth surfaces typically displayed single dominant peaks corresponding to specular reflection, while rougher surfaces exhibited multiple significant peaks arising from diverse scattering paths within the glistening zone. The spectral dispersion criterion provided additional discrimination power by quantifying the energy distribution pattern, with concentrated energy distributions indicating coherent reflections and broader distributions suggesting diffuse scattering mechanisms.

Figure 4.13 illustrates the performance characteristics of the red curve, which represents the main peak values corresponding to the first criterion, while the blue curve illustrates the sum of all significant peaks derived from the signal across the various surface types encountered during the flight experiment. The visualization demonstrates the clear separation achievable between different surface categories, with calm water bodies clustering in regions of high coherence and low spectral complexity, while land surfaces occupy regions characterized by reduced coherence and increased spectral dispersion. This separation provides the foundation for the subsequent classification analysis.

4.5.3.2 Enhanced Surface Classification Through Spectral Analysis

The integration of the three Doppler-based criteria into an unsupervised classification framework represented a significant advancement over the reflectivity-based approach employed in Chapter 3. The K-means clustering

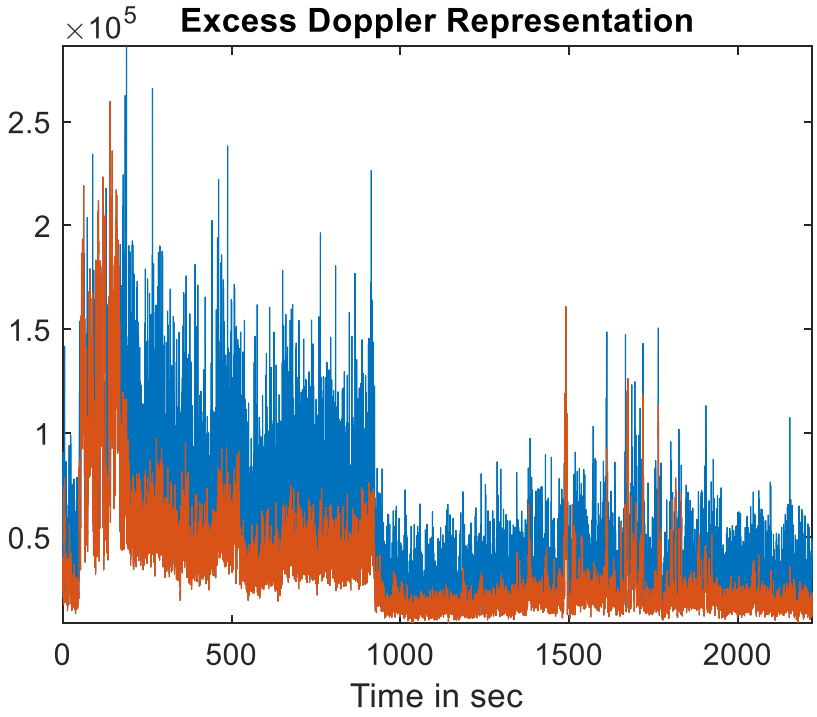


FIGURE 4.13: Excess Doppler spectrum representation of significant peaks (in blue) and maximum peaks (in red), showing peak variations across different surface types.

algorithm was implemented to evaluate the discriminative power of phase and spectral characteristics in surface type identification, utilizing the mean values of the three criteria computed for each homogeneous segment identified through the change point detection methodology.

The classification approach successfully identified four distinct surface classes, depicted in Figure 4.14, corresponding to the major surface types encountered during the flight experiment. Land surfaces were characterized by low coherence indicators and high spectral dispersion values, reflecting the complex scattering mechanisms associated with vegetated and rough terrestrial areas. Sand surfaces exhibited intermediate coherence levels with distinctive spectral signatures that distinguished them from both land and water categories. Sea surfaces demonstrated variable coherence characteristics depending on local wind conditions and surface roughness, with calmer areas showing higher coherence values. Inland water bodies consistently displayed high coherence indicators and concentrated spectral energy, confirming their specular reflection properties.

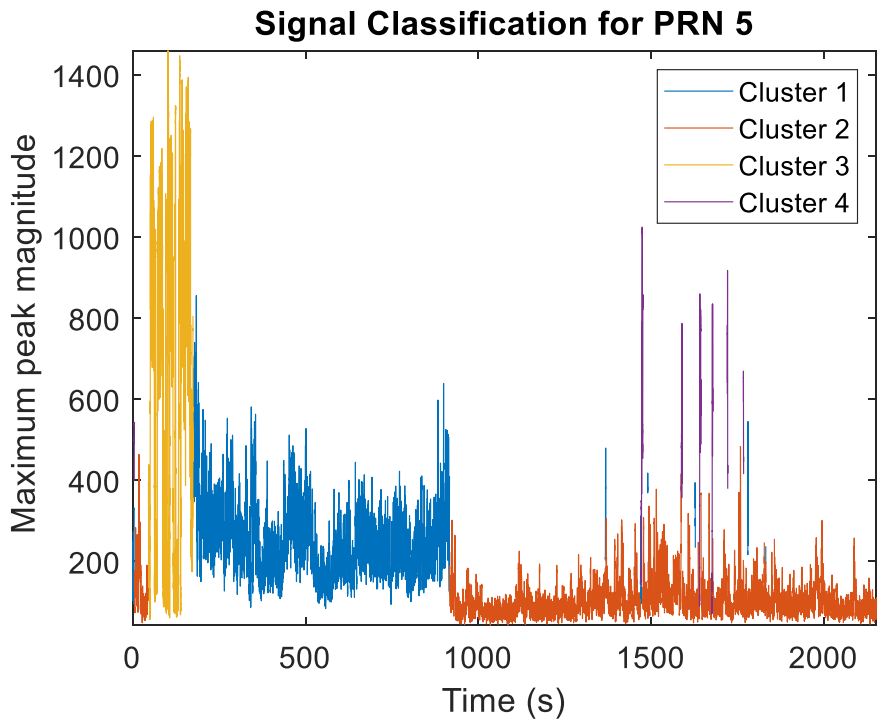


FIGURE 4.14: Classification of the Max Peaks Values.

Compared to reflectivity-based classification methodology, the incorporation of phase and spectral information enabled more reliable discrimination between surface types that exhibited similar reflectivity characteristics but fundamentally different scattering mechanisms. This enhancement was particularly evident in the separation of inland water bodies from other surface types, where the coherence-based criteria provided clearer discrimination than reflectivity measurements alone.

4.6 Conclusion

This chapter develops an integrated framework for GNSS-R phase modeling and excess Doppler analysis and thereby provides the basis for sophisticated phase-aware surface discrimination. The novel method successfully addresses a gap in GNSS-R signal processing by systematically exploiting phase information virtually ignored in conventional amplitude-centered methodologies.

Major theoretical contribution revolves around statistical phase model founded on von Mises distribution framework, describing transition from coherent to incoherent regimes of scattering as controlled by signal-to-noise ratio and surface characteristics. Obtained from principles by going through

whole signal processing chain, such a model offers rigorous mathematical backgrounds for operating characteristics of surfaces through phase behavior.

The processing of the excess Doppler spectrum yields three quantitative parameters that effectively characterize surface scattering: the maximum amplitude of the dominant Doppler component, the number of spectral peaks statistically significantly above the noise, and the degree of spectral dispersion. They provide supplementary information in addition to traditional reflectivity measurements and demonstrate greater discriminatory ability for surface classification-related applications.

Experimental evaluation is first conducted on synthetic GNSS-R data generated from the statistical I/Q model, enabling evaluation of the proposed criteria under controlled signal-to-noise conditions. Experimental verification performed using real airborne GNSS-R data confirms the effectiveness of the proposed framework. The apparent distinction behavior between motion over sandy surfaces and motion over rough surfaces validates both the theory and the practical implementation. An effective comparison between measured and estimated phase observations over coherent segments across a lake demonstrates the accuracy of the I/Q-based estimation method for phase measurements. The sophisticated surface classification outputs demonstrate increased discrimination among surfaces having similar reflectivity characteristics, which holds particularly true for lake waters where other methodologies are deficient. The multi-path reflection model method for creating synthesized data provides a priceless asset for algorithm development in cases where training data with appropriate labels is not abundant. The demonstrated fidelity between real and synthesized signal characteristics validates the physical models upon which these are based.

The described works lay the foundation both for theory and practical implementation issues concerning applications of machine learning from Chapter 5. They constitute fundamental prerequisites for building sophisticated automated classification systems whose full potential use of GNSS-R signals extends capabilities of Earth observation.

Chapter 5

Machine Learning Applications for GNSS-R Surface Classification

5.1 Introduction

The development of GNSS-R from demonstration of the concept to advanced remote sensing applications has been matched by a corresponding evolution of signal processing techniques. Although the previous chapters have laid the theoretical ground for phase-aware GNSS-R processing and formulated advanced statistical paradigms for signal segmentation and surface characterization, the real-world complexity of surface classification scenarios requires computational methods capable of managing the nonlinear relationship between the surface characteristics and the corresponding GNSS-R observables. This chapter presents the use of machine learning (ML) algorithms for the surface classification problem based on the previously derived phase-based signal models and the framework for generating the corresponding synthetic data presented in Chapter 4.

The primary problem in GNSS-R surface classification is the complex interdependencies between the properties of reflected signals and the scattering mechanisms that influence surface-signal interaction. Traditional techniques have been reliant on electromagnetic modeling and semi-empirical relationships which, though providing insightful physical interpretation, often struggle to detect surface variability under different environmental conditions. The statistical approach for phase modeling described in Chapter 4 revealed that surface-phase behavior provides rich information about coherent and incoherent scattering regimes for classification purposes, and it requires sophisticated pattern recognition capabilities that exceed the abilities of analytical methods.

Machine learning (ML) methods provide a compelling solution by presenting data-based methods capable of learning nonlinear mappings between input features and target classes without the need for explicit mathematical expressions. The ML algorithms can automatically learn patterns by operating in high-dimensional feature spaces, making them especially suited for GNSS-R applications due to the generation of complicated decision boundaries by the many physical parameters involved.

The proposed synthetic data generation approach addresses the widespread lack of large, labeled datasets in GNSS-R applications by providing the generation of comprehensive training datasets with complete ground-truth labeling for supervised learning applications. In this work, the testing framework is based on the airborne GNSS-R dataset from the Northern France campaign, augmented by the phase model-based generation of a synthetic dataset. This combination leads to a comprehensive evaluation of ML algorithms while addressing ground-truth limitations. The successful transfer from synthetic to real domains confirms the validity of the physical models used and offers a generalized extension for other remote sensing modalities.

The choice of the ML approach matches the characteristics of the GNSS-R data and real-time processing requirements, from traditional approaches like Random Forest (RF) and Support Vector Machines (SVM) to deep learning (DL) architectures like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) based ones that can automatically extract relevant features from raw signal data.

The chapter structure reflects a systematic development and evaluation of the ML-based classification framework. Following this introduction, Section 5.2 reviews machine learning applications in GNSS-R and related remote sensing domains. Section 5.3 presents the dual-dataset approach by merging the generated synthetic data with the real airborne measurements. Section 5.4 summarizes the preprocessing of the data, while Section 5.5 shows the chosen algorithms varying from classical approaches to deep learning models. Section 5.6 summarizes the training and the evaluation plans. Section 5.7 presents experimental results demonstrating classification performance on both synthetic and real datasets. Finally, Section 5.8 concludes with key findings and research directions.

5.2 Literature Review: Machine Learning in GNSS-R and Earth Observation

5.2.1 Machine Learning in Spaceborne GNSS-R

The integration of ML techniques with GNSS-R presents a paradigm shift in the methodologies used for Earth observation and remote sensing. As highlighted in [106], the emergence of multiple spaceborne GNSS-R missions and the simultaneous exponential growth in available datasets have created unprecedented opportunities for applying advanced computational techniques to extract geophysical information from backscattered GNSS signals. The Technology Demonstration Satellite (TDS-1), Cyclone Global Navigation Satellite System (CYGNSS), Bufeng-1 A/B, and Fengyun-3 series have together created datasets that enable machine learning (ML) and deep learning (DL) techniques to be applied across a variety of disciplines such as oceanography, land studies, cryospheric studies, atmospheric studies, hydrological studies, and environmental monitoring.

The fundamental advantage of ML approaches in GNSS-R applications lies in their ability to handle the inherent complexity and nonlinear relationships within reflected signal data. Traditional GNSS-R processing methods, while providing valuable physical insights through electromagnetic modeling and semi-empirical relationships, often struggle to fully exploit the information content available in modern high-resolution datasets. Machine learning algorithms, particularly deep learning architectures, offer the capability to automatically discover complex patterns and relationships that may not be readily apparent through conventional analytical approaches.

Bu et al. [106] establish their comprehensive review as the first literature specifically focusing on ML and DL algorithms within the spaceborne GNSS-R context for Earth observation and remote sensing, highlighting the transformative potential of these techniques in addressing complex, large-scale challenges across various remote sensing applications. This foundational work demonstrates that ML methods have become crucial for extracting valuable information from GNSS-R signals, enabling the retrieval of various geophysical parameters with improved accuracy and efficiency compared to traditional methods.

5.2.2 ML Applications in Ocean Remote Sensing

The ocean domain is one of the most developed applications of the techniques of ML to spaceborne GNSS-R. ANNs and CNNs have also shown important potential for accurately estimating wind speed at the sea surface. The ability of the networks to create intricate, nonlinear relationships between GNSS-R observables and ocean parameters was particularly desirable for wind speed retrieval.

Deep learning structures have led to significant improvements in significant wave height estimation, with CNN-ConvLSTM-FCN [12] and Deep CNN (DCNN) [107] achieving swell height root-mean-square error (RMSE) values of 0.32 m and below. These enhancements show that DL techniques may extract subtle information from Delay-Doppler Maps (DDMs) that correlate with wave properties. The success of these applications has established ocean remote sensing as a proving ground for advanced ML techniques in GNSS-R.

Sea surface salinity retrieval has also been improved using ML techniques, especially ANNs and Multi-Layer Perceptrons (MLPs) for simulating the complex nonlinear relations between different environmental parameters and salinity [108]. A significant advancement in high-precision salinity retrieval was achieved by integrating GNSS-R and L-band microwave radiometric measurements with a machine learning model, which successfully reduced the retrieval RMSE from 0.63 to 0.43 psu and increased the correlation coefficient R_c from 0.52 to 0.87 [109].

For estimating sea surface height, [110] shows the application of two of the most outstanding ML techniques: Principal Component Analysis combined with Support Vector Regression (PCA-SVR) and Convolutional Neural Networks (CNN). Based on TDS-1 data and cross-validated using the DTU15 sea surface height model, the PCA-SVR method resulted in a Mean Absolute Error (MAE) of 0.61 m and RMSE of 1.72 m, whereas the CNN model resulted in MAE of 0.71 m, RMSE of 1.27 m.

5.2.3 Land Surface Applications and Soil Moisture Retrieval

The use of ML techniques in GNSS-R for land surface studies is growing rapidly as well, with soil moisture retrieval emerging as the most widely

explored application. Due to the complexity of scattering over the land surface and its sensitivity to vegetation, topography, and soil parameters, this area is especially suitable for the application of ML methods capable of learning from multidimensional correlations between GNSS-R observables and surface parameters.

5.2.3.1 Ensemble Methods and Traditional ML Approaches

The effectiveness of the XGBoost algorithm in soil moisture retrieval was demonstrated in [111], focusing on its ability to handle complex, nonlinear regression problems inherent in GNSS-R missions. The superiority of XGBoost comes from its second-order Taylor expansion for loss function minimization, its capacity to handle missing and sparse data, and its ability to avoid overfitting with complexity regulation [112–114]. The study systematically evaluated input variable importance and found invariably the SNR to be the dominant parameter responsible for 40–70% of retrieval accuracy [111].

Random Forest (RF) and Support Vector Machine (SVM) techniques have also long been researched for soil moisture applications [115, 116]. Modeling and theoretical analysis of GNSS-R Soil Moisture Retrieval Based on the Random Forest and Support Vector Machine Learning Approach presents the fact that RF is advantageous over SVM in addressing the inversion problem for soil moisture retrieval with high accuracy and universality, especially when soil textures are non-homogeneous or unknown [117].

5.2.3.2 Deep Learning Approaches for Soil Moisture

A comprehensive machine learning framework applying Random Forest regression to extract soil moisture content from CYGNSS measurements was presented in [115, 118]. Their method fully leverages physics-based variables extracted from CYGNSS measurements, i.e., reflectivity, specular point incidence angle, and Trailing Edge Slope, in addition to supplementary data such as Normalized Difference Vegetation Index, Vegetation Water Content, elevation, and soil texture. The experiment results in a mean unbiased RMSE of 0.044 cm^3/cm^3 and a correlation coefficient R_c of 0.66 compared to SMAP observations [115].

[119] extended the applicability of the study by applying different machine learning techniques, especially ANN, RF, and Support Vector Regressors (SVR), to extract soil moisture using CYGNSS data. Their ANN model resulted in a correlation coefficient R_c of 0.89 and an RMSE of 0.066 m^3/m^3, with the RF model showing slightly improved performance with an R_c of 0.91 and an RMSE of 0.06 m^3/m^3.

5.2.3.3 Vegetation and Biomass Monitoring

ML algorithms also have immense possibilities for the estimation of vegetation parameters, whose potential applications include Above Ground Biomass, Canopy Height, and Vegetation Water Content. [120] depicts Above Ground Biomass retrieval from ANN algorithms using CYGNSS observables and auxiliary data, with R_c = 0.91 and RMSE = 56.7 t/ha when trained on 10% of the used data. This demonstrates the possibility of retrieving vegetation-related information from GNSS-R signals with the aid of the ML methods and indicates a promising prospect for applications of global vegetation monitoring.

5.2.4 Cryosphere Applications and Sea Ice Classification

The cryosphere application is also an important domain of application for ML with GNSS-R, and sea ice classification is a notably mature application area. Both the sensitivity of the GNSS signal to various types of ice and surface conditions and the difficult scattering processes involved make the cryosphere application area well-suited for an application of ML.

5.2.4.1 Sea Ice Type Classification

[121] conducts a detailed analysis of spaceborne GNSS-R for the classification of sea-ice using Random Forest and Support Vector Machine classifiers. Their work focuses on discriminating open water, first-year ice and multi-year ice from eight shortlisted GNSS-R features extracted from TDS-1 data. A two-class classification method is adopted, in which open water is discriminated from sea-ice first with accuracy of 98.83% for RF and 98.60% for SVM, and then first/multi-year ice are classified with accuracy of 84.82% for RF and 71.71% for SVM. Accuracy is defined in equation **??** in this chapter.

The method unveils the intrinsic role of feature engineering for GNSS-R systems and takes advantage of surface reflectivity, Delay-Doppler Map Average, and different waveform-based features, such as the normalized

central delay waveform, the normalized integrated delay waveform, and the differential delay waveform [122]. The features are able to obtain subtle variations of surface roughness and scattering properties and denote different types of sea ice.

5.2.4.2 Advanced Classification Techniques

Machine learning approaches have successfully been applied to sea ice remote sensing using SAR [123], altimeter [124], and passive microwave data [125]. In the GNSS-R domain, applying higher-level ML techniques to sea ice classification has shown particular promise. DNNs and CNNs have been employed in Sea Ice Concentration retrieval to considerable effect relative to traditional methods [126–128]. Sea Ice Thickness from GNSS-R remains an emerging area of application; however, ML methods such as CNN and SVR have been found useful to retrieve Sea Ice Thickness with good consistency with reference measurements.

A systematic review of sea ice remote sensing using GNSS-R recognizes a growing application of machine learning techniques [129]. This work notes that ML-based systems, particularly CNNs and SVMs, achieve high accuracy in detecting sea ice, often exceeding 97-98%. CNNs demonstrate particular strengths derived from convolutional layers capable of reducing DDM noise and preserving spatial relations.

5.2.5 Inland Water Body Detection and Hydrological Applications

Unsupervised learning methods have been found to demonstrate notable potential in both the detection and observation of inland water bodies. The application of K-means, agglomerative, and Density-based Spatial Clustering (DBSCAN) clustering methods to bistatic radar cross-section (BRCS) and SNR parameters derived from CYGNSS data has been proven to lead to the detection of inland water bodies.

One study reveals that K-Means achieved the highest accuracy (93.5%), while DBSCAN yielded the highest recall (83.1%) for water body detection [130]. Importantly, these methods successfully identified small tributaries that are often missed by lower-resolution land cover classification maps like those from Copernicus. This demonstrates the potential of GNSS-R combined

with ML techniques for high-resolution hydrological monitoring applications.

The research highlights both the capabilities and limitations of unsupervised approaches. Though the methods effectively capture water bodies at a high percentage correct, weaknesses are experienced in false positives in desert areas of near-flat topography and complexities in large waters affected by roughening caused by wind. These results have significant implications for designing better classification methods and characterizing the physical limitations of GNSS-R for the detection of water.

5.2.6 Forest Fire Disturbance Detection

The scope of application has been expanded to include the detection of forest fire disturbance using SVM, RF, and ANN to apply to classification as well as Burned Fraction extraction [119]. SVM and RF provided high detection efficiencies (higher than 80%) along with very low false positives (less than 1%). ANN regressor showed superior performance compared to other methods in Burned Fraction extraction, resulting in $R_c = 0.83$ compared to RF ($R_c = 0.78$). This application shows the versatility of ML techniques to apply to diverse environmental monitoring tasks through the utilization of GNSS-R data.

5.2.7 Methodological Advances and Algorithm Development

5.2.7.1 Traditional Machine Learning Algorithms

The literature provides a deep study of traditional machine learning methodologies employed in GNSS-R problems. Linear models like Polynomial Regression, Stepwise Linear Regression (SLR) and Ordinary Least Squares (OLS) set a baseline performance against which superior methodologies can be compared [106, 131, 132]. Support Vector Machines (SVM) with various kernel functions [133] have been extensively used to carry out classification and regression processes and are significantly effective in high-dimensional feature spaces typical of GNSS-R tasks.

Decision Trees like ID3, C4.5, and classification and regression trees (CART) have interpretability advantages but are prone to overfitting [134]. The evolution toward ensemble methods solves these limitations through techniques like bagging and boosting.

5.2.7.2 Ensemble Methods

Ensemble methods have proved to be particularly effective for GNSS-R scenarios. Averaging techniques of Random Forest, Bootstrap Aggregating, and Evolutionary Trees yield consistently good performances with various datasets [135]. Boosting techniques of AdaBoost, Gradient Boosting Decision Tree (GBDT), XGBoost, and LightGBM have proved their dominance over challenging retrieval situations [136].

The effectiveness of ensemble approaches to GNSS-R applications can be traced to their capability of addressing variability and noise inherent to reflected signal measurement while offering generalization to new data that is robust. The integration of several weak learners gives rise to outputs that are more stable and accurate, an effect that is very significant considering the diverse environmental conditions that affect GNSS-R signal characteristics.

5.2.7.3 Deep Learning Architectures

Deep learning techniques constitute the current frontier of GNSS-R signal processing. Multi-layer perceptrons and Artificial Neural Networks are the basis of DL schemes [137, 138]. CNNs were found to be very efficient for DDM data processing, with their capability to maintain spatial relations and hierarchical feature extraction.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks enable the analysis of the temporal features inherent to GNSS-R data to inspect time series and sequential patterns [139, 140]. Deep Convolutional Neural Networks (DCNN), with their increased layer counts, offer increased capability of learning abstract features within intricate patterns of signal [141].

New advances employ Transformer frameworks with self-attention mechanisms to handle sequence data and learn long-range relationships [142]. Hybrid models employing mixed structures (e.g., CNN-LSTM, CNN-MLP) take advantage of multiple approaches to achieve superior outputs [143].

5.2.8 Performance Evaluation and Comparative Analysis

5.2.8.1 Accuracy and Generalization

Comparative analyses of various ML strategies demonstrate consistent patterns of performance characteristics. Ensemble strategies such as Random Forest always result in correlation coefficients $R_c > 0.9$ with typically less than a 0.06 m^3/m^3 RMSE while tested against reference datasets [117, 119]. The deep learning strategies achieve comparable or higher levels of accuracy while offering the benefit of being able to automatically extract their features and process raw signal data.

Sea ice classification applications demonstrate the highest accuracy rates, with CNN and SVM approaches achieving 97–98% accuracy for basic ice/water discrimination [129]. More challenging tasks like FYI/MYI classification show reduced but still significant accuracy rates of 80-85% for the best-performing algorithms.

5.2.8.2 Computational Considerations

The computational speed of various ML strategies varies considerably, with significant operational implications. Classical strategies such as SVM and Random Forest provide short training times and low computational demands; therefore, they are applicable to real-time or near-real-time processing-based applications [106]. Deep learning strategies, while with large training computational demands, offer efficient inference with training and can provide better performance with intricate pattern recognition-based tasks.

Accuracy vs. computation efficiency trade-offs remain an area of research activity, with hybrid approaches seeking to bring together the desirable features of different approaches. The improvements in efficient architecture and optimization schemes keep broadening the scope of practical application of advanced ML methods to operational GNSS-R systems.

5.2.9 Challenges and Future Directions

5.2.9.1 Data Availability and Quality

A persistent challenge across GNSS-R ML applications is the limited availability of high-quality, labeled training data. The literature review indicates that research is mostly conducted using relatively small datasets or data

simulation to complement real measurements [115]. Model generalizability is impaired and requires verification schemes to achieve robust output against widely diverging environmental conditions.

The temporal and spatial uncertainty of the surface conditions and coverage of ground truth observations place specific requirements on developing globally transferable ML models. Future research opportunities are available for developing transfer learning approaches and domain adaptation methods to overcome these data limitations.

5.2.9.2 Physical Interpretation and Model Interpretability

Although machine learning methods have found higher efficiency within various GNSS-R applications, the black-box characteristics of a majority of algorithms limit thorough knowledge and interpretation of outcomes. Merging explainable artificial intelligence (XAI) methods with physics-informed machine learning strategies is one of the crucial future research directions within this domain.

The evolution of hybrid strategies that link physical modeling to data-driven learning promises to preserve interpretability while benefiting the pattern recognition capabilities of ML algorithms. Such strategies could afford better extrapolation to conditions not included in the set of training cases and afford the precision advantages of machine learning strategies.

5.2.9.3 Multi-Mission and Multi-Sensor Integration

Future developments of GNSS-R ML application will probably focus on the integration of data between multiple missions and sensor combinations to enhance coverage and retrieval accuracy. Synergistic exploitation of GNSS-R with other remote sensing modalities, including optical and SAR sensors, will potentially facilitate better characterization of the surface and validation [119].

The standardization of processing methods and the building of interoperable ML models among various GNSS-R missions will play a key role in advancing the domain to operational applications. This necessitates reconciling variation among various satellite systems in terms of instrument parameters, processing algorithms, and data formats.

The literature review demonstrates that machine learning has been an important GNSS-R data analysis and interpretation tool in various scenarios of Earth observation. Beginning with the earlier works that developed CNN and SVM-based schemes for basic parameters, the research area has rapidly evolved to use sophisticated ensemble methods and deep learning frameworks to derive higher-order surface classifications. The persistent finer improvements in accuracy, the capability to tackle intricate nonlinear connections, and the potential to automate feature extraction make the ML strategies key elements of future GNSS-R processing systems.

5.3 Data Collection and Generation

This section explains the collection of the datasets utilized for training, validating, and testing the supervised machine learning models of this chapter. The main difficulty associated with using supervised learning for GNSS-R classification is the limited availability of large, diverse, and accurately labeled real-world data. To address this difficulty, a dual-dataset approach was used. A large synthetic dataset, based on physical properties, was simulated to act as the main training and validation set. At the same time, a set of real airborne GNSS-R measurements was used twice: to provide guidance to the generation of realistic synthetic data and to serve as a final, independent testing dataset.

5.3.1 Real Airborne GNSS-R Dataset

The foundation of this experimental validation is a set of real-world GNSS-R data; the collection and instrumentation are outlined thoroughly in Section 1.4.1 of this thesis. To summarize, the data was gathered during a flight experiment over Northern France on October 19, 2020, using a gyrocopter-based system at an altitude of 315 m. The flight path covered approximately 70 km and included a variety of surface types, such as sandy regions, open sea, land, and over 50 inland water bodies.

For the machine learning applications described in this chapter, this dataset played two important roles. Primarily, we processed segments of the real data that corresponded to known types of surfaces to determine important statistical and physical parameters, including the number of reflections, characteristics of the Doppler frequency, and reflection amplitudes, in accordance with the Excess Doppler analysis presented in Chapter 4.

All of these factors were important for ensuring that the synthetic dataset was physically valid. Additionally, to ensure validation, the entire airborne dataset was reserved as the testing set.

The labels of the real GNSS-R signals were determined using ground truth observations, Google Earth imagery, and detailed IGN (Institut Géographique National) maps.

5.3.2 Synthetic Signal Generation

The synthetic dataset was created with the signal model (I/Q) presented in Section 4.5.1.2, which models GNSS-R reflections by a coherent sum of various components. To model surface-dependent realistic signals, important parameters were initially derived from actual data segments, such as the amount of reflection, their corresponding Doppler frequencies, and reflection intensities.

Using this model, a total of 1000 synthetic signal segments were generated for the purposes of supervised learning. In contrast to the synthetic signals used in Chapter 4 for model evaluation, the synthetic dataset employed here was designed to introduce additional variability representative of real measurement conditions. In particular, both the segment duration and the signal-to-noise ratio (SNR) were varied across the generated segments, while relying on the same underlying physical model and parameter ranges derived from real data.

The generated signals were evaluated by comparing their spectral behavior to that of real GNSS-R observations. The similarity in Doppler-domain structure across both datasets reinforces the physical validity of the model and confirms the suitability of the synthetic data for supervised training and evaluation.

5.3.3 Feature Engineering

To derive distinguishing features from GNSS-R reflections, we initially divided the signals into continuous segments that align with stable surface labels. For each segment, we considered two input channels: the raw amplitude time series and the max spectral peaks derived from the Fourier

transform of the complex signal.

We calculated a wide range of statistical features for both the amplitude signal and the max spectral peaks representation. These included count, sum, mean, standard deviation, minimum, maximum, range (maximum-minimum), coefficient of variation (standard deviation/mean), skewness, kurtosis (according to Fisher's definition), median, and interquartile range (75th percentile to 25th percentile).

The characteristics of these features include both time-domain texture and frequency-domain structure, offering a comprehensive and interpretable representation of the reflected signals across various surface types.

To assess the separability of classes in this enriched feature space, we generated two key scatter plots:

- The analysis of the coefficient of variation between the amplitude and the spectral peaks demonstrated the texture-driven variability that differentiates segments of sand, sea, land, and river (Figure 5.1).
- The relationship between the skewness and kurtosis of the amplitude revealed variations in waveform asymmetry and peakedness among different surface classes (Figure 5.2).

Using skewness and kurtosis specifically helped distinguish sea from river sections, which showed quite similar values depending just on fundamental metrics, including mean, maximum, or standard deviation. This made it possible to separate the two classes that would otherwise overlap. On the other hand, when using the statistical tests, sand and land classes were already rather identifiable.

These preliminary visualizations verified that the designed features contain significant class-specific information.

In contrast, the frameworks of deep learning (Bi-LSTM and 1D-CNN) utilized an end-to-end learning mechanism to avoid the necessity of manual feature engineering. In the case of these structures, fixed-sized windows of the input channels were taken from each segment and fed directly to the network to allow it to learn the relevant features automatically.

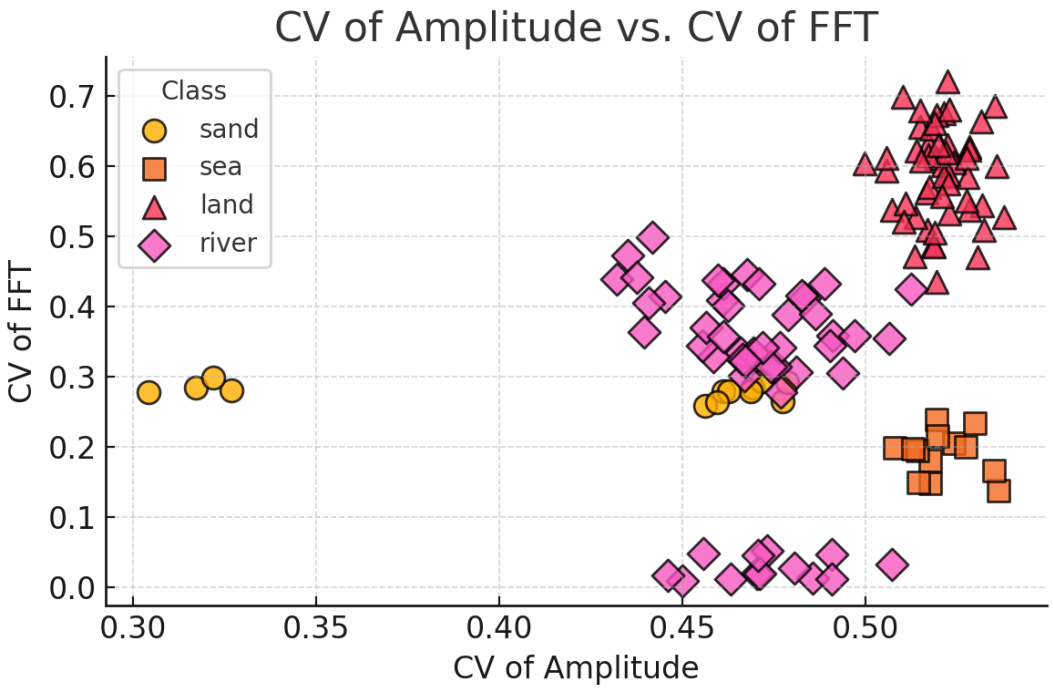


FIGURE 5.1: Coefficient of variation (CV) of the amplitude versus CV of the spectral peaks for different surface types.

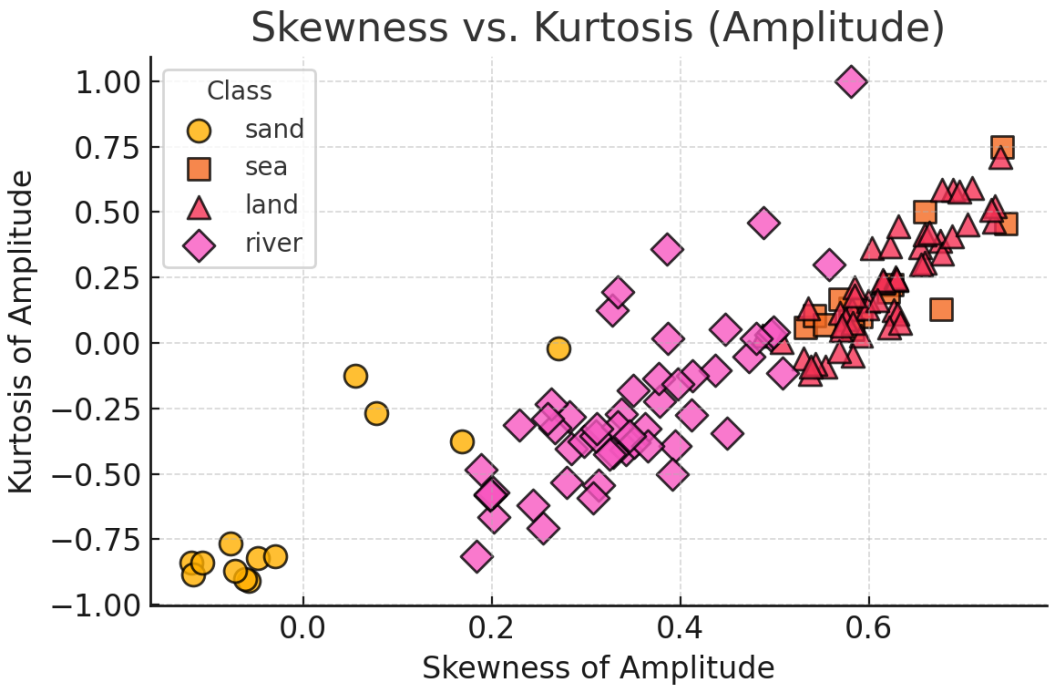


FIGURE 5.2: Amplitude skewness versus kurtosis across surface classes.

Finally, each segment of the two datasets received a label that belonged to one of four classes of surfaces. The complete synthetic dataset was used during the model training, with a 5-fold cross-validation process held out to

allow hyperparameter tuning.

5.4 Data Preprocessing

A set of preprocessing techniques were implemented to prepare the designed features for the machine learning models. To maintain uniformity among features and avoid the influence of specific variables dominating the learning process, numerical features were normalized to achieve a mean of zero and unit variance, using the standard scaling formula:

$$x_i' = \frac{x_i - \mu_i}{\sigma_i} \tag{5.1}$$

where μ_i and σ_i are the mean and standard deviation of feature x_i throughout the dataset. This transformation ensures that all features contribute equally to the classification process.

In addition, Min-Max normalization was applied to adjust features within the range $[0, 1]$. This method is particularly useful for certain algorithms, such as neural networks, as it ensures numerical stability and can improve the convergence of the training process while preserving the relative differences between values.

Since machine learning algorithms require numerical inputs, categorical surface type labels (sand, sea, land, river) were converted into their corresponding integer representations.

5.5 Model Selection

A variety of supervised learning models have been investigated for the classification of GNSS-R reflected signals, using features based on amplitude and spectral peak representations. The analysis included both traditional machine learning methods and deep learning architectures to provide a comprehensive comparison.

The classical, feature-based models selected for this study include:

- **Logistic Regression:** An effective linear classifier commonly used as a baseline. It applies to binary classification but can easily be extended to the multi-class framework.

- **Random Forest:** An algorithm that combines multiple decision trees, resulting in a forest of trees. It creates decision trees on the given data samples, gets prediction from each tree and selects the best solution by means of voting.
- **XGBoost (eXtreme Gradient Boosting):** A powerful and efficient implementation of gradient boosting based on sequential ensemble learning with decision trees. It iteratively corrects errors from previous models and incorporates advanced techniques like regularization.
- **Multilayer Perceptron (MLP):** A classic feedforward neural network consisting of fully connected layers capable of modeling non-linear relationships between features.

Additionally, two end-to-end deep learning models were also implemented, which learn features directly from the raw, segmented time-series data:

- **1D Convolutional Neural Network (1D-CNN):** A neural network architecture specifically designed to process one-dimensional sequential data, making it ideal for identifying local patterns within the GNSS-R signal segments.
- **Bidirectional Long Short-Term Memory (Bi-LSTM):** A type of recurrent neural network architecture that captures dependencies in both forward and backward directions across sequences.

This diverse selection enables a comparative analysis between traditional feature-based classifiers and deep learning approaches for GNSS-R surface type classification.

5.6 Model Training and Evaluation

5.6.1 Experimental Setup and Training Procedure

Training and testing for all machine learning and deep learning models were performed on the Google Colab platform, leveraging its cloud-based GPU computing capabilities to meet the computational requirements of the experiments.

The supervised models were only trained on the synthetically generated dataset. To allow for a fair and unbiased evaluation, the dataset was first split as a training set and a held-out testing set. This was an intentional selection as a means of determining if the models could learn the raw properties of the GNSS-R signals in a controlled physical environment and then apply this knowledge to real-life scenarios.

The training procedure followed two different paths based on the type of model:

- Classical Models (Random Forest, MLP, Logistic Regression, XGBoost): These models were trained on the preprocessed feature set built on the synthetic signals. For these models, a 5-fold cross-validation technique was employed on the synthetic training set to find their optimal hyperparameters.
- Deep Learning Models (1D-CNN, Bi-LSTM): These models were trained on raw synthetic signal channels' fixed-length segments without undergoing the manual feature engineering step, with end-to-end feature learning allowed. Both deep learning networks were trained using the Adam optimizer, with early stopping on validation loss to prevent overfitting. Their hyperparameters were tuned via a 5-fold cross-validation grid search over the learning rate, dropout rate, number of filters (for the 1D-CNN), and hidden size (for the Bi-LSTM).

5.6.2 Evaluation Strategy and Metrics

The trained models were evaluated by a two-stage process system constructed such that internal consistency could be determined as an index of real-world generalization:

Synthetic Data Testing: The models were then tested against a held-out set of the synthetic data. This step was used to confirm the models' ability to learn as well as set a baseline for performance in an ideal, noise-managed setting.

Real Data Validation: The trained models were evaluated on the full real airborne dataset, which was kept entirely unseen during training. It was an important step that assessed models' ability to transfer their learned

representations from synthetic to real-world signals.

To quantify and compare the accuracy in terms of classification, two standard metrics were used: Accuracy and the F1-score, defined in Chapter 4. These metrics were calculated for each model's performance on both the synthetic and real test sets, providing a comprehensive basis for comparison.

5.7 Results

The performance of the six supervised learning models, trained exclusively on the synthetic training dataset, was evaluated on both the held-out synthetic test set and the complete real-world airborne GNSS-R dataset. The comprehensive results, quantified by Accuracy and F1-score, are presented in Table 5.1.

5.7.1 Model Performance on Synthetic and Real Data

Table 5.1 shows that nearly every model achieved very high performance levels on the synthetic test set, with a majority reaching an accuracy and F1-score above 0.98. The implication here is that the models were effective at learning the underlying patterns and class distinctions characteristic of the data.

TABLE 5.1: Performance comparison of machine learning and deep learning models on synthetic and real GNSS-R data.

	Synthetic Testing Data		**Real Testing Data**	
Model	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**
Logistic Regression	0.98	0.98	0.85	0.83
Random Forest	0.99	0.99	0.88	0.86
XGBoost	0.99	0.99	0.90	0.89
MLP	0.98	0.98	0.87	0.85
1D-CNN	0.98	0.98	0.90	0.88
Bi-LSTM	0.93	0.92	0.82	0.80

The more critical test was the evaluation on the real airborne dataset, which measured the models' generalization capability. As expected, there was a performance drop for all models when transitioning from the controlled synthetic environment to the more complex real-world signals.

Despite the presence of this "sim-to-real" gap, several models showed remarkable generalization ability. The XGBoost and 1D-CNN models emerged as top performers by achieving an accuracy of 0.90 when put to the real dataset test. The Random Forest (0.88) and MLP (0.87) models also showed excellent performance, thus validating the quality of engineered features. In contrast, the Bi-LSTM model showed the greatest degradation in its performance, suggesting a vulnerability towards overfitting with regard to the specific sequential patterns of the synthetic dataset.

5.7.2 Visual Comparison of Generalization

The difference in performance between the synthetic and real test sets is visualized in Figures 5.3 and 5.4. The plots illustrate the high and consistent accuracy and F1-scores of all models on the synthetic data, followed by the varied levels of success in generalizing to the real data. It visually confirms that the XGBoost and 1D-CNN models maintained the highest accuracy in the real-world application.

5.8 Conclusion

The chapter reflects on the efficient application of machine learning approaches in GNSS-R surface classification by combining signal modeling with data-driven pattern recognition. The integration of traditional machine learning methods and deep learning models provides valuable insight into automated methods for bistatic radar reflectometry classification.

The dual-dataset approach addresses a persistent challenge for GNSS-R remote sensing: the lack of large, accurately labeled training datasets. By generating synthetic data for training, using the physical signal models outlined in Chapter 4, while preserving the real dataset for independent validation, this approach forms a rigorous framework for evaluating generalization abilities. The demonstrated success confirms both the validity

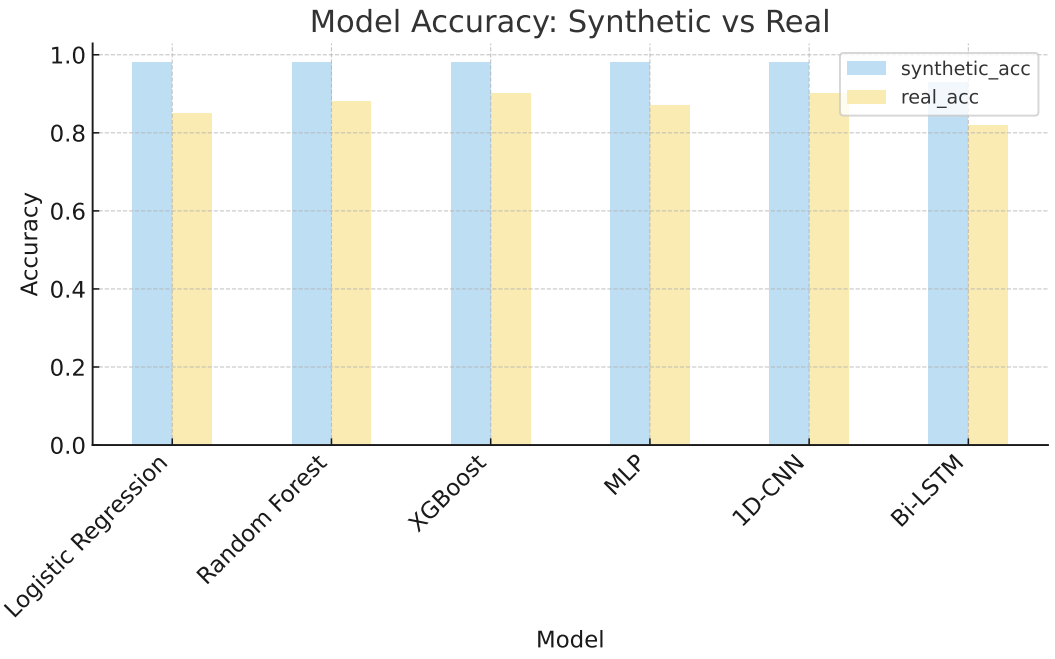


FIGURE 5.3: Model accuracy comparison between synthetic and real testing data.

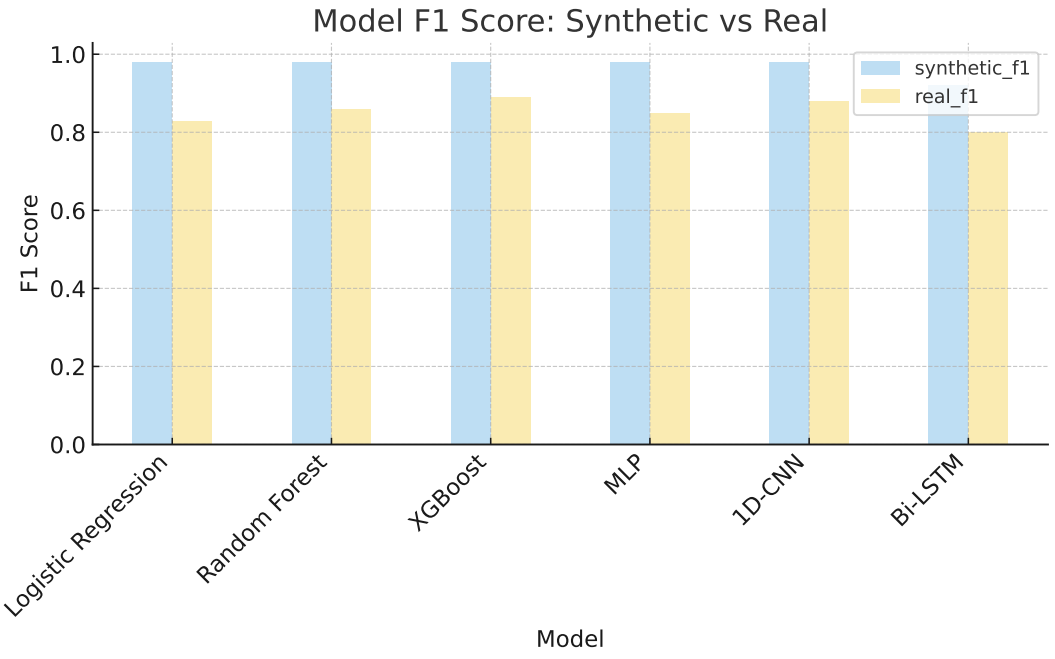


FIGURE 5.4: Model F1-score comparison between synthetic and real testing data.

of the ground physical models and the robustness of the model architectures, as top models achieved 90% accuracy on real data after being trained only on synthetic data.

The comparative analysis reveals important patterns for selecting algorithms, with ensemble methods, such as XGBoost and Random Forest, demonstrating strong performance and robustness across both datasets. Deep learning techniques, particularly 1D-CNN, offer automatic feature extraction, matching traditional methods.

The feature engineering method that integrates time-domain texture metrics and frequency-domain spectral features has shown significant efficiency. Coefficient of variation, skewness, and kurtosis-based statistical moments were capable of separating otherwise similar surface types successfully, as they showed the potential of the features for discrimination between coherent and incoherent scattering. These engineered features provide interpretable classification criteria in addition to deep learning techniques.

Future research directions in the field of applying machine learning to surface characterization might include domain adaptation techniques, semi-supervised learning approaches, physics-informed neural networks, and multi-temporal analysis for change detection applications.

Based on Chapter 3 segmentation modeling and Chapter 4 phase modeling, the contributions introduce operational GNSS-R classification as an applicable use of machine learning and create an end-to-end processing chain for raw signal inputs through to full automatic surface identification, advancing bistatic radar remote sensing capabilities for Earth observation.

General Conclusion

This thesis offered advanced processing methods for GNSS-R signals to automatically identify different surfaces using airborne GNSS-R data. We established a complete framework that combines statistical model-based segmentation using reflectivity, phase signal processing, and machine learning techniques for enhanced surface differentiation. Experimental campaign observations of airborne GNSS-R collected in October 2020 above Northern France served as the validation set, providing high-resolution observations for diverse surface cover like sandy beaches, open sea, inland water bodies, and land cover.

The thesis contained five chapters. The first chapter included an overview of the evolution of GNSS-R technology, from initial concept experiments to operational remote sensing applications. We addressed bistatic radar theory and signal processing techniques and, furthermore, examined the current state of the GNSS-R research, identifying the inefficient use of phase information as a key research gap. The experimental setup and airborne data acquisition campaign were detailed as the background for the verification work in the following chapters.

In the second chapter, a statistical modeling framework for multiplicative speckle noise associated with coherent radar measurements was established. Through comprehensive analysis, it was shown that the Gamma distribution serves as the suitable statistical model for multi-look GNSS-R reflectivity data, unlike the additive Gaussian noise assumptions that were employed in prior studies. A review of existing algorithms for change point detection highlighted significant limitations in current methodologies, especially concerning parameter sensitivity and computational complexity when dealing with Gamma noise conditions.

To overcome these limitations, we developed in the third chapter an integrated segmentation framework specifically designed for Gamma-distributed reflectivity signals. For this purpose, we derived the theoretical basis

for a homomorphic transform that turned multiplicative speckle noise to constant-variance additive noise, enabling us to fix a detection threshold. The methodology used online CUSUM detection for low-computational-overhead detection and offline Log-Gamma Maximum Likelihood Estimation for accurate change localization. To reduce false alarms and over-segmentation, we used a Wald test-based merging algorithm that effectively minimized false detections. Real-airborne measurement validation proved that the proposed system effectively segmented GNSS-R reflectivity signals, correctly identifying surface transitions along the entire 70 km flight path.

Following the results of the segmentation analysis, in the fourth chapter, a detailed framework was established for exploiting phase information in GNSS-R signals. The entire signal processing chain was derived with discrete-time integrations for both direct and reflected signals. This formed the mathematical foundations for phase extraction from I/Q components. A statistical phase model was established to characterize the transition from coherent to incoherent scattering regimes as a function of the signal-to-noise ratio, with concentration parameters that closely relate to signal coherence properties. In addition, we studied the excess Doppler spectrum analysis using three quantitative metrics that included peak amplitude, spectral complexity, and characteristics of energy dispersion. Analysis of real airborne data demonstrated that these metrics had significant potential in distinguishing surface types, exhibiting clear spectral differences between coherent surfaces (such as sand and inland water) and incoherent surfaces (such as land and rough sea) surfaces.

To address the limited availability of labeled real GNSS-R training data, we established an approach based on synthetic data generation from the multipath reflection model, in which surface-specific parameters were extracted from real airborne measurements. Spectral characteristics of the generated signals showed close agreement with real observations, thus validating the physical accuracy of the underlying models. K-means clustering on the three excess Doppler criteria demonstrated efficient unsupervised surface classification, leading to improved discrimination compared to purely reflectivity-dependent methods in chapter 3, and particularly in separating sea and river surfaces that, although sharing similar reflectivity, had inherently distinct scattering mechanisms.

We used machine learning algorithms with the established phase-based models and features in the fifth chapter. A dual-dataset approach was used to systematically compare six supervised learning algorithms. All of the models achieved more than 98% accuracy when tested on synthetic test sets, therefore validating their learning ability. The crucial validation performed on real airborne data showed that XGBoost and 1D-CNN achieved 90% classification accuracy even though they had only been trained on synthetic signals. Random Forest and MLP had also been found to show high performance with 88% and 87% accuracy, respectively. These results validated both the physical realism in the generation of the synthetic data and the resilience of the classification paradigm.

The work described in this thesis addressed two major challenges related to airborne GNSS-R signal processing effectively. First, the Gamma-based segmentation paradigm allowed for reliable, automated detection of surface transitions from reflectivity observations, leading to improved localization precision in comparison with currently used methods. This paradigm is phase-invariant and served as the basis for phase-aware, automated detection of surface interfaces. Secondly, the phase-aware processing paradigm leveraged thus-far untapped data within GNSS-R signals through the use of statistical phase modeling and excess Doppler spectral analysis, thus allowing for additional discrimination beyond reflectivity alone. Integration of these methods with machine learning techniques produced 90% classification accuracy on real airborne datasets. Above all, the physics-based generation approach for synthetic data resolved the critical problem of scarce labeled GNSS-R datasets, allowing for effective training and adaptation towards real-world observations.

Perspectives

One of the main perspectives of this work is to conduct additional flight campaigns in different environmental conditions, seasons, and geographically diverse locations to comprehensively evaluate the framework robustness. Our validation was performed on data collected in October 2020 with wind speeds of 24 km/h under rainy conditions with high humidity, and absolute reflectivity levels for surfaces like sea and sandy beaches are expected to vary significantly with meteorological parameters. High wind speeds increase surface roughness, leading to more distributed multipath scattering and lower

mean reflectivity. Therefore, while the segmentation approach successfully detects changes between surfaces, specific reflectivity thresholds defining the surface classes might need recalibration when used on data collected under different conditions. Future work should target adaptive threshold approaches including Rice or Rician-Gamma mixture models. This might improve surface characterization with enhanced detection sensitivity and reduced false alarm probability in coherent reflections.

The validation relied on Google Earth imagery and the Institut Géographique National (IGN) maps for verification of surface classifications, therefore providing a qualitative assessment of segmentation and classification performance. A critical perspective for future work involves conducting coordinated in-situ measurement campaigns as a quantitative ground truth. This ground truth dataset would facilitate the calculation of standard classification metrics for each surface type, enabling a statistical analysis of error localization along surface boundaries. Such validation would further support the development of physics-based correction factors relating reflectivity measurements to quantitative surface parameters such as soil moisture or vegetation water content.

Adding multi-frequency GNSS observations to the framework would take advantage of the better capabilities of modernized satellite signals. GPS L5 and Galileo E5 signals transmit higher power than L1 signals, improving the signal-to-noise ratio and provide deeper penetration into vegetation canopies to better characterize the land surface. Multi-frequency coherence analysis could show scattering characteristics that vary with wavelength, which provide supplementary discriminative features beyond single-frequency methodologies. To deal with multi-frequency data streams, the segmentation framework would need to be expanded. This might include cross-frequency change detection to find surfaces that show frequency-dependent scattering features.

Another direction to advance in this topic is combining GNSS-R data with SAR images to improve surface monitoring. GNSS-R provides continuous temporal coverage with passive reception and wide-area observations, while SAR offers higher spatial resolution through active illumination and coherent processing. Combining both measurements in data fusion methods improves classification accuracy and enables multi-scale surface

characterization. The characteristics of these techniques, as passive versus active radar and temporal versus spatial resolution, indicate the potential for hybrid categorization frameworks that utilize the positive aspects of both methodologies. This fusion would yield independent validation and facilitate the quantification of the relative benefits of passive and active radar remote sensing across various surface types and environmental conditions.

Another direction to advance this work is to delve into physics-informed neural networks that directly use the multi-path mechanism and Gamma distribution parameters as inputs into the network architecture or loss functions. This would combine data-driven learning with physical constraints. Explainable AI techniques, including SHAP values and attention mechanisms, can explain the features (reflectivity statistics, phase coherence, or spectral characteristics) that influence classification decisions for various surfaces. This interpretability would verify if the patterns learned fit with electromagnetic scattering theory and could even show new physical relationships. Domain adaptation methods might enable model transfer across several regions or weather conditions without requiring additional labeled data.

The developed framework successfully classified four main surface types: sand, sea, land, and inland water bodies. Future work might enhance the classification to identify more specific surface classes, including various vegetation types (forests, crops, grasslands), urban regions, wetlands, and snow or ice coverage. Combining reflectivity statistics with phase coherence metrics developed in this thesis provides a baseline for separating surfaces with different dielectric and roughness properties. Applying the methodology to measure surface parameters like soil moisture or vegetation biomass would make it a quantitative measurement tool rather than just a categorical classifier.

Appendix A

Signal Processing Derivations for Multi-Path GNSS-R Phase Modeling

The mathematical development in Chapter 4 relies on a detailed understanding of how to process the received GNSS-R signals across the entire receiver chain, from reception at the first point to estimation at the output point. The core chapter addresses the discrete-time implementation that is immediately translated into real data processing, yet the continuous-time mathematical development forms the underlying theory that justifies our approach and ensures the appropriateness of our numerical algorithm.

The continuous-time analysis becomes particularly important when dealing with the multi-path reflection scenarios that characterize GNSS-R observations over different surface types. As demonstrated in the chapter, the range between the coherently and incoherently scattering regimes depends essentially on the combination, at the level of the receiver antenna, of multiple reflection paths. The mathematical expressions thus obtained in continuous time accurately capture the underlying physics of this combination process, eliminating the discretization effects that could potentially mask important theoretical insights.

Additionally, the formulation in continuous time enables us to establish precise interconnections between the physical parameters of the scattering phenomenon—such as surface roughness, reflection coefficients, and geometrical path differences—and the statistical entities of the resulting phase measurements. The resulting relationships form the basis for the surface classification methodology proposed in Chapter 4, where we show that phase coherence is a reliable signature for a surface type.

A.1 Continuous-Time Reflected Signal Processing Model

The signal processing chain begins with the fundamental representation for the reflected GNSS signal in continuous time. For the reflected signal, considering multiple reflection paths within the glistening zone, consider the following digital signal, sampled at a frequency f_e:

$$s_r(t) = \sum_{i=1}^{M} A_i \, CA(t + \tau_i) \sin\left(2\pi f_1 t + \phi_i(t)\right) + \eta(t) \tag{A.1}$$

In this equation t corresponds to $1/f_e$, where t is an integer in the range $[1, N_c f_e T_c]$, with N_c representing the number of integration time T_c. $CA(\cdot) \in \{-1, 1\}$ represents the CDMA code of the GNSS signal

This model assumes M reflections, each with different Doppler shifts (i.e., different phase errors) and amplitudes A_i. The noise component $\eta(t)$ is assumed to originate only from the antenna, ensuring that it remains the same for all reflections.

The received frequency f_1 is fixed to match the frequency of the direct signal. The excess Doppler, defined in $\phi_i(t)$, represents the difference in carrier frequency between the direct and reflected signals. This difference arises due to the variation in signal paths in the bi-static GNSS-R configuration.

A.2 Statistical Phase Model

A.2.1 Linear Phase Assumption

For the reflected signal, assuming multiple reflection paths, each reflection i follows a linear phase model:

$$\phi_i(t) = \phi_{0_i} + \alpha_i t \tag{A.2}$$

where $\alpha_i = 2\pi f_{d_i}$ represents the Doppler shift of the ith reflection path in the glistening zone.

This linear evolution assumption for the phase becomes crucial for analytical tractability while ensuring physical realism. The assumption is acceptable

for the usual scenarios in GNSS-R observations, where satellite orbital dynamics and the motion of the platform make the Doppler frequencies vary slowly compared to the integration timescales used in the receiver chain.

A.2.2 In-phase and Quadrature Signal Modeling

Using the reflected signal model, we now develop the complete signal processing chain to extract the in-phase and quadrature components. The following steps demonstrate how the theoretical signal representations are processed to obtain the I_{kTc} and Q_{kTc} measurements that form the basis of our phase estimation methodology.

There are three successive stages in the methodology of signal processing, i.e., demodulation to extract baseband information, demultiplexing that isolates target components in the signal and correlation to integrate the signal in the interval of coherent processing. Each stage includes several mathematical processes that must be examined keenly in order to determine their impact on phase estimation error.

A.2.2.1 Demodulation

The demodulated in-phase and quadrature components are:

$$\begin{aligned} s_{dI}(t) &= \sin(2\pi f_1 t) \cdot s_r(t) \\ &= \sum_{i=1}^{M} CA(t+\tau_i) \cdot \frac{A_i}{2} \Big[\cos(\phi_i(t)) \\ &\quad - \cos(4\pi f_1 t + \phi_i(t)) \Big] + \sin(2\pi f_1 t) \cdot \eta(t) \end{aligned} \tag{A.3}$$

$$\begin{aligned} s_{dQ}(t) &= \cos(2\pi f_1 t) \cdot s_r(t) \\ &= \sum_{i=1}^{M} CA(t+\tau_i) \cdot \frac{A_i}{2} \Big[\sin(\phi_i(t)) \\ &\quad + \sin(4\pi f_1 t + \phi_i(t)) \Big] + \cos(2\pi f_1 t) \cdot \eta(t) \end{aligned} \tag{A.4}$$

The correlation separates the desired baseband components from the double-frequency terms that are filtered out. Then, we have:

$$s_{dI}(t) = \sum_{i=1}^{M} CA(t+\tau_i) \cdot \frac{A_i}{2} \cos(\phi_i(t)) + \sin(2\pi f_1 t) \cdot \eta(t) \tag{A.5}$$

$$s_{dQ}(t) = \sum_{i=1}^{M} CA(t+\tau_i) \cdot \frac{A_i}{2} \sin(\phi_i(t)) + \cos(2\pi f_1 t) \cdot \eta(t) \tag{A.6}$$

A.2.2.2 Demultiplexing

To isolate the reflected signal:

$$\begin{aligned} s_I(t) &= CA(t+\tau_l)\, s_{dI}(t) \\ &= \sum_{i=1}^{M} \frac{A_i}{2} CA(t+\tau_l) CA(t+\tau_i) \cos(\phi_i(t)) \Big|_{t\in\mathcal{S}_1} \\ &\quad - \sum_{i=1}^{M} \frac{A_i}{2} CA(t+\tau_l) CA(t+\tau_i) \cos(\phi_i(t)) \Big|_{t\in\mathcal{S}_2} \\ &\quad + CA(t+\tau_l)\ \sin(2\pi f_1 t) \cdot \eta(t) \end{aligned} \tag{A.7}$$

$$\begin{aligned} s_Q(t) &= CA\,(t+\tau_l)\, s_{dQ}(t) \\ &= \sum_{i=1}^{M} \frac{A_i}{2} CA(t+\tau_l) CA(t+\tau_i) \sin\,(\phi_i(t)) \Big|_{t\in S_1} \\ &\quad - \sum_{i=1}^{M} \frac{A_i}{2} CA(t+\tau_l) CA(t+\tau_i) \sin\,(\phi_i(t)) \Big|_{t\in S_2} \\ &\quad + CA\,(t+\tau_l) \cos\,(2\pi f_1 t)\, \eta(t) \end{aligned} \tag{A.8}$$

Here, S_1 and S_2 are the sets where $CA(t+\tau_i) = CA(t+\tau_l)$ and $CA(t+\tau_i) \neq CA(t+\tau_l)$, respectively.

A.2.2.3 Correlation

The correlation operation integrates the demodulated signals over the coherent processing interval T_c, we have:

$$I_{kT_c} = \int_{kT_c}^{(k+1)T_c} s_I(t)dt \tag{A.9}$$

$$Q_{kT_c} = \int_{kT_c}^{(k+1)T_c} s_Q(t)dt \tag{A.10}$$

Then,

$$\begin{aligned} I_{kT_c} = &\sum_{i=1}^{M} \frac{A_i}{2} R(\tau_i - \tau_l) \int_{kT_c}^{(k+1)T_c} \cos(\phi_i(t))dt \\ &+ \int_{kT_c}^{k+1T_c} CA(t+\tau_l)\sin(2\pi f_1 t)\ \eta(t)dt \end{aligned} \tag{A.11}$$

$$\begin{aligned} Q_{kT_c} = &\sum_{i=1}^{M} \frac{A_i}{2} R(\tau_i - \tau_l) \int_{kT_c}^{(k+1)T_c} \sin(\phi_i(t))dt \\ &+ \int_{kT_c}^{k+1T_c} CA(t+\tau_l)\cos(2\pi f_1 t)\ \eta(t)dt \end{aligned} \tag{A.12}$$

where $R(\cdot)$ represents the autocorrelation function of the spreading code.

Using the linear phase assumption in A.2.1, the integrals can be evaluated analytically:

$$\begin{aligned} \int_{kT_c}^{(k+1)T_c} \cos(\phi_{0_i} + \alpha_i t)\, dt &= \left[\frac{1}{\alpha_i}\sin(\phi_{0_i} + \alpha_i t)\right]_{kT_c}^{(k+1)T_c} \\ &= \frac{1}{\alpha_i}\left[\sin(\phi_{0_i} + \alpha_i(k+1)T_c) - \sin(\phi_{0_i} + \alpha_i k T_c)\right] \\ &= \frac{2}{\alpha_i}\cos\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right)\sin\left(\frac{\alpha_i T_c}{2}\right) \end{aligned} \tag{A.13}$$

$$\begin{aligned} \int_{kT_c}^{(k+1)T_c} \sin(\phi_{0_i} + \alpha_i t)\, dt &= \left[\frac{-1}{\alpha_i}\cos(\phi_{0_i} + \alpha_i t)\right]_{kT_c}^{(k+1)T_c} \\ &= \frac{-1}{\alpha_i}\left[\cos\left(\phi_{0_i} + \alpha_i(k+1)T_c\right) - \cos\left(\phi_{0_i} + \alpha_i k T_c\right)\right] \\ &= \frac{2}{\alpha_i}\sin\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right)\sin\left(\frac{\alpha_i T_c}{2}\right) \end{aligned} \tag{A.14}$$

The expressions of the in-phase and quadrature components for the reflected signal are given by:

$$I_{kT_c} = \sum_{i=1}^{M} \frac{A_i}{2} R(\tau_i - \tau_l) \frac{2}{\alpha_i} \sin\left(\frac{\alpha_i T_c}{2}\right) \cos\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right) + \eta_I \tag{A.15}$$

$$Q_{kT_c} = \sum_{i=1}^{M} \frac{A_i}{2} R(\tau_i - \tau_l) \frac{2}{\alpha_i} \sin\left(\frac{\alpha_i T_c}{2}\right) \sin\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right) + \eta_Q \tag{A.16}$$

Finally, we have:

$$I_{kT_c} = \sum_{i=1}^{M} a_i \cos\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right) + \eta_I \tag{A.17}$$

$$Q_{kT_c} = \sum_{i=1}^{M} a_i \sin\left(\frac{2\phi_{0_i} + 2\alpha_i k T_c + \alpha_i T_c}{2}\right) + \eta_Q \tag{A.18}$$

with $a_i = \frac{A_i}{2} R(\tau_i - \tau_l) \frac{2}{\alpha_i} \sin\left(\frac{\alpha_i T_c}{2}\right)$.

It is worth noting that the noise term η follows a Gaussian distribution. According to the law of large numbers, we can assume that the second term of I_{kT_c}, denoted as η_I, and respectively η_Q for Q_{kT_c}, are also distributed according to a Gaussian distribution.

Bibliography

[1] Alexander G Voronovich and Valery U Zavorotny. "Bistatic radar equation for signals of opportunity revisited". In: *IEEE Transactions on Geoscience and Remote Sensing* 56.4 (2017), pp. 1959–1968.

[2] Manuel Martin-Neira et al. "A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry". In: *ESA journal* 17.4 (1993), pp. 331–355.

[3] Joan Francesc Munoz-Martin et al. "Single-pass soil moisture retrieval using GNSS-R at L1 and L5 bands: Results from airborne experiment". In: *Remote Sensing* 13.4 (2021), p. 797.

[4] Joakim Strandberg, Thomas Hobiger, and Rüdiger Haas. "Coastal sea ice detection using ground-based GNSS-R". In: *IEEE geoscience and remote sensing letters* 14.9 (2017), pp. 1552–1556.

[5] Estel Cardellach et al. "Characterization of dry-snow sub-structure using GNSS reflected signals". In: *Remote Sensing of Environment* 124 (2012), pp. 122–134.

[6] Jiahua Zhang et al. "Mapping surface water extents using high-rate coherent spaceborne GNSS-R measurements". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15.

[7] Scott Gleason et al. "Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing". In: *IEEE Transactions on Geoscience and Remote Sensing* 43.6 (2005), pp. 1229–1241.

[8] Adriano Camps et al. "Sensitivity of TDS-1 GNSS-R reflectivity to soil moisture: Global and regional differences and impact of different spatial scales". In: *Remote Sensing* 10.11 (2018), p. 1856.

[9] C Chew et al. "The sensitivity of ground-reflected GNSS signals to near-surface soil moisture, as recorded by spaceborne receivers". In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2017, pp. 2661–2663.

[10] Baojian Liu et al. "Statistical analysis of CyGNSS speckle and its applications to surface water mapping". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15.

[11] Zhounan Dong and Shuanggen Jin. "Evaluation of the land GNSS-Reflected DDM coherence on soil moisture estimation from CYGNSS data". In: *Remote Sensing* 13.4 (2021), p. 570.

[12] Jinwei Bu, Qiulan Wang, and Jun Ni. "Estimating sea surface swell height using a hybrid model combining CNN, ConvLSTM, and FCN based on spaceborne GNSS-R data from the CYGNSS mission". In: *GPS Solutions* 28.3 (2024), p. 133.

[13] Hamza Issa. "Non-linear GNSS signal processing applied to land observation with high-rate airborne reflectometry". In: Université du Littoral Côte d'Opale. 2022.

[14] Hyuk Park et al. "Analysis of spaceborne GNSS-R delay-Doppler tracking". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.5 (2014), pp. 1481–1492.

[15] Martin Unwin, Scott Gleason, and Michael Bren. "The space GPS reflectometry experiment on the UK disaster monitoring constellation satellite". In: *Proceedings of the 16th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003)*. 2003, pp. 2656–2663.

[16] Martin Unwin et al. "Spaceborne GNSS-reflectometry on TechDemoSat-1: Early mission operations and exploitation". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.10 (2016), pp. 4525–4539.

[17] MP Clarizia et al. "Analysis of GNSS-R delay-Doppler maps from the UK-DMC satellite over the ocean". In: *Geophysical Research Letters* 36.2 (2009).

[18] Joakim Strandberg. *New methods and applications for interferometric GNSS reflectometry*. Chalmers Tekniska Hogskola (Sweden), 2020.

[19] Alberto Alonso-Arroyo, Valery U Zavorotny, and Adriano Camps. "Sea ice detection using UK TDS-1 GNSS-R data". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.9 (2017), pp. 4989–5001.

[20] Nereida Rodriguez-Alvarez et al. "Airborne GNSS-R wind retrievals using delay–Doppler maps". In: *IEEE transactions on geoscience and remote sensing* 51.1 (2012), pp. 626–641.

[21] Yang Nan et al. "Signal-to-noise ratio analyses of spaceborne GNSS-reflectometry from Galileo and BeiDou satellites". In: *Remote Sensing* 14.1 (2021), p. 35.

[22] Eric Loria et al. "Comparison of GNSS-R coherent reflection detection algorithms using simulated and measured CYGNSS data". In: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), pp. 1–16.

[23] Chris Oliver and Shaun Quegan. *Understanding synthetic aperture radar images*. SciTech Publishing, 2004.

[24] David Long and Fawwaz Ulaby. *Microwave radar and radiometric remote sensing*. Artech, 2015.

[25] Dusan Gleich and Mihai Datcu. "Wavelet-based despeckling of SAR images using Gauss–Markov random fields". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.12 (2007), pp. 4127–4143.

[26] Fawwaz T Ulaby et al. "Textural infornation in SAR images". In: *IEEE Transactions on Geoscience and Remote Sensing* 2 (2007), pp. 235–245.

[27] Joan Francesc Munoz-Martin et al. "Untangling the incoherent and coherent scattering components in GNSS-R and novel applications". In: *Remote Sensing* 12.7 (2020), p. 1208.

[28] Thomas Hobiger, Rüdiger Haas, and JS Löfgren. "GLONASS-R: GNSS reflectometry with a frequency division multiple access-based satellite navigation system". In: *Radio Science* 49.4 (2014), pp. 271–282.

[29] Vu A Nguyen et al. "Initial GNSS phase altimetry measurements from the spire satellite constellation". In: *Geophysical Research Letters* 47.15 (2020), e2020GL088308.

[30] Fran Fabra et al. "Phase altimetry with dual polarization GNSS-R over sea ice". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.6 (2011), pp. 2112–2121.

[31] AM Semmling et al. "Detection of Arctic Ocean tides using interferometric GNSS-R signals". In: *Geophysical Research Letters* 38.4 (2011).

[32] Weiqiang Li et al. "First spaceborne phase altimetry over sea ice using TechDemoSat-1 GNSS-R signals". In: *Geophysical Research Letters* 44.16 (2017), pp. 8369–8376.

[33] Vu Nguyen et al. "Sea Ice Altimetry and Classification using Grazing Angle Reflected GNSS Signals Measured by Spire's Nanosatellite Constellation". In: *AGU Fall Meeting Abstracts*. Vol. 2020. 2020, pp. C014–0014.

[34] LI Weiqiang et al. "First spaceborne demonstration of BeiDou-3 signals for GNSS reflectometry from CYGNSS constellation". In: *Chinese Journal of Aeronautics* 34.9 (2021), pp. 1–10.

[35] Jilun Peng et al. "Signal coherence and water detection algorithms for the ESA HydroGNSS mission". In: Institute of Electrical and Electronics Engineers. 2024.

[36] Carolyn J Roesler et al. "Coherent GNSS-reflections characterization over ocean and sea ice based on spire global CubeSat data". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), pp. 1–18.

[37] Stienne Georges. "Processing of circular signals applied to altimetry by the phase of GNSS signals". Theses. Université du Littoral Côte d'Opale, Dec. 2013. URL: https://theses.hal.science/tel-00950869.

[38] Ilaria Mara Russo et al. "Entropy-based coherence metric for land applications of GNSS-R". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), pp. 1–13.

[39] Hamza Issa et al. "Airborne GNSS reflectometry for water body detection". In: *Remote Sensing* 14.1 (2021), p. 163.

[40] Joseph W Goodman. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[41] Yves Delignon and Wojciech Pieczynski. "Modeling non-Rayleigh speckle distribution in SAR images". In: *IEEE Transactions on Geoscience and Remote sensing* 40.6 (2002), pp. 1430–1435.

[42] Heng-Chao Li et al. "An efficient and flexible statistical model based on generalized gamma distribution for amplitude SAR images". In: *IEEE Transactions on Geoscience and Remote Sensing* 48.6 (2010), pp. 2711–2722.

[43] Lorenzo Bruzzone et al. "An advanced system for the automatic classification of multitemporal SAR images". In: *IEEE Transactions on Geoscience and Remote Sensing* 42.6 (2004), pp. 1321–1334.

[44] Fawwaz T Ulaby, Thomas F Haddock, and Richard T Austin. "Fluctuation statistics of millimeter-wave scattering from distributed targets". In: *IEEE transactions on Geoscience and Remote Sensing* 26.3 (2002), pp. 268–281.

[45] Samuel F George. *The detection of nonfluctuating targets in log-normal clutter*. Tech. rep. 1968.

[46] A Farina, A Russo, and FA Studer. "Coherent radar detection in log-normal clutter". In: *IEE Proceedings F (Communications, Radar and Signal Processing)*. Vol. 133. 1. IET. 1986, pp. 39–53.

[47] GV Trunk. "Radar properties of non-Rayleigh sea clutter". In: *IEEE Transactions on Aerospace and Electronic Systems* 2 (2007), pp. 196–204.

[48] Mario A. T. Figueiredo and Anil K. Jain. "Unsupervised learning of finite mixture models". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.3 (2002), pp. 381–396.

[49] M Sekine et al. "Weibull-distributed ground clutter". In: *IEEE Transactions on Aerospace and Electronic Systems* 4 (1981), pp. 596–598.

[50] Matsuo Sekine et al. "On Weibull-distributed weather clutter". In: *IEEE Transactions on Aerospace Electronic Systems* 15 (1979), pp. 824–830.

[51] DC Schleher. "Radar detection in Weibull clutter". In: *IEEE Transactions on aerospace and electronic systems* 6 (2007), pp. 736–743.

[52] Gabriele Moser, Josiane Zerubia, and Sebastiano B Serpico. "SAR amplitude probability density function estimation based on a generalized Gaussian model". In: *IEEE transactions on image processing* 15.6 (2006), pp. 1429–1442.

[53] Edney W Stacy. "A generalization of the gamma distribution". In: *The Annals of mathematical statistics* (1962), pp. 1187–1192.

[54] E Webb Stacy and G Arthur Mihram. "Parameter estimation for a generalized gamma distribution". In: *Technometrics* 7.3 (1965), pp. 349–358.

[55] Zhenhong Du et al. "A new method for ship detection in SAR imagery based on combinatorial PNN model". In: *2008 First International Conference on Intelligent Networks and Intelligent Systems*. IEEE. 2008, pp. 531–534.

[56] Gui Gao. "Statistical modeling of SAR images: A survey". In: *Sensors* 10.1 (2010), pp. 775–795.

[57] Gabriele Moser, Josiane Zerubia, and Sebastiano B Serpico. "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation". In: *IEEE Transactions on Geoscience and Remote Sensing* 44.1 (2005), pp. 188–200.

[58] Puyang Wang, He Zhang, and Vishal M Patel. "SAR image despeckling using a convolutional neural network". In: *IEEE Signal Processing Letters* 24.12 (2017), pp. 1763–1767.

[59] Dong-Xiao Yue, Feng Xu, and Ya-Qiu Jin. "SAR despeckling neural network with logarithmic convolutional product model". In: *International journal of remote sensing* 39.21 (2018), pp. 7483–7505.

[60] Jong-Sen Lee and Eric Pottier. *Polarimetric radar imaging: from basics to applications*. CRC press, 2017.

[61] Eo Jakeman. "On the statistics of K-distributed noise". In: *Journal of Physics A: Mathematical and General* 13.1 (1980), p. 31.

[62] JGPW Clevers. "FT Ulaby and DG Long: Microwave Radar and Radiometric Remote Sensing". In: *International Journal of Applied Earth Observation and Geoinformation* 34 (2014), pp. 264–265.

[63] Donald E Kerr. *Propagation of short radio waves*. Vol. 24. IET, 1987.

[64] CJ Oliver. "Correlated K-distributed clutter models". In: *Optica Acta: International Journal of Optics* 32.12 (1985), pp. 1515–1547.

[65] Frank Spitzer. *Principles of random walk*. Vol. 34. Springer Science & Business Media, 2001.

[66] Ercan E Kuruoglu and Josiane Zerubia. "Modeling SAR images with a generalization of the Rayleigh distribution". In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 527–533.

[67] CJ Oliver and RJA Tough. "On the simulation of correlated K-distributed random clutter". In: *Optica Acta: International Journal of Optics* 33.3 (1986), pp. 223–250.

[68] CJ Oliver. "The interpretation and simulation of clutter textures in coherent images". In: *Inverse Problems* 2.4 (1986), p. 481.

[69] Torbjørn Eltoft. "The Rician inverse Gaussian distribution: A new model for non-Rayleigh signal amplitude statistics". In: *IEEE Transactions on Image Processing* 14.11 (2005), pp. 1722–1735.

[70] Stephen O Rice. "Mathematical analysis of random noise". In: *The Bell System Technical Journal* 23.3 (1944), pp. 282–332.

[71] E Jakeman. "K-distributed noise". In: *Journal of Optics A: Pure and Applied Optics* 1.S (1999), p. 784.

[72] Heng-Chao Li et al. "On the empirical-statistical modeling of SAR images with generalized gamma distribution". In: *IEEE Journal of selected topics in signal processing* 5.3 (2011), pp. 386–397.

[73] Michele Basseville and Igor V Nikiforov. "Detection of abrupt changes: theory and applications". In: *Journal of the Royal Statistical Society-Series A Statistics in Society* 158.1 (1995), p. 185.

[74] Fredrik Gustafsson. "The marginalized likelihood ratio test for detecting abrupt changes". In: *IEEE Transactions on automatic control* 41.1 (2002), pp. 66–78.

[75] Yixiao Li et al. "A review of changepoint detection models". In: *arXiv preprint arXiv:1908.07136* (2019).

[76] Erich Leo Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer, 2005.

[77] Thomas Kerr. "Decentralized filtering and redundancy management for multisensor navigation". In: *IEEE transactions on Aerospace and Electronic Systems* 1 (2007), pp. 83–119.

[78] Daniel Barry and John A Hartigan. "A Bayesian analysis for change point problems". In: *Journal of the American Statistical Association* 88.421 (1993), pp. 309–319.

[79] Jeremias Knoblauch and Theodoros Damoulas. "Spatio-temporal Bayesian on-line changepoint detection with model selection". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 2718–2727.

[80] Paul Fearnhead and Zhen Liu. "On-line inference for multiple changepoint problems". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 69.4 (2007), pp. 589–605.

[81] Ryan Prescott Adams and David JC MacKay. "Bayesian online changepoint detection". In: *arXiv preprint arXiv:0710.3742* (2007).

[82] Ulrich Paquet. "Empirical Bayesian change point detection". In: *Graphical Models* 1995 (2007), pp. 1–20.

[83] Jeremias Knoblauch, Jack E Jewson, and Theodoros Damoulas. "Doubly Robust Bayesian Inference for Non-Stationary Streaming Data with β-Divergences". In: *Advances in Neural Information Processing Systems* 31 (2018).

[84] Hamed Heydari and Sayyed Bagher Fatemi Nasrabadi. "Remote sensing change detection: a comparative study of spectral distances". In: *Geocarto International* 38.1 (2023), p. 2248059.

[85] Nobuyuki Otsu et al. "A threshold selection method from gray-level histograms". In: *Automatica* 11.285-296 (1975), pp. 23–27.

[86] Ewan S Page. "Continuous inspection schemes". In: *Biometrika* 41.1/2 (1954), pp. 100–115.

[87] Gary Lorden. "Procedures for reacting to a change in distribution". In: *The annals of mathematical statistics* (1971), pp. 1897–1908.

[88] Gerrit JJ Van den Burg and Christopher KI Williams. "An evaluation of change point detection algorithms". In: *arXiv preprint arXiv:2003.06222* (2020).

[89] Paul Fearnhead and Guillem Rigaill. "Changepoint detection in the presence of outliers". In: *Journal of the American Statistical Association* 114.525 (2019), pp. 169–183.

[90] Samaneh Aminikhanghahi and Diane J Cook. "A survey of methods for time series change point detection". In: *Knowledge and information systems* 51.2 (2017), pp. 339–367.

[91] Xun Xiao et al. "On computing multiple change points for the gamma distribution". In: *Journal of Quality Technology* 53.3 (2021), pp. 267–288.

[92] Stergios B Fotopoulos, Alex Paparas, and Venkata K Jandhyala. "Change point detection and estimation methods under gamma series of observations". In: *Statistical Papers* 63.3 (2022), pp. 723–754.

[93] DA Hsu. "Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis". In: *Journal of the American Statistical Association* 74.365 (1979), pp. 31–40.

[94] Ashit B Chakraborty and Anwer Khurshid. "A note on Cumulative Sum (CUSUM) Control Chart for Doubly Truncated GAMMA Distribution". In: ().

[95] Daniel R Jeske. "CUSUM algorithm for detecting translations in gamma distributions". In: *Quality and Reliability Engineering International* 32.3 (2016), pp. 1205–1211.

[96] Chang-chun Tan, Bai-qi Miao, and Xing-cai Zhou. "Statistical inference for the shape parameter change-point estimator in negative associated gamma distribution". In: *Journal of Inequalities and Applications* 2013.1 (2013), p. 161.

[97] Sudha Jain and RK Jain. "An application of SPRT for detecting change point in a reliability growth model". In: *Microelectronics Reliability* 34.5 (1994), pp. 811–814.

[98] Eman Mohammed S Alamer. "CUSUM control charts for censored gamma distributed quality measurements". PhD thesis. 2017.

[99] Joseph W Goodman. "Statistical properties of laser speckle patterns". In: *Laser speckle and related phenomena*. Springer, 1975, pp. 9–75.

[100] Adriano Camps and Joan Francesc Munoz-Martin. "Analytical computation of the spatial resolution in GNSS-R and experimental validation at L1 and L5". In: *Remote Sensing* 12.23 (2020), p. 3910.

[101] Leigh J Halliwell. "The log-gamma distribution and non-normal error". In: *Variance* 13.2 (2021), pp. 173–189.

[102] Ross L Prentice. "A log gamma model and its maximum likelihood estimation". In: *Biometrika* 61.3 (1974), pp. 539–544.

[103] Monir Azmani et al. "Un estimateur récursif des paramètres statistiques d'une distribution circulaire de von Mises et d'une distribution Log-normale: Application à l'estimation de la direction et de la vitesse du vent". In: *Journal européen des systèmes automatisés* 44.4-5 (2010), pp. 525–545.

[104] Kenneth D Anderson. "Determination of water level and tides using interferometric observations of GPS signals". In: *Journal of Atmospheric and Oceanic Technology* 17.8 (2000), pp. 1118–1127.

[105] Georges Stienne et al. "A multi-temporal multi-sensor circular fusion filter". In: *Information Fusion* 18 (2014), pp. 86–100.

[106] Jinwei Bu et al. "Machine learning methods for earth observation and remote sensing using spaceborne gnss reflectometry: Current status, challenges, and future prospects". In: *IEEE Geoscience and Remote Sensing Magazine* (2025).

[107] Jinwei Bu et al. "Combining ERA5 data and CYGNSS observations for the joint retrieval of global significant wave height of ocean swell and wind wave: A deep convolutional neural network approach". In: *Journal of Geodesy* 97.8 (2023), p. 81.

[108] Zheng Li et al. "Integrating spaceborne GNSS-R and SOS for sea surface salinity retrieval using artificial neural network". In: *GPS Solutions* 28.4 (2024), p. 162.

[109] Joan Francesc Munoz-Martin and Adriano Camps. "Sea surface salinity and wind speed retrievals using GNSS-R and L-band microwave radiometry data from FMPL-2 onboard the FSSCat mission". In: *Remote Sensing* 13.16 (2021), p. 3224.

[110] Yun Zhang et al. "Machine learning methods for spaceborne GNSS-R sea surface height measurement from TDS-1". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2021), pp. 1079–1088.

[111] Yan Jia et al. "GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation". In: *Remote sensing* 11.14 (2019), p. 1655.

[112] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232.

[113] Dahai Zhang et al. "A data-driven design for fault detection of wind turbines using random forests and XGboost". In: *Ieee Access* 6 (2018), pp. 21020–21031.

[114] Huiting Zheng, Jiabin Yuan, and Long Chen. "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation". In: *Energies* 10.8 (2017), p. 1168.

[115] Volkan Senyurek et al. "Evaluations of machine learning-based CYGNSS soil moisture estimates against SMAP observations". In: *Remote Sensing* 12.21 (2020), p. 3503.

[116] Fangni Lei et al. "Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations". In: *Remote Sensing of Environment* 276 (2022), p. 113041.

[117] Yan Jia et al. "Modeling and theoretical analysis of GNSS-R soil moisture retrieval based on the random forest and support vector machine learning approach". In: *Remote Sensing* 12.22 (2020), p. 3679.

[118] Volkan Senyurek et al. "Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS". In: *Remote Sensing* 12.7 (2020), p. 1168.

[119] Emanuele Santi et al. "Machine learning applications for classification and retrieval of surface parameters from GNSS-R". In: *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2023, pp. 1174–1177.

[120] Emanuele Santi et al. "Forest biomass estimate on local and global scales through GNSS reflectometry techniques". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2019, pp. 8680–8683.

[121] Yongchao Zhu et al. "Spaceborne GNSS-R for sea ice classification using machine learning classifiers". In: *Remote Sensing* 13.22 (2021), p. 4577.

[122] Yongchao Zhu et al. "Machine learning-aided sea ice monitoring using feature sequences extracted from spaceborne GNSS-reflectometry data". In: *Remote Sensing* 12.22 (2020), p. 3751.

[123] Lei Wang et al. "Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.8 (2016), pp. 4524–4533.

[124] Xiaoyi Shen et al. "Sea ice classification using Cryosat-2 altimeter data by optimal classifier–feature assembly". In: *IEEE Geoscience and Remote Sensing Letters* 14.11 (2017), pp. 1948–1952.

[125] Leonid P Bobylev et al. "Neural-Network based algorithm for ice concentration retrievals from satellite passive microwave data". In: *2008 Microwave Radiometry and Remote Sensing of the Environment*. IEEE. 2008, pp. 1–4.

[126] Qingyun Yan, Weimin Huang, and Cecilia Moloney. "Neural networks based sea ice detection and concentration retrieval from GNSS-R delay-Doppler maps". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.8 (2017), pp. 3789–3798.

[127] Qingyun Yan and Weimin Huang. "Sea ice sensing from GNSS-R data using convolutional neural networks". In: *IEEE geoscience and remote sensing letters* 15.10 (2018), pp. 1510–1514.

[128] Lu Yang et al. "DNN-Based Retrieval of Arctic Sea Ice Concentration From GNSS-R and Its Effects on the Synoptic-Scale Forecasting as Supplementary Observation Source". In: *Geophysical Research Letters* 50.14 (2023), e2023GL104219.

[129] Qingyun Yan and Weimin Huang. "Sea ice remote sensing using GNSS-R: A review". In: *Remote Sensing* 11.21 (2019), p. 2565.

[130] Stylianos Kossieris, Milad Asgarimehr, and Jens Wickert. "Unsupervised machine learning for GNSS reflectometry inland water body detection". In: *Remote Sensing* 15.12 (2023), p. 3206.

[131] Steven J Miller. "The method of least squares". In: *Mathematics Department Brown University* 8.1 (2006), pp. 5–11.

[132] Jinwei Bu and Kegen Yu. "Significant wave height retrieval method based on spaceborne GNSS reflectometry". In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5.

[133] Mahdi Movahedian Moghaddam, Alireza Afzal Aghaei, and Kourosh Parand. "Rational Jacobi Kernel Functions: A novel massively parallelizable orthogonal kernel for support vector machines". In: *2024 Third International Conference on Distributed Computing and High Performance Computing (DCHPC)*. IEEE. 2024, pp. 1–8.

[134] Arundhati Navada et al. "Overview of use of decision tree algorithms in machine learning". In: *2011 IEEE control and system graduate research colloquium*. IEEE. 2011, pp. 37–42.

[135] Leo Breiman. "Bagging predictors". In: *Machine learning* 24.2 (1996), pp. 123–140.

[136] Ji Zhu et al. "Multi-class adaboost". In: *Statistics and its Interface* 2.3 (2009), pp. 349–360.

[137] Emanuele Santi et al. "Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors". In: *International journal of applied earth observation and geoinformation* 48 (2016), pp. 61–73.

[138] Orhan Eroglu et al. "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks". In: *Remote sensing* 11.19 (2019), p. 2272.

[139] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[140] Xin Wang et al. "The prediction of oceanic mesoscale eddy properties and propagation trajectories based on machine learning". In: *Water* 12.9 (2020), p. 2521.

[141] Jinwei Bu et al. "Joint retrieval of sea surface rainfall intensity, wind speed, and wave height based on spaceborne GNSS-R: a case study of the oceans near China". In: *Remote Sensing* 15.11 (2023), p. 2757.

[142] Xin Qiao and Weimin Huang. "WaveTransNet: A transformer-based network for global significant wave height retrieval from spaceborne GNSS-R data". In: *IEEE Transactions on Geoscience and Remote Sensing* (2024).

[143] Cuixian Lu et al. "Global ocean wind speed retrieval from GNSS reflectometry using CNN-LSTM network". In: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), pp. 1–12.

Résumé Étendu de la Thèse

Introduction

La réflectométrie GNSS (GNSS-R : Global Navigation Satellite System Reflectometry) s'est imposée comme un paradigme révolutionnaire pour l'observation de la Terre, exploitant les signaux de navigation réfléchis pour la télédétection radar bistatique. Sa couverture globale continue et son architecture de réception passive, utilisant les signaux de navigation satellitaires existants, font du GNSS-R un complément économiquement viable aux techniques de télédétection micro-ondes établies.

La technologie GNSS-R s'est développée depuis plus de trois décennies, approfondissant la compréhension du contenu informationnel des signaux de navigation réfléchis. La proposition de Martin-Neira en 1993 du système *Passive Reflectometry and Interferometry System* (PARIS) a posé les bases théoriques de l'utilisation des signaux GPS (Global Positioning System) comme radar bistatique. Depuis, les applications se sont étendues au-delà des environnements marins vers les surfaces terrestres, la cryosphère et les eaux continentales, et sont passées d'expériences au sol à des missions spatiales telles que UK-DMC, TechDemoSat-1 et la constellation CYGNSS.

Malgré ces avancées, les travaux de recherche se sont principalement concentrés sur les observables basées sur l'amplitude, à travers l'étude de cartes Délai-Doppler et de mesures de réflectivité. Bien qu'efficaces pour l'estimation de la vitesse du vent en mer, de la hauteur de surface ou de l'humidité du sol, ces approches négligent une composante essentielle du signal : la composante de phase, encore peu exploitée, reflète l'évolution temporelle des interférences issues de mécanismes de diffusion complexes, ouvrant ainsi des opportunités pour une meilleure caractérisation des surfaces et une précision accrue des mesures.

Les réflexions cohérentes se produisent principalement sur des surfaces

planes, présentant une stabilité de phase permettant un traitement interférométrique et des mesures altimétriques précises. Les propriétés statistiques des mesures de phases incohérentes observées lors de réflexion sur surfaces rugueuses contiennent des informations sur les mécanismes de diffusion et les caractéristiques des surfaces. La distinction entre réflexions cohérentes et non cohérentes dépend de la longueur d'onde, de la rugosité de surface et de l'angle d'observation, ce qui rend l'analyse de la phase des signaux réfléchis particulièrement efficace pour la classification automatique des surfaces.

Cette thèse vise à combler l'écart entre le potentiel théorique du GNSS-R utilisant les mesures de phase et ses applications pratiques en observation de la Terre. Un cadre intégré combinant modélisation statistique, algorithmes de traitement du signal et apprentissage automatique est développé pour améliorer la classification des surfaces. Il est appliqué à des observations GNSS-R aéroportées acquises en nord de France à la verticale de différents types de surfaces tels que plages de sable, plans d'eau, mer et surfaces variées de sols.

Le bruit affectant les données GNSS-R est traité comme un bruit multiplicatif de type speckle, ce qui implique des traitements statistiques différents de ceux développés pour des modèles additifs classiques. Des modèles spécialisés basés sur la distribution Log-Gamma sont développés, permettant de concevoir des algorithmes de détection à seuils statistiques ne nécessitant pas de réglage empirique des paramètres.

L'analyse de la phase des signaux réfléchis nécessite l'utilisation de techniques avancées de segmentation pour identifier les transitions de surface tout en minimisant les fausses alarmes. L'approche proposée combine des méthodes de détection de type somme cumulative (CUSUM) et de localisation par estimation du maximum de vraisemblance adaptées au modèle de distribution Log-Gamma des mesures.

Les objectifs généraux et les contributions de ce travail peuvent être résumés comme suit :

– Développement d'un système automatique de segmentation des signaux GNSS-R modélisés suivant la distribution Gamma et transformation homomorphique associée, pour la détection de rupture de continuité des surfaces dans des mesures aéroportées.

- Conception de modèles statistiques tenant compte de la phase pour le traitement des signaux GNSS-R, établissant la relation entre cohérence de phase et propriétés de diffusion des surfaces pour une meilleure discrimination des surfaces.
- Introduction d'une méthode analyse spectrale d'excès Doppler basée sur des critères complémentaires aux techniques fondées sur la réflectivité.
- Développement d'une méthode de génération de données synthétiques basée sur des modèles physiques permettant de surmonter la rareté des jeux de données GNSS-R annotés.
- Implémentation et validation d'algorithmes d'apprentissage automatique pour la classification des surfaces, avec des modèles entraînés sur des données synthétiques et appliqués avec succès à des observations réelles.

La thèse est composée de cinq chapitres. Les Chapitres 1 et 2 sont bibliographiques, tandis que les Chapitres 3 à 5 présentent les contributions de recherche. Le Chapitre 1 décrit l'évolution des techniques de traitement des signaux GNSS-R, la théorie du radar bistatique ainsi que l'expérimentation aéroportée de laquelle sont issues les données qui seront traitées dans la thèse. Le Chapitre 2 établit la modélisation statistique du bruit de speckle, notamment la distribution Gamma, et passe en revue différentes méthodes existantes de détection de ruptures en identifiant leurs limites.

Le Chapitre 3 propose un modèle de segmentation séparant les signaux GNSS par morceaux stationnaires associés à différentes zones de réflexion. Il définit le modèle de mesures de réflectivité suivant la distribution Gamma, justifie la transformation homomorphique utilisée et décrit le système intégré de détection-localisation de ruptures combinant CUSUM et estimation du maximum de vraisemblance Log-Gamma. Un test de Wald est utilisé pour pour fusionner les segments, diminuant la sur-segmentation liée à de fausses détections.

Le Chapitre 4 présente les fondements théoriques des traitement GNSS-R tenant compte de la phase des signaux suivant un modèle de réflexion multi-trajets. Ce chapitre présente l'architecture de traitement du signal, le modèle de phase ainsi développé suivant la loi de von Mises ainsi que l'analyse spectrale de l'excès Doppler des signaux GNSS-R, pour laquelle

travers trois critères quantitatifs sont définis. Un algorithme de clustering K-means utilisant ces critères d'excès Doppler est appliqué aux données aéroportées réelles. Une méthode de génération de données synthétiques GNSS-R suivant ces modèles et paramétré à partir des données réelles est également présenté.

Le Chapitre 5 applique différents algorithmes d'apprentissage automatique aux modèles de phase précédemment définis. Il met en place une approche à double jeu de données, où les données synthétiques servent à l'entraînement de modèles d'apprentissage classiques et profonds, et les données réelles à leur test. Six modèles, allant de la Régression Logistique au Bi-LSTM, sont évalués de manière systématique, démontrant la pertinence de l'entraînement sur données synthétiques pour la classification réelle des surfaces observées.

1 Modèle statistique de la réflectivité GNSS-R

Cette section est consacrée à la modélisation statistique des mesures de réflectivité GNSS-R. Dans un système d'observation radar GNSS-R, la réflectivité est définie comme le rapport entre l'intensité du signal réfléchi et celle du signal direct émis par le satellite GNSS. Après l'étape de démultiplexage, le signal direct traité par la boucle fermée de poursuite de la phase PLL (Phase Lock Loop) pour chaque satellite GNSS peut s'exprimer comme :

$$s_d(t) = a \sin(2\pi f_d t + \phi) + \eta \tag{1}$$

où a désigne l'amplitude du signal, f_d représente sa fréquence Doppler, ϕ sa phase et η correspond à un bruit additif.

La boucle fermée de poursuite du délai de code (DLL) traite les deux composantes de corrélation :

$$I_d(\tau) = \int_0^{T_c} s_d(t) \sin\left(2\pi \hat{f}_1 (t+\tau) + \hat{\phi}\right) dt \tag{2}$$

$$Q_d(\tau) = \int_0^{T_c} s_d(t) \cos\left(2\pi \hat{f}_1 (t+\tau) + \hat{\phi}\right) dt \tag{3}$$

où $I_d(0)$ et $Q_d(0)$ représentent les deux composantes en quadrature de la corrélation et sont assumées être des variables aléatoires suivant une distribution gaussienne. τ est le délai de code. À partir de ces deux composantes,

l'intensité du signal est calculée selon :

$$T_d = I_d(0)^2 + Q_d(0)^2 \tag{4}$$

L'écart-type (σ) de la distribution gaussienne dépend de la puissance du bruit sur le signal direct.

Pour le signal réfléchi, le démultiplexage est effectué pour un délai modélisé à partir du délai de code du signal direct en supposant connue la position du récepteur. Après démultiplexage, l'expression du signal réfléchi est donnée par :

$$s_r(t) = \sum_{i=1}^{L} a_i \sin(2\pi f_i t + \phi_i) + \eta \tag{5}$$

où a_i, f_i et ϕ_i représentent les paramètres d'amplitude, fréquence et phase des réflexions multiples du signal GNSS sur le sol atteignant l'antenne réceptrice. La poursuite en boucle ouverte assistée du signal reçu traite les deux composantes en quadrature, et l'intensité du signal réfléchi est exprimée par :

$$T_r = I_r(0)^2 + Q_r(0)^2 \tag{6}$$

Lorsque le signal GNSS est réfléchi sur une surface rugueuse, les phases sont supposées aléatoires et suivent une distribution uniforme entre 0 et 2π, conformément au modèle de Goodman. Les composantes de corrélation sont considérées comme des variables aléatoires suivant une distribution normale centrée :

$$I_r(\tau) = \int_0^{T_c} s_r(t) \sin\left(2\pi \hat{f}_1 (t+\tau)\right) dt \tag{7}$$

$$Q_r(\tau) = \int_0^{T_c} s_r(t) \cos\left(2\pi \hat{f}_1 (t+\tau)\right) dt \tag{8}$$

Dans notre implémentation, nous considérons l'intensité moyenne du signal reçu dans une fenêtre de travail de taille N :

$$T_r = \frac{1}{N} \sum_{i=1}^{N} \left(I_r(0)^2 + Q_r(0)^2\right)_i \tag{9}$$

D'après les équations précédentes, chaque mesure individuelle d'intensité (T) suit une loi du khi-deux à deux degrés de liberté, avec un paramètre d'échelle σ^2. Lorsque N variables aléatoires khi-deux indépendantes sont moyennées, leur somme suit une distribution Gamma. Étant donné que T_r

correspond à la moyenne de ces variables, on obtient :

$$T_r \sim Gamma\left(N, \frac{2\sigma^2}{N}\right) \tag{10}$$

Pour normaliser l'observation du signal réfléchi, la réflectivité est définie par :

$$R = \frac{T_r}{\mathbb{E}[T_d]} \tag{11}$$

où l'intensité moyenne du signal direct $\mathbb{E}[T_d]$ est supposée constante pendant toute la durée de l'expérimentation. Dans le cas d'un modèle multiplicatif, la réflectivité suit une distribution Gamma, comme conséquence directe de la chaîne de traitement du signal et des propriétés statistiques des composantes de corrélation :

$$R(N, \lambda) \sim Gamma\left(N, \frac{\lambda}{N}\right) \tag{12}$$

La nature multiplicative du bruit de speckle introduit des difficultés fondamentales pour les algorithmes de détection de ruptures conçus pour des scénarios de bruit additif. L'opération homomorphique de transformation logarithmique permet de convertir le bruit multiplicatif en bruit additif de puissance constante. Cette transformation logarithmique est définie par :

$$W(N, \lambda) = \log R(N, \lambda) \tag{13}$$

avec :

$$W(N, \lambda) \sim \text{Log-Gamma}\left(N, \frac{\lambda}{N}\right) \tag{14}$$

où la variance de $W(N, \lambda)$ est constante et indépendante de la moyenne, laquelle varie avec la puissance du signal.

2 Segmentation de la réflectivité GNSS-R

2.1 Modèle de transition

La réflectivité des signaux GNSS est directement proportionnelle aux caractéristiques de diffusion du sol à l'intérieur de la première zone de Fresnel. Elle augmente lors de réflexions sur des plans d'eau en raison de la diffusion spéculaire et diminue sur les surfaces terrestres rugueuses impliquant des réflexions multiples. En s'appuyant sur le modèle statistique de distribution Gamma précédemment établi, nous développons un système de détection de

ruptures permettant d'identifier automatiquement ces transitions de surface. La propriété de variance constante obtenue par transformation logarithmique constitue la base théorique pour la conception d'un algorithme de détection robuste à seuil fixe.

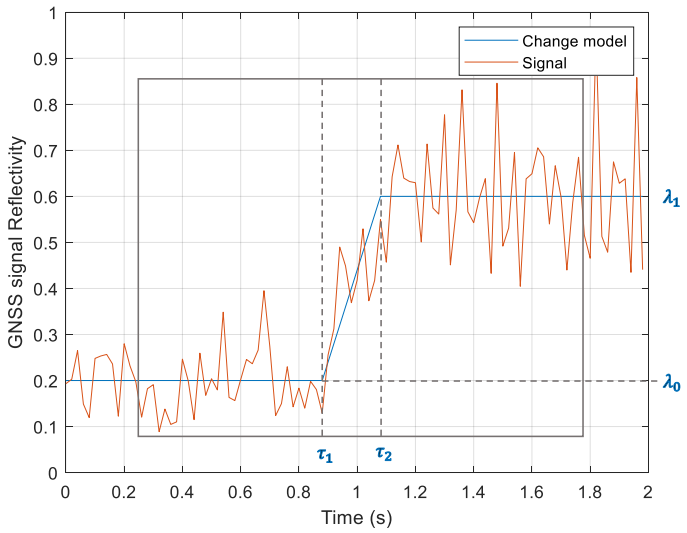


FIGURE A.1: Modèle du signal dans une fenêtre de travail.

Comme illustrée Figure refmodel, lorsque la valeur moyenne de la réflectivité GNSS est égale à λ_0, le signal réfléchi provient d'un premier type de surface ; lorsqu'elle évolue vers λ_1, la réflexion est issue d'un second type de surface.

2.2 Système de détection de ruptures en ligne et hors ligne

2.2.1 Détection de rupture

La Figure A.2 illustre le système de détection de ruptures en ligne proposé. Le processus de détection utilise les observations de réflectivité transformées par logarithme (w_t), qui sont traitées au moyen d'un filtre récursif estimant la moyenne locale du signal. Ce filtre suit de manière adaptative les variations du niveau du signal tout en maintenant une estimation courante de la moyenne du processus. Le terme d'innovation (i_t), représentant l'écart entre l'observation courante et la moyenne estimée, constitue ensuite l'entrée du détecteur CUSUM.

L'algorithme CUSUM accumule ces innovations au cours du temps, formant deux statistiques qui suivent respectivement les dérives ascendantes et descendantes du signal. Tant que le processus reste stationnaire (hypothèse

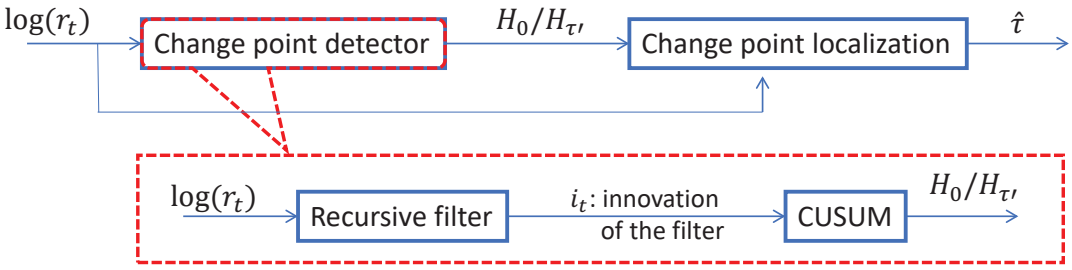


FIGURE A.2: Système de détection de ruptures.

H_0), ces statistiques demeurent bornées et proches de zéro. Lorsqu'une rupture se produit, les écarts accumulés commencent à croître de manière systématique, entraînant le dépassement par la statistique CUSUM d'un seuil prédéfini. Ce seuil est dérivé statistiquement en fonction de la longueur moyenne de fonctionnement sans alarme (*Average Run Length*, ARL(0)), qui contrôle le taux de fausses alertes. Lorsque la statistique CUSUM franchit le seuil, le détecteur déclenche une alarme à l'instant τ', indiquant qu'une rupture de moyenne de réflectivité a été détecté. Cette approche de détection en ligne permet une identification en temps réel des transitions de surface avec un coût calculatoire minimale, ce qui la rend particulièrement adaptée au traitement continu des données GNSS-R aéroportées.

2.2.2 Localisation de la rupture

Comme le détecteur CUSUM présente un certain délai de détection par rapport à l'instant réel de la rupture, une approche de localisation hors ligne est utilisée pour affiner l'estimation initiale. Une fois qu'une rupture est identifiée par le détecteur CUSUM, cette approche hors ligne permet d'affiner la localisation dans une fenêtre de travail. Nous proposons une méthode d'estimation du maximum de vraisemblance sous hypothèse Log-Gamma (LGMLE : Log-Gamma Maximum Likelihood Estimate) afin de déterminer avec précision les bornes de la transition de rupture. La localisation suppose une transition progressive entre deux régimes stationnaires, caractérisés par un instant de début τ_1 et un instant de fin τ_2, au cours desquels la moyenne de la réflectivité évolue linéairement de λ_0 à λ_1. Les paramètres optimaux sont obtenus en maximisant la fonction de log-vraisemblance totale:

$$(\hat{\tau_1}, \hat{\tau_2}, \hat{\lambda_0}, \hat{\lambda_1}) = \arg\max_{\tau_1,\tau_2} \log L_{\text{total}}(\tau_1, \tau_2, \lambda_0, \lambda_1) \tag{15}$$

où $\mathcal{L}_{\text{total}}$ représente la somme des log-vraisemblances sur trois segments : le segment pré-transition ($t < \tau_1$), le segment de transition ($\tau_1 \leq t \leq \tau_2$) et le segment après transition ($t > \tau_2$).

Cette optimisation permet d'identifier les paramètres de transition les plus probables expliquant les mesures observées de réflectivité transformée par logarithme. L'approche LGMLE offre une précision de localisation supérieure à celle obtenue par la détection CUSUM initiale, en particulier pour les transitions progressives de surface observées lors des mesures aéroportées.

2.2.3 Fusion des ruptures

Suite à la détection et de la localisation des ruptures, le signal de réflectivité est souvent sur-segmenté en raison de détections erronées induites par le bruit. Nous proposons un algorithme de fusion fondé sur un test d'hypothèse statistique permettant de combiner les segments adjacents tout en préservant les véritables transitions de surface. Le processus de fusion évalue si des segments consécutifs proviennent de la même population statistique en testant l'égalité de leurs paramètres d'échelle sous le modèle Gamma.

Pour deux segments adjacents dont les observations de réflectivité suivent respectivement $Gamma(N, \lambda_i/N)$ et $Gamma(N, \lambda_{i+1}/N)$, nous appliquons un test de Wald afin d'évaluer l'hypothèse :

$$H_0 : \lambda_i = \lambda_{i+1} \quad \text{vs.} \quad H_1 : \lambda_i \neq \lambda_{i+1} \tag{16}$$

La statistique du test de Wald est calculée selon :

$$W_T = \frac{(\hat{\lambda}_i - \hat{\lambda}_{i+1})^2}{\text{Var}(\hat{\lambda}_i) + \text{Var}(\hat{\lambda}_{i+1})} \tag{17}$$

Les segments sont fusionnés lorsque $W_T < \chi^2_{1,\alpha}$ au niveau de signification $\alpha = 0.05$. De plus, les segments dont la longueur est inférieure à un seuil minimal $l_{\min} = \max(2N, 10)$ sont automatiquement fusionnés avec les segments voisins afin de garantir la validité statistique des estimations de paramètres.

2.3 Application aux données GNSS-R aéroportées

Nous appliquons la technique de segmentation du signal radar à la détection de plans d'eau. La Figure A.3 illustre la segmentation du signal radar pour

les mesures GNSS aéroportées acquises le long de la trajectoire de vol. Les images Google Earth présentées dans ces figures montrent les traces des empreintes au sol des satellites, représentées par des localisations de points spéculaires espacées de 20 ms. Le code couleur adopté est le suivant : jaune pour les zones terrestres et bleu pour les plans d'eau. Ces couleurs sont associées automatiquement à différentes valeurs moyennes de réflectivité. Nous présentons également les mesures de réflectivité GNSS correspondant à chaque signal satellitaire et la segmentation du signal obtenue par la méthode radar proposée.

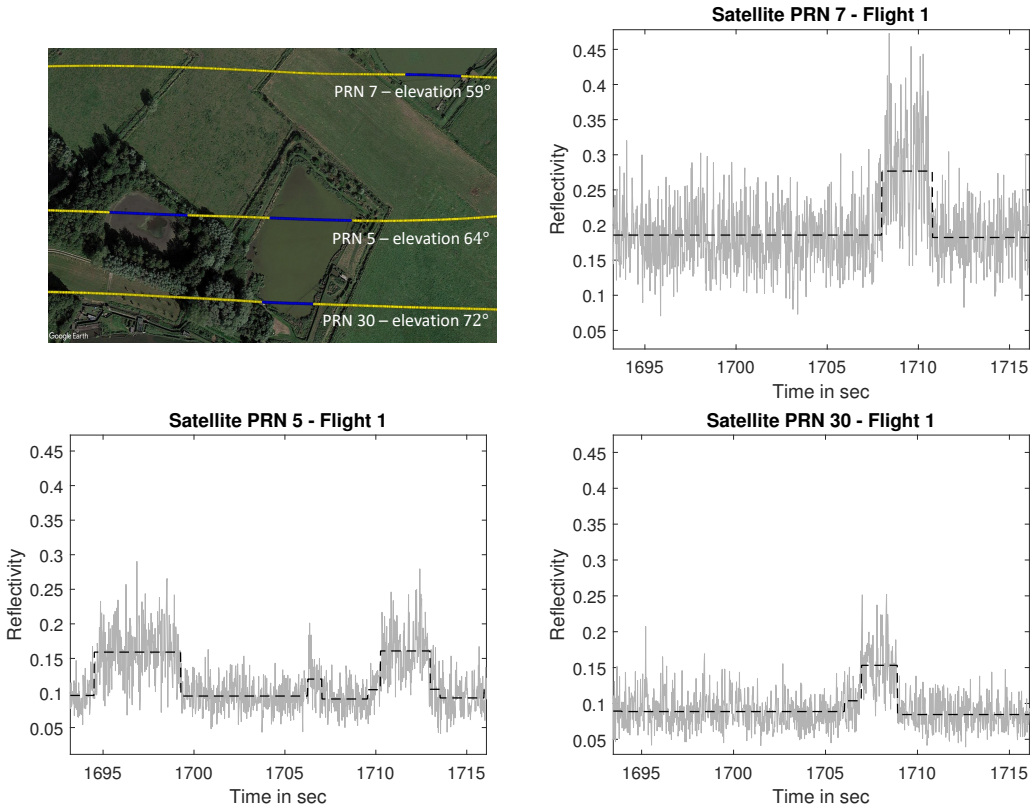
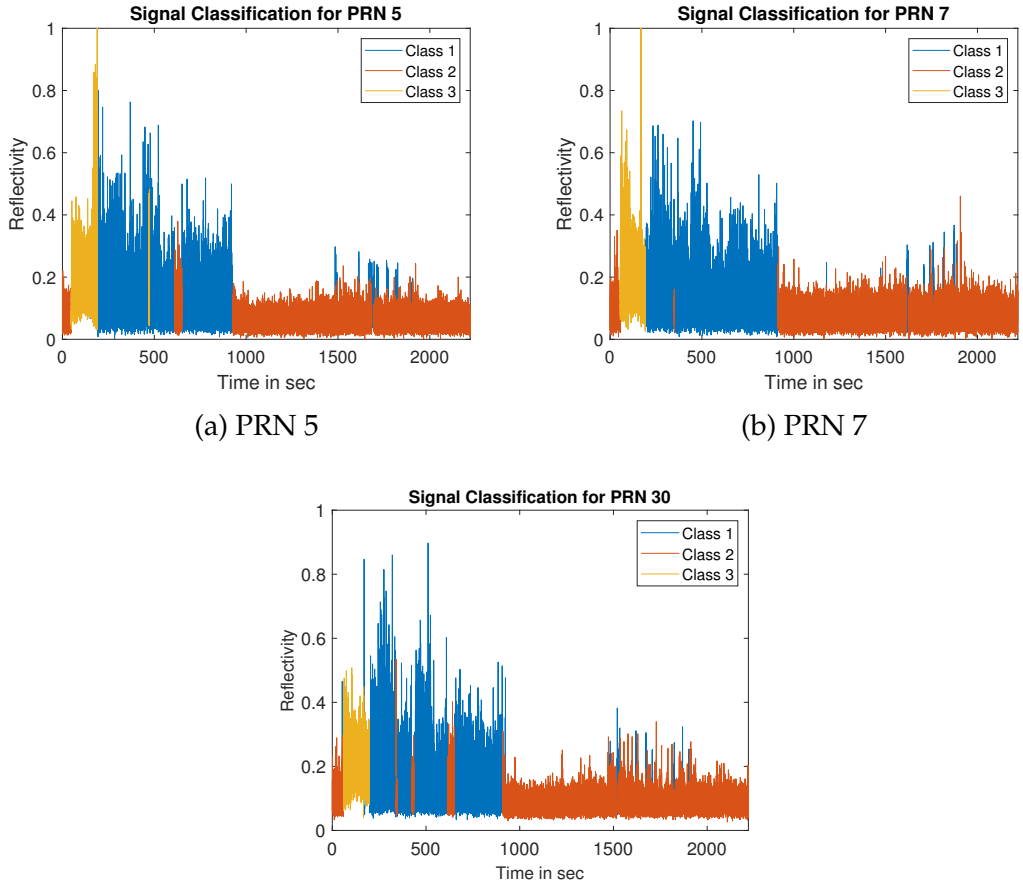


FIGURE A.3: Segmentation des mesures GNSS à (50.877390°, 1.933011°).

Les propriétés statistiques de chaque segment, notamment leur moyenne et écart-type, sont utilisées comme descripteurs d'entrée pour un algorithme K-means de classification automatique des segments détectés selon les types de surface. La Figure A.4 présente les résultats de classification K-means pour plusieurs observations satellites, illustrant une identification cohérente du sable, de la mer, des sols et des plans d'eau continentaux à travers différents satellites.



(a) PRN 5

(b) PRN 7

(c) PRN 30

FIGURE A.4: Classification de la réflectivité pour les satellites PRN 5, 7 et 30.

3 Modélisation de la phase GNSS-R et analyse d'excès Doppler

La cohérence de phase des signaux réfléchis contient des informations encore inexploitées sur les mécanismes de diffusion de surface. Cette section développe des méthodes de traitement tenant compte de la phase, basées sur la modélisation statistique et l'analyse spectrale d'excès Doppler, en vue d'une classification automatique des surfaces.

3.1 Traitement du signal et estimation de la phase

Pour un signal réfléchi issu de réflexions multiples, les composantes en phase et en quadrature s'écrivent :

$$I_{kT_c} = \sum_{i=1}^{M} a_i \cos\left(\phi_{i,0} + \alpha_i T_c(k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_I \tag{18}$$

$$Q_{kT_c} = \sum_{i=1}^{M} a_i \sin\left(\phi_{i,0} + \alpha_i T_c (k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_Q \tag{19}$$

où a_i représente l'amplitude, $\phi_{i,0}$ la phase initiale, et α_i le décalage Doppler de chaque réflexion. La phase estimée est extraite selon :

$$\hat{\phi}_{kT_c} = \arctan(Q_{kT_c} / I_{kT_c}) \tag{20}$$

La somme pondérée de la phase dans le domaine circulaire est définie par:

$$\hat{\phi}_{kT_c} = \arctan 2 \left(\frac{\sum_{i=1}^{M} a_i \sin\left(\phi_{i,0} + \alpha_i T_c (k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_Q}{\sum_{i=1}^{M} a_i \cos\left(\phi_{i,0} + \alpha_i T_c (k-1) + \frac{\alpha_i T_c}{2}\right) f_e \frac{T_c}{2} + \eta_I} \right) \tag{21}$$

où $atan2$ désigne la fonction arctangente sur quatre quadrants.

Les propriétés de cohérence de phase du signal réfléchi constituent un indicateur direct du régime de diffusion sous-jacent. Nous utilisons un modèle statistique de phase fondé sur la distribution de von Mises afin de caractériser la transition entre les régimes cohérents et incohérents. Ce modèle met en évidence les relations entre les propriétés de cohérence, le rapport signal sur bruit et les amplitudes relatives des composantes de réflexion.

Pour les surfaces lisses, la diffusion spéculaire produit des variations de phase cohérentes. Dans ce régime, la phase reste stable car la réflexion est dominée par un seul trajet bien défini, induisant des relations de phase continues qui évoluent de manière régulière avec les changements de géométrie d'observation. À l'inverse, pour les surfaces rugueuses, la présence de multiples trajets de diffusion engendre des interférences complexes conduisant à des variations incohérentes de la phase. La somme vectorielle de ces trajets à phases aléatoires génère un signal composite de faible amplitude, caractérisé par une phase hautement variable dont la distribution tend vers un comportement uniforme. Cette distinction constitue un puissant critère discriminant pour la classification automatique des surfaces.

Afin de pallier la rareté des jeux de données GNSS-R réelles annotées, une méthodologie de génération de données synthétiques basée sur des modèles physiques a été développée. Celle-ci est paramétrée à partir de caractéristiques

spécifiques aux surfaces, extraites de mesures aéroportées réelles. L'approche repose sur un modèle de réflexion multi-trajets, dans lequel les signaux synthétiques sont générés comme une superposition cohérente de composantes multi-trajets, chacune définie par des intervalles empiriques du nombre de réflexions, des décalages Doppler et des niveaux de puissance caractéristiques de différentes surfaces telles que les sols, les plans d'eau, la mer et le sable. L'accord entre les signaux réels et synthétiques confirme la validité physique des modèles sur lesquels repose cette approche.

3.2 Analyse du spectre d'excès Doppler

Pour étudier le comportement Doppler des signaux GNSS-R réfléchis, une transformée de Fourier rapide (FFT) est appliquée au signal complexe de bande de base $I + jQ$. La représentation spectrale de l'excès Doppler normalisé s'exprime comme :

$$\hat{\Gamma}(f)_{kT_c} = \frac{feT_c}{2} \sum_{i=1}^{w} a_i \delta\left(f - fd_i\right) + \Psi \tag{22}$$

où $\delta(f)$ est la fonction delta de Dirac, Ψ est un terme de bruit multiplicatif et f_{d_i} sont les fréquences des pics maximums de la représentation spectrale. Le spectre illustre la répartition de l'énergie du signal réfléchi sur les différentes fréquences Doppler.

Trois critères quantitatifs sont introduits pour caractériser le spectre Doppler et extraire des informations significatives sur la surface cible :

- **Amplitude de pic de la composante Doppler dominante (C_1) :** Ce critère quantifie l'amplitude de la principale composante Doppler présente dans le spectre. Il est défini mathématiquement par :

$$C^1_{kT_c} = \underbrace{\max}_{f \in [-1/2T_c, 1/2T_c]} \hat{\Gamma}(f)_{kT_c} \tag{23}$$

  où T_c désigne le temps d'intégration cohérent utilisé dans le traitement radar. L'amplitude maximale constitue une mesure de l'effet de diffusion cohérente de la surface : des valeurs élevées traduisent généralement un comportement de réflexion plus spéculaire. Ce critère est particulièrement sensible aux mécanismes de diffusion dominants et fournit une indication utile du rapport signal sur bruit global de la mesure.

- **Nombre de pics spectraux statistiquement significatifs (C_2) :** Ce paramètre estime le nombre de pics Doppler du spectre considérés comme significatifs d'un point de vue statistique. Un pic spectral est jugé significatif lorsque son amplitude dépasse un seuil fixé à la moitié de l'amplitude maximale :

$$C^2_{kT_c} = \sum_{f=-1/2T_c}^{1/2T_c} \sum_{i=1}^{p} \gamma(i)\delta\left(f - f_{d_i}\right) \tag{24}$$

$$\gamma(i) = \begin{cases} 1, & \text{si } a_i \geq C^1_{kT_c}/2 \\ 0, & \text{sinon} \end{cases} \tag{25}$$

où $\gamma(i)$ est une fonction permettant de distinguer les pics issus du signal de ceux générés par le bruit, et a_i représente l'amplitude de chaque pic.

En ne retenant que les pics dont le contenu énergétique est significatif, cette méthode fondée sur un seuil permet d'éliminer efficacement les fluctuations dues au bruit. Ce critère agit comme un indicateur de la complexité spectrale et reflète la diversité du milieu de diffusion. Les surfaces présentant une plus grande rugosité ou une combinaison de mécanismes de diffusion affichent en général un nombre plus élevé de pics spectraux, correspondant à une distribution plus large de fréquences de diffusion.

- **Dispersion spectrale des pics (C_3) :** Ce paramètre quantifie la répartition de la puissance spectrale totale entre plusieurs pics significatifs plutôt que sa concentration dans une seule composante dominante. Il est obtenu en soustrayant l'amplitude du pic maximal (C^1) de la somme cumulée des amplitudes des pics significatifs. En utilisant $\gamma(i)$, il s'écrit:

$$C^3_{kT_c} = \frac{feT_c}{2} \sum_{i=1}^{p} \gamma(i)a_i - \underbrace{\max}_{f\in[-1/2T_c, 1/2T_c]} \hat{\Gamma}(f)_{kT_c} \tag{26}$$

La dispersion spectrale traduit la répartition de l'énergie à travers l'ensemble du spectre Doppler. Des valeurs élevées de C^3 indiquent que l'énergie réfléchie est plus largement distribuée sur plusieurs fréquences Doppler, phénomène typique des surfaces fortement rugueuses ou impliquant des mécanismes de diffusion multiples. Cette mesure s'avère

particulièrement utile pour différencier des surfaces présentant des amplitudes de pic similaires mais des propriétés de diffusion distinctes.

Ces trois métriques permettent de distinguer efficacement les surfaces cohérentes (sable, plans d'eau) des surfaces incohérentes (sols, mer agitée) à partir de leur distribution d'énergie spectrale. L'analyse de données aéroportées réelles met en évidence des différences spectrales marquées : les surfaces d'eau calmes présentent des pics dominants d'amplitude élevée et une énergie concentrée, tandis que les surfaces terrestres rugueuses exhibent plusieurs pics associés à une dispersion énergétique plus large.

3.3 Résultats de classification des surfaces

L'intégration de ces trois critères Doppler dans un cadre de classification non-supervisée par K-means constitue une avancée notable par rapport aux méthodes reposant uniquement sur la réflectivité. L'application du regroupement K-means aux trois critères d'excès Doppler a permis d'identifier avec succès quatre classes de surface distinctes, offrant une discrimination améliorée, notamment pour différencier la mer et les rivières, qui présentent des réflectivités similaires mais des mécanismes de diffusion distincts. Les critères tenant compte de la phase offrent une discrimination plus claire que les seules mesures de réflectivité, comme illustré dans la Figure A.5.

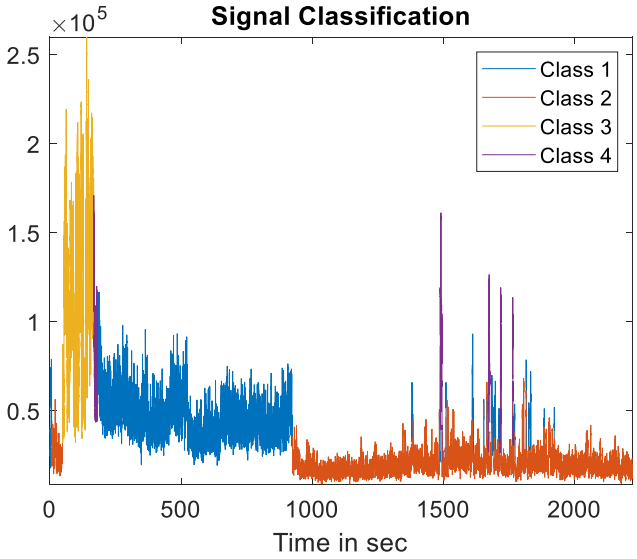


FIGURE A.5: Classification des valeurs de pics maximaux.

Le modèle de réflexion multi-trajets développé, ainsi que la méthodologie de génération de données synthétiques fondée sur l'extraction de paramètres

spécifiques aux surfaces à partir de mesures réelles aéroportées, se sont révélés particulièrement utiles pour le développement d'algorithmes dans un contexte de rareté des données annotées.

4 Apprentissage automatique pour la classification des surfaces

Ce dernier chapitre résume les capacités du cadre proposé, en intégrant les approches issues du traitement physique du signal aux techniques de reconnaissance automatique des formes basées sur l'apprentissage automatique (Machine Learning, ML). Les méthodes analytiques classiques peinent souvent à modéliser les relations non linéaires et la complexité inhérente aux données GNSS-R réelles, faisant du ML une solution adaptée à la classification des surfaces. L'intégration d'algorithmes d'apprentissage automatique traditionnels (Random Forest, XGBoost) et d'architectures d'apprentissage profond (Deep Learning, DL) telles que les réseaux 1D-CNN et Bi-LSTM constitue une méthodologie robuste pour la classification des signaux réfléchis GNSS-R.

4.1 Approche bi-dataset et ingénierie des caractéristiques

Pour pallier le manque généralisé de grands jeux de données réelles GNSS-R précisément étiquetées, une approche fondée sur deux ensembles de données a été adoptée. Le jeu de données synthétiques, utilisé exclusivement pour l'apprentissage, a été généré à partir du modèle de réflexion multi-trajets, en utilisant des paramètres spécifiques à chaque type de surface (nombre de réflexions, caractéristiques fréquentielles Doppler et amplitudes de réflexion) extraits des mesures réelles aéroportées.

Ce processus garantit que les signaux synthétiques conservent les caractéristiques physiques et la complexité observées dans les mesures GNSS-R réelles, renforçant ainsi la validité physique du modèle. Le jeu de données complet issu des mesures réelles a été réservé à la phase de validation indépendante, permettant une évaluation rigoureuse de la capacité de généralisation des modèles.

Une phase d'ingénierie des caractéristiques (feature engineering) a permis d'extraire des métriques statistiques discriminantes à partir de deux canaux

d'entrée : la série temporelle d'amplitude brute et les pics spectraux maximaux dérivés de la transformée de Fourier du signal complexe. Les caractéristiques calculées comprennent le nombre, la somme, la moyenne, l'écart-type, les valeurs minimale et maximale, l'étendue et le coefficient de variation (CV). Leur formulation générale est résumée comme suit :

$$\text{Features} = \text{mean, std, skewness, kurtosis, range, CV, } \ldots \tag{27}$$

L'analyse des statistiques d'ordre supérieur, telles que l'asymétrie (skewness) et la kurtosis de l'amplitude, s'est révélée essentielle, ces métriques mettant en évidence des variations de forme d'onde et de concentration qui distinguent efficacement les classes de surface — par exemple la séparation entre mer et rivière, qui présentent des moyennes et écarts-types similaires. De plus, la corrélation entre le coefficient de variation (CV) de l'amplitude et celui des pics spectraux a permis d'identifier des différences de texture caractéristiques des surfaces de sable, de mer, de sols et de plans d'eau.

4.2 Prétraitement des données et sélection des modèles

Les caractéristiques numériques ont été normalisées par une mise à l'échelle standard (Standard Scaling) afin d'obtenir une moyenne nulle (μ_i) et une variance unitaire (σ_i), garantissant une homogénéité des variables et empêchant certaines d'entre elles de dominer le processus d'apprentissage :

$$x_i' = \frac{x_i - \mu_i}{\sigma_i} \tag{28}$$

Une normalisation Min-Max a également été appliquée pour contraindre les caractéristiques dans l'intervalle $[0, 1]$, ce qui s'avère particulièrement bénéfique pour les modèles d'apprentissage profond, en assurant la stabilité numérique et en améliorant la convergence lors de l'entraînement.

Six modèles d'apprentissage supervisé ont été évalués, englobant à la fois des approches classiques fondées sur des caractéristiques extraites et des architectures de Deep Learning de bout en bout :

1. **Modèles classiques (apprentissage sur caractéristiques extraites)** : Régression logistique, Random Forest, XGBoost et Perceptron multicouche (MLP). XGBoost, une implémentation efficace du gradient boosting, a été choisi pour sa capacité démontrée à traiter des problèmes complexes de régression et de classification non linéaires.

2. **Modèles d'apprentissage profond (apprentissage sur signaux bruts)** : Le réseau de neurones convolutif unidimensionnel (1D-CNN) a été sélectionné pour sa capacité à identifier automatiquement des motifs locaux dans des données séquentielles unidimensionnelles, tandis que le réseau à mémoire longue bidirectionnelle (Bi-LSTM) a été employé pour capturer les dépendances temporelles dans les deux directions le long des séquences.

Les modèles classiques ont été optimisés par validation croisée à 5 plis sur l'ensemble synthétique d'apprentissage, tandis que les modèles d'apprentissage profond ont été entraînés à l'aide de l'optimiseur Adam avec arrêt anticipé (*early stopping*).

4.3 Performance et analyse de classification

Les modèles ont d'abord été testés sur un sous-ensemble de validation synthétique indépendant afin d'établir une référence de performance, puis validés sur les données GNSS-R réelles aéroportées pour évaluer leur capacité de généralisation. Tous les modèles ont atteint une précision supérieure à 98% sur les données synthétiques, confirmant leur aptitude à apprendre les motifs issus des modèles physiques.

La phase critique a consisté à évaluer les performances sur les données réelles. Comme attendu, une baisse de performance — souvent appelée *écart simulation–réalité* (sim-to-real gap) — a été observée, en raison de la complexité et du bruit non contrôlé présents dans les signaux réels.

4.3.1 Comparaison des performances des modèles

Les résultats, résumés dans le Table A.1, mettent en évidence le transfert d'apprentissage réussi du domaine synthétique au domaine réel, les meilleurs modèles atteignant une précision de 90% malgré un entraînement exclusivement effectué sur des données synthétiques.

Les modèles XGBoost (méthode d'ensemble) et 1D-CNN (méthode d'apprentissage profond) se distinguent comme présentant les meilleures performances, atteignant tous deux une précision de 0.90 sur les données réelles. Ce résultat valide à la fois le réalisme physique de la génération de données synthétiques et la robustesse des architectures de classification. Les modèles Random Forest (0.88) et MLP (0.87) montrent également d'excellentes

TABLE A.1: Comparaison des performances des modèles d'apprentissage automatique et profond sur les données GNSS-R synthétiques et réelles.

	Données de test synthétiques		**Données de test réelles**	
Modèle	**Précision**	**F1 Score**	**Précision**	**F1 Score**
Régression logistique	0.98	0.98	0.85	0.83
Random Forest	0.99	0.99	0.88	0.86
XGBoost	0.99	0.99	0.90	0.89
MLP	0.98	0.98	0.87	0.85
1D-CNN	0.98	0.98	0.90	0.88
Bi-LSTM	0.93	0.92	0.82	0.80

performances, confirmant la pertinence des caractéristiques conçues. En revanche, le modèle Bi-LSTM présente la plus forte dégradation, suggérant une sensibilité accrue au surapprentissage des motifs séquentiels spécifiques aux données synthétiques.

4.3.2 Visualisation de la généralisation

La capacité de généralisation, mesurée par l'écart de performance entre les ensembles de test synthétiques et réels, est illustrée dans les figures suivantes. La Figure A.6 compare les précisions, tandis que la Figure A.7 présente les F1-scores, métrique plus robuste en cas de déséquilibre des classes. Ces représentations confirment visuellement que les modèles XGBoost et 1D-CNN conservent la meilleure généralisation en conditions réelles.

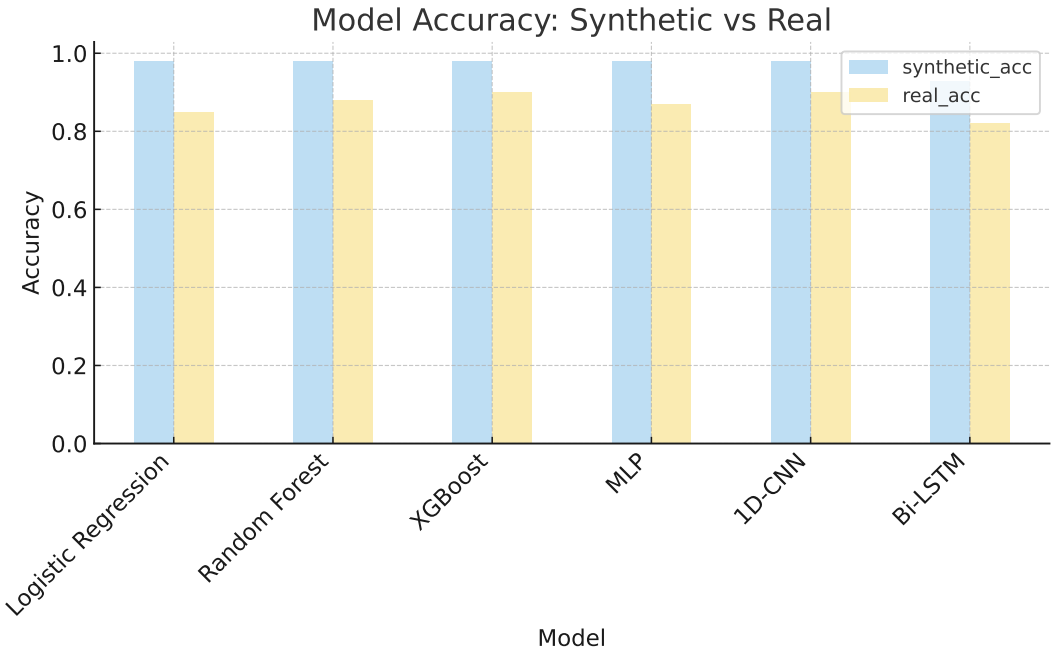


FIGURE A.6: Comparaison de la précision des modèles entre les données de test synthétiques et réelles.

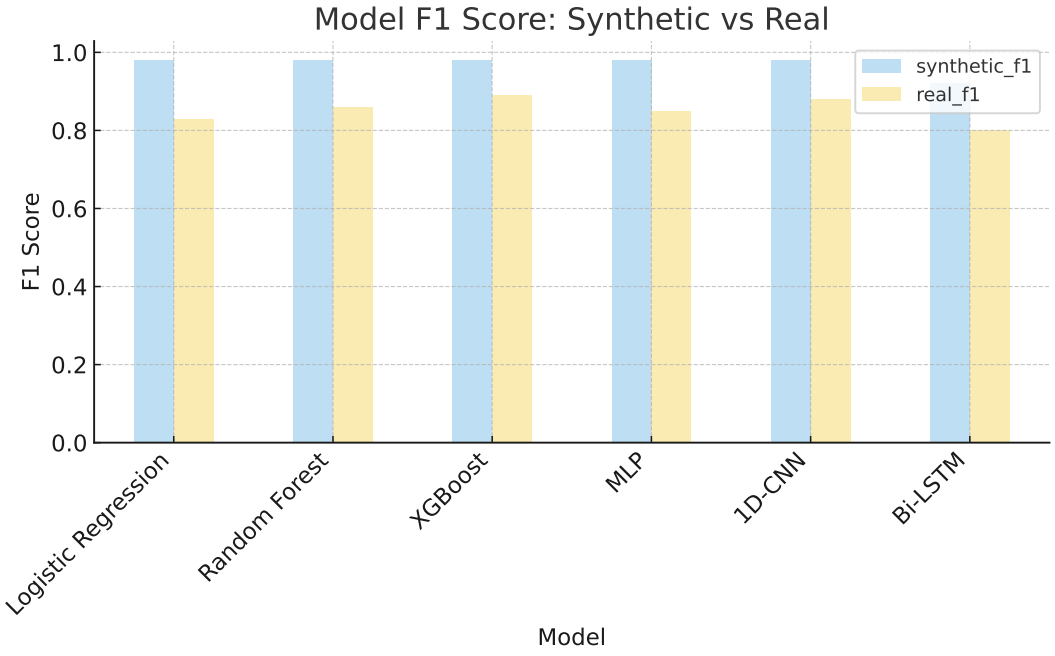


FIGURE A.7: Comparaison du F1-score des modèles entre les données de test synthétiques et réelles.

Résumé

La caractérisation des sols par télédétection est devenue un enjeu majeur pour l'observation de la Terre. La réflectométrie GNSS (GNSS-Reflectometry, GNSS-R), technique radar bistatique exploitant les signaux de navigation réfléchis des constellations satellitaires GNSS existantes, est prometteuse dans ce cadre. Les méthodes GNSS-R conventionnelles reposent principalement sur l'analyse de l'amplitude, tandis que les informations de phase et les caractéristiques fréquentielles offrent un potentiel encore sous-exploité pour une différenciation plus fine des surfaces.

Cette thèse propose un cadre de segmentation fondé sur la distribution Gamma, utilisant une transformation homomorphique pour convertir le bruit multiplicatif affectant les données en bruit additif à variance constante. Cette approche permet l'application d'un algorithme de détection de ruptures CUSUM avec des seuils déterminés statistiquement ainsi que d'une technique de localisation précise fondée sur l'estimation du maximum de vraisemblance Log-Gamma. Un modèle de phase basé sur la distribution de von Mises est proposé pour décrire la transition entre diffusions cohérente et incohérente. Une étude spectrale définissant trois critères Doppler d'amplitude, de complexité spectrale et de dispersion d'énergie est également proposée. Une méthodologie de génération de données synthétiques, fondée sur un modèle physique et paramétrée à partir de mesures réelles, pallie la rareté des jeux de données GNSS-R annotés pour l'application de modèles d'apprentissage automatique.

Les trois approches proposées sont confrontées à des mesures GNSS-R acquises lors d'une mission apportée réalisée le long d'une trajectoire de 70km en nord de France, démontrant leurs capacités à différencier plages, mer, plans d'eau et sols. Ce travail étend les capacités de la réflectométrie GNSS au-delà des applications océanographiques, ouvrant la voie à une observation à haute résolution des milieux terrestres et hydrologiques, avec un potentiel démontré pour l'estimation du taux d'humidité des sol, la caractérisation de la végétation et la détection des changements environnementaux.

Mots-clés: Réflectométrie GNSS ; radar bistatique ; traitements phase-aware ; distribution Gamma ; détection de rupture ; analyse Doppler ; apprentissage automatique ; classification de surface ; génération de données synthétiques.

Abstract

Surface characterization through remote sensing has become increasingly critical for Earth observation applications, with GNSS-Reflectometry emerging as a promising bistatic radar technique that utilizes reflected navigation signals from existing satellite infrastructure. Conventional GNSS-R methods have relied mainly on amplitude-based observations, whereas phase information and frequency characteristics of reflected signals represent an underexploited potential for enhanced surface differentiation. Advancing beyond these approaches requires sophisticated frameworks that can extract and use this rich information content for automated classification across diverse terrestrial and aquatic environments.

The main methodological contributions provide a Gamma distribution-focused segmentation framework using homomorphic transformation to transform multiplicative noise to constant-variance additive noise, allowing CUSUM change point detection with statistically determined thresholds and Log Gamma Maximum Likelihood Estimation to achieve accurate localization. A von Mises statistical phase model characterizes the coherent-to-incoherent scattering transition. This is supplemented by three excess Doppler spectrum criteria that quantify peak amplitude, spectral complexity, and energy dispersion. Additionally, a physics-based synthetic data generation methodology is parameterized using real measurements, addressing the challenge of limited labeled real datasets.

Airborne GNSS-R measurements acquired during a 70 km flight mission over northern France demonstrate robust performance of the proposed approach across diverse surface types, including sandy beaches, open ocean, inland water, and varied land covers. The segmentation scheme demonstrated a high level of consistency in classification across several satellite observations, while machine learning algorithms, trained only with synthetic data, achieved 90% accuracy on real-world measurements. This work advances GNSS-R capabilities beyond oceanographic applications to enable high-resolution terrestrial and hydrological monitoring with demonstrated potential for soil moisture estimation, vegetation characterization, and environmental change detection.

Keywords: GNSS-Reflectometry; bistatic radar; phase-aware processing; Gamma distribution; change point detection; excess Doppler analysis; machine learning; surface classification; synthetic data generation.