Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(34th cycle)

# Global Navigation Satellite Systems as Signals of Opportunity for Environmental Applications

## Reflectometry and Scintillation

By

**Rayan Imam**

******

**Supervisor(s):**
Prof. Fabio Dovis, Supervisor

**Doctoral Examination Committee:**
Prof. A.B. , Referee, University of...
Prof. C.D, Referee, University of...
Prof. E.F, University of...
Prof. G.H, University of...
Prof. I.J, University of...

Politecnico di Torino
2022

# Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Rayan Imam
2022

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

*I would like to dedicate this thesis*

*to my parents, for empowering and supporting me through my formal education journey and beyond,*

*and to the memory of my dear uncle Farah Hajnour.*

# Acknowledgements

I would like to express my endless gratitude to my supervisor, Professor Fabio Dovis, for his guidance and support throughout my four years study at the Navigation Signal Analysis and Simulation group. Thank you for sharing the knowledge and experience in GNSS and related applications, and for the opportunities you opened for me during this PhD and beyond.

I would like to acknowledge the contribution of my colleagues at Politecncio di Torino, Nicola Linty, Wenjian Qin and Caner Savas, for our collaboration on GNSS scintillation. I would like to acknowledge the role of Nicola in manually labelling the scintillation data to be used for training and validating the machine learning models.

I would like to acknowledge the contribution of Marco Pini, Micaela Troglia Gamba, Mario Nicola, Gianluca Falco and Gianluca Marucco, at LINKS foundation's Space and Navigation lab, for our collaboration on GNSS reflectometry. The in-field data collection was provided by Marco Pini and Gianluca Marucco, while the in-lab calibration was carried out in collaboration with Micaela Troglia Gamba, Mario Nicola and Gianluca Falco. I would like to aknowledge also Gabriella Povero, at LINKs foundation's Space and Navigation lab, for providing me assistance with equatorial scintillation data.

I would like to acknowledge Lucilla Alfonsi, Luca Spogli, Claudio Cesaroni, Giorgiana De Franceschi and Vincenzo Romano, at Istituto Nazionale di Geofisica e Vulcanologia (INGV), for supporting my work on high latitude Scintillation. I would like also to acknowledge their contribution to my work while hosting my research mobility at INGV including data curation and manuscript reviewing.

Finally, I would like to thank my colleagues Alex Minetto and Andrea Nardin for our countless discussions on GNSS topics, especially the signal processing and the receiver. Thank you for being the best PhD colleagues and the brightest GNSS researchers/engineers.

# Abstract

The proposals on opportunistically utilizing Global Navigation Satellite Systems' (GNSS) signals for environmental monitoring came to light hand in hand with proposing the GNSS systems themselves. GNSS satellites broadcast signals that can be exploited for positioning, navigation and timing anywhere and at any time on Earth. The high temporal and geographic availability of the GNSS satellites, together with particular properties of the GNSS signals, made them an ideal remote sensing tool. Here, the satellites are considered as radio signals sources to be utilized by passive remote sensing sensors. Many applications have been successfully deployed based on this concept including GNSS-Reflectometry (GNSS-R), and GNSS scintillation.

Today, less than 50 years after the first GNSS satellite was launched and less than 30 years after the first GNSS system was declared fully operational, GNSS systems contribution to environmental monitoring is evident in our daily life. Opportunistic use of GNSS signals as remote sensing data sources is already contributing to weather forecast and environmental hazards monitoring, including hurricanes and floods. Moreover, the upper atmosphere environment, that is affected by both space and Earth environmental conditions, is monitored using GNSS signals. Monitoring the ionospheric conditions is important because it affects directly and indirectly many of the critical infrastructures on Earth. Scintillation, which is rapid random fluctuations in the amplitude and phase of radio waves resulting from the signal passing through plasma density irregularities in the ionosphere, is the most important source of ionospheric errors that affect GNSS systems in particular.

In this thesis, detecting floods using special GNSS receivers and a technique known as GNSS–R is discussed. Also, detecting ionospheric perturbed conditions leading to ionospheric scintillations using another type of special GNSS receivers, known as Ionospheric Scintillation Monitoring (ISM) receivers, is discussed.

For floods detection, the feasibility of using data from GNSS–R receivers mounted onboard small Unmanned Aerial Vehicles (UAVs) are investigated. Taking into account the constraints imposed on/by such platforms, a signal processing methodology that respects such constraints when processing the signals reflected by the potential floods is implemented. The method roughly estimates the reflected signal strength with respect to the noise floor, and it is demonstrated that such rough estimation is capable of accomplishing the water detection task with high accuracy. Also, the importance of UAV-based multi-constellation GNSS–R receivers for accurate estimation of water extents is investigated. Moreover, the importance of the UAV-based GNSS–R sensor calibration is discussed. It is demonstrated that, for floods detection, the calibration is not strictly necessary. However, for estimating the amount of reflected power and the height of the water surface (i.e extracting more parameters of the water/floods), calibrating the sensor is necessary.

For scintillation detection, the utilization of Machine Learning (ML) algorithms for detecting scintillation in GNSS data to overcome the known limitations of the scintillation detection metrics is discussed. ML techniques has demonstrated success in many scientific fields including GNSS signal processing and scintillation modelling. It is demonstrated that ML models are able to perform better than the scintillation metrics thresholds in scintillation detection. It is also demonstrated that ML models can detect scintillation in environments where interference, multipath in particular, exists. In this regard, 98% detection accuracy, 2% scintillation miss-detection and 2% scintillation false alarm is demonstrated for a model developed for equatorial scintillation, while 95% accuracy, 5% scintillation miss detection and 5% scintillation false alarm are demonstrated for high latitude phase scintillation detection. The abilities of various ML algorithms, with different hyper-parameters, in carrying out the detection task is also investigated. It is found that bagged decision trees give superior performance. Finally, various measurements available in ionospheric scintillation monitoring records as well as in high rate post-correlation data are investigated. For low latitude scintillation, the signal intensity measurements at high rate (50 Hz) are essential and sufficient. On the other hand, it is found that 3-minutes Total Electron Content (TEC) data are able to surpass the scintillation metrics in detecting polar scintillations. This could open the door for unprecedented expansion in global scintillation monitoring by utilizing a wide range of professional GNSS receivers, other than the ISM receivers, that are already providing TEC measurements.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Environmental factors are known to affect Global Navigation Satellite Systems (GNSS) signals, including the effects of multipath, electromagnetic interference, atmospheric delays and ionospheric scintillations [1]. Research on mitigating these effects, or reducing their impact, has been going on for decades. On the other hand, many applications look at these effects as an opportunity to monitor the environment by observing the aforementioned factors. For example, multipath/reflections from water surfaces can actually be utilized to monitor water bodies and retrieve some of the water parameters.

GNSS has been utilized for decades in environmental applications. We differentiate here between applications that utilize the geo-location information provided by GNSS receivers and applications that process the GNSS radio signal itself for environmental monitoring. The former utilize geo-referencing to tag environmental measurements. For example, GNSS positioning has been utilized to geo-reference shorelines and then monitor spatial variation in the coast line to identify coastal erosion. This utilization of GNSS services for environmental monitoring is not the focus of this thesis and more information can be found for example in [2] and the references therein.

On the other hand, processing the GNSS radio signal for environmental applications, is the focus of this thesis. Here, we monitor the GNSS signal itself to extract information about the environment. We thus use the GNSS signal as a remote sensing signal, that interacts with the environment of interest, and therefore it holds information about this environment. An example of this remote sensing concept is the multipath from the water surfaces mentioned above, where the reflected signal interacted with the water body and thus it carries some information about the reflecting surface.

On this regard, GNSS systems have been utilized as a source of remote sensing signals in many configurations [3] including Radio Occultation (RO), GNSS–Reflectometry (GNSS–R) and Ionospheric Scintillation sounding. In GNSS RO, atmospheric profiles (e.g. temperature, height and pressure) are derived from GNSS signals that are received by a satellite behind the Earth. These signals were refracted (bend) by the Earth's atmosphere and thus they hold data about the atmospheric environment they have passed through. ROs are beyond the scope of this thesis, but their applications in environmental monitoring are so evident in our life today, including their contribution to weather forecast and tropical cyclones prediction services [4].

GNSS–R is a remote sensing concept where GNSS signals reflected by objects like water and ice are processed to extract measurements about the reflecting surface. GNSS–R has been operational in various environmental monitoring applications, including sea surface wind speed retrieval [5–7], river flow estimation [8], soil moisture retrieval [9, 10], hurricanes monitoring and tracking [11], sea ice remote sensing [12] and flood inundation monitoring [13–19]. The latter application in particular will be the focus of Part I of this thesis.

These days, such applications leverage the amount of data available mainly from the space–borne GNSS–R sensors [20], some of them are part of GNSS–R constellations [21–23], besides air-borne and ground based GNSS–R sensors. In this thesis, we will be investigating air-borne GNSS–R sensors to support flood monitoring applications. Here the sensor can be mounted on-board a small Unmanned Aerial Vehicle (UAV) that flood assessment operations can utilize on demand to assess the situation during and after floods.

For the ionospheric remote sensing domain, GNSS provide a valuable tool for studying the upper layers of the atmosphere. GNSS trans–ionospheric signals carry information about the propagation environment they pass through [24]. Thus, they have been utilized as signals of opportunity for observing the ionosphere including monitoring the Total Electron Content (TEC) [25], ionospheric storms [26], Travelling Ionospheric Disturbance (TID)s [27] and ionospheric electron density irregularities [28, 29]. This monitoring is essential for understanding the environment at our planet's upper atmosphere, which directly and indirectly affect life on the lower parts of the atmosphere and vice-versa.

In fact the ionosphere is a complex part of our planet, and the weather and climate at this layer is affected by both the space above and the atmospheric layers and lithosphere beneath, making it a really complicated and interesting environment to study.

In Part II of this thesis we will focus on monitoring and detecting ionospheric scintillation which occurs when the ionosphere is in disturbed states.

## 1.1 Problem Definition

The expansion in research on GNSS as signals of opportunity for remote sensing applications created various opportunities for utilizing these signals, and introduced challenges for researchers to enable this utilization. In particular, the increase in the amount of available data collected by the various satellite constellations and ground monitoring systems invites researchers to benefit from these infrastructures, as well as introduced challenges in processing these data for new applications that were not reported in the original plans, or when the data are new to the researchers. Also, being based on GNSS theory and technology, it is the duty of GNSS researchers to facilitate the utilization of these data and make them available in higher level more suitable for scientific utilization.

GNSS–R in particular was first introduced for ocean studies. In these studies the spatial, temporal resolution is on a level different that those targeted by recent flood studies that demand meter level resolutions in case of river stretches for example. Furthermore, in flood scenarios, the data is expected to include more than just water, for example buildings and cars, not like in the case of oceans, and thus new demands from the signal processing are introduced. Finally, being proposed to be implemented on small UAVs, new aspects are to be investigated that were not the case for space-borne GNSS–R for example the constraints introduced by the small size of the UAV platforms.

Scintillation data on the other hand has expanded tremendously in the last decade. With almost 30 years of GNSS-based ionospheric scintillation monitoring, there are considerably big amount of scintillation data that are attractive for scientific studies but challenging to curate. The main reason for this challenge is the fact that GNSS-based scintillation monitoring is affected by many environmental factors including interference from the environment and also anomalies in the GNSS satellites and the receiver itself. This sensitivity of the GNSS signals is the main enabler for utilizing the signal for opportunistic monitoring of the ionosphere, but also invites extensive curation of the data before they are used for scientific applications. If we want to open the door for trusted automated processing of the repositories of ionospheric scintillation GNSS measurements for ionospheric modelling and scientific research in general, we need to provide trusted tools for curing the scintillation data that researchers can utilize. For this reason, we will be focusing on detecting scintillation in GNSS-based measurements. We will be considering data from commercial receivers as well as from custom built receivers.

## 1.2 Objectives and Methodology

The objective of this thesis is to investigate two of the state-of-the-art topics in GNSS remote sensing for environmental applications: GNSS–R for floods monitoring and GNSS scintillation detection using Machine Learning (ML). More precisely, the feasibility of detecting floods using GNSS–R sensors mounted on–board small UAVs will be investigated. These platforms are frequently used for floods assessment in general. By including such sensors on–board these platforms, data from GNSS–R can be integrated with other measurements for floods assessment. GNSS–R is considered a passive sensor, and its size is relatively small. Thus it can be mounted easily on–board UAVs. It also works on all weather conditions, and thus it is not limited by the bad weather states (the UAV might be). For this, we will be processing GNSS signals received by a GNSS–R sensor custom build for the i–REACT project that developed solutions for efficient responses to emergencies. The study will investigate detecting water bodies as small as ponds and artificial lakes for similarity to water extents in case of floods. Also, simulations will be carried out to analyse the expected capabilities of multi–constellation GNSS–R (multi–GNSS–R) sensors in floods detection.

For ionospheric scintillation, we investigate the recent applications of ML algorithms for scintillation detection. The motivation behind using these techniques rise from the need to overcome some of the limitations of the traditional metrics for scintillation detection. Being the result of a phenomenon driven by a complex process, namely the evolution and decaying of ionospheric electron density irregularity, monitored using a signal that is sensitive to many factors in the propagation environment, made the scintillation detection task difficult and the metrics susceptible to interference from the environment and the instruments themselves. ML techniques are well established in modelling systems where the physics underlying the phenomenon is complex. For this we investigate ML for detecting amplitude and phase scintillation. We investigate in details the selection of the GNSS measurements suitable for scintillation detection using ML models, i.e. we investigate the feature engineering part of the ML modelling cycle. We show statistical results of applying the developed models on scintillation detection.

## 1.3 Thesis Contribution and Outline

The main contributions of this thesis can be summarized as follows:

- Assess the feasibility of UAV-based GNSS–R for flood monitoring operations. The thesis focuses on the UAV-based GNSS–R sensor capability to estimate the existence of GNSS signals reflected by flood water.

- Simplify the signal processing in order to meet the constraints introduced on/by the small UAV platforms forseen for this GNSS–R application. In particular, reflected signal acquisition assisted by the direct signal tracking is proposed. A metric from the acquisition theory, $\alpha_{mean}$ is adopted as the metric for detecting water.

- Assess the feasibility of estimating the reflected signal strength and the possibility to achieve flood water detection relying only on processing the reflected signal, without estimating the exact amount of direct and reflected powers. The calibration needed in order to estimate the latter was demonstrated.

- Justify the importance of knowledge about the terrain in the area under study in order to geo–reference the UAV-based GNSS–R sensor measurements to the reflecting surface.

- Assess the ability of the UAV-based GNSS-R sensor to the estimate the flight height, the needed sensor internal delay calibration, the foreseen signal processing to refine the flight height estimation, and the feasibility of achieving flood water detection without the need to implement the flight height estimator.

- Investigating supervised ML techniques for polar and equatorial scintillation detection. It is demonstrated that a ML model able to achieve this task is obtainable. In particular, bagged decision trees, which are known in the literature to succeed in modelling noisy data, was demonstrated to achieve superior performance for scintillation detection.
- Various ISM receivers measurements were investigated as input features for the ML model. It is found that, for detecting equatorial scintillation, high rate (50 Hz) signal intensity measurements are essential. On the other hand, TEC measurements, which are provided by a wide range of GNSS receivers, not only ISM receivers, is demonstrated to achieve the best results in detecting high latitude phase scintillation. This invites future research to consider scintillation monitoring using GNSS receivers that have higher geographic and temporal distribution with respect to ISM receivers.
- The feasibility of training a ML model able to distinguish scintillation from multipth, which is the main impairment that contaminate scintillation measurements mimicking amplitude scintillation, is demonstrated. It is shown that implementing such model increased the amount of scintillation data that would have been discarded by conservative multipath avoidance thresholds.

The original GNSS–R sensor used in this work [30] was developed with the objective of retrieving the moist land water content. The original sensor implements two down–looking antennas to receive the Left-Hand Circular Polarized (LHCP) and Right-Hand Circular Polarized (RHCP) reflected GNSS signals. In this work, a simplified version of the sensor that uses only one down–looking LHCP antenna is used. All the GNSS–R signal processing proposed as part of this thesis is developed accordingly.

The scintillation data utilized in this thesis were retrieved from previous scintillation data collection campaigns. The equatorial scintillation data were collected in Hanoi as part of the Equatorial Ionosphere Characterization in Asia (ERICA) project [31, 32]. The Antarctic data is continuously being collected at SANAE IV base as part of the DemoGRAPE project [33].

The rest of this thesis is organized as follows:

- In Chapter 2, GNSS is briefly introduced with focus on the aspects that allow the GNSS signals to be utilized opportunistically for remote sensing. This includes the systems' global coverage, the number of available GNSS satellites, and the various signals frequencies. Furthermore, the concept of radio navigation is introduced and the GNSS radio signal structure is presented. The Chapter is concluded by presenting the equation of the complex signal at the output of the GNSS signal processing block. This complex signal is the basic information that was utilized as input for both the reflectometry and scintillation analysis that was done in this thesis.

- The body of the thesis is grouped in two parts. Part I (Chapter 3 and 4) for GNSS–R:

  - In Chapter 3 the basic concepts of GNSS–R are covered with focus on applying these concept to UAV-borne GNSS–R sensors. In particular, the estimation of the specular point position from the GNSS–R sensor coordinates is presented. The importance of knowledge about the terrain in the area under study is justified for detecting floods extends with high accuracy from UAV-based GNSS–R measurements. The feasibility of estimating the flight height is presented justifying the need for sensor internal delay calibration and further signal processing implementation. The feasibility of detecting water presence using UAV-based GNSS–R was detailed. The possibility to further estimate the amount of reflected power from UAV-based GNSS–R was presented, justifying the need for well calibrating the sensor internal gains.
  - The detailed GNSS–R experiment methodology and the results obtained are shown in Chapter 4. The original sensor developed in [30] is briefly described. The simplified version of the sensor used in this thesis is explained. The post processing methodology developed in this thesis is detailed including the parameter proposed as the metric for estimating the reflected signal strength. Results of the proposed signal processing are shown with three case studies of water detection that prove the feasibility of flood–like water detection. Results of estimating the water extends, the water surface area, and the benefits of multi–GNSS–R are presented. The results for the sensor delay calibration are shown.

- Part II (Chapter 5 - 8) for scintillation:

  - GNSS scintillation literature and theoretical background are given in Chapter 5. A brief introduction about the ionosphere is given, describing the scintillation effect and why monitoring scintillation is important. The effects of the ionosphere on GNSS signals are detailed. The metrics for measuring scintillation are detailed. Then, the GNSS measurements at the various GNSS receiver stages are introduced justifying their with potential for scintillation detection. Finally, the scintillation detection mechanism implemented in the literature are presented.
  - The aspects of ML essential for this thesis are introduced in Chapter 6. The cycle of a ML model from idea to deployment is described. The ML algorithms utilized in this thesis are introduced. Then the choices on the ML implementation are justified.
  - Chapter 7 shows the methodology and results of high latitude phase scintillation detection using ML. The data preparation is described and the input features to the models are introduced. Results of scintillation detection are presented and the model performance is compared with the standard scintillation detection methods.
  - Chapter 8 is dedicated to the methodology and results of equatorial amplitude scintillation detection. The data preparation steps are detailed. Statistics about the data set are presented. The input features to ML models are introduced. Results from more than 40 trained models are shown and compared. Results of testing the best obtained model in operative scenarios are shown.

- The thesis is concluded in Chapter 9 giving a summary of all the results and final remarks.

# GNSS

# Chapter 2

# Global Navigation Satellite Systems

## 2.1 A Brief History of GNSS

The history of the GNSS systems we know today dates back to the seventies of the last century when in orbit experiments and validations for the American Global Positioning System (GPS) and the Soviet/Russian GLobal Orbiting NAvigation Satellite System (GLONASS) were taking place.

The early GPS experiments launched radio navigation satellites in the period spanning 1974-1989 as what is known in the literature as Navigation Technology Satellites (NTS) and Navigation Development Satellites (NDS). A total of 13 satellites were launched in that phase providing global navigation and timing with limited availability and continuity. These satellites prepared for the successful full deployment of the system with 24 active satellites in the period of 1989-1995, when GPS was declared operational. Around the same period, GLONASS system was developed and launched. GLONASS was also declared fully operational in 1995 with 24 active satellites. Before the end of the millennium, another two GNSS constellations were already announced, the European Global Navigation Satellite System (Galileo) [34] and the Chinese BeiDou Navigation Satellite System (BDS or BeiDou).

Even if the main purpose of GNSS systems is the provision of positioning and navigation services, opportunistic utilization of the GNSS signals for remote sensing purposes came to surface even before the GPS and GLONASS systems were declared fully operational. In 1980, Rino et.al [35] demonstrated the ability to measure amplitude and phase scintillation indexes using a custom built GPS ionospheric monitoring receiver. The foreseen global coverage of dual-frequency L-band trans–ionospheric GPS signals derived the motivation behind the study. By 1993 [36], commercial GPS receivers for ionospheric scintillation monitoring were presented. Today, GNSS based ionospheric monitoring is providing various ionospheric monitoring data including TEC and ionospheric scintillation monitoring metrics. For the latter, hundreds of professional GNSS receivers dedicated for scintillation monitoring, known as Ionospheric Scintillation Monitoring (ISM) receivers, have been installed over the globe. These receivers have been providing continuous measurements of the ionospheric scintillation indexes for the last two decades.

At the same time, studies proposing utilizing GNSS signals as Radio Detection And Ranging (RADAR) transmitters were first published in 1988 [37], 1993 [38] and 1996 [39] for possible applications in ocean wind surface estimation and altimetry. By 1997, Garrison et. al [40] proved that they were able to process GPS signals reflected by ground surfaces using a conventional GPS receiver mounted on-board an aircraft. By 2003 [41], passive GNSS radar sensors on-board satellites have been demonstrated proofing the feasibility of the sensor on space-borne platforms and providing measurements on a global scale. This concept, known as GNSS–R, has found wide applications in environmental monitoring activities specially ocean wind estimation and ice monitoring. Today, GNSS–R sensors are installed on fleets of satellites as well as fixed monitoring stations. On demand and experimental GNSS–R equipped flights are also available. More deployments are planned and new applications have been proposed in the last five years.

The objective of this thesis is the investigation of modern techniques for the use of GNSS for environmental monitoring. In particular, the interest is focused on the use of advanced techniques for the monitoring of the ionosphere, by using artificial intelligence algorithms. A second field of applications that has been investigated is the use of GNSS–R for water detection. This latter has direct application in assessing the effects of natural disasters originating from Earth, floods in this study.

In the rest of this chapter, a brief summary about the current status of the GNSS constellations that are operational today is given in order to understand the abundance in radio sources that inspired the opportunistic use of GNSS signals for remote sensing applications. Then a description of the signals broadcast by GNSS satellites is provided to clarify also why utilizing these signals is attractive and feasible. Then the basic operational concept of GNSS for navigation purposes is given (although GNSS signals are not utilized for this purpose here, this section is essential to understand the basic concept of radio navigation that made all these applications possible). Finally, the concept of GNSS for remote sensing is detailed focusing on reflectometry and ionospheric scintillation.

## 2.2 GNSS Constellations

Deployed to provide navigation signals on a global scale, there are 4 GNSS systems operational today: Galileo, GPS ( originally Navstar GPS), GLONASS and BeiDou . There exist also other systems that broadcast navigation signals on regional scale (Regional Navigation Satellite Systems (RNSS)) including Japan's QZSS, India's IRNSS (also known as NavIC), Europe's EGNOS, America's WAAS and China's Beidou regional.

The four GNSS constellations are deployed in Medium Earth Orbit (MEO), except for BeiDou which has satellites also in Inclined Geosynchronous Orbit (IGSO) and Geostationary Earth Orbit (GEO). The current status of the four constellations is reported in Table 2.1 [42]. The table shows the status of GPS (As of April 12, 2022 [43]), GLONASS [44], Galileo [45] and BeiDou [46]. Note that early GLONASS-M satellites only transmit Frequency Division Multiple Access (FDMA) signals on G1 and G2. GLONASS-M and GLONASS-M+ launched after 2014 broadcast a Code Division Multiple Access (CDMA) signal on G3.

Over the last 20 years alone, the number of operational GNSS satellites have increased from only 44 satellites in 2002, to 123 satellites as of 2022. as shown in Figure 2.1 (Figure source [47]). This increase in the number of satellite has direct benefits not only to the navigation and timing users, but to other applications that utilize the GNSS signals opportunisticly.

From a user on ground point of view, this high number of satellites in orbit translates to a high number of visible satellites as seen by a receiver at a given time. The average number of visible GNSS satellites at any location on Earth is shown in the map of Figure 2.2 (Figure source: [42]). Figure 2.2a shows the satellite visibility for the stand alone systems, while Figure 2.2b shows the map for the multi–constellation GNSS (multi–GNSS) case when all the global constellations are

Table 2.1 Current status of the four GNSS constellations.

| System | Orbits | Operational Satellites | Civil Signals Broadcast |
|---|---|---|---|
| GPS | MEO | 7 Block IIR | L1 C/A |
| | | 7 Block IIR-M | L1 C/A + L2C |
| | | 12 Block IIF | L1 C/A + L2C + L5 |
| | | 4 GPS III/IIIF | L1 C/A + L2C + L5 + L1C |
| GLONASS | MEO | 24 GLONASS-M, | G1 + G2 + [G3] |
| | | 2 GLONASS-M+ | G1 + G2 + [G3] |
| | | 2 GLONASS-K1 | G1 + G2 + G3 + G2a |
| Galileo | MEO | 24 operational | E1, E5a, E5b, E5(a + b), and E6 |
| BeiDou | GEO | 6 | B1I, B2I and B3I |
| | MEO | 3 BDS-2 | |
| | IGSO | 7 | |
| | GEO | 3 | B1I, B1C, B2a, B2b, and B3I |
| | MEO | 26 BDS-3 | |
| | IGSO | 5 | |

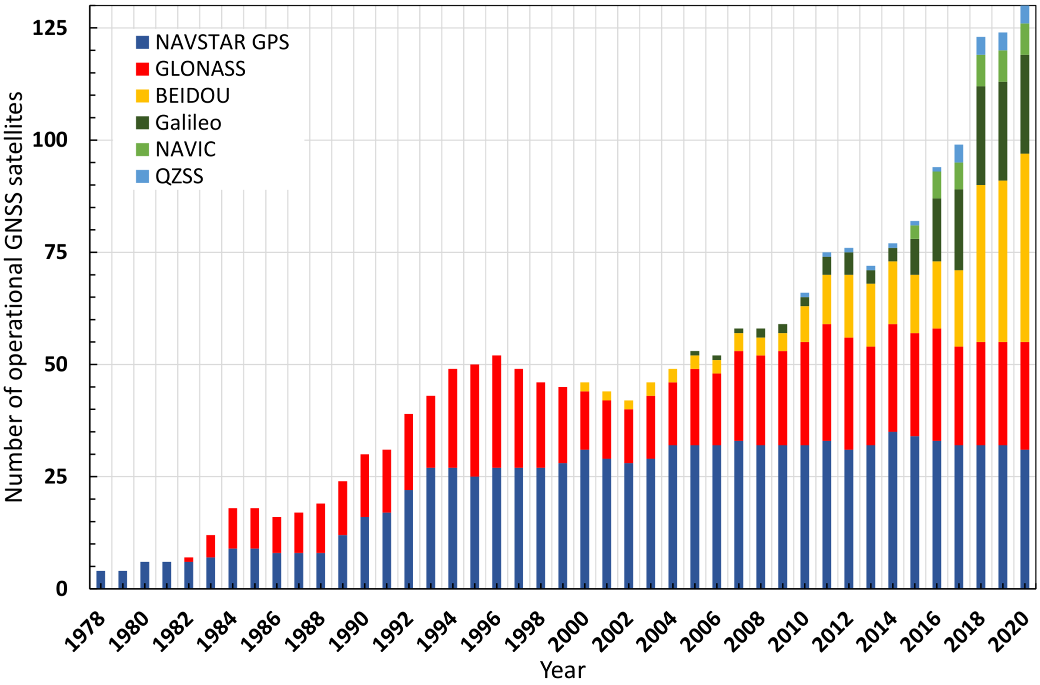


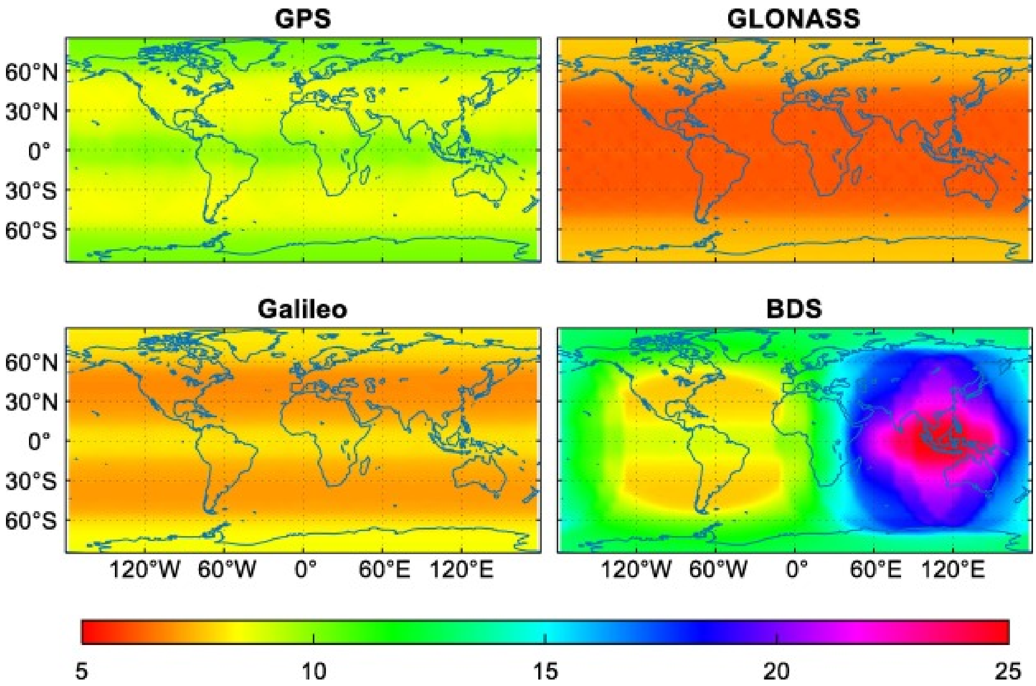
Fig. 2.1 Number of operational GNSS and RNSS satellites per constellation from 1978-2020.

utilized. It can be seen that the number of visible satellites is much higher in the far east compared to the west because of Beidou's coverage in this figure, in addition to the concentrated existence of Japan's, China's and India's regional systems, not shown in Figure 2.2 but reported in Figure 2.1. Also, in general, the number of visible satellites near the equator is high compared to the high latitudes because of the inclination of the GNSS satellites' orbits. Nevertheless, the number of visible satellites is always much greater than 16 (4 satellites from each constellation) which is the minimum number of satellites needed to compute a stand alone positioning using the various constellations independently.

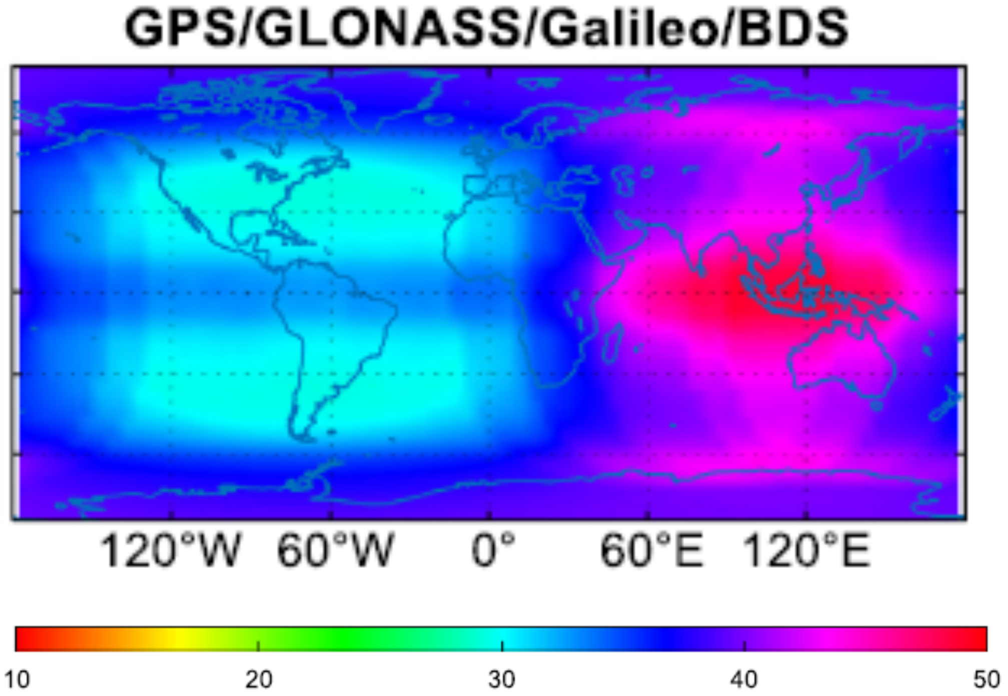
This high number of satellites results in increased availability and accuracy of the multi–GNSS navigation service (i.e. when a multi-constellation receiver is used) because of many factors. First of all, the high number of satellites provides better geometrical distribution of the satellites in the sky which translates to a smaller Geometric Delusion of Precision (GDOP) error (the more spread the satellites, the lower the error). Moreover, in places with reduced sky visibility, e.g. mountains and urban canyons, the availability of the positioning solution is increased. Regarding the accuracy, the authors in [48] show that the standard deviation of the positioning error of an open service receiver working in L1/E1 band decreases from 75 to 40 to 25 to 18 meters when the number of visible satellites increase from 4 to 8 to 12 to 16 relatively.

Another beneficiary of this increased number of GNSS satellites is the field of remote sensing. For these applications, the increase in the number of radio sources automatically increases the amount of remote sensing measurements. In these cases, Figure 2.2 translates to 25-50 simultaneous ionospheric measurements, and the same number of ground reflections as observed by a receiver. In Section 4.2.1 we discuss in detail the benefit of multi–GNSS for UAV-based reflectometry. We do not discuss the benefits of multi–GNSS for scintillation studies as they are beyond the scope of this work. We refer the reader to [49–51] for examples on why multi–GNSS is important for receivers working under scintillation conditions, and also for receivers observing ionospheric irregularities.

Next, we explore the navigation signals transmitted by these satellites and we explain why are these signals useful for reflectometry and scintillation.



(a) Stand alone systems cases



(b) Multi-GNSS case.

Fig. 2.2 Map display of the average number of visible satellites for the four GNSS systems. Sampling interval 5 minutes. cuttoff elevation angle 10°. Spatial resolution $2.5° lat, 5° lon$ ($72 \times 72$ grids). Virtual receiver at the centre of each grid at height 25m. Analysis period (25-31 Oct 2020).

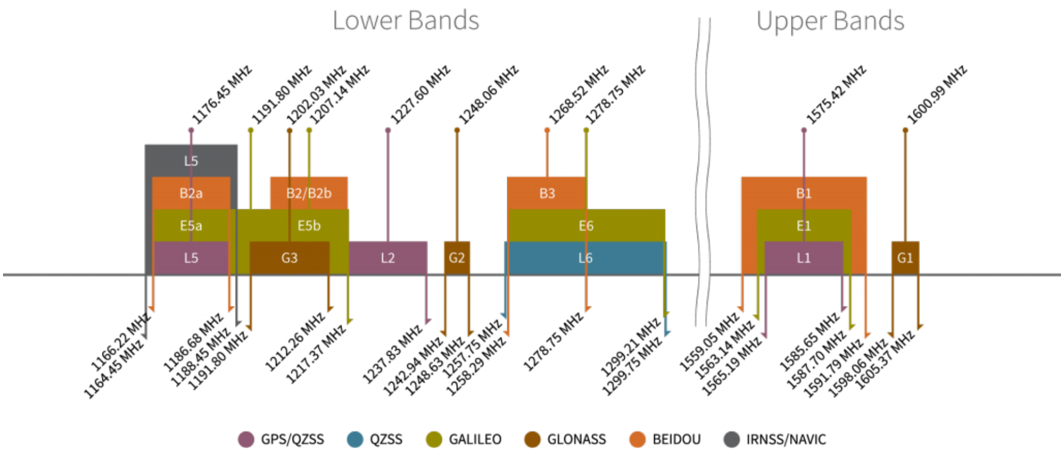


Fig. 2.3 GNSS navigational frequency bands.

## 2.3 GNSS Navigation Signals Frequencies

GNSS satellites transmit multiple navigation signals in the frequencies dedicated to Radio Navigation Satellite Services at the L–band (1-2GHz). These frequencies are illustrated in Figure 2.3 (image credit [52]) where three bands in particular can be highlighted:

- L5: between 1164 and 1215 MHz
- L2: between 1215 and 1300 MHz
- L1: between 1559 and 1610 MHz

The different constellations refer to the exact bands they utilize using different names (for example E5a, E5b, E6 and E1 in the case of Galileo) with the central frequencies and bandwidths shown in Figure 2.3. Moreover, multiple signals (can be at the same frequency) are transmitted by each GNSS system. For example, a GPS satellite transmits 5 signals at 3 different L–band frequencies at the same time. The complete list of navigation signals transmitted by the different constellations are shown in Table 2.2 (source: [53]). Some of these signals are restricted to specific users while others are available for free and commercial use.

L–band signals are able to penetrate rain, clouds, fog and storms which make them ideal for providing navigation signals even under severe weather conditions. However, these signals are affected by vegetation (where their power is attenuated)

Table 2.2 GNSS Frequencies and Signals.

| System | Signal | Frequency (MHz) |
|---|---|---|
| GPS | L1 C/A | 1575.42 |
| | L1C | 1575.42 |
| | L2 C | 1227.6 |
| | L2 P | 1227.6 |
| | L5 | 1176.45 |
| GLONASS | L1 C/A | 1598.0625-1609.3125 |
| | L2 C | 1242.9375-1251.6875 |
| | L2 P | 1242.9375-1251.6875 |
| | L3 OC | 1202.025 |
| Galileo | E1 | 1575.42 |
| | E5a | 1176.45 |
| | E5b | 1207.14 |
| | E5 AltBOC | 1191.795 |
| | E6 | 1278.75 |
| BeiDou | B1l | 1561.098 |
| | B2l | 1207.14 |
| | B3l | 1268.52 |
| | B1C | 1575.42 |
| | B2a | 1176.45 |
| | B2b | 1207.14 |

and buildings (where they get diffracted and reflected). This leads to degraded navigation signals where the open sky conditions are unavailable. Moreover, the signals in this band are susceptible to passage into the ionosphere which introduces delays and/or random fluctuations in the L-band received signals when the geospace is perturbed (see Section 5.4).

Special measures to counteract the corruption of the signals due to ionospheric propagation and multipath from the surrounding environment (buildings, mountains, ice, lakes, etc) can be set up at receiver level or post-processing the data. At the same time, such weakness make the GNSS signals ideal for the remote sensing applications we are targeting in this thesis, i.e ionospheric scintillation and reflectometry [54] as will be discussed in Section 5.4 and Section 3.1 respectively.

On the other hand, the motivation behind broadcasting signals at multiple frequencies by each satellite is related to ionospheric propagation errors. The GNSS services are based on estimating the propagation time of signals from multiple satellites to a receiver. Since the satellites are orbiting at a height of approximately $20,000$ km in space, their signals pass through the ionosphere and the troposphere before arriving at a receiver on ground. These mediums introduce changes in the propagation of the radio wave that, if not accounted for, will introduce errors in the estimation of the propagation time. The most effective way to eliminate these errors is to use signals broadcast on multiple frequencies from the same source. This elimination is generally applied by dual frequency GNSS receivers that are able to mitigate propagation errors without external aids. For ionospheric sciences, this actually translates to observing the effect of the ionosphere (the bulk of the information stands into the aforementioned errors) as will be discussed in Section 5.4. For reflectometry applications, these multi-frequency signals provide the chance to monitor a phenomenon at different frequencies. It has demonstrated benefits in many applications specially in ground-based reflectometry sensors. Multi-frequency GNSS–R is beyond the scope of this thesis, but recent examples of its implementation can be found for example in [55, 56].

In conclusion, the GNSS systems provide abundance of signals that are well suited for ionospheric remote sensing as well as for sensing reflected signals by a wide category of ground objects including water surfaces.

## 2.4 Navigation Using GNSS

In this section a brief description of the operation of the GNSS systems is provided. In its core, GNSS is a timing instrument that is able to provide precisely timed signals that can be exploited for *cm* accuracy measurements as well as a time transfer accuracy in the order of $1-2$ *ns* [57]. This capability is the key enabler for the exploitation of GNSS signals for scientific uses beyond the positioning and navigation services, for example for geodesy, timing and remote sensing.

In radio navigation, the distance between the 3-dimensional position of the receiver ($P_R$) and a satellite $i$ ($P_{S_i}$) can be indicated by (2.1):

$$d_R^{S_i} = |P_{S_i} - P_R| = c(t_R - t_{S_i}), \tag{2.1}$$

where $c$ is the speed of light, $t_{s_i}$ is the epoch at which the signal left satellite $i$, and $t_R$ is the epoch at which the signal arrived at the receiver. The position of the satellite is known as well as the epoch of transmission. The epoch of reception is measured and thus the only unknown is the position of the receiver. An error of $1\mu s$ in measuring this time difference, translates to approximately $300m$ error in the estimation of the distance between a satellite and a receiver. For this reason, GNSS receivers need to estimate with high accuracy the signal flight time, and also the GNSS systems need to provide the satellite coordinates and timing with high accuracy.

However, the receiver's clock is actually not synchronized with the satellite's clock. On the other hand, the GNSS satellites are well synchronized among each other and the overall GNSS system offset from the reference Coordinated Universal Time (UTC) is known. To achieve this synchronization and reference time keeping, GNSS systems rely on atomic clocks on-board the satellites.

The GNSS satellites broadcast their Coordinated Universal Time (UTC) time offset in the navigation message in addition to the epoch of transmission. Because of the bias in the receiver clock (and other propagation delays), the receiver then estimates not the actual range, but a pseudo-range from satellite $i$ to the receiver ($\rho_i$). This translates (2.1) to:

$$\rho_i = c[(t_R - t_{ref}) - (t_{S_i} - t_{ref})] + \varepsilon_d, \tag{2.2}$$

where $t_{ref}$ is the reference UTC time and $\varepsilon_d$ is a term that incorporates the additional delays due to the ionosphere (as it will be seen in section 5.4), troposphere and the receiver hardware [57]. The receiver clock offset is unknown and will be estimated by the receiver as a fourth unknown besides the unknown 3-dimensional coordinates.

Now that we have introduced the measurements that the receiver needs to obtain in order to estimate the range to a satellite, and the 4 unknowns that it needs to solve, we next introduce how the positioning and navigation equations are constructed.

The principle of estimating the position of a receiver using GNSS satellites is illustrated in Figure 2.4. The main job of a GNSS receiver is to estimate the propagation time $\tau_i$ of the signal from satellite $i$ to the receiver antenna. The satellite coordinates $(x_s^i, y_s^i, z_s^i)$ and time are known, therefore four $\tau_i$ measurements are needed to estimate the coordinates of the receiver $(x_R, y_R, z_R)$ and the receiver clock offset $(\Delta\tau)$. Finally, using at least four measured $\tau_i$ values and applying the equation of distance between two points in 3–Dimensional space, the coordinates and time offset of the receiver clock with respect to the GNSS time scale can be estimated.

$$\begin{aligned}
\rho_1 &= \sqrt{(x_R - x_s^1)^2 + (y_R - y_s^1)^2 + (z_R - z_s^1)^2} + b_1, \\
\rho_2 &= \sqrt{(x_R - x_s^2)^2 + (y_R - y_s^2)^2 + (z_R - z_s^2)^2} + b_2, \\
\rho_3 &= \sqrt{(x_R - x_s^3)^2 + (y_R - y_s^3)^2 + (z_R - z_s^3)^2} + b_3, \\
\rho_4 &= \sqrt{(x_R - x_s^4)^2 + (y_R - y_s^4)^2 + (z_R - z_s^4)^2} + b_4, \\
&. \\
&. \\
\rho_n &= \sqrt{(x_R - x_s^n)^2 + (y_R - y_s^n)^2 + (z_R - z_s^n)^2} + b_n
\end{aligned}$$

where $\rho_i$ is the pseudo-range between satellite $i$ and the receiver, and $b_i$ is the extra range due to the receiver clock bias and the other errors. $b_n$ for each satellite is different from the others because the different satellites' signal follow different propagation paths specially during their propagation in the ionosphere.

In fact, in Figure 2.4, it is explicitly illustrated that the GNSS signal passes through a couple of propagation mediums before it reaches a receiver on Earth. Thus, $\tau_i$ measured by the receiver does not represent signal travelling with the speed

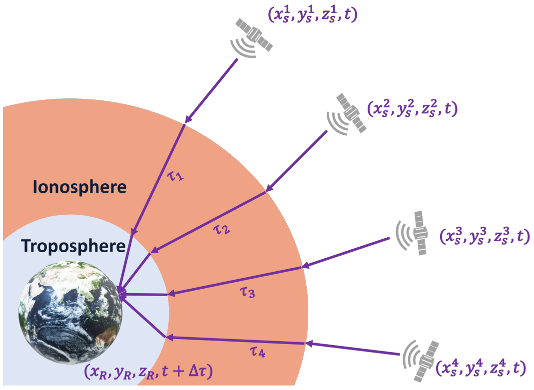


Fig. 2.4 Illustration of GNSS based position estimation concept

of light in a straight line between the satellite and the receiver. The bending is significant when low elevation (less than 15° [58]) signals are to be processed, for example in RO. It will not be the case in this thesis and thus the bending effects will not be further addressed. Most importantly, the propagation medium affects the speed of the signal. This eventually affects the estimation of the distance between the satellite and the receiver. For scintillation, this propagation delay is important and it will be addressed in Section 5.4. For our GNSS-R application, however, we will be comparing the delay of the signal received from the satellite with the signal reflected by the surface. Since the sensor is flying at low latitudes, the atmospheric propagation delays are equivalent for both signals and thus they will cancel each other, as will be illustrated in Section 3.3. Thus, for the GNSS–R part, atmospheric propagation delays will not be further discussed.

Finally, it is important to highlight that many simplifications were applied to make this topic focused on the applications that this thesis is addressing. These simplifications do not address for example, and not limited to, relativistic effects, satellite clock stability, satellite orbital errors, etc [59]. These simplifications are acceptable for this study because, as will be seen in Chapter 3 and 5, we will focus on aspects that affect the power of the received signal and/or those that dual frequency receivers can not totally eliminate.

Next, the principles for how does the GNSS receiver processes the signal to estimate $\tau_i$ is presented.

### 2.4.1 The GNSS Signal

GNSS satellites broadcast communication signals that are designed specifically to be used for Positioning, Navigation and Timing (PNT). For a generic GNSS signal, the signal transmitted by satellite $i$ has the structure:

$$x_{RF,i}(t) = \sqrt{2P_{TX}}\; c_i(t)\; s_c(t)\; d_i(t)\; \cos(2\pi f_i t + \theta_0), \tag{2.3}$$

where $P_{TX}$ is the transmitted signal power. $c_i(t)$ is the ranging code which is unique for satellite $i$. $s_c(t)$ is a sub–carrier that modulates some of the modern GNSS signals (if no sub–carrier is present, $s_c = 1$). $d_i(t)$ is the navigation data, which is a low rate message carrying information about the satellite health status in addition to some parameters that are needed to estimate the range from the receiver to the satellite. $f_i$ is the carrier frequency and $\theta_0$ is a random initial phase.

This signal propagates from the satellite to a receiver on Earth passing through space and the different layers of the atmosphere.

The signal reaches the receiver attenuated, delayed, and with a different phase with respect to the transmitted signal. The signal described by (2.3) at the receiving antenna input becomes:

$$y_{RF,i}(t) = \sqrt{2P_{RX,i}(t)}c_i(t-\tau_i(t))s_c(t-\tau_i(t))d_i(t-\tau_i(t))\cos\gamma_i(t), \tag{2.4}$$

where $P_{RX,i}(t)$ is the received power at time $t$. $\tau_i(t)$ is the time the signal took to travel from the satellite to the receiver. The signal arrives with frequency and phase $\gamma_i(t) = 2\pi(f_i + f_{D,i})t + \theta_i(t)$. $f_{D,i}$ is Doppler shift resulting from the relative motion between the satellite and the receiver. $\theta_i(t)$ is the phase of the received signal at time $t$. Received power can be described as [60]:

$$P_{RX,i}(t) = P_{TX}L_{LOS,i}(t)L_{atm,i}(t), \tag{2.5}$$

where $P_{TX}$ is the nominal transmitted power. $L_{LOS,i}(t)$ is the free space loss across the line of sight range, $L_{atm,i}(t)$ is the power loss due to atmospheric propagation.

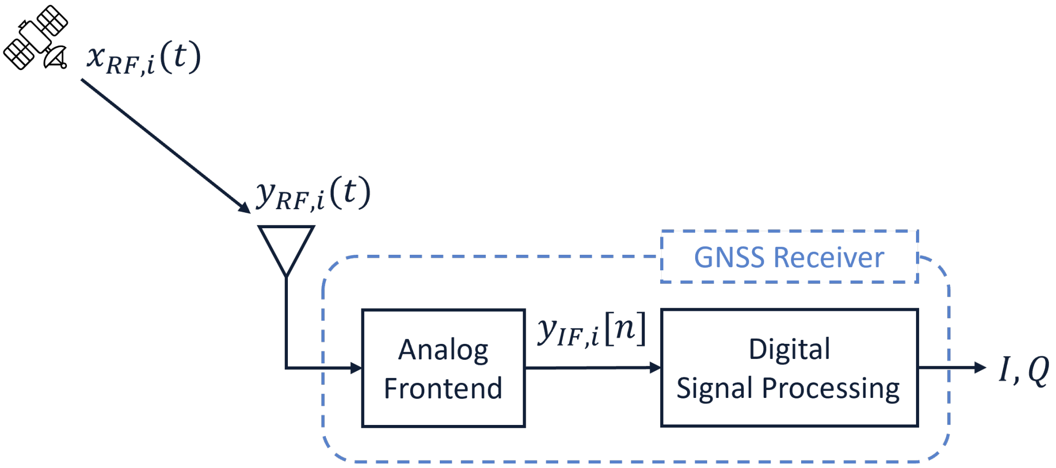


Fig. 2.5 GNSS receiver architecture

### 2.4.2 The GNSS Receiver

Figure 2.5 shows the general GNSS receivers architecture, without the PNT unit. The antenna receives the electromagnetic signals from the visible GNSS satellites, and feeds them at the input of the front–end as electrical signals. The receiver front–end amplifies the signal power, down–converts the carrier frequency to an intermediate frequency, and digitizes the received signal. The front–end, on the other hand, adds undesirable thermal noise to the received signals. The output of the front–end is thus the digital signal:

$$y_{IF}[n] = \sum_{i \in Vis} y_{IF,i}[n] + \eta[n], \tag{2.6}$$

where $Vis$ is the set of visible satellites, and $\eta[n]$ is the thermal noise. The signal for satellite $i$ at the output of the front–end $y_{IF,i}[n] is$:

$$y_{IF,i}[n] = \alpha\sqrt{2P_{RX,i}[n]}\; c_i[n-\tau_i[n]]\; d_i[n-\tau_i[n]] \cos\, \phi_i[n], \tag{2.7}$$

where $\alpha$ is the power gain introduced by the receiver antenna, front end and digitization process, and $\phi_i[n]$ is the phase of the signal down–converted to intermediate frequency. The next step is digital signal processing, where the signal parameters are estimated.

### 2.4.3 GNSS Signal Processing

The core job of a GNSS receiver is to estimate the propagation time of the signal. It does so by estimating the delay of the code $\hat{\tau}_i[n]$, and the phase of the carrier, $\hat{\phi}_i[n]$ for each of the visible satellites. For simplicity we drop $[n]$ from $\{\tau_i[n], \hat{\tau}_i[n], \phi_i[n], \hat{\phi}_i[n], P_{RX,i}[n]\}$ for the rest of the thesis.

To retrieve the signal for satellite $i$, the receiver multiplies the received signal by a complex local replica of the satellite's code and carrier $c_i[n-\hat{\tau}_i].e^{j\hat{\phi}_i}$, and filters out the high frequency component. The resulting signal is:

$$s_i[n] = \langle y_{IF}[n]\ c_i[n-\hat{\tau}_i]e^{j\hat{\phi}_i}\rangle = A\ \sqrt{P_{RX,i}}\ d_i[n-\tau_i]c_i[n-\tau_i]c_i[n-\hat{\tau}_i]e^{j\delta\hat{\phi}_i} + \xi[n], \tag{2.8}$$

where $\langle . \rangle$ is the low pass filter, $\delta\hat{\phi}_i = \phi_i - \hat{\phi}_i$, and $A$ is a term that collects the constants and the amplification factor resulting from the filter and the front–end. $\xi$ includes all the noise terms including the thermal noise and the signals from the other satellites.

The signal $s_i[n]$ is processed by the receiver in chunks that are multiples of the code length, for example multiples of 1ms for GPS L1/CA signal. The receiver integrates and dumps all the samples in the chunk and outputs two values $I_i[m]$ and $Q_i[m]$:

$$r_i[m] = I_i[m] + jQ_i[m] = \frac{1}{L}\sum_{n=1}^{L} s_i[n] = Ad_i[m-\tau_i]\frac{1}{L}\sum_{n=1}^{L} B\sqrt{P_{RX,i}}e^{j\delta\hat{\phi}_i} + \varepsilon[m], \tag{2.9}$$

where $L$ is the number of samples in the chunk of signal. $B = c_i[n-\tau_i]\ c_i[n-\hat{\tau}_i] \in \{1,-1\}$. The real and imaginary parts of $r_i[m]$ are known as the in–phase ($I$) and quadrature–phase ($Q$) components of the signal respectively. The amplitude of $r_i[m]$ is thus an estimation of the average received power, multiplied by a receiver dependent factor, and the navigation message ($d_i[n] \in \{1,-1\}$). The quality of this estimation is affected by the receiver's capability to estimate $\tau_i$ and $\phi_i$.

If $\delta\hat{\phi}_i = 0$ and $\delta\hat{\tau}_i = \tau_i - \hat{\tau}_i = 0$, i.e. the receiver is perfectly estimating the code delay and the carrier phase, and taking into account that $\frac{1}{L}\sum_{n=1}^{L} c_i[n-\tau]\ c_i[n-\hat{\tau}]|_{\delta\hat{\tau}=0} = 1$ (properties of ranging codes) , then:

$$r_i[m]|_{\delta\hat{\phi}_i=0,\delta\hat{\tau}_i=0} = Ad_i[m-\tau_i]\frac{1}{L}\sum_{n=1}^{L}\sqrt{P_{RX,i}} + \varepsilon[m], \tag{2.10}$$

i.e. $r_i[n]|_{\delta\hat{\phi}_i=0,\delta\hat{\tau}_i=0}$ would be a real signal that only contains the navigation message multiplied by a factor proportional to the average received power. The general expression for the raw signal intensity is:

$$SI_{i,raw}[m] = \sqrt{I_i^2[m] + Q_i^2[m]}. \tag{2.11}$$

GNSS receivers are designed to continuously refine $\hat{\phi}_i$ and $\hat{\tau}_i$ keeping $I$ as maximum as possible and $Q$ almost zero. This tracking capability degrades in presence of interference like Radio Frequency Intereference (RFI), multipath and scintillation. What looks like the failure of the receiver to perfectly estimate $\hat{\tau}_i$ and $\hat{\phi}_i$ is in fact also an estimation of the dynamics of the signal as perceived by a GNSS receiver.

In this thesis we focus on this $SI_{i,raw}$ and utilize it for two different applications. The first is for GNSS–R, where our objective is to observe this value and infer it to the surface that reflected the signal on the ground. The second application is for scintillation, where we observe how this signal intensity changes depending on the condition of the ionosphere that it passes through.

For this reason two different receivers will be used, that provide this same value but they are for different applications. The GNSS-R receiver is designed to look at GNSS signals that are reflected by ground objects. It deals in particular with very weak GNSS signals (weaker than the already weak GNSS signal) and they implement multiple antennas to monitor the direct and reflected GNSS signals. On the other hand, ISM receivers implement advanced signal processing techniques to maintain tracking the signal under challenging scintillation conditions [61]. Both the GNSS–R receiver and the ISM receiver process the GNSS signal, not to obtain the receiver position but rather leverage the known position of the receiver to aid processing the GNSS signal.

The details of processing the GNSS signal for reflectometry and scintillation monitoring will be detailed in Chapter 3 and 5 respectively.

# Part I

# GNSS-Reflectometry

# Chapter 3

# GNSS Reflectometry

## 3.1 GNSS-Reflectometry Concept

A navigation GNSS receiver uses an antenna to look at the direct signals broadcast by GNSS satellites. If we point the GNSS antenna towards the ground, and record the GNSS signals reflected by ground objects, this is called GNSS reflectometry. A GNSS–R receiver is interested in observing the signals reflected by ground objects. For this, it implements a down–looking antenna, besides the up–looking antenna, to observe the reflected GNSS signals.

The concept of the GNSS–R is depicted in Figure 3.1. In this illustration, the signal from a GNSS satellite gets reflected by water, and is received by an airborne GNSS sensor. The satellite signal is incident on all the ground area and is strongly reflected by water with reflection angle equal to the incident angle. Thus, the location of the receiver determines which ground measurement is observable. By moving the sensor, for example on a drone along the trajectory illustrated in black dashed lines, ground measurements are collected from the different reflection points.

From remote sensing perspective, the fact that a GNSS signal was reflected, indicates that an object that reflects signals in L-band exists. These objects include water, ice and moist soil, and do not include dry soil and trees for example. We can estimate the position of this reflecting surface (Section 3.2), the height of the flight above the reflecting surface (Section 3.3) and the amount of power reflected by the surface (Section 3.4).

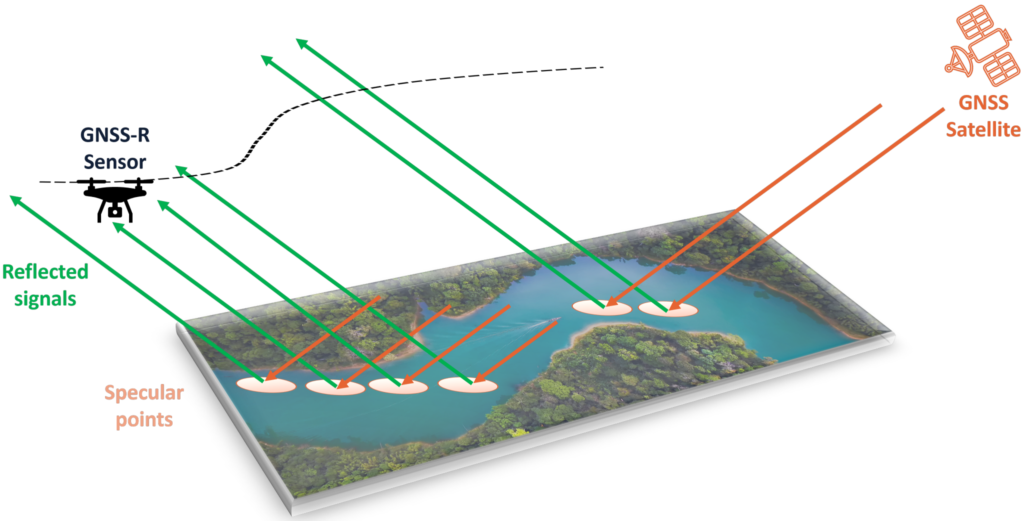


Fig. 3.1 Illustration of the GNSS–R concept

## 3.2 The Specular Point

We can estimate the position of the reflecting surface (i.e. the specular point) knowing the position of the sensor (precisely the GNSS–R antennas), the elevation angle of the transmitter (the GNSS satellite broadcasting that specific signal), and the topology of the terrain in the area of interest. Figure 3.2 illustrates the estimation of the position of the reflecting surface indicated with P in the figure. Here, the satellite elevation $\theta_{EL}$ is known, as well as the receiver position $(x_R, y_R, z_R)$. Being a specular point, the reflection angle $\theta_r$ is equal to the incident angle $\theta_i$. Therefore, its possible locations are restricted to the set of points on terrain surface that fulfil the incident/reflection angles equality. These are the points where the equation of a point in 3-dimentional with a slope corresponding to the reflected signal intersect with the receiver position. From 3D Euclidean geometry, we recall the equation of a line with direction vector $\vec{\boldsymbol{d}} = (l, m, n)$, i.e. the direction of the reflected signal line, that passes through the specular point $(x_P, y_P, z_P)$, i.e. the known point in the line, is:

$$\frac{x - x_P}{l} = \frac{y - y_P}{m} = \frac{z - z_P}{n} \tag{3.1}$$

Thus, if the receiver position $(x_R, y_R, z_R)$ is on the line (3.1) then the signal reflected by this specular point is indeed received by the GNSS–R sensor. The set of all specular points for which the receiver fulfils their reflected signal line equation constitute the set of specular points from a satellite signal (since many points in the ground can reflect a satellite signal in the direction of the receiver simultaneously specially in mountains and similar topologies).

The solution of the previous equation is, however, not feasible because of the unlimited number of specular points that we have to evaluate to construct the set of reflection points from one satellite at each time step. The problem can be simplified by assuming that only one specular point exists from each satellite signal. This is possible if we assume that the terrain is flat in the region of interest for example in the ocean or a lake. In this case, we can solve (3.1) to find the specular point given the known location of the GNSS–R sensor, and the known incident angle, i.e:

$$\frac{x - x_R}{l} = \frac{y - y_R}{m} = \frac{z - z_R}{n} \tag{3.2}$$

The value of $(x_P, y_P, z_P)$ can be estimated from the intersection between the terrain topology and (3.2).

The approach in the literature to solve (3.2) starts from estimating the specular point assuming the reflection happened at the Earth ellipsoid, shown with the red solid line in Figure 3.2 , which is a geometric model that best fits the entire Earth surface. Therefore, we will be first estimating point E in Figure 3.2. Then this estimation is refined with the knowledge of the geoid, the purple solid line in Figure 3.2, which is the equipotential surface best fitting the Earth gravity field, to estimate G in Figure 3.2 [38, 62]. The difference between the geoid and the ellipsoid, indicated by the purple dashed line in Figure 3.2, is known as the geoid height $N$. The position of the specular point is further refined with knowledge of the topology of the terrain to estimate point P in Figure 3.2. This refinement is essential for inland reflectometry studies [63] and other applications where accuracy of the specular point estimation is needed [64]. The difference between the ellipsoid and the specular point on Earth's surface, shown in red dashed lines, is known as the ellipsoid height. Finally, the vertical distance between the geoid and the specular point, shown in dashed blue lines in Figure 3.2, is known as the orthometric height.

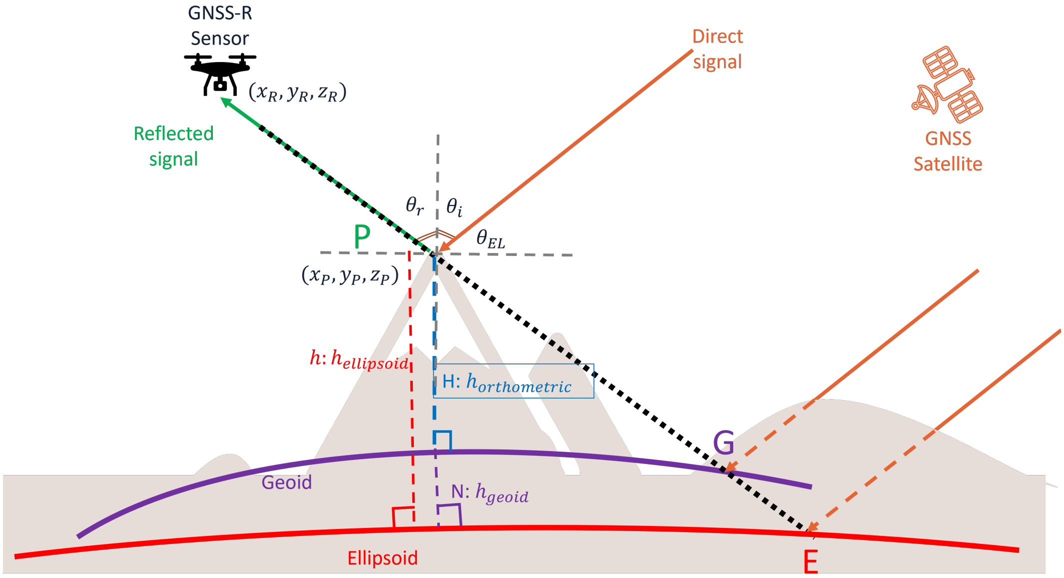


Fig. 3.2 Estimating the specular point position

In summary, with the help of orthometric terrain heights, Earth geoid and ellipsoid, the location of the specular points can be obtained from the satellite and receiver locations given by GNSS measurements.

## 3.3 Flight Height Estimation

Figure 3.3 illustrates the concept of height estimation from a geometry point of view. Here, we can measure the delay of the reflected signal with respect to the direct signal, and thus we can measure the extra distance travelled by the reflected signal. This extra distance is a function of the satellite elevation (known) and the height of the surface (to be estimated) as will be shown in this Section.

In Figure 3.3, since the GNSS satellites (the transmitters) are very far (20,000km) compared to an airborne GNSS-R receiver (10km for an aeroplane, less that 1km for a drone), we can assume that the signal ray reaching the direct antenna is parallel to the signal ray incident on the ground surface. Thus, the wave front reaching the up–looking antenna at time $t_0$ reaches the specular point at $t_S$:

$$\begin{cases} t_S = t_0, & \theta_{EL} = 45^\circ \\ t_S > t_0, & \theta_{EL} > 45^\circ \\ t_S < t_0, & \theta_{EL} < 45^\circ \end{cases} \tag{3.3}$$

Since we are processing GNSS signals, the delay $\delta$ between the direct and reflected signals, $t_0$ and $t_1$ respectively, is measurable as:

$$\delta = t_1 - t_0 = (t_1 - t_s) - (t_0 - t_s)[s] \tag{3.4}$$

From Figure 3.3, (3.4) can be rewritten in [m] as:

$$\delta = c\,(t_1 - t_s) - c\,(t_0 - t_s) = y - x[m] \tag{3.5}$$

where $c$ is the speed of light. $x$ is the distance between the specular point and the incident wavefront at $t = t_0$. $y$ is the distance between the specular point and the GNSS–R sensor. The distances $x$ and $y$ are shown in Figure 3.3. From the geometry illustrated in Figure 3.3, we can estimate the values of x and y:

$$x = y \sin\left(\frac{\pi}{2} - 2\theta_{EL}\right) \tag{3.6}$$

$$y = \frac{h}{\sin\theta_{EL}} \tag{3.7}$$

where h is the vertical height of he GNSS–R sensor above the specular point and $\theta_{EL}$ is the elevation of the satellite as depicted in Figure 3.3. Substituting (3.6) and (3.7) in (3.5), the height of the flight can be estimated from the delay between the direct and reflected signals as:

$$\delta = y - x = \frac{h}{\sin\theta_{EL}}\left(1 - \sin\left(\frac{\pi}{2} - 2\theta_{EL}\right)\right)$$

$$h = \frac{\delta \sin\theta_{EL}}{1 - \sin\left(\frac{\pi}{2} - 2\theta_{EL}\right)} \tag{3.8}$$

where $c$ is the speed of light, $h$ is the height of the GNSS-R sensor above the reflecting surface and $\theta_{EL}$ is the elevation of the satellite as depicted in Figure 3.3.

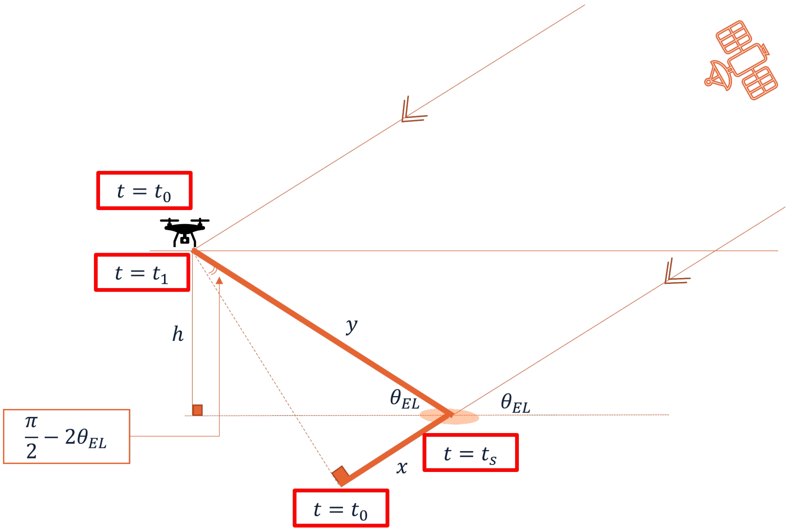


Fig. 3.3 Estimating height from GNSS–R measurements

In summary, by measuring the delay $\delta$ of the reflected signal with respect to the direct signal, we can estimate the height of the flight above the reflecting surface.

## 3.4 Reflected Power Estimation

If we observe the power of the signal broadcast by a satellite as measured by the up–looking antenna, and the same satellite signal as measured by the down–looking antenna, we can estimate the amount of power reflected by the ground surface as follows.

Figure 3.4 illustrates the concept of estimating the reflected power using GNSS–R. Starting from (2.5), the power of the signal incident on the up–looking antenna ($S_D$) and the power of the signal incident on the reflecting surface ($S_i$) are:

$$S_D = P_{TX} L_{LOS,D} L_{atm,D}, \tag{3.9}$$

$$S_i = P_{TX} L_{LOS,i} L_{atm,i}, \tag{3.10}$$

where $P_{TX}$ is the nominal transmitted power from a satellite. $L_{atm,D}$ and $L_{atm,i}$ are the power loss due to atmospheric propagation of the direcct and incident signals. For low altitude flights, since the two signals cross almost the same portion of the atmosphere, they get attenuated by the same atmospheric environment, and thus the atmospheric loss is equal for both signals. $L_{LOS,D}$ and $L_{LOS,i}$ are the free space loss (the signal attenuation due to the distance travelled by the signal) across the line of sight given by:

$$L_{LOS} = \left(\frac{4\pi d}{\lambda}\right)^2, \tag{3.11}$$

where d is the distance from the satellite to the receiver/surface and $\lambda$ is the signal wavelength. For low altitude flights, for example, if the direct GNSS signal travelled $d_D = 20,000$km, and the incident signal travelled an extra distance of 2km (i.e $d_i = d_D + 0.0001d_D$), and substituting in (3.11):

$$L_{LOS,i} = \left(\frac{4\pi}{\lambda}\right)^2 \times (1.0001d_D)^2 \approx L_{LOS,D} \tag{3.12}$$

Therefore, combining (3.9) and (3.10):

$$S_i \approx S_D \tag{3.13}$$

i.e. the signal incident on the surface has the same power of the signal incident on the up–looking antenna.

Similarly for the power received by the down–looking antenna $S_R$:

$$S_R = \alpha S_i L_{LOS,R} L_{atm,R} \tag{3.14}$$

where $\alpha$ is the fraction of $s_i$ that was reflected by the surface, $L_{atm,R}$ is the power loss due to the atmosphere between the reflecting surface and the sensor. Since the sensor is flying at low altitude, i.e. in the troposphere, we can assume $L_{atm,R}$ to be negligible for such a small distance in this environment. For the free space loss from the specular point to the down–looking antenna $L_{LOS,R}$, assuming a distance of $d_R = 2$km travelled by the signal from the specular point to the down–looking antenna, the free space loss is:

$$L_{LOS,R} = \left(\frac{4\pi}{\lambda}\right)^2 \times (0.0001d_D)^2 = L_{LOS,D} \times 10^{-8} \tag{3.15}$$

i.e. for low altitude flights, the free space loss due to this final signal path is negligible compared to the loss that has already occurred due to the signal travelling from the

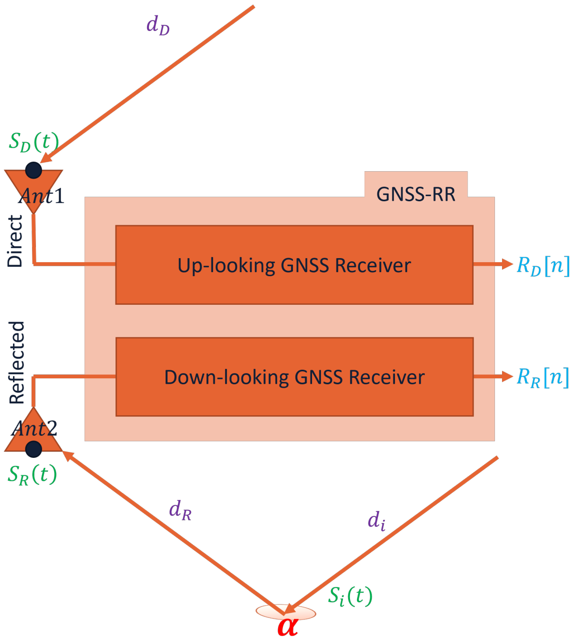


Fig. 3.4 Estimating the reflected power from GNSS–R measurements

satellite to the up–looking antenna/reflector. Therefore, (3.14) becomes:

$$S_R = \alpha S_i \approx \alpha S_D \tag{3.16}$$

In summary, when the distance from the sensor to the GNSS satellites is much larger than the distance from the sensor to the reflecting surfaces, we can assume that power of the direct and incident signals are equal. Therefore, to estimate the reflectivity of a surface, we only need to measure the direct and reflected signal powers ($S_D$ and $S_R$) incident on the up–looking and down–looking antennas respectively.

In the next Section we show how these two measurements are estimated and calibrated for a GNSS–R sensor.

## 3.5 The GNSS-R Receiver

In a GNSS-R receiver, two independent GNSS receivers are implemented to receive the direct and reflected signals. The reason is that the reflecting surface scatters and absorbs the incident signal resulting in the reflected signal becoming much weaker than the direct signal. For this reason the reflected signal receiver is in particular carefully designed to handle weak GNSS signals. Figure 3.5 shows the detailed block diagram of the GNSS–R receiver. Here, Ant1, LNA1 and ADC1 are the antenna, Low-Noise Amplifier (LNA) and Analogue to Digital Converter (ADC), respectively, for the direct signal channel. Ant2, LNA2 and ADC2 are the antenna, LNA and ADC, respectively, for the reflected signal channel. The ADC blocks usually incorporate an Automatic Gain Control (AGC) unit which is responsible for adjusting the gain of the ADC depending on the signal power. Looking at the equations of the direct and reflected signals we can understand the effect of these hardware differences on estimating the reflected power. The direct and reflected signal powers ($R_D[n]$ and $R_R[n]$) at the output of the ADCs are:

$$R_D[n] = G_{Ant1}G_{LNA1}G_{ADC1}S_D[n-\tau_1] = \ AS_D[n-\tau_1] \tag{3.17}$$

$$R_R[n] = G_{Ant2}G_{LNA2}G_{ADC2}S_R[n-\tau_2] = \alpha BS_D[n-\tau_2-\tau] \tag{3.18}$$

where $G$ is the gain for the various components (Ant, LNA and ADC), $\tau$ is the delay of the reflected signal w.r.t the direct signal, and $\tau_1$ and $\tau_2$ are the delays introduced by the direct and reflected receiver channels, i.e. it is the delay at the output of the receiver w.r.t. the incident signal on the antenna. $A$ and $B$ are terms that incorporate all the gains on the direct and reflected channels respectively. Finally, $\alpha$ is the fraction of power reflected by the surface.

In (3.17) and (3.18), $\alpha$ and $\tau$ are the two quantities we want to estimate, while $R_D[n]$ and $R_R[n]$ are the two quantities we are able to measure at the output of the ADC. The values of $\frac{A}{B}$ and $\tau_1 - \tau_2$ are obtainable in calibration phase.

The calibration of a GNSS-R sensor generally consists of two phases:

- Lab calibration where the above mentioned values ($\frac{A}{B}$ and $\tau_1 - \tau_2$) are calibrated.
- Field calibration where $\alpha$ and $\tau$ values estimated by the sensor are validated against true measurements.

The general idea of the in-lab calibration is to provide the same signal to both the direct and reflected channels as shown in Figure 3.6. This means the delay between the direct and reflected signals $\tau$ is 0 and the direct and reflected powers are the same, i.e. $\alpha$ is 1. However, since we need to calibrate the whole sensor including the antennas, we will need to broadcast the GNSS signals in a controlled environment

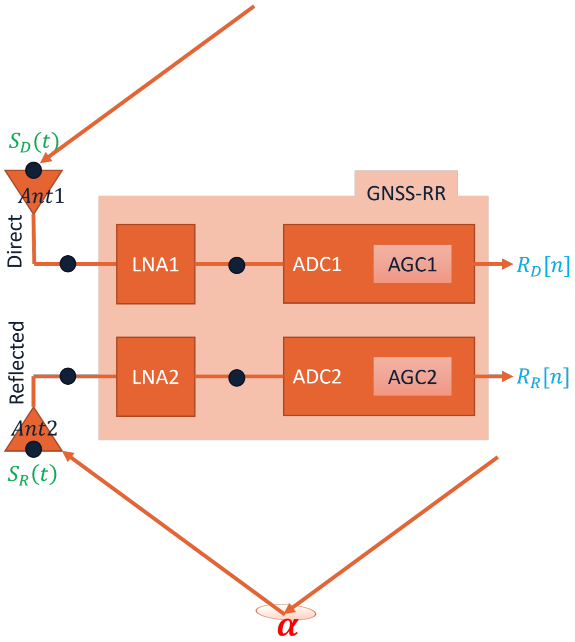


Fig. 3.5 Block diagram of the GNSS-R Receiver hardware

and , most importantly, we need to obtain the LHCP version of the GNSS signal and feed it to the down-looking antenna.

On the other hand, for the field calibration, we need to know the exact height of the reflecting objects (to estimate the true delay) and we also need in-situ measurements for the reflectivity of the surfaces.

In this work (1) we show results of the field calibration, and (2) we focus on calibrating the delay. The motivation behind those two choices are as follows. For application in flood detection, it is beneficial to estimate the height of the water and compare it with the normal water level. At the same time, the true height of the reflecting surfaces can be obtained from Digital Elevation Models, which are usually available for post flood events assessment. Furthermore, since our objective is to detect water existence and its extent, we do not dive into estimating the amount of reflected power $\alpha$ and thus we will not show results of calibrating $\frac{A}{B}$. The latter allows us also to activate the AGC in the front ends, to maximize the received power, without needing to calibrate the effect of the changing gain of the front ends. Finally, since we are proposing simple sensors to install in small UAV platforms, we deliberately simplify the pre-processing and post processing work to be done and we prove that it is still feasible to obtain the target results.

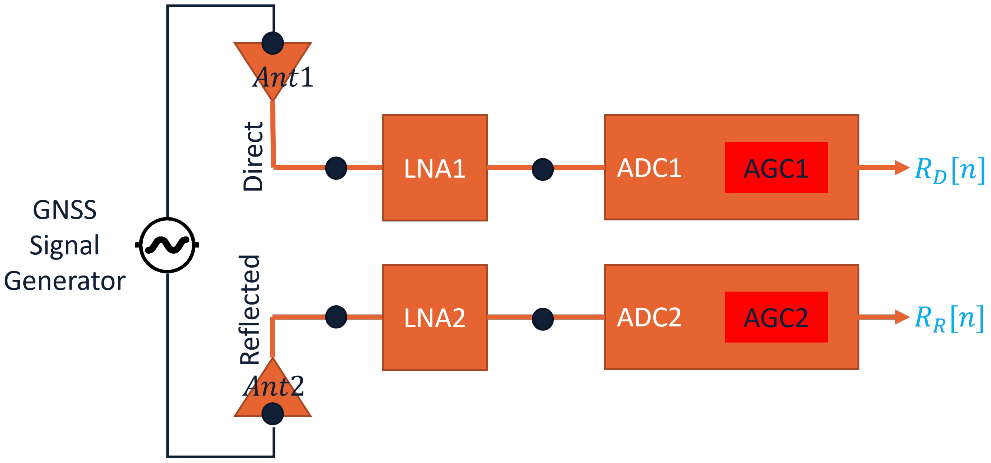


Fig. 3.6 GNSS-R Sensor calibration.

## 3.6 GNSS-R Application in Water Detection

Remote sensing has been used in studying floods for more than 40 years [65], where data from air- and space-borne optical, thermal and microwave sensors [66] are exploited to support the different stages of flood risk management [67]. Each of the sensors categories has advantages and disadvantages, and thus the selection of the most appropriate type of devices and data for the monitoring of the specific events can be not easy.

Space-borne data are widely used for flood monitoring providing different spatial and temporal resolutions. Satellite optical images, although the most straightforward to interpret, are usually not available during floods because of the cloud cover during these events [68]. Synthetic Aperture Radar (SAR) techniques on the other hand overcome this weather limitation, and also provide day/night visibility, [69] but are challenged to detect inundations in urban areas [70] and under vegetation [71]. The latter is overcome with passive microwave sensors in low microwave frequency which have better penetration through dense vegetation [72]. GNSS-R falls in the latter category of passive low frequency radars, and had proven successful in detecting wetlands during floods (e.g [73]). In general, space-borne remote sensing data for flood monitoring have the disadvantage of long revisit time (>2.5 days) [74] and thus they do not provide enough repeating visits needed during these events, and in worst cases they even miss the peak of the event [75] like in flash flooding (for example [67]). Also, the satellites with short revisit times have low resolution (in kilometers), which make them not sensitive enough to detect the details needed for floods [76]. Moreover, even the meter level space-borne resolution sensors are challenged in delineating inundation in headwater regions which require sub-meter image resolution [71].

From a general perspective, aircraft flights are able to overcome the weather limitations of optical sensors and the resolution and revisit limitation of space-borne remote sensing, but they are expensive and sometimes, due to the need of runways and the difficult topology of the areas, they are not feasible [77]. Moreover, they are not able to provide the temporal resolution provided today by space-borne sensors that scan the Earth continuously. In fact, today hybrid data from different space- and air-borne sensors are fused together to complement each other and provide the different aspects of global flood monitoring, local detailed inundation maps, temporal data availability and quick data availability of floods [78, 79]. For comprehensive literature on floods remote sensing using the various platforms and sensors refer to [67, 65, 69, 66, 71, 74].

The processing of GNSS signals that are forward-scattered by the ground allows for using GNSS-R in monitoring the parameters of the Earth's surface. It is a passive radar where all the GNSS satellites are the transmitters, and any sensor capable of processing their reflected GNSS signals is the receiver [3].

GNSS-R advantages include, beside penetration of vegetation cover and resilience to clouds and smoke, the frequent global coverage of the transmitters from multiple constellations which avoids the limitation of current active remote sensing techniques that rely on a single transmitting satellite [80]. Also, the low cost of the receivers and the small size of the antennas regardless of the targeted resolution, makes GNSS-R an interesting sensing technique.

A wide range of applications has been reported in the literature for GNSS-R including: water basins detection (e.g., [81, 82]), river level monitoring (e.g., [83]), estimation of the surface roughness and wind retrieval (e.g., [84, 85]), measurement of the soil moisture (e.g., [86, 87, 63, 88, 89]), altimetry (e.g., [90, 91]), monitoring the presence of vegetation (e.g., [92, 93]), and estimation of snow/ice thickness (e.g., [94]).

GNSS-R sensors have been deployed onboard spaceborne platforms, for example the GPS reflectometry experiment onboard the UK-Disaster Monitoring Mission (UK-DMC) [95], the Cyclone Global Navigation Satellite System (CYGNSS) [96], the International Space Station (ISS) [97] and the UK TechDemoSat-1 mission [98].

They proved to be a valuable source of remote sensing data for both land [99–102] and ocean [98] remote sensing. However, space-borne GNSS-R have few limitations. As of today they provide 0.5–1.0km resolution which is not satisfactory for detailed inundation maps. Also, due to the nature of the flight trajectories of satellites, the measurements they produce are in form of surface tracks rather than instantaneous images [80]. These two limitations can be overcome by air-borne GNSS-R, and they are addressed in this work by flying a GNSS-R onboard a UAV.

The diffusion of Unmanned Aerial Vehicles today is a result of the great positive impact it has proven in modern societies [103]. UAVs are offering solutions for a broad range of applications including support after natural disasters operations [104] such as floods (e.g., [105]), monitoring of environmental pollution, security and surveillance, and critical infrastructures diagnostic. UAVs can be equipped with several types of remote sensing sensors like hyperspectral and optical cameras (e.g., [106]), Synthetic Aperture Radar (SAR) (e.g., [107–109]), and GNSS-R sensors.

GNSS-based passive radars on-board commercial UAVs introduce constraints on the size, weight and power consumption of the GNSS-R sensor, especially if a small commercial quad- or hex- copter carries out the experiment. Many designs for airborne GNSS-R sensors have been presented in the literature. For example the sensor presented by Troglia Gamba et al. [30] implements a SDR based GNSS-R on an ODROID–X2 microprocessor able of receiving both the LHCP and the RHCP reflected signals using two front–ends streams. Esterhuizen and Akos [110] also presented a miniaturized receiver based on two GPS L1 front ends and a Nano-ITX Single Board Computer (SBC) to store raw signal samples, which were analyzed in post-processing [111]. Marchan-Hernandez et al. [112] on the other hand, designed an FPGA-based GNSS–R that is capable of computing the Delay Doppler Maps (DDMs), with update rate of 1 ms.

The majority of the GNSS-based passive radar sensors in the literature are composed of (1) a zenith pointing RHCP antenna for the reception of the direct GNSS signals, and (2) a nadir-pointing antenna to receive the ground reflected GNSS signals. The latter can be LHCP, assuming the reflected signal underwent a complete polarization flip, or dual polarized.

The work presented in this thesis was done under the I-REACT project [113]. This project investigates solutions that can help improving the response of decision makers and also rescuers to extreme events. The overall objective of this study is to investigate the feasibility of using data collected by UAV-based GNSS-R sensors to support flood monitoring operations. In particular, we show the ability to detect the presence of various water bodies on ground using a custom made UAV-based GNSS-R sensor. We focus on:

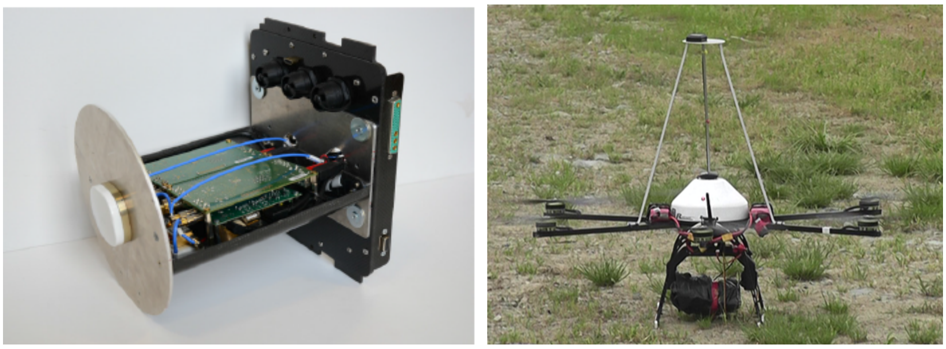
1. Investigate the possibility to equip small UAVs with GNSS-based passive radar capabilities, to be used for water detection in post-mission assessments.
2. Estimate the performance of the GNSS-based passive radar developed by Troglia Gamba et. al. [30] in monitoring water surfaces on ground, when it is mounted on board a small UAV. This assessment could set-up a further source of geospatial data for the system developed in the I-REACT project [113].

# Chapter 4

# Feasibility of UAV-Based GNSS-R for Water Detection as a Support to Flood Monitoring Operations

## 4.1 UAV-Based Data Collections and Processing

To investigate the feasibility of detecting water surfaces using GNSS-based passive radar carried on small UAVs, we implemented a GNSS-R sensor complying with the strict requirements of a UAV payload in terms of weight, size and power consumption. In post-processing, we targeted water surfaces detection, and for that we optimized all the parameters needed in order to increase the reliably of water detection from GNSS signals. GNSS signals are transmitted with RHCP, and the carrier frequency is in the order of 1 GHz ( 1.150–1.620 GHz). Their polarization is reversed when a smooth surface reflects them becoming LHCP. If the reflecting surface is non-specular, the reflected signal will be a mixture of LHCP and RHCP signals. Thus, it is necessary to use antennas capable of discriminating between the two kinds of polarization, and selecting the desired one. The synchronous recording of both the direct and the reflected GNSS signals acts as an accurate geo-referencing mechanism for the GNSS-R measurements, which is important for post mission assessments. Figure 4.1 shows the sensor (left) and a photo from one of the data collection campaigns showing the sensor mounted on-board a UAV (right). In the following subsections, we will describe the sensor used for data collection, then we will



(a) The sensor (b) The sensor mounted onboard a UAV

Fig. 4.1 The GNSS–R sensor used in this study, and a picture from one of the data collection campaigns.

explain the data collection campaigns, after that we will detail the signal processing procedure, and finally we will introduce the data sets processed in this work.

### 4.1.1 The Utilized GNSS-R Sensor

In this section, the original version of the GNSS–R sensor used for data collection is described, then the modifications introduced to achieve the low-complexity version are explained. The complete details of the sensor were published in [30], however it is presented briefly here for sake of clarity of the experimental setup and the post-processing methodology.

The original sensor was designed with the aim of detecting water surfaces and land water content. It is intended for small UAVs with weight of 3 kg and memory limit able to store 30 min of data collection. The sensor receives both direct and reflected GNSS signlas using three separate antennas: (1) an up-looking RHCP antenna, (2) a down-looking LHCP antenna and (3) a down-looking RHCP antenna.

The sensor features four synchronized RF channels connecting the direct signal and the LHCP reflected signal to a front end, while the RHCP reflected signal with the direct signal are connected to another front end. Both front-ends down convert the direct signals to baseband (BB), while the reflected signals are down converted to an intermediate frequency (IF) in sake of low noise levels for the already weak reflected signals. An embedded microprocessor controls the flow of the digital samples of all channels.

In this work, in order to reduce the total mass of the payload, a simplified version of the sensor that embeds (1) the RHCP antenna for the direct signal and (2) the LHCP antenna for the reflected ray, was used.

### 4.1.2 Data Collection Campaign

The data collection flights were over different types of water surfaces (rivers, lakes, ponds, etc.) selected to test the sensor in detecting well known water surfaces. Nevertheless, these data were considered valuable by this study to challenge the sensor in detecting flood-similar water presence. The criteria for water detection were defined as being able to detect small water surfaces, distinguish narrow water flows and estimating water surfaces area. With that in mind, three test cases were chosen from the data campaign area:

1. A 0.89 km$^2$ lake with a known basin where we could challenge the ability to estimate the area covered by water and its boundary.
2. A river stream to challenge detecting narrow water streams.
3. Small ponds of water to challenge the detection of small and unexpected water content on ground.

Being not flood events, we were able to overlap the GNSS-R measurements on available orthophotos without the need for field measurements or satellite images taken on the same dates and time of the test flights.

### 4.1.3 Post Processing Methodology

Figure 4.2 shows the block diagram of the post-processing steps. First, we processed the digital samples of the direct RHCP signal using a GNSS software receiver. Here we extracted the UAV trajectory. We also retrieved the list of satellites in view during the data collection campaign, as shown in the upper chain of Figure 4.2. Then, we calculated the lines of specular points on the ground for all the visible satellites for the whole flight duration, utilizing the receiver trajectory extracted in the previous step, and the known satellites positions, as shown in the middle chain of Figure 4.2. We calculated the specular points using the algorithms in [3, 90].

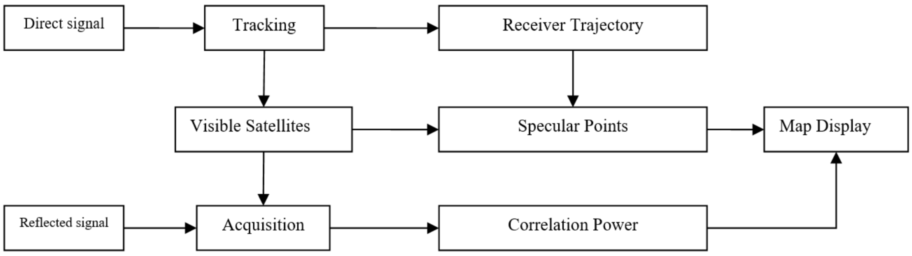


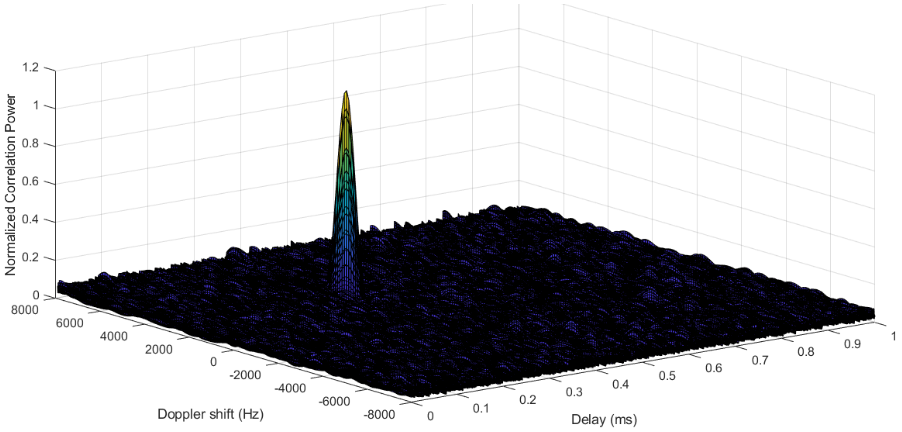
Fig. 4.2 Block diagram of the post-processing steps.

We processed samples of the LHCP signal to measure the power reflected by the surfaces, as shown in the bottom chain of Figure 4.2. We estimated the reflected power for all the visible satellites, by evaluating the cross-ambiguity function (CAF) [114] over a reduced search space [115]. The CAF is generally computed and compared to a threshold to detect the presence of a GNSS satellite signal, and in our case the detection of a GNSS signal reflected by a ground surface. The CAF was evaluated over a set of Doppler-delay values that define the search space where the reflected signal was cross-correlated with a local replica of the code. Figure 4.3a shows an example of the CAF of a visible satellite using the direct signal, and Figure 4.3b shows the CAF for the same satellite at the same time using the reflected signal. It can be noticed that the direct signal as expected has a peak that is well separated from the noise floor, while the CAF associated to the reflected signal shows a weaker peak but it is still visible.

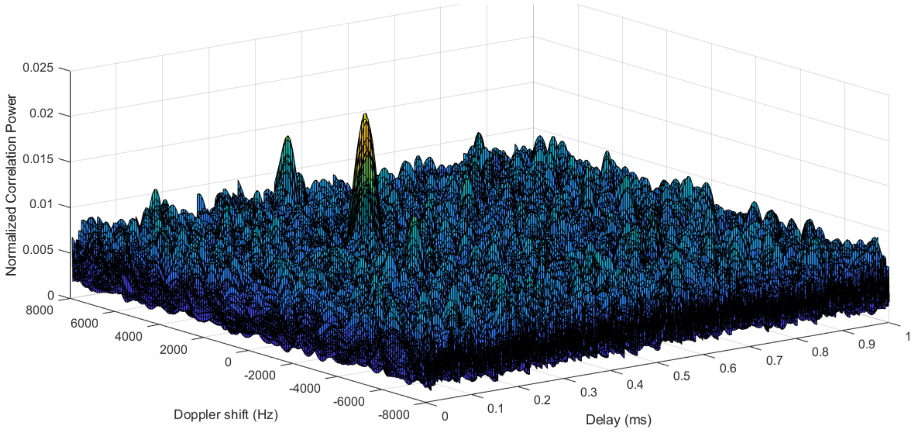
We choose the peak-to-noise-floor separation $\alpha_{mean}$ [116] as a measure of the reflected signal strength:

$$\alpha_{mean} = \frac{R_P}{M_c} \tag{4.1}$$

where $R_P$ is the correlation peak value, and $M_c$ is the mean value of the correlation noise. $M_c$ was calculated from the peak values of the search space obtained cross-correlating the received signal with an orthogonal code not used by the constellation for the whole flight duration. The non-coherent integration time was fixed to 10 ms, and the coherent integration time was 1 ms i.e., we averaged 10 consecutive coherent integrations of 1 ms long. We evaluated the CAF at various rates: 1, 10 and 20 Hz, which corresponded to different levels of resolution along the lines of specular points. In this work we are interested in understanding if this peak corresponds to reflection



(a) Using the direct signal



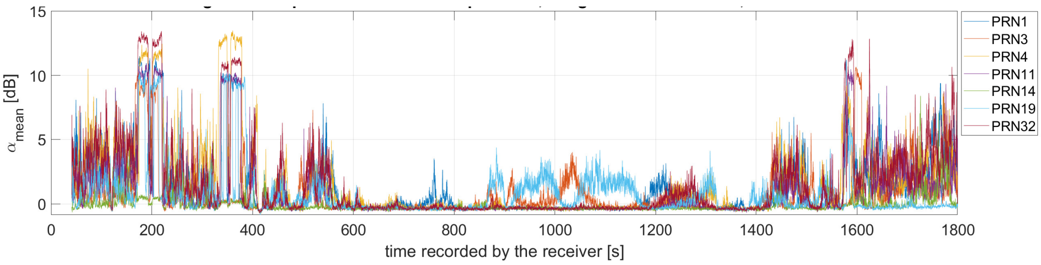
(b) Using the reflected signal

Fig. 4.3 Examples of the search space for a visible satellite when (**a**) evaluated from the direct signal, and (**b**) evaluated from the reflected signal.

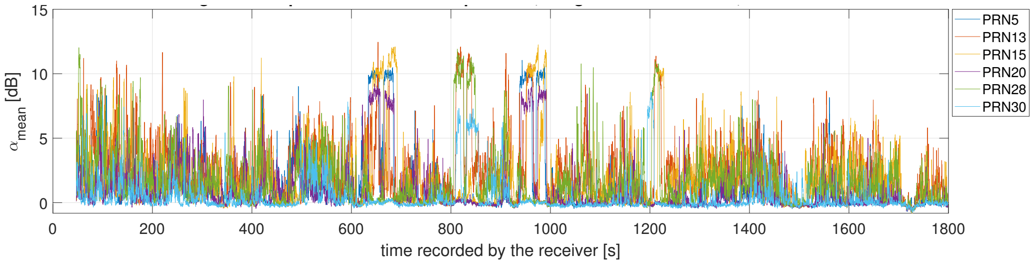
from a water surface. As will be shown later, when the signal is reflected from water, $\alpha_{mean}$ will generally be above the threshold we used to discriminate the presence of reflections.

### 4.1.4 The Data Sets

In this article, we processed two data sets collected during different flights, with different passes over the test area in different days and different seasons. The first flight was in December, 2013 and the second flight was in the following May. We processed both flight data using the same methodology, comparing the results for



(a) data set (1)



(b) data set (2)

Fig. 4.4 Peak-to-noise-floor separation $\alpha_{mean}$ of the reflected signal for a subset of visible GPS satellites during (**a**) flight (1) December, 2013 and (**b**) flight (2) May, 2014

what concerns water detection on ground. In this section, we describe the data sets used.

Figure 4.4 shows the peak-to-noise-floor separation $\alpha_{mean}$ of data set (1) and data set (2), in dB, for a subset of the visible pseudo-random noise (PRN) codes. We computed $\alpha_{mean}$ at a rate of 20 Hz and we used a first order low pass digital Butterworth filter with a cutoff frequency of 1 Hz to smooth the measurements. The x-axis reports the time, in seconds, from the beginning of the data set. It can be noticed that $\alpha_{mean}$ has a time-variant trend. Focusing on data set (1), some epochs are characterized by $\alpha_{mean}$ values greater than 5 dB (e.g., 0–400 s), indicating potential signal reflections. Other epochs, such as those associated to PRN 32 in 600–820 s, have small values that hardly reach 2 dB . Variations can be observed among satellites, as noticeable for example between 900 s and 1200 s, where $\alpha_{mean}$ values related to PRN 32 are lower than the values associated to PRN 19 and 3. The same can be observed on the plot referencing to data set (2) which, however, had higher reflection values through out the whole flight duration.

In the next section, we reference each of these reflections to their ground reflecting points, with focus on reflections generated from water.

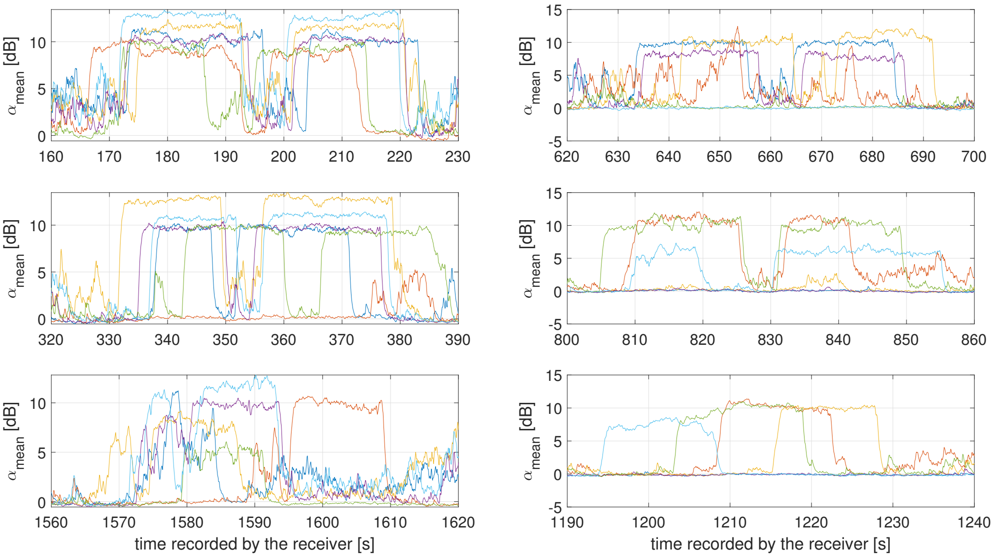
## 4.2 Results and Discussion

In this section we are going to present three case studies, that we have chosen to emphasise particular features and highlight the potential of UAV based GNSS-R for monitoring surface water. The first case describes detecting the Avigliana lakes which are two adjacent lakes in Northern Italy. We took them as reference for calibrating both the GNSS–R sensor and the post–processing algorithms. The other two cases relate to smaller and narrower water surfaces, where the performance of the technique is challenged in detecting unexpected water contents on ground and in recognizing narrow river streams.

### 4.2.1 Case-Study I: Lakes

The UAV flew over the Avigliana lakes, which are wide water bodies convenient for validating the sensor and the post–processing algorithms. The approximate dimensions of the largest lake is 970 m on the widest east–west direction, and 1200 m on the north–south direction. The sensor passed over the lakes three times: north–south, south–north, and west–east. The average height of the UAV was 450 m over the lake surface with average speed of 50 ms$^{-1}$.

In data set (1) that originated Figure 4.4a, these passes correspond to epochs 170–230 s, 330–390 s and 1560–1620 s. Reflections were observed, as expected, during these epochs, as shown in Figure 4.5a, which is a zoom view into the epochs corresponding to the passes over the lakes. The figure shows that for some PRNs the values of $\alpha_{mean}$ increase up to roughly 10 dB and remain constant for many seconds. This trend can be noticed for the satellites that have specular points falling on the lakes and it confirms detecting GNSS signals reflected from the lakes. The periods of these reflections depend on the lake width, the sensor speed and the satellite-sensor geometry. Indeed, the value of $\alpha_{mean}$ stays high as long as the specular points are on water. It can be noticed also that the reflection value corresponding to water detection differs by some dBs for the different satellites. This is due to the different received power from the different satellites on the first place, which when reflected correspond to different surface properties. The same reflections can be noticed looking at data set (2) where the flight, again, passed over the lakes three times: north–south, south–north and west–east. The corresponding passes are 600–700



(**a**) Data set (1) (**b**) Data set (2)

Fig. 4.5 Comparison of the peak-to-noise-floor separation of the reflected signal for (**a**) data set (1), and (**b**) data set (2), when the sensor crossed the lakes on the different passes: north to south (**top**), south to north (**middle**), and west to east (**bottom**).

s, 800–860 s and 1190–1240 s as reported in Figure 4.4b and zoom-viewed in Figure 4.5b.

This concept can be appreciated in Figure 4.6a,b, which show the specular points lines of all the PRNs in view, superimposed into orthophotomaps. These maps correspond to data set (1) and data set (2) respectively. The flight trajectory is represented by the black dots and it is displayed at a rate of 1Hz. The colored lines are the specular points of the satellites which showed enhanced values of $\alpha_{mean}$. The colors of the specular points reflect on the values of $\alpha_{mean}$ associated to them. From Figure 4.6, it can be noticed that the boundary between land and water is clearly distinguishable and it correlates with the orthophotomap. Concentrating on the area between the two lakes, the steep increments and decays in the values of the peak-to-noise-floor separation along time, that were commented in Figure 4.5, actually correspond to the water–land boundaries. The resolution of the water edge detection is at the order of tens of meters, as expected and reported in similar investigations available in the scientific literature. This demonstrates the suitability of the GNSS-R onbaord UAVs in detecting water at a resolution space-borne GNSS-R are not able to provide at the moment. Note that we processed these measurements at 20 Hz rate, with an integration time of 10 ms only. The measurements rate could be set up to 100 Hz to provide an even higher resolution.

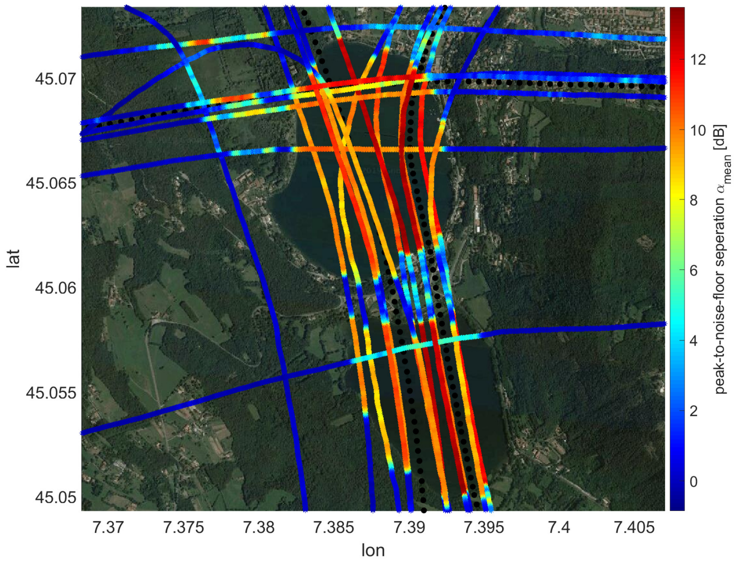
Now that we know which reflections were from the lake, we want to have an insight on the reflected power as a function of the satellite elevation. The reflected power depends on many factors: the transmitted power by the satellite, the satellite elevation, the receiving antenna pattern, the properties of the reflecting surface, among others. Figure 4.7 reports the elevation angle for the reflections observed over the lake. Each point is averaged from five consecutive seconds during which the specular points were falling on the lakes. We used samples from both data sets (1) and (2) to create Figure 4.7. We also reported the PRN associated to each of these data points. It can be noticed that the reflected power from water roughly increases linearly with the incident angle (the satellite elevation). It can also be noticed that although we did not set a cut-off for the elevation angle, the useful measurements i.e., the measurements from water, were for angles greater than 30 degrees in our data collection. Measurements from satellites with less then 30° simply did not fall on the lakes during this data collection, and thus did not show significant reflections.

The previous results and discussion prove the possibility of detecting the presence of water using our low cost, software radio GNSS-R sensor. They also prove that the edges of water are well detected. In the following sub-sections, we are investigating the collected data to extract information about the area of the detected water surfaces. Then, we are analyzing the benefit of multi–GNSS for GNSS–R, by investigating the improvement in water surface area estimation when multi–GNSS are considered.

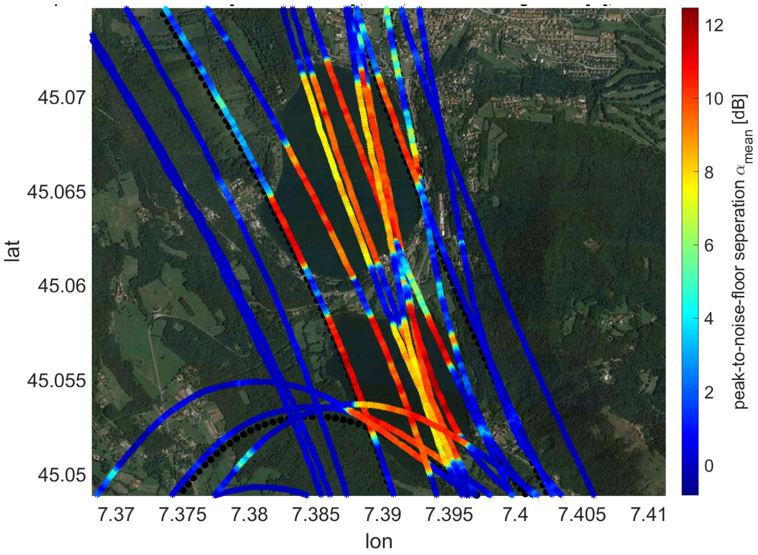
### Water Surface Area Estimation and Benefit of multi-GNSS

Figure 4.8 reports the lines of specular points for which the value of $\alpha_{mean}$ exceeded 6 dB. Here, we superimposed the specular points into the orthophotomap of the upper lake and we linked the extreme points in order to form a polygon. This roughly creates an estimation of the water surface using GNSS-R. This particular encircled area is 0.822 km$^2$. It is marginally smaller than the true surface of the lake which is approximately 0.89 km$^2$ as reported in [117].

This estimate of the surface water area is about 7.5% less than the true area. But, since we are processing only the reflected GPS signals, the accuracy of such an estimate could be enhanced by taking advantage of multi–GNSS signals.



(a) data set (1)



(b) data set (2)

Fig. 4.6 GNSS signals specular reflection points superimposed into an orthophotomap. The colorgrades indicate the measured reflected power $\alpha_{mean}$.

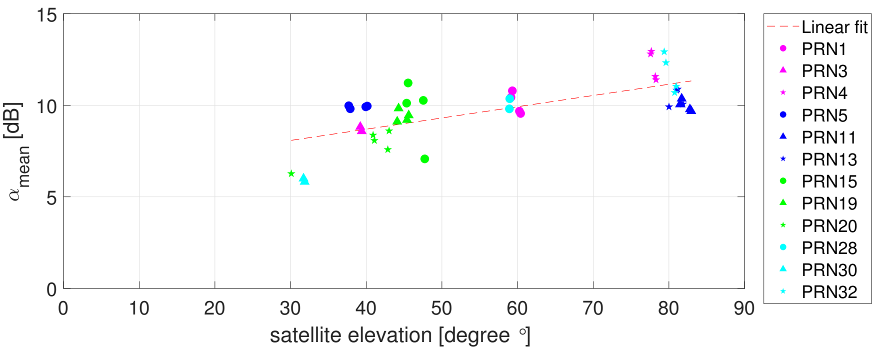


Fig. 4.7 The reflected power from the lake $\alpha_{mean}$ VS satellite elevation.

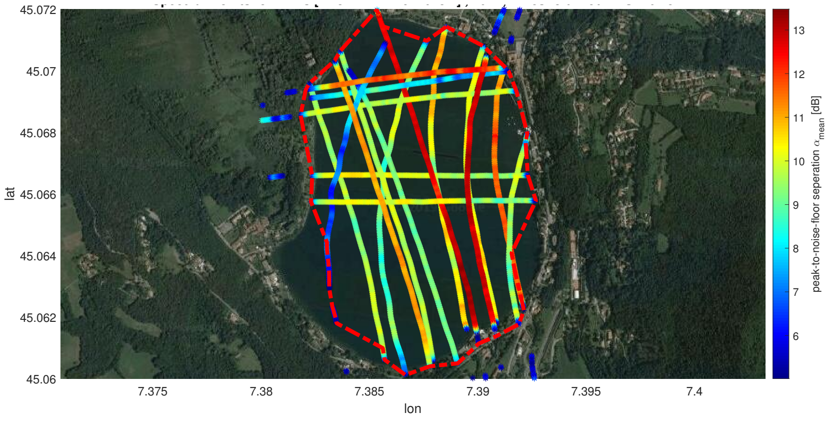


Fig. 4.8 Estimation of the Northern Avigliana lake surface area using GNSS-R. The red dotted lines represent the estimated boundary of the lake from the GNSS signals reflected from the water surface.

In order to investigate the improvement introduced by Galileo, GLONASS and BeiDou signals, we simulated a realistic flight trajectory and evaluated the increase in the number of specular points on the Northern lake. Figure 4.9a reports the sensor trajectory we choose for the simulation. It passes the Northern lake roughly in the middle, both on the north–south and west–east directions. We calculated the lines of specular points for the satellites belonging to the various constellations. We evaluated this trajectory at 100 and 200 m heights above the Earth surface, which are typical values for small UAV flights. The reason for simulating two heights is to ensure that the simulated height of the flight is not biasing the results we obtain, due to the fact that we are measuring a specific lake area; since the flight height should be proportional to the size of the targeted water body.

Figure 4.9b–e show samples of the results, with clear enhancement in the number of specular points lines when constellations are added. Here we report the results of the 200 m flight height above the Northern lake only. The advantages of considering multi–GNSS–R were also quantified in terms of the accuracy of the estimation of the surface area of the Northern lake, evaluated as the ratio between the estimated area and the true area. Figure 4.9f reports the estimation accuracy versus the number of visible satellites, which improves when adding more GNSS signals. It can be seen that the accuracy of the estimation using GPS only remains at about 80% for the two heights considered for the UAV, but increases to 90% when including Galileo. It further improves when GLONASS is added, and reaches 98% when Beidou is added. Notice that the curves for the different heights follow the same trend which supports our conclusion that the more satellites, the better the area estimation accuracy.

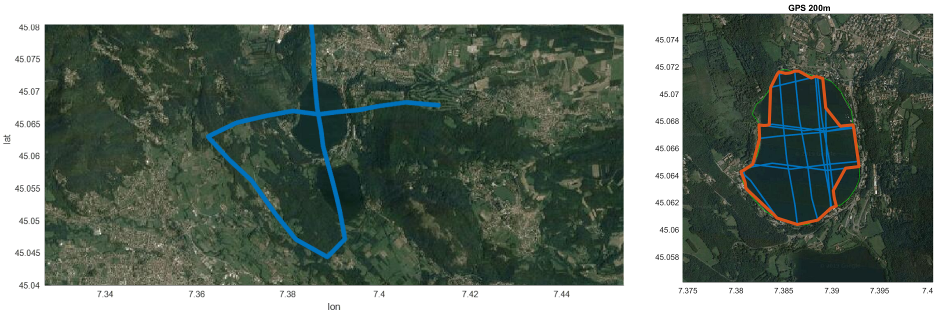
In fact, the high–end multi–GNSS receivers, are already the state-of-the-art for the majority of the civil applications and their use on–board UAVs is growing. Indeed, the results presented in this sub–section refer to a particular case study (and thus they rely on the properties of this water body). Nevertheless, they demonstrate the expected advantages of processing reflected multi–GNSS signals. Processing multi–GNSS signals allows increasing the number of specular reflection points over a certain area. As an example, by, adding Galileo, we expect to double the number of visible satellites compared to what we could have if we were using GPS alone. At mid latitudes, the number of visible satellites can be greater than 40 when GLONASS and Beidou are considered too. This results in increasing the number of polygon vertices and therefore creates a more accurate estimation of the water surfaces.

Undoubtedly, the method presented here is an effective and simple way for estimating surface water. It can be integrated with other sources of data, or even to be used when orthophotos or maps are not available. Surely any reduction or expansion in the water surface can be detected when comparing against historic data. Indeed, utilizing UAV-based GNSS-R sensors, combined with ad hoc processing routines, proves to be an innovative tool for water monitoring. The results shown here were obtained computing the specular points at a rate of 20 Hz, but finer resolutions are attainable with a moderate increase in the complexity.

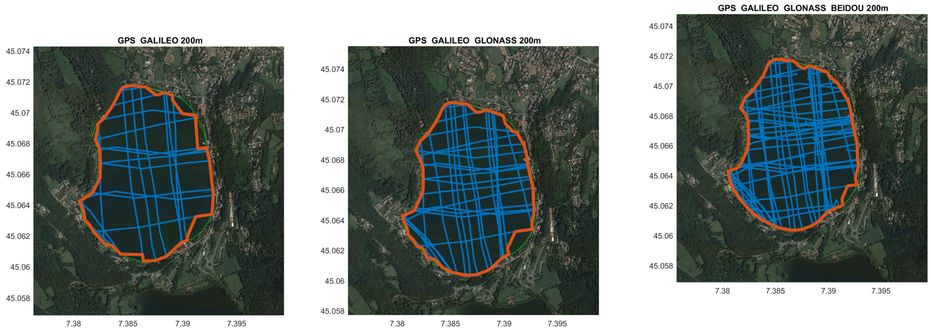
### 4.2.2 Case-Study II: River Stretches

The second case study presents detecting and measuring the Dora river stretches, which have widths that vary between few meters and tens of meters.

The values of $\alpha_{mean}$ for PRN 1 and PRN 32 from data set (1) are reported in Figure 4.10a, for one pass crossing the river course. These PRNs have been selected because they show an evident increment of $\alpha_{mean}$ in the observed time window. Using the same approach followed for the previous case study, we superimpose the lines of specular points to an orthophotomap in Figure 4.10b. Again, the black points show the flight trajectory displayed at a rate of 1 Hz. The colored lines are the specular points displayed with a rate of 20Hz. The values of $\alpha_{mean}$ associated to these specular points are indicated by the colors of the lines. The reflected signals recorded from the river can be appreciated from Figure 4.10b, where the boundaries between land and water correlate well with the orthophotomap, particularly the reflections from the specular points associated to PRN1 (see the left part of Figure 4.10b). The width of the river where the specular points are red is approximately 30 m. However, differences in the reflected power can be noticed by observing $\alpha_{mean}$ values for different satellites. Here, the values of $\alpha_{mean}$ for PRN32 are lower compared the values associated to PRN1. This is reasonably due to the different features of the reflecting surface. While for PRN1 the line of specular points crosses a portion of the river with sharp transitions between land and water, the same ways as in case study 1 (i.e., the red portion in the figure), the line of specular points related to PRN32 falls on a portion of the river following an artificial dam. Here the roughness of the water is different than the other section of the river crossed by PRN1 mentioned above, with low water depth and stones outcropping on the river level.



(a) the sensor trajectory (b) GPS only



(c) GPS and Galileo (d) GPS, Galileo, GLONASS (e) GPS, Galileo, GLONASS, BEIDOU

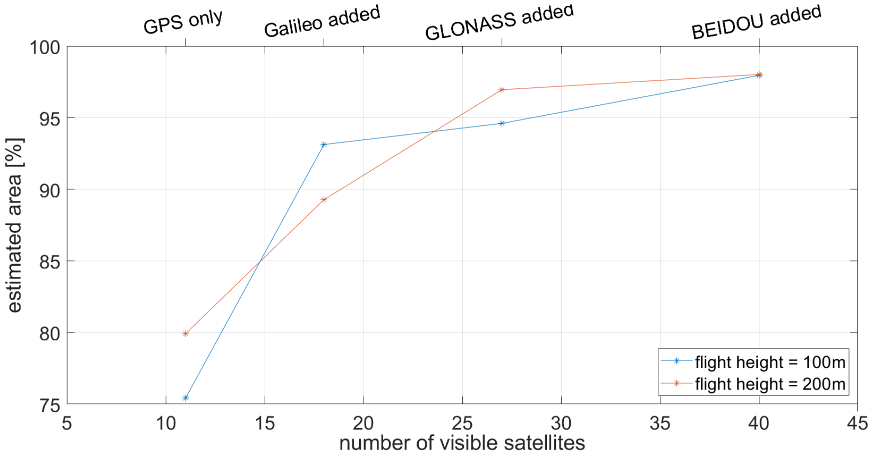
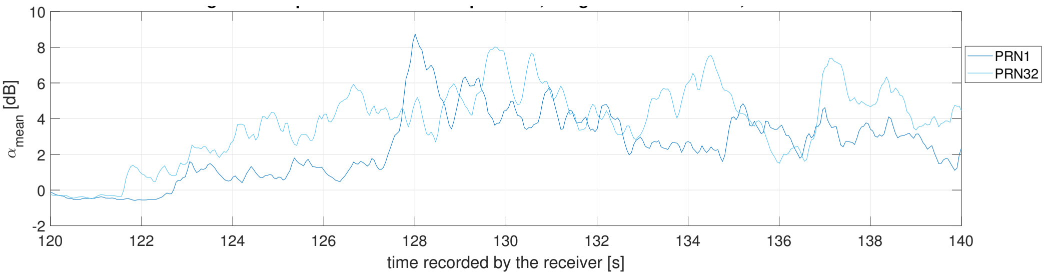
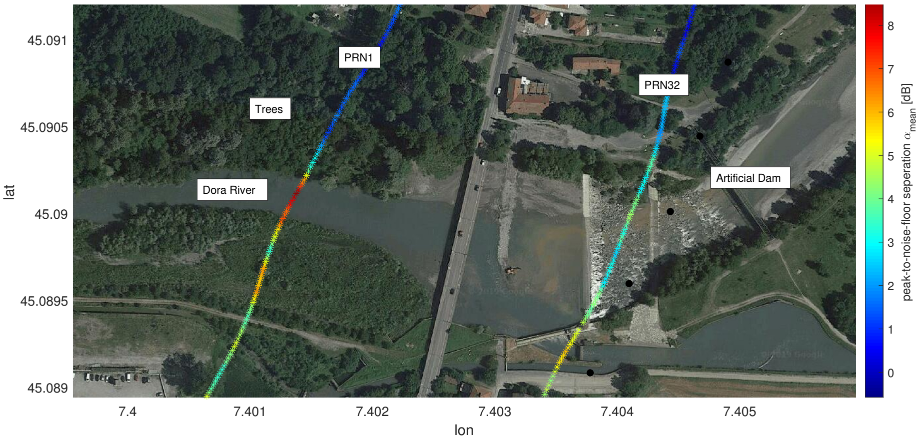


Fig. 4.9 Simulation results of multi-GNSS reflections: (**a**) the sensor trajectory considered, (**b**–**e**) the specular points lines using the different GNSS constellations, (**f**) the ratio between the estimated and the real area of the Northern Avigliana lake versus the number of satellites in view, two different heights of the sensor are considered.

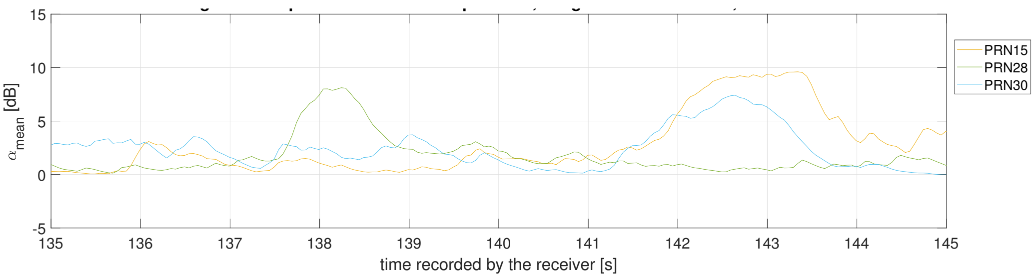


(a) Peak-to-noise-floor-separation $\alpha_{mean}$ of the reflected signals when the specular points were over the river for PRN 1 and PRN 32.

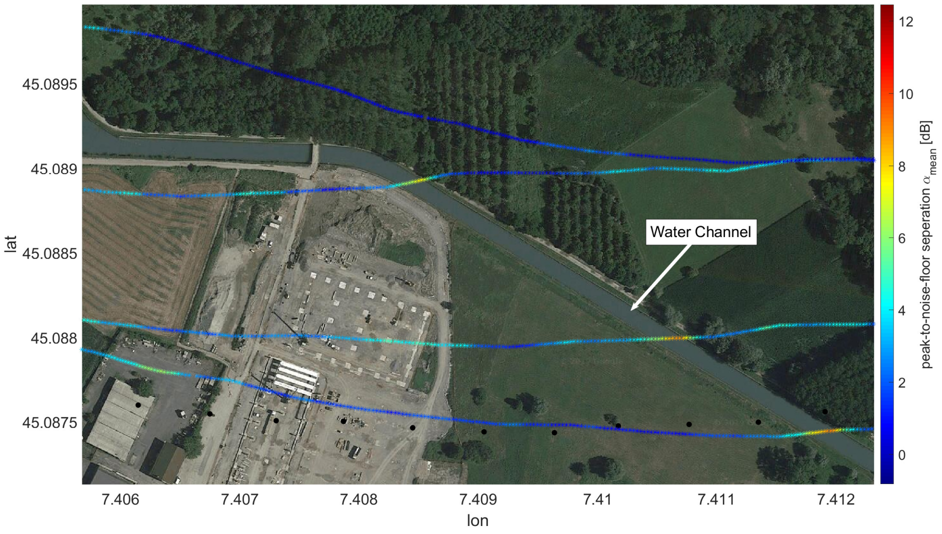


(b) Map display of the river, with the specular points superimposed on it. The colorgrade indicates the peak-to-noise-floor separation $\alpha_{mean}$ of the reflected signal.

Fig. 4.10 Example of river width detection taken from data set (1).



(a) Peak-to-noise-floor-separation $\alpha_{mean}$ of the reflected signals when the specular points were over the channel for PRN15, PRN28 and PRN32.



(b) Map display of the water channel, with the specular points superimposed on it. The colorgrade indicates the peak-to-noise-floor separation $\alpha_{mean}$ of the reflected signal

Fig. 4.11 Example of water channel width detection taken from data set (2).

Figure 4.11 reports the reflections from another portion of the river. These were measured from data set (2). The width of this water channel is approximately 12 m. The sensor and the post-processing methodology are the same, so is the result (i.e., water channel detection): boundaries of the river stretch can be detected, and its width eventually estimated.

### 4.2.3 Case-Study III: Small Artificial Water Basins

The third case study presents the detection of the presence of small water surfaces, a small lake in the backyard of building and two small adjacent artificial ponds in a golf court. Starting with the golf court water basins, Figure 4.12a shows the presence of reflected signals, in terms of peak-to-noise-floor separation for PRN 11 and PRN 32. These showed an increment of the values that could be associated to reflections. Following the same approach used for the previous case studies, the specular points lines were overlapped to the orthophotomap. Figure 4.12b shows that the edges of the pond approximately correspond to the increase (decrease) of the peak-to-noise-floor separation. The width of the upper pond is approximately 120 m while the smaller pond is approximately 35 m. This indicates that GNSS-R, in addition to the monitoring of lakes and rivers, can enable the detection of small water surfaces.

The reflections from the small backyard lake of Figure 4.13 were recorded on flight (2). The approximate width of this lake is 45 m. Like in the case of the golf court basin, the presence of the pond was clearly detected and the width of the lake is clearly distinguishable processing PRN13. These types of detection can provide valuable data to the potential application of supporting operations performed after floods, especially for post-mission analysis and in case images from satellites are not available.

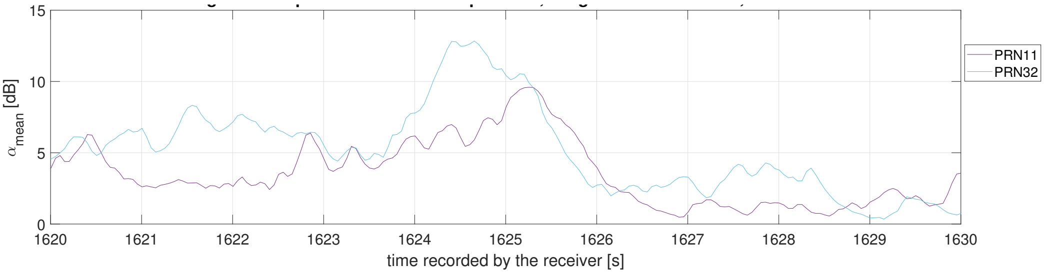
## 4.3 Estimating the Flight Height

In this section we calibrate the delay of the GNSS-R sensor (including the antennas), and accordingly estimate the height of the flight above water. We then compare the height estimated from the GNSS–R measurements to the actual height obtained from DEM.

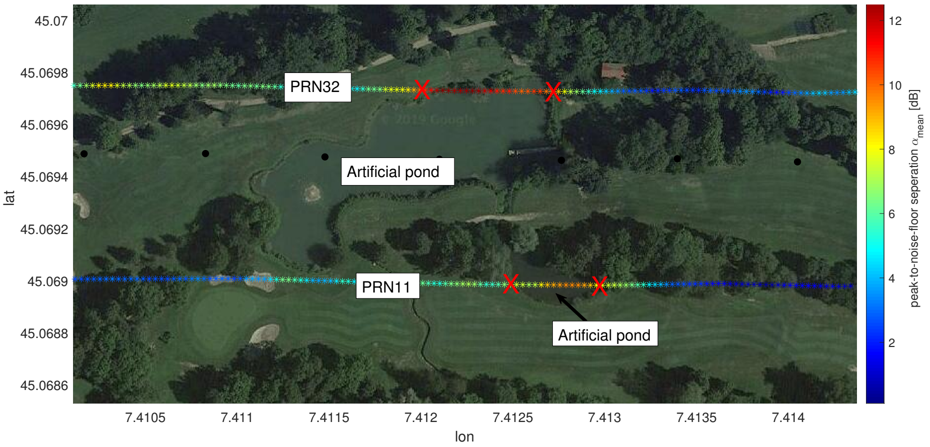
The calibration delay $\tau_{Calibration}$ is the difference between the theoretical delay, given by (3.8) and the measured delay ($\delta_\tau$):

$$\tau_{Calibration} = \delta_\tau - \frac{h}{\sin(\theta_{EL})}(1 - \sin(\frac{\pi}{2} - 2\theta_{EL}))[s] \tag{4.2}$$

For this calibration, we choose few samples of reflected signals from the various water bodies in our dataset as shown in Table 4.1. The table shows the PRN and elevation of the satellite associated with each sample, in addition to the height of the specular point. We measured the delay of the reflected signal with respect to the direct one from the CAF. Finally we estimated the calibration curve for the different elevation angles as shown in Figure 4.14. The x-axis is the elevation. The y-axis is the calibration delay in seconds (left) and meters (right).



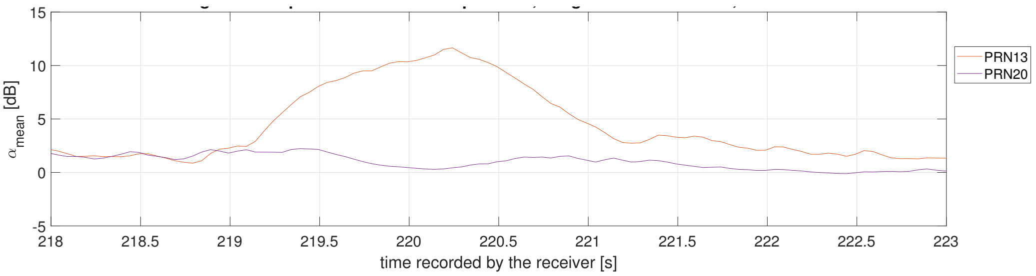
(a) Peak-to-noise-floor-separation $\alpha_{mean}$ of the reflected signals when the specular points were over the artificial pond



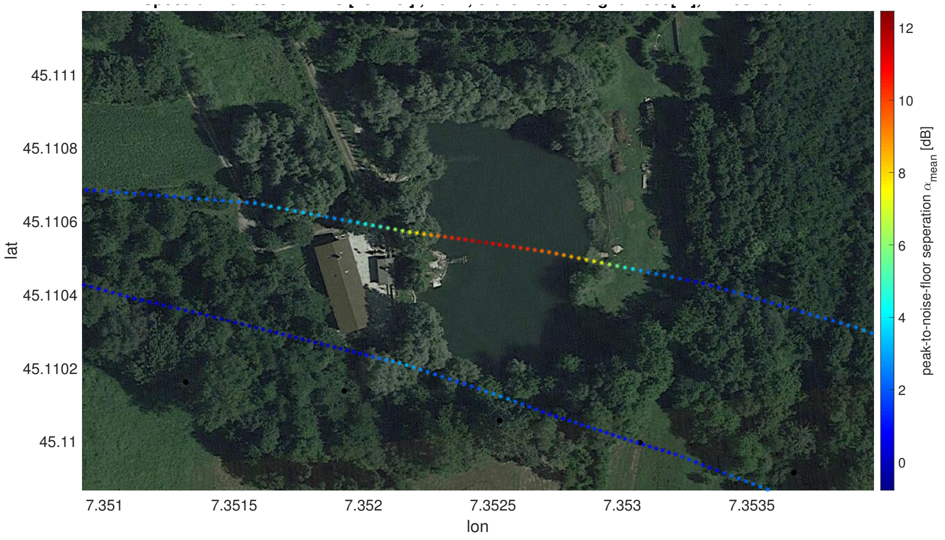
(b) Map display of the artificial pond, with the specular points superimposed on it. The colorgrade indicates the peak-to-noise-floor separation $\alpha_{mean}$ of the reflected signal

(c) Image of the two ponds

Fig. 4.12 Example of the detection of an artificial pond taken from data set (1).



(a) Peak-to-noise-floor-separation $\alpha_{mean}$ of the reflected signals when the specular points were over the backyard-pond pond



(b) Map display of the backyard-pond, with the specular points superimposed on it. The colorgrade indicates the peak-to-noise-floor separation $\alpha_{mean}$ of the reflected signal.

Fig. 4.13 Example of the detection of an backyard pond taken from data set (2).

Table 4.1 Calibration points

| PRN | EL | h |
|---|---|---|
| 01 | 59 | 355 |
| 03 | 39 | 355 |
| 04 | 77 | 580 |
| 04 | 78 | 355 |
| 11 | 83 | 580 |
| 11 | 81 | 355 |
| 19 | 46 | 355 |
| 32 | 81 | 580 |
| 32 | 79 | 355 |
| 19 | 44 | 580 |
| 01 | 60 | 580 |

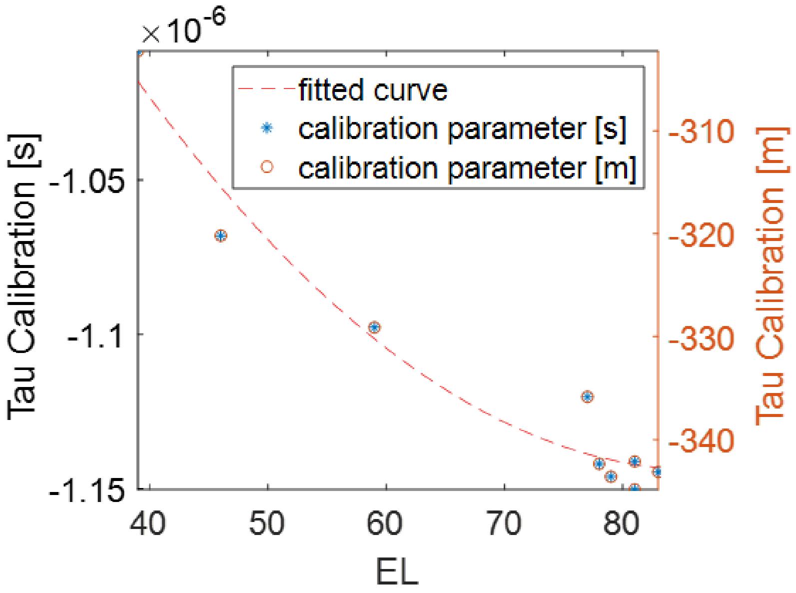


Fig. 4.14 GNSS-R sensor calibration curve.

To validate this result, we estimate the height of the flight from all the dataset, where strong reflected power was detected. In Figure 4.15 we show an example of the reflecting surface height estimation (bottom plot) for PRN3 which had the elevations shown in the upper plot. The x-axis is the sample number. Here we only selected the samples where strong reflection was recorded.

To assess the height estimation results, we show in Figure 4.16 the mean square error (MSE) of the height estimation for all the data collection (all satellites with reflections from water bodies) versus the average elevation for each satellite. As expected, for low elevations, the MSE is higher than the high elevations. The reason for this high error is the simple signal processing mechanism where we did not fine estimate the correlation peak for the reflected signal. In fact, for a 13MHz sampling frequency, the correlation peak estimation error for the C/A code with 1ms code period, 1024 chips per code and 300m per chip is:

$$\pm\frac{1}{2}\frac{1024}{13e6\times 1e-3}\times 300 = \pm 11.8[m]$$

which explains why the mean square error is taking these values.

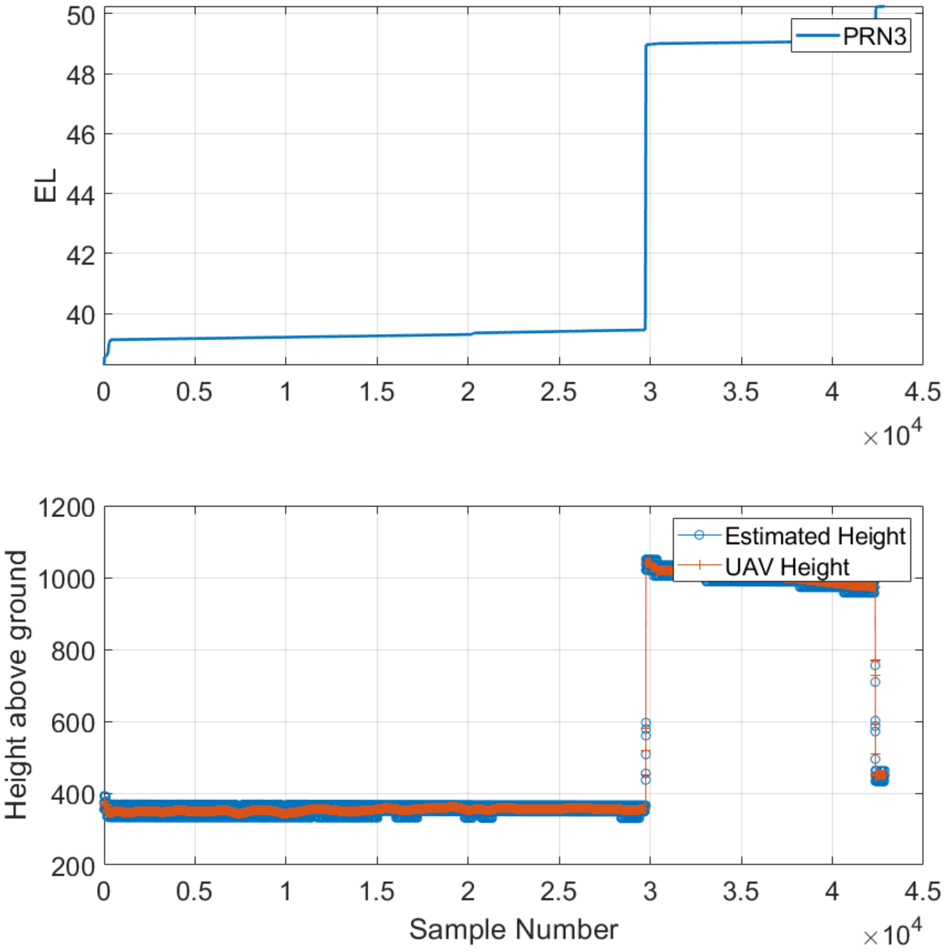


Fig. 4.15 Height estimation results

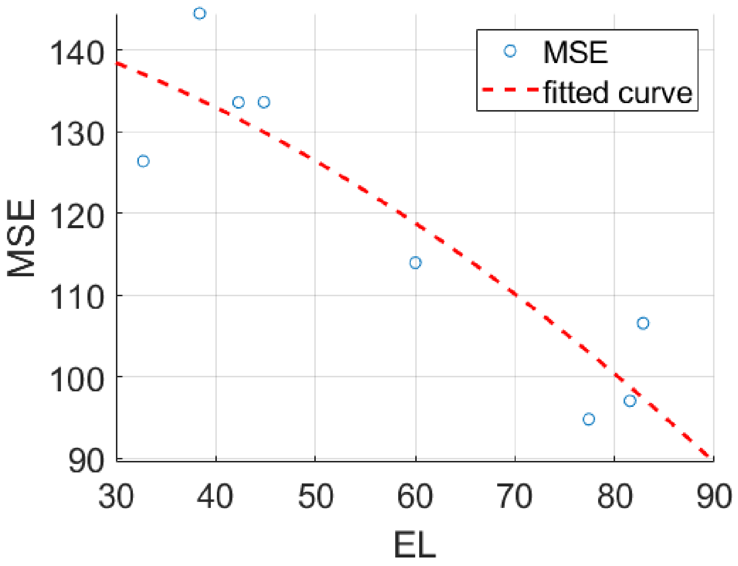


Fig. 4.16 MSE of the height estimation for the various elevations.

## 4.4 Summary

The preliminary results of investigating the feasibility of using data from UAV-based GNSS-R sensors for water detection and with a potential application in supporting flood monitoring operations were presented. The GNSS-R sensor used in this work is a custom made sensor, built using low cost commercial of the shelf components. It allowed for collecting samples of ground-reflected GNSS signals, which revealed the presence of water surfaces on ground when post processed.

Three different case studies were investigated through the processing of two different data sets collected in different seasons. The case studies are: a lake, a river and artificial water basins. These cases were selected because they challenged the sensor in detecting flood-like water presence. Performing the data collections over the lakes (and in general over large water surfaces), we were able to detect the boundaries between ground and water with few tens of meters accuracy and to estimate the extension of the water surface. Moreover, It was shown that the multi–GNSS approach (i.e.: processing of GPS, Galileo, GLONASS and Beidou signals) could even improve the estimation accuracy, without extra costs on the sensor hardware, but at the expenses of a moderate increase in the complexity of the software used in the off-line analysis. The data collections from the UAV based GNSS-R and the consequent signal processing demonstrated to be effective also for the detection of narrower water surfaces, like river stretches and water channels. In addition to that, small unexpected water presence on the ground were localized including an artificial pond in a golf court and a backyard pond. The experiments reported in this work confirmed that the use of GNSS-R onboard UAVs is a valid remote sensing tool to be used for water detection, and potentially for flood monitoring operations. However, to extend these results to flood monitoring, weather conditions affecting both the UAV operations and the water surface roughness need to be further analyzed. The UAV stability, attitude and orientation are to be considered because the geo-referencing of the measurements is affected when the RHCP and the LHCP antennas might be no longer pointing at Zenith and Nadir respectively. Moreover, investigation of the roughness and wind speed from small and narrow water bodies could be of interest, in a similar way as GNSS-R has been used for ocean surface roughness and wind speed measurements e.g in [81, 82, 84, 85]. Furthermore, since flight trajectory optimization is a common practice in UAV-based data collections, we recommend investigating how GNSS-R measurements will perform in an optimized data collection scenario. Also investigating the optimization of the UAV trajectory and the data collection time of the day to achieve optimum multi–GNSS–R measurements is suggested. Finally, for flood detection in the future, we recommend also investigating the possibility to integrate data from the digital elevation model of the terrain with data coming from GNSS-R sensors. Indeed, GNSS-R should be considered as a valuable source of geospatial data after floods, for UAV-based reconnaissance of remote areas and for environmental monitoring.

# Part II

# GNSS Scintillation

# Chapter 5

# GNSS Scintillation

Monitoring and studying the ionospheric environment has been important for decades for applications that observe and utilize transionospheric signals, like radio astronomy, satellite telecommunications and GNSS. It has also been important for terrestrial communications that rely on the ionosphere for reflecting the signal back to the ground, like the signals utilized in aviation systems. In recent years, with the increased dependency on space-based systems in our daily life, monitoring and modelling ionospheric climate and weather has become even more relevant to a wider range of applications that directly and indirectly get affected by the ionospheric conditions, including power grids, mobile communications and financial systems.

Opportunistic use of GNSS signals for monitoring the ionosphere is a well established domain now with a wide range of ground- and space-based instruments that monitor the ionosphere with a high temporal and spatial resolutions. GNSS has been one of the key enablers for the expansion in the infrastructure monitoring the ionosphere in the last three decades, due to the nature of the well distributed constellations of multi-frequency L-band radio sources (i.e. GNSS satellites) as discussed in Chapter 2. In fact, GNSS has become the standard method for monitoring the ionosphere, for example for monitoring TEC, TIDs and the ionospheric effects of space weather [118].

Studying the ionospheric environment, however, is not easy because of the complexity of this environment. The ionosphere, situated between the space and the terrestrial environments, is affected by a wide range of factors from above and beneath, including Sun conditions, the Earth magnetic field, earthquakes, volcanic eruptions and solar eclipses. For such environment, noise and interference on the measurements can make the monitoring task even more complicated.

This part of the thesis addresses the methods and algorithms for monitoring the ionosphere exploiting GNSS signals. We focus, in particular, on the scintillation effect which occurs when the ionosphere presents an irregular distribution of free electrons. When scintillation occurs, trans-ionospheric signals can be degraded or even lost. This directly and indirectly affects many of the essential infrastructures including radio communications [119–121], remote sensing [122] and global navigation [123].

Monitoring ionospheric scintillations through GNSS signals has been studied for almost 40 years now. Nevertheless, there are many issues that are still challenging. First, the metrics for detecting scintillation are sensitive to interference from the environment around the receiver, including multipath effects (mimicking scintillation patterns in the received signals) from objects around the receiver [124, 125] and RFI from intentional and unintentional sources [126, 127]. Furthermore, these metrics depend on the detrending schemes adopted to retrieve the scintillation indexes. Also, these metrics do not give definitive measures to classify the severity of the scintillation event in terms of the phenomenum underlying it. In particular, scintillations are classified usually as weak, moderate and strong based on thresholds set on the scintillation indexes. These thresholds are not standard, and even the same authors use different threshold values on different studies [61]. This invites research on robust scintillation detection mechanisms that can overcome the limitations of the current detection metrics and methodologies. Moreover, scintillation monitoring receivers are very expensive, and their distribution around the globe is sparse. Although half of the Earth surface is susceptible to this phenomenon, the number of scintillation monitors are concentrated in few regions specially those with high commercial demand for ionospheric data [128, 129]. This invites both maximizing the utilization of the available scintillation data and also the search for economic solutions for scintillation monitoring. Furthermore, GNSS based scintillation monitoring has been continuously operating for more than 20 years now. This resulted in huge repositories of various scintillation data in various formats. The automatic processing of these data, to curate and clean the repositories from non scintillation related data for example, has not gained the same momentum. This results in great effort from ionospheric researchers to deal with the GNSS scintillation data. This invites research on utilizing state-of-the-art data mining techniques to assist in curing, exploring and utilizing these data repositories.

In this thesis we tackle these issues together by developing ML models able to detect GNSS scintillation. We address the two types of scintillation events: polar scintillation (Chapter 7) and equatorial scintillation (Chapter 8). For the latter, we distinguish scintillation from multipath in order to address one of the main contaminators of scintillation measurements. Multipath events are often roughly filtered by excluding low elevation satellites, which are more likely to be affected by the phenomenon, from the observations. By implementing smarter criteria and algorithms for the detection, we can increase the amount of useful scintillation data by lowering the elevation mask and thus increasing the visible sky. As far as the polar scintillations are concerned, we investigate alternatives for the scintillation metric that can be used for scintillation detection. This could open the door for monitoring scintillation using a wide range of GNSS receivers beside those specially configured for scintillation monitoring.

In this chapter we describe the ionospheric scintillation and how it affects the GNSS signal, and thus how the GNSS receiver handles scintillation measurements. The chapter is organized as follows. First we describe the ionosphere (Section 5.1), ionospheric scintillation (Section 5.2) and the importance of scintillation monitoring (Section 5.3). Then the effects of the ionosphere, focusing on the diffraction and refraction effect, on L-band signals are presented in Section 5.4. Finally, we describe how state-of-the-art GNSS receivers compute the scintillation indexes (Section 5.5) and the other measurements, that could be related to ionospheric monitoring, provided by these receivers (Section 5.6). We conclude the Chapter in Section 5.7 by discussing how GNSS scintillation detection is implemented.

## 5.1 The Ionosphere

The ionosphere is a layer of ionized gases that surrounds the Earth from about 50 to 1000 km above surface. It is constituted by electrons and ions, thus it is a plasma that is globally neutral. The Sun is the main driver and influencer behind this layer because the ionization is mainly caused by the Sun's radiation. As a result:

- The state of the ionosphere at the day side of the Earth is different from the night side.
- The state of the ionosphere changes with the seasons.
- The state of the ionosphere depends on the solar conditions.

The electron density in the ionosphere varies along the altitude, with the maxima and minima of the electron density defining the ionospheric regions and layers: D-, E-, and F-layers (from Earth up). Figure 5.1 (image credit [130]) shows a sketch of the ionospheric layers at the day and night sides. The x–axis is the electron density and the y–axis is the altitude. The electron density and height of the layers differ from day to night with the D- layer completely disappearing (or merging with the E-layer) at night and the F-layer splitting into two at the days side. Scintillations are hypothesized to occur due to disturbances in the E- and F-layers mainly [131].

Being an ionized medium, the ionosphere is affected by the Earth magnetic field. Therefore, the state of the ionosphere also differs depending on the geomagnetic latitude.

The day/night effects of the Sun are interesting especially near sunset when the radiation source disappears and thus the ionization changes. At low latitudes in particular, the state of the ionosphere after sunset is characterized as disturbed, and the disturbance can continue for many hours in the night. The ionosphere state around local sunrise is not as disturbed as the sunset, and few studies reported noticeable disturbance at that time of the day [132].

On the other hand, the effects of the solar storms on the state of the ionosphere near the geomagnetic equator and the geomagnetic poles are totally different. Near the geomagnetic poles, where the geomagnetic field lines are vertical, the ionosphere is frequently disturbed, while near the geomagnetic equator, the geomagnetic storms can either enhance or suppress the usually disturbed post-sunset time [133–135].

The state of the ionosphere at mid-latitude is characterized generally as quiet day and night, however, it can be significantly disturbed during solar storms as reported for example in [136–138]. It can also be affected by disturbances like those represented by TIDs [139, 140].

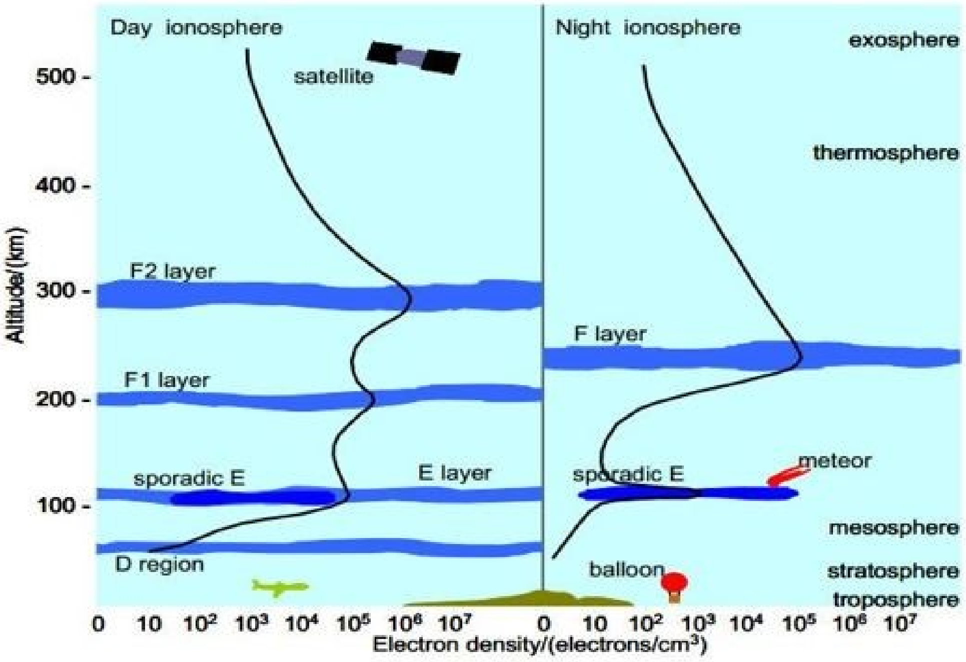


Fig. 5.1 Day night structure of the ionosphere

## 5.2 Ionospheric Scintillation

Ionospheric scintillations are rapid random fluctuations in the amplitude and phase of electromagnetic wave signals as observed by a radio receiver. It happens when the signals pass through small scale irregularities (with respect to the signal wavelength) in the electron density distribution of the ionosphere. Scintillation affects trans-ionospheric radio wave propagation in the range of kHz to GHz including satellite communications, astronomy observations [141], radar remote sensing, and GNSS signals.

Ionospheric irregularities are concentrated near the magnetic equator (mainly observed pre-midnight period), in auroral zone (during the night time period), and in the polar region (at all local times). Thus, Scintillation happens mainly around these regions. The mechanism behind the equatorial and polar scintillations are different. The former is mainly derived by irregularities in the electron density that form after local sunset when the main source of ionization, the Sun, disappears.

At low latitude the daytime ionosphere presents two electron density maxima, called crests, on average located at $+/-20°$ off the equator, and a minimum located in between the crests, the trough. Such configuration is known as the Equatorial Ionization Anomaly (EIA) and it is essentially formed from the removal of ionospheric plasma from around the equator by the upward ExB drift (motion of charged particles in a magnetic field) ([142] and references therein). After sunset, Equatorial Plasma Bubbles (EPB) which are localized depletions in the plasma density, can form. In their evolution and decay process, small-scale irregularities can form, mainly because of Rayleigh-Taylor instability, causing scintillations on the radio signals received at ground. Under geomagnetically perturbed conditions the EIA can be significantly modified and the formation of the irregularities can increase or suppress. Equatorial scintillations thus have seasonal (it is more dense at the equinoxes) and diurnal dependency.

Polar (or high latitude) scintillation is associated to magnetic storms and thus it shows dependency on space weather and the Sun conditions. Here, ionospheric irregularities happen due to magnetic storms, when batches of the ionosphere are pushed from the day side over the polar caps (the projection of the two lobes of the magnetic tail onto the atmosphere) into the night side creating irregularities in the ionosphere [143]. The scale sizes of high latitude irregularities range goes from hundreds of kilometers down to few centimeters. Their structures are utterly dynamic, with the Interplanetary Magnetic Field (IMF) controlling their convective motion. In the night-time auroral oval and the cusp, medium to large scale structures (structures with scales larger than $\approx 50$ km) are associated with structured fluxes of precipitating electrons. Small scale irregularities are typically due to instabilities and turbulent processes in the cusp, where a tongue of ionization (TOI) is broken into batches [144]. These structures are unstable and they can produce smaller irregularities through the current-convective instability or E×B, cascading and wave-wave interaction. High latitude large-scale structures have long lifetimes and can be convected long away from their origin places. [145].

## 5.3 Why Scintillation Monitoring

For navigation, scintillation results in degraded performance of the service and it might lead to complete loss of the service. This is not acceptable for many applications where the availability and integrity of the system is crucial. For example, in power grids and telecommunication networks, where the GNSS signal is utilized for

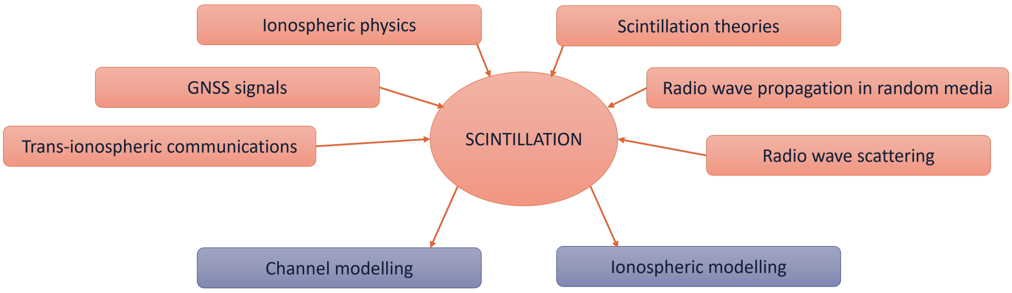


Fig. 5.2 Scintillation studies are beneficial for both ionospheric sciences and GNSS systems.

timing purposes, the availability the GNSS service is compelling. In precise positioning applications, for example in oil&gas and surveying industries, the accuracy of the systems (sometimes beside availability) are essential.

These users require to be alerted about scintillation occurrence, at least in real-time, so that they can abort operations or use alternative solutions until the GNSS signal is reliable again. The possibility to forecast scintillation is further desirable to help these users schedule operations in advance.

On the other hand, the study of the physical phenomenon behind scintillation is important for the scientific community. The ionosphere is an important part of our planet, and constitutes the boundary between space and the lower atmosphere where life exists. By observing scintillations, we can better understand the physics of the upper atmosphere, the Sun-Earth interaction and the effects of space weather. Monitoring scintillations will enable refining the ionospheric models of today with accurate parameters for the effect of the irregularities.

For the radio telecommunication sciences, understanding scintillation is important for modelling the trans-ionospheric communication channels and also for designing communication signals that are robust against scintillations. For GNSS in particular, understanding scintillation is further important in order to design receivers robust against their effects and/or to mitigate the effects of the ionosphere on GNSS measurements.

In summary, studying the ionosphere is mutually beneficial for the ionospheric sciences and the GNSS community. Knowledge about the physics of the scintillation phenomenon as well as about the GNSS-based measurement instruments are essential to achieve this understanding [146, 147]. We illustrate this dependency in Figure 5.2.

A map display for the frequency of GNSS scintillation occurrence is shown in Figure 5.3a. The Equatorial disturbance can affect the signal up to 100 days per year [28]. In the map, it can be seen that large part of South America, Africa and South East Asia beside the high latitude regions are affected by scintillation. For comparison, we show in Figure 5.3b a map display of the world population density [148] as of 2022[1], where the high frequency of scintillation occurrence at some of the regions with high population density can be noticed. This affects the navigation service for a considerable portion of the Earth population, but it does not motivate monitoring scintillation unless there is an economic derive and demand behind this. For this reason, we show in Figure 5.3c a map display of the economic activity per km$^2$, where it can be noticed that the economic activity is not high for most of the regions with high scintillation occurrence. Although this might indicate less economic impact for scintillation, it also less motivates installing scintillation monitoring stations where the demand for scintillation data will be low. This negatively affect the geographic distribution of ionospheric scintillation measurements that rely on expensive scintillation monitoring stations, and motivates the search for alternative scintillation monitoring technologies using other GNSS infrastructure that are available for broader economic uses for example geodetic receivers and the IGS network shown in Figure 5.4 (image source [149]).

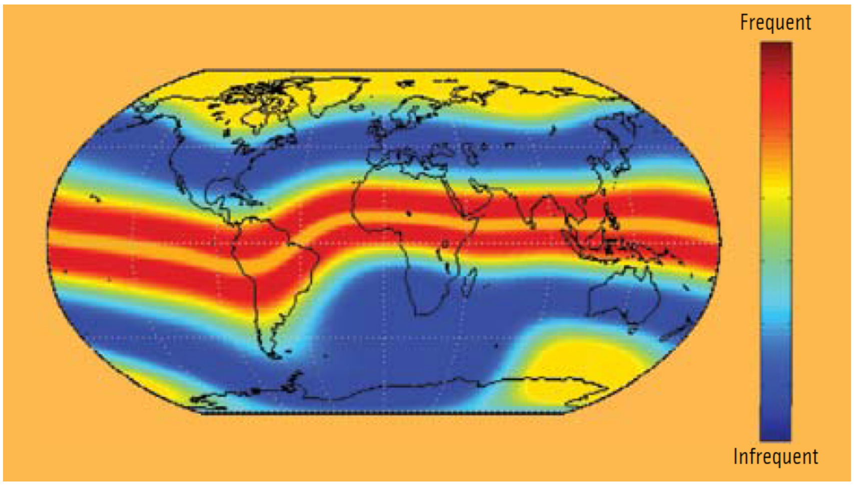
For this reason, we explore using GNSS measurements other that the standard scintillation indexes (will be introduced in Section 5.5) for scintillation detection.

## 5.4 Ionospheric Effects on the GNSS Signals

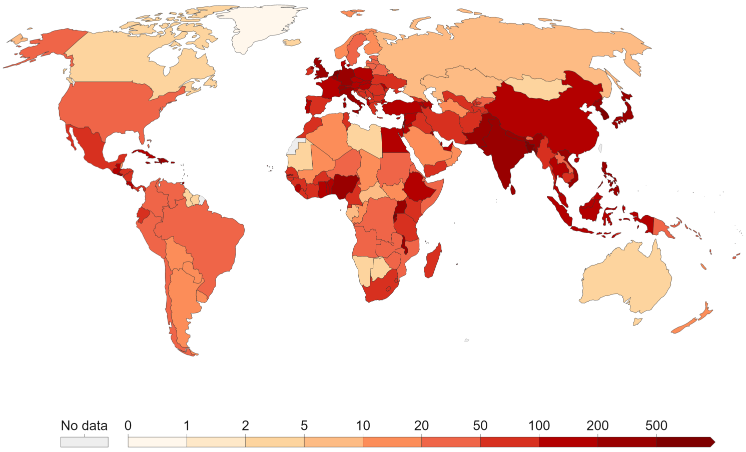
As mentioned in the GNSS frequencies section (Section 2.3), L-Band was selected for the GNSS ranging signals because in this band it can provide navigation service regardless of the weather conditions. We report in Figure 5.5 (image credit [151]) the atmospheric opacity for a wide range of frequencies (we report here in wavelength) [151]. It can be seen that the atmosphere is transparent for L-Band signals (wavelength 19cm for L1/CA). In fact, even during scintillation, the GNSS signal does not suffer from ionospheric absorption. The fluctuation in the signal power associated with scintillation comes from the signal interfering with itself resulting

---

[1] https://ourworldindata.org/world-population-growth

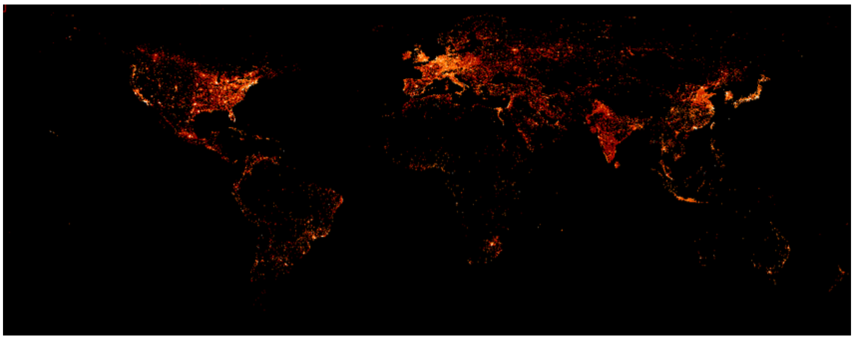


(a) Recurrence of scintillation at solar maximum [image credit [28]]



Source: World Bank (2021), Gapminder (v6), HYDE (v3.2), UN (2019)  OurWorldInData.org/world-population-growth • CC BY

(b) World population density, 2022 [image credit: World Population Growth]



(c) World total economic activity in millions of dollars per $km^2$ pixel [image credit: [150]]

Fig. 5.3 Map display of the scintillation occurrence frequency and relating it to the population density and economic activity maps

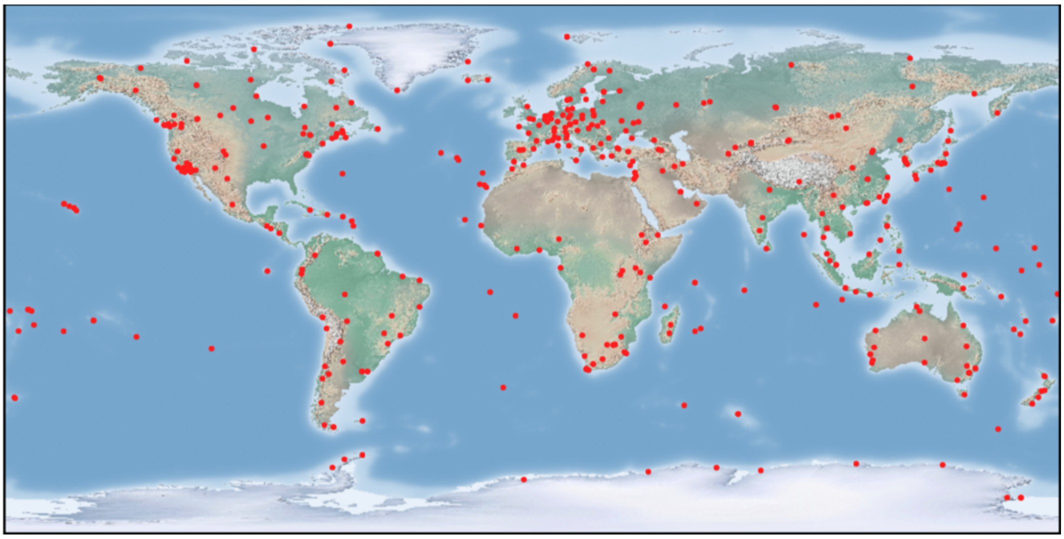


Fig. 5.4 IGS network with 512 geodetic receivers as of 2019.

in constructive and distructive combination of the signal at the receiver as will be explained in this section.

If we follow the journey of a GNSS signal from a satellite to a receiver on the ground, focusing on the propagation through the ionosphere (which represents 5% of the propagation environment), as depicted in Figure 5.6, we can recall the following from electromagnetic waves theory. The existence of different media along the ray path will result in bending the signal, just like the bending effect experienced by light travelling from air to water. But most importantly, the speed of propagation of the electromagnetic wave will change. As will be seen in the next paragraphs, this change of speed has implications that are crucial to GNSS signals processing. This is called the refractive effect of the medium. The refractive effect will always be experienced by the GNSS signals as they pass from space to the ionosphere as well as while exiting from the ionosphere into the troposphere. Above that, when the electron density of the ionosphere changes, as depicted by the gradient color in the cartoon, the refraction will change. Furthermore, depending on the size of the media relative to the wavelength of the ray, the signal might also experience scattering just like a ray of light scattering by rain drops to form rainbows. The later are called the diffractive effects of the ionosphere and also it is the most dangerous effect of the ionosphere on GNSS systems. In Section 5.4.1 and Section 5.4.2 we explain the refractive and diffractive effects of the ionosphere also known as the deterministic and stochastic effects, relatively.

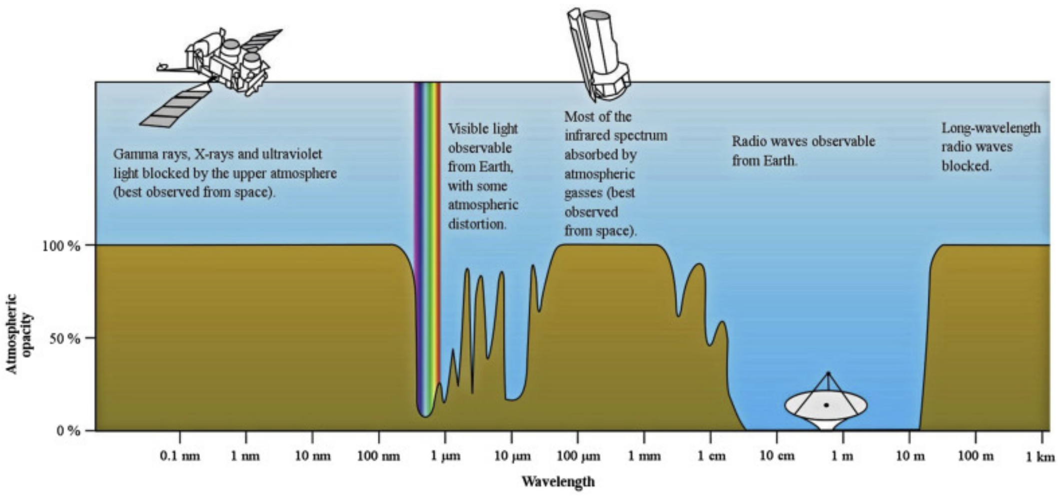


Fig. 5.5 atmospheric opacity.

### 5.4.1 Refractive Effects

The refractive effect of the ionosphere on GNSS signals mainly affects the delay of the signals.

From electromagnetic waves propagation physics, we know that the speed of propagation of a wave within a certain medium depends on the refractive index of the medium. The refractive index of the ionosphere depends on the electron density, so that the speed of a radio wave propagating through the ionosphere depends on the number of free electrons along its path. This speed is related to the speed of light via the refractive index of the medium ($n$) defined as:

$$n = \frac{\text{speed of propagation in vacuum}}{\text{speed of propagation in the medium}} = \frac{c}{v} \geq 1 \tag{5.1}$$

As a result, the refractive index of the ionosphere (how fast the electromagnetic wave travels through the ionosphere) changes with the change in the ionization level along the ionospheric propagation path. This results in two refractive effects of the ionosphere that are experienced by GNSS signals:

1. Group delay: which results from the signal travelling with a speed slower than the speed of light, and thus the measured delay (receive time minus transmit time) being greater that the free-space propagation delay, and thus this extra delay is recorded as the delay due to the ionosphere.

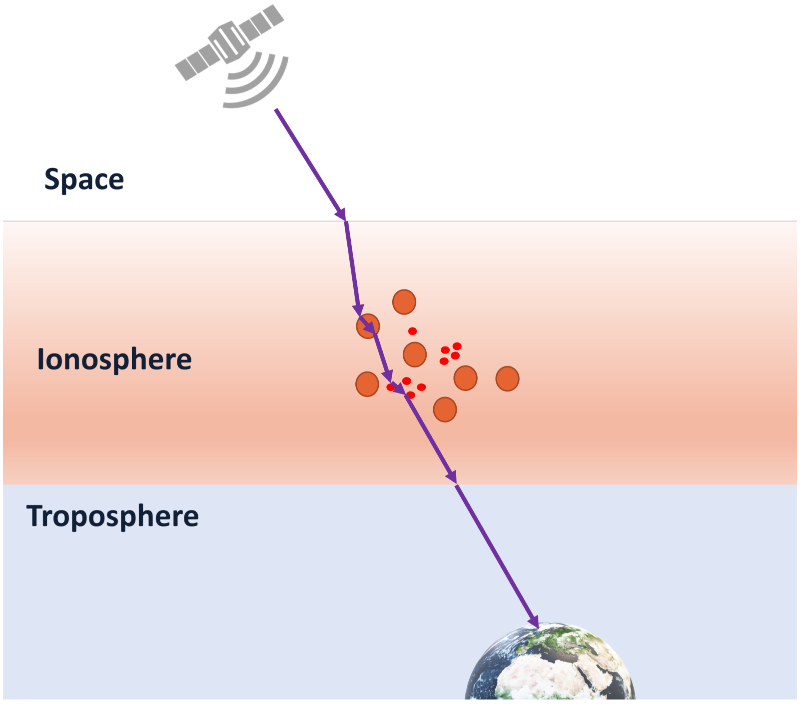


Fig. 5.6 Ionospheric Propagation

2. Phase advance: which results from the carrier travelling with a speed slower than the speed of light. This translates to the wavelength of the signal (the distance travelled by one cycle of the carrier) becoming smaller ($\lambda = \frac{v}{f}$) where $v < c$ and $f$ is a constant because the frequency (the number of cycles of the carrier in a second) of the signal transmitted by the satellite is fixed. Since the distance between the satellite and the receiver is the same regardless of the speed of propagation, this means for the carrier to travel this distance passing through the ionosphere it will arrive with more cycles i.e with an advanced phase. Figure 5.7 illustrates the concept of phase advance.

Form the previous paragraphs, it can be understood that:

- The ionosphere is a dispersive medium: i.e. the effects induced by the ionosphere is frequency dependent. This means that the ionospheric delay can be estimated by observing signals with different frequencies. This is exactly the reason why satellite based radio navigation systems broadcast their signals in multiple frequencies as anticipated in Section 2.3.

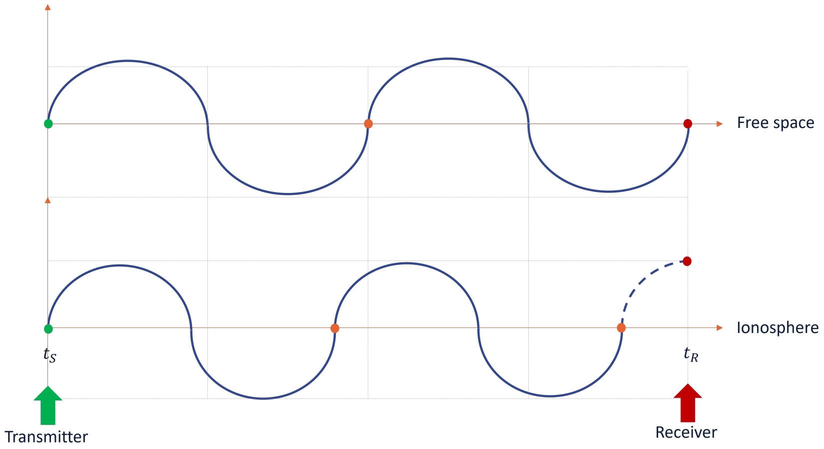


Fig. 5.7 Illustration of GNSS signal phase advance due to the ionosphere

- The refractive index for the code is different than the carrier of the GNSS signal and delays experienced by the code and carrier have different signs. Thus, we introduce the code index of refraction and phase index of refraction respectively.

The ionospheric group delay ($I_P(f)$) and phase advance ($I_L(f)$) can be estimated by integrating the index of refraction along the signal path from the satellite $S$ to the receiver $R$ as [152] [153]:

$$I_P(f) = \int_R^S (n_P(f) - 1)\, dl, \tag{5.2}$$

$$I_L(f) = \int_R^S (n_L(f) - 1)\, dl, \tag{5.3}$$

where $n_P$ and $n_L(f)$ are the group and phase index of refraction respectively. The evaluation of this integration has been studied by many authors for various applications (precise positioning, scintillation monitoring, geodesy) under different ionospheric conditions, at different latitudes, making relevant assumptions on $n$, the path between the satellite and the receiver, the signal, the ionosphere hight, the geomagnetic effects, etc. We adopt here $n_P$ and $n_L$ evaluated by [152] [153] because they exclude the sub millimetre terms, which as will be clear in section 5.6, are beyond the capabilities of GNSS receivers under disturbed ionospheric conditions when the receivers experience errors that reach tens of centimetres. $n_P$ and $n_L$ for the

RHCP GNSS signals are calculated from [152] [153]:

$$n_P(f) = 1 + \frac{f_p^2}{2f^2} + \frac{f_p^2 f_g \cos\theta}{f^3} + \frac{f_p^4}{f^4}, \tag{5.4}$$

$$n_L(f) = 1 - \frac{f_p^2}{2f^2} - \frac{1}{2}\frac{f_p^2 f_g \cos\theta}{f^3} - \frac{1}{3}\frac{f_p^4}{f^4}, \tag{5.5}$$

where $f_p$ is the plasma frequency, $f_g$ is the electron gyrofrequency, and $\theta$ is the angle between the geomagnetic field vector and the wave propagation direction.

Plasma frequency, $f_p$, is the frequency at which the electrons in the plasma naturally oscillate relative to the ions. For the ionosphere, it has typical values between 2 and 20 MHz [154]. Electron gyrofrequency, $f_g$, is the angular frequency of the circular motion of an electron in the plane perpendicular to the magnetic field. $f_p$ and $f_g$ are computed from:

$$f_p = \sqrt{\frac{e^2 n_e}{4\pi^2 m_e \varepsilon_0}}, \tag{5.6}$$

and

$$f_g = \frac{|e|B}{2\pi m_e}, \tag{5.7}$$

where $e = 1.60217662 \times 10^{-19}$ $c$ is the electron charge, $n_e$ is the electron density, $m_e = 9.10938356 \times 10^{-31} kg$ is the electron mass, $\varepsilon_0 = 8.85 \times 10^{-12}$ $F$ is the free space permittivity, and $B$ is the geomagnetic induction. Combining (5.2) - (5.7) and substituting the constants:

$$I_P(f) = -\frac{40.3}{f^2}\int_S^R n_e\, ds - \frac{2.26 \times 10^{12}}{f^3}\int_S^R B\cos\theta n_e\, ds - \frac{2439.42}{f^4}\int_S^R n_e^2\, ds, \tag{5.8}$$

$$I_L(f) = \frac{40.3}{f^2}\int_S^R n_e\, ds + \frac{1}{2}\frac{2.26 \times 10^{12}}{f^3}\int_S^R B\cos\theta n_e\, ds + \frac{1}{3}\frac{2439.42}{f^4}\int_S^R n_e^2\, ds, \tag{5.9}$$

The integral $\int_S^R n_e\, ds$ along the ray path corresponds to the TEC which is defined as the number of electrons in a cylinder of $1m^2$ cross section from a receiver to a satellite. TEC is one of the parameters that are widely used to characterize the ionosphere. It is measured in TEC units (TECU) where $1TECU = 1 \times 10^{16} electrons/m^2$. The typical values of TEC are in the range $1 - 100$ TECU, reaching higher values under very disturbed ionospheric conditions. Rewriting (5.8) and (5.9) in terms of TEC:

$$I_P(f) = -\frac{40.3TEC}{f^2} - \frac{2.26\times 10^{12}}{f^3}\int_S^R B\cos\theta n_e\, ds - \frac{2439.42}{f^4}\int_S^R n_e^2\, ds, \quad (5.10)$$

$$I_L(f) = \frac{40.3TEC}{f^2} + \frac{1}{2}\frac{2.26\times 10^{12}}{f^3}\int_S^R B\cos\theta n_e\, ds + \frac{1}{3}\frac{2439.42}{f^4}\int_S^R n_e^2\, ds, \quad (5.11)$$

The first order delay ($\frac{40.3TEC}{f^2}$), is $0.1625m/TECU$ for L1 frequency (1575.42 MHz) for example. Dual frequency GNSS receivers are able to estimate this delay and thus remove its effect.

The integral $\int_S^R B\cos\theta n_e\, ds$, on the other hand, corresponds to integrating the product of the magnetic field and the electron density which are both varying along the path of propagation. To calculate the exact value of this second order delay term, knowledge of the exact distribution of the electron density along the propagation path in needed. This is generally not available to dual frequency GNSS receivers. However, by assuming that the ionosphere is a thin layer located at a fixed altitude (generally 350km), an approximation of this second order delay is achievable with RMS of 1.25cm.

The last part of (5.10) and (5.11), the third order delay, corresponds only to the square of the electron density along the signal path. It can be approximated to [155]:

$$\int_S^R n_e^2\, ds = \eta N_{max}\int_S^R n_e\, ds = \eta N_{max} TEC, \quad (5.12)$$

where $N_{max}$ is the peak electron density along the signal path and $\eta$ is the shape parameter which is a value that counts for the different electron density distributions along the ray path. Thus, this third order delay is directly affected by the max electron density in addition to TEC. It can reach 10 centimetres on disturbed ionospheric conditions.

To conclude this discussion about the 1st, 2nd and 3rd order ionospheric delays, we show in Table 5.1 (Source: [152]) an example of typical delay values that were observed by [152] on quiet and disturbed ionospheric conditions. We report the max values observed in addition to the 1 $\sigma$ value. It can be appreciated that the delays on disturbed days are an order of magnitude more than on quiet days. More importantly, the 2nd and 3rd order delays which are not removed by double

Table 5.1 Higher order Ionospheric delays and Residual Errors.

| | | | Quiet day | Storm day |
|---|---|---|---|---|
| Group delay [mm] | 68th percentile | 1st | 15 m | 10 m |
| | | 2nd | 5 mm | 5 mm |
| | | 3rd | 1 mm | 1 mm |
| | Max | 1st | 35 m | 110 m |
| | | 2nd | 25 mm | 85 mm |
| | | 3rd | 5 mm | 100 mm |
| Max Residual Double Difference Code Error [mm] | | | 30 | 185 |
| Max Residual Double Difference Phase Error [mm] | | | 15 | 75 |

differencing techniques are not negligible values. In this case 185mm and 75mm residual code and phase errors, respectively, were observed by the authors.

From positioning point of view, these delays (if not counted for) will result in errors in estimating the position. However, from the point of view of ionospheric monitoring, information about the ionosphere can be extracted from observing the residuals in pseudo-range measurements which act as the extra delay faced by GNSS signals as they pass through the ionosphere.

### 5.4.2 Diffractive Effects

The differactive effects of the ionosphere are the reason for the appearance of what is generally described as "scintillations". Diffraction, the spreading/scattering of waves as they pass through or around obstacles, occurs when the size of the obstacle is of the same order of magnitude as the wavelength of the incident wave (i.e. below Fresnel's scale).

If the ionosphere is disturbed with irregular distribution of ions, as illustrated in Figure 5.8, where the different colours of the irregularities indicate different ions concentration and different irregularities sizes, then the signal will be suffering from many of the above two effects (delays and phase advance). It will thus undergo different refractions and diffractions along its path according to the corresponding refractive index of each ions irregularity they pass through and the size of the irregularity . The end result will be a wave front that is a combination of many signals that travelled different paths, on each path it went through many refractions and diffractions that changed the speed and direction of propagation. It is important

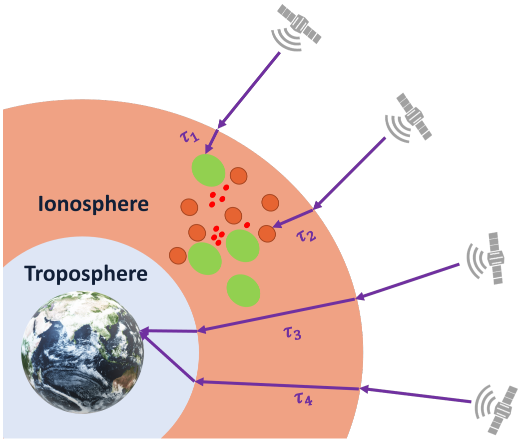


Fig. 5.8 Illustration of GNSS scintillation.

to mention here that there are no ionospheric absorptions due to the irregularities. At the receiver, these signals will add constructively and destructively, resulting in a total signal with randomly and rapidly fluctuating phase and amplitude. This fluctuation in phase and amplitude is called phase and amplitude scintillation respectively.

From the above description of the scintillation phenomena, we can describe the final signal emerging from the ionosphere at time $t$ as the sum of many rays of the satellite signal, each one has its delay and phase as:

$$r^i(t) = \sum_{n=0}^{\infty} s_n^i(\tau_n, \phi_n) \tag{5.13}$$

where $s_n^i$ is the signal broadcast by satellite $i$ and followed the path number $n$ undergoing a delay $\tau_n$ and a phase advance $\phi_n$. In radio wave propagation terms this equation can be expressed and then solved using the theory of radio wave propagation in random media. The vector wave equation can be expressed in the scalar wave equation form:

$$\nabla^2 E + k_0^2 < \varepsilon > [1 + \varepsilon_1(\mathbf{r}, t)]E = 0, \tag{5.14}$$

where $k_0$ is the wave number in free space (the number of complete wave cycles per unit distance), $< \varepsilon >$ is the average dielectric permittivity, $\varepsilon_1(\mathbf{r}, t)$ is the fluctuating part of the dielectric permittivity caused by electron density irregularities, and $(\mathbf{r}, t)$ specifies the location of the irregularity is space and time. Equation (5.14), a differential equation with randomly fluctuating coefficients, forms the basis of the scintillation theory.

Using the phase screen model [156], and solving (5.14), the phase $\phi(\mathbf{r},t)$ and amplitude $u(\mathbf{r},t)$ of the wave $E$ can be approximated to [157]:

$$\phi(\mathbf{r},t) = -\lambda r_e \Delta N_T(\mathbf{r},t), \tag{5.15}$$

$$u(\mathbf{r},t) = A_0 exp[-i\phi(\mathbf{r},t)], \tag{5.16}$$

where $\lambda$ is the wave length of the carrier, $r_e$ is the classical electron radius, and $\Delta N_T$ is the fluctuation of the electron content between the transmitter and the receiver.

Being a stochastic phenomena, statistics about the signal like the average, standard deviation, rms, etc, are the best way to describe the general phenomena. The standard deviation of the signal amplitude and phase were adopted by the ionospheric studies community and two indexes were defined: the phase scintillation index $\sigma_\phi$ and the amplitude scintillation index $S_4$.

In phase screen theory, The phase scintillation index and amplitude scintillation index are given by [156]:

$$\sigma_\phi = \sqrt{r_e^2\lambda^2\sec\theta}\sqrt{GLC_s q_0^{-2\nu+1}\frac{\Gamma(\nu-1/2)}{4\pi\Gamma(\nu+1/2)}}, \tag{5.17}$$

$$S_4 = \sqrt{r_e^2\lambda^2\sec\theta}\sqrt{\mathscr{G}LC_s Z^{\nu-1/2}\frac{\Gamma([2.5-\nu]/2)}{\sqrt{4\pi}\Gamma([\nu+1/2]/2)(\nu-1/2)}}, \tag{5.18}$$

where $r_e$ is the classical electron radius, $\lambda$ is the wavelength of the satellite signal, $\theta$ is the satellite zenith angle, $G$ is a factor that relies on the satellite zenith angle and the axial ratios of the irregularities (equals 1 in the case of isotropic irregularities), $L$ is the thickness of the irregularity slab, $C_s$ is the turbulence strength parameter (defined at that height where the phase screen is assumed to be located), $q_0$ is the outer-scale cutoff wave number, $\nu$ is the 3-D spectral index, $\mathscr{G}$ is a factor that is a function of the propagation angles relative to the irregularity principal axis as well as the axial ratio. $Z$ is the Fresnel-zone parameter.

The first part of (5.17) and (5.18) ( $\sqrt{\lambda^2\sec\theta}$) are know values for a satellite. The second part of the equations encompass the terms that describe the irregularity (shape, thickness, strength, etc). The standard deviation of the signal phase and amplitude are among the parameters that can be provided in near–real time by those GNSS receivers capable to sample the signals at high sampling rates (50 to

100 Hz) and equipped with an integrated firmware able to derive the indices from the measurements. The indices can be also derived in post–processing from the high sampling rate measurements (see Section 5.5). From there, the ionospheric irregularities' parameters can be estimated.

## 5.5 Scintillation Metrics

Measurements of ionospheric scintillation have been continuously collected since the first artificial satellites in the 1950s, and the preceding radio star measurements in the 1940s [158]. To interpret these measurements, many theories of radio wave scattering have been applied [158] [157] and many ionospheric scintillation models have been developed [159–161].

### 5.5.1 Amplitude Scintillation Index

The Amplitude scintillation index ($S_4$) is the standard deviation of the detrended received signal intensity normalized to the average signal intensity [162]. This $S_4$ definition is applicable to any trans-ionospheric communication signal, transmitted by any satellite at any orbit. When applied to GNSS-based scintillation monitoring, the terms detrending and signal intensity are not trivial. In the previous section, we showed that the signal intensity estimated by a GNSS receiver is affected by receiver dependent noise and gain (2.9). These can be removed by subtracting the noise value (can be estimated by the receiver) and normalizing the signal intensity. The signal intensity is also affected by the propagation losses that are induced by factors other than the irregularities in the ionosphere. Detrending aims to remove the latter effects. In the following paragraphs we explain how $S_4$ metric is computed taking into account these effects.

#### Detrending the Signal Intensity

When detrending the signal intensity, a low pass version of $SI$ is subtracted from $SI$ with the objective of removing the non scintillation–related components. The filter applied is usually a sixth order Butterworth low pass filter with 0.1 Hz cutoff frequency [35, 36], but other detrending methods are also present in the literature.

For example, [163] exploits the wide and narrow band power, $WBP$ and $NBP$ respectively, to detrend the signal. NBP, WBP and the detrended signal intensity are computed as:

$$WBP = \sum_{m=1}^{M} (I_m^2 + Q_m^2), \tag{5.19}$$

$$NBP = \sum_{m=1}^{M} I_m^2 + \sum_{m=1}^{M} Q_m^2. \tag{5.20}$$

$$SI = \frac{NBP - WBP}{\langle NBP - WBP \rangle}, \tag{5.21}$$

where $M$ is the number of samples in the integration window (20ms is common for GPS L1/CA signals and for Galileo) and $\langle . \rangle$ is the average over 60s intervals.

Filtering techniques, other than the sixth order Butterworth filter, claim to give better estimation of $S_4$ [164, 165], and thus $S_4$ estimation changes with the detrending method in use [164, 166]. Furthermore, detrending the signal intensity is essential for Low Earth Orbit (LEO) satellites but it is not mandatory for MEO satellites, such as GPS and Galileo (as of today), where the slow power variation is not apparent over short time windows. Nevertheless, most GNSS scintillation monitoring receivers choose to implement the filter.

In short, detrending the signal intensity is implemented by most scintillation monitoring receivers, using different techniques, with arguable benefits and known disadvantages in the case of MEO GNSS signals.

### Noise Removal from $SI$

As discussed and shown in (2.9), the signal intensity value incorporates a noise term that scintillation monitoring receivers choose to counter in various ways. Some receivers eliminate the receiver noise from (2.9) before applying the detrending filter (e.g. [167]). Other receivers choose to remove the receiver noise by subtracting the predicted $S_4$ due to noise factor from the standard deviation in $SI$ (e.g. [36]). Some receivers implement neither. In the following paragraphs we show how $S_4$ is finally calculated.

**Estimation of $S_4$**

$S_4$ over observation window $T_w$ is given by [168]:

$$S_4 = \sqrt{\frac{\langle SI^2\rangle - \langle SI\rangle^2}{\langle SI\rangle^2}}, \tag{5.22}$$

where $\langle . \rangle$ is the average over $T_w$ and $SI$ is the detrended signal intensity. $S_4$ corresponds then to the amplitude scintillation index of the previous time window. $S_4$ can also be calculated, taking into account the effect of the ambient noise (if not already removed from $SI$), as [36]:

$$S_4 = \sqrt{\frac{\langle SI^2\rangle - \langle SI\rangle^2}{\langle SI\rangle^2} - \frac{100}{C/N_0}[1 + \frac{500}{19C/N_0}]} \tag{5.23}$$

where $C/N_0$ is the carrier to noise ratio estimated by the receiver, and the overall term $\frac{100}{C/N_0}[1 + \frac{500}{19C/N_0}]$ corresponds to the noise jitter variance [169, 170].

A good comparison of the calculated $S_4$ by a selection of ISM receivers can be found in [171]. Here, we just highlight that estimating the noise free signal intensity and detrending it are not trivial. The type of receiver in use controls how $S_4$ was calculated. Finally, if $I$ and $Q$ are available, an estimation of intensity of the fluctuation in the received GNSS signal is obtainable.

In this thesis, and with the use ofML, we investigate if it is possible to detect scintillation, knowing that all these effects are present in our measurements.

### 5.5.2 Phase Scintillation Index

For GNSS-based measurements, and radio wave measurements in general, the phase scintillation index $\sigma_\phi$ is defined as the variance of the detrended phase. It is computed from:

$$\sigma_\phi^2 = \int_{f_c}^{\infty} P_\phi(f)\,df \approx \frac{2T}{(p-1)f_c^{p-1}} =< \phi^2 > - < \phi >^2, \tag{5.24}$$

where $f_c$ is the lowest freq admissible by the system, $P_\phi$ is the phase spectrum, $T$ is the phase spectral power, $P$ is the spectral index, $\phi$ is the raw phase, and $< . >$ is the detrending operation. This detrending and the previously defined $f_c$ have been a topic of debate among the scientific community for a long time specially for high latitude scintillations (see [172] and [173]). It is not the objective of this work to investigate detrending, so we will stick to the method most used in the literature because this is how the majority of historic scintillation data are presented.

ISM receivers implement the phase detrending in various ways. For example, the receiver in [167] implements a polynomial fits to remove the high frequency noise, and uses a third order polynomial subtraction to filter out the low-frequency phase variations. Then, it applies (5.24) on a 100 s cadence to estimate $\sigma_\phi$. However, the most common way to detrend $\sigma_\phi$ uses a sixth-order high-pass digital Butterworth filter with 0.1 Hz cutoff frequency. Then estimating $\sigma_\phi$ from (5.24) on a 60s cadence.

## 5.6 Scintillation Signal Processing Outputs

The receiver deals with the GNSS signal in two main stages: acquisition and tracking. In acquisition, the receiver objective is to detect if the signal is available and if yes, it roughly estimates the delay and Doppler of the signal. Then the acquired signals are further processed in the tracking stage where the fine delay and phase of the signal are estimated.

If the scintillated signal is too weak, the acquisition stage will fail to detect the signal and thus it will fail to acquire the signal. During the tracking stage, if the signal to noise ratio of the scintillated signal is below a certain threshold, the receiver considers the signal too weak to process and quits processing the signal. Otherwise, if the phase of the signal is fluctuating beyond the tracking ability of the receiver processing power (i.e the tracking loop bandwidth) the receiver considers this signal too dynamic and quits processing the signal.

Scintillation monitoring GNSS receivers are carefully designed to handle the dynamic, noisy and weak scintillation signals in order to guarantee acquiring the signal and also to avoid the receiver dropping processing the signal. When the receiver completely fails to keep up with the signal after all, this is called acquisition failure and tracking loss, it means the loss of valuable scintillation measurements.

The main measurements of an ionospheric scintillation monitoring receiver are obtained from the outputs of the tracking stage. The immediate outputs of the tracking are the signal amplitude (both the in–phase and quadrature–phase components) and the accumulated phase. These are called the raw measurements, and from there the amplitude and phase scintillation indexes are obtained. Since we mentioned earlier that the design of the receiver (and implicitly the tracking ability of the receiver) affect the measurements, these receiver effects on the indexes computation have to be calibrated.

The pseudorange is estimated by further processing the raw measurements in addition to decoding the navigation message and other information e.g the known receiver position, tropospheric corrections, etc. From this pseudorange measurements the other ionospheric measurements can be obtained, i.e. TEC.

### 5.6.1 Example of a Scintillation Event

In this section we present an example of a scintillation event, and how the GNSS receiver acquires and tracks the signal under this condition.

The data we are using were collected by a software defined radio receiver (SDR) installed in parallel to a professional ISM receiver in Antarctica (at SANAE IV station). The two receivers share the same antenna and they were deployed in the frame of the DemoGRAPE project [33]. The SDR records the GNSS spectrum, down-converted (from GHz) to an intermediate frequency (few MHz). This is called the raw signal and it contains the actual GNSS spectrum in digital format without any signal processing.

At high latitudes, the ionospheric region within the auroral oval is often perturbed, giving rise also to scintillation events. The oval represents the footprint of the boundaries between open and closed geomagnetic field lines. The oval location indicates the zone where electron irregularities are present with electron fluxes energies varying in a wide range of $\sim 0.1$ to $\sim 20$ keV in the upper atmosphere [174]. Thus, scintillation is more likely to occur when the satellite signals pass through the auroral ionosphere.

The data were recorded on 27 September 2020 at 22:31:22 UTC and lasted for 50 minutes. The scintillation monitoring receiver as well as other instruments confirm that scintillation likely affected some of the GNSS signals at that moment. We report in Figure 5.9 the Dst index (which is an index of magnetic activity derived from near-equatorial geomagnetic observations) for the whole month of September 2020

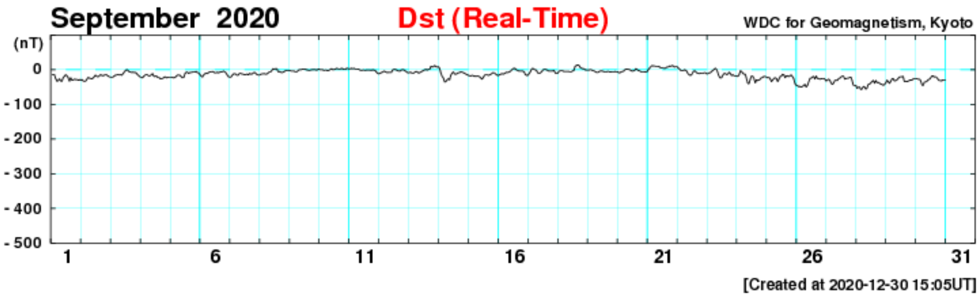


Fig. 5.9 Dst Index for September 2020.

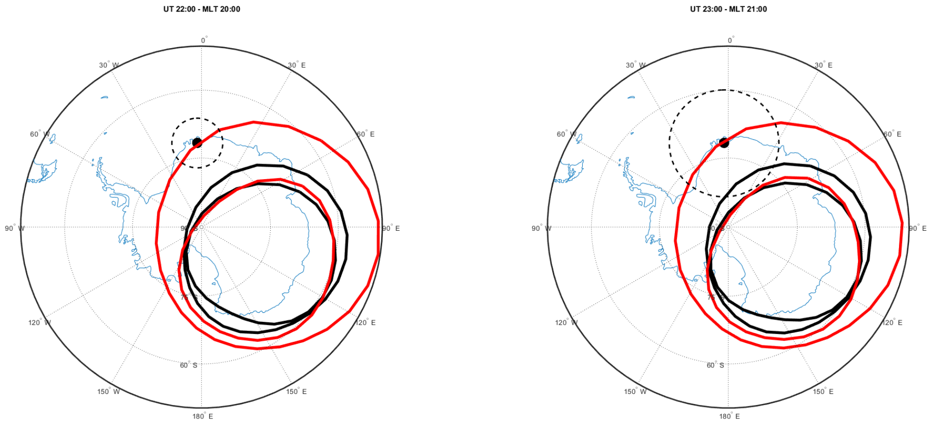
(Figure source [175]). It can be seen that the index went negative at the end of September, 27th indicating a geomagnetic storm at that time.

We report in Figure 5.10a and Figure 5.10b the Antarctic auroral oval as predicted by Feldstein model at 22:00 UTC. The oval between the two solid black circles indicate the oval when the disturbance level is weak, while the one between the two red circles indicate the oval under strong disturbance. The dotted black circles in Figure 5.10a and 5.10b indicate the visible sky from SANAE IV station with elevation masks 30° and 10°, respectively. Therefore, when the auroral oval overlaps within this dotted circle, scintillation could be observed by the receiver. But also inside the inner auroral boundary, the polar cap can host scintillations.

To complete the picture, we show in Figure 5.10c the skyplot reporting the visible GPS satellites at the time of our measurements. It can be seen that on disturbed days, a good part of the oval is well within the visibility of SANAE station specially if the elevation mask is lowered to 15°. It is important to mention here that (1) the Feldstein oval underestimates strong disturbance conditions [176], and (2) the description of the oval is climatological, while the case we are considering is a single event.

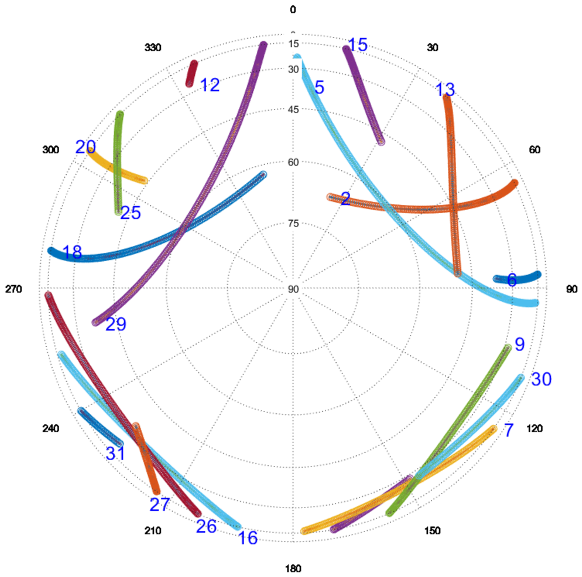
The phase and amplitude scintillation indexes measured by the ISM receiver are shown in Figure 5.11. The figure report multi–GNSS $\sigma_\phi$ (radians) and $S_4$ at L1 (top) and $\sigma_\phi$ (radians) and $S_4$ at L2/E5a (bottom) for the whole day. The different colours indicate different satellite signals. It can be seen that strong phase scintillation (up to 1 rad) was observed on many satellite signals from 20:00 UTC until midnight.

Next, we will select some of the scintillated signals and see how the receiver processes them.



(a) Auroral Oval. Elevation mask 30°

(b) Auroral Oval. Elevation mask 15°



(c) Visible Satellites. Elevation mask 10°

Fig. 5.10 Feldstein Auroral Oval and the visible GPS satellites from SANAE IV station on 27 Sept 2020 22:00 UTC . Elevation mask 30° and 15° for Feldstein ovals, and 15° for the visible satellites

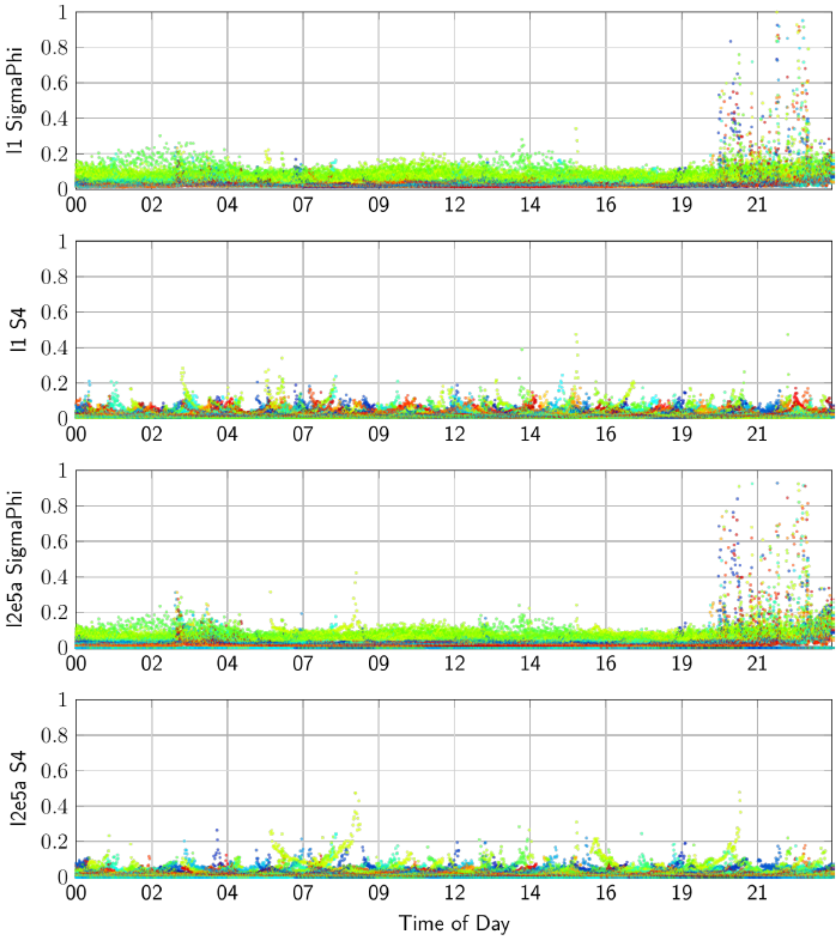


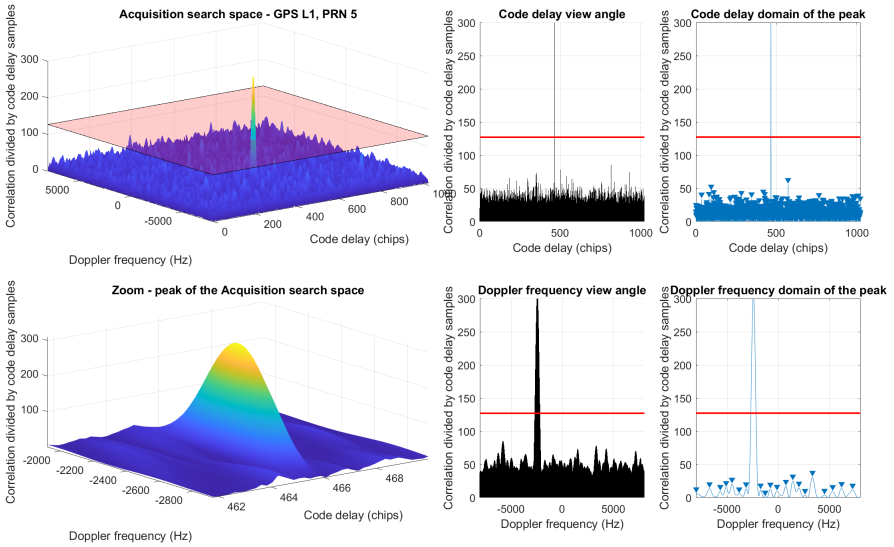
Fig. 5.11 $\sigma_\phi$ (radians) and $S_4$ for multi–GNSS signals measured by Septentrio PolaRxS professional ISM receiver at L1 and L2/E5a on 27 Sept 2020 at SANAE IV Antarctic station. The different colours indicate different satellite signals.

### 5.6.2 Acquisition Stage Outputs: CAF

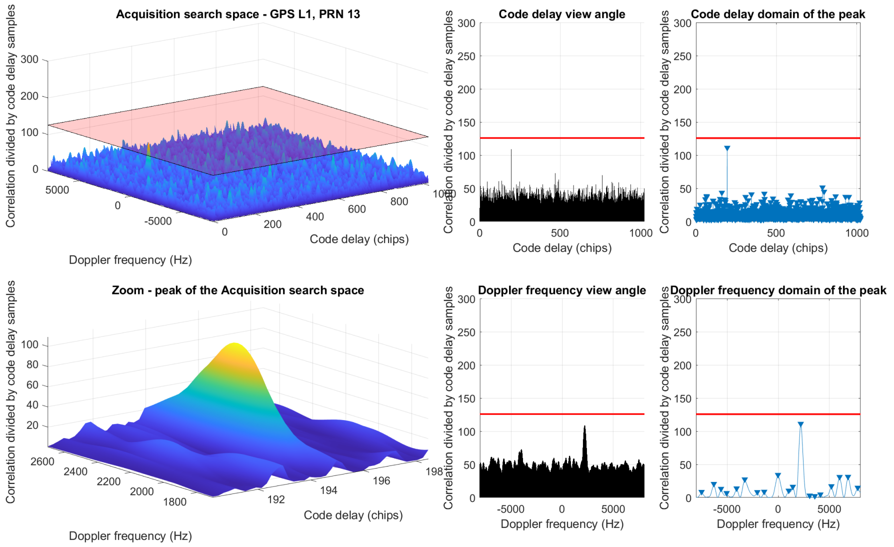
The raw signal recorded by the SDR is the output of the front end, down converted to an intermediate frequency and digitized. It contains the sum of all the GNSS (and possibly non-GNSS) signals in the working frequency of the front-end. Being pseudo-random code-modulated, the spectrum of the GNSS band is basically noise. To extract information about any of the GNSS satellite signals, we need to decode the satellite signal, i.e. cross-correlate the signal with the satellite pseudo random code. For this reason, we will directly move to inspecting a certain satellite and not the whole spectrum (the latter can be useful for studying RFI for example).

The tool we use for this inspection is the standard cross ambiguity function (CAF) of the signal acquisition theory plotted in 3D delay doppler map representation. Here, the spectrum is cross-correlated with all the possible pairs of code delays and carrier frequencies. In other words, we calculate the amplitude of the complex signal in (2.9) for all the possible sets of $\hat{\phi}_i, \hat{\tau}_i$. In case of a perfect signal, the result should be a peak power for the pair $\delta\hat{\phi}_i = 0, \delta\hat{\tau}_i$ and much smaller values for noise. Since the scintillated signal is self interfering, we do not expect to see a peak as strong as the ideal case and in strong scintillation cases, we expect the signal peak to be at the level of the noise floor.

In Figure 5.12a and 5.12b we show the CAF for GPS C/A PRN05 (not affected by scintillations) and PRN13 (scintillating), respectively, evaluated at 22:40 UTC. We evaluated the CAF with coherent integration time 1, 3 and 5 ms and in this figure we are reporting only the case of 3 ms which had better CAF than n=1 (similar result to other works in the literature, e.g [177]. We also made sure we are not reporting the CAF at a data bit flip by sliding the evaluation within 20 ms of data and selecting the best CAF. The top left figure shows the CAF in 3D while the bottom left figure shows a zoom view to the peak of this CAF. The red plane is the acquisition threshold estimated by the receiver. The middle top and bottom plots show side views of the 3D CAF. Finally, the right top and bottom plots show the code and doppler domains of the CAF peak. For the signal not affected by scintillations, the peak of the CAF is well separated from the noise floor while for the scintillated signal the peak is below the threshold and there is a chance that this peak is just noise. In the latter case, if the receiver tries to acquire the signal at this moment, it will declare PRN13 not present.



(a) PRN05 non scintillating signal



(b) PRN13 scintillating signal

Fig. 5.12 Examples of the CAF for non-scintillated and scintillated signals

For scintillation studies, although the raw signal carries all the information about the scintillation that affected the signal, but almost zero studies tried to utilize the signal at this stage of the receiver to study scintillation, for many reasons. First, the GNSS signal before removing the PRN code, is basically undistinguishable from noise, and this is one of the features of CDMA code modulation. Second, and most importantly, processing the GNSS signal with the CAF is an expensive mathematical operation requiring evaluation of the cross-correlation millions of times, while in the next stages of the receiver, only 3 or 5 correlations (depending on the number of early-late correlators) are performed. For example, in Figure 5.12b the data was sampled at 13MHz. We implemented 75 Hz Doppler steps covering ±8 KHz Doppler range. Thus, approximately 1.7 billion cross correlations were evaluated per each 1ms of data to obtain the CAF. This is computationally and memory consuming operation, that provides a detected peak way less accurate compared to the obtainable peak in the next stage of the receiver.

Finally, the availability of data is another limiting factor for working with the acquisition results. To the best of our knowledge, no ISM receiver is providing CAF as a ready product at the moment (for example most reflectometry receivers discussed in Part I provide DDM/CAF) as a product. Although many scintillation repositories are available, they are not as abundant as the other scintillation data as will be seen in the next paragraphs. Consequently, we will not further process the GNSS signal at this stage of the receiver.

### 5.6.3 Tracking Stage Outputs: Raw Measurements

This stage of the receiver fine estimates the delay and phase of the GNSS signal. It outputs the accumulated phase measurement, besides the I and Q components of the complex signal as described by (2.9). The typical rate for I and Q depends on the GNSS signal type, for example for GPS L1 C/A the rate is 50Hz and 20ms coherent integration is usually implemented. Nevertheless, higher rate raw measurements are also available. Besides the low demand for computation with respect to acquisition, these measurements occupy significantly less storage than the raw signal or CAF measurements, since only three measurements are recorded for each epoch.

Phase scintillation index is estimated directly from the accumulated phase (60 s of phase measurements is the most common), while the amplitude scintillation index is calculated from I and Q (also usually at 1 minute cadence). The delay and pseudorange are estimated from I & Q in the subsequent PNT unit.

I, Q and the phase are called the raw measurements of the receiver. They provide the best estimate for the correlation peak. Their accuracy depends on the capability of the tracking loop to keep up with the signal dynamics. Thus, these measurements have an extra noise term that reflects on the tracking performance.

In Chapter 8, we will focus on these measurement and apply ML techniques to detect scintillation using raw measurements.

### 5.6.4 Scintillation Engine Outputs: Scintillation Metrics

This stage is present in ISM receivers. It is responsible of computing the scintillation indexes as well as other related ionospheric parameters like TEC. Most of the literature on the utilization of GNSS measurements for scintillation studies use data produced by this unit. The indexes provided by this stage are the same indexes that have been in use among the ionospheric monitoring scientific community and thus it forms a measurement compatible with measurement from other historic instruments.

However, since the design of the receiver and the signal processing implemented affects the measurements estimation, as discussed in Section 5.5, the use of these indexes is not standard. For example, in [171] the authors compared the scintillation indexes reported by various ISM receivers that were simultaneously monitoring the same scintillation event and found that they indeed report different values for the index. This is directly related to the receiver hardware design and the signal processing implemented. Since not all of the design details of ISM receivers are available to the end user, specially the commercial ones, then treating the metrics for generalized applications is challenging. This makes large scale scintillation studies rely on ISM receivers provided by the same monitoring network or the same manufacturer.

Above that, ISM receivers scintillation metrics have been criticized in the literature for raising false scintillation alarms [61, 178]. In particular, the amplitude scintillation index behaves the same when scintillation as well as multipath occur. This is expected from signal point of view, because multipath resulting from the signal rays following different paths caused by reflections and bouncing from objects around the receiver and then adding up constructively and destructively at the receiver antenna, is the same behaviour of scintillation where signal also follows different paths due to diffraction that happened up in the ionosphere. On the other hand, the phase scintillation index is criticized for incorporating both the refractive and diffractive effects of the ionosphere if it was not detrended properly [172]. Finally, since most of the accumulated scintillation measurements metrics data suffer from both the previous effects, any attempt to propose new techniques should be able to compensate these effects or better, utilize these data in spite of their known defects. The reason for this is, despite of all these defects, they are still valuable data collected over more than two decades of measurements campaigns.

In Chapter 7, we use the scintillation measurements provided by a Septentrio PolaRxS ISM receiver. In Appendix A, we list the measurements provided in the ISM records (i.e. the scintillation engine outputs) which include other scintillation measurements beside the indexes, evaluated for different GNSS frequencies.

### 5.6.5 PNT Unit Outputs: Pseudorange Measurements

These are the measurements provided by the positioning, navigation and timing unit available in all types of GNSS receivers (because these are the measurements needed to obtain the positioning solution, the main utilization of the GNSS systems). Although ISM receivers do not need to estimate their positions, they still provide the pseudorange to the end user. This is useful to estimate the error in estimating the receiver position from the scintillation–affected signals and compare it with the known position of the ISM receiver. The main usage for these measurements in ionospheric studies is calculating TEC, which utilizes range measurements from different frequencies as described in Section 5.4.1.

The biggest advantage of these measurements is that they are provided by all GNSS receivers. The challenge of utilizing these measurements for scintillation monitoring lies in the low quality of these observables when non ISM receives are utilized. Since the effects of the receiver has to be removed (or estimated) from the measurements before calculating the scintillation indexes, the wide utilization of these measurements from non ISM receivers is challenging. Nevertheless, many studies have conducted experiments on utilizing these measurements from multi-frequency receivers including professional geodetic receivers [179, 180] and mass market smartphones [181, 182].

The move towards utilizing these measurements for scintillation studies will provide wider and denser geographic distribution of GNSS-based scintillation measurements unprecedented by any of the professional ISM receivers networks. They are also less expensive than ISM receivers which, for the latter reason, are mainly installed around the magnetic poles and equator where the demand for scintillation measurements is the highest.

In Chapter 7 we utilize the TEC values provided by professional ISM receiver for scintillation detection. We do not consider TEC provided by non-ISM receivers at the moment because (1) TEC from ISM receivers are abundant already and deserve thorough investigation as potential features for detecting scintillations using ML techniques, and (2) if TEC from non-ISM receivers are to be considered for scintillation detection, we need first to co-locate such a receiver with a reference ISM receiver and compare the TEC values provided both receivers. This is beyond the objective of this investigation.

## 5.7 Scintillation Detection

Now that we have introduced the scintillation phenomena, and how to utilize GNSS for observing the ionosphere, we discuss how scintillation is actually detected in the collected remote sensing measurements.

In ionospheric scintillation remote sensing, the focus of this work, the users (ionospheric scientist) are usually interested in extracting data that are affected by scintillation among possibly huge collection of GNSS data. To do this, they need to rely on detection mechanisms that are able to identify scintillation. This motivates the need for having a reliable scintillation detection method.

Moreover, on the side of designing monitoring stations, it might be required to collect measurements only when scintillation events happen. This is important specially for scintillation monitoring stations that collect the raw GNSS samples (see Section 5.6) and also to monitoring stations with limited data storage and transfer capabilities (e.g remote monitoring stations). In these stations, a reliable scintillation detection method is needed to lower the scintillation miss-detection rate, that leads to loss of valuable scintillation measurements, and also to lower the false scintillation alarm (that leads to excess use of the storage and communication resources mentioned earlier).

Furthermore, for real time scintillation monitoring, it is important to rely on detection schemes that are able to reliably raise scintillation alarms without false notifications. Real-time monitoring is critical for alerting high end GNSS users of scintillation risks.

Finally, as scintillation monitoring has been going on for decades, huge repositories of data has been accumulated over the years. This accumulation enables studies that analyse data with long time span, but also brings the need for data mining techniques that are able to handle such data. This motivate the use of data mining techniques and artificial intelligence to reliably extract information from big data.

### 5.7.1 Scintillation Indexes Threshold

Historically, the method for scintillation detection relies on thresholds on the scintillation index. In this method, scintillation is declared present if the scintillation index exceeds a predefined threshold. The thresholds on the indexes are not fixed among the scientific community with various studies adopting different threshold on the amplitude and phase scintillation indexes for the different scintillation monitoring receivers and also for the different scintillation severity. Acceptable threshold on $S_4$ and $\sigma_\phi$ can be defined as [61]:

$$\begin{cases} S_4 > 0.4 & \text{strong scintillation} \\ S_4 > 0.2 \text{ and } < 0.4 & \text{moderate scintillation} \\ S_4 < 0.2 & \text{weak-to-no scintillation} \end{cases} \tag{5.25}$$

$$\begin{cases} \sigma_\phi > 0.25 & \text{strong scintillation} \\ \sigma_\phi > 0.15 \text{ and } < 0.25 & \text{moderate scintillation} \\ \sigma_\phi < 0.15 & \text{weak-to-no scintillation} \end{cases} \tag{5.26}$$

However, the GNSS amplitude and phase measurements that are the base for scintillation metrics, are affected not only by scintillation, but the whole GNSS propagation environment (multipath and interference), the receiver clock anomalies, and the satellite clock anomalies. The latter effects, if not counted for, can lead to contamination of the scintillation data extracted to be processed for further ionospheric studies.

By expert visual inspection, and consulting GNSS data from different days and satellites, scintillation can be discriminated from the other impairments affecting the scintillation metric as follows. We show next how these impairments can be detected using the scintillation indexes, and thus we show how they can be discriminated from scintillation.

### Multipath Detection Using Scintillation Indexes

In a fixed scintillation monitoring receiver, multipath can be detected by observing repetitive scintillation indexes inflation that have a geospatial pattern. If $S_4$ for a certain satellite signal inflates every time the satellite is at a certain azimuth and elevation with respect to the receiver, we suspect that the signal fluctuates due an obstacle in the environment around the receiver. Scintillation is random by definition and, although we expect to observe it at certain ionospheric conditions, we do not expect the repetitive nature at the same location and time on consecutive days.

Thus, to distinguish if the cause beneath the scintillation index inflation is multipath, we can compare the indexes for several days and observe when the scintillation index is inflated. in Chapter 8 we show an example of multipath like $S_4$ inflation.

### Receiver Oscillator Errors Detection

It happens that the oscillators (on the receivers and the satellites) experience noise for short periods. These result in generating noisy carriers at the receiver and thus the carrier removal operation of the receiver leaves a residual phase. This residual phase translates to inflation in the $\sigma_\phi$ as observed by an ISM receiver.

Since it happens at the receiver oscillator, this phase noise affects the processing of all the GNSS signals in view. Thus, it can be detected by observing $\sigma_\phi$ of all the processed signals, and if phase inflations occurred at all the signals at the same time, then this is declared as receiver oscillator anomaly.

### Satellite Clock Errors Detection

Contrary to the receiver oscillator anomalies, the satellite oscillator anomalies signature is that a phase index inflation is observed at all the signals of a certain satellite with the same magnitude, but not all the visible satellites at the same time. To distinguish it from scintillation, as discussed in Section 5.4, ionospheric effects on the phase are frequency dependent, and thus they do not affect similarly the signals of a satellite at multiple frequencies. Thus by consulting $\sigma_\phi$ from a satellite at multiple frequencies, and making sure that it is not a receiver clock anomaly, satellite clock errors are detectable.

In summary, to detect the anomalies in the GNSS scintillation measurements, we compare the scintillation indexes from:

- Several consecutive days,
- All the visible satellites, and
- Multiple GNSS frequencies.

Although this yields accurate scintillation detection, it is not feasible for neither real time application, nor large amounts of data. At the same time, a simple threshold on the scintillation index is not reliable.

In particular, a threshold on the indexes leads to loss in identifying scintillation at the edges of the index inflation, i.e. at the beginning and end of the index disturbance, and when the signal experiences irregularities that cause strong then weak then strong etc scintillations.

Nevertheless, the scintillation index threshold can be beneficial as a rough detector for scintillation events. For example, we can use a simple threshold to quickly identify all the data with potential scintillation measurements. Then, further processing can be done to identify if scintillation was the cause behind the index inflation, by visual inspection (as shown in Figure 5.13) or by implementing a mechanism able to distinguish scintillation from the anomalies as was implemented in [183] and [184] (and the recent improved version of the latter [185]).

These limitations in the traditional metrics lead to investigating alternative methods for scintillation detection (other than the simple threshold on the metrics) and also to investigating other metrics for scintillation detection.

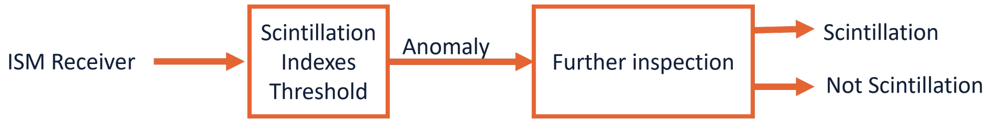


Fig. 5.13 Scintillation detection using the indexes

### 5.7.2 Other Detection Mechanisms

ISM receivers rely on elevation mask, beside multipath rejecting antennas, to eliminate all the potential multipath contaminated measurements, accepting the loss of a good portion of the sky visibility.

In [186], the authors proposed to overcome this by implementing a smart elevation mask where only portions of the sky with known multipath are masked. 35-45% gain in the collected scintillation data was demonstrated. However, such techniques requires surveying the station area and generating a mask for each installed station. Moreover, like the consecutive days data check, this mask is not sensitive to mobile object and changes to the environment around the receiver.

Another approach was implemented in [187], where multi measurements were consulted to eliminate multipath. By combining the elevation angle, signal to noise ratio, with the scintillation index. Similar to the elevation mask, this method reduced the visible sky even more than the original threshold on the elevation angle. The gain of this method is the low false scintillation alarm rate.

Then came the Artificial Intelligence (AI) and ML approach. With the success demonstrated by the various ML models in various fields of science including GNSS signals processing and also scintillation studies, these techniques found their way to scintillation detection as well.

ML techniques has demonstrated superior performance to human in the fields of classification and anomaly detection in general. For example, in classifying images and detecting malware. In the last two decades, with the advancements in computation power and in the increased availability of data in all fields of science, ML has found applications in almost every field of life.

In GNSS receivers in general, ML has been utilized for [188] multipath detection [189], shadow matching [190], GNSS–R [191], RFI source classification [192], spoofing detection [193], and scintillation forecasting [194, 195] and detection [196–201, 185].

Finally, focusing on the topic of this thesis, scintillation detection, ML has been applied to detect both amplitude and phase scintillation, in order to overcome the limitations of the techniques discussed earlier in this Section.

In Chapter 6, we briefly introduce ML, and how it is utilized for scintillation detection.

In conclusion, the need for reliable scintillation detection techniques able to handle huge repositories of data is needed to enable mining the repositories of scintillation data that are available to the scientific community today. This, combined with the arguable reliability of the scintillation detection based on a threshold on the traditional scintillation indexes, demands the search for alternative detection methods compatible with the available data today. In particular, the ML techniques have gained interest in the last decade due to their success in the many GNSS and non-GNSS fields, ionospheric science in general, and GNSS-based scintillation detection in particular. For this, we focus on the latter and investigate phase and amplitude scintillation detection using ML models.

# Chapter 6

# A Brief Introduction to Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that includes techniques and algorithms which learn from data and improve from experience [202]. AI, the big umbrella under which ML lies, in itself is a science and technology discipline that incorporates a collection of techniques that emulate learning, reasoning and self-correction (thus the name intelligence) to develop solutions for computers, robotics and machines in general (thus the term artificial). AI consists of a wide range of sub-categories including computer vision, robotics, natural language processing, and ML. Being related to data discovery, AI techniques have been closely linked to Data Mining (DM) and Data Science (DS) in general. Figure 6.1 (adapted from [203]) illustrates the relationship between AI, ML, DM and DS. We included in this illustration also Deep Learning (DL), a subset of ML techniques that utilize Neural Network (NN) architectures and has been achieving exceptional results in the last few years, leveraging the advanced affordable computing resources and the increasing availability of huge amounts of data.

ML has gained success in a wide range of applications including remote sensing, telecommunications and scientific fields. Their ability to learn without prior knowledge about the underlying physical phenomenon attracted their use in science and engineering specially when it is complex to model the physics that generated the data. A ML model can discover patterns in the data, or predict values of some variables based on measurements and previous data.

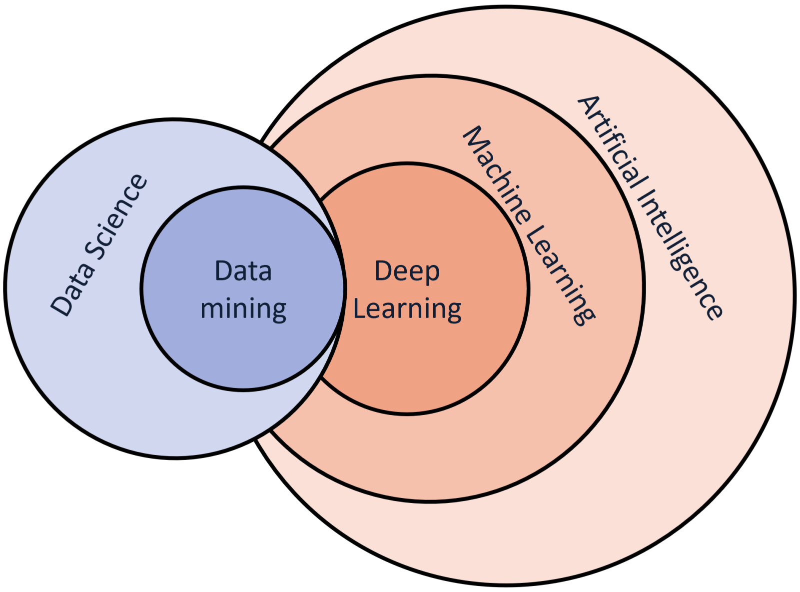


Fig. 6.1 The relationship between Data Science (DS), Data Mining (DM), Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL).

In this Section, we briefly introduce the concept of ML and discuss the ML algorithms that are used in this thesis.

## 6.1 Understanding the ML Task

When ML is utilized, the aim is to obtain a model that learns from the data how to perform the task, for example how to detect scintillation. An important task to obtain the model is to perform the model training, which requires expertise in the application domain, ionospheric sciences for example. Then, this model needs to be validated, similar to any model used in the scientific community. The outcome of the ML training process is a data driven model that predicts the future outcomes from the input data. For the scintillation detection task depicted in Figure 5.13, ML replaces the processing flow into the new form depicted in Figure 6.2 where we replace the threshold and expert inspection of the data with one ML model block that takes the ISM receiver outputs and predicts if scintillation occurred. The methodology to obtain this model is explained in Section 6.2.

There are many ways to categorize ML techniques. Based on the availability of the input data at the time of training the model, the learning can be offline, online

Fig. 6.2 Scintillation detection using ML models

or active as illustrated in Figure 6.3a (adapted from [203–205]). Here, the data that the model needs might not be available before hand, for example in online advertisements, where data arrive as the user is browsing a website. For our study, we will be using repositories of scintillation data, and we will extract the input features beforehand. Thus we will focus only on offline learning.

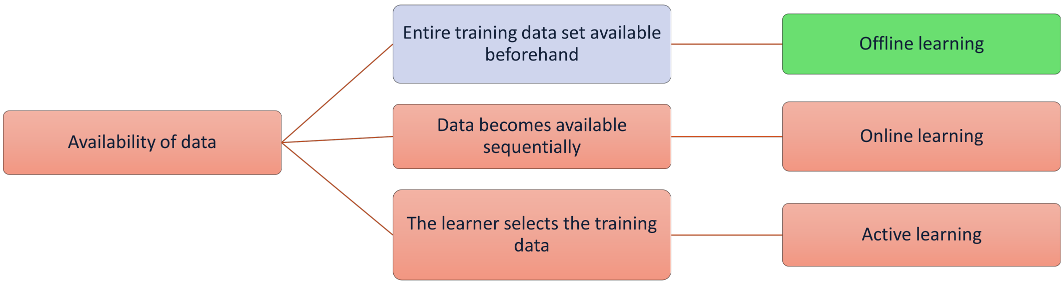
Moreover, depending on how much labelled data are provided to the ML algorithm, ML techniques can be divided into supervised, semi-supervised, reinforcement and unsupervised learning as illustrated in Figure 6.3b (adapted from [203–205]). In supervised learning (see Section 6.3), we inform the ML algorithm about the target outputs we expect the algorithm to predict. Thus, the model fits the data somehow to obtain the desired output. Supervised ML predicts the class corresponding to the input data if the output is categorical data, or performs regression if it has to predict a numerical value.

Unsupervised learning on the other hand expects the ML model to explore the data and understand how these data are related or how they can be grouped together. In unsupervised ML, the task of the model is either clustering the data that are similar, or reducing dimensionality and thus indicating which data are redundant to each other.

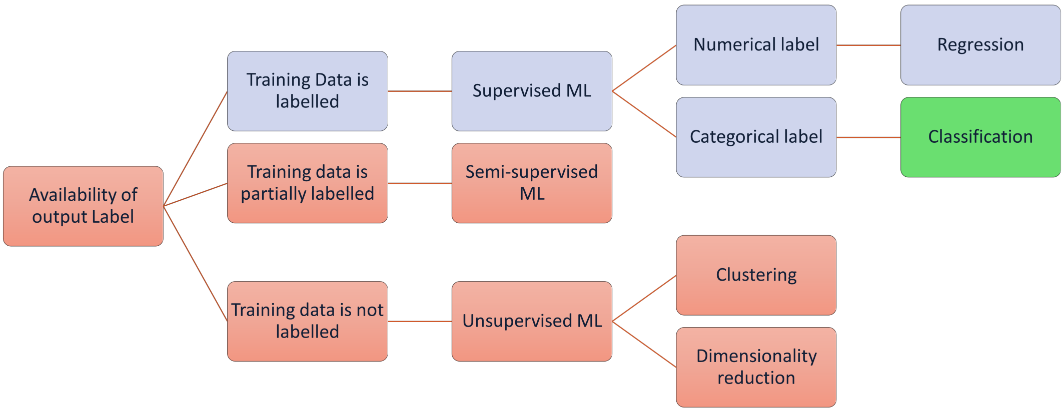
In between supervised and unsupervised ML, semi-supervised learning. Here, the label of some of the data is made available to the algorithm. Thus the algorithm receives a priori guidelines on what clusters it is expected to find.

## 6.2 Machine Learning Modelling Cycle

To develop a ML model, certain steps has to be followed before the model is declared operational. These steps are illustrated in Figure 6.4 (adapted from [203]) and explained in the following subsections.



(a) based on time of data availability



(b) Based on output label availability

Fig. 6.3 Machine Learning Model selection.

### 6.2.1 Feature Engineering

Feature engineering, also known as feature discovery and feature extraction, is the process of selecting, transofrming and manipulating features, also knows as attributes, from the raw data using domain knowledge in order to improve the learning of the models (the data is thus represented by feature vectors $x \in R^n$). Components of Feature Engineering include:

- Data cleaning, where corrupted samples and outliers are removed from the data set. This step is essential to ensure that the model will not be trained on corrupted data. Since the ML model is data-driven model, the quality of the model relies on the quality of the data set. For example in GNSS scintillation data, cycle slips and oscilator anomalies might be considered as outliers and thus removed from the data. However, they can also be considered as valuable inputs if the objective of the model is to detect outliers in the measurements. A careful definition of the objective of the model and the problem statement is stressed here because of its importance in developing a model able to accomplish the task. In this thesis, since we are processing data off-line from repositories of GNSS scintillation, we consider cycle slips and oscillator anomalies as outliers and we remove them from the datasets. For future implementations we recommend considering them as valuable data and train the model to recognize them. The latter is important for implementing the detection mechanism in real-time scintillation monitoring receivers where the detection model should be able to handle this type of outliers.
- Data normalization to limit the range that the data set values can take. This is important for some ML algorithms, like the k-Nearest Neighbours (k-NN) (see Section 6.3.2), and not essential for other algorithms, like the trees (see Section 6.3.3). The k-NN in particular replies on evaluating the distance between samples in the feature sets domains and thus it is important to guarantee that no feature will dominate the evaluation of the distance because its values take wider range than the other features. Normalization is a good practice in ML generally and we adopt it in preparing all the data for ML modelling..
- Transformation, where the data are presented in other domains that best represents the data for example in frequency domain. The choice of the transformation required domain knowledge where experience on the transforms that

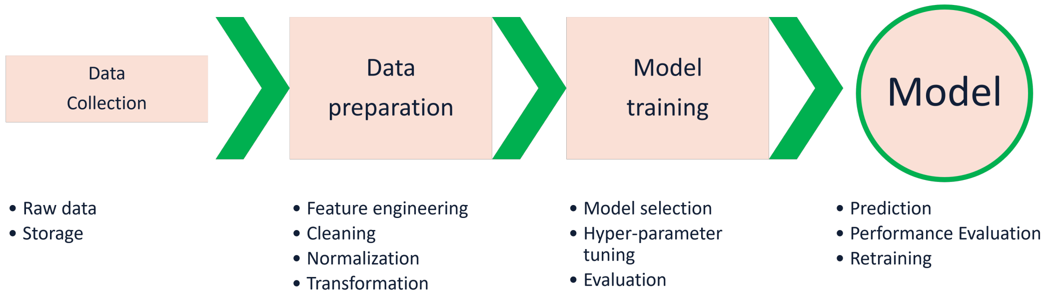


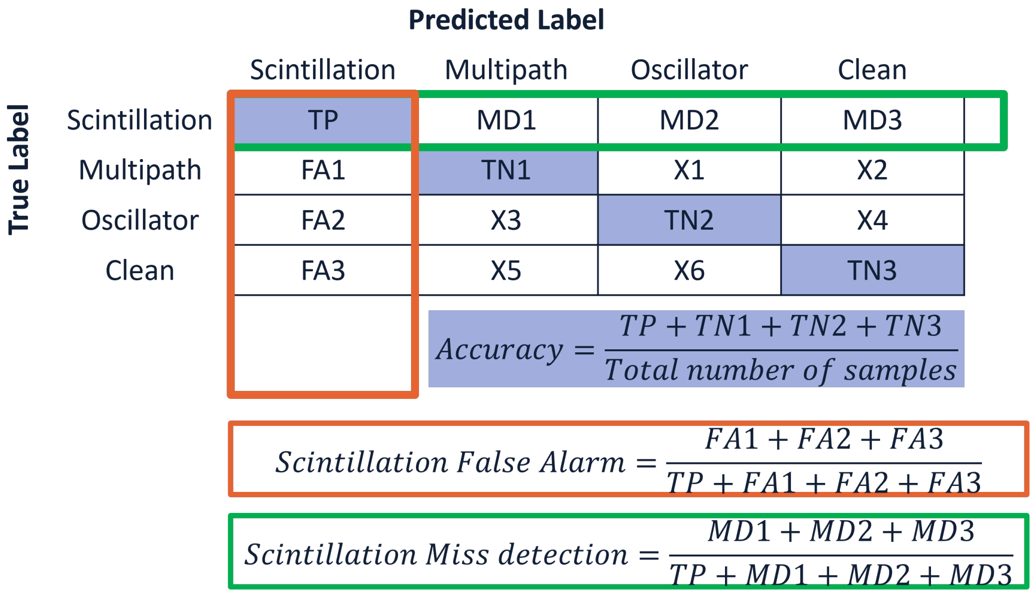
Fig. 6.4 Machine Learning Modelling Cycle.

best suit the data is needed. For our scintillation detection models, we did not implement any feature transformation and relied on the features as provided by the receivers.

### 6.2.2 Model Evaluation

Figure 6.5 shows the confusion matrix, a popular method to visualize ML models performance, for a multi-class classification model. Here we hypothize that the model gives 4 class labels (scintillation, multipath, oscillator anomaly and clean GNSS signal). Since our interest is detecting scintillation we consider the positive class to be scintillation, and the other three classes to be negative. We define the following terms:

- True positive (TP): the true label is positive (i.e. scintillation), and the model correctly predicts the label to be positive.
- True negative (TN): the true label is negative (i.e. no scintillation), and the model correctly predicts the label to be negative.
- False alarm (FA): the model wrongly predicts scintillation (i.e. raises scintillation alarm) but the true label is not positive.
- Miss detection (MD): the true label is positive but the model does not detect it (i.e. wrongly labels the data as negative).
- Don't care predictions (X): these are the miss classifications that we do not care about in this application. For example, when multipath is classified as oscillator anomaly or clean signal.



**Predicted Label**

**True Label**

| | Scintillation | Multipath | Oscillator | Clean |
|---|---|---|---|---|
| Scintillation | TP | MD1 | MD2 | MD3 |
| Multipath | FA1 | TN1 | X1 | X2 |
| Oscillator | FA2 | X3 | TN2 | X4 |
| Clean | FA3 | X5 | X6 | TN3 |

$$Accuracy = \frac{TP + TN1 + TN2 + TN3}{Total\ number\ of\ samples}$$

$$Scintillation\ False\ Alarm = \frac{FA1 + FA2 + FA3}{TP + FA1 + FA2 + FA3}$$

$$Scintillation\ Miss\ detection = \frac{MD1 + MD2 + MD3}{TP + MD1 + MD2 + MD3}$$

Fig. 6.5 confusion matrix for a multi-label scintillation detector

The accuracy, scintillation miss detection and false scintillation alarm metrics are defined as follows:

- Accuracy is the ratio of the number of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN1 + TN2 + TN3}{\text{Total number of samples}}$$

- Scintillation Miss detection rate is the ratio of the number of scintillation samples wrongly classified to the total number of scintillation samples.

$$\text{Scintillation miss detection} = \frac{MD1 + MD2 + MD3}{TP + MD1 + MD2 + MD3}$$

- False scintillation alarm is the ratio of the number of samples wrongly classified as scintillation to the total number of samples classified as scintillation.

$$\text{Scintillation false alarm rate} = \frac{FA1 + FA2 + FA3}{TP + FA1 + FA2 + FA3}$$

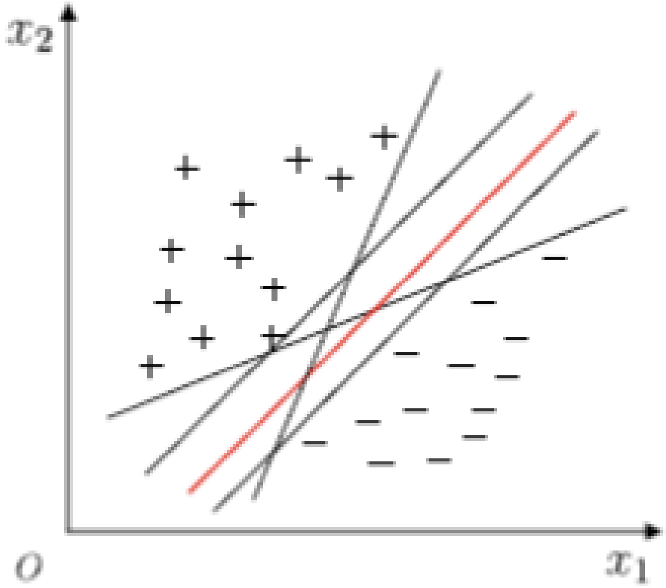


Fig. 6.6 Illustration of the hyperplanes separating two classes '+' and '-'.

## 6.3 Supervised Learning Overview

Overviews of the algorithms considered for scintillation detection (i.e. classification) in this work are given in this Section.

### 6.3.1 SVM Algorithm

Support Vector Machines (SVM) algorithm, one of the most influential supervised learning approach associated with kernel trick, aims to classify the samples using a separating hyperplane and it output a class identity [206].

The basic idea of SVM classification is to use the training set to learn the best separating hyperplane in the sample space that can separate samples of the different classes. Figure 6.6 (Figure source [207]) shows a 2–D illustration of the multiple hyperplanes that can separate the samples in the two classes denoted as '+' and '-'. The best separating hyperplane can be defined as the one with the most robust classification results strongest generalization ability [207].

A separating hyperplane $y = f(w,b)$ in the sample space can be expressed as the linear function [207]:

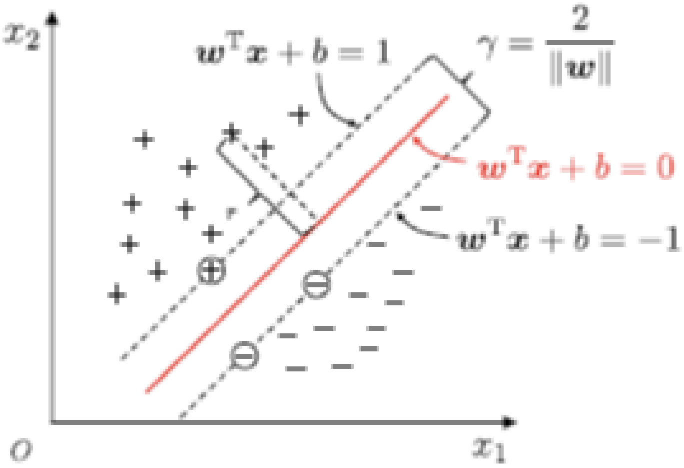


Fig. 6.7 Illustration of the hyperplanes separating two classes '+' and '-'.

$$y = f(w,b) = w^T x + b = 0, \tag{6.1}$$

where $w$ is the normal vector that controls the direction of the hyperplane, and $b$ (the bias) is the distance between the hyperplane and the origin.

For a hyperplane that can correctly classify the training samples, there exist $w^T x_i + b > 0$ when $y_i = +1$, and $w^T x_i + b < 0$ when $y_i = -1$. These hyperplane are called support vectors. Figure 6.7 (Figure source [207]) shows an example of SVM for two-class classification. The total distance between the two support vectors is called the margin $\gamma$. Finding the separating hyperplane with the maximum margin is equivalent to finding the parameters $w$ and $b$ that maximize $\gamma$.

If the training samples are not linearly separable in the sample space, then the samples $x_i$ can me mapped from the original feature space to a higher dimensional feature space, $\varphi(x)$, where the samples become linearly separable. In this case, maximizing $\gamma$ involves the calculation of $\varphi(x_i)^T \varphi(x_j)$. Since the mapped feature space can have high dimensionality, it can be computationally difficult to estimate $\varphi(x_i)^T \varphi(x_j)$ directly. This can be avoided by assuming that there exists a function in the following form:

$$\kappa(x_i,x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle = \varphi(x_i)^T \varphi(x_j), \tag{6.2}$$

i.e. the inner product of $x_i$ and $x_j$ in the feature space can be calculated in the sample space using the function $\kappa(.,.)$ [207].

Since the best kernel function for a certain dataset in unknown, the choice of the kernel is the greatest uncertainty in applying SVM. Some common kernel functions include the linear, the Gaussian and the polynomial kernels.

In cases where the samples are linearly separable, the linear kernel function can be used, defined as:

$$\kappa(x_i,x_j) = x_i^T x_j \tag{6.3}$$

where the kernel function $\kappa$ provides a mapping from the instance space to a feature space associated the kernel. Hence, it enables to find another hyperplane in the kernel space and to achieve non–linear separation in the feature space [208].

The most commonly used kernel is the Gaussian kernel, also known as Radial Basis Function (RBF), defined as:

$$\begin{aligned}\kappa(x_i,x_j) &= \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \\ &= \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right)\end{aligned} \tag{6.4}$$

where $\sigma > 0$ is the width of the kernel and it has to be properly selected. If it gets closer to zero, it might cause over-fitting. However, a bigger value of $\sigma$ might lead to under-fitting and ends up with classifying all the instances into one class [208]. Selection of kernel scale parameter $\gamma$ in the RBF has the similar issues as well.

Another well-known kernel function is the polynomial:

$$\kappa(x_i,x_j) = \left(1 + x_i^T x_j\right)^p \tag{6.5}$$

where $d \geq 1$ is the degree of the polynomial, and addition of 1 provides in-homogeneity.

Finally, The class predictions are made using the function [206]:

$$y = f(x) = b + \sum_{i=1}^{N_s} w_i\,\kappa(x,x^i) \tag{6.6}$$

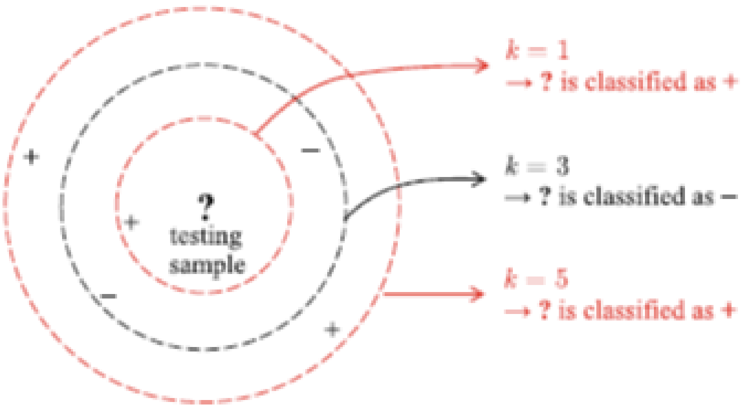


Fig. 6.8 k-NN classification concept and example 1.

where $f(x)$ describes the approximate relationship between input $x$ and corresponding target output values $y$. $x^i$ is a sample of training input data-set given $\left\{x^1, x^2, \cdots, x^{N_s}\right\}$ where $x^i \in \mathbb{R}^n$. $b$ is the parameter of the optimum hyperplane shown in Figure 6.7.

### 6.3.2 k-NN Algorithm

k-NN is a type of non-probabilities supervised algorithm and it is generally used for classification or regression.

Given a positive integer $k$ and a sample $x$, the classifier first identifies $k$ points in the training data closest to $x$. At test time, it calculates the distances between $x$ and all training data. The class is thus found by maximizing the number of neighbours with the same label. Figure 6.8 (Figure source [207]) illustrated the concept of k-NN classification where the class label changes from '+' to '-' to '+' by increasing the value of $k$ from 1 to 3 to 5 respectively.

k-NN algorithm looks at the close neighbourhood of the input example in the feature space and labels it that as seen in this close neighborhood [209]. In other words, a test sample ($x^{(i)}$) is classified considering majority class of its neighbours:

$$y = \arg\max_{t^{(i)}} \sum_{r=1}^{k} d(x^{(i)}, t^{(r)}) \tag{6.7}$$

where $t^{(i)}$ is a class label and $d$ is the distance metric. The algorithm is generalized through a distance metric to measure the distance or similarity between training

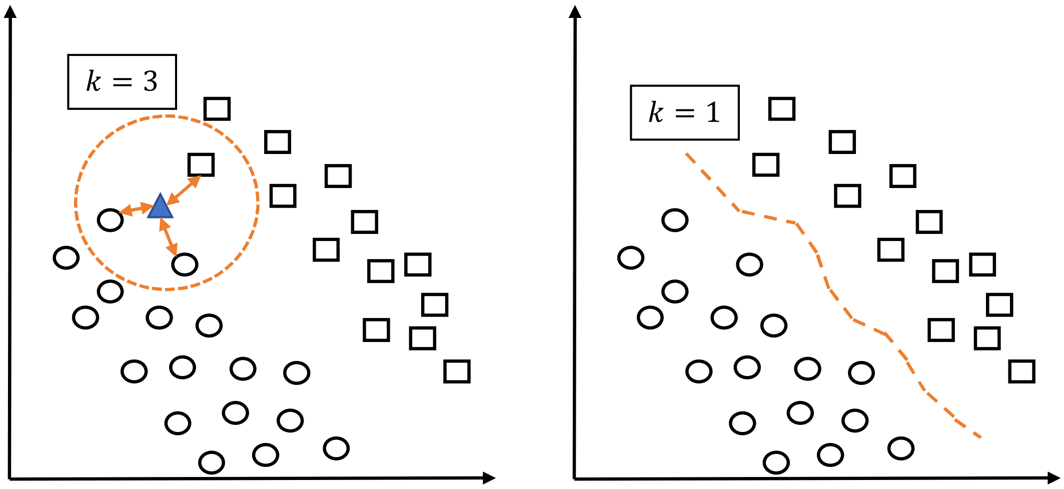


Fig. 6.9 k-NN classification concept and example 2.

samples and test examples. Figure 6.9 shows another example of k-NN classification cases for $k = 1$ and $k = 3$, where $k$ is the number of training examples closest to the considered input sample.

As it is observed in Figure 6.9, if $k > 1$, there are multiple training samples describing an example input test sample that is shown as a blue triangle mark. When $k = 1$, it creates a locally constant surface computed cell-by-cell in which the feature space is divided. There are various distance metrics and one of them used for real-valued vector spaces is Minkowski distance [210]:

$$d(x_i, x_j) = \left( \sum_{i=1}^{l} \left| x_i - x_j \right|^r \right)^{1/r} \tag{6.8}$$

where $l$ is the number of dimensions. $x_i$ and $x_j$ are the data points. When $p = 1$ it is Manhattan distance and if $p$ is set to 2, Euclidean distance that is most popular among distance metrics is got. It represents the root of the sum of the square of differences in vectors [210]. Euclidean distance can be weighted

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^{m} w_i \left( x_i - x_j \right)^2}, \tag{6.9}$$

where $w_i$ is the weight that influences the distance of instance $x_i = [x_{i1}, x_{i2}, \cdots, x_{im}]$ to the nearest neighbor instance $x_j = [x_{j1}, x_{j2}, \cdots, x_{jm}]$. $w_i$ represents the feature weighting that consists of $m$ weight coefficient for $m$ features. Furthermore, a distance weighted k-NN can also be applied directly in (6.7) before distance function $d(x^{(i)}, t^{(r)})$. One of the widely applied weighting is the inverse squared distance.

Cosine distance that is also called angular distance is a type of similarity measure [210]:

$$d(x_i, x_j) = \frac{\vec{x}_i \cdot \vec{x}_j}{\|\vec{x}_i\| \, \|\vec{x}_j\|}, \tag{6.10}$$

where $\cdot$ represents the dot product between two vectors and it is normalized by their magnitude.

The choice of the distance metric and the value of $k$ should be made carefully. In order have powerful k-NN classification, a proper value for $k$ should be selected. If it is set too small, k-NN becomes sensitive to class noise. However, selecting $k$ too large leads to include many neighbor points and hence increases the bias. A performance comparison of selected different $k$ values and distance metrics in terms of scintillation detection accuracy through carried out experimental tests is discussed in Section 7.2.

### 6.3.3 Decision Tree Learning

Decision tree learning is based on tree structures, defined by recursively partitioning the input space, as depicted in Figure 6.10. Decision tree is an acyclic graph in which each branching node a decision is made by examining a specific feature vector and depending on the decision the right or left branch is followed [209]. In other words, the learning takes place along the branches and nodes by means of applied functions for the decision criteria in each node [198].

A classifier can be represented by $C(S, x)$ where $x$ is the input point and $S$ is the training data having a set of labelled data:

$$S \overset{\text{def}}{=} \{(x_i, y_i)\}_{i=1}^{N} \tag{6.11}$$

where it also denotes the start node that contains all examples [209].

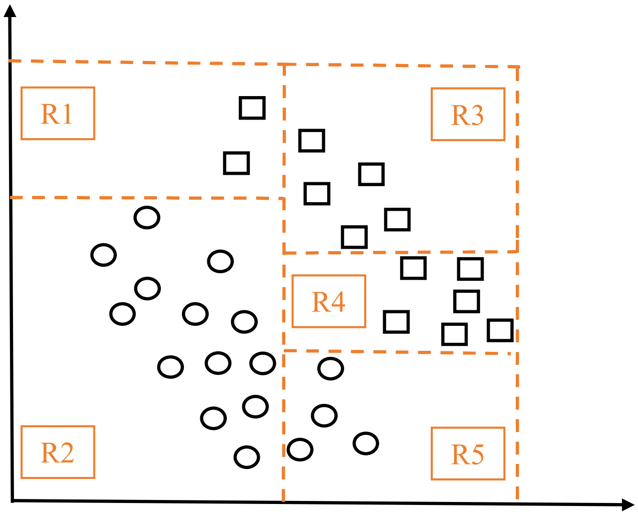


Fig. 6.10 An example of feature space partitioning in decision tree.

## 6.4 Ensemble Learning

ML methods sometimes are not able to obtain adequate performances when dealing with complex data, like noisy imbalanced high-dimensional data. In such cases, these methods might fail to capture the underlying multiple characteristics and structures of the data [211].

Ensemble learning methods exploit multiple ML algorithms to produce many predictive models based on diverse features, and fuse results of individual models using a selected voting mechanisms. By doing this, ensemble learning aims to achieve better performances than that obtained by the individual child algorithms. Increasing the complexity of the ensemble model decreases the model error until reaching a certain complexity, after it the error just increases as shown in Figure 6.11 (Figure source [211]). The trends of the variance, however,is opposite and it increases with the increase in the complexity of the ensemble. Thus, increasing the complexity of the ensemble is not always the way to achieve better models, but a balance between bias and variance is what we search for [211].

Ensemble learning combines a group of ML algorithms into one final model. This way, the knowledge of each algorithm is complimented and effectively utilized to gain superior performance for the overall ensemble. Figure 6.12 (adapted from [212]) illustrates the main idea of a typical ensemble classification model, which consists of two steps: (1) generating classification results using multiple classifiers

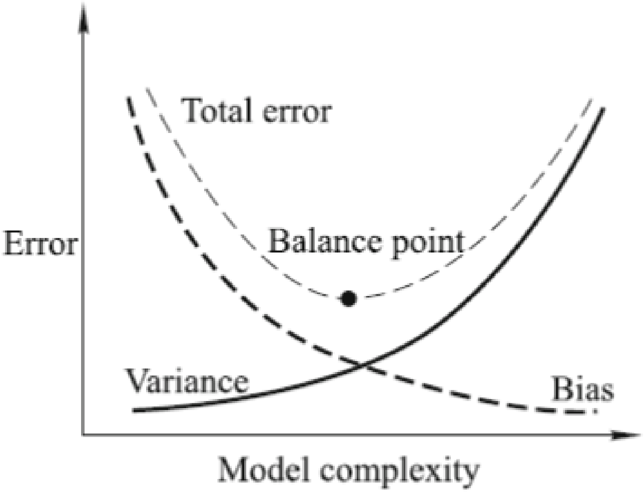


Fig. 6.11 The relationship between learning curve and model complexity.

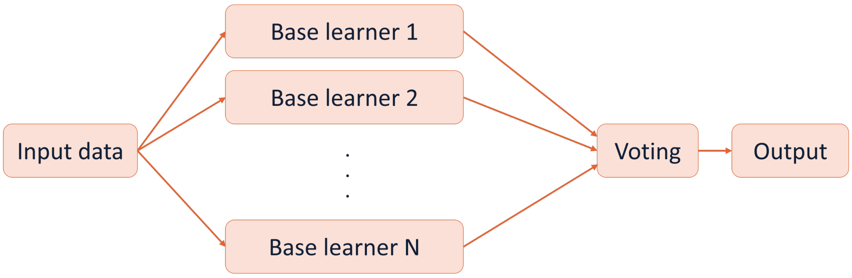


Fig. 6.12 The workflow of ensemble learning.

(called base learners), and (2) integrating multiple results into a consistency function to get the final result with voting schemes. The widely-used ensemble classification methods include bagged trees, AdaBoost, gradient boosting, random forest and random subspace [211].

The term Bagged trees comes from Bootstrap aggregation. It was introduced in 1996 [213] to reduce the variance of statistical learning methods, by building separate prediction models using subsets of the training set, i.e. bootstrapping the data. Then the average of all the models outputs in case of regression, or the voting result in case of classification, is taken as the final output of the model. Since each sub–model (or sub–tree) is trained using a subset of the training data, the remaining data of each subset (referred to as out-of-bag OOB observations) are used for validating that trained sub-model. This way, bagged trees usually increase the classification trees accuracy, but on the other hand we loose the interpretability advantage of decision trees. We fixed the number of sub–trees to 30 through out this work to avoid investigating the optimum number of sub–trees.

Bagging trees utilize an ensemble technique that creates a classifier from training a number of tree classifiers. Bagging trees classify by majority vote

$$C(x) = \text{Majority Vote}\left\{C\left(S^{(b)},x\right)\right\}_{b=1}^{B} \tag{6.12}$$

where $B$ is the number of decision trees in the ensemble.

Boosting is another ensemble technique that classifies by weighted majority vote

$$C(x) = \text{sign}\left[\sum_{m=1}^{M} w_m C_m(x)\right] \tag{6.13}$$

where $M$ is the number of decision trees in the ensemble.

The tree structure should be pruned to an optimal size through evaluations of cross-validation results.

Ensemble methods are considered the state-of-the art solution for many ML challenges. The objective is to compensate the errors of a single learner by other learners, and to reach an overall better prediction performance of the ensemble with respect to the single inducers [214]. The reasons behind this improvement of ensemble methods performance can be summarized as [214]:

- Overfitting avoidance: When only few data are available, a ML algorithm is likely to find enough hypothesis to predict perfectly all the training data but making poor generalization of the prediction. Averaging different hypothesis reduces the risk of overfitting and thus improves the overall predictive capability.
- local minimum avoidance: ensemble learners decrease the risk of obtaining a local minimum, that base learners might easily get stuck on.
- Extended search space: when the optimal classifier is outside the space of any single model, combining different models is equivalent to extending the search and hence better fitting the data.

## 6.5 Machine Learning for GNSS-based Scintillation Detection

From the discussion of the properties of the scintillation phenomenon (Section 5.2), the GNSS receiver measurements (Section 5.6), and the introduction to ML techniques, we summarize the drivers that motivated investigating ML techniques to tackle measurements from GNSS receivers to detect scintillation as follows:

- Scintillation, being a phenomenon with complex physics underlying it, is foreseen to benefit from ML techniques with their capabilities to handle complex measurements (eg. multipath affected scintillation measurements), similar to the other disciplines of science that successfully exploited these techniques to advance their understanding of the relevant phenomena.

- Scintillation events, being random and occasional (almost rare among the data continuously collected by scintillation monitoring stations) are the type of events that ML techniques are excellent in detecting with their data discovery and prediction capabilities. ML can be well trained to discover patterns in the data that might now be as frequent as other classes present in the data.

- Scintillation, being a phenomenon of interest for many users that require warning for occurrence (and even forecasting in advance), without a reliable mechanism for such warning, is still in search for a detection mechanism and thus ML with their anomaly detection capabilities can take part in this undergoing search for a detection mechanism.

- GNSS-based data, being measured using a signal that is susceptible to many environmental factors besides scintillation, that have properties difficult to process using mathematical models alone, are suited for data driven models.

- GNSS measurements, being sensitive to anomalies in the hardware of the transmitters and the receivers, are suitable candidates for ML techniques that are capable of handling noisy data.

- With the increase in the amount of scintillation measurements, they have become suitable to be exploited by ML techniques that require abundant data in first place to work successfully.

- Also, the abundant amount of scintillation data covering the last 2-3 solar cycles, invites exploring the data with ML techniques that are able to discover hidden structures in data that span such long periods and have complex physics.

Referring to the ML modelling cycle in Figure 6.4, we follow the journey of a model from data preparation to prediction. For data preparation, we focus on the feature engineering. The objective is to explore the features available from the GNSS-based scintillation receivers (discussed in Section 5.6) only. We will not include features from other instruments, like the Dst index, for the following reasons:

- The features available from the GNSS receivers are already abundant with measurements in multiple frequencies for various applications (positioning, TEC, scintillation indexes, etc). Thus, a thorough investigation of GNSS-only features is needed.
- All the anomlies we are distinguishing in this receiver (multipath and oscillaor anomalies) are actually GNSS anomlies. Thus, investigating the GNSS signal itself for more understanding of the signature of these anomalies on the GNSS scintillation measurements is needed. Since the physics underlying the various effects are different, we challenge the ML models to be able to distinguish that these are different clusters/classes.
- By relying on GNSS measurements only, we target developing models that do not need to consult external measurements to process GNSS data, for example external databases for retrieving the geomagnetic indexes and the space weather conditions. This important for autonomous operation of scintillation detection specially for real-time scintillation detection.
- Finally, since the GNSS signal is rich with propagation environment, and it is already sensitive to the scintillation phenomenon, we focus on expanding the investigation of the signature of scintillation on the signal, and focus on different aspects of the GNSS measurements to derive the scintillation detector.

Among the ML disciplines, in this thesis we will focus on supervised ML techniques, and not unsupervised for the following reasons:

- Unsupervised clustering requires deep investigation of the results of the clustering task to understand why certain groups of data were clustered together.

This usually is aided by visual inspection and other similar techniques. For scintillation, following such inspection method, and due to the complex phenomenon that required us to check many data sources and visualize the data in many formats, this lead us to gradually label the data, after such extensive inspections, with the multipath and oscillator anomalies classes besides scintillation. At the end the amount of available labelled data was enough to follow the supervised learning paradigms which we show in this thesis. We do not show results from the investigation of unsupervised ML techniques because the subsequent model validation and deployment steps were not followed.

- Unsupervised dimensionality reduction, was not critical for the the scintillation detection task using data stored in scintillation repositories. Nevertheless, for deploying the ML-based detectors in future GNSS-based scintillation monitoring stations, such techniques are worth investigation to adapt with limited computation and storage resources at receiver level. For this thesis, such techniques will not be investigated.

Finally, among the supervised ML techniques, we will not consider DL for the following reasons:

- DL required much more data that the other supervised ML techniques. This indeed requires labelling for a considerably large training set.
- Although DL has proven to be the game changer in ML in many applications, we choose to investigate the capabilities of the other ML techniques as a first step and explore their capability to do this task. Indeed, DL techniques are the next step to expand and improve upon the current capabilities that will be shown in the results obtained in this thesis.

# Chapter 7

# Detecting Phase Scintillation at High Latitudes Using Ionospheric Scintillation Monitoring Records and Machine Learning Techniques

In this Chapter, we present a bagged tree model able to detect phase scintillation at high latitudes with 95% accuracy, 5% scintillation miss-detection and 5% scintillation false alarm. The input to the model is a series of 3 minutes of TEC, 3 minutes of the change in TEC (dTEC), and the satellite elevation. These values are extracted from Ionospheric Scintillation Monitoring Record (ISMR) logged by ISM receivers. We compare the performance of this model to Support Vector Machine (SVM) models, k-Nearest Neighbors (k-NN) models, and also to other decision tree models. Furthermore, we assess the ability of the TEC and dTEC features to detect scintillation independently of the scintillation indexes. For this, we compare the above decision trees, kNN and SVM models to the same models but trained using scintillation indexes as additional inputs. Moreover, we show the results of testing the proposed model using a novel data set. Finally, we compare the accuracy of the ML model to the performance of a detector based on the phase scintillation index $\sigma_\phi$ threshold.

## 7.1 Introduction

ISM receivers have been continuously recording ionospheric measurements for the last couple of decades. This resulted in a rich repository of ISMR data, which are logged at 1 minute rate, as well as other records with higher logging rate (for example raw correlator outputs are typically logged at 50 Hz or 100 Hz and raw Intermediate Frequency (IF) GNSS signals are logged at several MHz).

For such a high volume of data, automatic detection of scintillation is necessary, and in fact it has always been implemented by ISM receivers and ionospheric studies' researchers. However, phase scintillation detection, and GNSS scintillation in general, has not been a trivial task because of the scintillation-like anomalies that affect the scintillation indexes. Satellite and receiver clock anomalies, and multipath are the main sources for false scintillation alarms in GNSS measurements [61].

In recent years, ML models to detect phase scintillation have been proposed in the literature. For example, in [201], the high-rate raw correlator measurements from ISM receivers are utilized to train ML models able to detect phase scintillation and the performance of SVM-based implementations for phase and amplitude scintillation detection is evaluated. In [215], amplitude and phase scintillation indexes provided in ISMR files are utilized to carry out the detection task. In this Chapter we utilize the other measurements in ISMR files to detect phase scintillation. We investigate the feasibility of detecting phase scintillation relying on TEC and the rate of change of TEC over time (dTEC).

The motivation behind using ISMR data is that they are available with almost continuous monitoring for decades. That makes them a rich resource of scintillation monitoring data. In this Chapter we focus on exploiting TEC and dTEC measurements calculated from dual frequency pseudorange measurements and carrier phase measurements, respectively. Furthermore, besides SVM [215] and decision tree [198][197] learning methods, the usage of k-NN algorithm for phase scintillation detection is investigated through a comparative performance analysis. Because, beforehand it's quite hard to choose directly the correct method in most cases in high dimensional spaces on the selection of SVM, k-NN, and decision trees for the classification problems. Most of the time, a validation data-set is used to not only optimize hyperparameters of the algorithms but also to choose between algorithms, as is investigated in this work.

## 7.2 Implementation and Test Results

ISMR data collected in Antarctica was utilized to train and test the ML models. In this section we describe the data preparation for the ML task. Then we present the ML models obtained in this Chapter. Finally we show and discuss the testing results of testing these models.

### 7.2.1 Data Preparation

The data utilized to train, validate and test the models were collected at the South African Antarctic research base (SANAE IV, 71.67 S, 2.84 W) using a Septentrio PolaRxS Receiver between 23-29 August 2018. These days were selected because on 26-27 August 2018 phase scintillation occurred at high latitudes due to a G3 geomagnetic storm [216].

The data was manually labelled using visual inspection after consulting data from consecutive days (to check for multipath), data from all visible satellites (to check for receiver clock errors), and data from different frequency bands (to check for satellite clock errors). To read more about checking for multipath and oscillator anomalies refer to [197] and [184] respectively. The data is labelled as scintillated if the phase measurements were randomly fluctuating, but no multipath or clock anomaly was observed during the anomaly check mentioned above. Otherwise, the data was labelled as non-scintillated. Approximately $16,000$ labelled samples were extracted, of which $50\,\%$ were scintillated. The samples were randomly split into training and testing sets, with 70% of the data in the training set. Again, 50% of the training set was scintillation. Validation using 5-fold cross-validation was also implemented during the training phase.

Two sets of attributes were considered for training the ML models:

- F1: contains a 3-minutes series of TEC, dTEC and elevation values. A sliding window step of 1 minute was implemented. In other words, the label is given at a rate of 1 minute, and we evaluate 3 minutes of data to give the label. TEC and dTEC are logged in ISMR files every 15s.

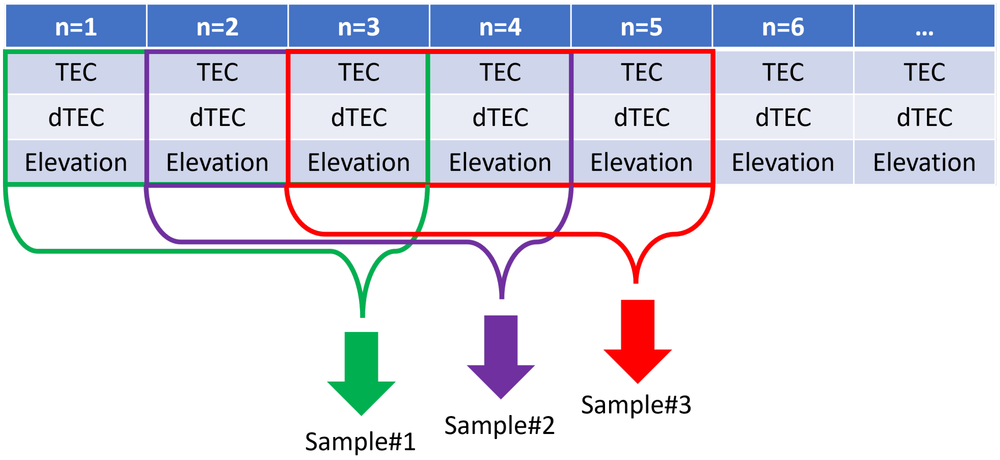


Fig. 7.1 Data preparation.

- F2: contains all the measurements related to L1 CA signal in the ISMR file, also grouped in blocks of 3-minutes overlapping measurements, with sliding window step of 1 minute.

Figure 7.1 depicts how the 3-minute overlapping window was prepared. In this figure, we showed only 3 features for demonstration purpose, but the concept is general.

In [217], the analysis carried out through collected data shows that the mean duration of the phase scintillation events in the polar region is 5.6 minutes. In this work, so as not to miss the below average duration scintillation events, time-window is adjusted to 3 minutes through an experimental analysis. Furthermore, the tests carried out applying an extended time window (e.g. 5-min) has not provided a significant improvement in the accuracy.

### 7.2.2 The Machine Learning Models

Table 7.1 summarizes the applied ML models having different settings. The first column shows the general ML models name. The second column shows the name of the variation of the model. By model variation we mean different hyperparameter settings of the model. The third column describes the model variation indicating the hyperparameters values.

Table 7.1 The Machine Learning models considered

| Model | Variation | Description |
|---|---|---|
| **Decision Tree Learning** | Fine | Max number of splits = 100 |
| | Medium | Max number of splits = 20 |
| | Coarse | Max number of splits = 4 |
| | Bagged | Ensemble method: Bag,<br>Number of learners = 30<br>Max number of splits = No limit |
| | Boosted | Ensemble method: AdaBoost<br>Number of learners = 30<br>Max number of splits = 20 |
| **SVM** | Linear | Kernel function: Linear |
| | Quadratic | Kernel function: Quadratic |
| | Cubic | Kernel function. Cubic |
| | Fine | Kernel function: Gaussian<br>Kernel Scale = $\sqrt(n)/4$<br>where $n$ is the number of predictors |
| | Medium | Kernel function: Gaussian<br>Kernel Scale = $\sqrt(n)$ |
| | Coarse | Kernel function: Gaussian<br>Kernel Scale = $4\sqrt(n)$ |
| **k-NN** | Fine | Distance metric: Euclidean<br>Number of neighbors = 1 |
| | Medium | Distance metric: Euclidean<br>Number of neighbors = 10 |
| | Coarse | Distance Metric: Euclidean<br>Number of neighbors = 100 |
| | Cosine | Distance metric: Cosine<br>Number of neighbors = 10 |
| | Cubic | Distance metric: Minkowski<br>Number of neighbors = 10 |
| | Weighted | Distance metric: Euclidean<br>Number of neighbors = 10<br>Distance weight : Square inverse |

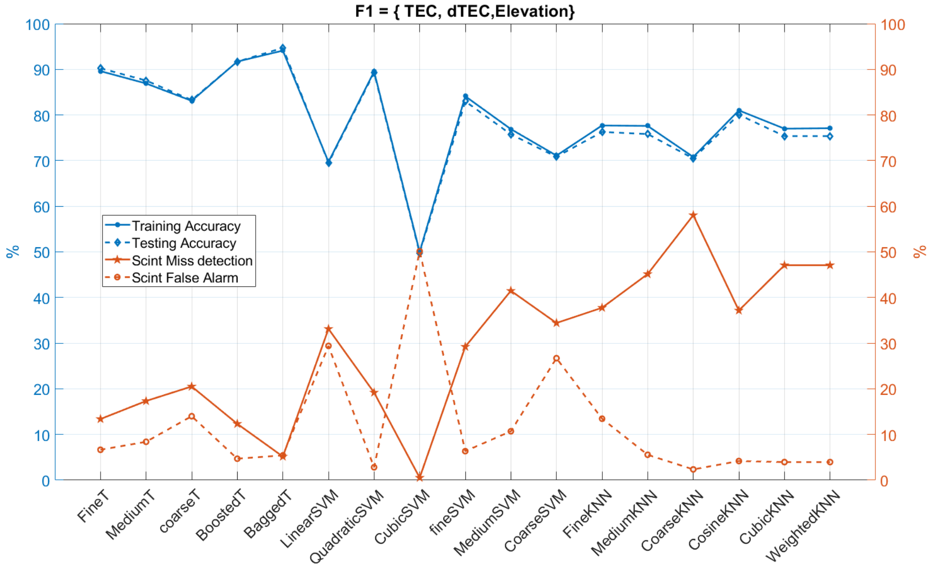


Fig. 7.2 Results of training and testing the models using feature set F1.

### 7.2.3 Experimental Test Results

Figure 7.2 reports the results of training and testing the models using feature set F1, while Figure 7.3 reports the results of training and testing the models using Feature set F2. The x-axis reports the models. The solid and dashed blue lines report the training and testing accuracy, respectively. The solid and dashed red lines report the miss-detection and false alarm rate respectively for the scintillation class.

Focusing on the bagged tree models (BaggedT) in Figure 7.2 and Figure 7.3, the results show that it is possible to obtain a model with 95% accuracy, 5% false alarm, and 5% scintillation miss detection, relying on TEC, dTEC and satellite elevation measurements alone (i.e. F1). Moreover, using F2 which includes all ISMR measurements did not give superior results to using F1 only. This indicates that TEC and dTEC are utilizable to detect phase scintillation without the need for phase scintillation indexes. Furthermore, many of the ML models achieved comparable good results in terms of accuracy, however the bagged tree demonstrated the highest accuracy.

The k-NN models in general reported lower accuracy and higher scintillation miss-detection than the trees. The decrease observed in the accuracy of coarse k-NN is expected since the number of neighbors might be accepted as the limit considering the rule of thumb $k < \sqrt(m)$, where $m$ is the number of training examples. At

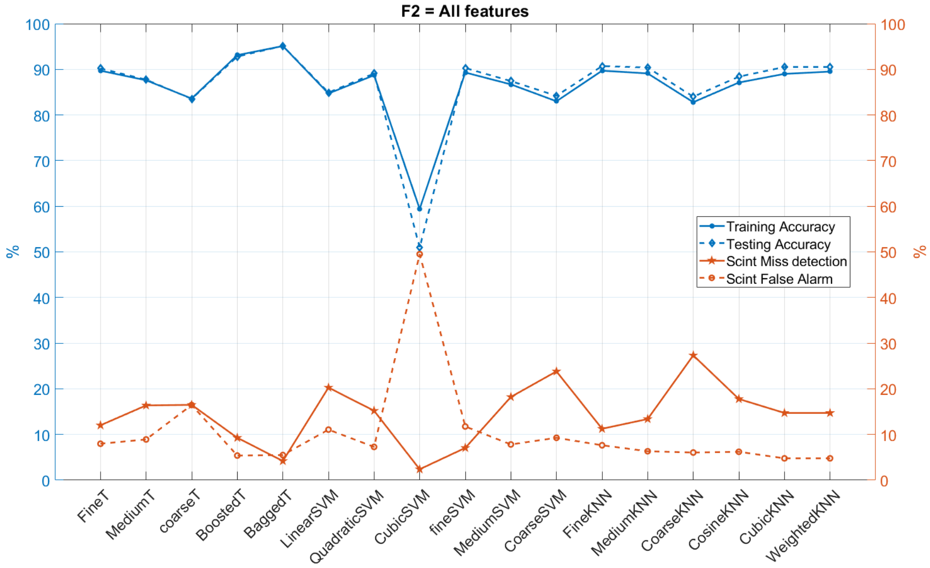


Fig. 7.3 Results of training and testing the models using feature set F2.

the same time, the cosine distance function, which is used when the magnitude between vectors does not matter but the orientation, could track better the trend of the scintillation indices (i.e., consecutive decrease/increase).

Finally, the SVM models, except the cubic SVM, reported a performance comparable to the trees. A drop in the accuracy of the cubic SVM is expected at an acceptable range observed in F1 and F2, considering the possibility that higher-degree polynomial might lead to over-fitting in the training test set. However, the same is not valid in cubic k-NN, because the classification output is computed through the majority class in that case. Furthermore, Manhattan distance has not been applied since it would work as thresholding.

## 7.2.4 Comparison with Standard Method

We compare the performance of the bagged tree model trained with F1 to the performance of a standard method in the literature; the threshold on the phase scintillation metric $\sigma_\phi$. For the Septentrio PolaRx5S Receiver installed at SANAE IV station, a threshold of $\sigma_\phi = 0.15$ *rad* is acceptable to detect moderate scintillation. To detect weak scintillation, lower threshold value is needed. For this reason, we tested different threshold values: 0.05, 0.1 and 0.15 *rad* as shown in Table 7.2. In

Table 7.2 Comparing the performance of the bagged tree model to standard threshold method

| **Method** | **Miss detection [%]** | **False alarm [%]** | **Accuracy [%]** |
|---|---|---|---|
| **Threshold=0.05** | 27.8 | 16.5 | 79.0 |
| **Threshold=0.10** | 57.2 | 3.2 | 70.7 |
| **Threshold=0.15** | 71.4 | 1.3 | 64.1 |
| **Bagged Tree Model** | 5.2 | 5.4 | 94.7 |

the table we report the scintillation miss-detection and false alarm beside the overall accuracy. We can see that with the threshold technique, the accuracy is below 80% while with the bagged tree model it reaches $\sim 95\%$. The improvement comes mainly from the significantly low miss-detection rate of the ML model (5.2%) compared to 27.8% for the best case in threshold method. The false alarm is slightly higher than a 0.1 rad threshold, but it is acceptable when looking at the overall accuracy.

## 7.3 Summary

In this Chapter we touch on the feasibility of using TEC, an ionospheric metric that is provided by dual- and multi-frequency receivers, for detecting scintillation. We use ML techniques to achieve the detection task.

We compare the performance of SVM, k-NN and Tree models trained to detect high latitude phase scintillation. We propose relying on series of 3-minutes TEC and dTEC measurements extracted directly from ISMR files to train the models. We compare these models to models trained with 3-minute samples that contain all L1CA related measurements in the ISMR records, in addition to the TEC and dTEC values. Finally, we show examples of testing the models with a novel data set.

We show that bagged trees relying on 3-minutes of TEC and dTEC measurements are able to detect phase scintillation with 95% accuracy, 5% scintillation miss detection and 5% scintillation false alarm.

The potential of this result is to question if it is possible to train a ML model to infer the value of the phase scintillation index from TEC measurements. This could open the door for utilizing professional receivers (like the IGS network) for estimating the scintillation index from the TEC measurements that they already provide. Such GNSS networks, have wider geographic distribution than the ISM receivers as of today, higher data availability and longer history of ionospheric measurements. Thus, utilizing data from these receivers to measure scintillations, or at least detect scintillations occurrence, could significantly increase the geographic and temporal availability of scintillation measurements.

# Chapter 8

# A Machine Learning Approach to Distinguish Between Scintillation and Multipath in GNSS Signals

This chapter presents a ML approach to distinguish between ionospheric scintillation and multipath in GNSS-based scintillation monitoring data. First, we investigate the feasibility of the classification task using measurements provided by various commercial and software–defined scintillation monitoring receivers. In particular we differentiate between receivers that log measurements at low rate (i.e. they log scintillation metrics and observables) and high rate (i.e. correlator outputs). We find that the high rate correlator measurements are enough to carryout the classification task with high accuracy, without relying on scintillation indexes. This was further investigated by training another twenty four models based on combinations of features derived from the receiver correlation stage outputs alone. These features were also investigated for different averaging windows ranging between 1 and 180 seconds. The results of the model with the best performance are presented. The overall accuracy of the model is 96% with 2% miss-detection rate and a negligible false alarm rate for the scintillation class in particular. The gain in the amount of scintillation data is up to 17.5% that would have been discarded if an elevation mask of 30° was implemented.

## 8.1 Introduction

Being based on the assessment of the variations of the amplitude and phase of the signals, scintillation indices are triggered by causes other than scintillation, which make these detection schemes suffer from false alarms. Thus, after the detection, further inspection of scintillation suspected the data is needed to verify if the source that triggered the indexes was scintillation or not. Multipath, which results from signal reflection and refraction by the environment around the receiver, is the most common and the most dangerous cause because it inflates $S_4$ index in a similar way to scintillation [178, 61]. In the literature, there are few techniques to avoid multipath in scintillation monitoring data. The first, which is the most reliable, depends on characterizing the environment around the scintillation monitoring receiver and thus mapping the sources of multipath and consequenctly excluding the data contaminated by these sources [178, 186]. This method is location-dependent and has to be repeated for every monitoring station. Another drawback of this method is that it is not sensitive to changes of the surrounding environment, that may occur. The second technique relies on eliminating data below an elevation threshold where most data with potential multipath contamination exist. This method has two drawbacks: i) it results in reduced visibility of the sky [187] and thus reduces the amount of useful scintillation data up to 35-45% [186], and ii) it actually does not guarantee multipath free samples by assuming all the signals above the elevation threshold are multipath free. A common elevation threshold is 30° and, as will be shown in the results Section, sometimes multipath signals exceed that threshold resulting in false scintillation alarms. In this Chapter, we target those two drawbacks by developing a ML model, able to distinguish between multipath and scintillation in scintillation monitoring data.

The boom in using ML techniques in the last decade found its way to both scintillation forecast and GNSS–based scintillation detection activities. The use of ML techniques to predict near–future ionospheric scintillation occurrences was pioneered in 2010 [218]. The proposed ML model relied on data from GNSS receivers and Digital Portable Sound (DPS) Digisondes installed in Brazil to forecast the ionospheric scintillation index $S_4$ for up to one day ahead. The bagged regression tree model developed also takes as inputs: i) the time of the day, ii) Kp data from World Data Center for Geomagnetism in Kyoto, and iii) the F10.7 solar flux with a resolution of 1 day obtained from National Geophysical Data Center. Since then, many articles investigated developing ML based models to predict ionospheric scintillations based on GNSS measurements and/or non–GNSS based measurements (e.g [219, 194, 220]).

On the other hand, detecting ionospheric scintillations is of great interest for GNSS–based scintillation monitoring activities. The ability to detect scintillation based on GNSS measurements alone is particularly favored because it eliminates the need for installing multiple instruments and/or for aiding the instruments with external data. In [200] and [201] support vector machines were trained to detect amplitude and phase scintillations respectively, while in [198] a decision tree was adopted for the classification task. Recently, [199] used Convolutional Neural Networks to accomplish the same task. All these models report high detection accuracy. However, in these models multipath effects were not considered, and thus the developed models were not trained to detect multipath. This limited the potential of implementing the models in operative scenarios where multipath is for sure expected to exist.

In this Chapter, we investigate the possibility to distinguish between multipath and scintillation using ML models. Our objective is to rely only on measurements provided by GNSS receivers. The motivations behind this objective are:

1. The GNSS signal has proved to be a powerful remote sensing tool with successful deployment in many enviornmental applications. We stress this advantage in challenging the ML model to process the GNSS signal alone and differentiate between scintillations that happen up in the ionosphere, and multipath that happens in the vicinity of the receiver.
2. Since the mechanisms causing scintillation and multipath are different, their effects on the signal are also different. Unfortunately, the scintillation indexes are not able to differentiate between these two effects. Thus, we want to search for another metric/method/technique able to carry out this task and distinguish between the two effects. In other words, we want to search for a scintillation detection method that is selective and sensitive to scintillations, even if multipath is present in the environment. Indeed if we could not achieve a satisfactory ML model that relies only on GNSS measurements, we will have to consider other alternative methods and data.

3. We would like to reduce/cut the need for external data sources that might require the model to consult instruments or data sources other than the GNSS receiver. Fusing such external measurements might lead to denying the model from working in real time in the future. It also might put constraints on the installation of the model in monitoring stations at locations with limited budgets and/or internet connection.

Furthermore, we consider the possibility of achieving this objective of distinguishing scintillation from multipath based on different measurements sets provided by commercial and software–defined scintillation monitoring receivers. The motivation behind this consideration is that different measurements are available based on the type of receiver and/or the configuration of the receiver output measurements. We differentiate between receivers that provide measurements with high logging rate (50Hz for example) and other receivers that provide the measurements at lower logging rate (1Hz for example). In particular, we consider if the user is able to access the correlators measurements I and Q or not. Here we assume the end user has more options on the measurements they can extract, and we propose solutions that exploit this advantage.

Finally, we highlight that our objective is to detect scintillation, and to distinguish it from multipath in particular. For this reason, we tolerate if the model will not be able to distinguish between multipath and clean signals.

## 8.2 Data Preparation

In this Section we describe the data we used in this work. We focus on how these data were prepared for the ML task.

### 8.2.1 The Data Set

Approximately 45 hours of data captured in Hanoi, Vietnam in 2015 are exploited to train, test and validate the ML models trained in this chapter. The data were collected in March, April, July and October 2015 in Hanoi, Vietnam, as part of the ERICA project [31, 32]. The data collection contains days with strong, moderate and weak–to–no scintillation. Moreover, some of the data collections coincide with days where geomagnetic storms were recorded. The occurrence of geomagnetic storms was checked by consulting Space Weather Services [221] and Solar Influences Data Analysis Center [222].

We used a software defined GNSS receiver (SDR) to process the data and extract the post-correlation outputs of the tracking loop, i.e. the in–phase $I$ and quadrature-phase $Q$ outputs, at a rate of $50Hz$ for all the visible GPS satellites. The SDR also estimates the raw signal intensity $SI$, the amplitude scintillation metric $S_4$ and the carrier to noise ratio $C/N_0$. It also calculates the azimuth ($\theta_{AZ}$) and elevation ($\theta_{EL}$) of the satellite.

We group the data into 3 categories: training data, testing data, and operative scenarios data.

### 8.2.2 Training and Testing Data Labelling

To label the data we compare data from consecutive days for a satellite to identify signal distortions induced by multipath. As discussed in Section 5.7, in presence of multipath the $S_4$ value is inflated taking values in order usually between the weak to moderate scintillation threshold with a spatio-temporal pattern. For GPS signals for example, looking at the $S_4$ behaviour over several days, if the multipath is generated by a static obstacle, we notice the same pattern each sidereal day (day minus 4 minutes). Therefore, if $S_4$ is inflated and takes a similar trend in multiple days, this is an indication of multipath presence. If $S_4$ is inflated without sidereal repetition, we label it as scintillation. Finally, we label a sample as clean if the $S_4$ index is not inflated.

To avoid including in the training and testing data signals that are affected by both multipath and scintillation at the same time, we label multipath data from days with reported no scintillation. Thus, in the training and testing data we only include data we could confidently label as either multipath or scintillation. As a result of such a strict labelling methodology, and to make sure we have balanced classes, we have a total of 14.5 hours of manually labelled data for each class: multipath, scintillation and clean. We used 10 hours for training and the remaining 4.5 hours for testing the models.

### 8.2.3 Statistics of the Training and Testing Data Sets

In this Section, we examine the statistical distribution of the features in the data that we have selected to train and test the ML models. This investigation is important for ML data preparation to ensure that the data that were selected are actually representative of their classes. More details about the good practices for data preparation can be found in [223]. For example, when we check that we have clean signals that were collected at various elevations, we ensure that we are not selecting clean signals only from high elevation and thus implicitly leading the model to give higher probability of multipath for signals with low elevation or low signal to noise ratio for example. By doing this representative data selection, we ensure that the model will look at the properties of the multipath/scintillation affected signal in general and the changes that happen to the signal rather than the location of the signal in the sky. The same applies for scintillation, where we want to ensure including scintillation samples at low elevations in particular in the training set. We will carry out this investigation by examining the histograms of $S_4$, $C/N_0$, $SI$ and $\theta_{EL}$, shown in Figure 8.1 and Figure 8.2 and discussed below.

$S_4$ **and** $C/N_0$

The first plot to investigate, shown in Figure 8.1, is the bivariate histogram for $S_4$ and $C/N_0$ of the training set . The x–axis shows $C/N_0$ and the y–axis shows $S_4$. The blue color is for scintillation, red for multipath and green for the clean signals. As expected from clean signals, they are all with $S_4$ values below 0.15 and they do not drop to the very low $C/N_0$ values that are associated to degraded signals. For scintillation and multipath classes however, we see considerable overlapping between their histograms. In our training set, scintillation values are everywhere on $S_4$ scale indicating various severity of the scintillation events from weak to strong scintillations. They are also spread on all $C/N_0$ values indicating severe fluctuations in the signal intensity that lead to weak as well as enhanced signal to noise ratios. On the other hand, multipath signals have smaller values of $C/N_0$ compared to the clean signals, because multipath signals are degraded signals with signal to noise ratios typically lower than the expected values. On $S_4$ scale, we have values up to $S_4 = 0.5$ which will certainly pass the $S_4$ threshold of scintillation monitoring stations.

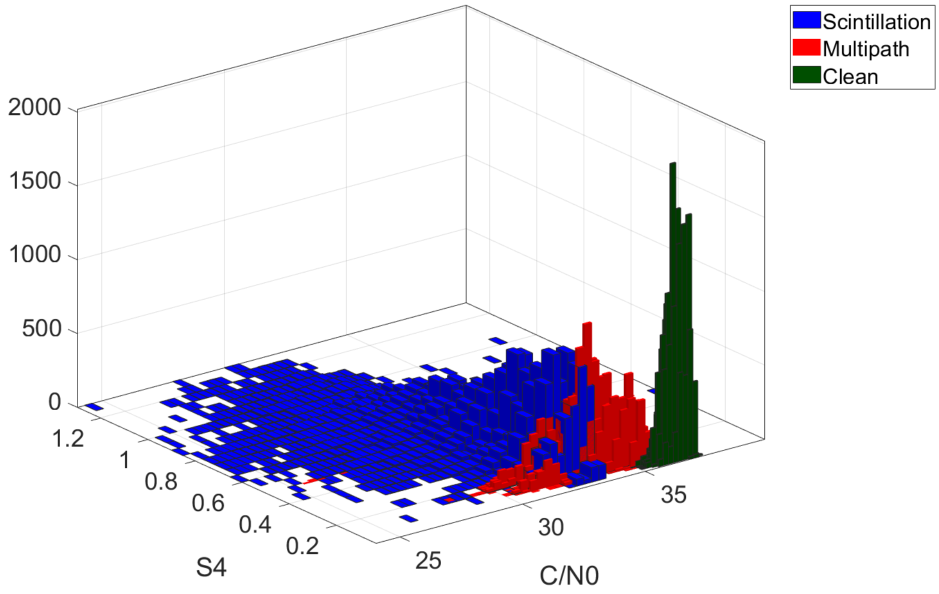


Fig. 8.1 Histogram of $S_4$ and $C/N_0$ of the training set.

### Signal Intensity (SI) and Satellite Elevation (EL)

The second bivariate histogram to investigate, shown in Figure 8.2, is for the elevation angle ($EL$) [°] and signal intensity ($SI$) of the training set. The x–axis plots $SI$ while the y–axis plots $EL$. Again, the blue color is for scintillation, red for multipath and green for the clean signals. In this histogram, the clean signals and the multipath signals occupy different areas of the $EL$–$SI$ space, while the scintillation class is widely spread all over the space. The signal intensity increases consistently with elevation for clean signals as they should be. While for scintillation, enhanced and degraded signal intensities are found at all elevations as a result of constructive and destructive signal interference. The multipath class is centred at low elevations but it reached up to 40° in this data set. This is above the elevation cut-off angles adopted in literature.

In conclusion, the training data set looks representative of the range of values that the signals can occupy in terms of $\theta_{EL}$, $C/N_0$, $SI$ and $S_4$. We are confident to use this data set to train our ML models.

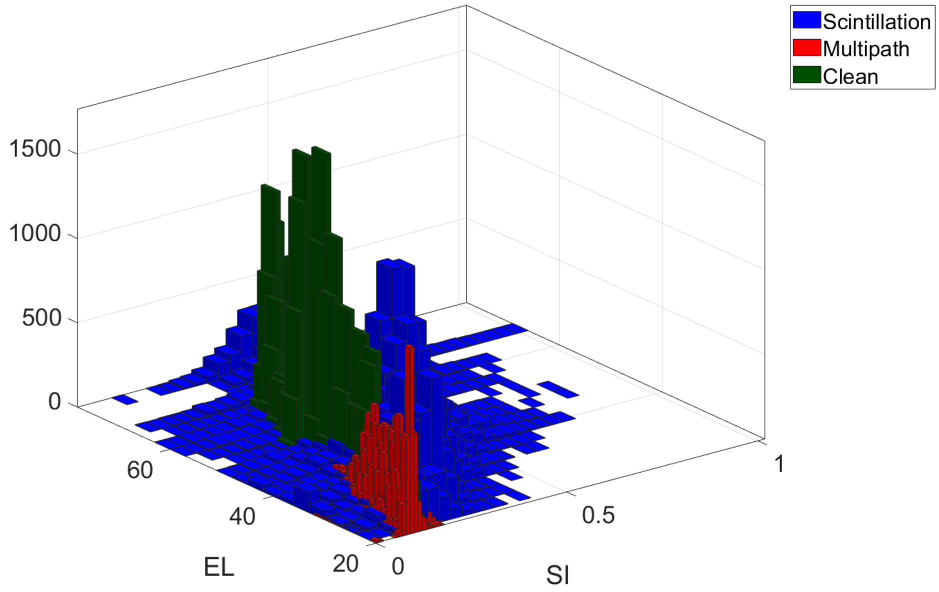


Fig. 8.2 Histogram of the elevation angle and SI of the training set.

## 8.2.4 Machine Learning Attributes

In this Chapter, we consider three options for the GNSS receivers that provide the scintillation measurements. The first one is a receiver that logs $\{SI, S_4, C/N_0, \theta_{AZ}, \theta_{EL}\}$ with a slow rate. We choose 1 Hz, but the main assumption is not to compute further statistics (average, variance, etc) from the measurements because it is already at a low rate. The second is a receiver that provides the same measurements above with a high logging rate (50 Hz). The third is a receiver that provides $I$ and $Q$ in addition to the above values at 50 Hz. For the later two types of receivers, we calculate the following additional values from the 50 Hz logs to use them as input features for the ML models.

From the definition of scintillation as fluctuations in signal intensity (and phase) we decide to use features that can capture fluctuations in general as inputs to train the ML algorithm. Thus we choose the average signal intensity, the variance in the signal intensity and the covariance between $I$ and $Q$ over a time window. We choose to adopt the raw signal intensity given by (2.11) without further processing it with detrending or noise removal. The average value of the signal intensity is calculated from $SI_i[n]$ values over the previous time window of $N$ samples as:

$$\overline{SI}_i = \frac{1}{N}\sum_{n=1}^{N} SI_i[n] \tag{8.1}$$

Similarly, the variance of the signal intensity is calculated over the same time window as:

$$\sigma_{SI,i} = \frac{1}{N-1}\sum_{n=1}^{N} (SI_i[n] - \overline{SI}_i)^2 \tag{8.2}$$

In the case of the covariance, we choose to calculate the covariance between $I^2$ and $Q^2$ instead of $I$ and $Q$ to avoid the effect of the data bits that we have not removed from the received signal (eq.2.9), and thus the covariance was calculated as:

$$cov_i(I^2,Q^2) = \frac{1}{N-1}\sum_{n=1}^{N} (I_i[n]^2 - \overline{I_i^2})(Q_i[n]^2 - \overline{Q_i^2}) \tag{8.3}$$

where $\overline{I_i^2}$ and $\overline{Q_i^2}$ are the average values of $I^2$ and $Q^2$ calculated over the same time window.

The value of the averaging time window $T_w$ will be fixed to 1 minute for the investigation done in Section 8.4, and in Section 8.5 we present an investigation into effect of the averaging window on the performance of the ML models.

## 8.3 Selection of the Machine Learning Algorithm

Since we have labelled data, and we want to train a ML model to perform a classification task, we choose a supervised ML model, a classification learner in particular, to train our models. In this Chapter we do not compare different types of ML algorithms, but rather fix the ML algorithm and focus on investigating the type of measurements that will allow obtaining a model able to distinguish scintillation from multipath. This will also facilitate focusing the discussion on the useful measurements to distinguish scintillation from multipath. We thus investigate one type of models in this work, and leave investigating the best classification model or ensemble model for future work.

We choose the Bagged trees model [224] because in Chapter 7 it has been proven to give superior performance compared to the other tree, SVM and k-NN models that we have tested (see Table 7.1).

Figure 8.3 illustrated the task the ML model should achieve. The model takes the input features that were prepared from the outputs of an ISM receiver, and outputs whether the data represent scintillation–affected, multipath–affected or clean GNSS signals.

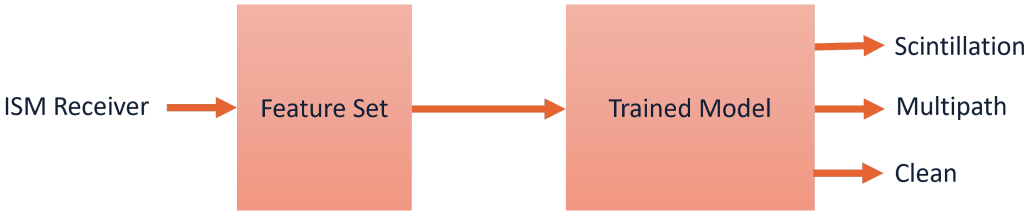


Fig. 8.3 Illustration of the task of the ML model to be developed.

## 8.4 Training Models Based on Different Receivers Logging Rates

In this Section we train 8 ML models based on assumptions about the logging rate of the available scintillation data from the receiver. These assumptions come from the fact that different receivers log the data at different rates. Even for the same receiver, sometimes it is possible to configure the receiver to output the data at custom rates. Since one of our objectives is to develop models that will assist researchers curating scintillation data repositories, we take into consideration that the repositories might contain data with different rates. Furthermore, such investigation sheds the light on the feasibility to carry out the data cleaning on data sets with low/sub-sampled rate. Such sub–sampling is beneficial in exploring the data repositories and identifying the outliers if present by retrieving data at lower rate which also implies faster scanning for the data repositories.

### 8.4.1 Machine Learning Attributes

Table 8.1 summarizes all the features that we use in this section to train the 8 models to distinguish between scintillation and multipath based on the assumed 3 types of receivers, dropping the subscript $i$ for the satellite number from the measurements described by (8.1), (8.3) and (8.2). Combinations of these features will be used as inputs to the various models as will be explained. The value of the averaging time window was fixed to 1 minute for the this part of the experiments.

| | | | |
|---|---|---|---|
| I Q | $C/N_0$ $S_4$ | AZ EL | SI $\overline{SI}$ $\sigma_{SI}$ $cov(I^2,Q^2)$ |

Table 8.1 Features (Attributes) to be considered for training

### 8.4.2 Training the Machine Learning Models

Table 8.2 reports the eight combinations of measurement groups that we want to investigate. The first column shows the name of the model trained with each measurement group. The second column lists the measurements that are utilized by each ML model. The last column gives a brief description of the measurements group.

$\text{MDL}_1$ is for a receiver where the measurements $F_1$ are logged with a low rate. $\text{MDL}_2$ on the other hand, works for receivers that log the measurements $F_2$ with a high rate. $\text{MDL}_3$ represents receivers where $I$ and $Q$ are available for the end user.

The other 5 feature sets shown in Table 8.2 represent variations of the above mentioned 3 feature groups, where each one tests a certain combination of measurements. $\text{MDL}_4$ tests if a ML model that does not rely on $\{S_4, C/N_0, \theta_{AZ}, \theta_{EL}\}$ can be developed. This is important because the calculation of $S_4$, as we discussed in Section 5.5, depends on the receiver manufacturer. Also, the estimation of $C/N_0$ depends on the receiver. The azimuth and elevation angles on the other hand are strongly biased towards multipath in the environment. Since all of the obstacles in this data set are fixed (or at least they were present in the data set for consecutive days), they can lead the model to understand that multipath can be identified with high accuracy based on the direction of the satellite in the sky. $\text{MDL}_5$ and $\text{MDL}_6$ test if excluding the azimuth and elevation will affect training the models for receivers with high and low logging rates respectively. $\text{MDL}_7$ tests what performance a model trained only on the azimuth and elevation can achieve. This model is similar to mapping the obstacles in a scintillation monitoring site. Finally, $\text{MDL}_8$ tests what

Table 8.2 The trained models for the considered low and high logging receivers and the Feature set (i.e. input measurements) considered for each model.

| Model | Inputs to Model (Feature set) | Feature set description |
|---|---|---|
| MDL$_1$ | $F_1 = \{S_4, C/N_0, AZ, EL\} + \{SI\}$ | All features from a receiver with low logging rate |
| MDL$_2$ | $F_2 = \{S_4, C/N_0, AZ, EL\} + \{SI, \overline{SI}, \sigma_{SI}\}$ | All features from a receiver with high logging rate, I and Q are not available |
| MDL$_3$ | $F_3 = \{S_4, C/N_0, AZ, EL\} + \{SI, \overline{SI}, \sigma_{SI}\} + \{I, Q, cov(I^2, Q^2)\}$ | All features from a receiver with high logging rate, I and Q are available |
| MDL$_4$ | $F_4 = \{Sl, \overline{SI}, \sigma_{SI}\} + \{I, Q, cov(I^2, Q^2)\}$ | Excluding $\{S_4, C/N_0, AZ, EL\}$ from the feature set of a receiver with high logging rate |
| MDL$_5$ | $F_5 = \{S_4, C/N_0\} + \{SI, \overline{SI}, \sigma_{SI}\} + \{I, Q, cov(I^2, Q^2)\}$ | Excluding $\{AZ, EL\}$ from the feature set of a receiver with high logging rate |
| MDL$_6$ | $F_6 = \{S_4, C/N_0\} + \{SI\}$ | Excluding $\{AZ, EL\}$ from the feature set of a receiver with low logging rate |
| MDL$_7$ | $F_7 = \{AZ, EL\}$ | Feature set to test the power of $\{AZ, EL\}$ in training the model |
| MDL$_8$ | $F_8 = \{S_4\}$ | Feature set to test the power of $S_4$ in training the model |

performance a model trained only on $S_4$ can achieve. This model is similar to finding the optimum $S_4$ threshold.

### 8.4.3 Performance of the Trained Models

Figure 8.4 reports the performances of the 8 bagged trees models that were trained using the training set described in section 8.2 . The x–axis indicates the model number as shown in Table 8.2. The y–axis on the left reports the overall accuracy

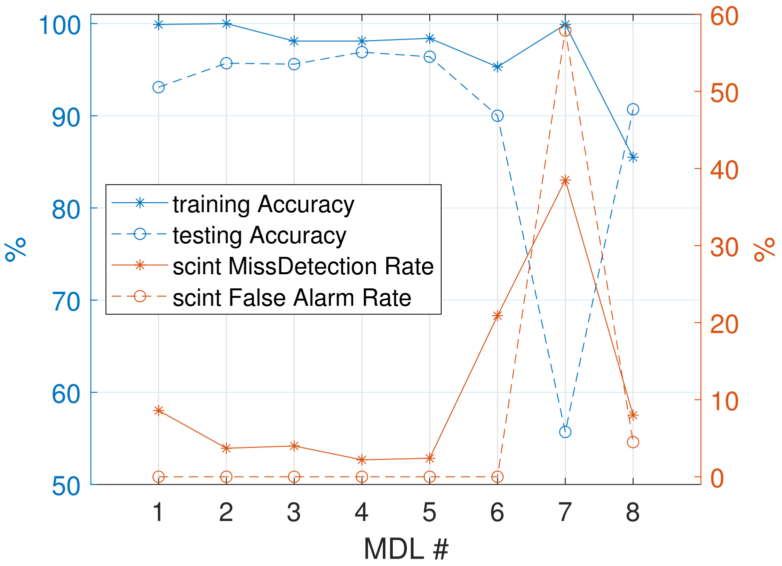


Fig. 8.4 Performances of the models trained based on the feature sets listed in Table 8.2

of the model in percentage, and the right y–axis shows both the miss detection and false alarm rates in percentage for the scintillation class in particular.

The blue solid line shows the accuracy achieved by the models during the training phase. It can be seen that all the models reported high accuracy (above 95%) at this phase except RX8 which achieved 85% only. To verify that the models can also achieve this high accuracy in other scenarios, the models were tested with an independent testing set (the testing set described in Section 8.2), from the same monitoring receiver in Hanoi. From this testing we report:

- The testing accuracy (blue dashed line) which is the ratio of the number of correctly classified samples to the total number of samples.
- Scintillation miss detection rate (the red solid line) which is the ratio of the number of scintillation samples wrongly classified to the total number of scintillation samples.
- False scintillation alarm (red dashed line) which is the ratio of the number of samples wrongly classified as scintillation to the total number of samples classified as scintillation.

### 8.4.4 Comparison of the Models' Performances

Below we summarize the results reported in Figure 8.4 for the models listed in Table 8.2:

- The first 3 models have high (above 90%) training and testing accuracy (blue lines), as well as acceptable (below 10%) miss-detection and false alarm rates (red lines). The deviation between the accuracy reported by the training phase and the testing phase are not huge enough (limited to 7% deviation between the accuracy reported during the model training and the one reported during the model testing) to suspect over–fitting. This indicates that for each of the three types of receivers we are investigating, a ML model to distinguish between scintillation and multipath can be developed using the measurements available from each receiver in addition to the ones that can be calculated from the available receiver outputs. Moreover, adding the variance and the average features to $\mathrm{MDL}_2$ outputs enhanced the testing accuracy as well as the miss detection rate with respect to $\mathrm{MDL}_1$. This means that a receiver with high logging rate has an advantage over the ones that log $\{SI, S_4, C/N_0\}$ at low rates.

- The performance of $\mathrm{MDL}_3$ that logs $I$ and $Q$ measurements is not better than $\mathrm{MDL}_2$ that does not include these measurements. In particular, the testing accuracy, miss detection and false alarm are similar for both models. This indicates that the availability of $I$ and $Q$ as an input to the ML model might not be an advantage as long as the receiver is logging $SI$ at a high rate.

- The availability of $\{S_4, C/N_0, AZ, El\}$ might not be necessary as indicated from the performance of $\mathrm{MDL}_4$ which takes as input $I$ and $Q$ that were logged at a high rate but doesn't consider $\{S_4, C/N_0, EL, AZ\}$ as inputs to the ML model. $\mathrm{MDL}_4$ actually has higher testing accuracy than the previous 3 models, and lower miss detection rate. From this result, we confirm that if a receiver logs only $I$ and $Q$ at a high rate, this should be sufficient to extract features that will train a model to distinguish between scintillation and multipath, without relying on $\{S_4, C/N_0, AZ, EL\}$. However, we still need to investigate if a receiver that logs only $SI$ at a high rate is sufficient, rather than logging $I$ and $Q$. Of course, if the receiver has a low logging rate, excluding all the 4 measurements $\{S_4, C/N_0, AZ, EL\}$ from the model training will leave only one feature $SI$ which is not enough to perform the classification task. For such receivers, $\text{MDL}_4$ is not feasible.

- To investigate the importance of azimuth and elevation measurements, we look at $\text{MDL}_5$ and $\text{MDL}_6$ which exclude $\{AZ, EL\}$ from the measurements set provided by receivers that log at high and low rate respectively. For the former, the performance of the model was not degraded with respect to $\text{MDL}_3$. For the slow logging receiver on the other hand, excluding $\{AZ, EL\}$ from the measurements set, left the model inputs to be only $\{S_4, C/N_0, SI\}$, and the accuracy of the model degraded to 90% while the scintillation miss detection rate jumped from 9 to 21% with respect to $\text{MDL}_1$. This indicates that, while $\{AZ, EL\}$ might not be important for receivers with high logging rate, they can enhance the performance of models developed for receivers with slow logging rate.

- $\text{MDL}_7$ gives a further hint on the effect of the location dependent measurements azimuth and elevation by training a model that takes as inputs only $\{AZ, EL\}$. Here we see that $\text{MDL}_7$ is able to achieve a high training accuracy, 100% actually, but fails the testing with accuracy of 55% only (blue dashed line). This indicates that the model was over-fit to the training data. Also the miss detection and false alarm rates are high (40% and 60% respectively), which is closer to a random guess made by the model. This poor performance is expected because such a model might learn where multipath comes from most of the time (i.e. fixed obstacles) but it can not differentiate between scintillation and clean signals based on where the signal comes from in the sky.

- Finally, $\text{MDL}_8$, which utilized $S_4$ only, was able to learn the best threshold on $S_4$ that will yield the highest accuracy. The training and testing accuracies are 85% and 90% respectively, while the miss detection and false alarm rates are below 10%. This good performance is expected from the metric adopted by the literature for scintillation detection, but as seen from the other models, it can be improved when combined with other measurements, and it can even be omitted.

In conclusion, for the three types of receivers we are considering, which represent both commercial and software-defined receivers, ML models can be developed to distinguish between scintillation and multipath. However, the receivers that log the signal intensity with a high rate have better potential than those that log scintillation measurements at a slow rate. The former can utilize the average *SI* and the variance in *SI* as extra features with respect to the slow logging receivers. Moreover, for fast logging receivers, $\{S_4, C/N_0, AZ, EL\}$ proved to be not crucial for the classification task. However, it is not clear yet if a receiver that logs $I$ and $Q$ has an advantage. Therefore, we devote the next section to answering this question. We consider only receivers that log measurements at 50 Hz and we exclude $\{S_4, C/N_0, AZ, EL\}$ from the features set.

## 8.5 Training Models for Receivers with High Logging Rate

In this section we further investigate the ML models for receivers that log measurements at 50 Hz. The objective is to compare models for receivers that log $I$ and $Q$, and those that log raw *SI* instead. We start from MDL$_4$ described in Table 8.2 that has the best performance despite of excluding the measurements $\{S_4, C/N_0, AZ, EL\}$. The measurements we consider in this section are $\{SI, \overline{SI}, \sigma_{SI}, cov(I^2, Q^2)\}$. We test 4 combinations of these measurements. We also investigate the choice of the value for the time window ($T_w$) over which the average, variance and covariance are calculated. We conclude the section showing results for testing the best of these models in different scenarios.

### 8.5.1 Machine Learning Features

The choice of the averaging time window considered for calculating the average, variance and co-variance features given by (8.1), (8.3) and (8.2) respectively, is investigated in this section. The considered values of the averaging windows $T_w \in \{1, 10, 30, 60, 90, 180\}s$, where 60 s is the typical time window used in literature to calculate $S_4$, while 180 s is the duration of a scintillation sample used by authors in similar studies [200]. So, we investigate the averaging windows around those two values.

Table 8.3 The trained models for fast logging receivers and the Features (measurements) considered for each

| Model | Features | Averaging Window | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 10 | 30 | 60 | 90 | 180 |
| $Model_1$ | $F_1 = \{SI, \overline{SI}, \sigma_{SI}, cov(I^2, Q^2)\}$ | variation (1) | variation (2) | variation (3) | variation (4) | variation (5) | variation (6) |
| $Model_2$ | $F_2 = \{SI, \overline{SI}, \sigma_{SI}\}$ | | | | | | |
| $Model_3$ | $F_3 = \{\overline{SI}, \sigma_{SI}, cov(I^2, Q^2)\}$ | | | | | | |
| $Model_4$ | $F_4 = \{\overline{SI}, \sigma_{SI}\}$ | | | | | | |

The feature sets we are investigating in this Section are shown in Table 8.3. The first column shows the model number. The features for each model are listed in the second column. $Model_1$ and $Model_3$ are applicable to receivers that log $I$ and $Q$, while $Model_2$ and $Model_4$ can be applied if the receiver is logging $I$ and $Q$ or logs $SI$ directly but in high rate.

We trained 24 bagged tree classifiers, with 30 learners each, using the training set of $\sim 36,000$ samples for each class as follows. First we formed the combination sets of sub–features shown in Table 8.3. We then prepared the training set 6 times, each time by calculating the average, variance and co–variance using the averaging windows of $T_w \in \{1, 10, 30, 60, 90, 180\}$ seconds. Therefore, the training for a Model$_x$ was carried using feature group $F_x$ where $x \in \{1,2,3,4\}$. This $Model_x$ had been investigated for 6 different values of $T_w$ previously introduced.

## 8.5.2 The Trained Models

In this section we report the results of testing the 24 models with the testing set. For scintillation detection, we are interested in two metrics: how many clean/multipath data were wrongly labelled as scintillation (false alarm), and how many scintillation data were wrongly classified as clean/multipath (miss–detection). We report here the testing accuracy (to give the big picture about the overall performance of the model) and the scintillation miss detection rate. We omit reporting the training accuracy and the scintillation false alarm rate because they were around 99% and 0% respectively for all the 24 models.

Figure 8.5 shows the overall testing accuracy (up) and the miss–detection rate of the scintillation class in particular (bottom) for the 24 models. The different line colors correspond to the 4 different feature sets listed in Table 8.3. The x–axis corresponds the 6 values of $T_w$ considered. The results shown in Figure 8.5 can be summarized as follows:

- All the models, except for $Model_1$ (blue) has similar performance in terms of accuracy and miss detection rate.
- Models with $T_w = 30, 60, 90s$ gave comparable good performance, with accuracy above 95% and miss detection rate below 2%.
- $Model_3$ (yellow) that is based on receivers that log $I$ and $Q$ has marginally better accuracy than the other two receivers that do not log $I$ and $Q$ (red and purple).
- Although $Model_4$ (purple) only uses two measurements to perform the classification task, it still has almost the same performance of $RX4$ which gave the best testing accuracy in the previous section as well as $Model_3$ which has the best accuracy in this section.

From this we draw the conclusion that if the receiver is logging $SI$ at a high rate, this measurement is enough to train a ML model to distinguish between scintillation, multipath and clean GNSS signals using the average $SI$ and the variance in $SI$ over $T_w$ between 30 and 90 seconds. Also, if $I$ and $Q$ are provided to the end user, this slightly increases the overall accuracy of the model, but it is not crucial.

In the next section we show further results of testing these models. For this purpose, we show only the results for $Model_3$, which has the highest overall accuracy, and we refer to it as the model. We also use the 60 s averaging window variation of the model, since it is the value in use in literature for calculating $S_4$.

### 8.5.3 Further Testing of $Model_3$

In this Section we show with more details the results of testing $Model_3$ using the novel testing set. Then, we demonstrate $Model_3$'s capabilities in distinguishing

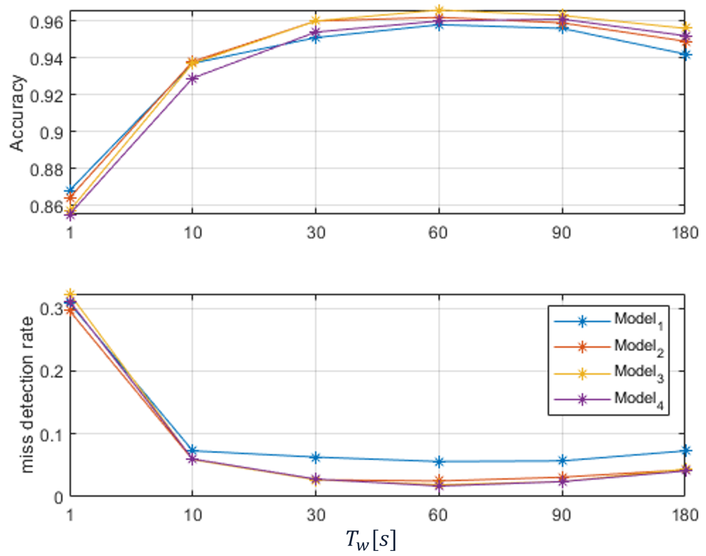
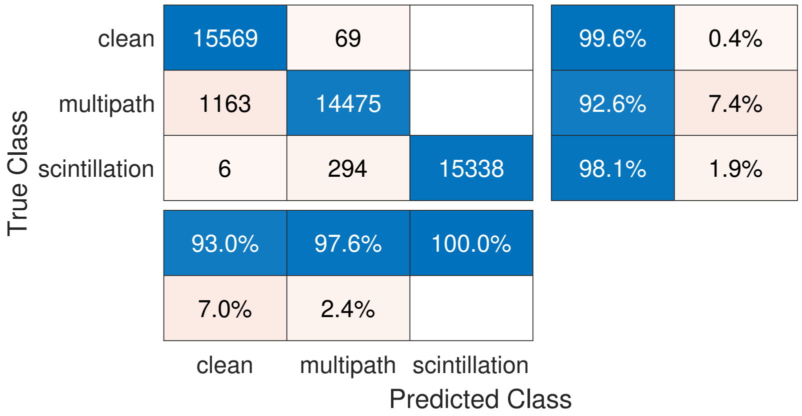


Fig. 8.5 Overall accuracy (up) and scintillation miss–detection (bottom) for the models in Table 8.3 trained with data with different $T_w$.

between scintillation and multipath using two scenarios i) detecting scintillation at low elevations, and ii) detecting multipath at above elevation threshold. These two scenarios challenge the model, because the first is the scintillation data that usually get discarded by the elevation angle threshold, while the second is the multipath data that raise false alarms in scintillation monitoring stations. We conclude the section by showing statistical results of a further testing of the model that we have done in operative scenarios.

### Confusion Matrix for $Model_3$

Figure 8.6 reports the confusion matrix obtained from testing $Model_3$ using the testing set. The rows represent the true (manual) label of the samples, while the columns are the label given by $Model_3$. It can be seen that the model gave no false alarm for scintillation. The miss-detection rate of scintillation is particularly low, with only 1.9% of the scintillation data miss-detected. The 0.0% false alarm for the scintillation class indicates that the model only labels a sample as scintillation if it is truly scintillation. This could be beneficial for post analysis of scintillation data, and for data storage systems with low false alarms. However, the model wrongly classified multipath as clean signals with miss detection rate of 7.4%. Since our



| True Class \ Predicted Class | clean | multipath | scintillation | | |
|---|---|---|---|---|---|
| clean | 15569 | 69 | | 99.6% | 0.4% |
| multipath | 1163 | 14475 | | 92.6% | 7.4% |
| scintillation | 6 | 294 | 15338 | 98.1% | 1.9% |
| | 93.0% | 97.6% | 100.0% | | |
| | 7.0% | 2.4% | | | |

Fig. 8.6 The confusion Matrix of the testing dataset using $Model_3$.

objective is to detect scintillation, and to distinguish scintillation from multipath in particular, this mix between the clean and multipath signals was acceptable for this work, and further investigating it can be the objective of a future work. Overall, the testing of the model with novel data verifies the trained model, and thus we can rely on it to distinguish multipath from scintillation in operative scenarios.

### Detecting Scintillation at Low Elevation

Figure 8.7 shows an example of detecting scintillation at low elevation for GPS signal PRN30. This signal was not part of the training or testing sets. Here, the signal intensity is rapidly fluctuating ($2^{\text{nd}}$ plot from top), leading to high $S_4$ ($4^{\text{th}}$ plot from top). However, the elevation of this signal is below 30° (bottom plot) and thus this data will probably be discarded by a scintillation monitoring station that relies on elevation masks to avoid multipath. The manual label for all this data set is scintillation, after consulting data from consecutive days. Our model correctly classifies this data as scintillation as shown in the top plot of Figure 8.7, with a couple of miss–classified signals as multipath. Using our model, scintillation monitoring could gain more scintillation data by including low elevation measurements. In fact, in our testing dataset, by including low elevation data, we gained 17.5% more scintillation data that would have been discarded by a 30° elevation mask.

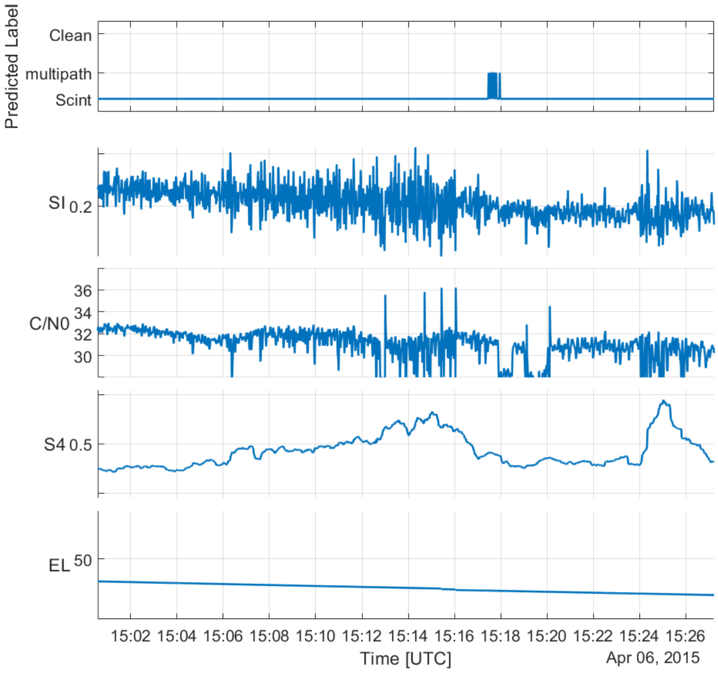


Fig. 8.7 Detecting Scintillation at Low Elevation.

### Detecting Multipath at Above Elevation Threshold

In Figure 8.8 we show an example of a multipath signal. Here, the $S_4$ index is inflated (second plot from bottom) due to multipath up to an elevation angle of 35° (bottom plot) at 14:33 UTC. In fact, our model suspects all the data up to 14:45 UTC to be multipath affected (top plot), which means the slight enhancement on $S_4$ at 40° elevation is due to multipath. Using the model will thus reduce false scintillation alarms. This in return will lead to better automation of scintillation monitoring activities and better management of scintillation data repositories.

### Statistical Results of Operative Scenarios.

Now that we have tested the model using labelled data, we want to test how it will perform in real operative scenarios. In operative scenarios we will provide the whole log file as it was generated by the receiver to the model. The log files we have can be up to four hours. One log file is generated per each visible satellite in the observation window. The satellite signal might suffer multipath as the satellite is

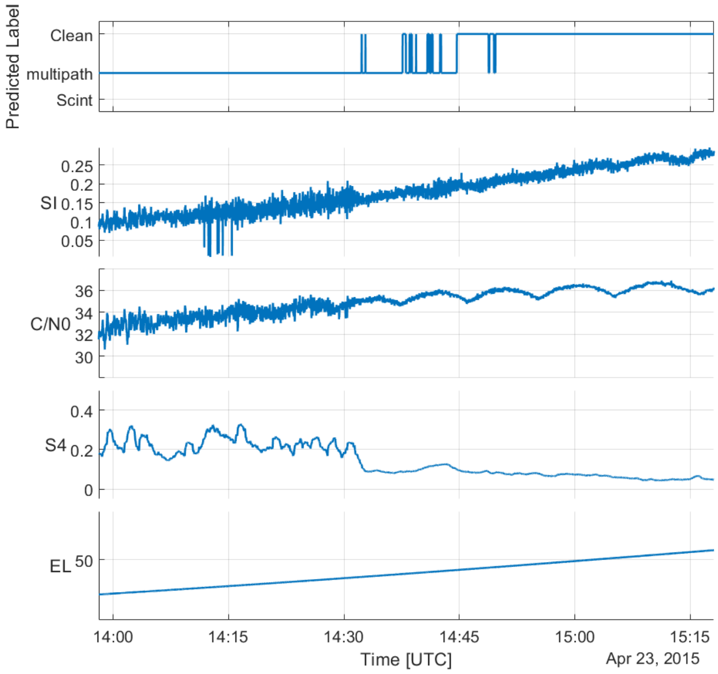


Fig. 8.8 Detecting Multipath.

rising (or setting), then the signal becomes clean as the satellite rises above the obstacles. Scintillation might affect the signal at any elevation. The model will decide the label second by second for the whole file. Forty seven log files were picked for this operative scenarios testing set, and labelled using the ML model. The results were then visually inspected to verify the results. Discarding the miss-classification between multipath affected and clean signals, we found that only two of the scintillation events were wrongly classified as multipath, while no multipath events raised false scintillation alarms. This agrees with the model performance reported above.

In Figure 8.9 we report one of the two miss detected scintillation events which is PRN06 on March, 14th. This is a strong scintillation event that started when the satellite was at low elevation, followed by moderate-to-low scintillation. On the top plot we report the second by second label given by the model. The model was able to capture the strong scintillation event, but not the moderate scintillation which was wrongly classified as multipath. However, on the same data set there were moderate-to-low scintillations and they were all correctly classified as scintillation.

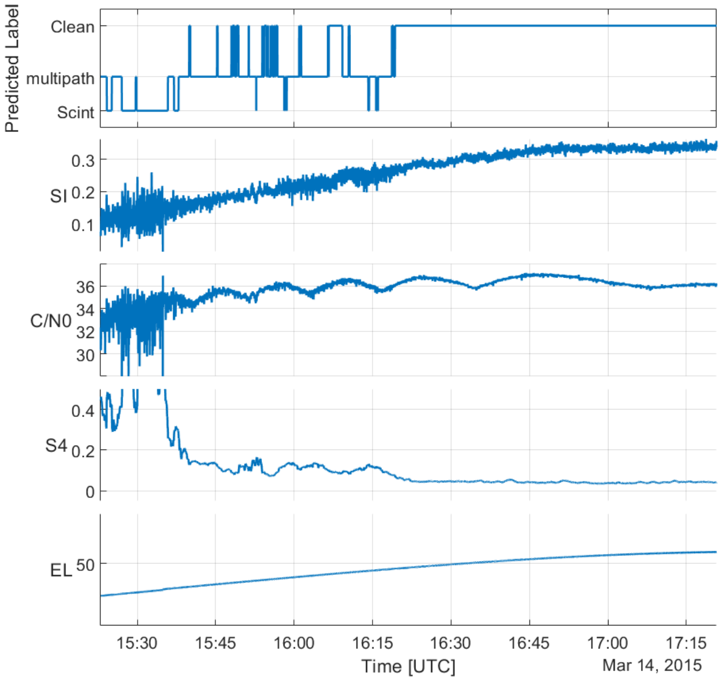


Fig. 8.9 Example of scintillation miss detection.

## 8.6 Summary

In this chapter, we showed that ML models can be trained to distinguish between scintillation, multipath and clean GNSS signals for scintillation monitoring receivers that log the measurements at low rate as well as for receivers that log at high rate. The measurements considered are $\{SI, S_4, C/N_0, AZ, EL\}$ for the slow logging receiver, and $\{SI, S_4, C/N_0, AZ, EL, I, Q, \overline{SI}, \sigma_{SI}, cov(I^2, Q^2)\}$ for the fast logging one. We showed that in particular for the latter, models that do not rely on $\{S_4, C/N_0, AZ, EL\}$ can achieve the best performance.

Furthermore, we investigated if fast logging receivers that log $I$ and $Q$ have advantage over the ones that do not log $I$ and $Q$. We found that if I and Q are available, the covariance between $I^2$ and $Q^2$ can be an extra feature that will increase the accuracy of the model. However, this enhancement in the accuracy is not big enough to give receivers that log $I$ and $Q$ great advantage over receivers that do not log $I$ and $Q$. For the latter, the average and variance of the signal intensity are enough to distinguish between scintillation and multipath. The averaging window was also investigated, and found that averaging windows between 30-90 s give comparable good performance.

The accuracy of our best model, which uses as inputs $\{\overline{SI}, \sigma_{SI}, cov(I^2, Q^2)\}$ calculated over 60s time window, is 96% with miss–detection rate on the scintillation class less than 2% and no false alarms when tested with a novel data set. In this dataset, 17.5% scintillation data gain was observed, resulting from including scintillation data that would have been discarded by a 30° elevation mask. Moreover, two examples of challenging the model at low elevations were shown, one for detecting scintillation at low elevations, and the other for identifying multipath signals that would have triggered a false alarm for a 30° elevation threshold system. Finally, the model was tested in operative scenario with data from 5 days in March, April and October. Here the model labelled the data, second by second, for all the visible satellites. By visual inspection, we found that the model still achieved 95.7% accuracy for this test,which further validated the model.

# Final Remarks

# Chapter 9

# Conclusions and Final Remarks

Two state-of-the-art utilizations of GNSS signals for environmental monitoring were discussed in this thesis. The first is utilizing GNSS–R for floods assessment using UAV-borne GNSS–R sensors and the second is applying ML techniques for detecting ionospheric scintillations.

For GNSS–R, data from a custom built GNSS–R sensor were utilized to investigate the feasibility of using such sensors in flood detection. The findings and recommendations are summarized as follows:

- The signal processing methodology, which was developed in this work to be implemented on sensors that will be mounted on–board small platforms, was detailed. The reflected signal processing was aided with Doppler and delay measurements from the direct signal processing. The peak of the reflected signal above the noise floor was roughly estimated. The signal processing methodology was shown to be adequate for detecting water from such sensors.
- Comparison between water detection identified by the GNSS–R and orthophotomaps was provided as a qualitative investigation of the methods applied. It was shown that data from GNSS–R sensors have good agreement with maps and the water edges in particular are well detected by the GNSS–R sensor.
- As a quantitative metric, the accuracy in estimating the water extends from GNSS–R measurements was analysed and compared to the ground truth data. It was found that the estimated area is 7.5% less than the ground truth. Such metric could be improved by optimizing the flight trajectory. Flight trajectory

optimization is a common practice in UAVs and it was out of the scope of this work. The estimation can be improved also using multi–GNSS signals.

- The advantages of using multi–GNSS in GNSS–R were investigated by observing the improvement in water area estimation with the increase in the number of visible satellites. For this, a simulation that increases the number of utilized signals, to simulate increasing the utilized GNSS constellations by the sensor, was performed. It was shown that 98% accuracy in estimating the water surface area is achievable with multi–GNSS. Here, again the flight trajectory was not optimized, however such optimization is expected to be beneficial for future studies for two reasons. The first is that the geometry distribution of the satellites in the sky might indicate optimized data collection at certain times. The second is the shape of the water body might require planning the UAV trajectory in a way that maximized the specular points locations to fall on certain water bodies (for example rivers).
- The feasibility of estimating the flight height above terrain utilizing the GNSS–R data was investigated. Such investigation is important for estimating the change in the water height in floods detection. It was shown that for such measurement, the height above the reflecting surface, the sensor has to be calibrated. For the water detection task the calibration was not needed.
- For the height estimation applications, it was also demonstrated that the signal processing technique needs to fine estimate the delay of the reflected signal. Such fine estimation was not needed for the water detection task, and a simplified GNSS–R receiver proved to be adequate for the detection task.
- Finally, for estimating the amount of the reflected power, calibrating the sensor is essential, as demonstrated in this work, and it was beyond the planned experiments on this sensor.

For scintillation, data from ionospheric scintillation monitoring receivers were utilized to train ML models able to detect scintillation with accuracy that exceeds the traditional metrics performance. Equatorial and polar scintillations were considered separately. The results and recommendations of this work are summarized as follows:

- A collection of the widely used supervised ML algorithms that has shown good performance in ML modelling literature were explored. Three algorithms (decision trees, SVMs and k-NN) were considered. Ensemble techniques as well as various models' hyper–parameters tunings were investigated (in total, 17 ML models were compared). The comparison was carried out on polar scintillation data and it was found that the bagged trees give superior performance among all the models with 95% accuracy, 5% scintillation miss detection, and 5% scintillation false alarm. From this investigation, the bagged tree was adopted as the method to train the models for detecting equatorial scintillation too.

- Various GNSS measurements were investigated as inputs to the ML models. For the polar scintillations, the data contained in the ionospheric scintillation monitoring records were investigated. It was found that TEC measurements are capable of detecting scintillation when the ML is trained on samples of 3–minutes of TEC data. TEC gave better results than the scintillation metrics themselves. This result invites investigating TEC measurements from non-ISM receivers for detecting scintillations. Such receivers will open the door for exploring data from GNSS networks wider than the ISM receivers networks.

- For equatorial scintillation, data contained in the ISM records as well as post-correlation measurements available at high rate from commercial ISM receivers and from SDR receivers were investigated. Cases where multipath is known to affect the scintillation data were considered. It was found that for such scenarios, the high rate post-correlation data are better than the scintillation metrics in detecting scintillation. In particular, the bagged tree model trained for these data is able to detect scintillation with 96% accuracy, 2% scintillation miss detection and 2% scintillation false alarm.

- Finally, It was demonstrated that such models trained to detect scintillation even in presence of interference from the environment, multipath in this case, are able to increase the amount of useful scintillation data by detecting scintillations that are usually discarded by the conservative elevation masks. A total of 17% scintillation data gain in the data set was achieved.

# References

[1] Mark Frederick Knight. *Ionospheric scintillation effects on global positioning system receivers*. PhD thesis, The University of Adelaide, 2000.

[2] Joseph L Awange. *Environmental monitoring using GNSS: Global navigation satellite systems*. Springer, 2012.

[3] Demoz Gebre-Egziabher and Scott Gleason. *GNSS applications and methods*. Artech House, 2009.

[4] Weihua Bai, Nan Deng, Yueqiang Sun, Qifei Du, Junming Xia, Xianyi Wang, Xiangguang Meng, Danyang Zhao, Congliang Liu, Guangyuan Tan, Ziyan Liu, and Xiaoxu Liu. Applications of GNSS-RO to numerical weather prediction and tropical cyclone forecast. *Atmosphere*, 11(11), 2020.

[5] Jin Xing, Baoguo Yu, Dongkai Yang, Jie Li, Zhejia Shi, Guodong Zhang, and Feng Wang. A real-time GNSS-R system for monitoring sea surface wind speed and significant wave height. *Sensors*, 22(10), 2022.

[6] 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. *Remote Sensing*, 13(16), 2021.

[7] F. Soulat, M. Caparrini, O. Germain, P. Lopez-Dekker, M. Taani, and G. Ruffini. Sea state monitoring using coastal GNSS-R. *Geophysical Research Letters*, 31(21), 2004.

[8] April Warnock and Christopher Ruf. Response to variations in river flowrate by a spaceborne GNSS-R river width estimator. *Remote Sensing*, 11(20), 2019.

[9] Xuerui Wu, Wenxiao Ma, Junming Xia, Weihua Bai, Shuanggen Jin, and Andrés Calabia. Spaceborne GNSS-R soil moisture retrieval: Status, development opportunities, and challenges. *Remote Sensing*, 13(1), 2021.

[10] Komi Edokossi, Andres Calabia, Shuanggen Jin, and Iñigo Molina. GNSS-Reflectometry and remote sensing of soil moisture: A review of measurement techniques, methods, and applications. *Remote Sensing*, 12(4), 2020.

[11] Meshal Alshaye, Faisal Alawwad, and Ibrahim Elshafiey. Hurricane tracking using multi-GNSS-R and deep learning. In *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, pages 1–4, 2020.

[12] Qingyun Yan and Weimin Huang. Sea ice remote sensing using GNSS-R: A review. *Remote Sensing*, 11(21), 2019.

[13] Wentao Yang, Fan Gao, Tianhe Xu, Nazi Wang, Jinsheng Tu, Lili Jing, and Yahui Kong. Daily flood monitoring based on spaceborne GNSS-R data: A case study on Henan, China. *Remote Sensing*, 13(22), 2021.

[14] Cinzia Zuffada, Brandi Downs, Ilaria Mara Russo, Eric Loria, Andrew O'Brien, Carmela Galdi, Maurizio di Bisceglie, Valery Zavorotny, Marco Lavalle, and Mary Morris. State of the art in GNSS-R capabilities over inland waters. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 950–953, 2021.

[15] Mahmoud Rajabi, Hossein Nahavandchi, and Mostafa Hoseini. Evaluation of CYGNSS observations for flood detection and mapping during Sistan and Baluchestan torrential rain in 2020. *Water*, 12(7), 2020.

[16] S L Kesav Unnithan, Basudev Biswal, and Christoph Rüdiger. Flood inundation mapping by combining GNSS-R signals with topographical information. *Remote Sensing*, 12(18), 2020.

[17] Nereida Rodriguez-Alvarez, Erika Podest, Katherine Jensen, and Kyle C. McDonald. Classifying inundation in a tropical wetlands complex with GNSS-R. *Remote Sensing*, 11(9), 2019.

[18] Clara C Chew, Eric E Small, Samir Belabbes, and Laura Read. Opportunities and challenges in flood monitoring using GNSS-R. In *AGU Fall Meeting Abstracts*, volume 2019, pages IN14A–03, 2019.

[19] Clara Chew and Eric Small. Mapping flooding and inundation dynamics using spaceborne GNSS-R observations. Technical report, Copernicus Meetings, 2022.

[20] Kegen Yu, Shuai Han, Jinwei Bu, Yuhang An, Zhewen Zhou, Changyang Wang, Sajad Tabibi, and Joon Wayn Cheong. Spaceborne GNSS reflectometry. *Remote Sensing*, 14(7), 2022.

[21] Scott Gleason, Andrew O'Brien, Anthony Russel, Mohammad M. Al-Khaldi, and Joel T. Johnson. Geolocation, calibration and surface resolution of CYGNSS GNSS-R land observations. *Remote Sensing*, 12(8), 2020.

[22] Guanglin Yang, Weihua Bai, Jinsong Wang, Xiuqing Hu, Peng Zhang, Yueqiang Sun, Na Xu, Xiaochun Zhai, Xianjun Xiao, Junming Xia, Feixiong Huang, Cong Yin, Qifei Du, Xianyi Wang, Yuerong Cai, Xiangguang Meng, Guangyuan Tan, Peng Hu, and Congliang Liu. FY3E GNOS II GNSS reflectometry: Mission review and first results. *Remote Sensing*, 14(4), 2022.

[23] Cheng Jing, Xinliang Niu, Chongdi Duan, Feng Lu, Guodong Di, and Xiaofeng Yang. Sea surface wind speed retrieval from the first chinese GNSS-R mission: Technique and preliminary results. *Remote Sensing*, 11(24), 2019.

[24] Elliott Kaplan and Christopher Hegarty. *Understanding GPS: principles and applications*. Artech house, 2005.

[25] J. Feltens. The International GPS Service (igs) ionosphere working group. *Advances in Space Research*, 31(3):635–644, 2003. Description of the Low Latitude and Equatorial Ionosphere in the International Reference Ionosphere.

[26] Michael Mendillo. Storms in the ionosphere: Patterns and processes for total electron content. *Reviews of Geophysics*, 44(4), 2006.

[27] Takuya Tsugawa, Nobuki Kotake, Yuichi Otsuka, and Akinori Saito. Medium-scale traveling ionospheric disturbances observed by GPS receiver network in Japan: A short review. *GPS Solutions*, 11(2):139–144, 2007.

[28] PM Kintner, Todd Humphreys, Joanna Hinks, et al. GNSS and ionospheric scintillation. *Inside GNSS*, 4(4):22–30, 2009.

[29] X Pi, AJ Mannucci, UJ Lindqwister, and CM Ho. Monitoring of global ionospheric irregularities using the worldwide GPS network. *Geophysical Research Letters*, 24(18):2283–2286, 1997.

[30] Micaela Troglia Gamba, Gianluca Marucco, Marco Pini, Sabrina Ugazio, Emanuela Falletti, and Letizia Lo Presti. Prototyping a GNSS-based passive radar for UAVs: An instrument to classify the water content feature of lands. *Sensors*, 15(11):28287–28313, 2015.

[31] G. Povero, M. Pini, F. Dovis, R. Romero, P. Abadi, L. Alfonsi, L. Spogli, D. Di Mauro, Le Huy Minh, La The Vinh, and N. Floury. Ionosphere monitoring in South East Asia: Activities in GINESTRA and ERICA projects. In *2015 International Association of Institutes of Navigation World Congress (IAIN)*, pages 1–7, 2015.

[32] Luca Spogli, Claudio Cesaroni, Domenico Di Mauro, Michael Pezzopane, Lucilla Alfonsi, Elvira Musicò, Gabriella Povero, Marco Pini, Fabio Dovis, Rodrigo Romero, Nicola Linty, Prayitno Abadi, Fitri Nuraeni, Asnawi Husin, Minh Le Huy, Tran Thi Lan, The Vinh La, Valdir Gil Pillat, and Nicolas Floury. Formation of ionospheric irregularities over Southeast Asia during the 2015 St. Patrick's Day storm. *Journal of Geophysical Research: Space Physics*, 121(12):12,211–12,233, 2016.

[33] L. Alfonsi, P. J. Cilliers, V. Romano, I. Hunstad, E. Correia, N. Linty, F. Dovis, O. Terzo, P. Ruiu, J. Ward, and P. Riley. First observations of GNSS ionospheric scintillations from DemoGRAPE project. *Space Weather*, 14(10):704–709, 2016.

[34] J. Wolfrum, M. Healy, J. P. Provenzano, and T. Sassorossi. Galileo – Europe's contribution to the next generation of GNSS. In *Proceedings of the 12th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1999)*, pages 1381 – 1390, 1999.

[35] CL Rino, MD Cousins, and JA Klobuchar. Amplitude and phase scintillation measurements using the Global Positioning System. In *In: Symposium on the Effect of the Ionosphere on Radiowave Systems*, 1981.

[36] AJ Van Dierendonck, John Klobuchar, and Quyen Hua. Ionospheric scintillation monitoring using commercial single frequency C/A code receivers. In *proceedings of ION GPS*, volume 93, pages 1333–1342, 1993.

[37] C.D. Hall and R.A. Cordey. Multistatic scatterometry. In *International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Moving Toward the 21st Century'.*, volume 1, pages 561–562, 1988.

[38] Manuel Martin-Neira et al. A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry. *ESA journal*, 17(4):331–355, 1993.

[39] Katzberg Stephen J. and Jr James L. Garrison. Utilizing GPS to determine ionospheric delay over the ocean. Technical report, 1996.

[40] J.L. Garrison and S.J. Katzberg. Detection of ocean reflected GPS signals: theory and experiment. In *Proceedings IEEE SOUTHEASTCON '97. 'Engineering the New Century'*, pages 290–294, 1997.

[41] S. Gleason, S. Hodgart, Yiping Sun, C. Gommenginger, S. Mackin, M. Adjrad, and M. Unwin. Detection and processing of bistatically reflected GPS signals from low Earth orbit for the purpose of ocean remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1229–1241, 2005.

[42] Zhehao Zhang and Lin Pan. Current performance of open position service with almost fully deployed multi-GNSS constellations: GPS, GLONASS, Galileo, BDS-2, and BDS-3. *Advances in Space Research*, 69(5):1994–2019, 2022.

[43] GPS space segment. https://www.gps.gov/systems/gps/space/, June 2022.

[44] GLONASS constellation status. https://www.glonass-iac.ru/en/sostavOG/, June 2022.

[45] Galileo constellation information. https://www.gsc-europa.eu/system-service-status/constellation-information, June 2022.

[46] BeiDou system status. http://www.csno-tarc.cn/en/system/introduction, June 2022.

[47] Thomas Purfürst. Evaluation of static autonomous GNSS positioning accuracy using single-, dual-, and tri-frequency smartphones in forest canopy environments. *Sensors*, 22(3), 2022.

[48] Elena Simona Lohan and Kai Borre. Accuracy limits in multi-GNSS. *IEEE Transactions on Aerospace and Electronic Systems*, 52(5):2477–2494, 2016.

[49] V.K.D. Srinivasu, N. Dashora, D.S.V.V.D. Prasad, K. Niranjan, and S. Gopi Krishna. On the occurrence and strength of multi-frequency multi-GNSS ionospheric scintillations in Indian sector during declining phase of solar cycle 24. *Advances in Space Research*, 61(7):1761–1775, 2018. Studies on Mesosphere, Thermosphere and Ionosphere from Equatorial to Mid Latitudes - Recent Investigations and Improvements - Part 2.

[50] Khanitin Seechai, Punyawi Jamjareegulgarn, Ningbo Wang, and Pornchai Supnithi. Statistical analysis of scintillation index and ROTI based on multi-GNSS data at chumphon, thailand. In *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, pages 1–4, 2021.

[51] Jun Wang, Y. Jade Morton, and Donald Hampton. New results on ionospheric irregularity drift velocity estimation using multi-GNSS spaced-receiver array during high-latitude phase scintillation. *Radio Science*, 53(2):228–240, 2018.

[52] GNSS Constellations, Radio Frequencies and Signals. https://www.tallysman.com/gnss-constellations-radio-frequencies-and-signals/, April 2022.

[53] GNSS Frequencies and Signals. https://novatel.com/support/known-solutions/gnss-frequencies-and-signals, April 2022.

[54] Jean-Claude Auber, Alain Bibaut, and J.-M. Rigal. Characterization of multipath on land and sea at GPS frequencies. In *Proceedings of the 7th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1994)*, pages 1155–1171, 1994.

[55] Su-Kyung Kim and Jihye Park. Monitoring a storm surge during Hurricane Harvey using multi-constellation GNSS-Reflectometry. *GPS Solutions*, 25(63), 2021.

[56] Víctor Puente and Marcelino Valdés. Sea level determination in the Spanish coast using GNSS-R. *Proceedings*, 19(1), 2019.

[57] Patrizia Tavella and Gérard Petit. Precise time scales and navigation systems: mutual benefits of timekeeping and positioning. *Satellite Navigation*, 1:1–12, 2020.

[58] Gregor Möller and Daniel Landskron. Atmospheric bending effects in GNSS tomography. *Atmospheric Measurement Techniques*, 12(1):23–34, 2019.

[59] Malek Karaim, Mohamed Elsheikh, Aboelmagd Noureldin, and RB Rustamov. GNSS error sources. *Multifunctional Operation and Application of GPS0*, pages 69–85, 2018.

[60] J T Curran, Michele Bavaro, Aiden Morrison, and Joaquim Fortuny. Developing an ionospheric scintillation monitoring receiver. *Inside GNSS*, Sep 2014.

[61] Jordi Vilà-Valls, Nicola Linty, Pau Closas, Fabio Dovis, and James T. Curran. Survey on signal processing for GNSS under ionospheric scintillation: Detection, monitoring, and mitigation. *NAVIGATION: Journal of the Institute of Navigation*, 67(3):511–536, 2020.

[62] Sien-Chong Wu, Thomas Meehan, and Larry Young. Potential use of GPS signals as ocean altimetry observables. In Anon, editor, *Proceedings of the National Technical Meeting, Institute of Navigation*, pages 543–550. Inst of Navigation, Alexandria, VA, United States, 1997.

[63] DS Masters. *Surface remote sensing applications of GNSS bistatic radar: Soil moisture and aircraft altimetry*. PhD thesis, University of Colorado, 2004.

[64] Lucinda King, Martin Unwin, Jonathan Rawlinson, Raffaella Guida, and Craig Underwood. Towards a topographically-accurate reflection point prediction algorithm for operational spaceborne GNSS reflectometry—development and verification. *Remote Sensing*, 13(5), 2021.

[65] Guy J.-P. Schumann. *Remote Sensing of Floods*. Oxford University Press, 04 2017.

[66] G.J.-P. Schumann and D.K. Moller. Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth*, 83-84:84–95, 2015.

[67] M.S. Rahman and L. Di. The state of the art of spaceborne remote sensing in flood management. *Natural Hazards*, 85(2):1223–1248, 2017.

[68] H. Cao, H. Zhang, C. Wang, and B. Zhang. Operational flood detection using Sentinel-1 SAR data over large areas. *Water (Switzerland)*, 11(4), 2019.

[69] M. Guo, J. Li, C. Sheng, J. Xu, and L. Wu. A review of wetland remote sensing. *Sensors (Switzerland)*, 17(4), 2017.

[70] Y.N. Lin, S.-H. Yun, A. Bhardwaj, and E.M. Hill. Urban flood detection with Sentinel-1multi-temporal Synthetic Aperture Radar (SAR) observations in a Bayesian framework: A case study for Hurricane Matthew. *Remote Sensing*, 11(15), 2019.

[71] X. Shen, D. Wang, K. Mao, E. Anagnostou, and Y. Hong. Inundation extent mapping by synthetic aperture radar: A review. *Remote Sensing*, 11(7), 2019.

[72] Z. Kugler, S.V. Nghiem, and G.R. Brakenridge. L-band passive microwave data from SMOS for river gauging observations in tropical climates. *Remote Sensing*, 11(7), 2019.

[73] C. Chew, J.T. Reager, and E. Small. CYGNSS data map flood inundation during the 2017 Atlantic hurricane season. *Scientific Reports*, 8(1), 2018.

[74] D.E. Alsdorf, E. Rodríguez, and D.P. Lettenmaier. Measuring surface water from space. *Reviews of Geophysics*, 45(2), 2007.

[75] A. Benoudjit and R. Guida. A novel fully automated mapping of the flood extent on SAR images using a supervised classifier. *Remote Sensing*, 11(7), 2019.

[76] T. Lacava, V. Cuomo, E.V. Di Leo, N. Pergola, F. Romano, and V. Tramutoli. Improving soil wetness variations monitoring from passive microwave satellite data: The case of April 2000 Hungary flood. *Remote Sensing of Environment*, 96(2):135–148, 2005.

[77] A. Refice, D. Capolongo, G. Pasquariello, A. Daaddabbo, F. Bovenga, R. Nutricato, F.P. Lovergine, and L. Pietranera. SAR and InSAR for flood monitoring: Examples with COSMO-SkyMed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(7):2711–2722, 2014.

[78] M. Faruolo, I. Coviello, T. Lacava, N. Pergola, and V. Tramutoli. A multi-sensor exportable approach for automatic flooded areas detection and monitoring by a composite satellite constellation. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):2136–2149, 2013.

[79] D.C. Mason, M.S. Horritt, J.T. Dall'Amico, T.R. Scott, and P.D. Bates. Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3932–3943, 2007.

[80] S.V. Nghiem, C. Zuffada, R. Shah, C. Chew, S.T. Lowe, A.J. Mannucci, E. Cardellach, G.R. Brakenridge, G. Geller, and A. Rosenqvist. Wetland monitoring with global navigation satellite system reflectometry. *Earth and Space Science*, 4(1):16–39, 2017.

[81] J.L. Garrison, S.J. Katzberg, and M.I. Hill. Effect of sea roughness on bistatically scattered range coded signals from the Global Positioning System. *Geophysical Research Letters*, 25(13):2257–2260, 1998.

[82] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger. Spaceborne GNSS-R minimum variance wind speed estimator. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):6829–6843, Nov 2014.

[83] J. Beckheinrich, G. Beyerle, S. Schoen, H. Apel, M. Semmling, and J. Wickert. WISDOM: GNSS-R based flood monitoring. In *2012 Workshop on Reflectometry Using GNSS and Other Signals of Opportunity, GNSS+R 2012*, 2012.

[84] S. Esterhuizen, D. Masters, D. Akos, and E. Vinande. Experimental characterization of land-reflected GPS signals. In *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation, ION GNSS 2005*, volume 2005, pages 1670–1678, 2005.

[85] J. Kainulainen, S. Salo, J. Lahtinen, G. M. Calves, J. Seppanen, J. Praks, T. Hakala, Y. Chen, J. Hyyppa, M. Unwin, P. Jales, G. Ressler, T. Casal, and J. Rosello. Airborne wind vector scatterometer for sea surface measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–7, 2019.

[86] Y. Pei, R. Notarpietro, P. Savi, M. Cucca, and F. Dovis. A fully software Global Navigation Satellite System reflectometry (GNSS-R) receiver for soil monitoring. *International Journal of Remote Sensing*, 35(6):2378–2391, 2014.

[87] Y. Pei, R. Notarpietro, S. De Mattia, P. Savi, F. Dovis, and M. Pini. Remote sensing of soil based on a compact and fully software GNSS-R receiver. In *26th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS 2013*, volume 1, pages 56–61. Institute of Navigation, 2013.

[88] D. Masters, S. Katzberg, and P. Axelrad. Airborne GPS bistatic radar soil moisture measurements during SMEX02. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 2, pages 896–898, 2003.

[89] L.-C. Shen, J.-C. Juang, C.-L. Tsai, C.-C. Chang, P.-Y. Ko, and C.-L. Tseng. Stream soil moisture estimation by reflected GPS signals with ground truth measurements. *IEEE Transactions on Instrumentation and Measurement*, 58(3):730–737, 2009.

[90] A. Helm. Ground-based GPS altimetry with the L1 OpenGPS receiver using carrier phase-delay observations of reflected GPS signals. Technical Report STR 08/10, Deutsches GeoForschungsZentrum GFZ, Berlin, oct 2008.

[91] E. Vinande, D. Akos, D. Masters, P. Axelrad, and S. Esterhuizen. GPS bistatic radar measurements of aircraft altitude and ground objects with a software receiver. In *Proceedings of the Annual Meeting - Institute of Navigation*, pages 528–534, 2005.

[92] E.E. Small, K.M. Larson, and J.J. Braun. Sensing vegetation growth with reflected GPS signals. *Geophysical Research Letters*, 37(12), 2010.

[93] Mehrez Zribi, Erwan Motte, Nicolas Baghdadi, Frédéric Baup, Sylvia Dayau, Pascal Fanise, Dominique Guyon, Mireille Huc, and Jean Pierre Wigneron. Potential applications of GNSS-R observations over agricultural areas: Results from the GLORI airborne campaign. *Remote Sensing*, 10(8), 2018.

[94] N. Rodriguez-Alvarez, A. Aguasca, E. Valencia, X. Bosch-Lluis, A. Camps, I. Ramos-Perez, H. Park, and M. Vall-Llossera. Snow thickness monitoring using GNSS measurements. *IEEE Geoscience and Remote Sensing Letters*, 9(6):1109–1113, 2012.

[95] M Unwin, Scott Gleason, and Michael Brennan. 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)*, pages 2656 – 2663, Portland, OR, September 2003.

[96] C. Ruf, M. Unwin, J. Dickinson, R. Rose, D. Rose, M. Vincent, and A. Lyons. CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):52–67, June 2013.

[97] J. Wickert, E. Cardellach, M. Martín-Neira, J. Bandeiras, L. Bertino, O. B. Andersen, A. Camps, N. Catarino, B. Chapron, F. Fabra, N. Floury, G. Foti, C. Gommenginger, J. Hatton, P. Høeg, A. Jäggi, M. Kern, T. Lee, Z. Li, H. Park, N. Pierdicca, G. Ressler, A. Rius, J. Roselló, J. Saynisch, F. Soulat, C. K. Shum, M. Semmling, A. Sousa, J. Xie, and C. Zuffada. GEROS-ISS: GNSS REflectometry, Radio Occultation, and Scatterometry onboard the International Space Station. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(10):4552–4581, Oct 2016.

[98] Giuseppe Foti, Christine Gommenginger, Philip Jales, Martin Unwin, Andrew Shaw, Colette Robertson, and Josep Roselló. Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission. *Geophysical Research Letters*, 42(13):5435–5441, 2015.

[99] Weiqiang Li, Estel Cardellach, Fran Fabra, Serni Ribó, and Antonio Rius. Lake level and surface topography measured with spaceborne GNSS-Reflectometry from CYGNSS mission: Example for the lake Qinghai. *Geophysical Research Letters*, 45(24):13,332–13,341, 2018.

[100] Hugo Carreno-Luengo, Stephen Lowe, Cinzia Zuffada, Stephan Esterhuizen, and Shadi Oveisgharan. Spaceborne GNSS-R from the SMAP mission: First assessment of polarimetric scatterometry over land and cryosphere. *Remote Sensing*, 9(4), 2017.

[101] M. P. Clarizia, N. Pierdicca, F. Costantini, and N. Floury. Analysis of CYGNSS data for soil moisture retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–9, 2019.

[102] M. M. Al-Khaldi, J. T. Johnson, A. J. O'Brien, A. Balenzano, and F. Mattia. Time-series retrieval of soil moisture using CYGNSS. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–10, 2019.

[103] Ken Whitehead and Chris H. Hugenholtz. Remote sensing of the environment with small unmanned aircraft systems (uass), part 1: a review of progress and challenges. *Journal of Unmanned Vehicle Systems*, 02(03):69–85, 2014.

[104] Ken Whitehead, Chris H. Hugenholtz, Stephen Myshak, Owen Brown, Adam LeClair, Aaron Tamminga, Thomas E. Barchyn, Brian Moorman, and Brett Eaton. Remote sensing of the environment with small unmanned aircraft systems (UASs), part 2: scientific and commercial applications. *Journal of Unmanned Vehicle Systems*, 02(03):86–102, 2014.

[105] A.D. Tamminga, B.C. Eaton, and C.H. Hugenholtz. UAS-based remote sensing of fluvial change following an extreme flood event. *Earth Surface Processes and Landforms*, 40(11):1464–1476, 2015.

[106] Q. Feng, J. Liu, and J. Gong. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier-A case of yuyao, china. *Water (Switzerland)*, 7(4):1437–1455, 2015.

[107] I. Colomina and P. Molina. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92:79–97, 2014.

[108] P.A. Rosen, S. Hensley, K. Wheeler, G. Sadowy, T. Miller, S. Shaffer, R. Muellerschoen, C. Jones, H. Zebker, and S. Madsen. Uavsar: A new NASA airborne SAR system for science and technology research. In *IEEE National Radar Conference - Proceedings*, pages 22–29, 2006.

[109] P.A. Rosen, S. Heasley, K. Wheeler, G. Sadowy, T. Miller, S. Shaffer, R. Muellerschoen, C. Jones, S. Madsen, and H. Zebker. UAVSAR: New NASA airborne SAR system for research. *IEEE Aerospace and Electronic Systems Magazine*, 22(11):21–28, 2007.

[110] Stephan Esterhuizen and Dennis Akos. The design, construction, and testing of a modular GPS bistatic radar software receiver for small platforms. Master's thesis, University of Colorado, 2006.

[111] M. Junered, S. Esterhuizen, D. Akos, and P. Axelrad. A modular GPS remote sensing software receiver for small platforms. In *Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006)*, pages 634–642, Fort Worth, TX, September 2006.

[112] J.F. Marchan-Hernandez, A. Camps, N. Rodriguez-Alvarez, X. Bosch-Lluis, I. Ramos-Perez, and E. Valencia. PAU/GNSS-R: Implementation, performance and first results of a real-time delay-doppler map reflectometer using global navigation satellite system signals. *Sensors*, 8(5):3005–3019, 2008.

[113] "I-REACT project". http://www.i-react.eu/. Accessed: 2019-06-14.

[114] Letizia Lo Presti, Maurizio Fantino, and Marco Pini. *Digital Signal Processing for GNSS Receivers*, chapter 21, pages 707–761. John Wiley & Sons, Ltd, 2019.

[115] Nicola Linty and Fabio Dovis. *An overview on Global Positioning Techniques for Harsh Environments*, chapter 23, pages 839–881. John Wiley & Sons, Ltd, 2019.

[116] Fabio Dovis. *GNSS Interference Threats and Countermeasures*. Artech House, 2015.

[117] Lucia Borasi and Alberto Maffiotti. *Alpine Lakes: Survey on climate change*. Arpa Piemonte, Torino - Italy, 2012.

[118] Yury V Yasyukevich, Alexander V Kiselev, Ilya V Zhivetiev, Ilya K Edemskiy, Semen V Syrovatskii, Boris M Maletckii, and Artem M Vesnin. SIMuRG: System for ionosphere monitoring and research from GNSS. *GPS Solutions*, 24(3):1–12, 2020.

[119] Lara Fernandez, Joan Adria Ruiz-De-Azua, Anna Calveras, and Adriano Camps. Assessing LoRa for satellite-to-Earth communications considering the impact of ionospheric scintillation. *IEEE Access*, 8:165570–165582, 2020.

[120] NJ Skinner, RF Kelleher, JB Hacking, and CW Benson. Scintillation fading of signals in the SHF band. *Nature Physical Science*, 232(27):19–21, 1971.

[121] K. M. Groves, S. Basu, E. J. Weber, M. Smitham, H. Kuenzler, C. E. Valladares, R. Sheehan, E. MacKenzie, J. A. Secan, P. Ning, W. J. McNeill, D. W. Moonan, and M. J. Kendra. Equatorial scintillation and systems support. *Radio Science*, 32(5):2047–2064, 1997.

[122] Franz J. Meyer, Kancham Chotoo, Susan D. Chotoo, Barton D. Huxtable, and Charles S. Carrano. The influence of equatorial scintillation on L-Band SAR image quality and phase. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):869–880, 2016.

[123] Lucas A Salles, Bruno C Vani, Alison Moraes, Emanoel Costa, and Eurico R de Paula. Investigating ionospheric scintillation effects on multifrequency GPS signals. *Surveys in Geophysics*, 42(4):999–1025, 2021.

[124] Vincenzo Romano, Luca Spogli, Marcio Aquino, Alan Dodson, Craig Hancock, and Biagio Forte. GNSS station characterisation for ionospheric scintillation applications. *Advances in Space Research*, 52(7):1237–1246, 2013.

[125] Steve Taylor, Yu Morton, Yu Jiao, Jeffrey Triplett, and Wouter Pelgrum. An improved ionosphere scintillation event detection and automatic trigger for GNSS data collection systems. In *Proceedings of the 2012 International Technical Meeting of The Institute of Navigation*, pages 1563–1569, 2012.

[126] Yanlan Liao and Yuhua Zou. Impact of radio frequency interference on GNSS ionospheric scintillation data analysis. In *2021 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, pages 1–3. IEEE, 2021.

[127] Melania Susi, Marcio Aquino, Rodrigo Romero, Fabio Dovis, and Marcus Andreotti. Design of a robust receiver architecture for scintillation monitoring. In *Proceedings of IEEE/ION PLANS 2014*, pages 73–81, 2014.

[128] Martin Kriegel and Jens Berdermann. Ionosphere monitoring and prediction center. In *2020 European Navigation Conference (ENC)*, pages 1–10. IEEE, 2020.

[129] J. Lee, Y. T. J. Morton, J. Lee, H. Moon, and J. Seo. Monitoring and mitigation of ionospheric anomalies for GNSS-based safety critical systems: A review of up-to-date signal processing techniques. *IEEE Signal Processing Magazine*, 34(5):96–110, 2017.

[130] Introduction to HF radio propagation. https://www.sws.bom.gov.au/Educational/5/2/2, April 2022.

[131] Seebany Datta-Barua, Pau Llado Prat, and Donald L. Hampton. Multiyear detection, classification and hypothesis of ionospheric layer causing GNSS scintillation. *Radio Science*, 56(12):e2021RS007328, 2021.

[132] K. Wu, J. Xu, X. Yue, C. Xiong, W. Wang, W. Yuan, C. Wang, Y. Zhu, and J. Luo. Equatorial plasma bubbles developing around sunrise observed by an all-sky imager and global navigation satellite system network during storm time. *Annales Geophysicae*, 38(1):163–177, 2020.

[133] J. P. McClure, W. B. Hanson, and J. H. Hoffman. Plasma bubbles and irregularities in the equatorial ionosphere. *Journal of Geophysical Research (1896-1977)*, 82(19):2650–2656, 1977.

[134] R. T. Tsunoda, R. C. Livingston, J. P. McClure, and W. B. Hanson. Equatorial plasma bubbles: Vertically elongated wedges from the bottomside F layer. *Journal of Geophysical Research: Space Physics*, 87(A11):9171–9180, 1982.

[135] J. D. Huba and G. Joyce. Global modeling of equatorial plasma bubbles. *Geophysical Research Letters*, 37(17), 2010.

[136] Y. Nishimura, S. Mrak, J. L. Semeter, A. J. Coster, P. T. Jayachandran, K. M. Groves, D. J. Knudsen, N. Nishitani, and J. M. Ruohoniemi. Evolution of mid-latitude density irregularities and scintillation in North America during the 7–8 September 2017 storm. *Journal of Geophysical Research: Space Physics*, 126(6):e2021JA029192, 2021.

[137] E.L. Afraimovich, E.I. Astafyeva, V.V. Demyanov, and I.F. Gamayunov. Mid-latitude amplitude scintillation of GPS signals and GPS performance slips. *Advances in Space Research*, 43(6):964–972, 2009.

[138] B. M. Ledvina, J. J. Makela, and P. M. Kintner. First observations of intense GPS L1 amplitude scintillations at midlatitude. *Geophysical Research Letters*, 29(14):4–1–4–4, 2002.

[139] Michael H. Reilly. Thin-phase screen estimates of TID effects on midlatitude transionospheric radio paths. *Radio Science*, 28(06):979–986, 1993.

[140] Ercha Aa, Wengeng Huang, Siqing Liu, Aaron Ridley, Shasha Zou, Liqin Shi, Yanhong Chen, Hua Shen, Tianjiao Yuan, Jianyong Li, and Tan Wang. Midlatitude plasma bubbles over China and adjacent areas during a magnetic storm on 8 September 2017. *Space Weather*, 16(3):321–331, 2018.

[141] J. Osborn, D. Föhring, V. S. Dhillon, and R. W. Wilson. Atmospheric scintillation in astronomical photometry. *Monthly Notices of the Royal Astronomical Society*, 452(2):1707–1716, 07 2015.

[142] N. Balan, J. Souza, and G.J. Bailey. Recent developments in the understanding of equatorial ionization anomaly: A review. *Journal of Atmospheric and Solar-Terrestrial Physics*, 171:3–11, 2018. Vertical Coupling in the Atmosphere-Ionosphere System: Recent Progress.

[143] P M Kintner, B M Ledvina, and ER De Paula. GPS and ionospheric scintillations. *Space weather*, 5(9), 2007.

[144] P. Prikryl, R. Ghoddousi-Fard, J. M. Weygand, A. Viljanen, M. Connors, D. W. Danskin, P. T. Jayachandran, K. S. Jacobsen, Y. L. Andalsvik, E. G. Thomas, J. M. Ruohoniemi, T. Durgonics, K. Oksavik, Y. Zhang, E. Spanswick, M. Aquino, and V. Sreeja. GPS phase scintillation at high latitudes during the geomagnetic storm of 17–18 March 2015. *Journal of Geophysical Research: Space Physics*, 121(10):10,448–10,465, 2016.

[145] Andrzej W Wernik, Lucilla Alfonsi, and Massimo Materassi. Ionospheric irregularities, scintillation and its effect on systems. *Acta geophysica polonica*, 52(2):237–249, 2004.

[146] Mike Hapgood. Ionospheric science: An example of the importance of diversity in approaches to scientific research. *Atmosphere*, 13(3), 2022.

[147] Charles Rino. *The theory of scintillation with applications in remote sensing*. John Wiley & Sons, 2011.

[148] Hannah Ritchie Max Roser and Esteban Ortiz-Ospina. World population growth. *Our World in Data*, 2013. https://ourworldindata.org/world-population-growth.

[149] IGS.org. https://igs.org/wp-content/uploads/2020/03/Screen-Shot-2019-08-22-at-3.36.52-PM-1.png, April 2022.

[150] Ghosh Tilottama, Rebecca Powell, Christopher Elvidge, Kimberly Baugh, and Paul Sutton. Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3:147–160, 2010.

[151] M. Fingas and C.E. Brown. Chapter 5 - oil spill remote sensing. In Mervin Fingas, editor, *Oil Spill Science and Technology (Second Edition)*, pages 305–385. Gulf Professional Publishing, Boston, second edition edition, 2017.

[152] S. Datta-Barua, T. Walter, J. Blanch, and P. Enge. Bounding higher-order ionosphere errors for the dual-frequency GPS user. *Radio Science*, 43(5), 2008.

[153] M. Mainul Hoque and N. Jakowski. Higher order ionospheric effects in precise GNSS positioning. *Journal of Geodesy*, 81(4):259–268, April 2007.

[154] Mike Hapgood. Chapter 1 - linking space weather science to impacts—the view from the Earth. In Natalia Buzulukova, editor, *Extreme Events in Geospace*, pages 3–34. Elsevier, 2018.

[155] M. Fritsche, R. Dietrich, C. Knöfel, A. Rülke, S. Vey, M. Rothacher, and P. Steigenberger. Impact of higher-order ionospheric terms on GPS estimates. *Geophysical Research Letters*, 32(23), 2005.

[156] C. L. Rino. A power law phase screen model for ionospheric scintillation: 1. weak scatter. *Radio Science*, 14(6):1135–1145, 1979.

[157] Shishir Priyadarshi. A review of ionospheric scintillation models. *Surveys in Geophysics*, 36:295 – 324, 2015.

[158] Kung Chie Yeh and Chao-Han Liu. Radio wave scintillations in the ionosphere. *Proceedings of the IEEE*, 70(4):324–360, 1982.

[159] A. W. Wernik, L. Alfonsi, and M. Materassi. Scintillation modeling using in situ data. *Radio Science*, 42(1):1–21, 2007.

[160] Y. Béniguel. Global Ionospheric propagation Model (GIM): A propagation model for scintillations of transmitted signals. *Radio Science*, 37(3):1–14, 2002.

[161] J. A. Secan, R. M. Bussey, E. J. Fremouw, and S. Basu. High-latitude upgrade to the wideband ionospheric scintillation model. *Radio Science*, 32(4):1567–1574, 1997.

[162] B.H. Briggs and I.A. Parkin. On the variation of radio star and satellite scintillations with zenith angle. *Journal of Atmospheric and Terrestrial Physics*, 25(6):339–366, 1963.

[163] A. J. Van Dierendonck and B. Arbesser-Rastburg. Measuring ionospheric scintillation in the equatorial region over Africa, including measurements from SBAS geostationary satellite signals. In *Proceedings of the 17th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2004)*, pages 316–324, 2004.

[164] M Najmafshar, S Skone, and F Ghafoori. GNSS data processing investigations for characterizing ionospheric scintillation. In *Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014)*, pages 1190–1202, 2014.

[165] Massimo Materassi and Cathryn N. Mitchell. Wavelet analysis of GPS amplitude scintillation: A case study. *Radio Science*, 42(1), 2007.

[166] Massimo Materassi, Lucilla Alfonsi, Giorgiana De Franceschi, Vincenzo Romano, Cathryn Mitchell, and Paolo Spalla. Detrend effect on the scalograms of GPS power scintillation. *Advances in Space Research*, 43(11):1740–1748, 2009. Ionosphere - Modelling, Forecasting, and Telecommunications I.

[167] Geoff Crowley, Gary S. Bust, Adam Reynolds, Irfan Azeem, Rick Wilder, Brady W. O'Hanlon, Mark L. Psiaki, Steven Powell, Todd E. Humphreyes, and Jahshan A. Bhatti. CASES: A novel low-cost ground-based dual-frequency GPS software receiver and space weather monitor. In *Proceedings of the 24th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2011)*, pages 1437–1446, 2011.

[168] E. J. Fremouw, R. L. Leadabrand, R. C. Livingston, M. D. Cousins, C. L. Rino, B. C. Fair, and R. A. Long. Early results from the DNA Wideband satellite experiment—complex-signal scintillation. *Radio Science*, 13(1):167–187, 1978.

[169] J. K. Holmes. *Coherent spread spectrum systems*. 1982.

[170] Igor Ponte Portella, Alison de O Moraes, Marcelo da Silva Pinho, Jonas Sousasantos, and Fabiano Rodrigues. Examining the tolerance of GNSS receiver phase tracking loop under the effects of severe ionospheric scintillation conditions based on its bandwidth. *Radio Science*, 56(6):1–11, 2021.

[171] E. R. de Paula, A. R. F. Martinon, A. O. Moraes, C. Carrano, A. C. Neto, P. Doherty, K. Groves, C. E. Valladares, G. Crowley, I. Azeem, A. Reynolds, D. M. Akos, T. Walter, T. L. Beach, and J.-M. Slewaegen. Performance of 6 different global navigation satellite system receivers at low latitude under moderate and strong scintillation. *Earth and Space Science*, 8(2):e2020EA001314, 2021. e2020EA001314 2020EA001314.

[172] B. Forte and S. M. Radicella. Problems in data treatment for ionospheric scintillation measurements. *Radio Science*, 37(6):8–1–8–5, 2002.

[173] Luca Spogli, Hossein Ghobadi, Antonio Cicone, Lucilla Alfonsi, Claudio Cesaroni, Nicola Linty, Vincenzo Romano, and Massimo Cafaro. Adaptive phase detrending for GNSS scintillation detection: A case study over Antarctica. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[174] Y. I. Feldstein. The discovery and the first studies of the auroral oval: A review. *Geomagnetism and Aeronomy*, 56(2):129–142, March 2016.

[175] WDC for geomagnetism, Kyoto. hourly equatorial Dst values (real-time) September 2020. https://wdc.kugi.kyoto-u.ac.jp/dst_realtime/202009/index.html, April 2022.

[176] L. Spogli, L. Alfonsi, G. De Franceschi, V. Romano, M. H. O. Aquino, and A. Dodson. Climatology of GPS ionospheric scintillations over high and mid-latitude European regions. *Annales Geophysicae*, 27(9):3429–3437, 2009.

[177] Caner Savas, Gianluca Falco, and Fabio Dovis. A comparative performance analysis of GPS L1 C/A, L5 acquisition and tracking stages under polar and equatorial scintillations. *IEEE Transactions on Aerospace and Electronic Systems*, 57(1):227–244, 2021.

[178] OJ Olwendo et al. Elimination of superimposed multipath effects on scintillations index on solar quiet ionosphere at low latitude over the Kenyan airspace from a lone positioned SCINDA system. In *Proceedings of the 23rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2010)*, 2010.

[179] Sebastijan Mrak, Joshua Semeter, Yukitoshi Nishimura, Fabiano S. Rodrigues, Anthea J. Coster, and Keith Groves. Leveraging geodetic GPS receivers for ionospheric scintillation science. *Radio Science*, 55(11):1–17, 2020.

[180] JM Juan, A Aragon-Angel, J Sanz, Guillermo González-Casado, and Adria Rovira-Garcia. A method for scintillation characterization using geodetic receivers operating at 1 Hz. *Journal of Geodesy*, 91(11):1383–1397, 2017.

[181] Jon Bruno, Francesco Darugna, Karl Bolmgren, Jannes B Wübbena, Cathryn Mitchell, and Martin Schmitz. Quality analysis of dual-frequency smartphone-based ionospheric TEC measurements: what can be achieved? *Annals of Geophysics*, 64(1):RS103–RS103, 2021.

[182] DA Kogogin, IA Nasyrov, DS Maksimov, and RV Zagretdinov. Using android smartphones with dual-frequency multi-GNSS receiver to measure the total electron content of the ionosphere. In *Journal of Physics: Conference Series*, volume 1991, page 012025. IOP Publishing, 2021.

[183] Fabio Dovis, Rayan Imam, Wenjian Qin, Caner Savas, and Hans Visser. Opportunistic use of GNSS signals to characterize the environment by means of machine learning based processing. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 9190–9194. IEEE, 2020.

[184] Yunxiang Liu and Y. Jade Morton. Automatic detection of ionospheric scintillation-like GNSS satellite oscillator anomaly using a machine-learning algorithm. *NAVIGATION*, 67(3):651–662, 2020.

[185] Yunxiang Liu and Y. Jade Morton. Improved automatic detection of GPS satellite oscillator anomaly using a machine learning algorithm. *NAVIGATION: Journal of the Institute of Navigation*, 69(1), 2022.

[186] Luca Spogli, Vincenzo Romano, Giorgiana De Franceschi, Lucilla Alfonsi, Eleftherios Plakidis, Claudio Cesaroni, Marcio Aquino, Alan Dodson, Joao Francisco Galera Monico, and Bruno Vani. A filtering method developed to improve GNSS receiver data quality in the CALIBRA project. In Riccardo Notarpietro, Fabio Dovis, Giorgiana De Franceschi, and Marcio Aquino, editors, *Mitigation of Ionospheric Threats to GNSS*, chapter 9. IntechOpen, Rijeka, 2014.

[187] Nicola Linty, Fabio Dovis, and Lucilla Alfonsi. Software-defined radio technology for GNSS scintillation analysis: Bring Antarctica to the lab. *GPS Solutions*, 22(4):96, 2018.

[188] Akpojoto Siemuri, Heidi Kuusniemi, Mohammed S. Elmusrati, Petri Välisuo, and Ahm Shamsuzzoha. Machine learning utilization in GNSS—use cases, challenges and future applications. In *2021 International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–6, 2021.

[189] Li-Ta Hsu. GNSS multipath detection using a machine learning approach. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2017.

[190] Haosheng Xu, Antonio Angrisano, Salvatore Gaglione, and Li-Ta Hsu. Machine learning based LOS/NLOS classifier and robust estimator for GNSS shadow matching. *Satellite Navigation*, 1(1):1–12, 2020.

[191] Yan Jia, Shuanggen Jin, Patrizia Savi, Yun Gao, Jing Tang, Yixiang Chen, and Wenmei Li. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation. *Remote sensing*, 11(14):1655, 2019.

[192] Ruben Morales Ferre, Alberto de la Fuente, and Elena Simona Lohan. Jammer classification in GNSS bands via machine learning algorithms. *Sensors*, 19(22):4841, 2019.

[193] Silvio Semanjski, Ivana Semanjski, Wim De Wilde, and Alain Muls. Use of supervised machine learning for GNSS signal spoofing detection with validation on real-world meaconing and spoofing data—part I. *Sensors*, 20(4):1171, 2020.

[194] Ryan M. McGranaghan, Anthony J. Mannucci, Brian Wilson, Chris A Mattmann, and Richard Chadwick. New capabilities for prediction of high-latitude ionospheric scintillation: A novel approach with machine learning. *Space Weather*, 16(11):1817–1846, 2018.

[195] Lei Liu, Y Jade Morton, and Yunxiang Liu. Machine learning prediction of storm-time high-latitude ionospheric irregularities from GNSS-derived ROTI maps. *Geophysical Research Letters*, 48(20):e2021GL095561, 2021.

[196] Rayan Imam, Caner Savas, and Fabio Dovis. Detecting phase scintillation at high latitudes using ionospheric scintillation monitoring records and machine learning techniques. In *2021 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, pages 37–42. IEEE, 2021.

[197] R. Imam and F. Dovis. Distinguishing ionospheric scintillation from multipath in GNSS signals using bagged decision trees algorithm. In *2020 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, pages 83–88, 2020.

[198] N. Linty, A. Farasin, A. Favenza, and F. Dovis. Detection of GNSS ionospheric scintillations based on machine learning decision tree. *IEEE Transactions on Aerospace and Electronic Systems*, 55(1):303–317, 2019.

[199] Giulio Franzese, Nicola Linty, and Fabio Dovis. Semi-supervised GNSS scintillations detection based on DeepInfomax. *Applied Sciences*, 10(1):381, 2020.

[200] Y. Jiao, J. J. Hall, and Y. T. Morton. Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 53(1):405–418, 2017.

[201] Yu Jiao, John J Hall, and Yu T Morton. Performance evaluation of an automatic GPS ionospheric phase scintillation detector using a machine-learning algorithm. *NAVIGATION: Journal of the Institute of Navigation*, 64(3):391–402, 2017.

[202] T Mitchell, B Buchanan, G DeJong, T Dietterich, P Rosenbloom, and A Waibel. Machine learning. *Annual Review of Computer Science*, 4(1):417–433, 1990.

[203] Merima Kulin, Tarik Kazaz, Eli De Poorter, and Ingrid Moerman. A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer. *Electronics*, 10(3), 2021.

[204] Solveig Badillo, Balazs Banfai, Fabian Birzele, Iakov I. Davydov, Lucy Hutchinson, Tony Kam-Thong, Juliane Siebourg-Polster, Bernhard Steiert, and Jitao David Zhang. An introduction to machine learning. *Clinical Pharmacology & Therapeutics*, 107(4):871–885, 2020.

[205] Qianyun Zhang, Kaveh Barri, Pengcheng Jiao, Hadi Salehi, and Amir Alavi. Genetic programming in civil engineering: advent, applications and future trends. *Artificial Intelligence Review*, 54:1863–1885, 03 2021.

[206] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[207] Zhi-Hua Zhou. *Machine Learning*. Springer Singapore, 2021.

[208] Zhiliang Liu, Ming Jian Zuo, Xiaomin Zhao, and Hongbing Xu. An analytical approach to fast parameter selection of gaussian RBF kernel for support vector machine. *J. Inf. Sci. Eng.*, 31(2):691–710, 2015.

[209] A. Burkov. *The Hundred-page Machine Learning Book*. Andriy Burkov, 2019.

[210] Haneen Arafat Abu Alfeilat, Ahmad BA Hassanat, Omar Lasassmeh, Ahmad S Tarawneh, Mahmoud Bashir Alhasanat, Hamzeh S Eyal Salman, and VB Surya Prasath. Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big data*, 7(4):221–248, 2019.

[211] Xibin Dong, Zhiwen Yu, Wenming Cao, Yifan Shi, and Qianli Ma. A survey on ensemble learning. *Frontiers of Computer Science*, 14(2):241–258, 2020.

[212] Zhi-Hua Zhou. Ensemble learning. In *Machine learning*, pages 181–210. Springer, 2021.

[213] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[214] Omer Sagi and Lior Rokach. Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4):e1249, 2018.

[215] Caner Savas and Fabio Dovis. The impact of different kernel functions on the performance of scintillation detection based on support vector machines. *Sensors*, 19(23), 2019.

[216] Francesco Palma and et al. The august 2018 geomagnetic storm observed by the high-energy particle detector on board the CSES-01 satellite. *Applied Sciences*, 11(12), 2021.

[217] Yu Jiao, Yu Morton, and Steve Taylor. Comparative studies of high-latitude and equatorial ionospheric scintillation characteristics of GPS signals. In *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, pages 37–42, 2014.

[218] L. F. C. Rezende, E. R. de Paula, S. Stephany, I. J. Kantor, M. T. A. H. Muella, P. M. de Siqueira, and K. S. Correa. Survey and prediction of the ionospheric scintillation using data mining techniques. *Space Weather*, 8(6), 2010.

[219] G. R. T. de Lima, S. Stephany, E. R. de Paula, I. S. Batista, and M. A. Abdu. Prediction of the level of ionospheric scintillation at equatorial latitudes in Brazil using a neural network. *Space Weather*, 13(8):446–457, 2015.

[220] Kara Lamb, Garima Malhotra, Athanasios Vlontzos, Edward Wagstaff, Atılım Günes Baydin, Anahita Bhiwandiwalla, Yarin Gal, Alfredo Kalaitzis, Anthony Reina, and Asti Bhatt. Prediction of GNSS phase scintillations: A machine learning approach. *arXiv preprint arXiv:1910.01570*, 2019.

[221] Space Weather Services. http://www.sws.bom.gov.au/, June 2022.

[222] Solar Influences Data Analysis Center. http://sidc.oma.be/, June 2022.

[223] Jason Brownlee. Data preparation for machine learning, 2022.

[224] Miroslav Kubat. *An Introduction to Machine Learning*. Springer International Publishing, 2015.

# Appendix A

# Septentrio PolaRx5S ISM record

Table A.1 Septentrio PolaRx5S ISM record

| | |
|---|---|
| x2ndlocktime | Sig2 Lock time used for the TEC computation (seconds) |
| averagel1 | Sig1 Average C/N0 over the last minute (dB-Hz) |
| avg_c_n0_l2c | Sig2 Average C/N0 over the last minute (dB-Hz) |
| avg_c_n0_l5 | Sig3 Average C/N0 over the last minute (dB-Hz) |
| avgccd_l2c | Sig2 AvgCCD (meters) |
| avgccd_l5 | Sig3 AvgCCD (meters) |
| avgccdl1 | Sig1 AvgCCD (meters) |
| avgcn2freqtec | Sig2 Averaged C/N0 used for the TEC computation (dB-Hz) |
| azimuth | Azimuth |
| correctionS4_L2C | Sig2 S4 thermal noise Correction |
| corrections4_l5 | Sig3 S4 thermal noise Correction |
| corrections4l1 | Sig1 S4 thermal noise Correction |
| dtec0 | dTEC from TOW-15s to TOW (TECU) |
| dtec30_15 | dTEC from TOW-30s to TOW-15s (TECU) |
| dtec45_30 | dTEC from TOW-45s to TOW-30s (TECU) |
| dtec60_45 | dTEC from TOW-60s to TOW-45s (TECU) |
| elevation | Elevation |
| locktime_l2c | Sig2 lock time (seconds) |
| locktime_l5 | Sig3 lock time (seconds) |
| locktimel1 | Sig1 lock time (seconds) |
| p_l2c | Sig2 p phase spectral slope |

| | |
|---|---|
| p_l5 | Sig3 p phase spectral slope |
| phi01_l2c | Sig2 Phi01 (radians) |
| phi01_l5 | Sig3 Phi01 (radians) |
| phi01l1 | Sig1 Phi01 |
| phi03_l2c | Sig2 Phi03 (radians) |
| phi03_l5 | Sig3 Phi03 (radians) |
| phi03l1 | Sig1 Phi03 |
| phi10_l2c | Sig2 Phi10 (radians) |
| phi10_l5 | Sig3 Phi10 (radians) |
| phi10l1 | Sig1 Phi10 |
| phi30_l2c | Sig2 Phi30 (radians) |
| phi30_l5 | Sig3 Phi30 (radians) |
| phi30l1 | Sig1 Phi30 |
| phi60_l2c | Sig2 Phi60 (radians) |
| phi60_l5 | Sig3 Phi60 (radians) |
| phi60l1slant | Sig1 Phi60 |
| pl1 | Sig1 p (spectral slope of detrended phase) |
| rxstate | RxState |
| s4_l2_slant | Sig2 S4 |
| s4_l5_slant | Sig3 Total S4 |
| si_l1_29 | Sig1 SI Index |
| si_l1_30 | Sig1 SI Index (dB) |
| si_l2c_43 | Sig2 SI Index |
| si_l2c_44 | Sig2 SI Index (dB) |
| si_l5_57 | Sig3 SI Index |
| si_l5_58 | Sig3 SI Index (dB) |
| sigmaccd_l2c | Sig2 SigmaCCD (meters) |
| sigmaccd_l5 | Sig3 SigmaCCD (meters) |
| sigmaccdl1 | Sig1 SigmaCCD (meters) |
| svid | SVID |
| t_l1 | T on Sig1 (rad^2/Hz) |
| t_l2c | T on Sig2 (rad^2/Hz) |
| t_l5 | T on Sig3 (rad^2/Hz) |
| tec0 | TEC at TOW (TECU) |
| tec15 | TEC at TOW-15s (TECU) |

| | |
|---|---|
| tec30 | TEC at TOW-30s (TECU) |
| tec45 | TEC at TOW-45s (TECU) |
| totals4l1 | Sig1 Total S4 |