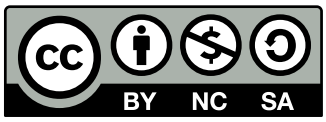
Aalto University
School of Electrical
Engineering

Master's Programme in Computer, Communication and Information Sciences

Soil Moisture Estimation with GNSS Interferometric Reflectometry and Multispectral Satellite Model

Nicolás Padrón

Master's Thesis
2024



Aalto University
School of Electrical
Engineering

Author Nicolás Padrón

Title Soil Moisture Estimation with GNSS Interferometric Reflectometry and Multispectral Satellite Model

Degree programme Computer, Communication and Information Sciences

Major Signal Processing and Data Science

Supervisor Prof. Sergiy Vorobyov

Advisors Dr Sergiy Vorobyov, Dr Andrea Gatti

Collaborative partner Geomatics Research and Development (GReD)

Date 13 July 2024 **Number of pages** 79 **Language** English

Abstract

During the last decade, GNSS Reflectometry (GNSS-R) together with its ground application, GNSS Interferometric Reflectometry (GNSS-IR), have been gaining momentum with satellite missions and ground campaigns. These passive remote sensing technique lies within the microwave remote sensing technology and makes use of satellite navigation signals for Earth Observation (EO) purposes.

Another passive EO technology is multispectral satellite imagery from missions such as Landsat-8. Multispectral imagery is an optical remote sensing technology and covers various wavelengths of the optical spectrum, allowing to analyze the spectral response of the surface materials.

This work focuses on GNSS Interferometric Reflectometry processing for soil moisture estimation, followed by a GNSS-IR aided multispectral model using Landsat-8 satellite data, with the aim of providing an accurate, cost-effective solution with wide-area coverage. In this research, a GNSS-IR processing chain is developed to estimate volumetric soil moisture surrounding static geodetic receivers. Results from this technique are used to fit a linear model with Landsat-8 data combining several optical indexes that have high correlation with soil moisture. Finally, the proof-of-concept of a multispectral imagery model, aided by local GNSS-IR results, is demonstrated and verified against data from the Soil Moisture Active Passive (SMAP) satellite mission for a wide-area coverage.

Keywords microwave remote sensing, GNSS-IR, volumetric soil moisture, multispectral satellite imagery, reflectometry, optical remote sensing

Preface

This Thesis marks the culmination of a chapter that, personally, spans beyond the two-year MSc degree. Prefaces are sometimes unnoticed, but it is important to me to dedicate it as much effort as the Conclusion section.

I want to thank Geomatics Research and Development (GReD) for allowing me to collaborate with them for the GNSS-IR development part of the Thesis. More than a year ago before starting this work, I saw a post on LinkedIn on Europe's MAGDA project about an Italian research group working in GNSS Reflectometry, topic in which I wanted to base my Thesis. Several months later, I decided to contact them looking for collaboration for my Thesis, and their answer was immediate. I have learnt a lot and I am glad to have seen that LinkedIn post in the right moment.

I also want to thank my current employer, Nordic Semiconductor, who hired me knowing that shortly I was going to be full-time both working and studying. Furthermore, although unrelated to this work, they supported me throughout the process allowing me to relax more the already hybrid working environment, providing me with more flexibility.

At first sight, it might appear that doing a MSc Thesis while full-time working in an unrelated domain would make one get out of the "comfort zone". Thankfully, I cannot say this in my case. I have enjoyed every moment of this research, from planing and designing, to testing and implementing. I feel thankful for having finished the MSc degree and delivering a Thesis that gave me learnings beyond my initial thoughts, learnings on water cycle processes and the strength of nature they symbolize. But this would have not been possible without the fundamental support.

I would like to thank my Mother, for her support and advice, example to follow of resilience and courage since I was a kid, and for teaching me that Home is not a physical place, but where one is accompanied by the loved ones. Of course, I also want to thank our four-legged family members Danko and Matilda, because no matter how tough a day is, they always soften it with their inconditional love.

Espoo, 13 July 2024

Nicolás V. Padrón

Contents

Abstract 3

Preface 4

Contents 5

Abbreviations 7

1 Introduction 11
- 1.1 Background . . . 11
- 1.2 Research Problem and Thesis Aim . . . 11
- 1.3 Structure of the Thesis . . . 12

2 Soil Moisture Estimation 13
- 2.1 Soil Moisture Concepts . . . 14

3 GNSS as Microwave Remote Sensing 15
- 3.1 Signal Characteristics . . . 15
- 3.2 GNSS Reflectometry . . . 17
  - 3.2.1 Ground Based . . . 17

4 GNSS-IR Fundamentals 18
- 4.1 Propagation effects . . . 19
- 4.2 Frequency Band . . . 20
- 4.3 Multipath modelling . . . 21
  - 4.3.1 Fresnel Zones . . . 24
  - 4.3.2 SNR region of interest . . . 26
- 4.4 Multipath SNR parameters . . . 27
  - 4.4.1 Antenna Height . . . 28
  - 4.4.2 Amplitude . . . 28
  - 4.4.3 Phase . . . 29

5 Multispectral Satellite Imagery 32
- 5.1 Landsat-8 Characteristics . . . 32
- 5.2 Spectrum bands . . . 33
- 5.3 Quantities of Interest for Soil Moisture . . . 35

6 Data 37
- 6.1 GNSS . . . 37
- 6.2 Landsat-8 . . . 37

7 GNSS-IR Processing Chain 39
- 7.1 Daily Processing . . . 39
  - 7.1.1 Arc Selection and Detrending . . . 39

7.1.2 Reflector Height Estimation . . . 39
7.1.2.1 Frequency Domain Representation . . . 41
7.1.2.2 Spectrum Estimation . . . 42
7.1.2.3 Lomb-Scargle Periodogram . . . 44
7.1.3 Amplitude Compensation and Phase Estimation . . . 45
7.1.3.1 Data smoothing . . . 47
7.1.3.2 Decaying factor estimation . . . 47
7.1.3.3 Phase calculation . . . 48
7.2 Post-Processing . . . 49
7.2.1 Outlier Filtering . . . 49
7.2.2 Filter High Vegetation Effects . . . 50
7.2.3 Phase Baseline Removal . . . 51
7.2.4 Satellite Averaging . . . 51
7.2.5 Vegetation Correction . . . 53
7.2.5.1 Influence on Amplitude and Height estimation . . . 54
7.2.5.2 Influence on Phase estimation . . . 54
7.2.6 Soil Moisture Estimation . . . 55

8 GNSS-IR aided Multispectral Model 56
8.1 Processing Chain . . . 57
8.2 Terrain-Specific Behavior . . . 60

9 Results 62
9.1 GNSS-IR only estimation . . . 62
9.1.1 Design Configuration . . . 63
9.2 GNSS-IR aided Multispectral VSM estimation . . . 69
9.2.1 Terrain-Specific Processing . . . 70

10 Conclusion 74
10.1 Future Work . . . 75

References 76

Abbreviations

BPSK	Binary Phase-Shift Keying
C/N0	Carrier-to-Noise Spectral Density
CYGNSS	Cyclone Global Navigation Satellite System
DFT	Discrete Fourier Transform
DSSS	Direct-Sequence Spread Spectrum
DTFT	Discrete Time Fourier Transform
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FIR	Finite Impulse Response
GEE	Google Earth Engine
GLONASS	Global Navigation Satellite System (Russia)
GNSS	Global Navigation Satellite Systems
GNSS-IR	GNSS Interferometric Reflectometry
GNSS-R	GNSS Reflectometry
GPS	Global Positioning System
IRNSS	Indian Regional Navigation Satellite System
LHCP	Left-Hand Circularly Polarized
LoS	Line of Sight
LSP	Lomb-Scargle Periodogram
LS	Least Squares
LST	Land Surface Temperature
MEO	Medium Earth Orbit
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
OLI	Operational Land Imager
PBO	Plate Boundary Observatory
PRN	Pseudorandom Noise
QZSS	Quasi-Zenith Satellite System
RHCP	Right-Hand Circularly Polarized
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SNR	Signal-to-Noise Ratio
SR	Surface Reflectance
SWIR	Short-Wave Infrared
TIRS	Thermal Infrared Sensor
TOA	Top of Atmosphere
VSDI	Vegetation Soil Drought Index
VSM	Volumetric Soil Moisture
VWC	Vegetation Water Content

List of Figures

1 C/N0 variation with time and low-elevation angles of interest from [25]. 19
2 Left: illustration of RHCP (blue) and LHCP (red) patterns in antenna design. Right: example of direct (blue) and reflection (red) components. Particular reflection surface shown in green, although signal reflects all around the antenna [13]. H_0 and E are the antenna height and satellite elevation angle, respectively. 20
3 Ground multipath signature from different GPS bands from [12]. . . 21
4 Illustration for construction of interferometric delay τ_i following [27] adjusting colors to same as in figure 2 for better understanding. . . . 22
5 Responses with respect to elevation angle of: power (direct and reflected), signal combinations (interferometric and composite), as well as interferometric and compositional phases form [3] 24
6 Reflection surface on the ground with respect to satellite elevation angle. Antenna illustrated as big cyan dot. 25
7 Illustration of ground reflections in the surroundings of the receiver antenna. 25
8 a) Illustration of ground Fresnel zones in 1D from [27]. b) Illustration of 2D Fresnel zones surrounding the antenna from [26]. On b), antenna height is 10m, and elevation angles (in degrees) are green (5), cyan (10), blue (15), magenta (20), and red (25). 26
9 Selection of region < 30 degrees and detrending by removing the influence of the direct component. Illustration from [29]. Notice that the x-axis on the left image is the elevation angle in degrees, while on the left image is the sine of elevation angle, which is the domain variable on equation (5) 27
10 Different frequencies from equation (6) by varying antenna heights [3]. 28
11 Vanishing amplitude with different surface roughness standard deviations from [3]. Recall that wavelengths from GPS bands L1, L2 and L5 are 19.05cm, 24.45cm and 25.48cm, respectively. 29
12 SNR phase from different reflecting materials and moisture levels [3]. 29
13 Relationship between phase and soil moisture [1]. Left) Low vegetation site. Right) Medium vegetation site. 30
14 Bare vs vegetated soil and its effect in SNR at low elevation angles [13]. 30
15 Typical multispectral reflectance spectrum response. Adapted from [33]. 34
16 Spectral response of different moistened soils and vegetation [5]. . . 36
17 Stations MFLE (top) and P267 (bottom) 38
18 Block diagram of GNSS-IR chain for daily processing and post-processing stages. 40
19 a) DFT over evenly sampled Gaussian pulse. b) DFT over unevenly sampled Gaussian pulse. [35]. 44
20 LSP computation for detrended SNR arc. 45
21 Sampling of positive peaks following detrended SNR in figure 9. . . 46

22 Illustration of raw data, smoothing, peak sampling and posterior decaying factor and phase estimation. 49
23 Outlier filtering based peak-to-noise ratio and variation of antenna height estimation. Blue denotes the raw data points, pink corresponds to after the 1st filtering, and black corresponds to after the 2nd filtering. 50
24 Baseline removal and normalized LSP peak monitoring. 52
25 Weighted average for phase-zeroed values for $\delta = 1$ (top) and $\delta = \frac{1}{4}$ (bottom). Phase-zeroed values ϕ_z are shown on the left y-axis in blue, and the corresponding weighting value is shown on the right y-axis in red. 53
26 Effect on vegetation in detrended SNR spectrum from LSP [13]. . . 54
27 Comparison among GNSS-IR, SMAP and Landsat. Positive characteristics for high-resolution soil moisture estimations are highlighted in green, and undesired characteristics in orange. "SM" stands for "Soil Moisture". 56
28 Data acquisition and processing chain for Landsat-8. 58
29 Example of yearly raw, filtered and monthly averaged LST and NDVI for P267 station. 60
30 Estimated parameters throughout the processing period. 62
31 Comparison of vegetation-corrected VSM vs non-corrected. Available daily measurements (top) and VSM (bottom). 63
32 Precipitation quantification during 2016 for MFLE geographical area. 63
33 Comparison of VSM with weighted vs non-weighted average. 64
34 Default experiment with parameters defined in row 1 of table 7. Available daily measurements (top) and VSM (bottom). 65
35 Experiment **A** (left) and **B** (right) defined in table 7. Available daily measurements (top) and VSM (bottom). 66
36 Experiment **C** defined in table 7. Available daily measurements (top) and VSM (bottom). 66
37 Experiment **D** defined in table 7. Available daily measurements (top) and VSM (bottom). 67
38 Experiment **E** defined in table 7. Available daily measurements (top) and VSM (bottom). 68
39 VSM in MFLE (left) and P267 (right) with design parameters in table 8. 69
40 Monthly averaged fitting (top) and whole period data (bottom). Area corresponds to approximately 100x100 meter for both MFLE (right) and P267 (left). 70
41 Result of multispectral model application on 10x10 km area around P267. Correlations with SMAP and GNSS-IR for 10x10km and 100x100m respectively. 71
42 NDVI for 10x10 km region in February (left) and June (right). . . . 72
43 Correlations between 10x10km VSM estimations from Landsat-8 and SMAP for different terrain types based on NDVI thresholds. 72

44 Color map images of VSM estimation for varying NDVI thresholds differentiating between low and medium vegetation subregions. VSM variations based on vegetation are more noticeable in Summer (June) than in Winter (February). 73

1 Introduction

1.1 Background

Estimating soil moisture is important for many purposes, including monitoring regions susceptible to natural hazards, planning the planting of crops, and optimizing irrigation in agriculture. Currently, in addition to ground-based and aerial methods, soil moisture has been globally estimated using Earth Observation (EO) satellites. EO satellites use remote sensing instruments to receive signals that are either explicitly designed and transmitted (active remote sensing) or can be leveraged as a secondary application from already existing sources (passive remote sensing).

Soil moisture has been estimated using either local (e.g., ground probes) or global (satellite observation) solutions. Ground probes can be inserted into the soil to estimate land water content. However, installing ground probes is not cost effective for monitoring relatively large areas. On the other hand, satellite observation such as NASA's Soil Moisture Active Passive (SMAP) mission records global daily measurements about soil moisture from the top 5-cm layer of soil. Despite its global coverage, the estimates provided by SMAP are limited to only km^2 resolution.

Over the last decade, GNSS Interferometric Reflectometry (GNSS-IR) has been introduced as a passive remote sensing method. GNSS-IR utilizes satellite navigation signals (GNSS) by exploiting multipath ground reflections captured by the GNSS receiver antenna. The GNSS-IR method has been widely used for soil moisture estimation [1, 2, 3]. Furthermore, since GNSS-IR is based on GNSS signals, it uses microwave remote sensing, thus enabling soil moisture top-layer estimates similar to SMAP but with m^2 resolution. GNSS-IR analyzes signal power anomalies influenced by characteristics of the ground, thus making it possible to infer parameters such as soil moisture. Therefore, GNSS-IR offers a valid alternative for providing accurate, meter-level and minimum-deployment solution.

1.2 Research Problem and Thesis Aim

Despite its benefits, GNSS-IR is still a local solution. To cover larger areas, GNSS-IR requires the installation of more antennas, which would prove to be impractical and expensive. Moreover, the electronic circuitry of its setup must be protected from environmental conditions.

Therefore, this thesis explores the feasibility of a global, high-resolution solution that combines accurate local GNSS-IR soil moisture estimates with meter-level multispectral satellite imagery. The outcome of this work is a GNSS-IR aided multispectral model utilizing NASA's Landsat-8 satellite data for providing an accurate, cost-effective, and mete-level global solution.

To achieve this goal, a GNSS-IR processing chain is developed to estimate soil moisture content in the surroundings of a static geodetic receiver. Results from this method are used to fit a linear model from Landsat data by combining several optical indexes that have a high correlation with soil moisture [4, 5]. Finally, the proof-of-concept of a multispectral satellite model aided with local GNSS-IR outputs

is demonstrated and verified against SMAP satellite data over extensive areas.

1.3 Structure of the Thesis

This thesis is structured as follows. Chapter 2 reviews state-of-the-art solutions for estimating soil moisture from satellite and ground sensors. Chapter 3 describes GNSS characteristics and its usage in passive remote sensing. Chapter 4 presents GNSS-IR covering both electromagnetic principles as well as the main parameters of interest. Chapter 5 introduces multispectral satellite imagery and the response of Landsat-8 optical bands to soil moisture. Chapter 6 presents the data and scenario characteristics. Chapter 7 describes the implementation of the GNSS-IR processing chain and its application for soil moisture estimation. Chapter 8 combines GNSS-IR outputs with multispectral data in the linear model to provide accurate coverage over wider areas. Chapter 9 presents the results from GNSS-IR standalone and the proposed multispectral model, and compares them to SMAP satellite data for verification. Finally, Chapter 10 concludes the thesis by evaluating the accuracy and scalability of the proposed solution and suggesting directions for future work.

2 Soil Moisture Estimation

Soil moisture has been the subject of numerous climate studies and it plays a crucial role in the energy exchange between the Earth's surface and the atmosphere, influencing vegetation growth, weather patterns, and hydrological cycles [6]. Soil moisture monitoring enables efficient management and use of water resources, having determinant importance for purposes including agricultural monitoring and planning, optimization of irrigation mechanisms and delimitation of risk areas susceptible to landslides. Multiple methods exist for estimating soil moisture either locally or remotely, varying from ground to space platforms.

In the gravimetric method, soil moisture content is directly determined by weighing a soil sample collected from the field before and after it is dried in the laboratory [7]. This method is considered the most accurate and is the only direct technique for measuring soil moisture. Despite its precision, the gravimetric method has significant limitations. It is not suitable for continuous monitoring because it is a destructive process, necessitating the collection of soil samples. Additionally, this method is not practical for assessing large areas because each measurement only represents the specific location from which the sample was taken, making it challenging to obtain a comprehensive view of soil moisture distribution across a broader region.

Other methods estimate soil moisture indirectly by assessing the physical properties of the soil, either through in situ or remote sensing techniques. In situ soil moisture probes facilitate extensive data collection and continuous monitoring. These probes work by measuring the soil's electric permittivity [8], which increases with higher moisture content. The electric permittivity is measured through the time delay in signal propagation or the frequency shift of an electromagnetic pulse. Despite the advantages this approach, it is limited in range since it only measures moisture content within a few centimeters of the sensor, restricting its ability to provide soil moisture estimations over larger areas.

Remote sensing technologies such as space-based sensors offer the advantage of global coverage, enabling the monitoring of soil moisture across wider areas. However, this broad coverage comes at the cost of lower spatial resolution. Active sensors typically have a spatial resolution of around 100 meters, while passive sensors have a much coarser resolution of approximately 10 kilometers [9]. This resolution refers to the spatial resolution of the data collected by the satellite's instruments, which is the scale at which the soil moisture measurements are represented on the Earth's surface.

Another limitation of satellite-based soil moisture measurement is the relatively low revisit time, which refers to the duration between consecutive passages of the satellite over the same ground location. This revisit time is considerably longer compared to the continuous data provided by dedicated local soil moisture monitoring networks. Despite not having the resolution and continuous monitoring capabilities of local sensors, their ability to cover large areas and provide consistent data over time makes them an essential tool in the study of Earth's hydrological processes. Therefore, space-based sensors are therefore considered reference soil moisture information sources on a large scale. Satellite missions such as NASA's Soil Moisture Active Passive (SMAP) [6, 10] and the ESA's Soil Moisture and Ocean Salinity (SMOS) [11]

utilize remote sensing technology to measure global soil moisture levels from space.

2.1 Soil Moisture Concepts

Several concepts exist for quantifying and interpreting soil moisture content. The concepts involved in this thesis are:

- Volumetric Soil Moisture (VSM): is defined as the ratio between the volume of water contained in a given volume of a soil sample [2, 12]. It has units of cm^3/cm^3 although can also be understood as a percentage.
- Vegetation Water Content (VWC): refers to the amount of water stored within the tissues of vegetation, measuring how much water vegetation can hold. It has units of kg/m^2 [13, 14].

3 GNSS as Microwave Remote Sensing

Satellite navigation constellations are widely deployed, their civilian service is free to use, and provide worldwide coverage with all four global (GPS, GLONASS, Galileo, BeiDou) and two regional (QZSS, IRNSS) systems. Despite being aimed to navigation purposes, GNSS presents highlightable features which make the system stand out as a prominent technology for remote sensing.

GNSS operating frequency band, known as L-band (around 1.2-1.5GHz) lies in the microwave spectrum, to which atmosphere is transparent. Furthermore, GNSS signals penetrates deeper into the ground than than other systems using higher frequencies [1, 9]. This classifies GNSS EO applications within the so called *microwave remote sensing* technologies. For these reasons, GNSS is an attractive candidate for Earth Observation due to availability, coverage, cost and propagation properties.

In contrast to instruments which actively sense the environment, GNSS is used as a *passive* remote sensing system. We talk about passive mode when an external system is leveraged as an input to the main application, in this case as an input signal-of-opportunity. GNSS has been widely used for secondary applications such as passive remote sensing, thus leading to less expensive solutions than tailor-made active systems.

The following subsection will present briefly the GNSS signals and the main characteristics that make them of special interest for remote sensing, and continuing subsection introduces the GNSS-Reflectometry technology.

3.1 Signal Characteristics

GNSS (Global Navigation Satellite Systems) signals are transmitted from Medium Earth Orbit (MEO) satellite constellations. In this thesis, only GPS system is used, and therefore this system and its signal characteristics will be briefly described.

GPS signals are modulated with Direct Sequence Spread Spectrum (DSSS) through pseudorandom (PRN) code sequences [15], since shall behave quasi-randomly to ensure white noise-like power spectral density (i.e. flat). This modulation behaves as an extension of Binary Phase Shift Key (BPSK) modulations, and it is common to treat GPS signals as BPSK-modulated.

GPS utilizes multiple frequencies, namely the bands L1, L2, and L5 frequencies. Signal characteristics and frequencies are described below [16, 17].

- **L1 Signal:**
  - **Carrier Frequency:** 1575.42 MHz.
  - **Wavelength:** 19.05 cm.
  - **PRN Code Frequency:** 1.023 MHz.
  - **Modulation:** BPSK at 1.023 kHz.
  - **Characteristics:**
    * Used for civilian GPS applications. It is the first and most common used signal.
    * Consists of two components: L1 C/A (Coarse/Acquisition) and L1 P (Precise) codes. L1 C/A code is of public usage, while L1 P is restricted to military usage. Therefore it is encrypted and provides higher accuracy than L1.
- **L2 Signal:**
  - **Carrier Frequency:** 1227.60 MHz.
  - **Wavelength:** 24.45 cm.
  - **PRN Code Frequency:** 511.5 kHz.
  - **Modulation:** BPSK at 1.023 kHz.
  - **Characteristics:**
    * Includes L2 C/A and L2 P codes, similar to L1.
    * Used for dual-frequency receivers. Similar precision as GPS L1 but main usage in combination with L1 to completely remove the ionospheric errors.
    * To PRNs both at 511.5kHz: short (Civil Moderate, "CM") and long (Civil Long "CL"). PRNs are multiplexed in time, leading to similar overall appearance and modulation as L1
- **L5 Signal:**
  - **Frequency:** 1176.45 MHz.
  - **Wavelength:** 25.48 cm.
  - **PRN Code Frequency:** 10.23 kHz.
  - **Modulation:** BPSK at 10.23 kHz.
  - **Characteristics:**
    * Intended primarily for civil aviation and safety-of-life applications. Also provides improved accuracy, integrity, and reliability compared to L1 and L2.
    * Supports modernized GPS capabilities, including robust signal tracking in challenging environments.

3.2 GNSS Reflectometry

In GNSS Reflectometry (GNSS-R), as other microwave remote sensing technologies, the signal is reflected from the Earth's surface. The aim is to exploit this reflection to infer the characteristics in the reflected area, also known as specular point in the case of a specular reflection (when the incidence and reflection angle match). Such reflections are captured by a receiver which can be located either in a space, aerial or land platform. In particular for soil moisture, microwave technologies can measure the soil's top 5 cm. Its accuracy is related to the large differences in dielectric properties between water and dry soil, with more moistened soils resulting in higher dielectric constants [18].

GNSS-R technology has been an active R&D field during the last decade, with distinguished mission and network campaigns such as NASA's CYGNSS in space, and NOAA's PBO H2O Network in land. Even though research in GNSS-R accounts for nearly 1% of published articles in Remote Sensing until 2023 [19], its performance has been demonstrated to provide trustful results to carry out space, aerial and land applications.

3.2.1 Ground Based

In GNSS Remote Sensing, an application of ground-based reflectometry is GNSS Interferometric Reflectometry, which exploits the satellites' signal reflections surrounding the receiver, therefore providing the lowest coverage area but sharpest resolution of few tenths of m^2.

GNSS-IR has been extensively demonstrated and validated for various applications, including the measurement of vegetation water content [20], detection of vegetation presence [13], surface soil moisture estimation [21, 1], and snow depth assessment [22]. GNSS-IR is currently integrated into the majority of GNSS stations within the National Science Foundation's (NSF) EarthScope Plate Boundary Observatory (PBO) network [23]. This integration serves to estimate changes in snow depth and surface soil moisture levels, both of which are available to download from the UNAVCO online dataset.

4 GNSS-IR Fundamentals

Different to space and aerial applications where the receiver is installed on a platform in movement, ground-based GNSS reflectometry requires the receiver installation to be on a few-meter tall static platform. This configuration is ideal for applications such as continuous real-time monitoring of a specific geographical area of interest.

GNSS-IR makes use of static receivers, which come in geodetic or commercial variants, to estimate Essential Climate Variables (ECVs) within areas of approximately 50-meter radius for a typical 2-meter tall antenna [12]. A notable advantage of this technique is the independence from specifically-designed GNSS equipment for capturing signal reflections, such as methodologies that rely on the Left-Hand Circularly Polarized (LHCP) component, necessitating specialized LHCP compatible antennas [3]. This flexibility not only simplifies equipment requirements but also extends the applicability of already-deployed GNSS stations to be used in parallel for both navigation and environmental sensing purposes.

In order to sense the surrounding area, GNSS-IR processes the outputs of the receiver station, in particular the carrier-to-noise-density ratio (C/N0) or Signal-to-Noise ratio SNR. It is important to clarify that C/N0 and SNR are different concepts.

$$SNR[\text{linear}] = \frac{P_S}{P_N} = \frac{C}{N_0 B} \rightarrow SNR[\text{dB}] = C/N0 - B \tag{1}$$

where $P_S = C$ and $P_N = N_0 B$ are the signal and noise powers, respectively. N_0 stands for the noise spectral density and B for the signal bandwidth. Therefore, due to the C/N0 and SNR relation, the bandwidth just acts as a scaling factor. For this reason it is common in GNSS-IR, also known as GNSS SNR multipath reflectometry [24], to use the terms C/N0 and SNR interchangeably, as is the case in this thesis.

A benefit from the use of SNR is that it is invariable to effects such as satellite orbit and ephemerides errors, clock errors and atmospheric delays which affect other GNSS observables and need to be taken into account for navigation applications [12]. In Remote Sensing, the SNR in GNSS can be used directly since the multipath effect of interest, the interferograms from direct-reflected waves, is already embedded on it.

C/N0 is calculated by satellite navigation receivers and used to monitor the signal quality and compute multiple metrics, such as pseudorange weighting for position fix, or detection of signal anomalies and interference [15]. In ground-based GNSS reflectometry, the C/N0 is a key parameter whose variability is studied to detect and analyze multipath, consequently post-processing it to estimate soil characteristics. It has been demonstrated that observations from GNSS multipath signals show strong correlations with environmental characteristics [3].

Nevertheless, although C/N0 measurements are collected throughout the passage of each satellite in the sky, GNSS-IR only makes use of certain time period. More specifically, since GNSS-IR is focused on ground reflections, it shall process the C/N0 values which are more influenced by ground multipath.

For this reason, the C/N0 values to analyze correspond to the time in which the satellite is rising or setting in the sky, limiting the observation period to the time in

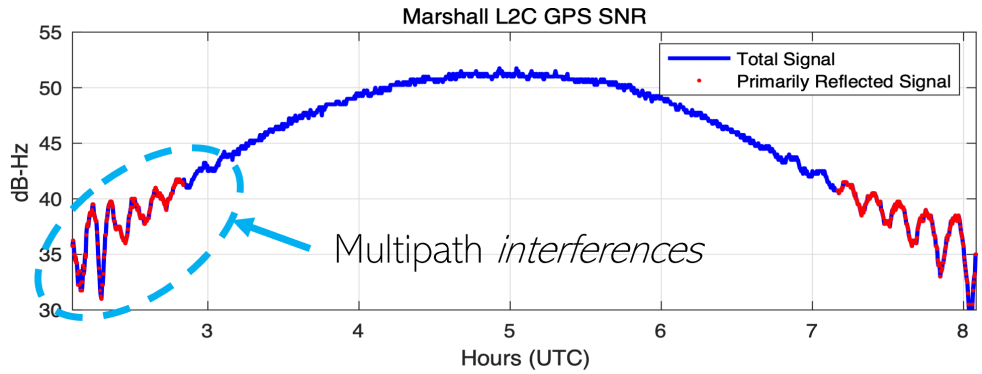


Figure 1: C/N0 variation with time and low-elevation angles of interest from [25].

which the satellites are at low elevation angles as shown in figure 1. Furthermore, GNSS-IR strictly depends on the geometry driving the satellite reflection on the ground surrounding the receiver as shown in figure 2. The details on antenna characteristics and multipath geometry driving this criteria will be explained in the coming subsections 4.1 and 4.3.

4.1 Propagation effects

GNSS signals have predominantly Right-Hand-Circular-Polarized (RHCP) polarization, since LHCP component is limited to 20% of RHCP [3]. The propagation path of the received signal can be divided into two parts: the direct link between satellite and receiver for the Line-of-Sight (LoS) component, and the link between the ground reflections and receiver for the multipath components. Both components are combined in the received input, since the antenna captures the overall signal, containing both LoS and reflections.

Ground reflections from microwave signals, such as other electromagnetic waves, can change their polarization when reflecting on a material. For instance, depending on the incidence angle and the dielectric properties of the surface material, RHCP signal will split between RHCP and LHCP components [3, 12].

From antenna design, the gain pattern of GNSS antennas has contributions from both right and left polarizations. An illustration of reflection surface is, as well as RHCP and LHCP radiation patterns is shown in figure 2. Notice that RHCP is maximum in boresight direction, which is the direction in which the antenna mainlobe is oriented, typically zenith direction.

In the receiver antenna, the RHCP gain pattern is omnidirectional in azimuth, meaning that all azimuth angles receive the same gain, and quasi-hemispherical with elevation angle, which means that gain increases for higher elevations (closer to the boresight direction). The LHCP gain pattern is not well defined but shall remain lower than RHCP gain in boresight direction, being less restrictive in anti-boresight [3]. Notice that the aforementioned restriction for LHCP magnitude to be less than 20% of RHCP applies for the GNSS transmitter on the satellite, not to the receiver. The exact LHCP-RHCP proportions on the receiver antenna design are specific for each product.

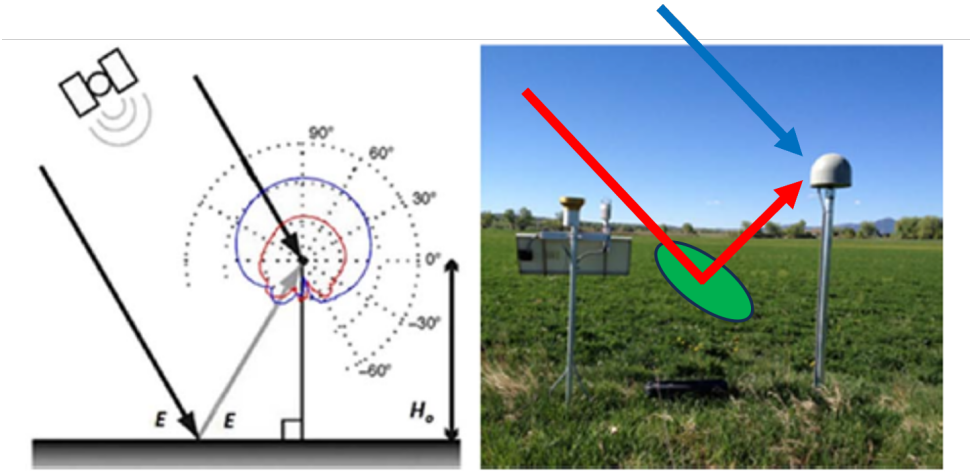


Figure 2: Left: illustration of RHCP (blue) and LHCP (red) patterns in antenna design. Right: example of direct (blue) and reflection (red) components. Particular reflection surface shown in green, although signal reflects all around the antenna [13]. H_0 and E are the antenna height and satellite elevation angle, respectively.

Due to the negligible LHCP power in LoS, the RHCP component predominates. Nevertheless, this is not the case when the signal reflects over a surface. In this case, part of the signal polarization will change to LHCP, thus making this component not negligible anymore. However, this behavior and the magnitude of the derived LHCP and RHCP components depends on the elevation angle and surface material. Satellites with higher elevation angles tend to have LHCP reflections, while low elevation satellites tend to have RHCP reflections [12]. This means that for a same geodetic or commercial receiver, whose radiation pattern follows principally RHCP polarization, it is best to analyze the low-elevation angles looking for multipath. This justifies why low elevation angles are preferred for GNSS-IR, because the same RHCP polarization predominates and therefore these components will cause clear interferograms with the direct signal in the C/N0.

4.2 Frequency Band

As previously mentioned, one of the benefits that microwave remote sensing has from the L-band in which GNSS operates is the transparency of atmosphere, propagation through vegetation and penetration in the top 5cm of soil. Nevertheless, not all GNSS bands behave similarly concerning multipath ground reflections. Therefore, it cannot be expected that the ECV to estimate is equally well embedded in signals from different frequencies.

Considering GPS system as an example, modern signals in L2 and L5 bands follow better theoretical models than legacy and oldest band L1. Details on GPS characteristics can be found in Data section 6. Figure 3 illustrates the ground multipath signature in low elevation angles from L1, L2 and L5 bands. Civilian signals L1 C/A, L2C and L5Q are public, while L2 P(Y) is a restricted for military usage. Multipath

oscillations should be evaluated only from civil signals, which are the ones that can be publicly used. Therefore, modern civil signals from L2 and L5 are excellent for GNSS-IR applications [26] and preferred over L1. Following the GNSS-IR soil moisture literature, this thesis uses L2 frequency band.

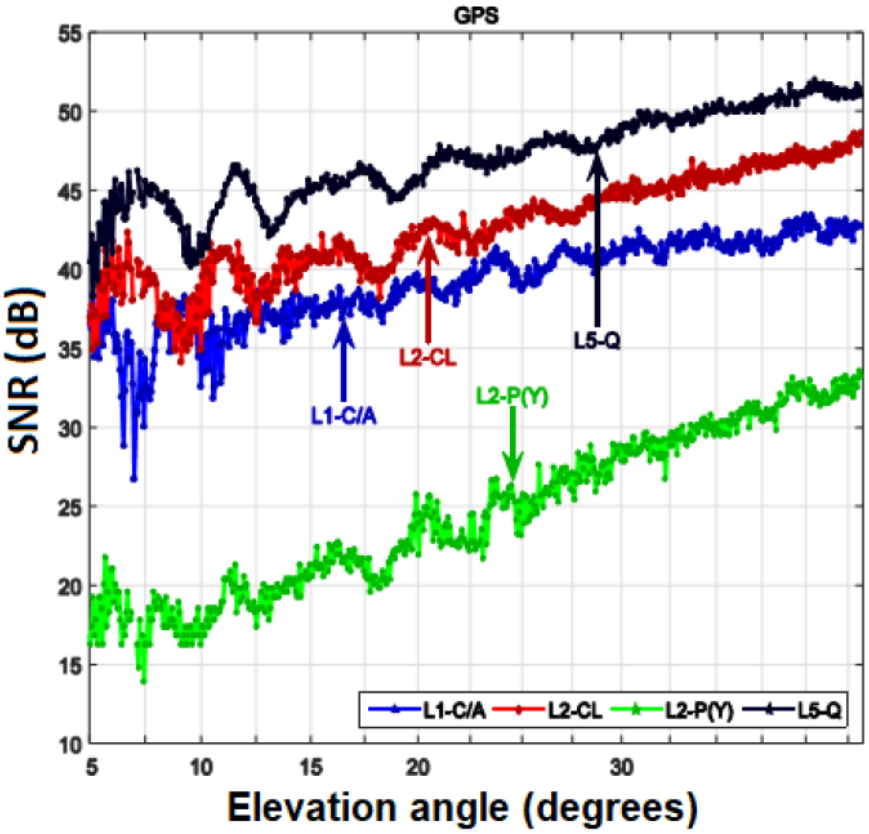


Figure 3: Ground multipath signature from different GPS bands from [12].

4.3 Multipath modelling

The composite signal resulting from the addition of the ground-reflected and the direct components leads to the interferometric pattern. Such pattern results in constructive/destructive additions due to the superposition of direct and reflected waves [3].

The composite signal is modelled with a two-ray model resulting from the reflected-direct geometry, as shown in figure 2. The interactions between reflected and direct waves produce oscillations in power, driven by three main characteristics:

- Interferometric power: it is defined as the ration between reflected and direct components, $P_i = \frac{P_r}{P_d}$, where P_r and P_d are the reflected and direct powers, respectively.

- Interferometric delay: defined as the excess delay between the reflected and direct propagation, $\tau_i = 2Hsin(\theta)$. A clearer illustration for this formulation can be obtained following the reasoning on figure 4, where same as in figure 2 H_0 (here called just H) and E (here called θ) are the antenna height and satellite elevation angle, respectively.

- Interferometric phase: corresponds to the excess phase from the reflected (Φ_r) with respect to the direct component (Φ_d), defined as $\Phi_i = \Phi_r - \Phi_d$.

In terms of received power, or more generally SNR if noise power P_n is accounted for, the interaction between reflected and direct signals can be modeled as [3, 27]

$$\begin{aligned} SNR &= \left(P_d + P_r + 2\sqrt{P_d P_r} cos(\Phi_i)\right) P_n^{-1} \\ &= P_d \left(1 + P_i + 2\sqrt{(P_i)} cos(\Phi_i)\right) P_n^{-1} \end{aligned} \tag{2}$$

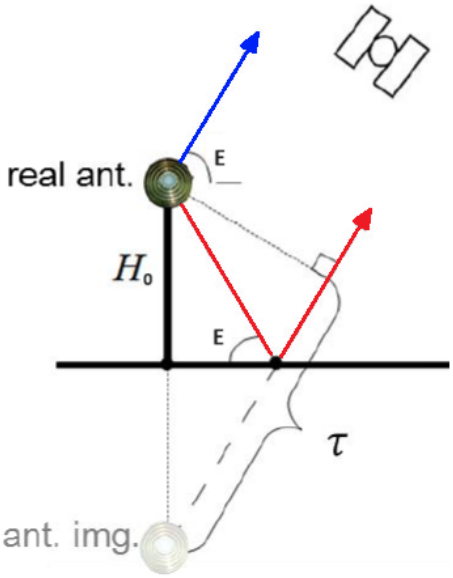


Figure 4: Illustration for construction of interferometric delay τ_i following [27] adjusting colors to same as in figure 2 for better understanding.

Focusing on interferometric phase, from [3, 28] Φ_i can be expanded to account for the interferometric delay τ_i and reflection-induced phase shift Φ_x

$$\Phi_i = \frac{2\pi}{\lambda}\tau_i + \Phi_X = \frac{4\pi H_0}{\lambda} sin(\theta) + \Phi_x \tag{3}$$

where H, λ and θ account for the antenna height, signal wavelength and elevation angle, respectively. The term Φ_x is a phase shift which strongly depends on the reflection surface characteristics, and is influenced by factors such as vegetation, soil moisture or snow accumulation. Therefore [3]:

- τ_i is another phase shift and is commonly called interferometric *geometric* delay.
- Φ_x is called *compositional* phase shift.

Combining equations (2) and (3), we obtain an expression for the interferometric SNR where the multipath influence is principally driven by the satellite elevation angle in the 2nd term.

$$SNR = P_d\,(1 + P_i) + 2P_d\sqrt{P_i}cos\left(\frac{4\pi H_0}{\lambda}sin(\theta) + \Phi_x\right) \tag{4}$$

Result from equation (4) is the detailed explanation of what we observe in figure 1. Therefore, SNR varies based on factors such as constructive or destructive additions, satellite elevation angle, receiver height, and receiver antenna gain pattern.

The interferograms exhibit oscillations in SNR at low satellite elevation angles, where the 2nd term of equation (4) dominates. These oscillations serve as a distinct marker of ground multipath interference, gradually becoming smoother as satellite elevation angle increases, thus making the 1st term of equation (4) the dominant. This smoothing effect occurs as the power of the reflected signal diminishes relative to the direct signal, until its influence on the interferogram becomes negligible.For a typical geodetic-quality GNSS antenna gain pattern, this transition occurs at $\sim$ 30 degrees.

It is worth mentioning that unlike navigation applications, which typically mask out satellites at elevations below 20 or 30 degrees, in reflectometry the region of interest lies precisely below 30 degrees. This emphasis on lower elevation angles stems from the higher influence that reflected components have in this region, that is where SNR oscillations are clearer, which is essential for accurate environmental data retrieval using reflectometry techniques.

On figure 5 we can see the effects of the interferometric characteristics such as power and phase with respect to the elevation angles. In the figure we can see how

- Direct and reflected components, P_d and P_r, respectively increase and decrease with elevation angle, consequently reducing the interferometric power P_i to zero.
- Composite power component $P_c = P_d + P_r$ converges to the direct component as reflected power tends to zero with increasing elevation angle.
- Consequently, error component $P_e = P_c/P_d$, tends to unity.
- Interferometric geometric delay τ_i varies from $\pm$180 degrees, while the compositional phase shift Φ_x varies smoother. This makes sense since Φ_x is driven by the material of the reflecting surface, typically uniform or quasi-uniform surrounding the receiver antenna.
- Finally, we can see how the error phase with respect to a reflection-free case (i.e. direct-only) reduces to zero as the reflected component vanishes with increasing elevation angle.

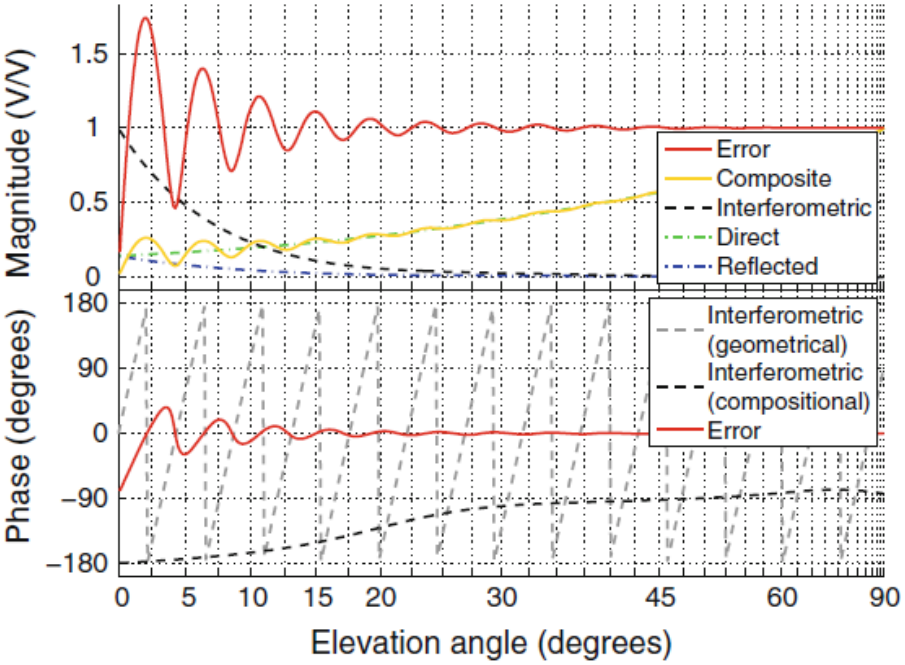


Figure 5: Responses with respect to elevation angle of: power (direct and reflected), signal combinations (interferometric and composite), as well as interferometric and compositional phases form [3]

As a final observation on the importance of elevation angle, from a geometrical perspective it is expected that at higher angles the reflection point on the surface on figure 2 (right image) will approach the antenna basement. As a result, the reflected component will be received closer to the anti-boresight direction of the antenna, where the gain pattern is minimum, consequently nulling the reflected wave. This explains why at higher elevation angles the reflected signal vanishes. Figure 6 illustrates how the reflection point on the ground approaches the antenna as the elevation angle increases, consequently being received at anti-boresight direction in figure 2 (right image).

4.3.1 Fresnel Zones

Ground reflections take place all around the antenna. Recall that GNSS antennas are omnidirectional in azimuth, thus they are capable of receiving signal from all azimuth directions at fairly equal gain. Figure 7 illustrates reflections surrounding the GNSS antenna. There is an empty region on the North which is the result from the satellite orbit inclinations and the North hemisphere where the antenna is located. This empty region would be on the South direction if the antenna was located in the South hemisphere [26].

Multipath reflections surrounding the antenna and their dependency on satellite elevation angle have been defined. Now the concept of Fresnel Zones can be introduced. Fresnel Zones are ellipses that denote the *ray thickness*, and the 1st Fresnel zone is

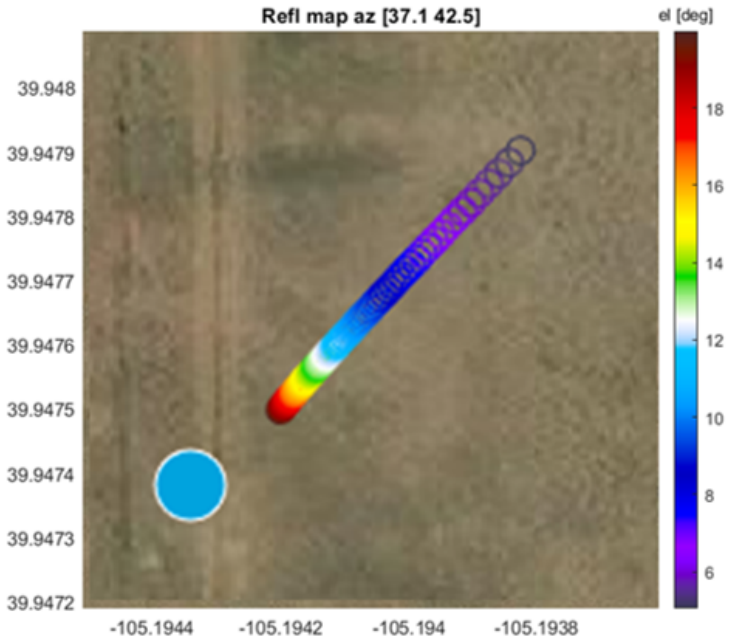


Figure 6: Reflection surface on the ground with respect to satellite elevation angle. Antenna illustrated as big cyan dot.

Figure 7: Illustration of ground reflections in the surroundings of the receiver antenna.

the region in which the signal is delayed up to half a wavelength [24]. Fresnel Zones can be thought as the sensing regions near the Earth's surface on which a particular satellite signal reflects off the ground to be later on received by the antenna. This is exemplified on figure 8 a).

Fresnel zones get smaller and closer to the antenna as the satellite elevation angle increases, and their size and orientation depends on the antenna height, satellite elevation angle, signal frequency and satellite azimuth angle [26]. This is exemplified

on figure 8 b).

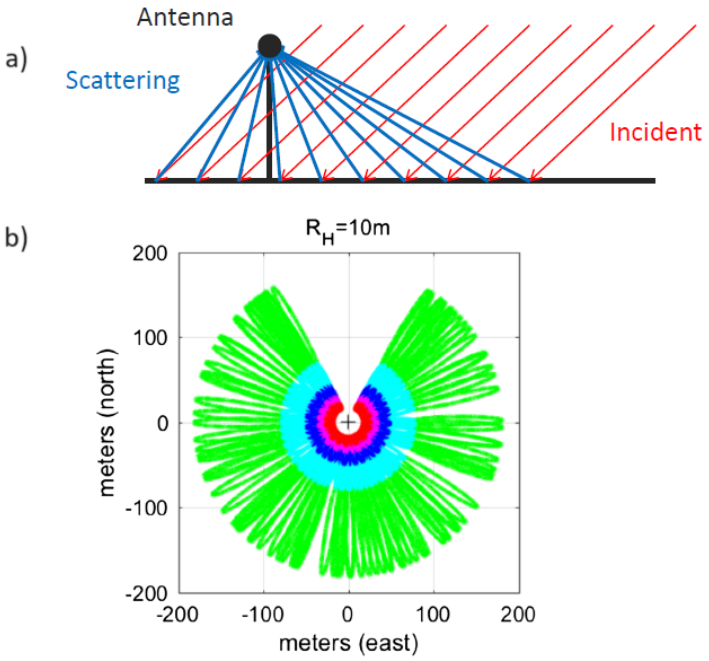


Figure 8: a) Illustration of ground Fresnel zones in 1D from [27]. b) Illustration of 2D Fresnel zones surrounding the antenna from [26]. On b), antenna height is 10m, and elevation angles (in degrees) are green (5), cyan (10), blue (15), magenta (20), and red (25).

4.3.2 SNR region of interest

Low elevation angles, also known as grazing angles, are the main regions of interest for estimating environmental characteristics whose effects are embedded in the signal power.We have seen that these angles are typically below 30 degrees. This region corresponds to the oscillations in figure 1, and whose behavior is described by equation (4). As was previously mentioned, the 1st term is principally driven by the direct component, while the 2nd term is mainly driven by the ground reflection. In low elevation angles, the 2nd term predominates, and therefore it is of our interest to extract this region on the SNR for post-processing to infer ECVs such as snow depth, vegetation presence or soil moisture.

In equation (4), the direct component acts as a *trend*, increasing the function when the satellite is rising, and decreasing it when the satellite is setting. It is essential to *detrend* this sequence, i.e. remove the effect of the direct component, to keep only the oscillations with zero-mean. This can be done with a low-order polynomial [26]. The

detrended SNR, representing ground reflection only, becomes

$$\begin{aligned} SNR_d =& 2P_d\sqrt{P_i}cos\left(\frac{4\pi H}{\lambda}sin(\theta)+\Phi_x\right) \\ =& A(\theta)cos\left(\frac{4\pi H}{\lambda}sin(\theta)+\Phi_x\right) \end{aligned} \tag{5}$$

where SNR_d denotes *detrended* SNR. Notice that the amplitude is now a function of the elevation angle (since interferometric power P_i is as well). The result of this deterend process is illustrated in figure 9.

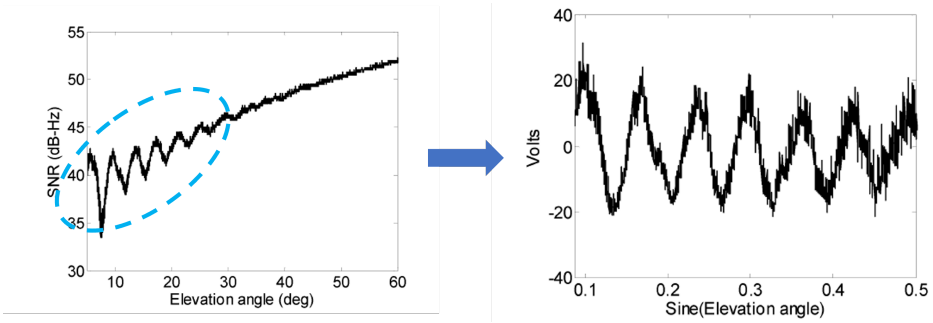


Figure 9: Selection of region < 30 degrees and detrending by removing the influence of the direct component. Illustration from [29]. Notice that the x-axis on the left image is the elevation angle in degrees, while on the left image is the sine of elevation angle, which is the domain variable on equation (5)

In order to keep coherency with literature on GNSS-IR soil moisture estimations [26, 1, 12], the cosine in equation (5) will be changed by a sine, which basically means a 180 degree shift, or sign inversion, to the result. This has no effects in the shape of C/N0 oscillations for ECV retrieval. For this reason, from now on the detrended SNR will be interpreted as follows

$$\begin{aligned} SNR_d =& A(\theta)cos\left(\frac{4\pi H}{\lambda}sin(\theta)+\Phi_x\right) \\ =& A(\theta)\cos\left(\frac{4\pi H}{\lambda}\sin(\theta)+\phi\right) \end{aligned} \tag{6}$$

where the environmental-influenced phase was renamed as $\phi = \Phi_x$ for notation simplicity.

After presenting the C/N0 shape (or equivalently SNR), the antenna pattern and multipath modelling influencing it, and introducing its mathematical expression and simplifications, we have arrived to equation (6), which is the starting point for GNSS Interferometric Reflectometry processing.

4.4 Multipath SNR parameters

GNSS-IR processes the C/N0 measurements following the model in equation (6), which resembles a signal modulated in amplitude, frequency and phase by the

parameters $A(\theta)$, H and ϕ. These parameters are strongly influenced by environmental characteristics and therefore they are the main attributes of interest to infer ECVs.

4.4.1 Antenna Height

The signal in equation (6) will slightly vary its frequency due to the $sin(\theta)$ multiplying as the domain variable. For this reason, the peaks in figure 9 are not equidistantly separated from each other, instead their separation increases slightly with the elevation angle.

However, the the sinusoidal argument can be rewritten as $2\pi\frac{2H}{\lambda}sin(\theta)$. This allows to define the so called *multipath frequency* [28], which is unique for the arc $0 < \theta < 30$ degrees.

$$f_{mp} = \frac{2H}{\lambda} \tag{7}$$

Since the antenna height is related to the frequency, we can expect faster oscillations on measurements captured from higher antennas. This is particularly useful in snow estimation [22]. Although the antenna height remains constant throughout all seasons, the snow accumulation will approach the apparent ground towards the antenna, thus changing the oscillation frequency since the receiver senses a lower antenna height. On figure 10 we can see different oscillations on equation (6) with varying antenna heights.

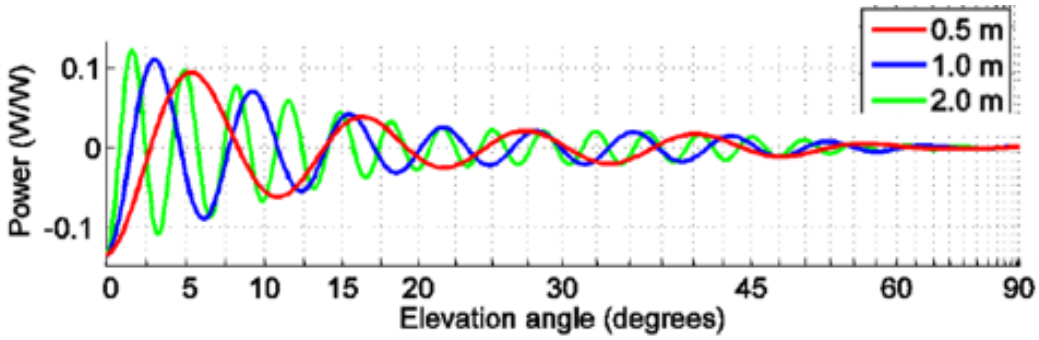


Figure 10: Different frequencies from equation (6) by varying antenna heights [3].

4.4.2 Amplitude

Amplitude is a good indicator on vegetation presence. Literature on GNSS-IR shows that although all three parameters are correlated with vegetation growth, amplitude presents good sensitivity for estimating vegetation cover, although this depends on how much moistened vegetation is [2].

On the other hand, SNR oscillations tend to vanish rapidly depending on the roughness of the reflecting surface. This is mainly a consequence of microwave signal propagation, since objects with similar size as the signal's wavelength will act as obstacles to the signal. The vanishing speed is increased with elevation angle, therefore

the already limited range of elevation angles below 30 degrees can be easily reduced to half depending on the surface roughness standard deviation [3]. This is illustrated on figure 11.

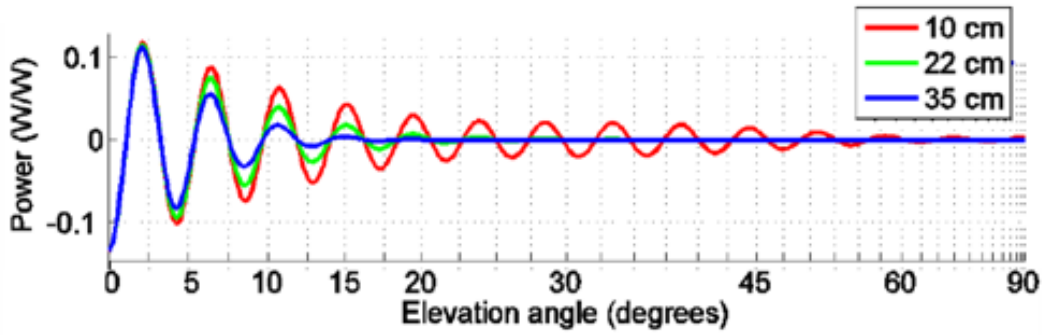


Figure 11: Vanishing amplitude with different surface roughness standard deviations from [3]. Recall that wavelengths from GPS bands L1, L2 and L5 are 19.05cm, 24.45cm and 25.48cm, respectively.

4.4.3 Phase

Literature shows the strong influence from vegetation on the phase ϕ [2, 13], but it is most widely used as an indicator for soil moisture [13, 1, 12]. Phase shifts caused by soil types and moisture levels are illustrated in figure 12.

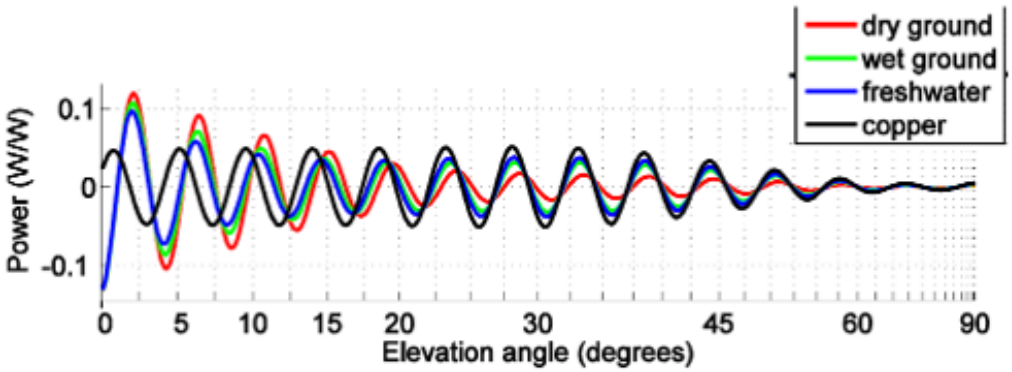


Figure 12: SNR phase from different reflecting materials and moisture levels [3].

Research done in phase estimation from equation (6) show a linear relationship between in-situ measurements collected with soil moisture probes and post-processed SNR phase, with R^2 of 0.91 measured in the average top 5cm of soil, and 0.997 measured on the soil surface [1]. Correlation between phase and in-situ soil moisture estimates are based on the top 5cm. This is relevant because other microwave-based remote sensing technologies, such as SMAP mission, also sense top 5cm, although in

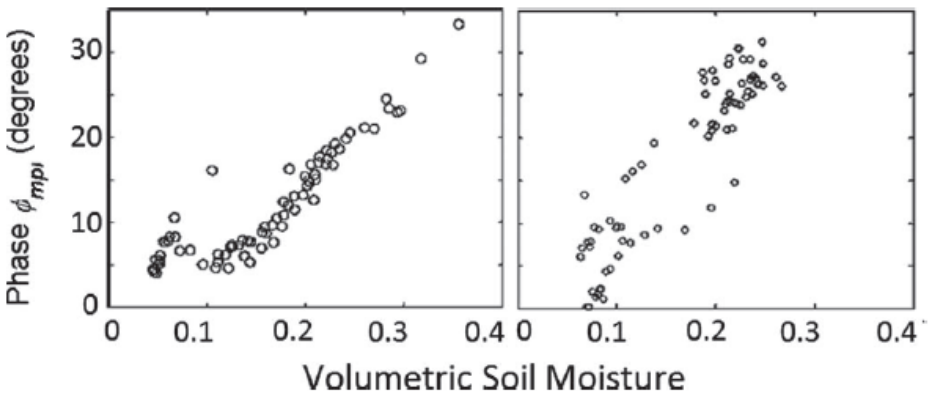


Figure 13: Relationship between phase and soil moisture [1]. Left) Low vegetation site. Right) Medium vegetation site.

a km-resolution scale. Linear relationship between soil moisture and phase values is shown in figure 13.

Phase post-processing will be detailed in section 7.2.6, but as illustrated on figure 13, a linear model can be constructed for translating phase values into soil moisture quantities. In [1, 2] this relationship is established with a scalar factor of 1.48 $\frac{cm^3}{cm^3 deg}$.

Despite providing high correlation with phase, soil moisture also presents good correlation with amplitude. Considering surface soil moisture, it has an R^2 of 0.997 and 0.86 with phase and amplitude, respectively [1]. On the other hand, vegetation presence also has high correlation in both phase and amplitude, with R^2 of 0.84 and 0.79, respectively [13]. Being the phase very sensitive to soil moisture variations, its utility for estimating vegetation presence is limited and amplitude is preferred. Figure 14 shows the vegetation effect on SNR in bare and vegetated soil scenarios.

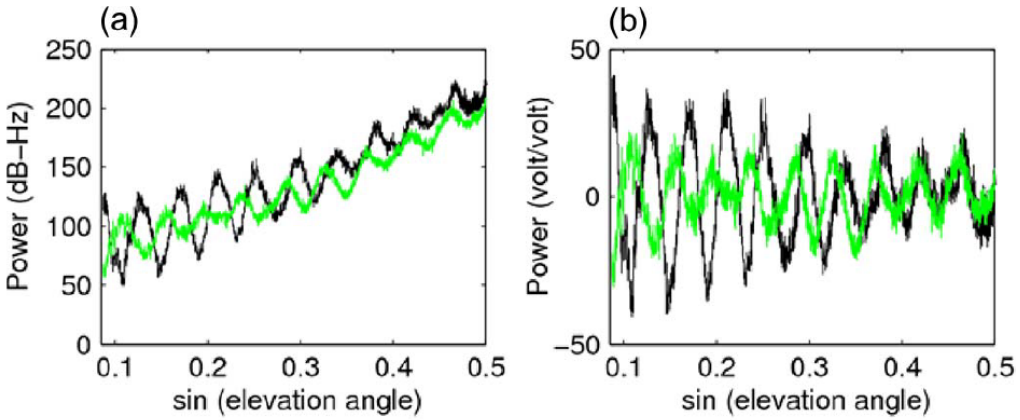


Figure 14: Bare vs vegetated soil and its effect in SNR at low elevation angles [13].

The strong correlation between SNR phase with soil moisture and vegetation indicate that there will be phase variations that are not only related to soil moisture, but also to vegetation cover. This is an important factor to take into account on phase-based

soil moisture retrieval algorithms, and for this reason phase values are *corrected* from vegetation [2, 13].

Section 7.2.5 describes how vegetation corrections are computed from amplitude values. Metrics based on amplitude estimations are used due to their high correlation with vegetation cover. With these metrics, a vegetation correction model is constructed to compensate for vegetation-induced phase variations, thus isolating moisture-specific phase values for a proper soil moisture retrieval algorithm [14].

5 Multispectral Satellite Imagery

The other technology used in this thesis is multispectral satellite imagery, which is a technology used for collecting data across various wavelengths of the optical electromagnetic spectrum. Unlike traditional photography, which captures images in visible light, multispectral imaging operates in both visible and non-visible wavelengths. This capability allows satellites to gather information about environmental features and processes whose perception is beyond the visible spectrum. Therefore, one of the key advantages of multispectral satellite imagery is its ability to extend our vision to different parts of the electromagnetic spectrum [18].

The use of multispectral imagery enables more accurate and comprehensive environmental assessments, providing crucial data for decision-making across numerous fields, including agriculture, forestry, environmental monitoring, and urban planning. In agriculture, for example, multispectral images can help farmers monitor crop health and optimize irrigation practices. In forestry, satellite data helps to track deforestation and forest health. Environmental monitoring benefits from the ability to detect changes in land use, water bodies, and pollution levels. Urban planners use this technology to map and manage urban sprawl, infrastructure development, diversification of green spaces and the so called urban heat islands. The coverage, continuity and data accessibility of multispectral imaging make it an invaluable tool in managing natural resources and planning sustainable development.

5.1 Landsat-8 Characteristics

Landsat 8 mission was launched in February 2013 and continues the Landsat program of Earth observation. The satellite the sensors Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). Landsat-8 captures data in the visible, near-infrared, shortwave infrared, and thermal infrared wavelengths.

The main characteristics of Landsat-8 are shown below [30].

- **Spectral Bands:** Landsat 8 collects data across 11 spectral bands, including visible, near-infrared, shortwave infrared, and thermal infrared.
- **Spatial Resolution:** The OLI sensor provides 30-meter resolution for visible, near-infrared, and shortwave infrared bands, and 15-meter resolution for the panchromatic band. TIRS offers 100-meter resolution for thermal infrared bands. Images are taken along approximately 185km by 180km.
- **Temporal Resolution:** 16 days revisit time.

Corresponding band names and spectrum range is presented in table 1.

In the visible range, the region is generalized to red, green and blue color bands, which conform the visible spectrum to the human eye. Beyond the visible range, the optical spectrum also includes near-infrared and shortwave-infrared wavelengths, which are not visible to the human eye but can be detected by specialized sensors.

Top of Atmosphere (TOA) and Surface Reflectance products are described below:

Band Number	Band Name	Wavelength (μm)	Resolution (m)
1	Coastal/Aerosol	0.43 - 0.45	30
2	Blue	0.45 - 0.51	30
3	Green	0.53 - 0.59	30
4	Red	0.64 - 0.67	30
5	Near Infrared (NIR)	0.85 - 0.88	30
6	Shortwave Infrared (SWIR) 1	1.57 - 1.65	30
7	Shortwave Infrared (SWIR) 2	2.11 - 2.29	30
8	Panchromatic	0.50 - 0.68	15
9	Cirrus	1.36 - 1.38	30
10	Thermal Infrared (TIRS) 1	10.60 - 11.19	100
11	Thermal Infrared (TIRS) 2	11.50 - 12.51	100

Table 1: Landsat-8 Bands, wavelengths, and resolutions [30].

- **Top of Atmosphere (TOA) Reflectance:** This product represents the reflectance measured at the satellite sensor, including atmospheric influences. TOA is useful for applications on the clouds and initial data assessments before atmospheric corrections.
- **Surface Reflectance (SR):** This product corrects for atmospheric effects to estimate the reflectance of the Earth's surface. Therefore, it provides more accurate data for vegetation, soil, and water bodies. The atmospheric corrections accounts for aerosols, water vapor, and ozone, which are crucial for applications requiring precise reflectance measurements.

In this implementation, the Surface Reflectance (SR) data is used, since the multispectral metrics are based on atmospherically-corrected data.

5.2 Spectrum bands

The optical spectrum refers to the range of electromagnetic wavelengths that are captured by sensors to produce images. This spectrum includes visible light as well as portions of the infrared and short-wave infrared spectra. Figure 15 illustrates spectrum regions for: ultraviolet (UV), visible (VIS), near-infrared (NIR), middle-infrared (MIR) and thermal-infrared (TIR). Although as described in 5.1, Landsat-8 has specific bands for SWIR, conceptually the combination of bands UV, VIS, NIR and MIR form the so called *showrtave* radiation, while TIR band forms the *longwave* radiation [18].

When illuminating an object light can be absorbed, transmitted, or reflected depending on the material's properties [18]:

$$\rho + \alpha + \tau = 1 \tag{8}$$

where ρ is the *albedo* or surface reflectance (ratio between input and output light radiance), α is the absorptance (absorbed energy) and τ is the transmittance (transmitted

energy). In spite of this, it is common in optical remote sensing literature to also only account for absorption and reflection.

Characteristics of the optical spectrum are crucial for understanding how light interacts with different materials. By capturing data in different regions of the spectrum, this technology can reveal subtle differences in surface materials and conditions. For instance, chlorophyll in plants absorbs most of the blue and red light for photosynthesis but reflects green light, which is why plants appear green [18]. In general, different materials, either natural or artificial, have unique spectral responses that can be used to identify them optical sensors with band combinations or spectral indexes.

Spectral indexes are metrics used to quantify features of interest. An example of this is the quantification of vegetation with the Normalized Differential Vegetation Index (NDVI), whose expression is presented on equation (9). NDVI which quantifies the amount of vegetation or *greenness* over a region and exploits the spectral signature of vegetation over red and NIR bands, since red band is highly absorbing while NIR reflects most of the incoming radiation.

The ability to distinguish between different materials and conditions based on their spectral characteristics makes multispectral imagery a powerful tool in remote sensing and environmental monitoring. Figure 15 illustrates the spectrum curves for different materials throughout the multispectral sensor range.

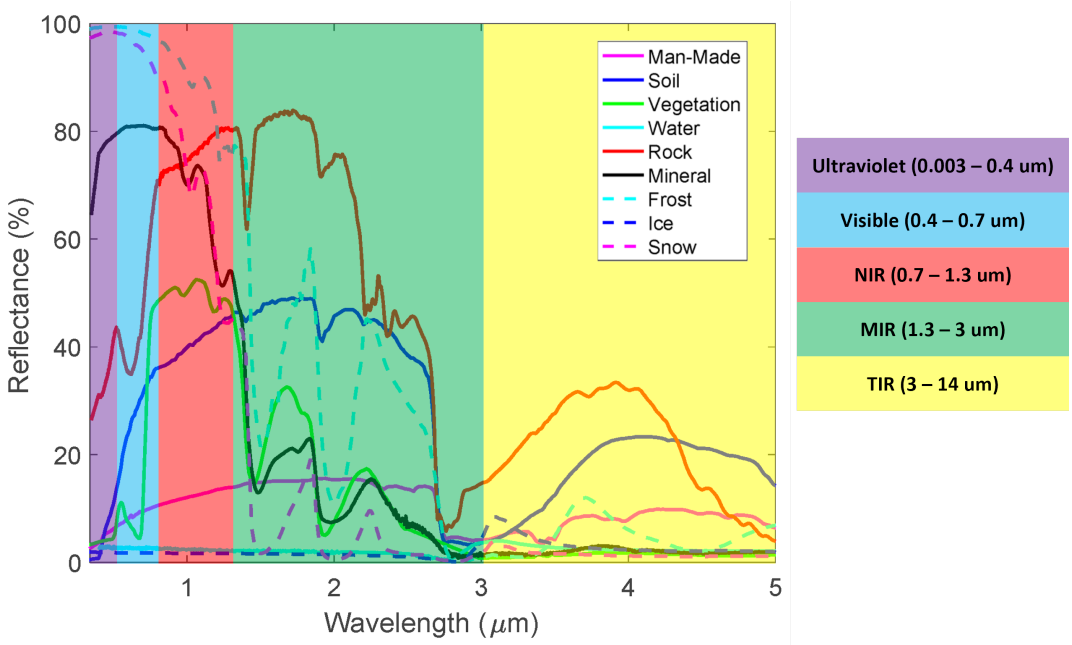


Figure 15: Typical multispectral reflectance spectrum response. Adapted from [33].

5.3 Quantities of Interest for Soil Moisture

In multispectral imagery, soil moisture can be detected and analyzed using specific spectral bands and indexes. Soil moisture affects the reflectance properties of soil in different spectral bands. Wet soils generally exhibit lower reflected energy in the visible (VIS) and near-infrared (NIR) regions compared to dry soils due to the higher absorption of water. This makes soils appear darker when moistened. The shortwave infrared (SWIR) bands in Landsat-8 is particularly sensitive to soil moisture content because water absorbs significantly in these wavelengths, especially in the water absorption bands centered at 1450 and 1900 nm [18].

Research in [4] shows that Land Surface Temperature (LST) and Normalized Differential Vegetation Index (NDVI) are highly correlated with soil moisture. LST is defined as the effective temperature of Earth's surface [31]. These quantities can be used to build models to quantify soil moisture such as the Thermal Optical Trapezoid Model (TOTRAM) [32], which estimates soil moisture based on the pixel distribution in the LST-NDVI domain. Alternatively to TOTRAM which uses LST, the Optical Trapezoid Model (OPTRAM) [32] uses the SWIR bands to compute a measure analogous to LST to be used in the optical domain.

An extensive comparison among several indexes which are indicators for soil moisture is done in [5]. This study concludes that OPTRAM and Visible and Shortwave infrared Drought Index (VSDI) present higher correlation with reference soil moisture values. These two measures reportedly outperform other indexes, and a reason is the spectrum characteristics towards soil and vegetation.

Expressions for NDVI and VSDI are given in (9) and LST is based on the Landsat-8 TIR band. OPTRAM and TOTRAM models defined in [32] are not utilized in this work. However, their inputs will be used used, NDVI and LST, as will be detailed in section 8.

$$\begin{aligned} NDVI &= \frac{NIR - RED}{NIR + RED} \\ VSDI &= 1 - (SWIR_2 - BLUE) + (RED - BLUE) \end{aligned} \tag{9}$$

where bands $BLUE$, RED, NIR and $SWIR_2$ are defined in 5.1.

Spectral response on different extents of dry and moistened soil and vegetation is shown in figure 16. The relationship between soil moisture and LST, NDVI and VSDI is related to the temperature, vegetation and water cycles as described below:

- Land Surface Temperature (LST): quantifies the thermal properties of the soil. Moist soils have higher thermal capacity and lower *surface emissivity*, resulting in lower surface temperatures compared to dry soils. Surface emissivity is a measure of how effectively a surface converts kinetic into radiant energy above the surface [31].

- Normalized Differential Vegetation Index (NDVI): since NDVI measures vegetation health and greenness, high values on NDVI typically indicate, indirectly, highly moistened soils.

- Visible and Shortwave infrared Drought Index (VSDI): combines visible and SWIR regions. In [5] that the spectral response due to soil moisture changes is

found to be relatively small in visible spectrum, but high in longer wavelengths, such as SWIR.

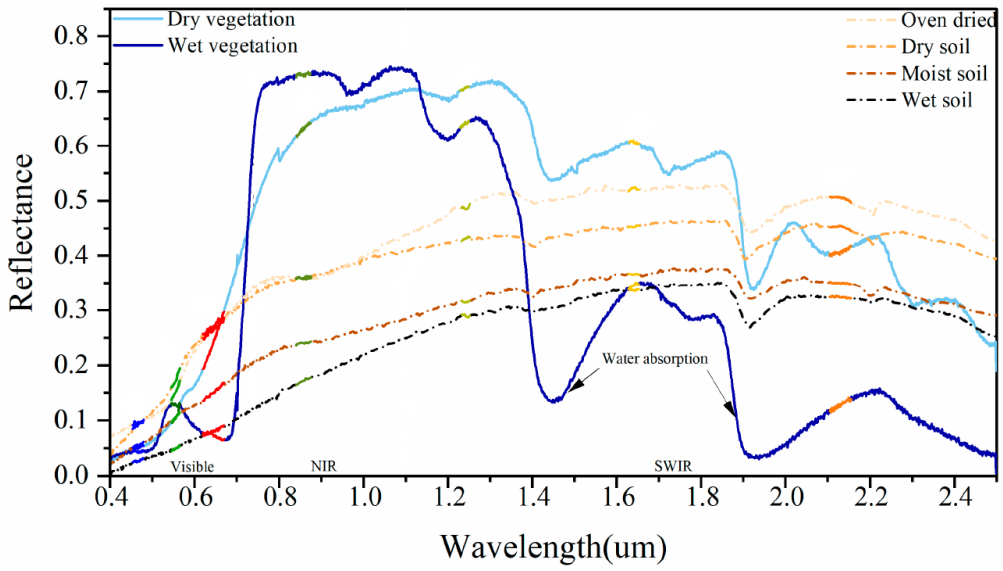


Figure 16: Spectral response of different moistened soils and vegetation [5].

6 Data

The data processed on GNSS-IR and multispectral model consist on GNSS and Landsat-8 Surface Reflectance data. This section explains each data source and its characteristics.

6.1 GNSS

GNSS data consists on public domain data recorded from reference stations with geodetic GNSS receivers. Data characteristics are shown in table 2.

Attribute	Description
Data Type	GNSS
Format	RINEX
Periodicity	Daily every 30 seconds
Data Source	UNAVCO
System	GPS
Used Frequency	L2
Collection Instrument	GNSS Geodetic Antenna
Coverage Area	Approximately radius of 50m [12]
Processing Tool	MATLAB

Table 2: GNSS Data Attributes

The stations belong to the UNAVCO network and the PBO H2O program [23]. Data processed in this work and whose results will be presented in section 9 correspond to the stations MFLE in Colorado and P267 in California. Table 3 shows the stations' coordinates and soil type, and figure 17 shows the GNSS stations, antennas and surrounding environment.

Station	Soil Type	Latitude	Longitude
MFLE	Arid, low vegetation	39.947640	-105.194439
P267	Medium vegetation	38.380335	-121.823234

Table 3: GNSS stations and sites

6.2 Landsat-8

Multispectral imagery used in this work corresponds to data from Landsat-8 satellite mission. In particular, Surface Reflectance dataset is used since that is already corrected for atmospheric effects, such as atmospheric particles or cirrus that could distort the captured radiance. Data characteristics are shown in table 4.

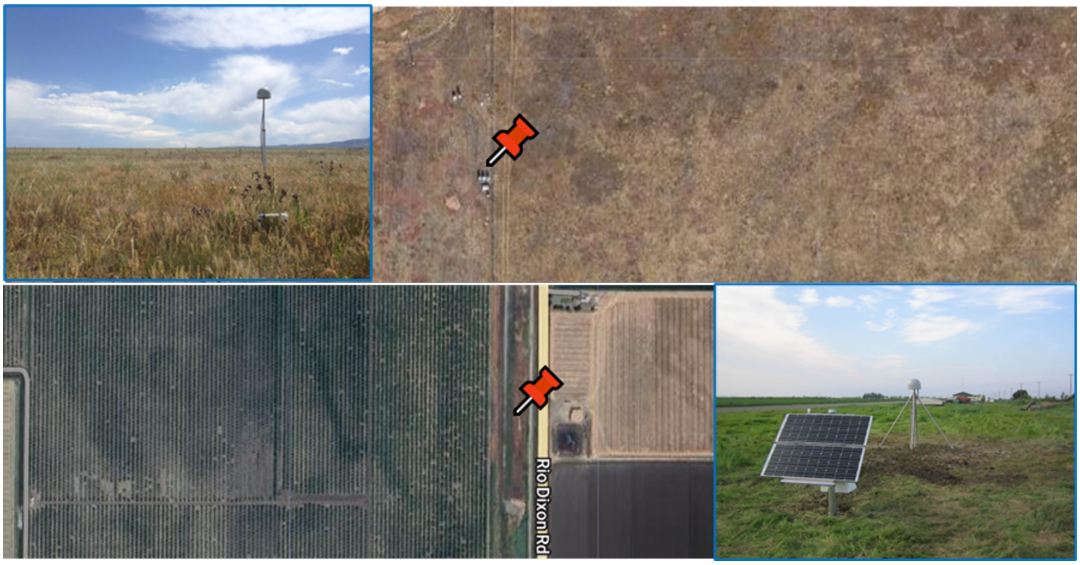


Figure 17: Stations MFLE (top) and P267 (bottom)

Attribute	Description
Data Type	Multispectral
Format	Google Earth Engine's Image Collection
Periodicity	Daily every 16 days
Data Source	USGS
System	Landsat-8
Used Frequency	VIS, NIR, SWIR and TIR bands.
Collection Instrument	GNSS Geodetic Antenna
Coverage Area	30 meter/pixel (TIR has 100 meter/pixel).
Processing Tool	Google Earth Engine

Table 4: GNSS Data Attributes

7 GNSS-IR Processing Chain

On section 4 the fundamentals of GNSS-IR were presented, covering aspects from the signal propagation and multipath perspectives, to the SNR influenced by ground multipath in low elevation satellites leading to the model in equation (6).

This section is focused on the GNSS-IR processing taking equation (6) as a starting point. The processing chain is responsible for selecting the SNR region of interest (i.e. low elevation angles), detrend it, estimate the SNR parameters described in section 4.4, and finally compute the vegetation-corrected soil moisture estimation from phase measurements. The processing to estimate a soil moisture product from GNSS-IR is divided in the following two steps and the block diagram is shown in figure 18.

1. **Daily Processing**: refers to the GNSS-IR processing from the daily RINEX inputs to estimate the parameters described in 4.4.

2. **Post-Processing**: where the daily estimated SNR parameters are gathered to compute the vegetation-corrected soil moisture.

7.1 Daily Processing

7.1.1 Arc Selection and Detrending

The information on the environmental characteristics from the reflecting surface is on the SNR oscillation. This is visible in the region of low elevation angles, predominantly below 30 degrees. Therefore, this region has to be extracted for further processing. This step corresponds to the *arc selection and detrending* and the aim is to extract the observations that follow the model in equation (6).

This can be done by just selecting the SNR observations that correspond to the elevation angles below 30 degrees. The resulting sequence follows an increasing trend. This is due to the presence of the 1st term in equation (4). Recall that the SNR has contributions from two terms: one term mainly driven by the direct component, and another driven by the ground multipath which causes oscillations. As previously explained, the term responsible of oscillations will vanish with increasing elevation angles, which occurs typically after 30 degrees.

For this reason, after extracting the region below 30 degrees, the influence of the direct component which causes the increasing trend has to be removed. This can be done by fitting and subtracting a low-order polynomial, typically with an order $n = 4$ [26]. After this operation, the 1st term in model (4) is removed, leading to the so-called detrended SNR that was presented in equation (6), illustrated in figure 9 (right image).

7.1.2 Reflector Height Estimation

Once SNR data is detrended, we can interpret the resulting sequence as a sinusoid signal modulated in amplitude, frequency and phase with the parameters of interest $A(\theta)$, H_0 and ϕ.

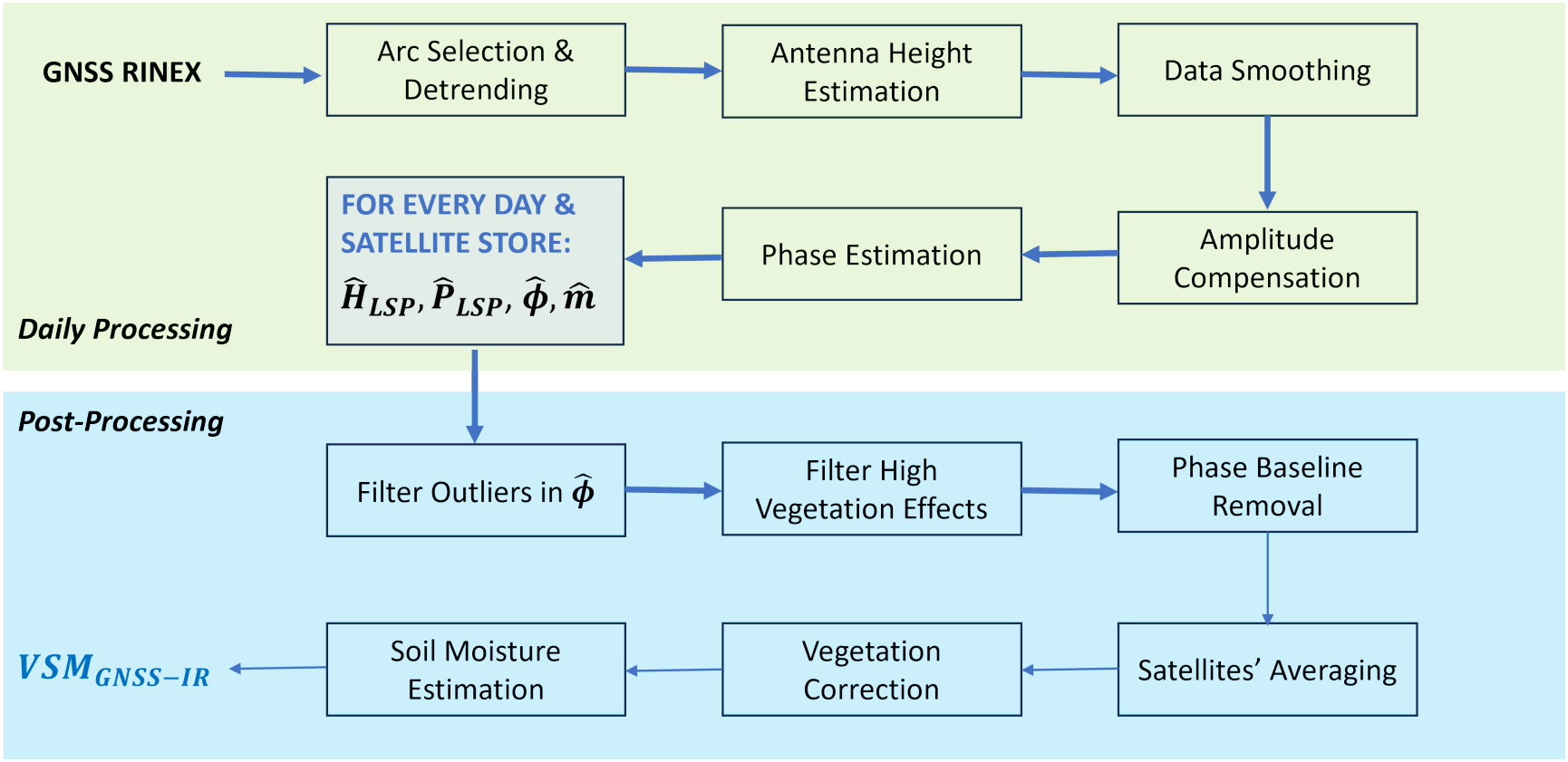


Figure 18: Block diagram of GNSS-IR chain for daily processing and post-processing stages.

Since soil moisture is highly correlated with the phase ϕ, this is the target parameter to estimate. However, the detrended SNR depends non-linearly on the phase, and any direct trial to estimate it would lead to inaccuracies in the final soil moisture product since the measurements also depend on amplitude and frequency (equivalently to the height). For this reason, the order of estimation matters, and the first parameter to estimate is the antenna height. After estimating the antenna height, we can proceed with the amplitude decaying factor estimation, compensate for it, and fit a least-squares model to calculate the phase.

As explained in section 4.4, the antenna height forms the multipath frequency of the detrended measurements which follow model in (6). Such frequency has the form detailed in equation (7), and it can be estimated as the sample index in the frequency domain with highest power. Following subsections will explain the reasoning for deriving the spectrum and estimate the antenna height.

7.1.2.1 Frequency Domain Representation

The frequency domain representation of a signal, also called spectrum, can be calculated with the Fourier transform. An infinite sinusoidal discrete signal with frequency w_0 and amplitude A_0

$$x[n] = A_0 \cos(w_0 n) \tag{10}$$

has a Discrete Time Fourier Transform (DTFT) that follows a Dirac delta function centered in the signal's frequency w_0. Dirac delta is defined as

$$\delta(z) = \begin{cases} 0 & \text{if } z \neq 0, \\ \infty & \text{if } z = 0, \end{cases} \tag{11}$$

The spectrum of signal $x[n]$ is defined as the Fourier transform in discrete domain (DTFT). DTFT maps discrete-time sequences into 2π periodic continuous functions and is defined as follows for infinite sequences [34]

$$X(e^{jw}) = \sum_{n=-\infty}^{\infty} x[n] e^{-jwn} \tag{12}$$

Since real signals are time-limited, such as the detrended SNR, the direct application of DTFT to understand the spectrum is not theoretically correct. However, we can interpret $x[n]$ as non-zero subset of length N part of an infinite set which is zero elsewhere

$$x_\infty[n] = \begin{cases} x[n] & \text{if } 0 < n < N, \\ 0 & \text{if elsewhere,} \end{cases} \tag{13}$$

Thus, the spectrum of $x[n]$ following DTFT is

$$X(e^{jw}) = \frac{1}{2} \left(\delta(w - w_0) + \delta(w + w_0) \right) \tag{14}$$

where $\delta(w \pm w_0)$ is the Dirac delta function defined in equation (11) with $z = w \pm w_0$. Therefore, the spectrum consists of two Dirac deltas centered at frequencies $w = \pm w_0$.

If the signal is limited in a time window, then the power concentrated in the Dirac spectral line centered around w_0 will be *spread* throughout a wider set of frequencies, leading to a *sinc*-shape rather than a Dirac delta. Thus, the spectrum of a time domain sinusoidal signal with a support of T seconds will be a *sinc* centered in w_0 in the frequency domain [34]. This is the result of the time-frequency property stating that shorter sequences in time domain result in wider spectrum in frequency domain. The *sinc* function is defined as

$$\text{sinc}(z) = \frac{\sin(\pi z)}{\pi z} \tag{15}$$

Thus, the spectrum defined in (14) turns into

$$X(e^{jw}) = \frac{1}{2}\left(\text{sinc}(w - w_0) + \text{sinc}(w + w_0)\right) \tag{16}$$

where due to the spectrum symmetry around zero for being a $\mathbb{R}$ signal [34], we can focus on the positive frequency term only, that is, the 2nd term.

It should be mentioned that regardless of the modelling in (13), which allows to write the DTFT result $X(e^{jw})$, the Discrete Fourier Transform (DFT) is employed for finite sequences

DTFT applies over infinite sets and has a continuous spectrum. On the other hand, DFT defined in (17), can be applied to finite sets and has a discrete spectrum since w is sampled based on the length N as $w = \frac{2\pi k}{N}$, where N is the discrete signal length, and k is the discrete frequency sample index (i.e. equivalent to n in time domain) [34]. Despite of this difference, DTFT and DFT share similarities regarding spectrum properties, although DTFT references to a frequency value, and DFT references to a frequency sample index.

$$X[k] = \sum_{n=0}^{N} x[n] e^{-j\frac{2\pi nk}{N}} \tag{17}$$

7.1.2.2 Spectrum Estimation

Now that the frequency domain representation of a sinusoidal was introduced and exemplified with DFT and DTFT, the explanation can be extrapolated to the GNSS-IR case. The detrended SNR which follows model in (6) could be thought as a time-limited sinusoidal with decaying amplitude centered at frequency $w_{mp} = 2\pi f_{mp}$. Such frequency has to be estimated to obtain the height estimation for H.

Spectrum estimation techniques based on DFT such as the classical periodogram can be utilized to estimate the frequency on which the maximum peak is located.

Classical periodogram is defined as [35]:

$$\begin{aligned} S_{classic}[k] =& \frac{1}{N}\left|X[k]\right| \\ =& \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi nk}{N}}\right|^2 \\ =& \frac{1}{N}\left[\left(\sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi nk}{N}\right)\right)^2 + \left(\sum_{n=0}^{N-1} x[n]\sin\left(\frac{2\pi nk}{N}\right)\right)^2\right] \end{aligned} \tag{18}$$

In our case, the sampling variable of the detrended SNR in (6) is the nonlinear an unevenly distributed term $\sin(\theta)$, not an integer sample index n. Therefore, notation in expression (18) will be slightly modified for two generalizations:

- frequency variable in terms of w.
- time domain sampling variable n replaced with t_n.

The expression of the classic periodogram in (18) with generalized frequency and sampling variable is defined in (19) (notice the replacement of brackets from "[]" to "()" to generalize the definition beyond discrete domain sequences):

$$S_{classic}(w) = \frac{1}{N}\left[\left(\sum_{n=0}^{N-1} x(t_n)\cos\left(wt_n\right)\right)^2 + \left(\sum_{n=0}^{N-1} x(t_n)\sin\left(wt_n\right)\right)^2\right] \tag{19}$$

However, there is one important observation to be done: classic periodogram is based on DFT, which is a function based on *evenly* sampled data, such as a signal sampled at a sampling frequency F_s, that is, one sample every $T_s = \frac{1}{F_s}$, known as the sampling interval. In the GNSS-IR case, the domain variable $\sin(\theta)$ is unevenly distributed.

The application of DFT, and consequently DFT-based spectrum estimators like the classical periodogram defined in (19) over *unevenly* sampled data lead to inaccurate spectrum estimations because irregular sampling affects the power spectrum [36]. The result shows increased noise and signs of aliasing (i.e. sampling frequencies smaller than the minimum acceptable Nyquist limit) due to the non-uniformity of the frequency during the signal snapshot [35]. Figure 19 exemplifies the spectrum distortion of an evenly and unevenly sampled Gaussian pulse.

For this reason, the signal's frequency cannot be estimated with classical periodogram, and an alternative method suitable for unevenly sampled data should be used to estimate f_{mp}, and consequently the parameter H. A common spectrum estimator used for unevenly sampled datasets is the Lomb-Scargle Periodogram.

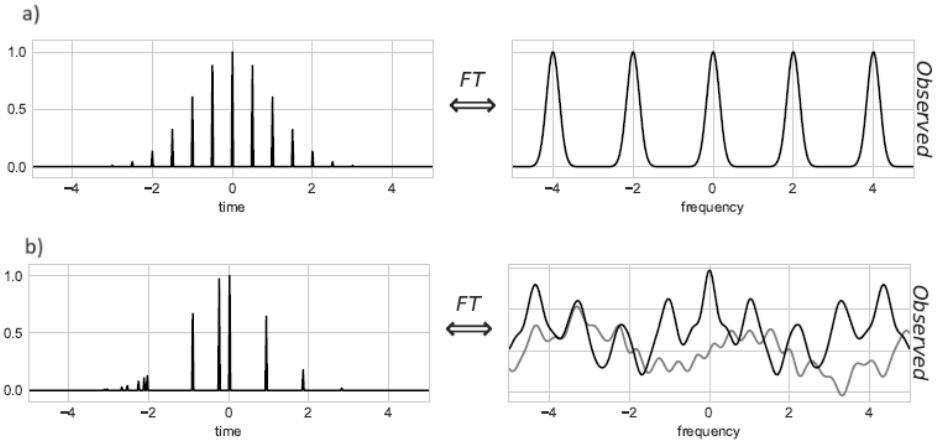


Figure 19: a) DFT over evenly sampled Gaussian pulse. b) DFT over unevenly sampled Gaussian pulse. [35].

7.1.2.3 Lomb-Scargle Periodogram

The Lomb-Scargle Periodogram (LSP) is a method used for efficiently computing the power spectrum of signals with unevenly sampled data. Unlike the classical periodogram, which assumes uniformly sampled data, the LSP adapts the Fourier transform to handle irregular sampling by fitting sine and cosine functions to the data through least-squares minimization [35].

Named after Nicholas Lomb and Jeffrey Scargle, the Lomb-Scargle Periodogram improves the sensitivity and accuracy of period detection, allowing for more robust identification of periodicity in datasets with gaps or uneven sampling intervals. LSP is particularly useful in astronomical time series analysis, where observations often have irregular time intervals. In the GNSS-IR case, LSP will be used to estimate the spectrum of the detrended SNR signal, and identify the height H as the index in x-axis that corresponds to the maximum peak.

To introduce the LSP expression, the same generalization is followed in terms of w and t_n as was done for the classical periodogram in equation (19). The LSP expression is defined in (20):

$$S_{LSP}(w) = \frac{A^2}{2}\left(\sum_{n=0}^{N-1} x(t_n)\cos(w(t_n-\tau))\right)^2 + \frac{B^2}{2}\left(\sum_{n=0}^{N-1} x(t_n)\sin(w(t_n-\tau))\right)^2 \quad (20)$$

where τ ensures the time-shift invariance of the result. Parameters A, B and τ are

functions of the frequency and observed sampling variables as

$$
\begin{aligned}
A^2 &= \frac{1}{\sum_{n=0}^{N-1} \cos^2(w(t_n - \tau))} \\
B^2 &= \frac{1}{\sum_{n=0}^{N-1} \sin^2(w(t_n - \tau))} \\
\tau &= \frac{1}{2w} \tan^{-1}\left(\frac{\sum_{n=0}^{N-1} \sin(2wt_n)}{\sum_{n=0}^{N-1} \cos(2wt_n)} \right)
\end{aligned}
\tag{21}
$$

These parameters allow the statistics of the periodogram to be analytically computable, and make the LSP converge to the classical periodogram in the case of evenly sampled data [35]. For instance, notice that if the data is evenly sampled, then $\tau = 0$ and parameters $A^2 = B^2 = \frac{2}{N}$. This makes the LSP expression in (20) to equal the classical periodogram in (19). Therefore, LSP can be utilized in both evenly and unevenly sampled sequences.

Figure 20 illustrates the application of LSP for the non-uniformly sampled SNR data in model (6). The LSP outputs are the estimated antenna height $\hat{H}_{LSP}$ and the estimated spectrum peak magnitude $\hat{P}_{LSP}$. The next processing stages for phase and soil moisture estimations.

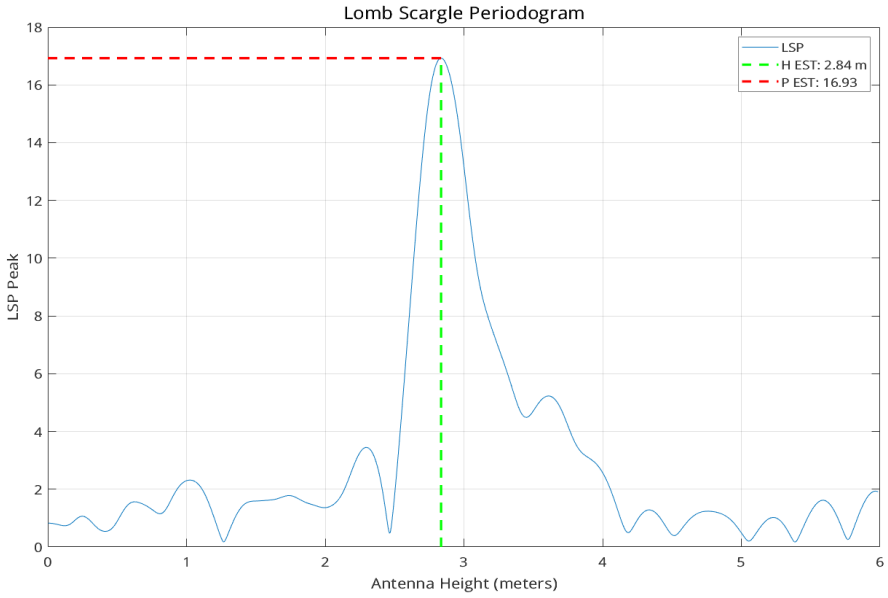


Figure 20: LSP computation for detrended SNR arc.

7.1.3 Amplitude Compensation and Phase Estimation

Once the antenna height is estimated, the decaying factor can be estimated and compensated prior to estimating the phase. Thus, the first steps in this stage are the

estimation of the initial amplitude and decaying factor. To do this, the attenuation in $A(\theta)$ has to be modelled. The model in (6) can be rewritten as

$$SNR_d = A_0 e^{m \sin(\theta)} \cos\left(\frac{4\pi H_0}{\lambda} \sin(\theta) + \phi\right) \tag{22}$$

where A_0 and m are the initial amplitude and decaying factor, respectively. These parameters, together with the phase ϕ will be estimated this stage. To properly estimate the decaying factor, the sinusoidal signal is sampled at the peaks as illustrated in figure 21, and a least-square fit is applied.

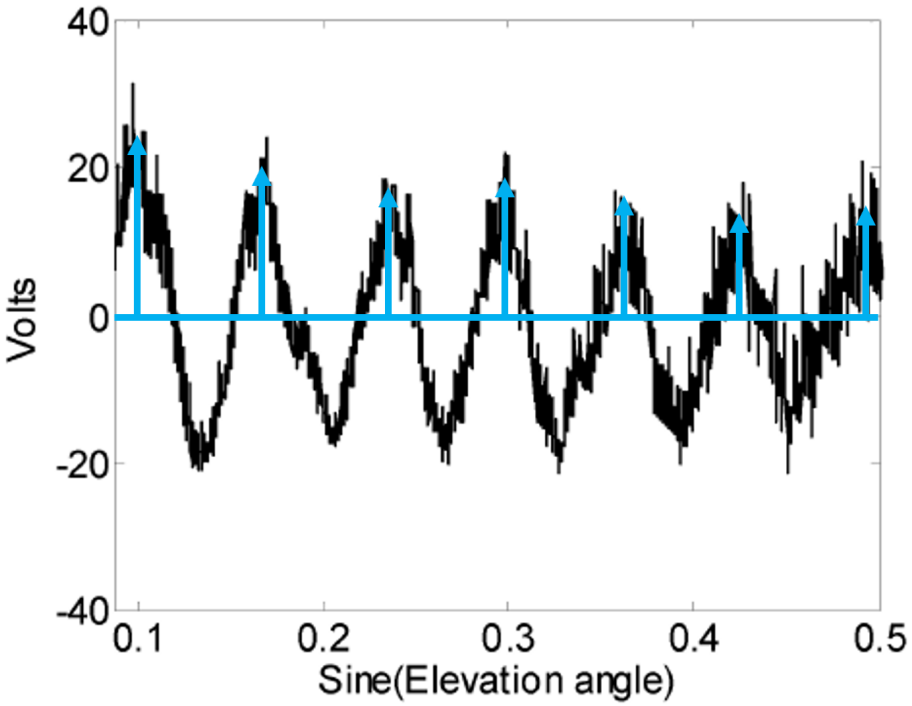


Figure 21: Sampling of positive peaks following detrended SNR in figure 9.

However, the data is noisy and therefore subsequent peaks are not strictly smaller than the previous, as one would expect in a noise-free sequence with decaying amplitude. For this reason, a signal smoothing is applied prior to the sampling process. The estimation of decaying factor $\hat{m}$, amplitude compensation and posterior phase estimation $\hat{\phi}$ is divided in three steps:

1. **Data Smoothing**: to smooth the noisy data before sampling the peaks. This is done to avoid wrong peak detection due to noise.

2. **Decaying Factor Estimation**: this samples the peaks and fits the decaying amplitude to estimate the factor $\hat{m}$. After doing so, amplitude can be compensated before estimating the phase.

3. **Phase Calculation**: this is the final step to estimate the phase from the smoothed and amplitude-compensated data.

7.1.3.1 Data smoothing

In general, a filtering process is needed to reduce the high variation due to noise. One could simply build a Finite Impulse Response (FIR) filter, but for this the sampling frequency and consequently signal's maximum frequency have to be taken into account. As stated in section 4.4, the detrended SNR will slightly increase the frequency as elevation angle increases. Due to the varying frequency, and to keep the processing simple, a different approach for smoothing is implemented. The applied smoothing follows an L_2 norm minimization of the data subject to minimizing the *energy* of the 2nd derivative [37]:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{D}\mathbf{x}\|_2^2 \tag{23}$$

where

- $\mathbf{y}$ is the noisy SNR data as an Nx1 vector.
- $\mathbf{x}$ is the smoothed data as an Nx1 vector. It is the unknown to calculate and upon which the peak sampling will be based to estimate the decaying factor.
- $\mathbf{D}$ NxN Toeplitz matrix of 2nd order derivative *taps*, such that Dx is a discrete form of the 2nd order derivative.
- λ is a regularization parameter which controls how smooth the solution will be. For this reason, $\lambda > 0$. Therefore, $\lambda = 0$ means no smoothing at all, and the solution $\mathbf{x}$ becomes smoother as λ increases.

Calculating the 1st derivative of (23) and equaling to 0 results in the expression for $\mathbf{x}$:

$$\mathbf{x} = (\mathbf{I} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{y} \tag{24}$$

7.1.3.2 Decaying factor estimation

After smoothing the data, the peaks can be sampled to estimate a least-square fit for the decaying factor m. Peak sampling is illustrated in figure 21.

Amplitude $A(\theta)$ is modeled in equation (22) as $A(\theta) = A_0 e^{\hat{m} \sin(\theta)}$. This model is utilized to build a least-squares fitting to estimate decaying factor $\hat{m}$ and initial amplitude $\hat{A}_0$. When applying the natural logarithm to the model we obtain:

$$\ln(A(\theta)) = \ln(A_0) + m \sin(\theta) \tag{25}$$

which in matrix form can be written as

$$\begin{bmatrix} \ln(A(\theta_0)) \\ \ln(A(\theta_1)) \\ \dots \\ \ln(A(\theta_N)) \end{bmatrix} = \begin{bmatrix} 1 & m \sin(\theta_o) \\ 1 & m \sin(\theta_1) \\ \dots & \dots \\ 1 & m \sin(\theta_N) \end{bmatrix} \begin{bmatrix} \ln(A_0) \\ m \end{bmatrix} \tag{26}$$

Let us now denote $\mathbf{s}$ as the vector of $\ln(A(\theta_i))$. For simplicity and with no impact on the performance, we can interpret it as the natural logarithm of the estimated

smoothed vector in equation (24), that is $\mathbf{s} = \ln(\mathbf{x})$. Denoting $\mathbf{H}$ as the matrix in the right-hand-side, and $\mathbf{z}^T = [\ln(A_0)\ m]$, the initial amplitude and decaying factor can be estimated with a least-squares fit as:

$$\begin{aligned} \hat{\mathbf{z}} =& (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{s} \\ \hat{\mathbf{z}}^T =& [\ln(\hat{A}_0)\ \hat{m}] \\ \text{so that} & \\ \hat{m} =& z_1 \\ \hat{A}_0 =& e^{z_0} \end{aligned} \tag{27}$$

where z_0 and z_1 are the 1st and 2nd items of $\mathbf{z}$ respectively.

7.1.3.3 Phase calculation

Now that the only unknown parameter on (22) is the phase ϕ, we can apply the sinusoidal trigonometric identity in (28) and rewrite the expression (22). The smoothed $\mathbf{x}$ is used rather than the original noisy data, as a design choice to work with a less noisy sequence:

$$\alpha\cos(A+B) = \alpha\cos(A)\cos(B) - \alpha\sin(A)\sin(B) = \mathbf{x} \tag{28}$$

we can identify $\alpha = \hat{A}_0 e^{\hat{m}\sin(\theta)}$, $A = \frac{4\pi\hat{H}_0}{\lambda}\sin(\theta)$ and $B = \hat{\phi}$.

To estimate B, the terms $\sin(B)$ and $\cos(B)$ will be treated as unknowns to be estimated from a least-square fit. After that, the inverse tangent can be applied to find the actual value of $B = \phi$. Notice that this isolation of ϕ is possible because the phase has a constant behavior [3], since is driven by the environmental conditions in a relatively small surface area. Modelling in matrix form:

$$\begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_N \end{bmatrix} = \begin{bmatrix} \hat{A}_0 e^{\hat{m}\sin(\theta_0)}\cos\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_0)\right) & -\hat{A}_0 e^{\hat{m}\sin(\theta_0)}\sin\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_0)\right) \\ \hat{A}_0 e^{\hat{m}\sin(\theta_1)}\cos\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_1)\right) & -\hat{A}_0 e^{\hat{m}\sin(\theta_1)}\sin\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_1)\right) \\ \dots & \dots \\ \hat{A}_0 e^{\hat{m}\sin(\theta_N)}\cos\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_N)\right) & -\hat{A}_0 e^{\hat{m}\sin(\theta_N)}\sin\left(\frac{4\pi\hat{H}_0}{\lambda}\sin(\theta_N)\right) \end{bmatrix} \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix} \tag{29}$$

So being $\mathbf{x}$ the smoothed data vector in the left-hand-side, $\mathbf{H}$ the matrix in the right-hand-side, and $\mathbf{z}^T = [\cos(\phi)\ \ \sin(\phi)]$, we can estimate the phase ϕ as:

$$\begin{aligned} \hat{\mathbf{z}} =& (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \\ \hat{\mathbf{z}}^T =& [\cos(\hat{\phi})\ \ \sin(\hat{\phi})] \\ \text{so that} & \\ \hat{\phi} =& tan^{-1}\left(\frac{z_1}{z_0}\right) \end{aligned} \tag{30}$$

where $z_0 = \cos(\phi)$ and $z_1 = \sin(\phi)$ are the 1st and 2nd items of $\mathbf{z}$ respectively.

Figure 22 shows a snapshot with real data illustrating the process for smoothing, and the estimations of decaying factor and phase. The raw SNR data is shown in blue.

Smoother data in orange, and sampled peaks in black. The estimated decaying factor is shown in green.

After the phase is estimated, a synthetic sinusoidal is generated which should match the noisy and smoothed data. Synthetic sinusoidal is shown in violet when the phase is calculated with $\hat{H}_{LSP}$ from this specific satellite track, that is, the LSP output. Synthetic sinusoidal shown in yellow corresponds to a signal with a phase estimated using a common reference $H_{REF} \neq \hat{H}_{LSP}$.

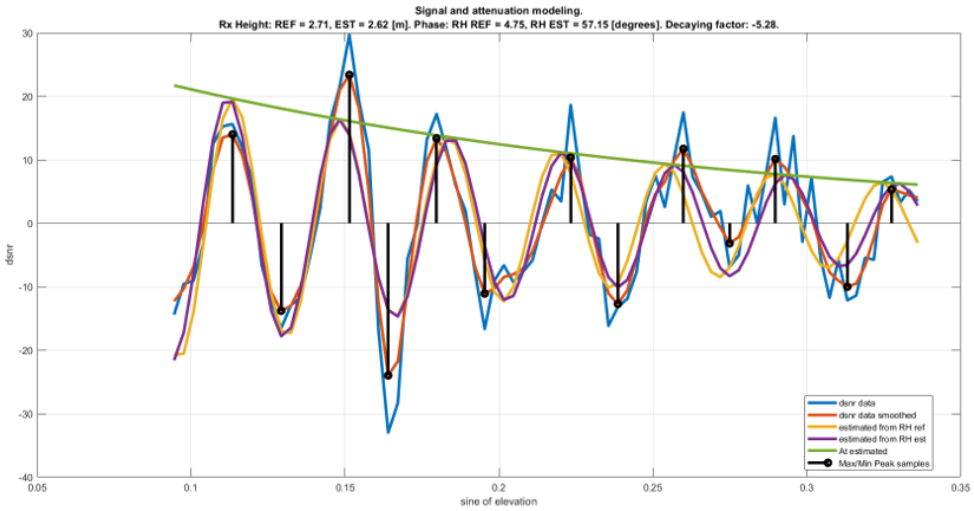


Figure 22: Illustration of raw data, smoothing, peak sampling and posterior decaying factor and phase estimation.

7.2 Post-Processing

7.2.1 Outlier Filtering

The phase estimations on which the soil moisture model is based can contain outliers. Apart from vegetation, whose effect can be corrected, events such as snow or heavy rainfall can also affect the phase, what results in wrong soil moisture estimates. To account for this, a two-step filtering is applied to remove outliers:

1. First, we can assume that LSP peak should meet certain peak-to-noise ratio, considering as *noise* any other secondary peak in the LSP. Therefore, a threshold peak-to-noise ratio $p2n$ is set such that days with lower $p2n$ values are discarded.

2. Second, a filtering based on the $\hat{H}_{LSP}$ is applied. This will discard any day whose estimated antenna height lies outside of the $k\sigma$ region from the median value, with $0 > k > 3$ to account for the desired k-sigma region around the median.

Given that the antenna height does not change throughout the entire data period, monitoring the $\hat{H}_{LSP}$ variation helps to detect possible snow accumulation or heavy

rainfall. Similarly, days whose measurements are extremely weak judging by the $p2n$ are discarded as well. The $p2n$ threshold and height's k-sigma regions are design parameters, and different configurations will be shown in section 9. Figure 23 shows the result of such filtering for a single satellite on one year of data with $p2n = 1$ and $k = \frac{1}{2}$.

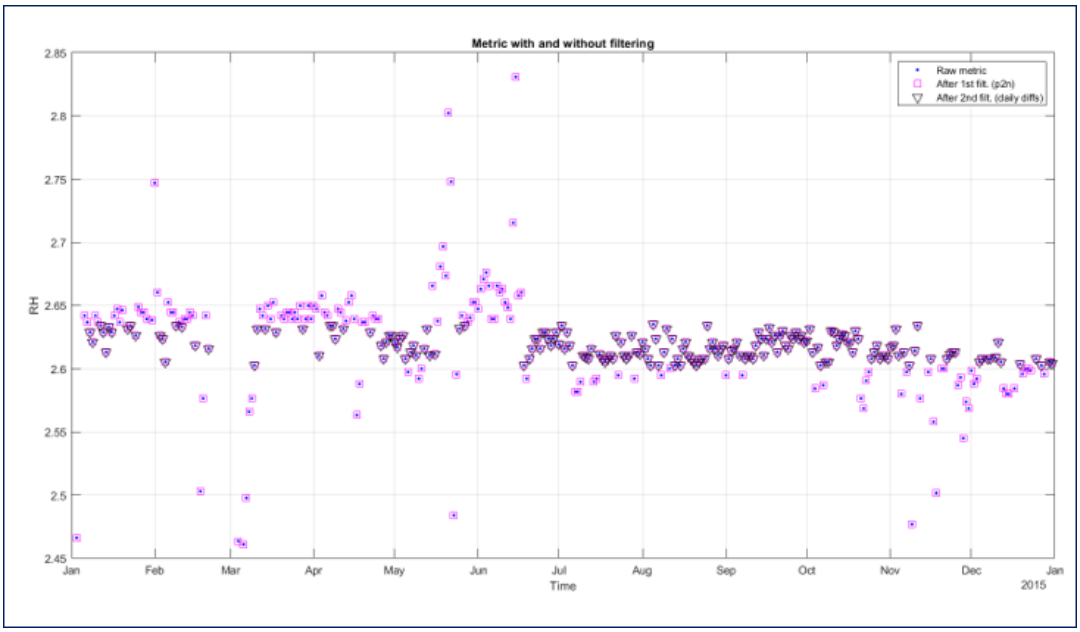


Figure 23: Outlier filtering based peak-to-noise ratio and variation of antenna height estimation. Blue denotes the raw data points, pink corresponds to after the 1st filtering, and black corresponds to after the 2nd filtering.

.

7.2.2 Filter High Vegetation Effects

Although vegetation effects will be corrected, as described in literature has shown that there is a limit in the vegetation density up to which the correction can be applied [2, 14]. For this reason, it is important to identify the days on which vegetation is high and discard them so that the correction algorithm can compensate for low and medium vegetation effects.

High corrupted values are identified with the stored LSP peak P_{LSP}, normalized to the maximum value throughout a period portion as

$$P_{LSP}^{norm} = \frac{P_{LSP}}{P_{LSP}^{XtopN\%,T}} \tag{31}$$

where $P_{LSP}^{XtopN\%,T}$ is the median (X operator) of the top N% of LSP peaks during a period T, where T can be, for instance, a year. The median of the top N% is used instead of the maximum to be more robust against sporadic outliers.

With normalization, P_{LSP}^{norm} is forced to range between 0 and 1. Values closer to 1 indicate little vegetation, and values closer to 0 indicate higher vegetation, since the LSP peak decreases with vegetation density. Therefore, periods in which the normalized LSP peak drops below certain threshold are considered highly vegetated days [2]. Following the literature, this threshold is set to approximately $0.7 - 0.8$, since the vegetation algorithm to be described in 7.2.5 was not originally designed for P_{LSP}^{norm} under those thresholds [14]. Further details on vegetation effects will be described in 7.2.5.

7.2.3 Phase Baseline Removal

Phase ϕ was estimated with equation (30) for each day and satellite. After all daily phases are calculated, each estimated phase is post-processed. In this step phases are uniformized by removing the baseline values. Baseline is considered as the median value calculated over the low-vegetation season [2, 14]. Thus, the baseline is calculated from the days on which P_{LSP}^{norm} is higher than 0.9. This gives place to the baseline-removed phase also known as *phase zeroed* ϕ_z:

$$\phi_z = \phi - X_{P_{LSP}^{norm} > 0.9}(\phi) \tag{32}$$

where X operator represents the *median* X, which corresponds to the baseline phase value to be removed. This is done to focus only on the phase variations, which is where information on soil moisture is.

From this point onwards in the processing chain, the terms phase-zeroed and phase are used interchangeably as is also done in the literature. However, it is important to keep in mind that the phase-zeroed from this current section 7.2.3, is nothing but the estimated phase in 7.1.3.3 with the baseline removed. This way, only phase variations are kept, from which soil moisture estimates will be calculated.

Figure 24 shows, in a four quadrant division, the baseline-removed phase values in blue. Furthermore, the corresponding normalized LSP peak values are shown in orange. Data processed for these results corresponds to the Northern hemisphere where summer season is in July.

Notice how the LSP peak values follow a trend which decreases towards the Summer season. This is because the vegetation grows during Spring. Some days contain sporadic $\hat{P}_{LSP}$ values which drift considerably from the rest, such as in Winter. These are remaining outliers due to snow or heavy rainfall which could not be filtered out with $\hat{H}_{LSP}$. However, in section 9 it will be shown how increasing the number of daily satellites to average can reduce the influence of outliers.

7.2.4 Satellite Averaging

Since multiple satellites are visible each day, many measurements for the parameters defined in 4.4 are available. If M satellite tracks are present, this leads to M measurements of phi_z and P_{LSP}^{norm}. This stage computes one single daily quantity for the phase and normalized LSP peak by averaging all M daily measurements. Previously

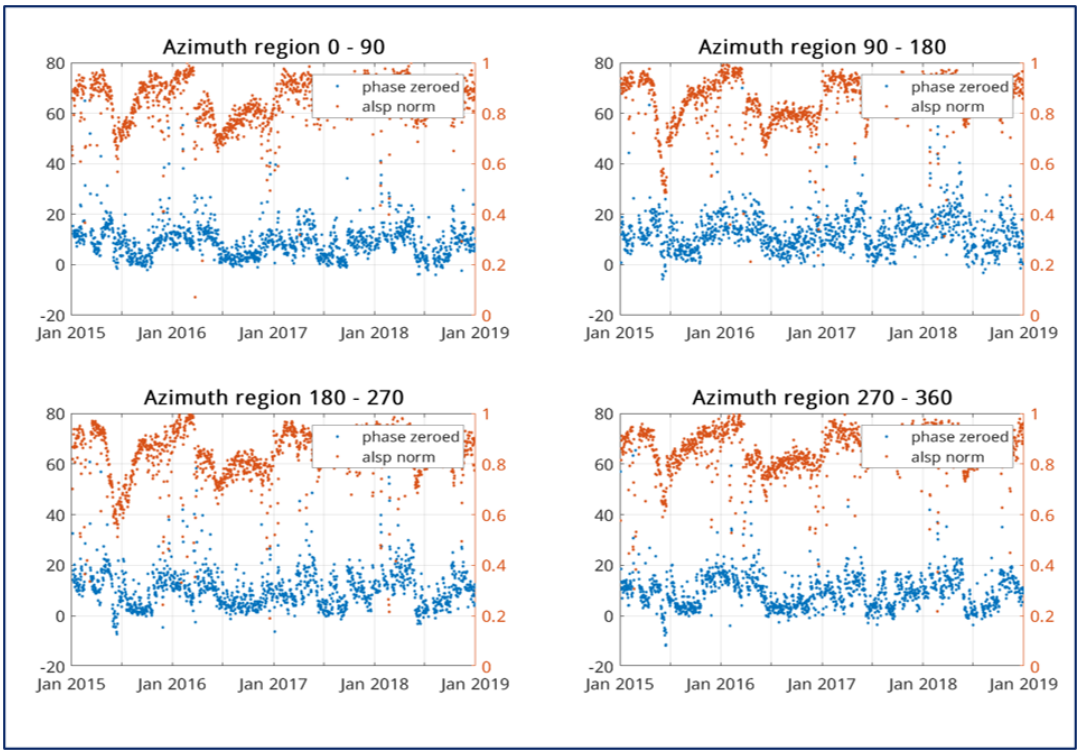


Figure 24: Baseline removal and normalized LSP peak monitoring.

it was shown how measurements can contain remaining outliers which might not be completely filtered out with $\hat{H}_{LSP}$ in 7.2.1, and averaging can help to reduce them.

Outliers can occur on any of the four quadrants where the signal reflects, such as snow accumulation or water puddles after a heavy rainfall. Nevertheless, not necessarily all measurements are contaminated by outliers, but just a few. For this reason, a weighting average is performed. Instead of weighting quadrants, this approach weights the measurements to avoid highly deviating values from the median. This is done following a Gaussian weighting function as:

$$
\begin{aligned}
w =& e^{\left(\frac{(y-X(y))^2}{\sigma^2}\right)} \\
&\text{with} \\
y =& \phi_z \text{ and } P_{LSP}^{norm} \\
X =& \text{median operator} \\
\sigma =& \delta * \sigma(y) = \delta\text{-tuned standard deviation,}
\end{aligned}
\tag{33}
$$

where $0 < \delta < 1$ is used to widen or shrink the range of acceptable values around the median (either ϕ_z or P_{LSP}^{norm}). By doing this, highly *uncommon* values that correspond to outliers or bad measurements are downweighted as illustrated on figure 25.

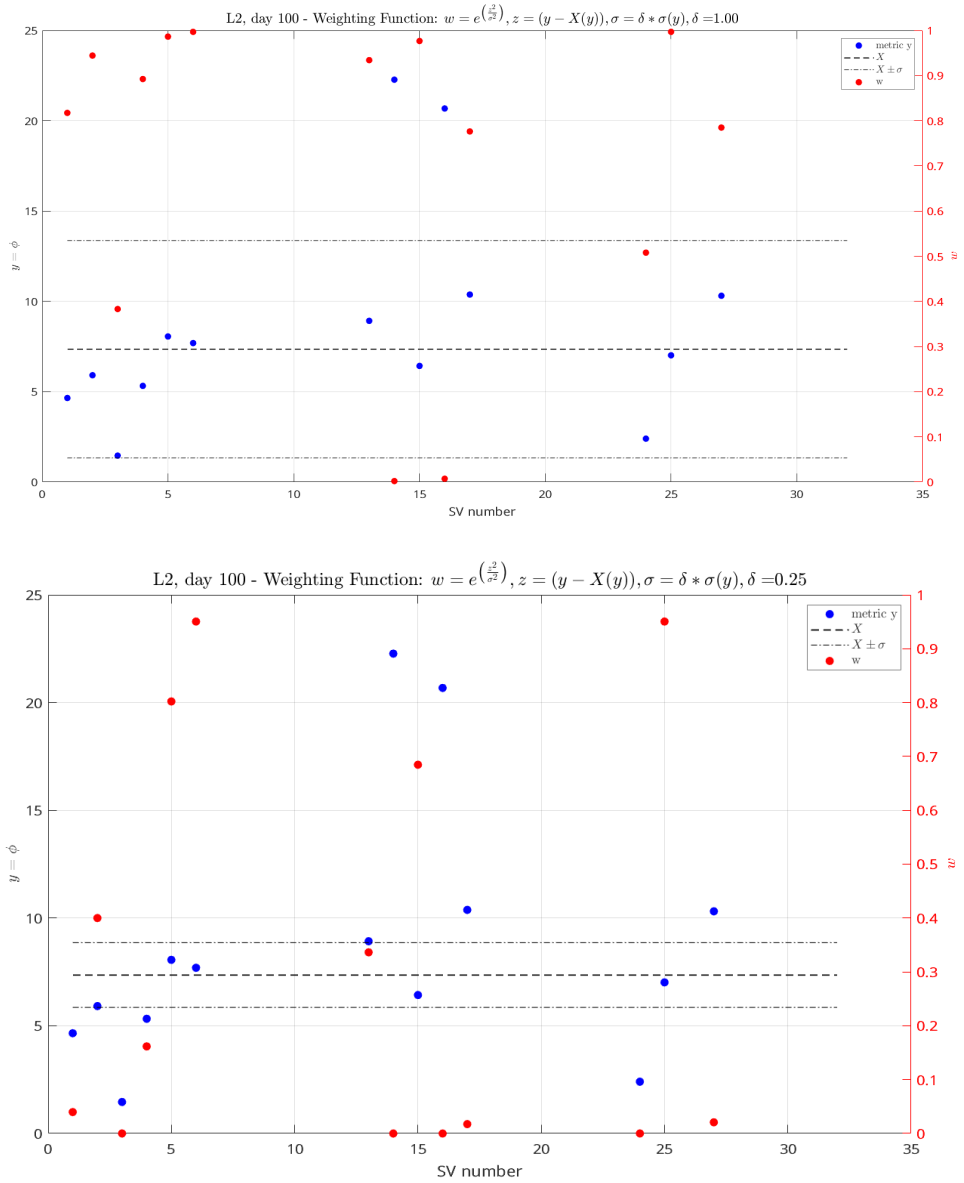


Figure 25: Weighted average for phase-zeroed values for $\delta = 1$ (top) and $\delta = \frac{1}{4}$ (bottom). Phase-zeroed values ϕ_z are shown on the left y-axis in blue, and the corresponding weighting value is shown on the right y-axis in red.

7.2.5 Vegetation Correction

Parameters on detrended SNR are strongly influenced by vegetation. As mentioned in section 4.4, vegetation affects phase similarly as soil moisture does and this effect has to be compensated. On figure 14 it was illustrated how vegetation affects the SNR shape.

The following two subsections explain the effects on amplitude and height estimation, and the influence on phase followed by the correction algorithm description.

7.2.5.1 Influence on Amplitude and Height estimation

At this point, after having explained the vegetation effects on the SNR and presenting the LSP, the effects of vegetation on spectrum can be detailed. Vegetation cover affects the shape of the LSP as shown in figure 26, where the antenna height decreases and shifts with respect to the bare soil. This makes sense because:

- Amplitude decreases with vegetation cover, also visible in the power spectrum peak due to the amplitude-power relation. This affects the estimation $\hat{P}_{LSP}$
- Phase also changes with vegetation cover, and following the Fourier transform properties, a phase change in time-domain is translated into a frequency-shift in the spectrum [34]. This affects the estimation $\hat{H}_{LSP}$.

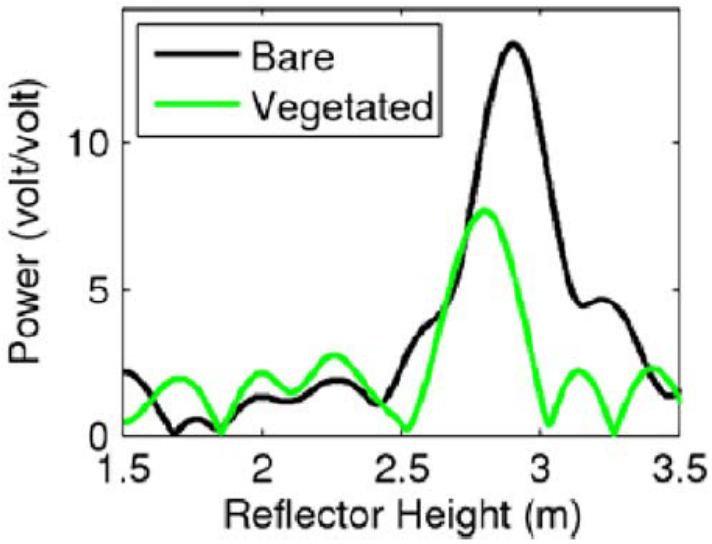


Figure 26: Effect on vegetation in detrended SNR spectrum from LSP [13].

7.2.5.2 Influence on Phase estimation

The vegetation correction algorithm from [14] is followed, which is based on model simulations of GNSS reflections on uniform vegetation canopies. This algorithm reportedly compensates for vegetation presence in VWC up to 1 kg/m^2. The method utilizes the normalized P_{LSP} measurements, estimates the amplitude-only VWC, and extrapolates the corresponding phase shift $\Delta\phi_{veg}$ to be used for correcting ϕ_z in (34). The phase correction is applied as:

$$\phi^* = \phi_z - \Delta\phi_{veg} \tag{34}$$

where ϕ^* represents the vegetation-corrected phase zeroed, and $\Delta\phi_{veg}$ is the computed phase correction.

Using the amplitude, in particular, the LSP peak P_{LSP} for the vegetation correction is a valid approach due to the direct relationship between spectrum power and time-domain amplitude. Therefore, the high correlation of amplitude with vegetation cover can be seen in both $A(\theta)$, the detrended SNR amplitude, and in P_{LSP}, the LSP peak (consequently also in the normalized peak P_{LSP}^{norm}).

The vegetation correction calculation is described below:

1. Smooth the LSP peak by applying an N-day filtering of the daily P_{LSP}^{norm}, where $N = 30$ days. Filtering is needed to remove the soil-moisture effects on amplitude.

2. Estimate amplitude-based VWC with 4th order polynomial:

$$VWC = \sum_{k=0}^{4} \alpha_k (P_{LSP}^{norm})^k \tag{35}$$

3. Predict phase change due to vegetation with 4th order polynomial:

$$\Delta\phi_{veg} = \sum_{k=0}^{4} \alpha_k VWC^k \tag{36}$$

The coefficients α_k depend on the vegetation parameters, such as canopy height, and the default parameters in [14] are used since they provide good results for this soil moisture application. Notice that for $k = 0$, the coefficient α_0 is just an offset for each model. Coefficients α_k for both equations (35) and (36) can be found in table 5.

	α_0	α_1	α_2	α_3	α_4
Equation (35)	5.24	-22.6	41.8	-34.9	10.6
Equation (36)	-2.37	20.4	-101	43.9	-5.65

Table 5: α_k coefficients for vegetation correction algorithm from [14].

7.2.6 Soil Moisture Estimation

The vegetation corrected phase values from equation (34) have to be converted into VSM [2, 14] following equation (37):

$$VSM = s(\phi^* - \phi_r) + VSM_r \tag{37}$$

where $s = 1.48 \frac{cm^3}{cm^3 deg}$ is a slope factor accounting for the linear relationship between phase and in-situ soil moisture measurements as described in 4.4, $\phi*$ is the vegetation-corrected phase values from equation (34), ϕ_r is a soil moisture residual which accounts as the median for the lowest N%, and VSM_r is a soil moisture residual which accounts for the soil texture. In this implementation, $N = 10\%$ is chosen. Regarding VSM_r, although it is location-specific, it can be found for multiple soil types surrounding PBO H2O network stations in UNAVCO databases [38, 23]. Results from (37) are shown in section 9.

8 GNSS-IR aided Multispectral Model

GNSS-IR is a reliable approach for estimating soil moisture both in terms of accuracy and meter-level resolution but it is still a local solution. Global alternatives for soil moisture estimation include SMAP satellite mission, which is a global reference for soil moisture monitoring, also based in microwave L-band similar to GNSS. However, SMAP provides km-level resolution, only suitable for large-scale monitoring.

An alternative for global but still meter-level coverage is needed. A solution for this is to use multispectral satellite imagery such as Landsat-8. Multispectral satellite data does not provide a direct moisture product, therefore it can be combined with local GNSS-IR results and extrapolate accordingly to larger areas. This approach provides accurate, high resolution (meter-level), economic and global soil moisture estimations, which can be used for applications including agriculture and landslide monitoring.

TECHNOLOGY	COVERAGE	RESOLUTION	DIRECT SM PRODUCT
GNSS-IR	Local	Meter	Yes
SMAP	Global	Km	Yes
LANDSAT	Global	Meter	No

Figure 27: Comparison among GNSS-IR, SMAP and Landsat. Positive characteristics for high-resolution soil moisture estimations are highlighted in green, and undesired characteristics in orange. "SM" stands for "Soil Moisture".

The comparison among GNSS-IR, SMAP and Landsat-8 is shown in figure 27 in terms of coverage, resolution and their direct applicability for soil moisture. Ideally it is desired to have a global coverage and meter level solution, but these cannot be obtained from SMAP or GNSS-IR. However, multispectral imagery such as Landsat provides these two, but requires to build a model for estimating soil moisture, which is done in this implementation using GNSS-IR as a local aiding source.

As we have seen in section 5, certain bands are more sensitive to soil moisture, allowing to construct indexes to quantify it. The aim is to develop a model capable of combining several indexes that are known to be fair indicators of soil moisture. Following literature findings described in 5.3 about VSM estimations from multispectral data, the measures of LST, NDVI and VSDI are of particular interest since they are highly correlated with soil moisture.

The implemented model is based on LST, NDVI, and VSDI. These parameters form the input observations to the multispectral soil moisture estimation algorithm. Besides these metrics, the difference between Red and Blue bands is also used to combined with LST, since it has been observed that this provides better correlation with VSM than LST standalone. Behavior of Red and Blue bands on wet soils can be seen in figure 16. Since Blue band presents more absorption than Red band, positive differences can indicate presence of moistened content in the reflected source. The difference between Red and Blue bands is named "REBL".

8.1 Processing Chain

The multispectral metrics (LST, REBL, NDVI and VSDI) are used to fit the coefficients of a linear regression model taking GNSS-IR estimated VSM as reference outputs. Figure 28 shows the Landsat-8 data acquisition and processing chain to fit the model and calculate a multispectral-based soil moisture estimate on larger land areas.

The utilized GNSS-IR data corresponds to stations MFLE and P267 described in section 6. Multispectral satellite images were clipped to an area of similar size than the covered by GNSS-IR. Since GNSS-IR coverage area radius is of approximately $\sim 50m$ for the $\sim 2m$-tall antenna, the Landsat-8 images were clipped to areas of approximately 100-by-100 meters.

Processing blocks in figure 28 corresponding to Google Earth Engine module:

1. **Masking of pixels with clouds, cloud shadow, water and snow**: given that surface reflectance values contain information on the surface properties, it is important to only gather the reflected light from such surface. For this reason, clouds and snow can lead to excessive reflectance values. Analogously, if there are shadows due to the clouds, reflections will present lower values and this can be mistakenly associated with surface absorption properties, while in fact there is less incoming energy due to the cloud shadow. Furthermore, we do not want to calculate a moisture level of a water body. Thus, pixels containing clouds, cloud shadow, water and snow have to be masked out.

2. **Clipping to desired area size**: as previously mentioned, Landsat-8 images cover sizes of 185km x 180km. For this reason, images are clipped to the area of interest. In this application, such area corresponds to an approximation of the GNSS-IR coverage area, so the images are clipped to 100m x 100m.

3. **Convert Digital Number (uint16) to physical quantity**: images on each band are coded on unsigned 16. This is a measure to quantify the values on the image, but is not a physical quantity. In order to link the surface reflectance to physical quantities, the digital number (also known as pixels in raw uint16 format) have to be converted following a linear model such as $s \cdot \text{px} + b$, where px is the pixel and s and b represent a scale factor and a bias, respectively. This conversion is done for all bands.

4. **Calculate parameters and average over area**: this represents the overall parameter and spectral index calculation to be post-processed and verify their relationship with GNSS-IR VSM. After calculating the parameters for each pixel, this is reduced to calculate an area average for each parameter, e.g. averaged NDVI over the 100m x 100m area.

5. **Multispectral VSM**: this calculates the multispectral-based VSM from the model with the obtained coefficients on a 100m x 100m area, extrapolating it to a larger area of similar terrain characteristics. The coefficients are calculated on the MATLAB module explained below.

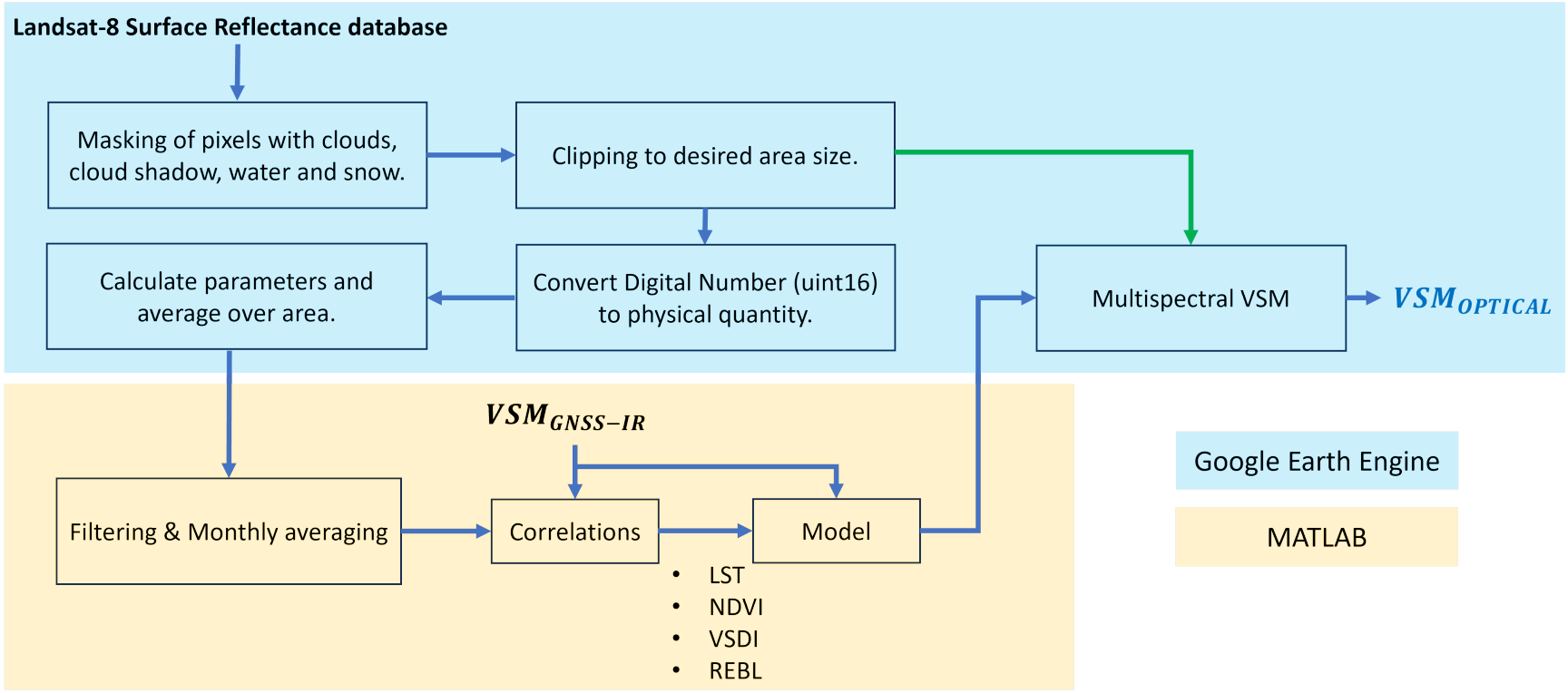


Figure 28: Data acquisition and processing chain for Landsat-8.

Processing blocks in figure 28 corresponding to MATLAB module:

1. **Filtering & Monthly averaging**: data from multiple years is collected and N-day filtered. N is chosen to cover one month of measurements, so it can vary between GNSS and Landsat data since GNSS measurements are daily, but Landsat-8 measurements are computed every 16 days. After filtering, an inter-annual median is computed for each month to reduce the multiple-year time series to a 12-month behavior. Median is chosen instead of the mean to avoid being affected by highly deviated values, such as remaining outliers. Figure 29 shows an example of this 12-month reduction for LST and NDVI for station P267.

2. **Correlations**: this block performs the correlations between Landsat-8 computed metrics and reference $VSM_{GNSS-IR}$. This is done to identify what are the most convenient multispectral metrics to be used for inputs to the algorithm.

3. **Model** this blocks performs the coefficient fit as:

$$VSM_{optical} = \alpha_1 LST \cdot REBL + \alpha_2 NDVI + \alpha_3 VSDI \tag{38}$$

   GNSS-IR soil moisture estimation used as a reference $VSM_{GNSS-IR}$ is also filtered and 12-month reduced. The result allows to calculate the multispectral $VSM_{optical}$ through the estimated coefficients α_1, α_2 and α_3.
   An offset parameter α_0 could be applied, but if other parameters (LST, REBL, NDVI and VSI) were zero, having an offset $\alpha_0 \neq 0$ means that there is a fixed "default" soil moisture, which has not physical meaning if the aim is to quantify the VSM. Therefore, the model is strictly limited to depend on multispectral parameters and coefficients calculated by fitting with GNSS-IR.

That the model in (38) written in matrix form is

$$\mathbf{VSM}_{optical} = \mathbf{H}\boldsymbol{\alpha} \tag{39}$$

and the coefficients α_i in model are estimated with a least-squares fitting using GNSS-IR results as

$$\boldsymbol{\alpha} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{VSM}_{GNSS-IR} \tag{40}$$

The vectors $\mathbf{VSM}_{optical}$, $\mathbf{VSM}_{GNSS-IR}$ and $\boldsymbol{\alpha}$ are of size Lx1, and $\mathbf{H}$ represents the LxK matrix of multispectral measurements. In this implementation, $L = 12$ since the monthly averages of spectral indexes and GNSS-IR VSM are used, and $K = 3$ since multispectral inputs are $LST \cdot REBL$, $NDVI$ and $VSDI$. The components of each

vector and measurement matrix is detailed below:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \ \mathbf{VSM}_{GNSS-IR} = \begin{bmatrix} VSM^{Jan}_{GNSS-IR} \\ VSM^{Feb}_{GNSS-IR} \\ \dots \\ VSM^{Dec}_{GNSS-IR} \end{bmatrix}, \ \mathbf{VSM}_{optical} = \begin{bmatrix} VSM^{Jan}_{optical} \\ VSM^{Feb}_{optical} \\ \dots \\ VSM^{Dec}_{optical} \end{bmatrix} \tag{41}$$

$$\mathbf{H} = \begin{bmatrix} LST_{Jan} \cdot REBL_{Jan} & NDVI_{Jan} & VSDI_{Jan} \\ LST_{Feb} \cdot REBL_{Feb} & NDVI_{Feb} & VSDI_{Feb} \\ \dots & & \\ LST_{Dec} \cdot REBL_{Dec} & NDVI_{Dec} & VSDI_{Dec} \end{bmatrix}$$

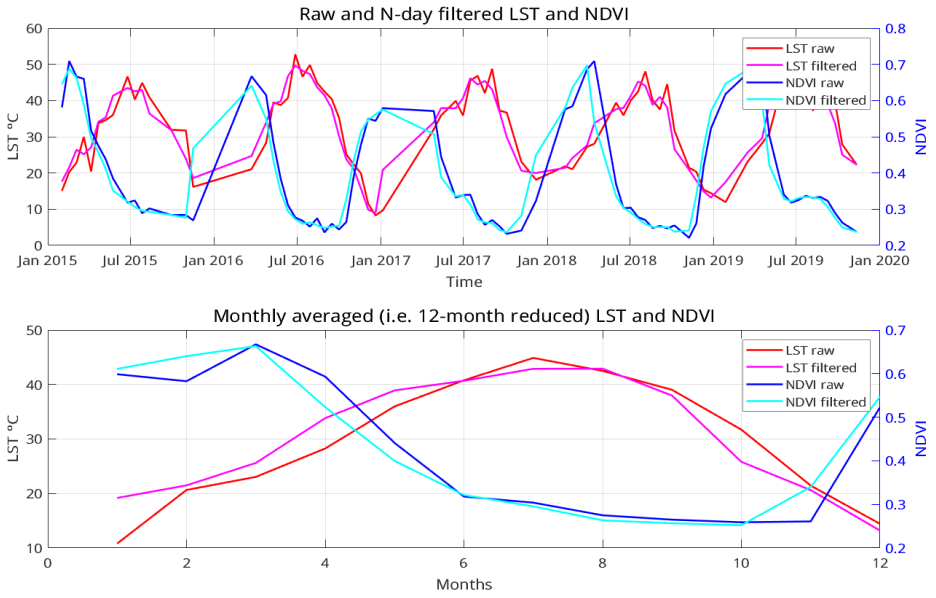


Figure 29: Example of yearly raw, filtered and monthly averaged LST and NDVI for P267 station.

8.2 Terrain-Specific Behavior

From figure 17 we can see that the environmental characteristics of both station sites is quite different from each other. MFLE is located in an arid region with little vegetation, while P267 is located in a medium-vegetated environment surrounded by crops. Table 6 shows the monthly averaged correlation values between GNSS-IR soil moisture and the multispectral data metrics on which the algorithm is based.

Notice that from table 6, station P267 has higher correlation values than MFLE, especially in NDVI and VSDI. The reason for this is that these indexes quantify vegetation density and drought, quantities with which soil moisture is positively and highly correlated due to the water need from vegetation to grow.

R with $VSM_{GNSS-IR}$	LST	$NDVI$	$VSDI$	$REBL$	$LST \cdot REBL$
MFLE	-0.44	0.03	0.37	-0.49	-0.53
P267	-0.67	0.87	0.90	-0.81	-0.75

Table 6: Correlation values between multispectral data and GNSS-IR soil moisture estimations on MFLE and P267 stations. For GNSS-IR, the optimal design configuration parameters in table 8 are used.

P267 is a mid-vegetated area, while MFLE is an arid region. Therefore, it is expected that NDVI and VSDI are highly correlated and have more dependency with soil moisture in terrains such as P267, opposite to arid regions such as MFLE. This little dependence with VSM in arid regions make the correlation become low or even negative. On the other hand, LST is negatively correlated with soil moisture in both terrain types. This is expected, since warmer seasons tend to have less moistened lands. Despite the negative correlation with LST, it should not be understood that high LST maps to low VSM.

Vegetation and drought also play an important role in VSM since without sufficient soil moisture vegetation cannot grow. Areas close to Equator are an example of this, where LST is not necessarily low and vegetation is high. Therefore, VSM is expected to be high accordingly. In arid regions such as MFLE vegetation is scarce, so among the analyzed parameters it is LST which will principally drive the soil moisture estimation in the model. Regarding p267 station, since it is a medium-vegetated site, all parameters will play a significant role in estimating VSM.

This suggests that the applicability of the model fitted with local estimations from GNSS-IR VSM can be extrapolated to larger areas, as long as the terrain characteristics remain sufficiently similar. This behavior is normal, since is equivalent to a model using a specific dataset for training purposes, and then restrict the testing or application to datasets sharing similar characteristics. Otherwise, the training would not be representative of the testing scenario, and a different training stage would be needed.

Therefore, there is a terrain-specific behavior on the model application and it cannot be extended without restrictions. Instead, different fittings are needed in order to create a *database* of coefficients applicable to different terrain types. This is also implemented, and the in model (38) is controlled with NDVI to determine the extent of vegetation and apply the corresponding coefficients:

$$VSM_{optical} = f(\alpha) \begin{cases} \alpha = \alpha_{\text{LOW VEG}} & NDVI < \gamma_{lm}, \\ \alpha = \alpha_{\text{MED VEG}} & \gamma_{lm} < NDVI < \gamma_{mh}, \\ \alpha = \alpha_{\text{HIGH VEG}} & NDVI > \gamma_{mh}, \end{cases} \tag{42}$$

where $f(\alpha)$ is model in (38). $\alpha_{\text{LOW VEG}}$, $\alpha_{\text{MED VEG}}$ and $\alpha_{\text{HIGH VEG}}$ correspond to the coefficients estimated for low, medium and high vegetated areas, respectively. Vegetation extent is determined with thresholds γ_{lm} (low-medium vegetation boundary) and γ_{lm} (medium-high vegetation boundary).

9 Results

This section presents soil moisture results obtained from GNSS-IR standalone in section 7, as well as from the multispectral model described in 8.

9.1 GNSS-IR only estimation

The processed data corresponds to GPS L2 frequency band between the years 2015 and 2019 for the MFLE station described in 6.

A snapshot of the parameters $\hat{H}_{LSP}, \hat{m}, \hat{\phi}_{LSP}$ and LSP peak-to-noise ratio $p2n$ is shown in figure 30.

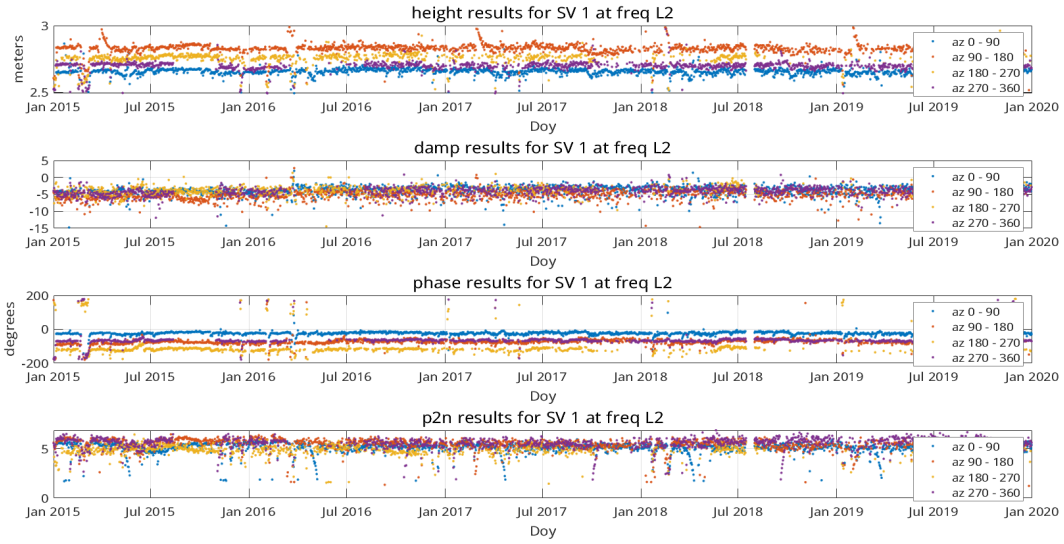


Figure 30: Estimated parameters throughout the processing period.

Figure 30 presents sporadic outliers that will cause biases to the soil moisture estimation. These outliers are related to environmental conditions and high vegetation which could not be corrected.

Figure 31 compares the vegetation-corrected VSM with the non-corrected version. Notice that the difference between both results is more highlightable in Spring, the vegetation growing season, and Summer, when vegetation is maximum. Due to the location of the GNSS stations used in the data, the Spring and Summer seasons corresponds to those of the North hemisphere.

Figure 31 presents spikes in the VSM estimation, for example between March and May in 2016. These events correspond to outliers and coincide with days in which available satellite measurements drops since less good quality measurements are present. Spikes are caused by environmental conditions such as snow or heavy rainfall, either from current day or previous days through snow or water accumulation on the ground. Figure 32 shows precipitation quantification for the MFLE area during 2016, notice the high activity coinciding with the period between March and May.

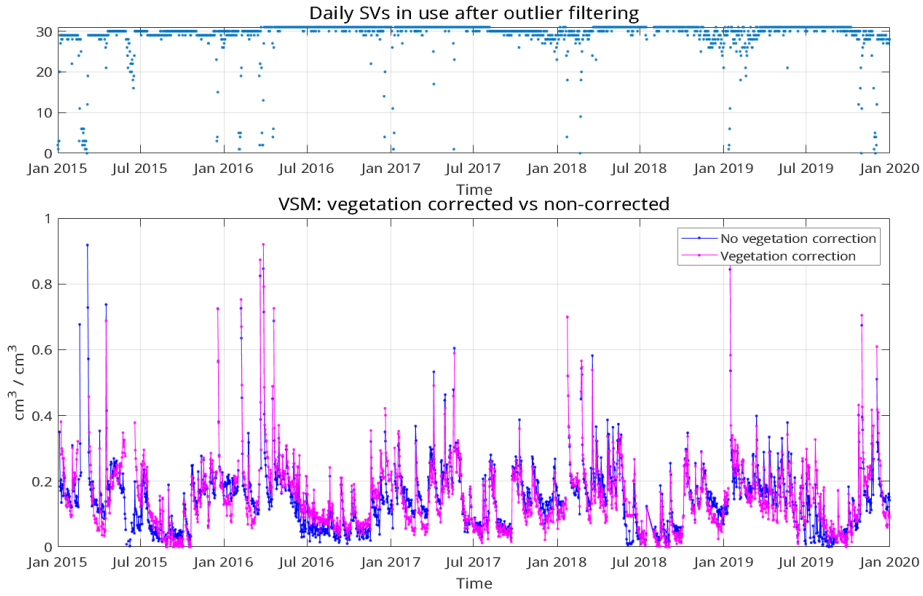


Figure 31: Comparison of vegetation-corrected VSM vs non-corrected. Available daily measurements (top) and VSM (bottom).

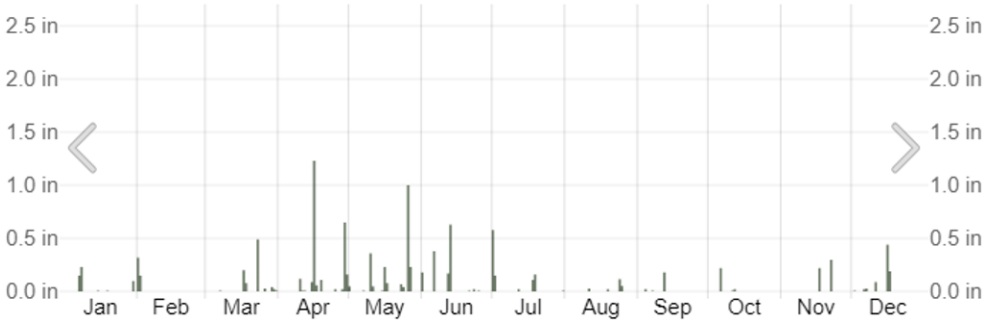


Figure 32: Precipitation quantification during 2016 for MFLE geographical area.

If outliers come from a particular satellite, the weighting average described in 7.2.4 could help to downweight it. However, in the case of environmental conditions all satellites are affected, which can be seen in figure 33 when comparing the weighted average vs non-weighted. As we can see, there is not much difference between both.

Therefore, outliers on figure 33 are removed by configuring several design parameters in the GNSS-IR processing chain, as described in next subsection 9.1.1.

9.1.1 Design Configuration

The soil moisture estimation from phase measurements following model in (6) is influenced by multiple design parameters. This customizable configuration allows to discard outliers and have a cleaner final result. These parameters are:

- k-sigma region and LSP peak-to-noise ratio described in 7.2.1.

- Number of averaged daily satellite measurements M described in section 7.2.4.
- Magnitude of $\Delta\phi_{veg}$ described in vegetation correction algorithm in 7.2.5.

In this subsection, an analysis is carried out for the impact that design configuration parameters have on VSM estimation at MFLE station. Different configuration parameters are defined in table 7. A default configuration is defined in row 1, which leads to the soil moisture results shown in figure 34 compared against UNAVCO reference VSM. In the figure, estimated VSM consists on the vegetation corrected (orange) and non-corrected (yellow), and is compared against UNAVCO reference data (blue).

Notice that, although data spans from 2015 - 2019, the plots compare estimated VSM with UNAVCO reference until 2018. This is done to make a fair comparison, since UNAVCO data is present until that date.

Experiment	Figure	Min. $p2n$	k	Min. M	Max. $\lvert\Delta\phi_{veg}\rvert$
Default	34	1	3	10	20
A	35 (left)	5	3	10	20
B	35 (right)	1	0.5	10	20
C	36	1	0.5	20	20
D	37	1	0.5	20	12
E	38	2	0.5	20	12

Table 7: Design configuration parameters

k-sigma region and LSP peak-to-noise ratio described in 7.2.1:
Increasing the peak-to-noise $p2n$ ratio ensures that we keep with cleaner measurements

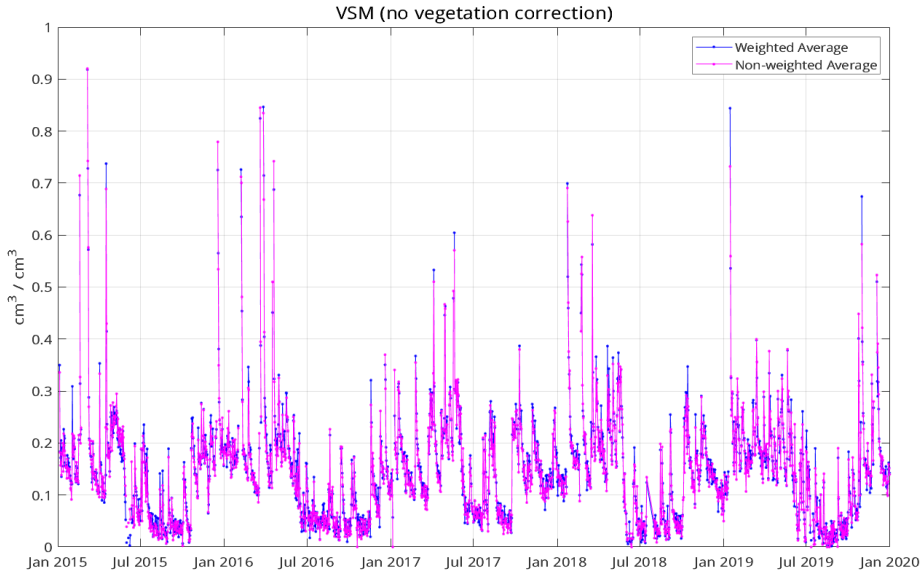


Figure 33: Comparison of VSM with weighted vs non-weighted average.

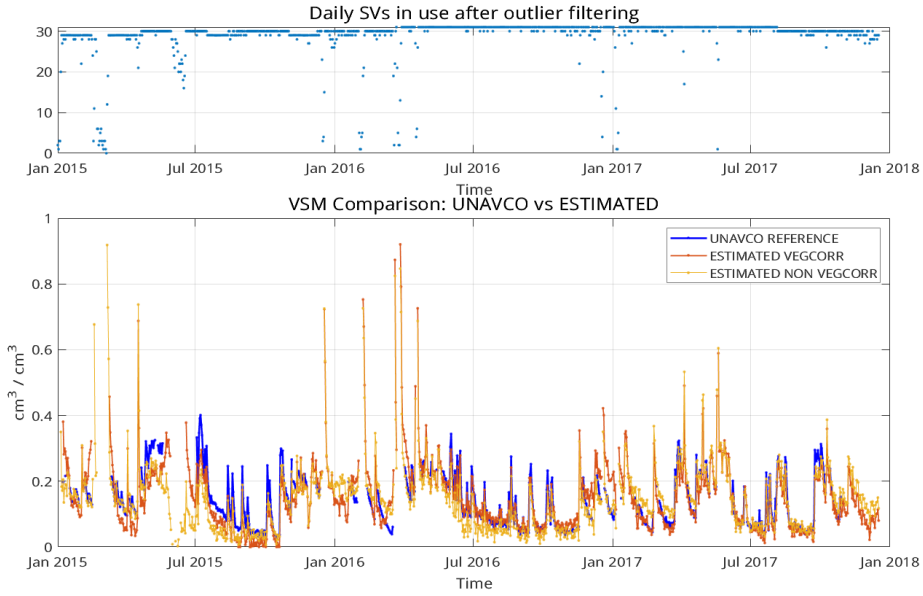


Figure 34: Default experiment with parameters defined in row 1 of table 7. Available daily measurements (top) and VSM (bottom).

where the SNR oscillations are well defined. However, doing this will also result in losing valuable information in days which measurements are noisier but still trustful, i.e. weak LSP peak but still distinguishable. Recall that vegetation presence can also reduce the peak, thus reduce the peak-to-noise. Therefore filtering too restrictively based on the peak-to-noise ratio can discard data excessively in periods such as medium to high vegetation seasons.

On the other hand, reducing the k-sigma region in the $\hat{H}_{LSP}$ filtering process helps to discard days where the height estimation varies highly. It has been shown that these days correspond to outliers such as snow or heavy rainfall, which can lead to misleading soil moisture results.

Figure 35 shows the VSM result when setting higher $p2n = 5$ (right, experiment **A**) and $k = 1$ (left, experiment **B**). Notice that these parameters have a direct impact on the number of available daily measurements M, since the measurements that do not meet the $p2n$ and k-sigma region will be discarded. From figure 35, the experiment **A** seems more restrictive than **B** since it reduces more the available measurements that meet the desired configuration: experiment **A** has an average daily M between 15-20, while experiment **B** has 25-30. In spite of this, a *minimum* number of daily measurements is guaranteed, since table 7 specifies the minimum daily M such that days with less measurements will be discarded. $M = 20$ for both experiments.

Number of averaged daily measurements M as described in section 7.2.4:
Increasing the minimum number of daily measurements ensures that only days with sufficient data will be considered. This helps to avoid being affected by outliers when M is low. In previous experiments, $p2n$ and k varied and M was set to 10. In this

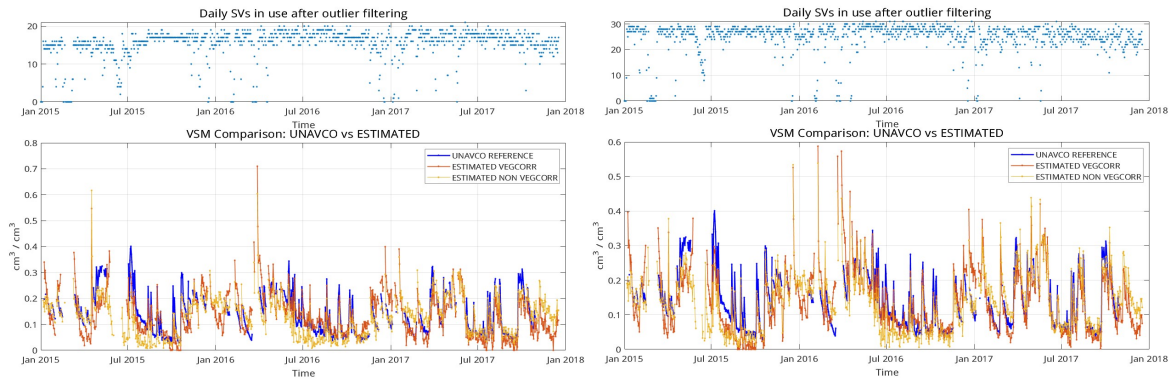


Figure 35: Experiment **A** (left) and **B** (right) defined in table 7. Available daily measurements (top) and VSM (bottom).

experiment, k was kept to 1 as in experiment **B**, but M is increased to 20, such that any day with less than 20 available satellite measurements is discarded.

Figure 36 shows the results for this experiment in table 7, but no significant improvements are observed when compared with experiment **B**. Since there are very few days whose available measurements are below 20, only few spikes are reduced in magnitude.

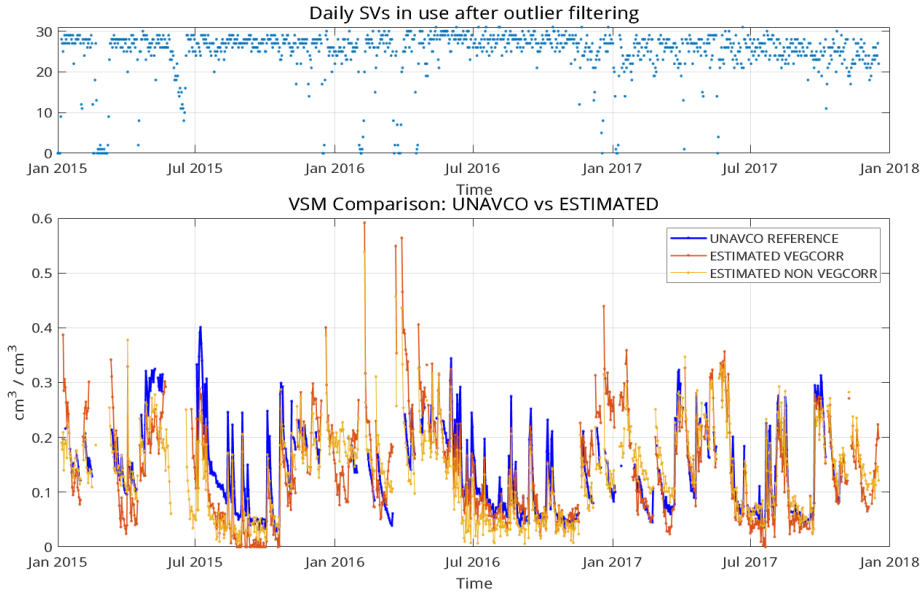


Figure 36: Experiment **C** defined in table 7. Available daily measurements (top) and VSM (bottom).

Magnitude of $\Delta\phi_{veg}$ described in vegetation correction algorithm in 7.2.5:
A very high correction means that the measurement is heavily affected by vegetation. In such case we could prefer to discard the phase estimation rather than correcting it,

since the correction itself might be inaccurate based on the mixture of soil moisture and vegetation in the amplitude and consequently in the LSP peak. Recall that the amplitude is affected by both vegetation and soil moisture as described in 4.4, and that the vegetation correction algorithm depends on P_{LSP}. Furthermore, it makes sense to discard estimations corresponding to high vegetation scenarios since the vegetation correction algorithm is capable of fairly correcting up to vegetation water contents of 1 kg/m^2 as described in 7.2.5.

Figure 37 shows the experiment **D** case in table 7. This reduces the acceptable maximum magnitude to correct for vegetation errors, otherwise discards the estimation since it is too affected by vegetation. In this case, same configuration as experiment **C** is applied, but maximum acceptable magnitude for vegetation correction is decreased to $|\Delta\phi_{veg}| = 13$ degrees. This reduction is possible since the vegetation surrounding the GNSS antenna is low. If vegetation were medium or high, it would be better to correct for higher magnitudes. This is aligned with findings in [14], since the experiments **A - D** show that selecting a right design parameter configuration can be sufficient for a low-vegetation site, while the vegetation correction algorithm can be applied in medium-vegetated scenarios.

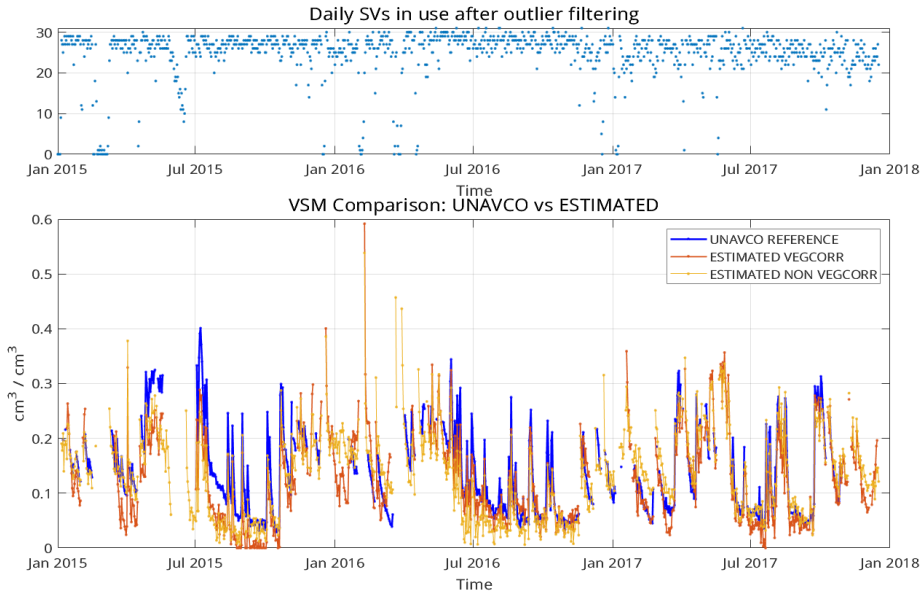


Figure 37: Experiment **D** defined in table 7. Available daily measurements (top) and VSM (bottom).

Final Parameters

After evaluating the effects of design parameters on experiments **A - D**, the configuration can be made stricter to reduce the remaining outlier spikes around January 2016. This can be done by reducing increasing the $p2n$ to 2 as shown in the last experiment in table 7. Results for this experiment **E** are presented in figure 38. The correlation of the VSM estimation with UNAVCO reference is also shown, resulting in $R = 0.87$ for

the vegetation-corrected case, and $R = 0.84$ for the non-corrected.

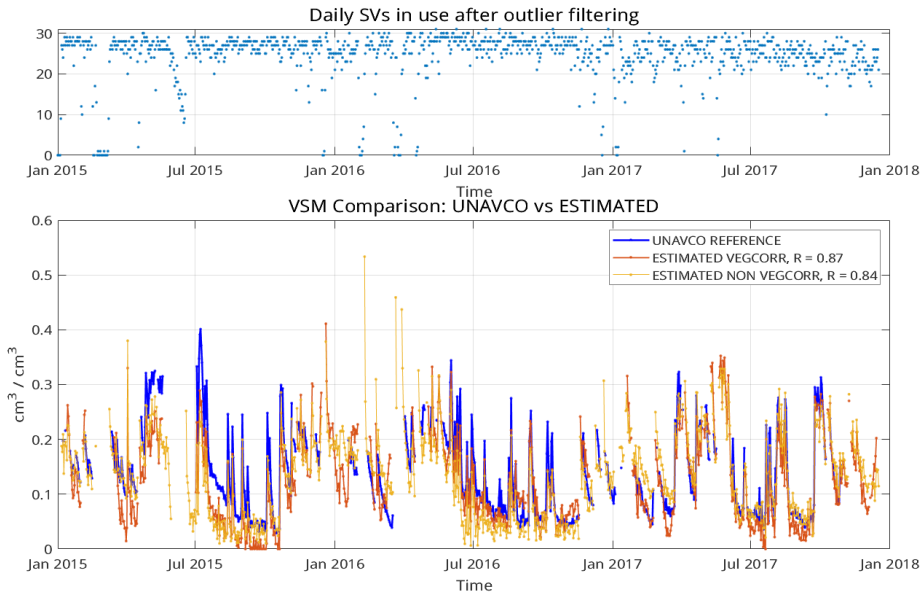


Figure 38: Experiment **E** defined in table 7. Available daily measurements (top) and VSM (bottom).

This result, although valid for outlier removal in MFLE station, is too strict and might not be generalized for P267 which has a different vegetation cycle. Therefore, it is important to relax the design constraints in medium-to-high vegetated locations.

For this reason, it is better to evaluate locations case by case and derive the best parameters accordingly. Table 8 shows the preferred design parameters for MFLE and P267 stations. Notice that the parameters for MFLE correspond to the experiment **E** in table 7, and parameters for P267 are more conservative than those in MFLE. This is because P267 is a medium-vegetated location while MFLE is low-vegetated.

Figure	Min. $p2n$	k	Min. M	Max. $\lvert\Delta\phi_{veg}\rvert$
MFLE	2	0.5	20	12
P267	0.5	1.5	10	20

Table 8: Design configuration parameters

Figure 39 shows the VSM for MFLE and P267 for vegetation-corrected and non-corrected. Notice how in P267 the VSM estimations follows a trend that, same as the available satellites M, is driven by the yearly seasons. These effects are more visible in medium vegetated (P267) than in low vegetation scenarios (MFLE). Minimums for VSM and available daily satellite measurements are reached in the high vegetation season. Furthermore, notice how the parameters in table 8 for MFLE considerably reduce the outliers.

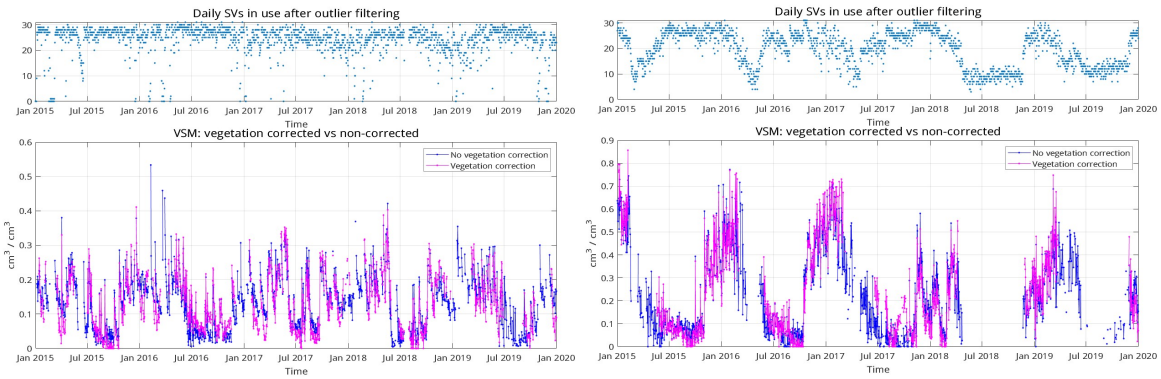


Figure 39: VSM in MFLE (left) and P267 (right) with design parameters in table 8.

9.2 GNSS-IR aided Multispectral VSM estimation

This subsection presents the results for the GNSS-IR aided multispectral model described in section 8. The same GNSS data processed to obtain the GNSS-IR standalone results in previous subsection is used, and the same design parameters in table 8 are applied. The GNSS-IR outputs are used to aid the algorithm and fit the coefficients in model (38).

Multispectral data corresponds to Landsat-8 imagery described in section 6, and also spans the period 2015 - 2019. In section 8.1 the processing chain for the multispectral VSM model was described, from data collection and processing in Google Earth Engine, to monthly averaging and model fitting in MATLAB. Below the results for MFLE and P267 stations are presented.

Figure 40 shows the results for MFLE and P267 fitting, respectively, and table 9 shows the coefficients estimated for each station. Blank periods for GNSS-IR VSM in figure 40 correspond to periods in which the available daily measurements were *lower* than the minimum M (table 8). As previously explained, the number of available measurements oscillates and reaches minimum in high vegetation seasons, since the measurements are more corrupted by vegetation in that period. This was also shown in figure 39. After calculating the coefficients to use with Landsat measurements, the model in (38) can be applied to *extend* the VSM solution beyond the GNSS-IR coverage area (approximately 100x100 meters).

Station	Type	α_1	α_2	α_3
MFLE	Low Vegetation	-0.0192	0.7293	0.0646
P267	Medium Vegetation	-0.0078	0.5569	0.0877

Table 9: Coefficients found for model in (38)

Recall that station P267 is surrounded by agricultural lands with similar terrain characteristics. Therefore, it is of interest to extend the VSM estimation for larger areas. In particular, an area of 10x10 km surrounding P267 station is analyzed and the estimated coefficients in table 9 are applied to estimate the VSM with Landsat-8 inputs.

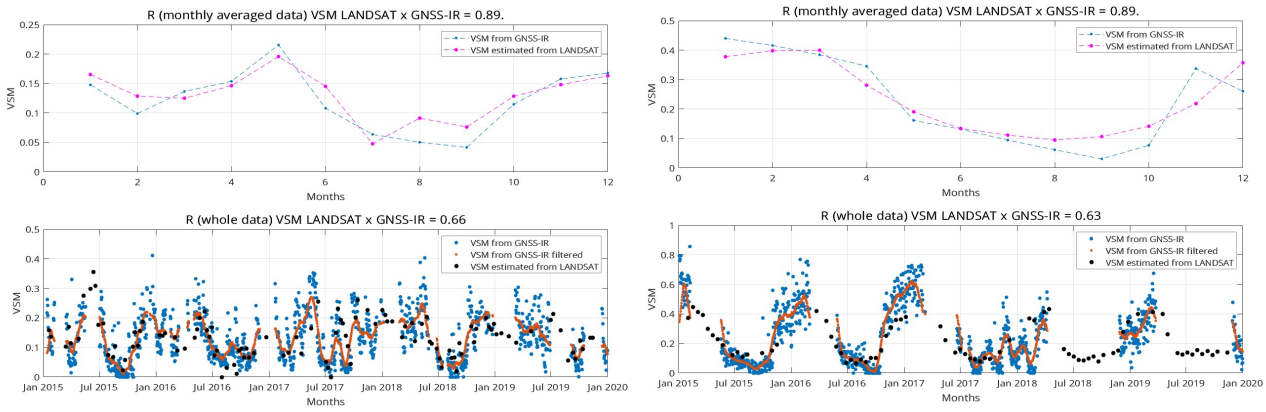


Figure 40: Monthly averaged fitting (top) and whole period data (bottom). Area corresponds to approximately 100x100 meter for both MFLE (right) and P267 (left).

After this, SMAP satellite data is retrieved to be used as reference for comparison in a monthly basis. SMAP has a resolution of 9km, thus the reason for choosing a grid of approximately 10x10 km to make a fair comparison.

Figure 41 presents the result for the estimated VSM over the extended area from Landsat-8 measurements using the model in equation (38) and compared with SMAP. Since SMAP provides an average over the entire area, the presented curves also correspond to the averaged VSM for the Landsat-8 estimates, although Landsat allows for finer resolution since pixels correspond to 30m. At this stage, since the algorithm utilizes the already estimated coefficients α_i, the calculations are done in Google Earth Engine. The blue and cyan curves are the Landsat-8 estimated and SMAP reference VSM, respectively, for a 10x10 km area centered in the GNSS station. On the other hand, dark-green and light-green curves are the Landsat-8 estimated and GNSS-IR reference VSM, respectively, for a 100x100 m area also centered in the GNSS station.

The obtained correlation for the 100x100 m area is $R = 0.87$, which matches the value obtained in figure 39. The correlation between SMAP and Landsat-8 VSM model for the 10x10 km is $R = 0.917$, which confirms the applicability of the model for extending local estimations of VSM to larger areas of similar terrain type.

With this result, since Landsat-8 resolution is 30 meters, different crops can be identified and individual lands can be processed to obtain VSM estimations, which is not possible with SMAP. On figure 41 the 10x10 km result is also shown as a color map image over the tested area, with color codes ranging from red to blue denoting low to highly soil moisture, respectively.

9.2.1 Terrain-Specific Processing

In 8.2 the terrain-specific behavior was explained and a variation of the model was described in equation (42) to account for different terrain types based on the vegetation quantification with NDVI.

The results in figure 41 do not have this feature enabled, and therefore the plain P267 coefficients from table 9 are applied to the entire terrain, regardless if certain

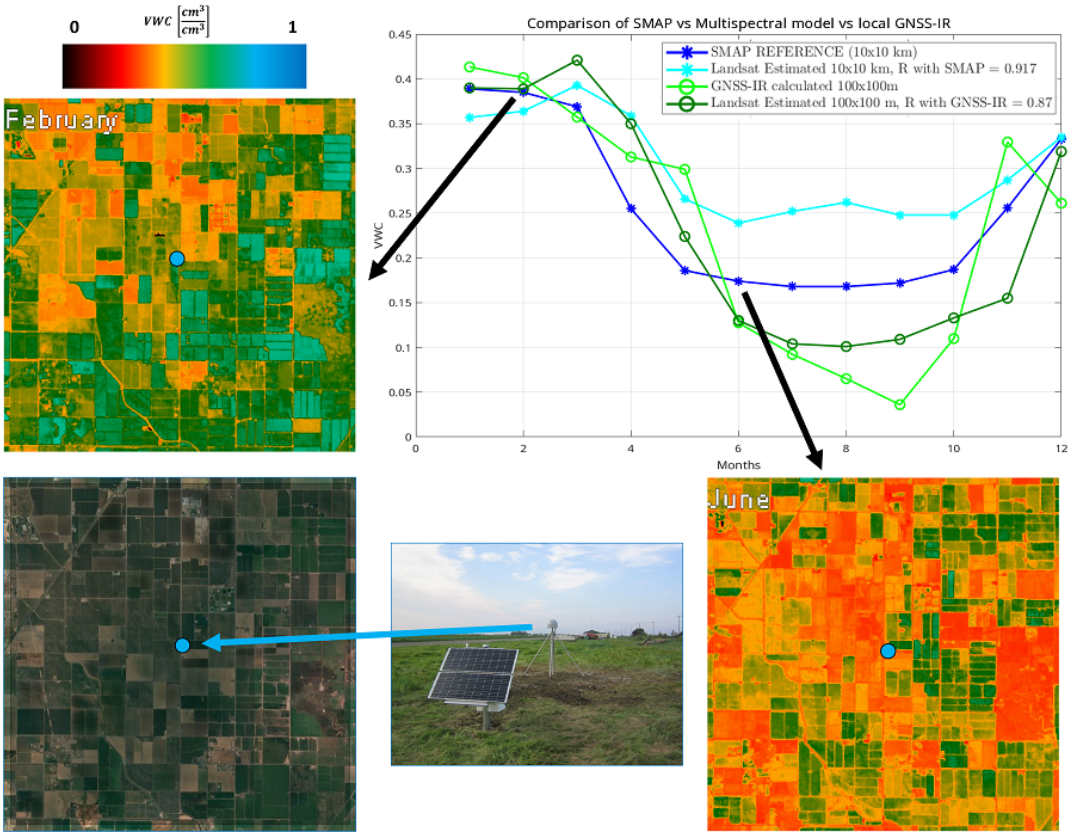


Figure 41: Result of multispectral model application on 10x10 km area around P267. Correlations with SMAP and GNSS-IR for 10x10km and 100x100m respectively.

regions are more or less vegetated. Figure 42 shows the NDVI for February (left) and June (right). We can see that, based on the considerable NDVI difference among different crops areas, it would worth it to vary the coefficients α accordingly.

In this experiment the understanding of "terrain type" is linked to the extent of vegetation. Variation described in equation (42) are applied. The NDVI threshold is set to $\gamma = 0.3$ and $\gamma = 0.5$ and only two alternatives are used: low and medium vegetation. Thus, $\gamma = \gamma_{lm}$. The MFLE coefficients are named $\alpha_{\text{LOW VEG}}$ and P267 coefficients are names $\alpha_{\text{MED VEG}}$. Coefficients $\alpha_{\text{HIGH VEG}}$ is not utilized.

Figure 43 presents the results for the two γ thresholds used. Notice that "No γ" case corresponds to same curves shown in figure 41. Results indicate that combining α coefficients estimated for different terrain types can represent more accurately the soil moisture temporal behavior. The obtained correlations for VSM are $R = 0.917$ for "No γ" (i.e. use the estimated α for the entire area around P267), $R = 0.921$ for $\gamma = 0.3$ and $R = 0.943$ for $\gamma = 0.5$. This avoids subregions with low NDVI to overestimate VSM if the same P267 coefficients are used.

Figure 44 shows the color map images with high resolution VSM estimations for the 10x10km area. Notice how low NDVI regions present lower VSM values when γ is applied. This helps to avoid overestimating soil moisture especially in dry periods, like Summer months. For this reason, after applying the NDVI threshold, the VSM

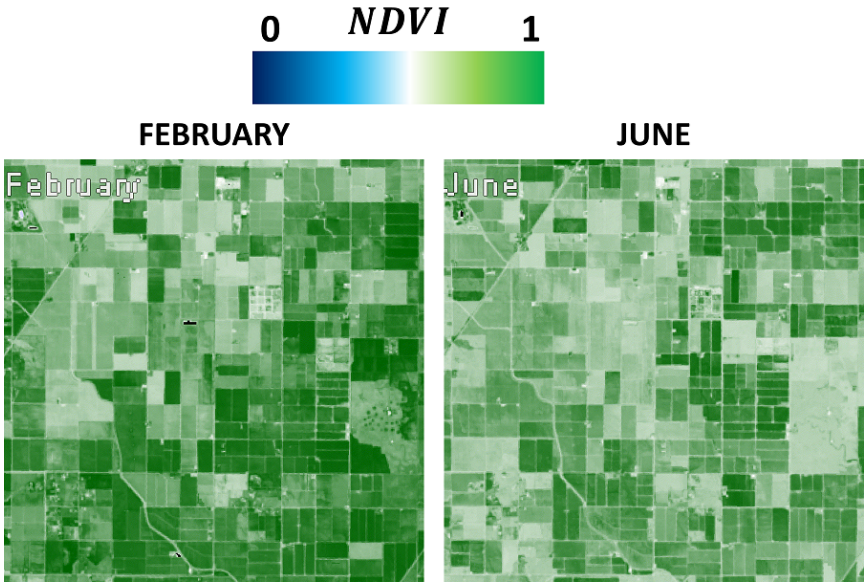


Figure 42: NDVI for 10x10 km region in February (left) and June (right).

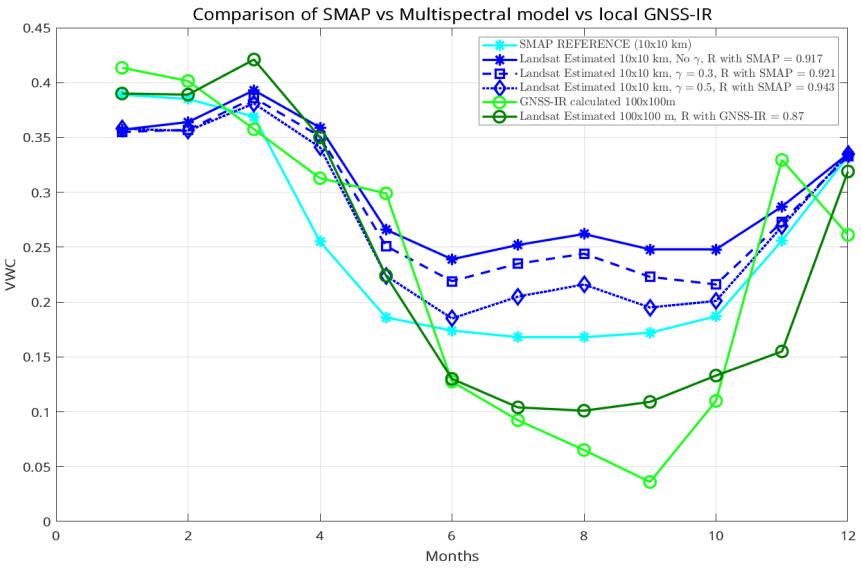


Figure 43: Correlations between 10x10km VSM estimations from Landsat-8 and SMAP for different terrain types based on NDVI thresholds.

variation over the tested area is more noticeable in June than in March. See yearly average curve on figure 29 (bottom in blue), where averaged NDVI is above 0.5 in February and below in June. Furthermore, certain land subregions can have higher or lower NDVI values depending on their specific terrain type.

The results confirm that the model can be used for VSM calculation over large areas and coefficient combinations help improving the estimation while accounting for different terrain-types.

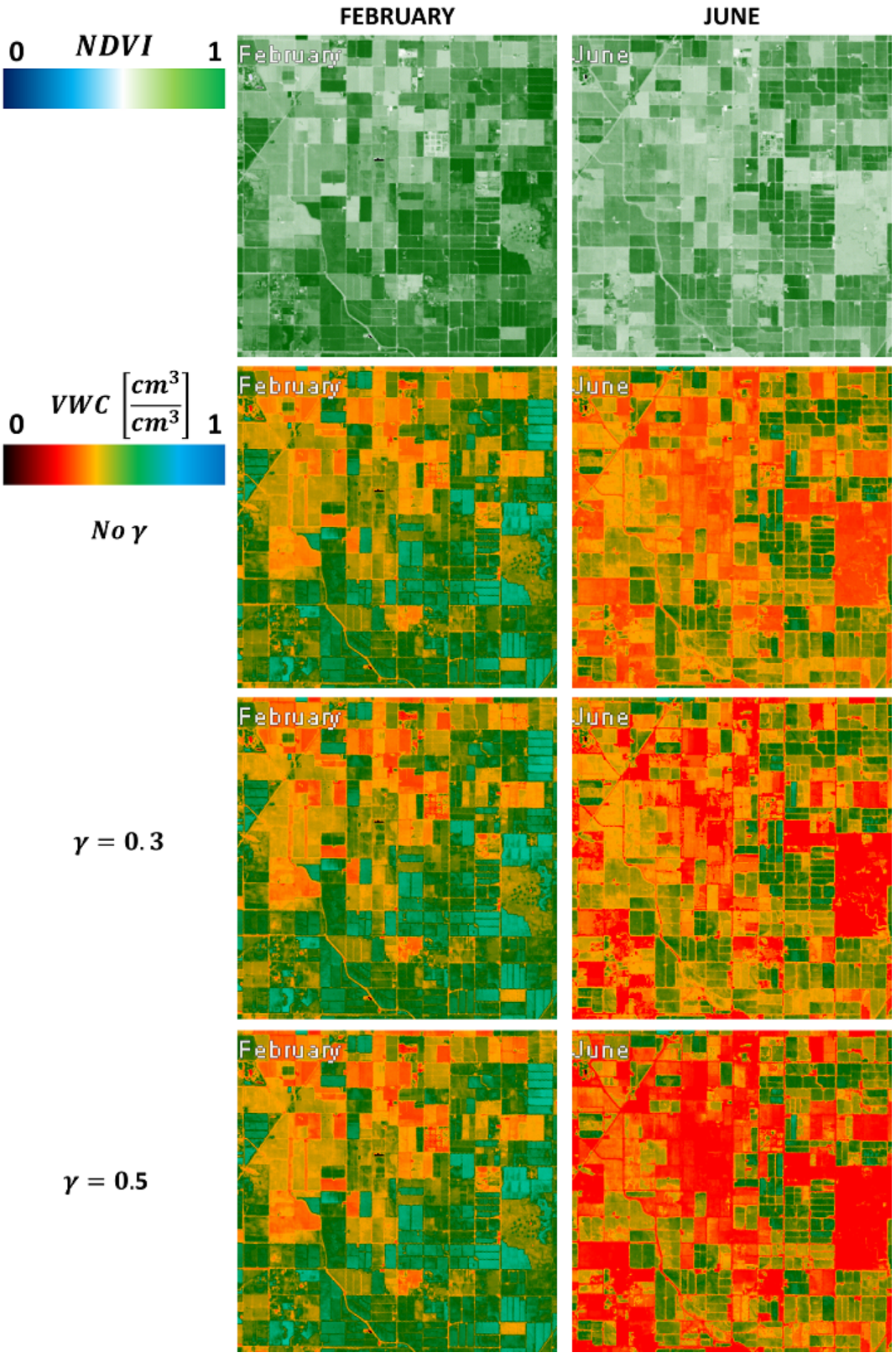


Figure 44: Color map images of VSM estimation for varying NDVI thresholds differentiating between low and medium vegetation subregions. VSM variations based on vegetation are more noticeable in Summer (June) than in Winter (February).

10 Conclusion

In this thesis, the the estimation of soil moisture from two different passive remote sensing technologies was reviewed. Microwave and optical spectra have different behaviors before moistened soils, and the algorithms to derive soil moisture content from these technologies must take the physical processes into account.

The effects of wet soils in GNSS-IR was reviewed both in literature and practice with the development of a GNSS-IR processing chain to estimate VSM. Although all three SNR parameters described in 4.4 are correlated with soil moisture, it is the phase which is more responsive and commonly used as an indicator for VSM due to its high correlation $R = 0.95$ [1].

In GNSS-IR, it has been shown that multiple design factors allow to reduce the outliers and refine the final VSM estimate considering their implications as described in section 9.1.1. It has been demonstrated that k-sigma region and LSP $p2n$ are good candidates to customize for removing outliers, typically from snow or rainfall, in both low and medium vegetated scenarios. Higher daily satellite measurements M produces more robust estimates, but its effect is negligible when this number is already high, such as in low vegetation areas. Therefore, M has a significant effect in medium-vegetated scenarios but not in low-vegetation areas. Special attention must be put in M in medium vegetated scenarios since daily availability of measurements depends on the environmental conditions of each location, and this was demonstrated for P267 station, where availability tends to decrease in highly vegetated seasons such as Spring or Summer.

It has been shown that corrections for vegetation effects in GNSS-IR VSM are needed, but these can be inaccurate if the vegetation is high, leading to excessive correction factors due to the correction algorithm limitation described in 7.2.5. Therefore, the magnitude of the vegetation correction $\Delta\phi_{veg}$ can be used as a monitoring metric to decide up to how much vegetation cover to correct confidently, and discard value requiring higher corrections.

In multispectral imagery, the response of dry and wet soil is correlated with metrics such as LST, VSDI, NDVI, and the difference between Red and Blue bands. LST and REBL are negatively correlated with VSM, while VSDI and NDVI are positively correlated. However, the correlation with VSDI and NDVI is more pronounced in medium-to-high vegetated scenarios, such as P267. The relation of these metrics with VSM makes them good candidates to be used in a soil moisture retrieval algorithm from multispectral data. It has been shown that terrain type plays a role in the relation between VSM and these quantities.

Finally, a GNSS-IR aided model using multispectral data as inputs was derived and fitted with GNSS-IR results. This helps to extend the VSM estimation coverage to , which is the regions bigger than $\sim$ 50-meter radius areas, which is the approximate coverage from GNSS-IR standalone for average 2-meter tall antennas. The model coefficients were calculated and a use-case in a 10x10 km agricultural region was presented.

The large area result was compared against data from Soil Moisture Active Passive (SMAP) mission, which is commonly used as a large-area reference for soil moisture

estimates. Obtained correlation with SMAP was $R = 0.917$ when the same coefficient were used over the entire 10x10km area. Correlation can increase when NDVI thresholds are applied to determine low and medium vegetated soils, applying MFLE coefficients for low vegetation and P267 coefficients for medium vegetation (table 9). After varying the coefficients within the 10x10 km area, correlations increased to $R = 0.921$ and $R = 0.943$ with NDVI thresholds of 0.3 and 0.5 respectively. Determining different subregions and applying coefficients accordingly is of particular interest in large areas such as agricultural lands with wide variety of crops.

This approach allows a sustainable way to directly monitor soil moisture content with multispectral data and requires little to no infrastructure deployment, apart from the installation of the GNSS antenna and receiver, in case that information from a specific site or soil type must be collected. GNSS-IR results and coefficient databases can be calculated with public data, retrieved from sources such as UNAVCO GNSS Network and Landsat-8 datasets. These results demonstrate that with the appropriate parameters, it is possible to extend local accurate soil moisture estimates to larger areas using multispectral satellite data, enabling global and accurate VSM monitoring while keeping meter-level resolution.

10.1 Future Work

This thesis combines microwave and optical remote sensing. Currently, only GPS L2 and Landsat-8 are used. The GNSS-IR processing can also take data from L5 band, or even other GNSS systems. On the other hand, using Sentinel-2C could provide higher resolution since visible and NIR bands cover 10 meter per pixel. Sentinel-2c, however, does not have a thermal band to compute LST, but other spectral indexes can be used as inputs to the algorithm such as OPTRAM [32], where LST is not necessary. Different inputs to the algorithm require re-estimate the coefficients α_i accordingly with (40).

The extrapolation of the local solution to bigger areas has been done by applying the model to the entire area or combine different coefficients based on soil characteristics. In this implementation, the terrain type was limited to the the extent of vegetation cover, but it can be expanded to other terrain types with bigger coefficient databases.

References

[1] Chew, C.; Small, Eric; Larson, Kristine; Zavorotny, Valery. Effects of Near-Surface Soil Moisture on GPS SNR Data: Development of a Retrieval Algorithm for Soil Moisture. Geoscience and Remote Sensing, IEEE Transactions on. 52. 537-543. 2014. DOI: 10.1109/TGRS.2013.2242332.

[2] Chew, C.; Small, E.; Larson, K. An algorithm for soil moisture estimation using GPS-interferometric reflectometry for bare and vegetated soil. GPS Solutions. 20. 2015. DOI: 10.1007/s10291-015-0462-4.

[3] Nievinski, F.G., Larson, K.M. Forward modeling of GPS multipath for near-surface reflectometry and positioning applications. GPS Solut 18, 309–322. 2014. Available: https://doi.org/10.1007/s10291-013-0331-y

[4] Entezari, M.; Esmaeily, A.; Niazmardi, S. "Estimation of Soil Moisture and Earth's Surface Temperature Using Landsat-8 Satellite Data". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. Available: https://doi.org/10.5194/isprs-archives-XLII-4-W18-327-2019

[5] Sun, H.; Liu, H.; Ma, Y.; Xia, Q. "Optical Remote Sensing Indexes of Soil Moisture: Evaluation and Improvement Based on Aircraft Experiment Observations". 2021. Available:https://doi.org/10.3390/rs13224638

[6] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, M. Neil, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, J. Van Zyl, "The soil moisture active passive (SMAP) mission," Proceedings of the IEEE, vol. 98, no. 5, pp. 704-716, May 2010. DOI: 10.1109/JPROC.2010.2043918.

[7] Hanson, B. Field Estimation of Soil Water Content: A Practical Guide to Methods, Instrumentation and Sensor Technology. Vadose Zone Journal, v. 8, n. 3, p. 628-759, 2009. DOI: 10.2136/vzj2008.0171.

[8] Babaeian, E.; Sadeghi, M.; Jones, S. B.; Montzka, C.; Vereecken, H.; Tuller, M.. Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. Reviews of Geophysics, v. 57, n. 2, p. 530-616, mar. 2019. DOI: 10.1029/2018RG000618.

[9] Edokossi, K.; Calabia, A.; Jin, S.; Molina, I. GNSS Reflectometry and Remote Sensing of Soil Moisture: A Review of Measurement Techniques, Methods, and Applications . Remote Sens. 2020, 12, 614. Available: https://doi.org/10.3390/rs12040614

[10] National Snow and Ice Data Centre. Soil Moisture Active Passive (SMAP). Available: https://nsidc.org/data/smap/data [Accessed: 11-Jul-2024].

[11] European Space Agency - Copernicus. Soil Moisture and Ocean Salinity (SMOS) Documentation. Available: https://documentation.dataspace.copernicus.eu/Data/ComplementaryData/SMOS.html [Accessed: 11-Jul-2024].

[12] Felipe Euriques, J.; Pereira Krueger, C.; Carrupt Machado, W.; Fernando Sapucci, L.; Geremia-Nievinski, F. Soil Moisture Estimation with GNSS Reflectometry: a Conceptual Review. Revista Brasileira de Cartografia. Vol. 73, n. 2, 2021. Available: https://doi.org/10.14393/rbcv73n2-55033

[13] Chew C.; Small E., Larson K. M. and Zavorotny V. U. Vegetation Sensing Using GPS-Interferometric Reflectometry: Theoretical Effects of Canopy Parameters on Signal-to-Noise Ratio Data. IEEE Transactions on Geoscience and Remote Sensing, Vol. 53, No. 5, May 2015. DOI: 10.1109/TGRS.2014.2364513

[14] Small E. E., Larson K. M., Chew C. C., Dong J. and Ochsner T. E. Validation of GPS-IR Soil Moisture Retrievals: Comparison of Different Algorithms to Remove Vegetation Effects. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015. DOI: 10.1109/JSTARS.2015.2504527

[15] Kaplan E., Hegarthy C. Understanding GPS: Principles and Applications. Second Edition, 2006.

[16] Navstar GPS Directorate. GPS ICD 200M. Interface Control Document (ICD), revision M. April 13, 2021.

[17] European Space Agency. GPS Signal Plan. Available: https://gssc.esa.int/navipedia/index.php/GPS_Signal_Plan [Accessed: 11-Jul-2024].

[18] Artiola J., Pepper I., Brusseau M. L., Environmental Monitoring and Characterization, 2004.

[19] Rodriguez-Alvarez, N.; Munoz-Martin, J.F.; Morris, M. Latest Advances in the Global Navigation Satellite System—Reflectometry (GNSS-R) Field. Remote Sens. 2023, 15, 2157. Available: https://doi.org/10.3390/rs15082157

[20] Wan, W., Larson, K.M., Small, E.E. et al. Using geodetic GPS receivers to measure vegetation water content. GPS Solut 19, 237–248. 2015. Available: https://doi.org/10.1007/s10291-014-0383-7

[21] Larson, K.M., Small E.E., Gutmann E., Bilich A., Braun J., Zavorotny V. Use of GPS receivers as a soil moisture network for water cycle studies. Geophys Res Lett 35:L24405. 2008. Available: https://doi.org/10.1029/2008GL036013

[22] Larson K.M., Gutmann E., Zavorotny V.U., Braun J.J., Williams M., Nievinski F.G. Can we measure snow depth with GPS receivers? Geophys Res Lett 36:L17502. 2009. Available: https://doi.org/10.1029/2009GL039430

[23] GPS Reflections Research Group. "PBO H2O Data Portal". Available: https://cires1.colorado.edu/portal/ [Accessed: 06-Jul-2024].

[24] Geremia-Nievinski, Felipe; Silva, Matheus; Boniface, Karen; Monico, Galera. GPS Diffractive Reflectometry: Footprint of a Coherent Radio Reflection Inferred From the Sensitivity Kernel of Multipath SNR. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 9. 1-8. 2016. DOI: 10.1109/JSTARS.2016.2579599.

[25] Roesler, C.; Larson, K.M. Software documentation of gnssrefl. ”Software tools for GNSS interferometric reflectometry (GNSS-IR)”. Available: https://gnssrefl.readthedocs.io/en/latest/index.html [Accessed: 10-Jun-2024].

[26] Roesler, C., Larson, K.M. Software tools for GNSS interferometric reflectometry (GNSS-IR). GPS Solut 22, 80. 2018. Available: https://doi.org/10.1007/s10291-018-0744-8

[27] Geremia-Nievinski, F. GNSS Inteferometric Reflectometry: Basic Theory. GNSS-IR Short Course - EarthScope Consortium, 2023.

[28] P. Axelrad, K. Larson, and B. Jones, “Use of the correct satellite repeat period to characterize and reduce site specific multipath errors,” in Proc. ION GNSS 18th Int. Tech. Meeting Satellite Div., 2005, pp. 2638–2648. Available: https://www.researchgate.net/publication/255606971

[29] Zavorotny V. U., Larson K. M., Braun J. J., Small E. E., Gutmann E. D. and Bilich A. L. A Physical Model for GPS Multipath Caused by Land Reflections: Toward Bare Soil Moisture Retrievals. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. Vol. 3 No. 1. 2010. DOI: 10.1109/JSTARS.2009.2033608

[30] U.S. Geological Survey. Landsat 8 (L8) Data Users Handbook. 2019 Environmental Monitoring and Characterization, 2004.

[31] Neinavaz E.; Skidmore, A. K.; Darvishzadeh R. "Effects of prediction accuracy of the proportion of vegetation cover on land surface emissivity and temperature using the NDVI threshold method", International Journal of Applied Earth Observation and Geoinformation, Volume 85, 2020. Available: https://doi.org/10.1016/j.jag.2019.101984

[32] Sadeghi, M.; Babaeian, E.; Tuller, M.; Jones, S.B. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. Remote Sens. Environ. 2017, 198, 52–68. Available: https://www.sciencedirect.com/science/article/abs/pii/S0034425717302493?via%3Dihub

[33] Yang Q.; Liu X.; Wu W. A Hyperspectral Bidirectional Reflectance Model for Land Surface. Sensors. 2020; 20(16):4456.

[34] Prandoni P., Vetterli M. Signal Processing for Communications. First Edition, 2008.

[35] VanderPlas JT. Understanding the Lomb Scargle Periodogram. 2017. Available: https://arxiv.org/pdf/1703.09824.pdf

[36] Eyer, L.; Bartholdi, P. Variable stars: Which Nyquist frequency? A&AS, Volume 135, Number 1. 1999. Available: https://doi.org/10.1051/aas:1999102

[37] Selesnick I. "Least Squares with Examples in Signal Processing". 2022. Available: https://eeweb.engineering.nyu.edu/iselesni/lecture_notes/least_squares/least_squares_SP.pdf

[38] EarthScope Consortium. "NOTA Network Monitoring". Available: https://www.unavco.org/instrumentation/networks/status/nota [Accessed: 11-Jul-2024].