

# **Application of physics-informed neural network approach in soil moisture retrieval using GNSS reflectometry**

Narin Gavilikilane

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE  
STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAMME IN EARTH AND SPACE SCIENCE  
YORK UNIVERSITY  
TORONTO, ONTARIO

December 2023

© Narin Gavilikilane, 2023

## Abstract

This research aims to obtain soil moisture from reflected GNSS signals using physics-informed neural networks (PINN). GNSS reflectometry (GNSS-R) signals can be considered as a new remote sensing source to study soil moisture. Despite the high sensitivity between GNSS reflected signal power and soil moisture, the model between measurements and parameters is difficult to solve mathematically due to the complexity of the electromagnetic relationships. Although Neural Network (NN) algorithms have been applied successfully in GNSS-R soil moisture retrieval, neural networks are trained without respecting any laws of physics. In this work, a new framework referred to as “physics-informed neural networks (PINN)” was used which adds governing physical relationships between data parameters to neural networks to generate more robust models, with less data.

The proposed research advances GNSS-R soil moisture estimations, exploiting Cyclone Global Navigation Satellite Systems (CYGNSS) satellite signals using PINN methodology. In PINN’s structure, reflected GPS signals from CYGNSS and land surface geophysical parameters are used as input features. Since reflected signal power variations are not only sensitive to changes in soil moisture, but also to changes in vegetation, surface roughness, soil texture, and elevation angle, the effects of land surface geophysical parameters involved in physical relationships are considered in the model. For reference data, soil moisture measurements of the International Soil Moisture Network (ISMN) were used in both training and validation.

The proposed PINN model generates daily soil moisture values with a root mean squared error (RMSE) of  $0.05 \text{ cm}^3 / \text{cm}^3$ , which is an improvement from  $0.0774 \text{ cm}^3 / \text{cm}^3$  for the underlying NN model due to adding physical models. Four different soil dielectric constant models (Dobson, Hallikainen, Mironov, and Wong models) have been used to investigate the impact of soil dielectric constant models as part of physical relations. The RMSE distinction and correlation coefficient difference of the best model (Hallikainen) and worst model (Mironov) is 0.02 and 0.13, respectively demonstrating PINN sensitivity to different soil dielectric constant models. Consequently, the soil dielectric constant model selection influences overall PINN results. Thus, calibration of soil dielectric models is necessary for GNSS-R soil moisture retrieval in the future.

## **Dedication**

To: Grandma, Mom, Dad, and Narges

## Acknowledgements

I could not have undertaken my PhD journey without generous support from my mentors, colleagues, friends, and family. I would like to thank the following people, without whom I would not have been able to complete this research.

First and foremost, I would like to express my gratitude and appreciation to my supervisor, Dr. Sunil Bisnath who mentored me in the development of this research work. I am incredibly thankful to him for his priceless guidance during my graduate studies and supporting me through the challenges of graduate school especially during COVID time. Words cannot express my gratitude to him for his patience and kindness during my academic years.

I am also indebted to Dr. Regina Lee, without whom I would not have been able to make it through my PhD degree. I am incredibly thankful for all the meetings and conversations with her that inspired me through my hardest challenges and worst moments. She has always been one of my greatest supporters who helped me to think about the bigger picture, which has kept me from falling too deeply into numerous rabbit holes.

I also wish to thank Dr. Franz Newland for his corridor consultations and insightful feedback. Next, I would like to thank my committee for their reviews and comments on this dissertation.

I would also like to thank Dr. Junchan Lee for his assistance and insightful comments. His contribution has been essential, specifically in the first stages of my PhD degree.

From the bottom of my heart, I would like to say big thank you to my amazing friends Sudha, Sogand, and Ding for their energy, understanding, and help throughout my project. It truly has been a very fun time in the lab and I'm going to miss our conversations.

Final words go to my lifelong supporters, my parents Adel and Homa to have always believed in the little girl with crazy dreams. I am the luckiest person for having them in my life.

# Table of Contents

Abstract .....	ii
Dedication .....	iv
Acknowledgements .....	v
Table of Contents .....	vii
List of Tables .....	x
List of Figures.....	xi
List of Acronyms.....	xiv
List of Symbols.....	xvi
1. INTRODUCTION .....	1
1.1 Soil moisture measurement techniques .....	1
1.1.1 In situ soil moisture measurements .....	3
1.1.2 Remote sensing techniques .....	9
1.2 Motivation for using GNSS-R in soil moisture retrieval .....	12
1.3 Machine learning in GNSS-R soil moisture studies.....	16
1.4 Research objectives.....	18
1.5 Research Novelty and significance .....	21
1.6 Dissertation outline .....	22
2. PRINCIPLES OF GNSS REFLECTOMETRY .....	24
2.1 GNSS-R theory .....	24
2.1.1 Global Navigation Satellites Systems .....	25
2.1.2 GNSS-R .....	28
2.2 GNSS-R observables.....	34
2.3. Scattering geometry .....	43
2.3.1 Specular point calculation .....	44

2.3.2 Resolution.....	49
3.    PHYSICAL MODELS IN GNSS-R SOIL MOISTURE RETRIEVAL .....	52
3.1    Theoretical relationships in soil moisture retrieval using GNSS-R .....	52
3.2    Soil moisture – dielectric constant models .....	56
3.2.1 Empirical models .....	58
3.2.2 Semi-empirical .....	63
4.    PHYSICS-INFORMED NEURAL NETWORK STRUCTURE .....	70
4.1 PINN applications .....	70
4.2 PINN design for GNSS-R.....	71
5. DATASETS .....	81
5.1 CYGNSS data.....	81
5.2 Soil moisture data .....	84
5.3 Ancillary data .....	85
5.3.1 Vegetation.....	87
5.3.2 Land cover and surface roughness.....	90
5.3.3 Soil texture.....	91
5.4 Bulk data processing.....	92
6.    GNSS-R SOIL MOISTURE RETRIEVAL BASED ON PINN APPROACH .....	98
6.1 PINN model performance analysis.....	98
6.1.1 PINN results for different NN structures.....	101
6.1.2 PINN sensitivity to different land covers .....	104
6.2 Discussion.....	108
7.    COMPARISON OF DIFFERENT SOIL DIELECTRIC CONSTANT MODELS FOR PINN SOIL MOISTURE ESTIMATES .....	117
7.1 Impact of different soil dielectric models on PINN performance.....	119



7.2 Investigation of different soil dielectric models' performance for different soil texture classes.....	125
7.3 Discussion.....	136
8. CONCLUSIONS AND FUTURE WORK.....	140
8.1 Conclusions .....	140
8.2 Future work .....	146
REFERENCES.....	151

## List of Tables

Table 1.1 Comparison between SMAP and SMOS mission .....	11
Table 2.1 Comparison of different GNSS missions .....	29
Table 3.1. Experimental data used in Hallikainen model development .....	59
Table 3.2. Experimental data used in Wang model development .....	63
Table 4.1 software libraries designed for physics-informed machine learning .....	71
Table 5.1 CYGNSS Level 1 data used .....	84
Table 5.2 Data measured at SCAN stations .....	86
Table 6.1 ISMN stations used in PINN .....	101
Table 6.2 PINN performance metric values for different NN and PINN designs .....	103
Table 6.3 PINN performance metric values for different train test ratios .....	104
Table 6.4 PINN performance for different land cover for two-layer PINN .....	107
Table 6.5 PINN performance for different land cover for three-layer PINN .....	107
Table 7.1 Comparison of soil dielectric models used in PINN algorithm .....	118
Table 7.2 ISMN stations used in comparison of soil dielectric models' performance in PINN algorithm .....	120
Table 7.3 Comparison of four soil dielectric models' performance used in PINN .....	121
Table 7.4 Classes defined for soil dielectric constant model's assessment .....	126
Table 7.5 ISMN stations used in comparison of soil dielectric models' performance in PINN algorithm .....	127
Table 7.6 RMSE results of different soil dielectric constant models for four different soil texture classes .....	129
Table 7.7 ubRMSE results of different soil dielectric constant models for four different soil texture classes .....	130
Table 7.8. Pearson correlation coefficient results of different soil dielectric constant models for four different soil texture classes .....	131

## List of Figures

Figure 1.1 Physics-informed machine learning comparison with physics-based models and data-driven machine learning .....	20
Figure 1.2 NN structure vs PINN structure .....	23
Figure 2.1 Russian GLONASS satellite .....	27
Figure 2.2 European Galileo satellite .....	27
Figure 2.3 China Beidou satellite .....	28
Figure 2.4 GPS operational satellites .....	30
Figure 2.5 GNSS-R technique as a bistatic system .....	32
Figure 2.6 GNSS-R as a multi-static system .....	32
Figure 2.7 Comparison of GNSS-IR and GNSS-R .....	33
Figure 2.8 Different platforms used for GNSS-R soil moisture retrieval .....	35
Figure 2.9 Delay Doppler Map calculation in a SDR .....	39
Figure 2.10 Delay waveform and DDMs for a smooth surface .....	41
Figure 2.11 Delay waveform and DDMs over a rough surface .....	41
Figure 2.12 Relationship between delay waveform and GNSS-R applications .....	43
Figure 2.13 Comparison of coherent DDM versus incoherent DDM .....	44
Figure 2.14 The geometry of GNSS reflected signals .....	45
Figure 2.15 Specular scattering from smooth surfaces versus diffuse scattering from rough surfaces .....	51
Figure 3.1 Soil as a mixture of sand grain, bund and free water and oil .....	60
Figure 4.1 Physics-informed neural network loss function calculations .....	74
Figure 4.2 Physics-informed Neural Network (PINN) structure for CYGNSS soil moisture estimation .....	75
Figure 4.3 Different learning rate occurrences .....	77
Figure 4.4 Three different states of epoch, batch, and iteration .....	79
Figure 5.1 Soil moisture sensors at a SCAN site .....	87
Figure 5.2 SCAN station sensors and data collectors .....	88
Figure 5.3 USDA Soil texture triangle .....	93
Figure 5.4 USDA Soil texture triangle visual assess .....	93

Figure 5.5 Diagram of components of the Earth Engine Code Editor from .....	96
Figure 5.6 Different data sources used in PINN .....	97
Figure 6.1 Soil moisture loss function changes over 1000 epochs .....	102
Figure 6.2 Reflectivity loss function changes over 1000 epochs .....	102
Figure 6.3 Correlation scatterplots of soil moisture retrievals from PINN with two different NN structures .....	105
Figure 6.4 Correlation scatterplots of soil moisture retrievals from PINN for two different train/ test ratios .....	106
Figure 6.5 Sensitivity of PINN Pearson correlation coefficient metric to different land covers .....	109
Figure 6.6 Sensitivity of PINN RMSE metric to different land covers .....	110
Figure 6.7 Correlation plots of PINN soil moisture retrievals and ISMN soil moisture for different land covers .....	111
Figure 6.8 Learning curve plots for a three-layer NN and PINN comparison .....	116
Figure 7.1 Correlation plots of PINN soil moisture retrievals for different soil dielectric models .....	122
Figure 7.2 RMSE and ubRMSE comparison of four soil dielectric constant models used in PINN algorithm .....	123
Figure 7.3 Pearson's correlation coefficient comparison of four soil dielectric constant models used in PINN algorithm .....	123
Figure 7.4 Dielectric constant calculated from soil moisture retrievals from PINN relationship with soil moisture reference ISMN data for 2020 .....	124
Figure 7.5 Dielectric constant calculated from soil moisture retrievals from PINN relationship with soil moisture reference ISMN data for 2021 .....	124
Figure 7.6 Four soil texture classes defined to investigate soil dielectric constant model's performance .....	126
Figure 7.7 RMSE comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes .....	129
Figure 7.8 ubRMSE comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes .....	130

Figure 7.9 Pearson's correlation coefficients comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes .....	131
Figure 7.10 Correlation plots of PINN soil moisture retrievals with Dobson soil dielectric constant model for four different soil texture classes .....	132
Figure 7.11 Correlation plots of PINN soil moisture retrievals with Hallikainen soil dielectric constant model for four different soil texture classes .....	133
Figure 7.11 Correlation plots of PINN soil moisture retrievals with Wang soil dielectric constant model for four different soil texture classes .....	134
Figure 7.11 Correlation plots of PINN soil moisture retrievals with Mironov soil dielectric constant model for four different soil texture classes .....	135

## List of Acronyms

AI	Artificial intelligence
API	Application Programming Interface
AWS	Amazon Web Services
Beidou	Chinese Navigation Satellite System
CRN	Climate Reference Network
CYGNSS	Cyclone Global Navigation Satellite System
DC	Dielectric constant
DDM	Delay Doppler Maps
EO	Earth observation
EROS	Earth Resources Observation and Science
ESA	European Space Agency
GALILEO	European Union Navigation Satellite System
GEE	Google Earth Engine
GLONASS	Russian Navigation Satellite System
GNSS	Global Navigation Satellite System
GNSS-R	Reflectometry using Global Navigation Satellite System signals
GPS	Global Positioning System, United States Navigation Satellite System
IGBP	International geosphere-biosphere program land cover classification data
ISMN	International Soil Moisture Network
JAXA	Japan Aerospace Exploration Agency
LEO	Low Earth orbit

LF	Loss factor
LHCP	Left-hand circularly polarized
LUT	Look-up-table
ML	Machine learning
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean squared error
NavIC	Indian Regional Navigation Satellite System
NDVI	Normalized difference vegetation index
NIR	Near-infrared
NN	Neural network
PDE	Partial differential equations
PINN	Physics-informed neural network
PODAAC	Physical Oceanography Distributed Active Archive Center
QZSS	Japanese Quasi-Zenith Satellite System
RHCP	Right-hand circularly polarized
RMSE	Root mean squared error
SCAN	Soil Climate Analysis Network
SMAP	Soil Moisture Active Passive, NASA environmental monitoring satellite that measures soil moisture across the planet
SMOS	Soil Moisture and Ocean Salinity, ESA mission making global observations of soil moisture over land and salinity over oceans
SNR	Signal-to-noise ratio
ubRMSE	Unbiased root mean squared error
USDA	United States Department of Agriculture

## List of Symbols

$A$	Multipath amplitude
$\phi$	Multipath phase
$M_{stones}$	Mass of the stone in soil sample
$M_{fines}$	Mass of fine soil fraction in soil sample
$\rho_{water}$	Water density in $\text{g/cm}^3$
$\rho_{soil}$	Soil bulk density in $\text{g/cm}^3$
$G_r$	Antenna gain of receiver
$G_t$	Antenna gain of transmitter
$L$	Left-hand circularly polarized (LHCP) downward looking antenna
$P_{RL}^{coh}$	Coherent received power
$P_{RL}^{inc}$	Bistatic received power including diffuse scattering over the surface
$P_t$	Peak power of the transmitted GNSS signals
$R$	Right-hand circularly polarized (RHCP) GNSS transmit antenna
$r_{st}$	Distance between specular point to transmitter
$r_{sr}$	Distance between specular point to receiver
$\Gamma_{RL}(\theta_i)$	Surface reflectivity at incidence angle of $\theta_i$
$\Re_{RL}(\theta_i)$	Fresnel reflection coefficient at incidence angle of $\theta_i$
$\theta_i$	Incidence angle
$\lambda$	Wavelength
$\sigma_{RL}$	Bi-static radar cross section (BRCS) in $m^2$
$\gamma$	Transmissivity parameter related to the wave attenuation due to vegetation canopy
$h$	h-parameter related the root-mean-height surface roughness



$\tau$	Vegetation optical depth
$VWC$	Vegetation water content
$b$	Land cover-based proportionality value, b-parameter
$NDVI$	Normalized difference vegetation index
$\Re_{HH}(\theta_i)$	Linear polarized horizontal Fresnel reflection coefficient
$\Re_{VV}(\theta_i)$	Linear polarized vertical Fresnel reflection coefficient
$\varepsilon_r$	Dielectric constant
$m_v$	Volumetric moisture content
$C$	Clay percentage
$S$	Sand percentage
$\varepsilon'_m$	Moist soil dielectric constant
$\varepsilon''_m$	Moist loss factor
$\varepsilon'_s$	Composite dielectric constant of the soil mineral contents
$\rho_b$	Bulk density in grams per cubic centimetre
$\rho_s$	Specific gravity of the solid soil particles
$\beta$	Empirically determined soil-type-dependent constant
$WP$	Wilting point of soil
$W_t$	Transition moisture
$W_c$	Volumetric water content
$P$	Porosity of the dry soil
$\varepsilon_a$	Dielectric constant of air
$\varepsilon_w$	Dielectric constant of water
$\varepsilon_r$	Dielectric constant of rock

$\varepsilon_i$	Dielectric constant of ice
$\varepsilon_x$	Dielectric constant of the initially absorbed water
$\varepsilon'_{m0}$	Dry-soil dielectric constant
$\varepsilon'_{fw}$	Dielectric constant of free water
$\varepsilon''_{fw}$	Loss factor of free water
$\varepsilon_0$	Dielectric constant of free space equal to $8.854 \times 10^{-12} F/m$
$\tau_w$	Relaxation time for free water
$f$	Frequency in hertz
$\varepsilon_{\omega 0}$	Low-frequency limit of dielectric constant for water equal to 4.9
$\sigma_{eff}$	Effective conductivity
$n$	Refractive index
$k$	Normalized attenuation coefficient
$P_r^t$	Transmitted RHCP power
$R_{ts}$	Distance between the transmitter and the specular reflection point
$R_{sr}$	Distance between the receiver and the specular point

# 1. INTRODUCTION

Soil moisture determination is essential in the environmental sciences. Global Navigation Satellite Systems (GNSS) reflected signal strength has shown correlations to moisture content of soil. Earth's surface soil moisture can be monitored using reflected signals of the Global Navigation Satellite System (GNSS) as a cost-effective method. In this chapter, the motivation for using GNSS reflectometry (GNSS-R) for soil moisture is discussed. Finally, the objective and novelty of the physics-informed neural network (PINN) method as the proposed method to model the relationship between reflected signal power and soil moisture is explained.

## 1.1 Soil moisture measurement techniques

The knowledge of soil moisture impacts different environmental industries and research fields such as hydrology, meteorology, climatology, and agriculture. The water content in soil is a key variable in the understanding of the terrestrial environment and energy fluxes between the land and atmosphere. Soil water content controls the extent of plants' sunlight exploitation; therefore, knowledge of soil moisture can be effective in agriculture and forestry. Soil moisture is the representative of water source and is useful in water resource management. Soil moisture plays an important role in the estimation of the different components of the water and energy transfer between the land and atmosphere, thus understanding the hydrological cycles (Brocca et al. 2017). The distribution of soil moisture controls the division of incoming radiant energy into heat fluxes via evaporation and transpiration (Space Studies Board and National Research Council 2000). Since

water in soil can be returned directly to the atmosphere via evaporation from land surfaces, and transpiration from plants, this returning moisture can itself affect the local weather. Therefore, soil moisture is a principal parameter in the development of weather patterns and desired input to weather prediction models (Njoku and Entekhabi 1996). Critical environmental issues such as floods and landslides can also be modelled with soil moisture.

Soil moisture is measured with two main methods: 1) in situ and 2) remote sensing methods. In situ methods such as gravimetric instrumentation and electromagnetic probes are popular for soil scientists and agronomists (Chew 2015). Remote sensing methods include microwave radars as active remote sensing and radiometers as passive remote sensing which are more popular for hydrologists and climatologists. The difference between radars and radiometers is that radar systems compare transmitted and received energy of signal; however, radiometers receive natural thermal emission of microwave radiation coming from soil passively (Njoku and Entekhabi 1996). In addition to traditional remote sensing tools, available signals of Global Navigation Satellite Systems (GNSS) can be used as a new remote sensing technique to survey the Earth's surface properties such as ocean height (Ruffini et al. 2004, Rius et al. 2012), ocean wind speed (Komjathy et al. 2004; Guan et al. 2018), ice depth (Fabra et al. 2011, Cervellera 2013), soil moisture content (Larson et al. 2008a; Larson et al. 2008b; Camps et al. 2016; Pierdicca et al. 2021) and vegetation (Chew 2015; Wu et al. 2012) measurements. This research focuses on the application of reflected GNSS signals for soil moisture assessment. In the following subsections different methods of soil moisture measurements are explained.

### 1.1.1 In situ soil moisture measurements

In-situ measurements are classified into direct and indirect methods. Direct or oven drying methods include gravimetric and volumetric soil moisture methods for soil moisture determination. Automated techniques for soil moisture estimation are indirect soil-water dielectric techniques such as time-domain and frequency-domain reflectometry. In the direct method, also known as a destructive method, soil moisture is estimated by drilling holes into the ground, disturbing the soil root zone and eventually affecting infiltration and drainage behaviour. In contrast, indirect methods estimate soil moisture by developing a functional relationship between the chemical and physical properties of soil moisture and soil matrix (Dirksen 1999). A soil matrix is a mass of soil consisting of solid particles and voids filled with water and air.

The gravimetric method for soil moisture estimation has been counted as the most reliable and robust method (Schmugge et al. 1980). Gravimetric soil moisture ( $m_g$ ) is defined as the ratio of the mass of the water ( $m_{water}$ ) in the soil sample to the mass of the dry soil ( $m_{soil}$ ) expressed in (kg/kg) or percentage. Gravimetric soil moisture can be calculated as follows (Dirksen 1999),

$$m_g = \frac{m_{water}}{m_{soil}} = \frac{m_{wet} - m_{dry}}{m_{dry}} \quad (1.1)$$

where  $m_{water}$  is the mass of the water in the soil sample, expressed as the difference between the weight of the soil sample before drying ( $m_{wet}$ ) and after drying ( $m_{dry}$ ).  $m_{soil}$  and  $m_{dry}$  are both the mass of the dried soil.

Volumetric soil moisture ( $m_v$ ) is more relevant than gravimetric soil moisture as it considers the soil's bulk density and porosity (Singh 2023). Volumetric soil moisture ( $m_v$ ) is the ratio of the water volume ( $V_{water}$ ) to the volume of the soil ( $V_{soil}$ ), which is expressed in both percentage and ( $m^3/m^3$ ). Mathematically,  $m_v$  is defined as (Singh 2023)

$$m_v = \frac{V_{water}}{V_{soil}} \quad (1.2)$$

Equation (1.2) can be extended to produce a relationship between volumetric soil moisture ( $m_v$ ) and gravimetric soil moisture ( $m_g$ ) as follows

$$m_v = \frac{m_{water}/\rho_{water}}{m_{soil}/\rho_{soil}} = m_g \times \frac{\rho_{soil}}{\rho_{water}} \quad (1.3)$$

where  $\rho_{soil}$  is the soil bulk density and  $\rho_{water}$  is the water density both expressed in ( $g/cm^3$ ). Since the density of water is about  $g/cm^3$ , Equation (1.3) can be reduced to

$$m_v \approx m_g \times \rho_{soil} \quad (1.4)$$

If the soil sample contains a substantial amount of stone or gravel, the estimated soil moisture content can be affected, as the stones that are present in the sample contribute directly to the mass measurement of the sample without contributing equally towards the soil porosity. To avoid the effect of stones, Klute (1988) proposed the following expression.

$$m_v = \frac{m_{water}/\rho_{water}}{m_{soil}/\rho_{soil}} = m_g \times \left( \frac{\rho_{soil}}{\rho_{water}} \right) \times \left( 1 + \frac{M_{stones}}{M_{fines}} \right) \quad (1.5)$$

where  $M_{stones}$  and  $M_{fines}$ , respectively represent the masses of the stone and fine soil fraction in the soil sample.

The concept of using dielectric properties of soil to measure soil moisture was first studied by Canadian geophysicists Baker and Allmaras (Baker and Allmaras 1990). Soil-water dielectric techniques such as time-domain reflectometry (TDR) and frequency-domain reflectometry (FDR) estimate soil moisture content indirectly. Both TDR and FDR measure soil dielectric constant and use soil moisture dielectric constant models to convert dielectric constant measurements to soil moisture values. Dielectric constant is defined as the ratio of the electric permeability of the material to the electric permeability of free space. The dielectric constant of soil-water mixture is between 2 to 40, while the dielectric constant of dry soil is 2 and the dielectric constant of pure water is 80 (Ulaby et al. 1986). For different electromagnetic wave frequencies ranging from 1 to 1000 MHz, the gradient of dielectric constant for soil-water mixture from dry soil to pure water is explained in (Ulaby et al. 1986).

First, TDR determines the bulk dielectric constant ( $\epsilon_b$ ) of the soil from travel time. The travel time ( $t$ ) of the electromagnetic wave initiated along the waveguide, formed by the parallel rods of length ( $L$ ) with soil as the dielectric material in between is calculated. The propagation velocity ( $v$ ) can be calculated, by  $v = (2L/t)$  through travel time analysis. Since electromagnetic waves travel the rod twice (down and back),  $2L$  length is considered. After calculating propagation velocity ( $v$ ), the bulk dielectric constant ( $\epsilon_b$ ) is calculated as follows (Brocca 2007)

$$\varepsilon_b = \left(\frac{c}{v}\right)^2 = \left(\frac{ct}{2L}\right)^2 \quad (1.6)$$

where  $c$  is the speed of light in vacuum and equal to  $3 \times 10^8$  m/s and  $t$  is the travel time for the pulse to traverse the length of the embedded waveguide ( $2L$ ).

The TDR sensors send an electromagnetic wave pulse into an electrode inserted in the soil. Electrodes are conductors used to make electrical contact with a non-metallic part. As the permittivity changes between electrode and soil, a reflection occurs with a delay and intensity depending on soil properties. A complete overview of TDR principles, equipment and procedures can be found in (Jones et al. 2002).

After measurement of dielectric constant, soil moisture can be calculated. Different factors influence dielectric constant measurements, including soil texture, soil porosity and bulk density, measurement frequency, temperature, water status (bound or free) and dipole moments induced by mineral, water, and air shapes. Several empirical and semi-empirical soil dielectric mixing models have been developed to relate soil water content to soil bulk dielectric constant ( $\varepsilon_b$ ). These soil moisture dielectric models are explained in detail in Section 3.2.

Advantages of TDR over direct methods are as follows: i) The TDR probes measure the soil moisture with high accuracy (within 1% or 2%), ii) TDR does not contain radiation hazards, iii) minimum calibration is required, for example soil specific calibration is not required, iv) TDR have minimal soil disturbance, and have high spatial and temporal resolution (Topp et al. 1980).

However, the TDR measurements can be affected by high saline and high conductive soils. High soil salinity or high clay contents in soil cause attenuation of the reflected



pulses resulting in overestimating the value of the dielectric constant and hence the soil moisture (Chen et al. 2007; Jonnalagadda 2004).

The frequency-domain reflectometer (FDR) method uses radio frequencies and the electrical capacitance of a capacitor, formed by using the soil and embedded rods as a dielectric for determining the dielectric constant (Munoz-Carpena 2005) and thus calculation of the soil water content using soil dielectric models. FDR and capacitance probes consist of two or more capacitors (rods, plates, or rings) that are inserted into the soil. When the capacitor is connected to an oscillator to form an electric circuit, changes in soil water can be detected by changes in the circuit's operating frequency. FDR method uses an oscillator to propagate an electromagnetic signal through a metal tine or other wave guide. The difference between the output wave and the return wave frequency is measured to determine soil moisture. In FDR, the oscillator frequency is controlled within a certain range to determine the resonant frequency at which the amplitude is greatest, which is a measure of the soil water content. FDR method outperforms TDR in following cases: i) FDR is more flexible in probe design, ii) FDR works better in highly salinity conditions, iii) FDR is less expensive than TDR (Kuncham and Rao 2014). Still, FDR performs poorly due to variations of soil bulk electrical conductivity and spatial variability of vegetation water content. Also, installation can greatly affect the results. Although FDR works better in high salinity conditions compared to TDR, FDR has an accuracy between 2.5% to 4% which is lower than TDR accuracy (Rasheed et al. 2022).

Networks of in situ sensors using in situ measurement techniques include the Delaware Environmental Observing System (DEOS; <http://www.deos.udel.edu/data/>), Enviro-weather (formerly Michigan Automated Weather Network; <https://mawn.geo.msu.edu/>),

the NOAA Hydrometeorological Testbed (NOAA; <https://hmt.noaa.gov>), the Oklahoma Mesonet (<https://www.mesonet.org/>), the USDA-Natural Resources Conservation Service Soil Climate Analysis Network (SCAN; <https://www.wcc.nrcd.usda.gov/scan/>), the USDA-Natural Resources Conservation Service SNOwpack TELemetry (SNOTel; <https://www.wcc.nrcs.usda.gov/snow/>), the Soil Moisture Sensing Controller and Optimal Estimator (SoilScape; <https://www.soilscape.usc.edu/>), and the West Texas Mesonet (WTX Mesonet, <http://www.depts.ttu.edu/nwi/research/facilities/wtm/index.php>). In this research, the SCAN data are used as in-situ measurements. More details on ISMN SCAN data can be found in Section 5.2.

Indirect soil moisture sensors such as time domain reflectometry sensors or frequency domain reflectometry or capacitance sensors are capable of measuring soil moisture continuously using a data logger, thereby enabling the final user to save time. However, they suffer from the typically low number of in situ sensors that provide an incomplete picture of a large area.

While measurements of in situ network sensors (such as USDA's Soil Climate Analysis Network (SCAN) or NOAA's Climate Reference Network (CRN) in the continental United States) have potentially high soil moisture measurement accuracy, measurements are spatially very sparse. The main disadvantage of in situ measurement methods is that these measurements are time-consuming; users must go to the field to collect soil samples and place them in the oven for a long time.

Remote sensing methods can provide soil moisture estimates over larger areas and at a global scale without the need to be in the field. Remote sensing measurements are taken

remotely, while in-situ measurements require direct contact to soil. In the next subsection, remote sensing techniques for soil moisture studies are explained.

### **1.1.2 Remote sensing techniques**

Active and passive microwave remote sensing provides a unique capability to obtain observations of soil moisture at global and regional scales that help satisfy the science and application needs for hydrology (Njoku and Entekhabi 1996; Shmugge et al 2002). However, coarse spatial and temporal resolution, along with limited depth signal penetration can be obstacles in hydrological applications of remote sensing. Another issue with some sensors used for remote sensing of soil moisture is that their signals are affected by cloud cover. In addition, launching, designing, and maintenance of remote sensing satellites are expensive.

Remote sensing can estimate soil moisture continuously over large areas. In remote sensing, soil moisture estimations refer to the near-surface soil moisture representing the first 5 cm (or less) of the topsoil profile. In recent years, remote sensing techniques have improved and diversified their estimation tools for monitoring soil moisture along with other variables such as the Normalized Difference Vegetation Index (NDVI) and the land surface temperature (LST), (Fontanet et al. 2018). Different satellite missions have been developed for estimation of near-surface soil moisture, for example the Soil Moisture Active Passive (SMAP) satellite, the Advanced Scatterometer (ASCAT) remote sensing instrument on board the Meteorological Operational (METOP) satellite, the Advanced Microwave Scanning Radiometer 2 (AMSR2) instrument on board the Global Change

Observation Mission 1-Water (GCOM-W1) satellite, and the Soil Moisture and Ocean Salinity (SMOS) satellite.

Additionally, C- and X-band radiometers (with frequency range of 6 to 11 GHz and wavelength range of 3 to 5 cm) such as EOS Advanced Multichannel Scanning Radiometer (AMSR-E) and WindSat can provide soil moisture measurements with shallow sensing depth (~ 1 cm), and coarse resolution (>50 km). Nevertheless, satellite-based C- and X-band radiometers have low sensitivity to soil moisture for even small amounts of vegetation, leading to high soil moisture retrieval errors (Entekhabi et al. 2014). Therefore, L-band signals have been recently considered for soil moisture studies (Entekhabi et al. 2014). At L-band signals (with frequency range of 1 to 2 GHz and wavelength range of 15–30 cm), the atmosphere is less opaque, and soil penetration is higher in L-band observations (up to 5 cm). C-band observations are also more affected by vegetation attenuation (Njoku and Li 1999). L-band is one of the chief operating ranges used by various applications such as radars, Global Positioning System (GPS), radio, telecommunications, and aircraft surveillance.

The European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) satellite, launched in November 2009, is the first wide-swath L-band soil moisture mission (operating at 1.4 GHz frequency and ~21 cm wavelength), and retrieves soil moisture over a much higher range of vegetation conditions at a spatial resolution of ~40 km with a sensing depth of ~5 cm (Kerr et al. 2001). The SMOS single payload, the Microwave Imaging Radiometer by Aperture Synthesis (MIRAS), is the first 2D synthetic aperture interferometric passive radiometer in space which operates in L-band at 1.4 GHz frequency and 21 cm wavelength. The spatial resolution of SMOS ranges from 35 to

55 km, depending on the incident angle. The SMOS mission can retrieve near-surface soil moisture with a target accuracy of  $0.04 \text{ m}^3/\text{m}^3$  (Kerr et al., 2012).

The SMAP mission is another soil moisture retrieval satellite mission with a real aperture radiometer on board which operates in an L-band channel. NASA'S SMAP satellite was launched on January 31, 2015, including an L-band radar and an L-band radiometer that provide global radar backscatter and brightness temperature measurements, respectively, every 2–3 days. SMAP aimed to retrieve the soil moisture content of the upper ~5 cm of soil and freeze–thaw state (Entekhabi et al. 2014) and provide three soil moisture products (Panciera et al. 2013) including: i) low-resolution (~36 km) radiometer-only (L2\_SM\_P), ii) high-resolution (~3 km) radar-only (L2\_SM\_A), and iii) intermediate-resolution (~9 km) combined radar/radiometer (L2\_SM\_AP). Table (1.1) demonstrates a comparison between two L-band soil moisture missions (SMAP and SMOS). The SMAP radar transmitter failed on July 7th, 2015, while the SMAP radiometer continues to function.

Table 1.1 Comparison between SMAP and SMOS mission (Ray et al. 2017)

Satellite mission	Launch date	Frequency	Revisit coverage (days)	Spatial resolution	Data level
SMOS	2009	1.4	3	25 km	L3
SMAP	2015	1.41	3	3, 9 and 36 km	L2

SMAP uses the combination of passive radiometer and active radar. Active sensors have higher spatial resolution. Therefore, SMAP provides higher resolutions than SMOS. However, the backscatter commonly has a low temporal resolution (around one week)

and may be significantly affected by soil roughness and the soil-covering vegetation canopy, which complicates the active-only soil moisture retrieval (Portal et al. 2020).

## **1.2 Motivation for using GNSS-R in soil moisture retrieval**

Beside traditional remote sensing techniques, reflected GNSS signals have been used for a variety of remote sensing applications, for instance altimetry (Rius et al. 2012; Li et al. 2018; Tabibi et al. 2021), soil moisture retrieval (Larson et al. 2008a; Larson et al. 2008b; Camps et al. 2016; Chew et al. 2016; Chew et al. 2018), snow depth and ice thickness determination (Rius et al. 2017; Ghiasi et al. 2020), and wind speed estimation (Clarizia et al. 2014; Foti et al. 2015; Asgarimehr et al. 2022). GNSS-R (GNSS reflectometry) is a remote sensing technique that uses surface-reflected GNSS signals to infer information about the Earth's surface (Jin and Komjathy 2010). GNSS-R leverages free available and ubiquitous signals of GNSS and is considered as a passive and low-cost system (Yu et al. 2014). Furthermore, GNSS systems use L-band microwave signals suitable for soil moisture studies because of their all-weather detection and vegetation penetration ability (Jin et al. 2014). The L-band signal emitted by the Earth surface can be related to land parameters like surface soil moisture, roughness, or vegetation characteristics. In addition, in comparison to radiometry, GNSS reflected signals are not affected by thermal variations of background.

Hall and Cordey (1988) were the pioneers of GNSS-R application in ocean remote sensing. Afterward, Martin-Neira (1993) suggested an ocean altimetry concept known as Passive Reflectometry and Interferometry System (PARIS). Then, research on the ocean surface roughness retrieval (Garrison et al. 1998) and wind speed and direction

estimation from the reflected Global Positioning System (GPS) signals (Garrison et al. 2002) were investigated. Similar to sea surface wind speed retrieval, it was later shown that model fitting between the theoretical and measured delay waveforms can be used to determine the roughness of the ice surface (Rivas 2007).

The first space-based GNSS-R signals were observed for altimetry in the SIR-C and SAC-C missions (Lowe et al. 2002). From 2003 to 2011, the UK-Disaster Monitoring Constellation (UK-DMC) experimented GNSS reflectometry for sea state and wind speed estimates (Gleason et al. 2005). Such early experiments have indicated the feasibility of GNSS-R space missions. The UK-DMC experiment was used in optimizing new version of Surrey Satellite Technology Ltd (SSTL) GNSS-R payload instrument (Foti et al. 2015), which have been used in next GNSS-R missions such as UK TDS-1 and CYGNSS missions. Following the successful detection of reflected signals by UK-DMC, SSTL launched the UK TDS-1 on 8 July 2014. UK TDS-1 has been used for several remote sensing applications such as ice altimetry (Hu et al. 2017; Rius et al. 2017), ocean altimetry (Mashburn et al. 2018), ocean wind (Foti et al. 2015), and soil moisture (Camps et al. 2018). However, data collection for this mission was challenging for scientists because this mission was active two days out of eight (repeat time ~10–35). UK TDS-1 mission ended operations after December 2018 (Unwin et al. 2016; Yang et al. 2020).

On December 15, 2016, eight satellites of Cyclone Global Navigation Satellite System (CYGNSS) were launched on a near circular orbit with the inclination of  $35^\circ$  and orbit height of about 500 km (Ruf et al. 2012). This inclination improves the spatial and temporal sampling in critical latitudes for tropical cyclones which is between  $35^\circ\text{N}$  and  $35^\circ\text{S}$  (Rose et al. 2013). CYGNSS satellites have been developed by the University of

Michigan, Southwest Research Institute and NASA for forecasting hurricanes and the wind speed measurement (Ruf et al. 2016).

One of the main applications of GNSS reflectometry and the focus of this research is soil moisture retrieval. CYGNSS and similar GNSS-R instruments have several characteristics that make them functional for soil moisture (SM) remote sensing. First, GNSS signals are in L-band, which is optimal for SM remote sensing due to the increased ability to penetrate vegetation relative to shorter wavelengths (De Roo et al. 2001). Second, available constellation of GNSS satellites receivers shorten revisit time and increases temporal and spatial resolution. Third, transmitted GNSS satellites already exist for other purposes, reducing the cost of the complete sensing system. Sensitivity of GNSS reflectometry data to soil and vegetation parameters have been studied by several researchers. Initially, focusing more exclusively on soil moisture application, Zavorotny and Voronovich (2000) presented the capability of GPS scattering signals for soil moisture remote sensing. Research into the possibility of using GNSS reflections for sensing soil moisture was continued at the University of Colorado (Masters 2004) and NASA Langley Research Center (Katzberg et al. 2006) using measurements from aircraft flights and Earth based platforms. For satellite-based platforms, Gleason et al. (2005) experimented detectability of GNSS reflectometry signals for soil moisture studies in low Earth orbits using UK-DMC data collected from land surfaces. Gleason (2006) studied possibilities and challenges of GNSS reflectometry over land surfaces. Later, sensitivity of TDS1 space mission GNSS-R reflectivity data to SMOS satellite soil moisture was examined in (Chew et al. 2016; Camps et al. 2018).



Although CYGNSS satellites were designed for ocean applications, they provide reflections over land enabling use for soil moisture studies using GNSS-Reflectometry data. For instance, SMAP satellite data were compared with reflected signal power of CYGNSS to show the relationship between the soil moisture and reflected signals in (Chew and Small 2018; Al-Khaldi et al. 2019).

While high correlation between reflected signal power and water content of soil has been shown, variations in reflected GNSS signals are not only sensitive to changes in soil moisture, but also to changes in vegetation, surface roughness, soil texture and elevation angle (Chew and Small 2018). Most published research works have focused on the correlation of soil moisture with signal return power (Chew et al. 2016; Chew and Small 2018; Camps et al. 2018); however, the model between soil moisture and signal power has been poorly understood.

In cases that the model between input and output parameters is difficult to be solved mathematically due to the complexity of the electromagnetic relationship between measurements, Artificial Neural Networks (ANNs) have shown high performance in model learning (Goodfellow et al. 2016, Ali et al. 2015). For example, Prakash et al. (2018) applied recurrent neural networks for soil moisture prediction. The performance of the predictor was evaluated for three different hydrological monitoring network datasets. Convolutional neural networks were also used in remote sensing for soil moisture retrieval using brightness temperature data from the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) in (Rodriguez-Fernandez et al. 2015). Application of neural network as a soil moisture retrieval algorithm for different passive and active sensors; for example, datasets from spaceborne radiometers (AMSR-E/AMSR2), SAR

(Envisat/ASAR, Cosmo-SkyMed) and real aperture radar (MetOP ASCAT) have been studied in (Santi et al. 2016).

Focusing on GNSS-R technique, relevant research has been dedicated to soil moisture retrieval using GNSS-R with machine learning. Nevertheless, the problem of GNSS-R soil moisture retrieval using machine learning techniques is challenging due to the following reasons:

- i) Feature selection: As different surface data parameters affect signal power and the relationship between these land parameters and reflected signal power is complex, choosing the best surface parameters can affect results significantly.
- ii) Sufficient data availability: Machine learning techniques' performance is highly dependent on data availability and lack of sufficient data can be problematic.
- iii) Scale imbalances and differences: Different datasets are in different scales and ranges which will affect machine learning results convergence.

Understanding physics of the problem and inclusion of physical models in training process of machine learning process can help us to select features that are involved in physical relationships between parameters and use less data. Therefore, the problem of soil moisture retrieval using GNSS reflected signals still needs further study in terms of understanding both physics of the problem and consideration of all surface data parameters in solving for soil moisture.

### **1.3 Machine learning in GNSS-R soil moisture studies**

Machine learning (ML) methods are one of the most developed technologies in recent years. Due to the growth and emergence of big data in reflectometry, ML methods have been widely used in GNSS-R scientific modelling. For example, a Neural Network application for soil moisture retrieval from reflectivity data of CYGNSS satellites was surveyed (Eroglu et al. 2019). In (Eroglu et al. 2019), the potential of soil moisture estimation using Neural Networks with CYGNSS observations without the need for spatial gridding or temporal averaging has been presented. Surface ancillary data introduced to NN in (Eroglu et al. 2019) as input features include normalized difference vegetation index (NDVI), vegetation water content (VWC), elevation, slope, and h-parameter (roughness parameter). The proposed NN in (Eroglu et al. 2019) was able to predict soil moisture values from reflected signals of CYGNSS with an unbiased root mean squared error of  $0.0544 \text{ cm}^3/\text{cm}^3$  and Pearson correlation coefficient of 0.9009 for 2017 and 2018 datasets.

A comparison between two different NN approaches in GNSS-R soil moisture studies was addressed by (Roberts et al. 2022). The performance of a Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) in CYGNSS soil moisture estimation have been compared using delay-Doppler maps (DDMs) data. Different soil moisture data sources can be used for training and validation of NNs. As in (Eroglu et al. 2019), International Soil Moisture Network (ISMN) data were used for both training and testing, whereas researchers in (Roberts et al. 2022) used SMAP data as a reference.

The capability of NN inversion in GNSS-interferometric reflectometry (GNSS-IR) soil moisture retrieval also has been studied by (Shi et al. 2021). GNSS-IR technique uses one antenna to capture both direct and reflected signals simultaneously for signal to noise

ratio (SNR) data (Larson 2016). However, GNSS reflectometry (GNSS-R) benefits from a receiver using two zenith and nadir antennas to acquire waveform (Zavorotny et al. 2014). While GNSS-IR studies changes in the interference pattern of the direct and reflected signals recorded in SNR data, GNSS-R use delay-Doppler maps (DDMs) related to acquired scattered GNSS signals (Zavorotny et al. 2014). In (Shi et al. 2021), ten surface parameters such as latitude and longitude information, precipitation, temperature, land cover type, elevation, slope, slope direction, and shading were used in NN inversion model as well as GNSS-IR data to retrieve soil moisture. The NN model in (Shi et al. 2021) produced soil moisture values with RMSE of  $0.0346 \text{ cm}^3/\text{cm}^3$  and Pearson correlation coefficient of 0.8724.

Research in GNSS-R soil moisture retrieval using machine learning are not limited to neural networks. The XGBoost method has been utilized in soil moisture retrieval using GNSS-R data (Jia et al. 2019). In addition, a comparison of different machine learning methods has been accomplished in (Senyurek et al. 2020) analyzing three widely used machine learning approaches: artificial neural network (ANN), random forest (RF), and support vector machine (SVM). It has been shown that random forest had better results in estimating soil moisture trained and validated from the International Soil Moisture Networks (ISMN).

## **1.4 Research objectives**

Although machine learning techniques have shown promising results in modelling the relationship between reflected signals of GNSS and soil moisture, the performance of such ML methods is highly dependent on data. Machine learning's capability is reliant on

both quality and quantity of data significantly. Traditional machine learnings that are only trained using data are called “Data-driven machine learning”. To work with data-driven machine learning, the role of quality of data as well as quantity of data should be considered such that majority of workload includes data preparation. Prior to high-level accessibility to big data, physics-based numerical models have been used more than data-driven MLs. With full knowledge of physics and algebraic relationships, scientific problems can be modelled with physics-based numerical models. On the other hand, data-driven ML eliminates the need to understand complicated and unsolvable physics-based models. When we have a complete understanding of the physics of the problem, physics-based numerical models are used. However, with the availability of big data, MLs are able to find patterns, predict and classify without physics parametric modelling (Tartakovsky et al. 2020). Given that in case of small data ML fails to converge and lack of knowledge on physics is problematic in most scenarios, both physics-based and ML models can be insufficient in some ways and in real cases.

Development of a model to retrieve soil moisture from GNSS-R signals in an incomplete physics knowledge and imperfect observational data situation deserves more studies. In GNSS-R soil moisture retrieval, some physics knowledge is available; however, the electromagnetic relationships between soil moisture and GNSS reflected signal power are difficult to solve with a mathematical parametric model. Therefore, a model is required to introduce available physics principles to a NN model while still benefiting from data driven NN capabilities in solving complex problems.

The main goal of this dissertation is then to form a GNSS-R soil moisture retrieval model that can work in less data regime due to inclusion of available physical principles. Physical

relationships between geometrical and soil physical properties affecting reflected signal power can be added to a data driven NN through a methodology called “physics-informed neural network (PINN)”. Figure 1.1 compares a data-driven NN with a physics-informed neural network (PINN). In a data driven NN model, the focus is on big data and physics relationships are not added to the modelling and training process. However, in a physics-informed neural network, some physical principles are added to modelling to reduce the need for big data and also provide some physical knowledge in the process of training and modelling. The application of a physics-informed neural network (PINN) model is proposed in this research to provide additional physics knowledge into traditional NN models and reduce a need for availability of large amounts of data.

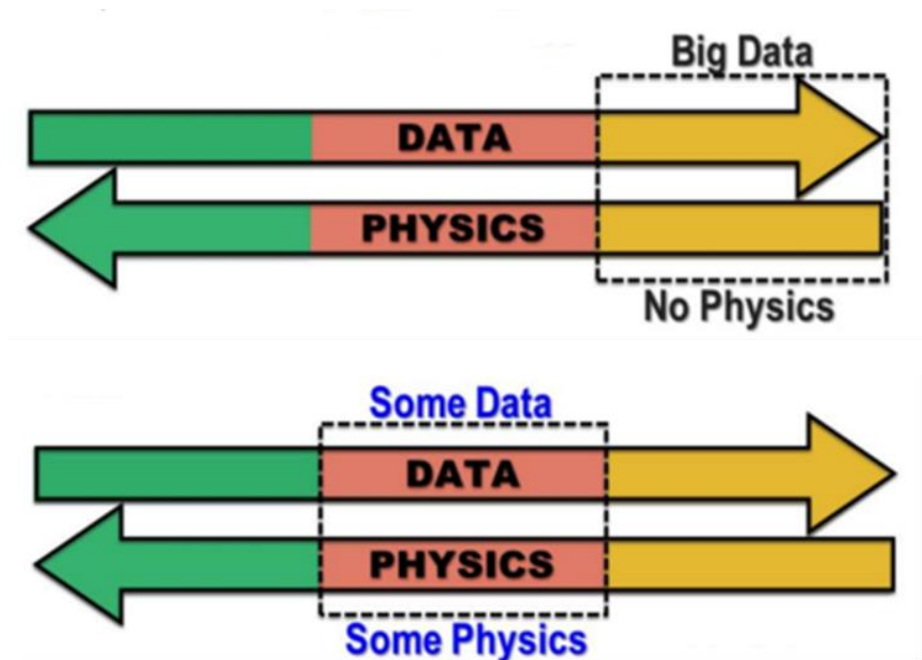


Figure 1.1. The top picture shows the data driven NN structure which lacks physics. For NN, big data is needed for training. In the physics informed NN structure, which is shown in the bottom picture, by adding some physics, the results can converge sooner with less data (Yang, Zhang and Karniadakis 2018).

## 1.5 Research Novelty and significance

A new emerged deep learning framework named as “Physics-Informed Neural Networks (PINN)”, (Raissi et al. 2019; Mao et al. 2020) is used in this research to model the relationship between soil moisture and GNSS reflected signals. This research is the first practice of PINN in GNSS-R soil moisture estimation, to the author’s knowledge. PINN algorithms have the advantage of being both data-driven to learn a model and also being able to ensure consistency with physics relationships between data parameters. As such, PINNs can generate more robust models, with less data. PINN algorithms are designed to consider any symmetries, invariances, or conservation principles originating from the physical relationships between observables (Shin et al. 2020). PINNs are trained while respecting any governing physics equations between datasets. A physics-data space can be defined to have a better comparison for different kinds of scientific modelling (See Figure (1.2)). Figure1.2 explains how physics-informed neural networks take advantage of both NN regression and physics-based laws at the same time in order to reduce the need for more / big data (Tartakovsky et al. 2020). In PINN, a neural network is trained with data that does not necessarily need to be large or complete due to the addition of physics laws.

The problem of GNSS-R soil moisture retrieval can be improved by physics-informed neural networks (PINN). Predictive modelling of soil moisture with GNSS-R data due to partially known physics and sparse data still presents a significant challenge. The main advantage of using PINN will be convergence of results with less data and more understanding of physics especially with comparison to NN algorithms. In NN modelling, the physical relationship of parameters is neglected. Therefore, the research aim is to

combine physics laws and NN in soil moisture retrieval problem to inform NN with physics of the problem using PINN as a new emerging methodology.

Although previous research has shown that neural networks can be used to retrieve soil moisture from GNSS-R signals, a single reliance on data in data-driven NNs may not be enough to truly understand a scientific problem. In this research, we investigate the suitability of PINNs to retrieve soil moisture from GNSS-R signals for the first time to the best of the author's knowledge. Here, the PINN model is learned using a data-driven supervised neural network, but also using physics equations to make sure it is consistent with the known physics. The advantages of using PINN in GNSS-R soil moisture estimation are that PINNs can learn a model from fewer data, and they can also ensure consistency with the physics and extrapolate beyond data.

## **1.6 Dissertation outline**

The integration of NN regression and electromagnetic physics relationships in PINN algorithm to produce soil moisture estimates from GNSS reflected signals is explained in this dissertation. The performance of PINN in different scenarios such as different soil texture and land cover classes are investigated. Chapter 2 explains GNSS-R theory and important GNSS-R observables. Chapter 3 describes physical models added to NN training in PINN modelling. Chapter 4 describes the structure of PINN for soil moisture retrieval using CYGNSS data and other geophysical parameters. Chapter 5 details the data used in PINN. Data used in this research includes GNSS-R data from CYGNSS satellites, surface parameters data such as vegetation, surface roughness, soil texture, soil temperature and land cover data.



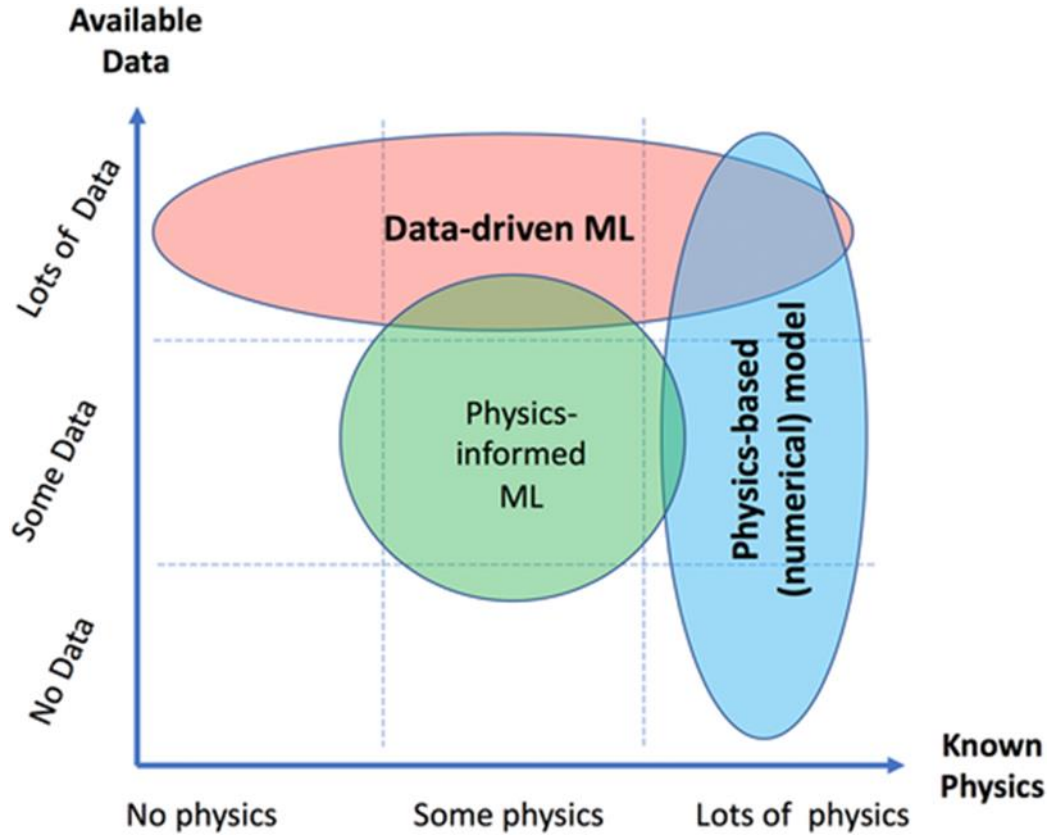


Figure 1.2 Physics-informed machine learning comparison with physics-based models and data-driven machine learning in data-physic space plot. Physics-informed Machine learning techniques work with less data and more physics knowledge as they integrate benefits of both machine learning techniques and physics laws (Tartakovsky et al. 2020).

PINN performance is discussed for different NN structures and land covers in Chapter 6. In Chapter 7, the effect of different soil dielectric models on PINN is investigated as soil dielectric constant models are an important part of physics dependencies of data. Finally, Chapter 8 summarizes the dissertation and suggests future possible research.

## **2. PRINCIPLES OF GNSS REFLECTOMETRY**

Due to the advancement of Global Navigation Satellite Systems (GNSS), more applications than navigation and positioning have been discovered for these systems. The freely-available signals of GNSS signals can be used as a remote sensing tool to survey the Earth's surface geophysical properties. GNSS-Reflectometry (GNSS-R) applies reflected GNSS signals for the Earth's surface characteristics assessments such as ocean and sea altitude (Rius et al. 2012; Ruffini et al. 2004), ocean wind speed measurements (Komjathy et al. 2004; Guan et al. 2018), soil moisture content (Camps et al. 2016; Chew et al. 2016; Chew et al. 2018;) and ice depth (Rius et al. 2017; Ghiasi et al. 2020). This research focuses on soil moisture studies application of GNSS-R. In this chapter, the basics of GNSS reflectometry are explained, then GNSS-R observables are described, and GNSS-R scattering geometry is detailed.

### **2.1 GNSS-R theory**

Reflected GNSS signals have been used in remote sensing of soil moisture due to following reasons: First GNSS signals are in L-band, which is able to penetrate in soil up to 5 cm and is sensitive to moisture variations in soil. Secondly, in contrast to microwave radiometry, even though variations in thermal background influence land bio-geophysical observables, they do not contaminate the GNSS reflected signals. Thirdly, GNSS scatterometry from space has potentially higher spatial resolution than microwave radiometry, due to the highly stable carrier and code modulations of GNSS signals.

Factors that affect the reflected signal power are categorized as geometric and physical properties. The effect of changing geometry defined by satellite elevation angle, transmitter and receiver position, and velocity on signal energy changes should be considered in the soil moisture retrieval algorithm. Physical surface parameter effects for instance surface roughness, vegetation, and soil texture must be better estimated in the soil moisture algorithm.

### **2.1.1 Global Navigation Satellites Systems**

Global Navigation Satellite Systems (GNSS) are constellations of satellites designed for providing time and position information. The worldwide GNSS satellites are the United States Global Positioning System (GPS), the Russian GLONASS, the European Union Galileo, and the Chinese Beidou. For regional positioning missions, the Japanese Quasi-Zenith Satellite System (QZSS), the Indian Regional Navigation Satellite System (NavIC), and the first Beidou constellation were launched. The first demonstration satellite of the Japanese Quasi-Zenith Satellite System (QZSS), QZS-1 was launched in 2010 by the Japan Aerospace Exploration Agency (JAXA). The building of NavIC was authorized by the Indian government in 2006. NavIC combines 3 geostationary and 4 geosynchronous satellites for a similar purpose. QZSS uses just 3 satellites, each 120° apart, in highly inclined, slightly elliptical, geosynchronous orbits, whose ground tracks are asymmetrical 8-like patterns, designed to ensure that at least one is always almost directly at an elevation of 60° or higher over Japan.

GLONASS was developed as a military system as a continuous study on the Doppler satellite system Tsikada by USSR the Union of Soviet Socialist Republics (USSR)

(Hofmann-Wellenhof et al. 2007). GLONASS satellite orbits have an altitude of 19,100 km with an inclination of 64.8 degrees and a period of 11 hours and 15 minutes (see Table 2.1). GLONASS transmits in Frequency Division Multiple Access (FDMA) on 15 channels. The signals are modulated with Direct Sequence Spread Spectrum (DSSS) and Binary Phase Shift Keying (BPSK) modulation (Franco 2013). Figure 2.1 shows a GLONASS satellite.

Galileo was built by the European Union (EU) and the European Space Agency (ESA), with a constellation of 30 satellites at 23.222 km of altitude, (see Table 2.1). The Galileo signal modulation is the BOC (Binary Off-set Carrier) modulation, which improves the tracking performance. Figure 2.2 illustrates a Galileo satellite.

BeiDou Navigation Satellite System (BDS) was developed by China for the Asian-Pacific region in 2012 (Gao et al. 2014). BeiDou constellation consists of 40 satellites including five geostationary (GEO) satellites, five satellites with an Inclined Geosynchronous Orbit (IGSO,) and 30 Medium Earth Orbit (MEO) satellites similar to GPS, GLONASS, and Galileo satellites, but with different orbital parameter values (Santerre et al. 2014), (see Table 2.1). Figure 2.3 shows a MEO BeiDou satellite.

For this research, the focus is on GPS as it is the most widely used GNSS, started by the United States in the early 1970s (Parkinson and Gilbert 1983; Spilker et al. 1996). GPS is an all-weather, space-based system initiated by the U.S. Department of Defense (DoD) for military use to determine position, velocity, and time in a common reference system. While the GPS primary goals were military, civil use of GPS was later promoted by the US congress (Hofmann-Wellenhof et al. 2007).



Figure 2.1 Russian GLONASS satellite ([www.russianspaceweb.com/uragan.html](http://www.russianspaceweb.com/uragan.html))

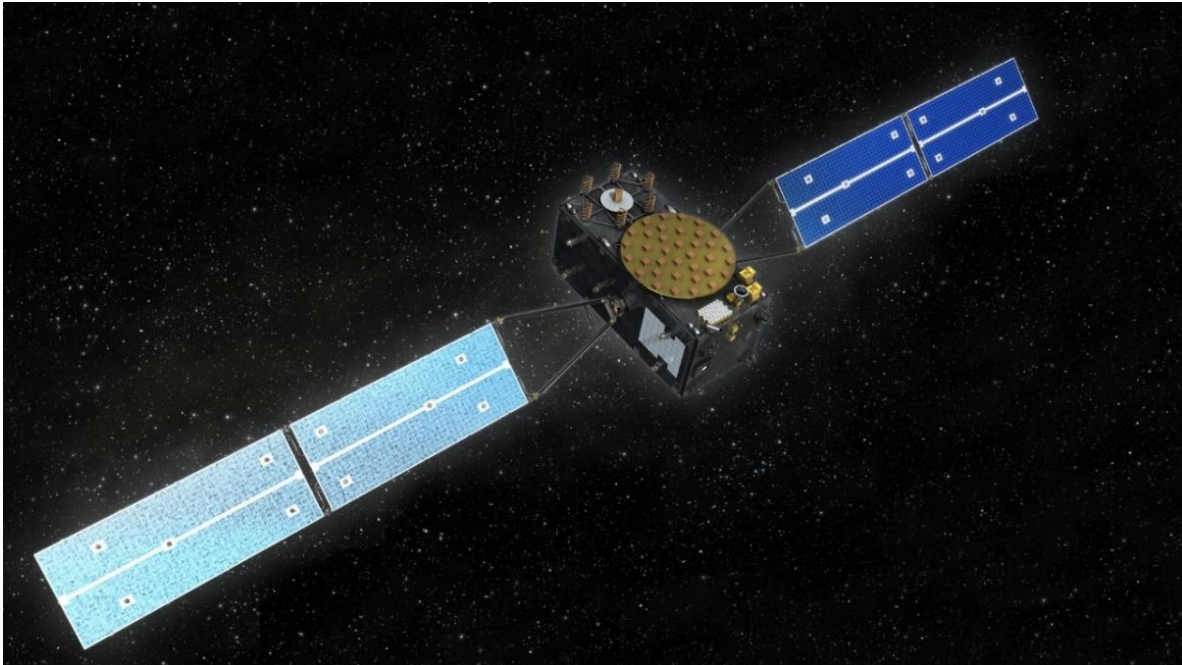


Figure 2.2. European Galileo satellite  
([https://www.esa.int/ESA\\_Multimedia/Images/2011/08/OHB-designed\\_Galileo\\_satellite](https://www.esa.int/ESA_Multimedia/Images/2011/08/OHB-designed_Galileo_satellite))

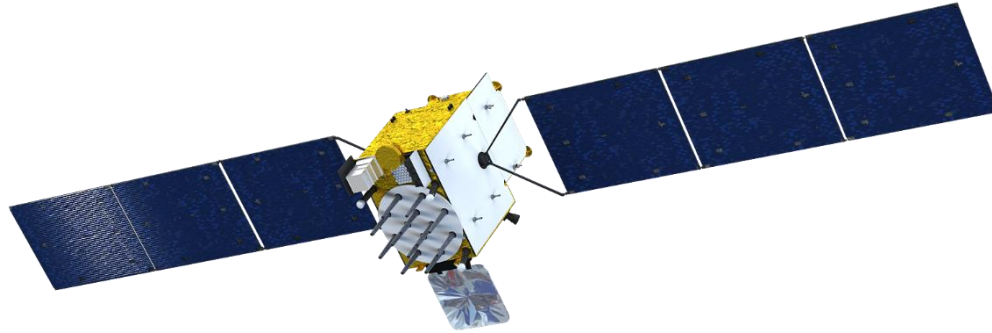


Figure 2.3. China Beidou satellite (<https://www.csno-tarc.cn/en/system/introduction>)

In 1993, the system achieved operational capability of 24 satellites with 4 satellites in each of 6 elliptical orbits. Each orbit is in nearly circular with an eccentricity of 0.01 and an inclination of  $55^\circ$  with a period of 12 sidereal hours, and the orbit has a semi-major axis of 26,560 km (Spilker et al. 1996; El-Rabbany 2002).

The GPS constellation is a mix of different satellites. Different generations of GPS satellites include Block IIA (2nd generation, "Advanced"), Block IIR ("Replenishment"), Block IIR-M ("Modernized"), Block IIF ("Follow-on"), GPS III, and GPS IIIF ("Follow-on"). Figure 2.4 summarizes the current operational GPS satellites.

### **2.1.2 GNSS-R**

GNSS reflectometry (GNSS-R) is the technique of studying GNSS scattered signals from the Earth's surface to assess geophysical parameters of Earth (Gleason and Gebre-

Egziabher 2009). GNSS-R signals are transmitted from GNSS satellites and received in a receiver with a considerable distance from the transmitters; therefore, GNSS-R geometry can be considered as a bistatic system (see Figure 2.5). However, the GNSS-R system has been described as a multi-static system by (Jin et al. 2014) due to the simultaneous use of multiple transmitters increasing the coverage and data samples (see Figure 2.6). So, GNSS-R utilizes multiple observations from free, versatile, and available signals of GNSS as a low-cost system for different Earth's surface studies. Reflected signals of GNSS satellites enable sensing the Earth's surface geophysical parameters as a new, continuous, all-weather remote sensing tool. Comparison of direct and reflected signals is the principle of GNSS reflectometry theory.

Table 2.1 Comparison of different GNSS constellations (Shi and Wei 2020).

	<b>GPS</b>	<b>GLONASS</b>	<b>Galileo</b>	<b>BeiDou</b>
<b>Owner</b>	United States	Russia	European Union	China
<b>First Launch</b>	1978-02-22	1982-10-12	2005-12-28	2017-11-05
<b>Number of satellites</b>	32 (at least 24 by design)	28 (at least 24 by design)	30	30 MEO And 5 GEO
<b>Number of orbital planes</b>	6	3	3	3 (MEO)
<b>Orbital inclination</b>	55°	64.8°	56°	55°
<b>Orbital altitude</b>	20,200	19,100	23,222	21,528
<b>Orbital period</b>	11 h 58 m	11 h 15 m	14 h 04 m	12 h 53 m

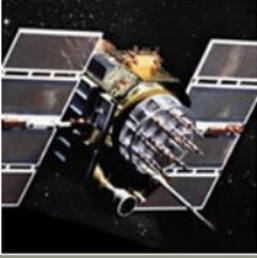
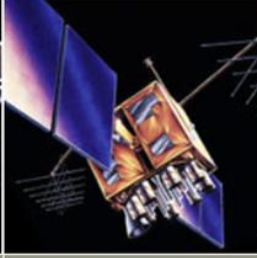
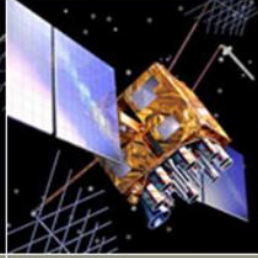
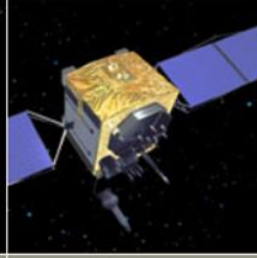
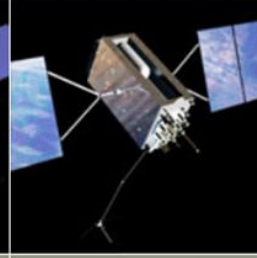
LEGACY SATELLITES		MODERNIZED SATELLITES		
				
BLOCK IIA	BLOCK IIR	BLOCK IIR-M	BLOCK IIF	GPS III/IIIF
0 operational	6 operational	7 operational	12 operational	6 operational

Figure 2.4 GPS operational satellites (<https://www.gps.gov/systems/gps/space/>).

Surface properties assessment using reflected GNSS signals includes two techniques based on the receiver antenna (Zhang et al. 2017). The first technique known as GNSS interferometric reflectometry (GNSS-IR) considers the use of one antenna and the resulting signal-to-noise ratio (SNR) data (Larson 2016) (see Figure 2.7a).

However, the second technique, GNSS reflectometry (GNSS-R), benefits from a receiver with two zenith and nadir antennas to acquire waveform (Zavorotny et al. 2014) (see Figure 2.7b). While GNSS-IR studies changes in the interference pattern of the direct and reflected signals recorded in SNR data, GNSS-R uses delay-Doppler maps (DDMs) related to acquired scattered GNSS signals (Zavorotny et al. 2014).

Using GNSS-IR, Larson et al. (2008 a) compared estimated multipath phase to measured Volumetric Water Content (VWC) from Water Content Reflectometers (WCR) and expressed their relationship with a 2<sup>nd</sup> order polynomial (Larson et al. 2008 a). In Larson et al. (2008b), a good correlation between the GPS multipath amplitude and a land model precipitation record was found.



In GNSS-IR, the main data is SNR of GNSS L band C/A signal. GPS multipath SNR can be defined in terms of the multipath amplitude ( $A$ ) and phase ( $\phi$ ) using Equation 2.1, (Larson et al. 2008b)

$$SNR = A \cos\left(\frac{4\pi h}{\lambda} \sin E + \phi\right) \quad (2.1)$$

where  $h$  is antenna height from the top of the soil,  $E$  is the elevation angle of the satellite and  $\lambda$  is the GPS wavelength. To calculate the multipath amplitude and phase from the SNR data, a least square estimation method can be applied as explained in (Larson et al. 2008a; Larson et al. 2008b). With the knowledge of multipath amplitude and phase, soil moisture changes can be studied. Larson et al. (2008a) demonstrated that the multipath SNR amplitude variations are a function of soil moisture content (Larson et al. 2008a). However, most studies practiced phase data for soil moisture retrieval since variations of shallow soil moisture owing to phase are larger than amplitude (Larson et al. 2008b). A strong relationship between phase and soil moisture measurements was later corroborated using a model of ground-reflected signal (Chew et al. 2013).

The gain pattern and multipath intensity factors influences are included in the amplitude term. Both the antenna's gain and multipath intensity vary with the elevation angle, however, the variations due to satellite elevation angle changes are not large in amplitude. Therefore, in an analysis of SNR data, it is assumed that amplitude does not vary with elevation angle (Larson et al. 2008a).

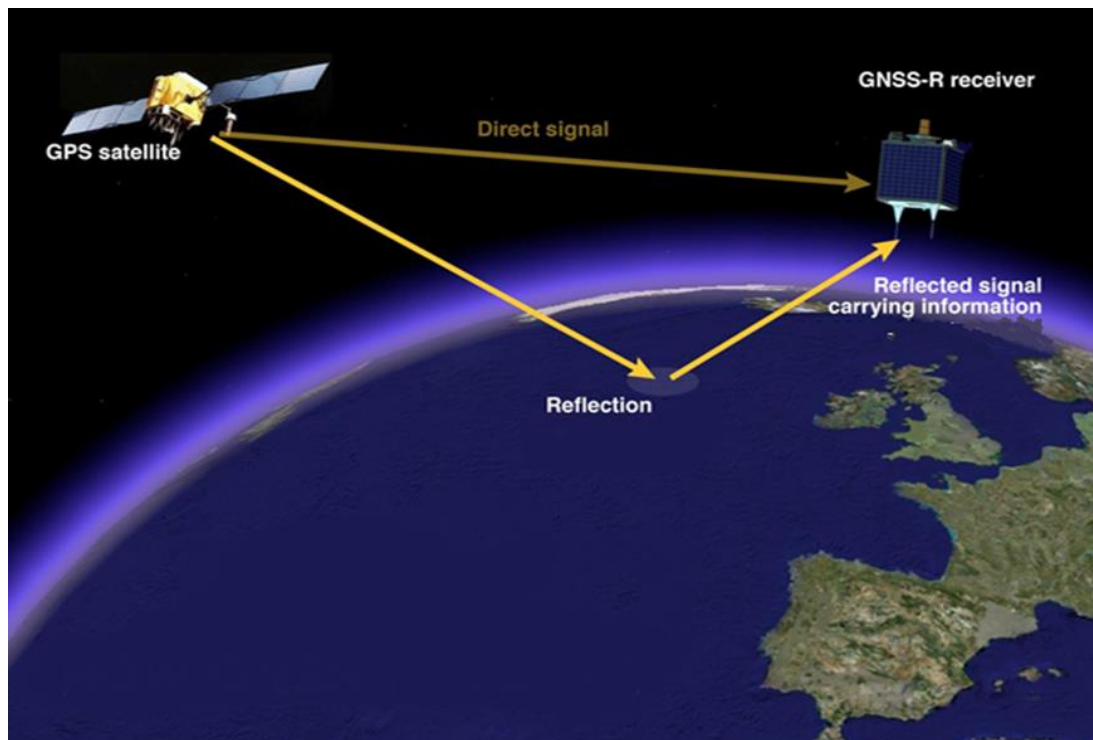


Figure 2.5. GNSS-R technique as a bistatic system

(<https://www.hydro-international.com/content/news/sea-level-mapped-from-space-with-gps-reflections>)

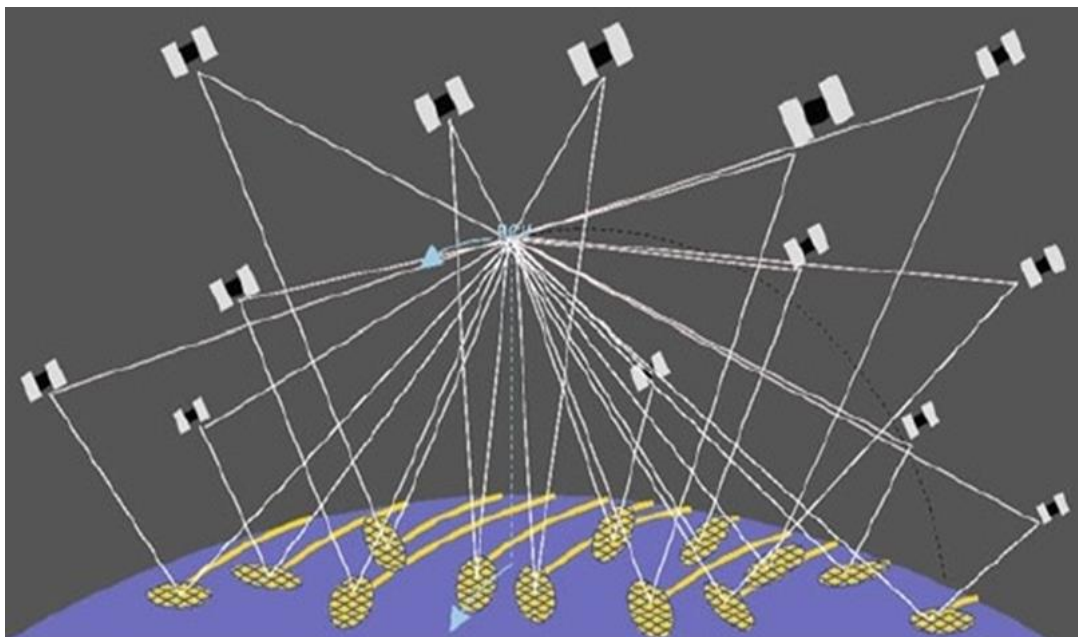
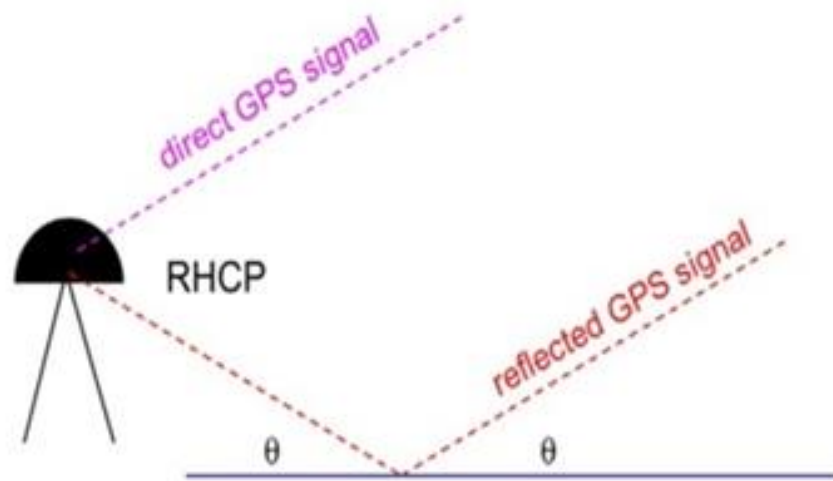
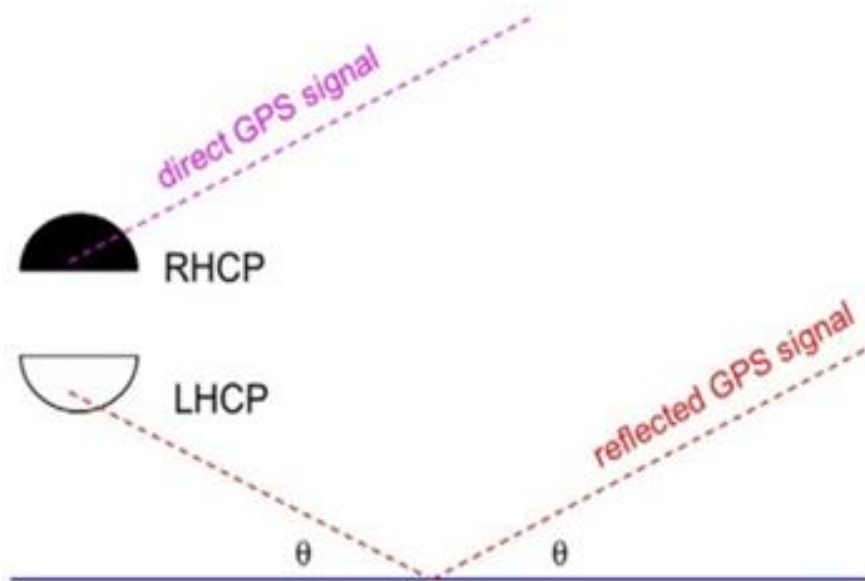


Figure 2.6. GNSS-R as a multi-static system (Jin et al. 2014)



(a)



(b)

Figure 2.7. Comparison of GNSS-IR and GNSS-R in terms of antenna use, a) GNSS-IR with one right hand circular polarized antenna (RHCP), and b) GNSS-R with two antennas including one zenith RHCP and one nadir left hand circular polarized antenna (LHCP).

GNSS-IR SNR data is not only used in soil moisture studies, but also for vegetation permittivity. Chew et al. (2016) suggested that for the changes in vegetation permittivity and canopy height, the amplitude is more effective metric than phase. The normalized amplitude time series can demonstrate the vegetation effects significance (Chew et al. 2016).

GNSS satellites transmit electromagnetic waves of the radio or microwave type with frequency in the range of 1 to 2 GHz, in the L band close to the frequencies used in SMOS and SMAP orbital missions. When such waves reflect on surfaces around a GNSS antenna, they can reach the receiver's antenna indirectly, delayed due to the additional distance travelled. Receiver antenna can be mounted on different platforms such as ground-based (Jia et al. 2019), UAV-based (Senyurek et al. 2021; Imam et al. 2019), aircraft-based (Jia et al. 2015; Lowe et al. 2020), and satellite-based GNSS receivers (Eroglu et al. 2019) (see Figure 2.8).

## **2.2 GNSS-R observables**

GNSS receivers have been traditionally more dependent on hardware. However, with a software GNSS receiver, hardware is still needed, known as the frontend, that digitizes the signal from the satellites. A software-defined radio (SDR) implements most processes. The SDR finds the transmitted signal by searching over a wide range of Doppler shifts and code delays (Gleason and Gebre-Egziabher 2009) (see Figure 2.9). The result of the SDR operation can be shown as Delay Doppler Maps (DDMs), delay waveforms, and a measurement of bistatic radar cross section (BRCS) which are the main observables in GNSS-R.

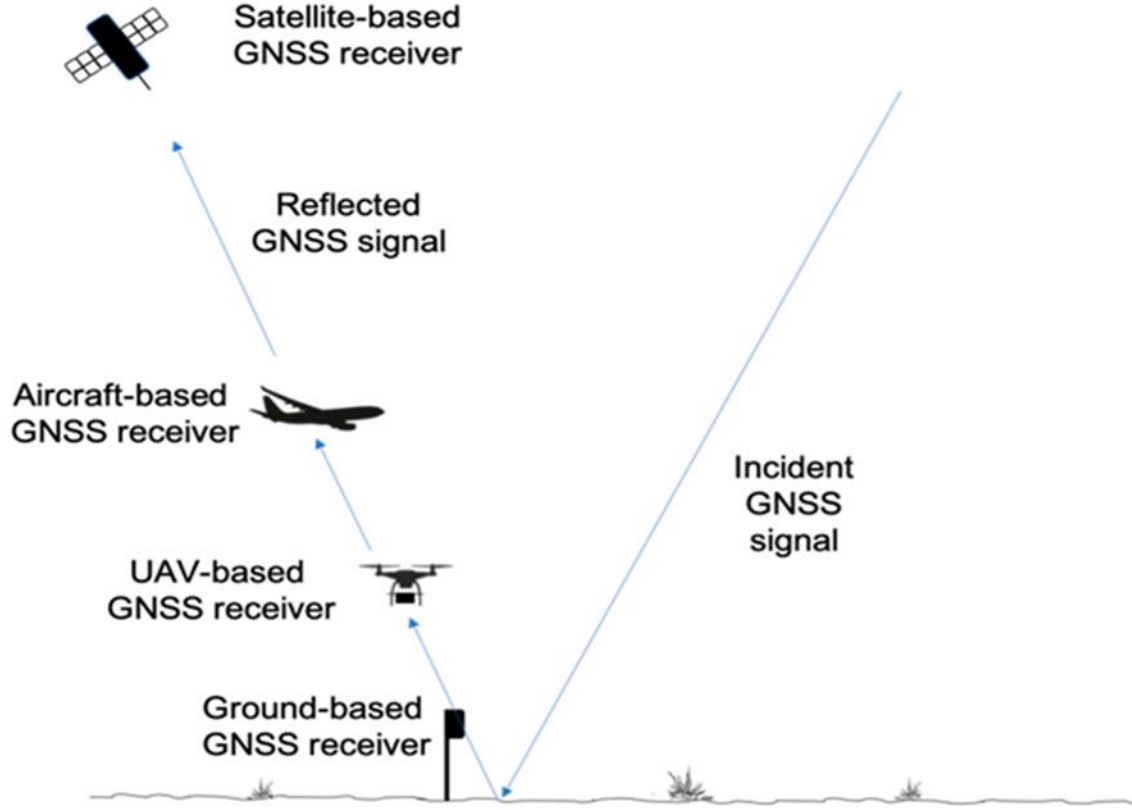


Figure 2.8. Different platforms that can be used for GNSS-R soil moisture retrieval (Rodriguez-Alvarez et al. 2023).

The GPS receiver cross-correlates the incoming signal ( $u$ ) at a time ( $t_0 + \tau$ ) with a replica of PRN code ( $a$ ) taken at a different time ( $t_0$ ) for a variety of time delays ( $t' + \tau$ ) to measure the similarity between two signals. When the maximum correlation  $y$  is found, the signal is aligned with the replica code. The result is expressed as follows (Parkinson 1996)

$$y(t_0, \tau, f_c) = \int_0^{T_i} a(t_0 + t')u(t_0 + t' + \tau)\exp(2\pi if_c t') dt' \quad (2.2)$$

where  $T_i$  is the coherent integration time,  $a(t)$  is the replica of the PRN code,  $\tau$  is the time delay of incoming signal with respect to the replica code. A GPS signal is received by multiplying the incoming signal ( $u$ ) by a replica code ( $a$ ) and carrier frequency ( $2\pi if_c t'$ ). For the signal to be received out of the noise the replica code and carrier must align with that in  $u$ . Carrier frequency compensates for the possible doppler shift of the signal with the oscillating factor including the code C/A frequency  $f_c$ .

The possible Doppler shift of the signal is compensated with the oscillating factor including the C/A code frequency  $f_c$  in local carrier term ( $\exp(2\pi if_c t')$ ). The cross-correlated signal power is then used to form delay waveforms, DDM, and BRCS which are the basic observables for the Earth's surface studies.

A delay Doppler map (DDM) is formed from the correlation of the incoming signal  $u(t)$  with a replica of the code  $a(t)$  for several Doppler and code values. A DDM shows the correlation power versus both the code delay and the Doppler frequency that is produced due to the relative motion between the transmitter and the receiver. A delay waveform represents correlation power distribution over code phase/delay for a constant Doppler (see Figure 2.9). Figure 2.9 demonstrates how correlation power  $y$  is calculated using Equation 2.2 over different delays and possible Doppler shifts.

After finding correlation power values, averaging of the signal over consecutive correlations is necessary due to fading caused by the surface scattering (Gleason and Gebre-Egziabher 2009). This process is called incoherent integration. In GPS signal processing a coherent integration of 1 ms is usually used due to the length of GPS PRN repeat sequence (Gleason 2006). The signal must be summed or averaged over consecutive 1 ms coherent integrations to increase signal power over fading noise. The only adjustment needed between consecutive averaged power outputs at a given

frequency and delay is the centre delay, which needs to be adjusted at each millisecond to correctly align the signal as it moves across the Earth's surface (Gleason 2006). No matter what the desired output is, averaging the signal is necessary in generating all products including delay waveforms, DDMs, BRCS. Incoherent integration is used both for GNSS signal acquisition and to estimate DDMs, delay waveforms, and BRCS as it increases the observable SNR and signal power and reduces fading noise (Gleason and Gebre-Egziabher 2009). In incoherent integration, signals must be summed or averaged over 1000 number of consecutive 1ms coherent correlations. From experience, it has been found that when the summation interval is limited to 1 second, the only term that has a noticeable effect on the shape of the returned signal is the first derivative of the code delay. The basic processing ignores all dynamics-related frequency terms over the summation interval, as well as higher order derivatives related to delay. Over one second, the relative centre frequency (the change in frequency between the first and 1000th look at the reflected signal) changes on the order of several 10's of Hz. For 1 second of summation, this is negligible, but needs to be taken into consideration for longer summation intervals. Therefore, 1 second of summation has been accepted by researchers (Marchan-Hernandez et al. 2008; Hu et al. 2019; Huang et al. 2021).

To estimate the GNSS signals behaviour after scattering, the Kirchhoff approximation model can be used (Bass and Fuks 2013). The received GNSS signal  $u(\vec{R}, t)$  at the receiver position of  $\vec{R}$  is computed as Equation 2.3

$$u(\vec{R}, t) = \int D(\vec{p}) a \left[ \frac{t - (R_t(t) + R_r(t))}{c} \right] g(\vec{p}, t) d^2p \quad (2.3)$$

where  $D(\vec{p})$  is the amplitude footprint of the receiver antenna,  $a(t)$  is the GNSS signal PRN code,  $R_t(t)$  is the distance to the transmitter and  $R_r(t)$  is the distance to the receiver for the point  $\vec{p}$ , and  $g(\vec{p}, t)$  defines propagation and scattering processes in Equation (2.4), with respect to Fresnel reflection coefficient ( $V$ ), and distances of reflection point to transmitter  $R_t(t)$ , and receiver  $R_r(t)$  (Gleason and Gebre-Egziabher 2009)

$$g(\vec{p}, t) = -V(\vec{p})q^2 \exp \frac{[ik(R_t(t) + R_r(t))]}{4\pi i R_t R_r q_z} \quad (2.4)$$

where  $V$  is the Fresnel reflection coefficient,  $q$  is the scattering vector, and  $k$  is the radio wave number ( $k = 2\pi/\lambda$ ). Fresnel reflection coefficient describe the reflection of light when incident on an interface between different optical media .After multiple calculations (Zavorotny and Voronovich 2000), the bistatic radar equation computing the power of a scattered GNSS signal is obtained as follows (Jin et al. 2014)

$$\langle |Y(\tau, f)|^2 \rangle = \frac{\lambda^2 T_i^2}{(4\pi)^3} P_t \int \frac{G_t G_r}{R_t^2 R_r^2} \Lambda^2(\tau - \tau(r)) |S(f_D(r) - f_c)|^2 \sigma^0 d^2 r \quad (2.5)$$

where:

- .  $T_i$  is the integration time
- .  $P_t$  is the transmitter power
- .  $\lambda$  is the electromagnetic wavelength
- .  $G_t$  and  $G_r$  are the transmitter and receiver antenna gains, respectively
- .  $\tau(r)$  is the delay of the ray path from the transmitter to the surface point  $r$  and from there to the receiver, and  $f_D(r)$  is the Doppler frequency of the ray path



- .  $R_t$  and  $R_r$  are the transmitter and receiver distances to point  $r$ , respectively
- .  $\Lambda^2$  function defines the annulus zone, and  $|S|^2$  function defines the Doppler zone
- .  $\sigma^0$  is the normalized bistatic radar cross-section (BRCS) of the rough surface

The BRCS function can be obtained from the Kirchhoff approximation through Equation ( $\sigma^0$ ). BRCS function is one of the GNSS-R observables that can be used to study the ocean wind speed and direction (Zavorotny and Voronovich 2000; Elfouhaily et al. 2002).

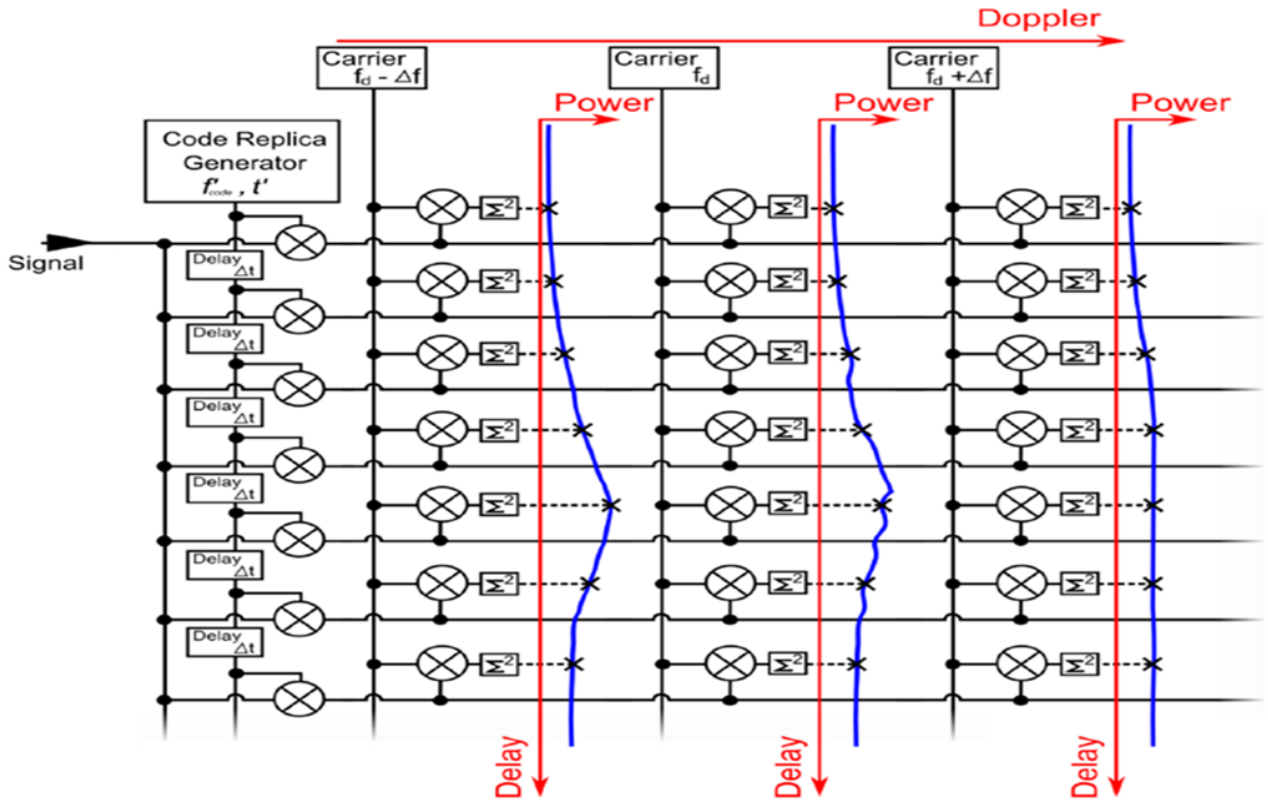


Figure 2.9. Delay Doppler Map (DDM) calculation in a software defined radio (SDR) (Jales 2012). For every delay and Doppler value, the correlated power can be drawn (blue plot). These blue plots are different delay waveforms while Doppler is constant.

Correlated powers drawn over delay and Doppler represented as DDMs are the primary data used in GNSS-R. Surface roughness can be declared with the distribution of correlative power across the DDM (Chew 2016). DDM shape shows the surface roughness. When the frequency is set to a constant value (normally the value at the reflection point), the delay waveform is obtained. The delay waveform is simply a two-dimensional representation of DDMs for a constant frequency (see Figures 2.10 and 2.11).

When the surface is smooth, the delay waveform has an undistorted triangular shape (similar to direct received signals with a time delay), and the reflection is mostly coherent coming from the area equal to the size of the Fresnel zone. For a smooth surface, DDM can have the shape of Figure 2.10.

When the surface is roughened (e.g., oceans), delay waveform has a horseshoe shape (Figure 2.11) and is mostly incoherent coming from a larger area known as glistening zone (Jin et al. 2014). In this case, the maximum power appears on the trailing edge of the waveform corresponding to the specular point. For points further than specular point, the reflection is not specular and the reflected power decreases.

The distribution of correlation power over delay ranges mapped as delay waveforms provide the basic measurements in GNSS-R. Figure 2.12 explains how different shapes for correlation power distribution over delay as waveforms can be applied to different GNSS-R applications. The delay of reflected signal with respect to the direct signal is the main concept of altimetry in GNSS-R (see Figure 2.12). The distribution of reflected power signal is the idea of surface roughness in the ocean or dielectric changes of soil.

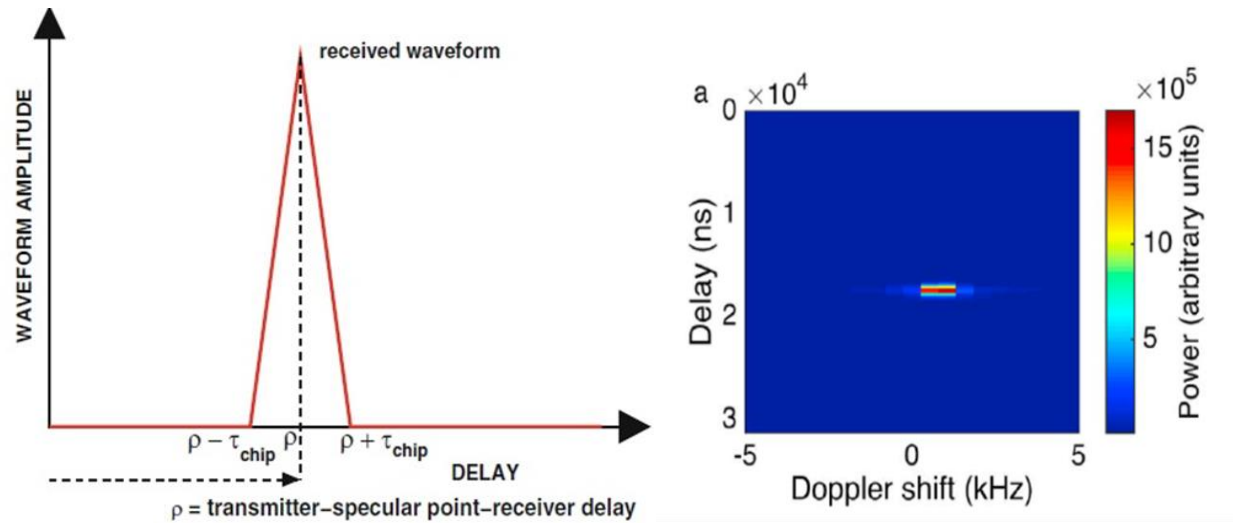


Figure 2.10. The left picture shows the delay waveform plot for a smooth surface with the specular reflection (Jin et al. 2014), and the right picture shows DDM over a smooth surface (Chew et al. 2016).

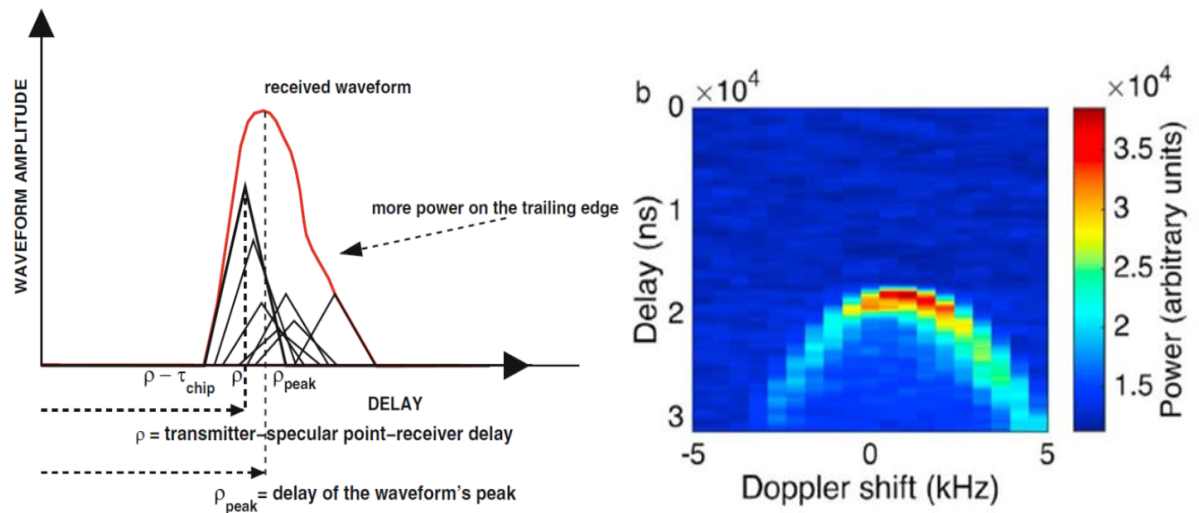


Figure 2.11. The left figure demonstrates the delay waveform for a rough surface reflection which can be expressed as coherent components showing in triangular shape (Jin et al. 2014), the right figure demonstrates DDM over a rough surface (Chew et al. 2016).

As the roughness increases the peak power of delay waveform decreases (Masters 2004). For example, bistatic cross section estimated from reflected power is a function of surface slope; therefore, a representative of roughness (Zavorotny and Voronovich 2000). Slope of delay waveform can later be related to trailing edge slope (TES) data, which is an indicator for coherency / incoherency of waveforms (Dong and Jin 2021; Rodriguez-Alvarez et al. 2019). Higher TES shows higher incoherency characteristics (see Figure 2.12).

The presence of coherent scattering is clearly seen in the shape of the DDMs over sea ice and over land, while over the ocean, DDMs spread over many delay (vertical axis) and Doppler (horizontal axis) cells (see Figure 2.13) (Clarizia et al. 2014). Some reasons explaining this behaviour are: i) for many scattering geometries (i.e., incidence angles), the long electromagnetic wavelength (e.g.,  $\lambda = 19$  cm at  $f_{L1} = 1575.42$  MHz) allows the surface to be considered “smooth,” despite natural surfaces exhibit a wide range of root mean square (rms) heights and correlation lengths; and ii) due to the long chip duration ( $T_{\text{chip}} \approx 1 \mu\text{s}$  for the GPS C/A code,  $c \cdot T_{\text{chip}} \approx 300$  m), volumetric effects are usually not resolved.

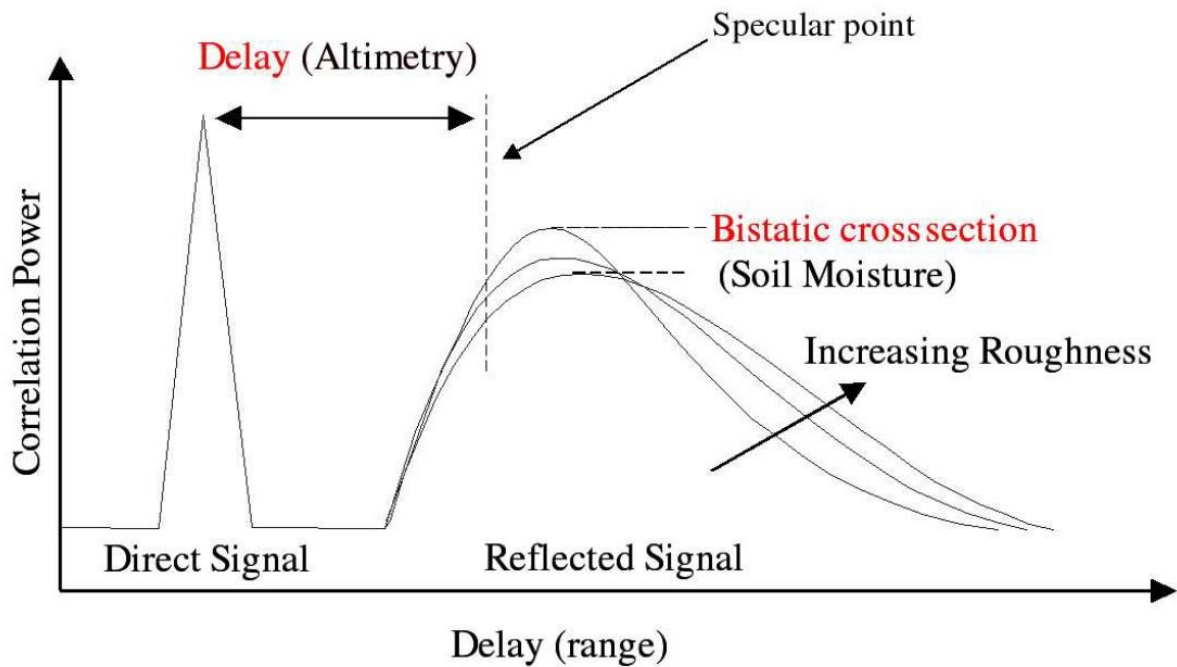


Figure 2.12. The relationship between delay waveform and GNSS-R applications (Masters 2004). Delay from the direct signal which can be expressed in phase difference is sensitive to receiver antenna relative height and provides altimetry information. The distribution of correlation power in waveform plots is the idea of studying surface properties such as ocean.

### 2.3. Scattering geometry

The magnitude of GNSS reflected signal power is highly affected by roughness of the scattering surface especially over the ocean. For example, with a fixed Doppler frequency, the correlation signal power of the reflected GNSS signal can be used to measure the roughness of the scattering area (Rodriguez-Alvarez et al. 2012). As surface roughness increases, the scattering zone area also increases. To understand the geometry of GNSS-R signals reflection, two physics terms should be studied: i) the specular point or reflection point explained in Subsection 2.3.1; and ii) the scattering zone or glistening zone explained in Subsection 2.3.2. The specular point and the zone around

it called the glistening zone represent the scattering geometry of the GNSS reflected power signal (Clarizia et al. 2009). Figure 2.14 shows basic geometry of GNSS reflected signals over land for specular points and glistening zone around it.

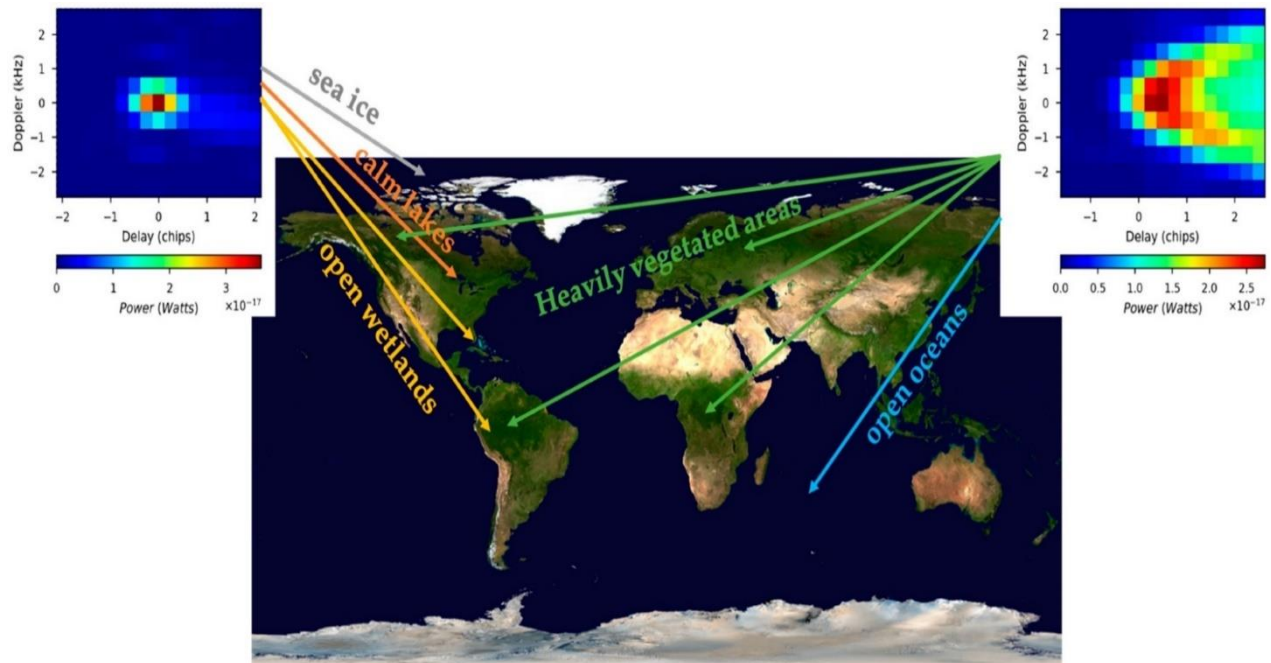


Figure 2.13. Comparison of coherent DDM for sea ice, calm lakes, and open wetlands versus incoherent DDM with horseshoe shape for open oceans and heavily vegetated areas. (Rodriguez-Alvarez et al. 2023).

### 2.3.1 Specular point calculation

Specular point is a reflection point on the earth's surface that satisfies two conditions: i) this point minimizes the total path between transmitter or GNSS satellite to reflection point and then to receiver; and ii) specular point satisfies the Snell's law as incident and reflected angles must be equal (see Figure 2.14).

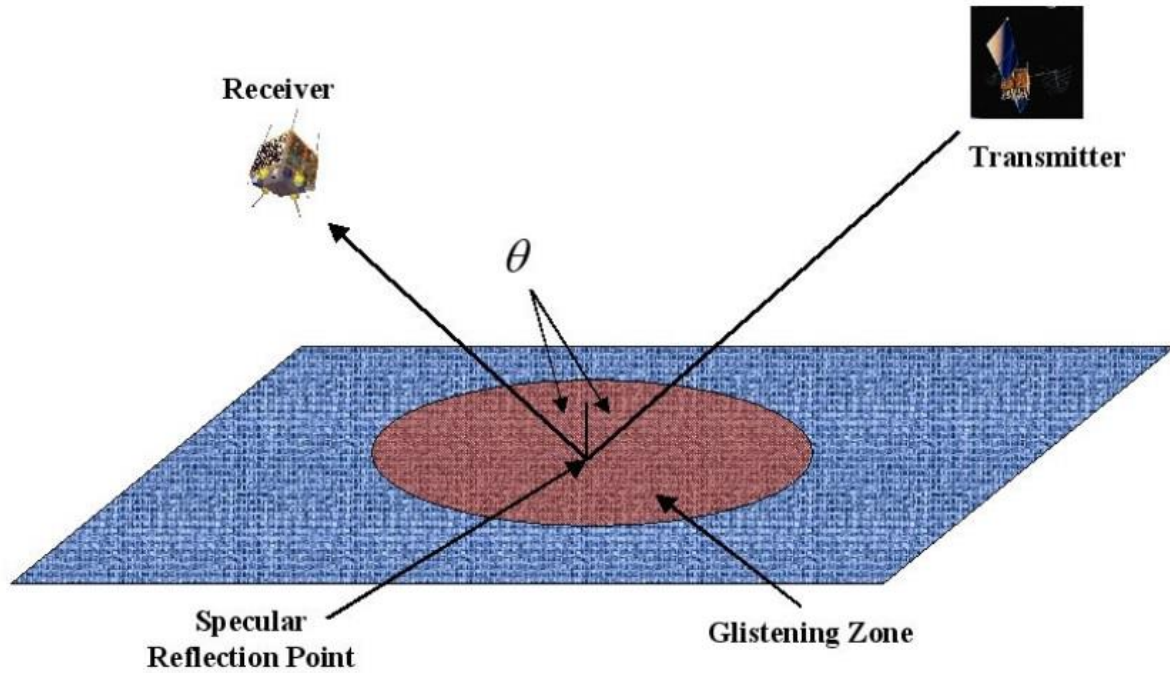


Figure 2.14. The geometry of GNSS reflected signals including specular point reflection and reflection area around it called glistening zone. For specular point, the incident and reflection angle are equal ( $\theta$ ) and the path between the receiver, reflection point, and transmitter is minimum (Gleason and Zavorotny 2006).

Different algorithms have been proposed to determine the positions of specular points. The first specular point position computation was proposed by Martin-Neira (1993) suggesting a quadric polynomial equation to find specular point on the spherical Earth. Then, a complete report on specular point position calculation was investigated by Roussel et al. (2014) that proposed three different algorithms based on the Earth's surface shape: the first considering the Earth as a sphere, the second as an ellipsoid, and the third with consideration of Earth's topography with digital elevation model (DEM). The precision of these methods is studied by Roussel et al. (2014). All these specular point

calculations use iterative methods to meet two conditions (path minimization and Snell's law).

Juang (2021) suggested an analytical methodology to determine specular points on the ellipsoidal Earth. Therefore, a sextic polynomial equation was found to determine specular point without a need for initial conditions and iterative methods (Juang 2021). More recent methodologies have focused on improving specular point position determination by adding geoid information and mean dynamic topography (Wu et al. 2021). In regions with high geoid undulations like the Greenland Sea, it's recommended to use a locally accurate Earth geoid model in specular point calculations.

The most famous approach is based on Gleason (2006). The computation of specular point is explained by Gleason (2010) in this subsection. To find the reflection point on the Earth's surface, the signal path  $P(S)$  is defined as a function of the specular point location as

$$P(S) = |(T - S) + (R - S)| \quad (2.8)$$

where T is the transmitter (GNSS satellite) position vector, R is the receiver position vector and S is specular point position vector in WGS84 reference frame. Equation 2.8 can be expanded to Equation 2.9, as (Gleason 2010)

$$P(S) = \sqrt{(T_x - S_x)^2 + (T_y - S_y)^2 + (T_z - S_z)^2} + \sqrt{(R_x - S_x)^2 + (R_y - S_y)^2 + (R_z - S_z)^2} \quad (2.9)$$



Receiver and transmitter locations are both known from navigation output of GNSS. Therefore, the coordinates of both receiver vector  $R$  and transmitter vector are known in WGS84 reference frame. As a result, specular point location is the only unknown parameter. Due to the non-linearity of Equation 2.9, to solve for specular position, an iterative approach with an initial guess is needed. To minimize the total path  $P(S)$  as the first condition of specular point, the partial derivatives of specular point with respect to  $x$ ,  $y$  and  $z$  should be calculated.

In order to minimize this path, we first take the partial derivatives of the specular point ( $S$ ) with respect to  $x$ ,  $y$ , and  $z$ . The partial derivative with respect to  $S_x$  is shown in Equation (2.10). The same approach can be applied for  $S_y$  and  $S_z$ .

$$\partial_{S_x} P(S) = \frac{(T_x - S_x)}{\sqrt{(T_x - S_x)^2 + (T_y - S_y)^2 + (T_z - S_z)^2}} + \frac{(R_x - S_x)}{\sqrt{(R_x - S_x)^2 + (R_y - S_y)^2 + (R_z - S_z)^2}} \quad (2.10)$$

Simplifying Equation (2.10) and including all three dimensions results in (Chan 2017)

$$dS = \partial_S P(S) = \frac{(T - S)}{|(T - S)|} + \frac{(R - S)}{|(R - S)|} \quad (2.11)$$

The Earth's radius with respect to WGS84 ellipsoid parameters is formulated as a function of the specular point  $z$  coordinate with  $\lambda = \sin\left(\frac{S_z}{|S|}\right)$  as follows,

$$r = a_{WGS84} \sqrt{\frac{1 - e_{WGS84}^2}{1 - (e_{WGS84}^2 + \cos(\lambda))}} \quad (2.12)$$

where  $a_{WGS84} = 6378137$  m and  $e_{WGS84} = 0.08181919084262$  are the semi-major axis and the eccentricity of the Earth's WGS84 ellipsoid, respectively.

To solve for specular point position that satisfies the conditions mentioned, a correction gain ( $K$ ) is considered. Magnitude of this correction gain depends on how far the sub-receiver Earth point can be from the actual specular point. Therefore, the initial guess for specular point location is sub-receiver Earth point and in each iterative step, the specular point location changes are modelled as (Chan 2017)

$$S_{temp} = (S + K \hat{S}) \quad (2.13)$$

where  $\hat{S} = \frac{\partial S}{|\partial S|}$  is the directional unit vector for the correction.

The new estimate for specular point  $S$  in each step of iteration will be calculated by Equation (2.14) with unit vector of  $S_{temp}$  scaled by the Earth's radius (Gleason 2010). Earth radius can be calculated from Equation (2.12).

$$S = r \hat{S}_{temp} = r \frac{S_{temp}}{|S_{temp}|} \quad (2.14)$$

Calculations will be repeated till the difference between the old and new values of  $S$  is below a predefined threshold. Finally, Snell's law as the last condition should be checked with consideration of the incoming and reflected vectors directions. Although Gleason's method has widely been used, choosing a suitable threshold for iterations and correction gain can affect the results of specular point position.

### 2.3.2 Resolution

The receiver experiences two scattering cases, hence affecting resolution of the scattering area in two different ways. With respect to the surface roughness, the scattering contains both coherent and incoherent types of reflection contributions (see Figure 2.15). From an electromagnetic point of view, surface roughness or smoothness definition is dependent on the transmitted wave frequency (or wavelength  $\lambda$ ) and incident angle ( $\theta$ ) which are not inherent properties of the surface itself. Considering Rayleigh criterion, the surface is defined smooth when the phase differences  $\Delta\phi$  between two reflected rays from two separate points on the surface is less than  $\pi/2$  radians (Moreira et al. 2013). The phase difference between reflected rays from two separate points with a roughness height standard deviation  $h$  is calculated as (Wolf 2007)

$$\Delta\phi = 2h \frac{2\pi}{\lambda} \cos\theta \quad (2.15)$$

Moreover, the reflection from a smooth surface is coherent indicating that the transmitted wave and received wave sources have a constant phase difference and the same frequency. Coherent scattering results in high spatial resolution GNSS reflectivity values because most of the power comes from the first Fresnel zone when coherent scattering occurs (see Figure 2.15).

Coherent reflection occurs in smooth flat surfaces where heights  $h$  of roughness are smaller than the wavelength  $\lambda$  of the signal. To be more precise in recognition of coherent reflection, if Rayleigh parameter calculated as Equation (2.15) is less than 1, the surface reflection is coherent (Zavorotny et al. 2014). The most power of reflected signals comes

mainly from the first Fresnel zone under coherent scattering (Beckmann and Spizzichino 1987). Fresnel zone is an ellipse with a semi-major axis  $a$  and semi-minor axis  $b$  defined as Equation (2.16) and Equation (2.17), respectively (Larson and Nievinski 2013).

$$a = \frac{b}{\sin(\epsilon)}, \quad (2.16)$$

$$b = \sqrt{\frac{\lambda h}{\sin(\epsilon)} + \left(\frac{\lambda}{2\sin(\epsilon)}\right)^2} \quad (2.17)$$

where  $\lambda$  is the wavelength (m),  $h$  is the receiver height (m) and  $\epsilon$  is the satellite's elevation angle seen from the specular point (rad).

Due to the weakness of the reflected GNSS signals, the GNSS bistatic radars only receive the strongest scattered signals. Therefore, receivers capture signals coming from specular point and the glistening zone angular area around it (see Figure 2.15).

In summary, for coherent scattering, the spatial resolution of reflection is high being as small as 400-500 m for a LEO satellite. However, as the roughness heights become significant with respect to the signal's wavelength, diffuse scattering occurs instead of specular reflection. Although diffuse scattering is still centered on the nominal specular point, reflections come from the larger area called glistening zone resulting in coarse spatial resolutions (see Figure 2.15).

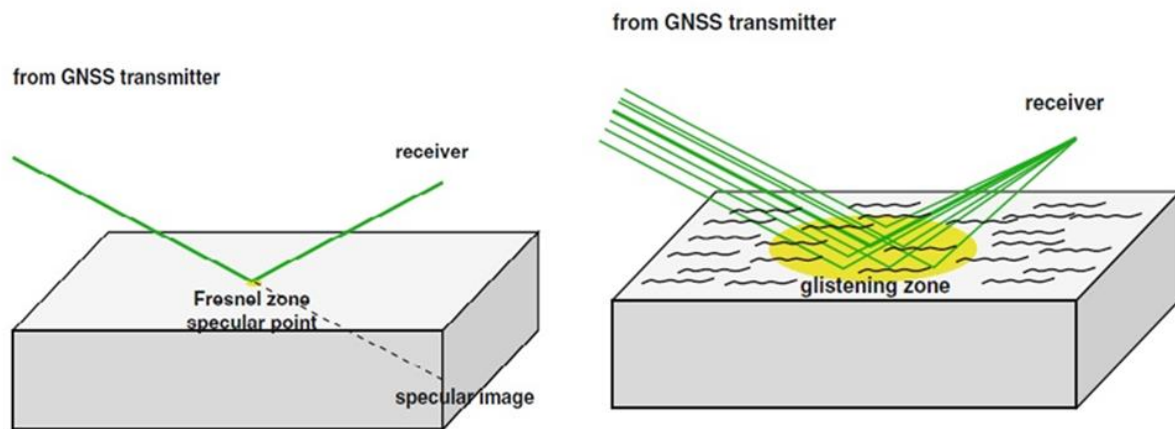


Figure 2.15. The left picture shows specular scattering or reflection from smooth surfaces which comes from the first iso-delay ellipsoid called Fresnel zone. The right picture demonstrates diffuse scattering from rough surfaces from larger area called glistening zone (Jin et al. 2014).

### **3. PHYSICAL MODELS IN GNSS-R SOIL MOISTURE RETRIEVAL**

In this Chapter, the theoretical relationship between soil moisture and reflected signal, and datasets employed in PINN problem are explained. For soil moisture retrieval using bistatic radar observations, theoretical physics relationships between different surface parameters must be understood. Theoretical physics relationships in Section 3.1 describe how datasets in PINN are connected in terms of physics models of the scattering surface.

#### **3.1 Theoretical relationships in soil moisture retrieval using GNSS-R**

The GNSS-R works as a bi-static system receiving GNSS L-band signals because the signals are transmitted from GNSS satellites and received by a CYGNSS receiver with a considerable distance from the transmitters as described in Subsection 2.1.2. Therefore, the bistatic radar equation describes the relationship between the bi-static received power and scattering properties of the surface in GNSS-R. The correlation between the reflected signal power and soil moisture is mostly due to the reflectivity coefficient. Surface reflectivity coefficient is one of the main data used in GNSS-R.

The total power of the forward scattered L-band signals is a sum of both coherent and incoherent components. However, it has been reported that reflections coming from the land surface are dominated by the coherent component (Nghiem et al. 2017).

The coherent component of the bistatic received power can be written as follows (Aubert et al. 2011; De Roo and Ulaby 1994; Voronovich and Zavorotny 2017):

$$P_{RL}^{coh} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P_t G_t G_r}{(r_{st} + r_{sr})^2} \Gamma_{RL}(\theta_i), \quad (3.1)$$

where  $P_{RL}^{coh}$  is the coherent received power,  $\lambda$  is the wavelength, and  $P_t$  is the peak power of the transmitted GNSS signals.  $G_t$  and  $G_r$  are the antenna gain of the transmitter and receiver, respectively.  $r_{st}$  and  $r_{sr}$  are the distance between specular point to transmitter and receiver, respectively. Finally,  $\Gamma_{RL}(\theta_i)$  is the surface reflectivity at incidence angle of  $\theta_i$ . The subscripts  $R$  and  $L$  denote the right-hand circularly polarized (RHCP) GNSS transmit antenna and the LHCP downward-looking, left-hand circularly polarized GNSS-R antenna, respectively.

The incoherent component of the bistatic received power can be written as follows (Zavorotny and Voronovich 2000):

$$P_{RL}^{inc} = \frac{(\lambda)^2}{(4\pi)^3} \frac{P_t G_t G_r}{r_{st}^2 r_{sr}^2} \sigma_{RL}. \quad (3.2)$$

where  $P_{RL}^{inc}$  is the bistatic received power including diffuse scattering over the surface.  $\sigma_{RL}$  denotes the bi-static radar cross section (BRCS) in  $m^2$ .

Similar to previous studies (Chew and Small 2018; Clarizia et al. 2019; Eroglu et al. 2019), we also assume that majority of the reflections recorded by CYGNSS originate from coherent reflections. With the assumption that surface is relatively flat, smooth, and less vegetation covered, the coherent component of the reflected power is significant, and the incoherent component calculated by (3.2) can be ignored. Then, the surface reflectivity is calculated by Equation (3.1), as follows (Chew and Small 2018; Eroglu et al. 2019):

$$\Gamma_{RL}(\theta_i) = \left(\frac{4\pi}{\lambda}\right)^2 \frac{P_{RL}^{coh}(r_{st} + r_{sr})^2}{P_t G_t G_r}. \quad (3.3)$$

For soil moisture retrieval, the Fresnel reflection coefficient ( $\Re_{RL}(\theta_i)$ ) should be acquired from reflectivity ( $\Gamma_{RL}(\theta_i)$ ), see Equation (3.4). Since near surface soil moisture affects the surface dielectric constant, and surface dielectric constant affects the Fresnel reflection coefficient (Beckmann and Spizzichino 1987).

$$\Gamma_{RL}(\theta_i) = \Re_{RL}(\theta_i)^2 \gamma^2 \exp(-h \cos^2(\theta_i)), \quad (3.4)$$

where  $\gamma$  is the transmissivity parameter related to the wave attenuation due to vegetation canopy, and the h-parameter is related the root-mean-height surface roughness (Choudhury et al. 1979). The transmissivity  $\gamma$ , depends on the vegetation optical depth ( $\tau$ ) and incidence angle ( $\theta_i$ ) through following equation (Eroglu et al. 2019):

$$\gamma = \exp(-\tau \sec(\theta_i)). \quad (3.5)$$

The vegetation optical depth can be calculated from vegetation water content (VWC) and land cover-based proportionality value (b) as Equation (3.6) (Chan et al. 2013).

$$\tau = b \times VWC \quad (3.6)$$

The VWC values can be obtained from normalized difference vegetation index (NDVI) and stem factor parameter as Equation (3.7). NDVI values come from Moderate



resolution imaging spectroradiometer (MODIS) satellites. Land cover-based proportionality b-parameter and stem factor values are provided in Soil moisture active passive (SMAP) mission look up table (LUT) (Chan et al. 2013).

$$VWC = (1.9134 \times NDVI^2 - 0.3215 \times NDVI) + stem\ factor \times \frac{NDVI_{max} - NDVI_{min}}{1 - NDVI_{min}}. \quad (3.7)$$

Finally, the Fresnel reflection coefficient ( $\Re_{RL}(\theta_i)$ ) is related to the dielectric constant of soil through Equations (3.8) to (3.10). This relationship between soil dielectric constant and Fresnel reflection coefficient explains the correlation between soil moisture content and reflected signal power. Equation (3.8) presents that the Fresnel reflection coefficient of circularly polarized waves is expressed as a combination of linear polarized horizontal ( $\Re_{HH}(\theta_i)$ ) and vertical ( $\Re_{VV}(\theta_i)$ ) Fresnel reflection coefficients.

$$\Re_{RL}(\theta_i) = \frac{1}{2}(\Re_{VV}(\theta_i) - \Re_{HH}(\theta_i)). \quad (3.8)$$

$$\Re_{HH}(\theta_i) = \frac{\cos\theta_i - \sqrt{\epsilon_r - \sin^2\theta_i}}{\cos\theta_i + \sqrt{\epsilon_r - \sin^2\theta_i}}. \quad (3.9)$$

$$\Re_{VV}(\theta_i) = \frac{\epsilon_r \cos\theta_i - \sqrt{\epsilon_r - \sin^2\theta_i}}{\epsilon_r \cos\theta_i + \sqrt{\epsilon_r - \sin^2\theta_i}}. \quad (3.10)$$

Finally, to derive soil moisture values from dielectric constant ( $\epsilon_r$ ) soil dielectric models should be used. The soil dielectric constant models define the relationship between the volumetric water content and the dielectric constant by consideration of soil properties

such as temperature, soil bulk density, clay, and sand percentage of soil. A complete explanation of soil dielectric constant models can be found in the next section.

### **3.2 Soil moisture – dielectric constant models**

Soil dielectric constant or permittivity consists of a real part and an imaginary part. The real part of permittivity represents the ability of the material to polarize in response to the electric field compared with the free space, whereas the imaginary part is the ability of the medium to absorb the wave (Ulaby et al. 1982). For L-band frequencies in GPS, the imaginary component of the soil dielectric constant is negligible (Ulaby et al. 1986); therefore, only the real part is considered in soil moisture studies using reflected GPS signals (Katzberg et al. 2006).

Soil water content is particularly related to soil dielectric constant. The dependence of the soil moisture on the dielectric constant increases as frequency decreases. Accordingly, for lower frequencies higher connection between soil moisture and dielectric constant has been seen (Ulaby et al. 1982; Ulaby et al. 1986).

At L-band frequencies, the difference between the dielectric constant of dry soil (~4) and pure water (~80) is very large (Ulaby et al. 1986). Hence, soil moisture and dielectric constant of soil are highly connected working with L-band instruments.

This large contrast between the dielectric constants of dry and moist soil in L-band has resulted in various soil moisture studies using microwave measurements (Schmugge, et al. 1986; Kerr 2007; De Jeu 2008). Dielectric constant is not only affected by moisture content in soil, but also by soil properties such as soil temperature, salinity, soil density, clay and sand percentage. For this reason, to understand the relationship between soil

moisture and soil dielectric constant, the effect of soil properties should be considered in modelling.

Three approaches can be used to compute soil moisture from dielectric constant: i) Empirical models such as the Hallikainen model (Hallikainen et al. 1985), and Wang–Schmugge model (Wang and Schmugge 1980), ii) Semi-empirical models like the Dobson (Dobson et al. 1985), and Mironov model (Mironov and Sergey 2009; Mironov et al. 2009) and iii) Complex analytical physical models (Shi et al. 2002; Lawrence et al. 2013). SMOS and SMAP missions used semi-empirical and empirical models since complex analytical models have not been manageable to a global scale (Kerr et al. 2012).

For passive L-band frequency soil moisture retrieval, dielectric models given by Mironov (Mironov and Sergey 2009; Mironov et al. 2009), Dobson (Dobson et al. 1985), Wang and Schmugge (Wang and Schmugge 1980), and Hallikainen (Hallikainen et al. 1985) are reported to have the best performances due to their easy implementation and requirement of relatively fewer input parameters for calculation of various soil properties such as bulk density and sand and clay fractions compared to other models.

First, Dobson model (Dobson et al. 1985) was chosen in soil moisture retrieval in SMOS as the best available option at that time. Mironov model has been proposed as the new semi-empirical model (Mironov and Sergey 2009; Mironov et al. 2009) afterward showing better performance verified by in-situ L-band measurements (Escorihuela et al. 2007; Wigneron et al. 2010; Wigneron et al. 2012). Mironov was developed on a wider range of data while Dobson was developed on five soil type datasets with sand ratios of less than 50%. Dobson and Mironov dielectric models have been implemented in the SMOS level-2 soil moisture processor version V6 (Kerr et al. 2012).

Dobson, Mironov and Wang models have been also widely used in SMAP mission due to their simple parameterizations and applicability at L-band frequencies (1.26-1.41 GHz). As part of SMAP calibration/validation, the performance of these dielectric models in soil moisture retrieval soil moisture was evaluated for production of SMAP data products (O'Neill et al. 2021). The SMAP L2\_SM\_P processing software has a switch that selects the dielectric model in the soil moisture retrieval. While all three dielectric models are coded and available in the SMAP software, L2\_SM\_P currently uses the Mironov model. For comparison, SMOS mission currently uses land cover classification to choose the appropriate dielectric model between Dobson and Mironov. In this research, empirical models (Hallikainen and Wang) and semi-empirical models (Dobson and Mironov) are used since they have been used widely in SMAP and SMOS soil moisture retrieval algorithms.

### **3.2.1 Empirical models**

The most popular empirical models are the Wang and Schmugge (Wang and Schmugge 1980) and Hallikainen (Hallikainen et al. 1985) models.

The Hallikainen model is an empirical model based on frequency and soil texture. The Hallikainen model was developed from experimental data of soil permittivity for five different soil types at frequencies between 1.4 to 18 GHz frequency regions (Hallikainen et al. 1985). Table 3.1 includes soil texture data for soil experimental data used in the Hallikainen model development.

Hallikainen et al (1985) have shown that dielectric constant of a soil-water mixture is a function of its volumetric moisture content ( $m_v$ ) and of the soils texture compositions such

as clay percentage ( $C$ ) and sand percentage ( $S$ ). Equation (3.11) is the general polynomial function expressing the relationship between soil moisture ( $m_v$ ), relative complex dielectric constant ( $\varepsilon$ ) and sand ( $S$ ) and clay textural ( $C$ ) components.

$$\varepsilon = (a_0 + a_1S + a_2C) + (b_0 + b_1S + b_2C)m_v + (c_0 + c_1S + c_2C)m_v^2 \quad (3.11)$$

Coefficients of polynomial expressions ( $a_0, a_1, a_2, \dots$ ), ( $b_0, b_1, b_2, \dots$ ), ( $c_0, c_1, c_2, \dots$ ) were generated by Hallikainen measurements for each soil type and frequency parameters. Experimental values for these coefficients for different frequencies can be found in (Hallikainen et al. 1985).

Table 3.1. Experimental data used in Hallikainen model development (Hallikainen et al. 1985).

<b>Designation</b>	<b>Sand %</b>	<b>Silt %</b>	<b>Clay %</b>
<b>Field 1</b>	51.51	35.06	13.43
<b>Field 2</b>	41.96	49.51	8.53
<b>Field 3</b>	30.63	55.89	13.48
<b>Field 4</b>	17.16	63.84	19.00
<b>Field 5</b>	5.02	47.60	47.38

Moist soil is a mixture of soil particles, air voids, and liquid water. The water in soil is divided into two fractions in hydrology (Baver et al. 1972): 1) bound water and 2) free water (see Figure 3.1). Bound water is an extremely thin layer of water surrounding soil particles tightly held by the soil particles due to influence of matric and osmotic forces (see Figure 3.1). Bound water includes clay-bound water and capillary-bound water. While bound water is held by electro-molecular and molecular forces, free water is only related to gravitational forces and can be evaporated without external energy.

The Wang model considers the impacts of both bound water and free water. However, Wang model determines the effect of bound water on soil dielectric constant by a linear combination of free water and ice, without consideration of the substantial characteristics of bound water (Jin et al. 2017). Wang and Schmugge (1980) suggested a value for bound water dielectric constant between free water and ice which is calculated by a linear combination of both.

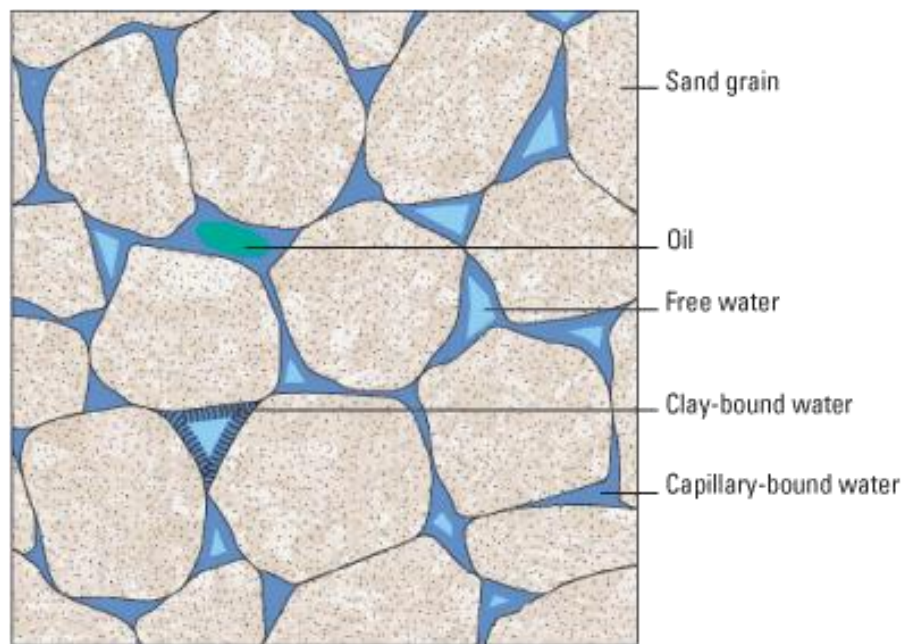


Figure 3.1. Soil as a mixture of sand grain, bund and free water and oil ([https://glossary.slb.com/en/terms/c/clay-bound\\_water](https://glossary.slb.com/en/terms/c/clay-bound_water))

Wang and Schmugge (1980) also used soil wilting point and its relationship to soil texture in their modeling. Wilting point is defined as the minimum amount of water in the soil that the plant requires not to wilt. The wilting point ( $WP$ ) of soils in terms of volumetric water content ( $\text{cm}^3/\text{cm}^3$ ) was derived from a multiple regression of over 100 soil humidity data sets by (Schmugge et al. 1974) as follows,

$$WP = 0.06774 - 0.00064 \times S + 0.00478 \times C \quad (3.12)$$

where C and S are the clay and sand contents in percent of dry weight of a soil.

Wang introduced a parameter called transition moisture ( $W_t$ ) to divide data to two categories since changes of dielectric constant due to volumetric water content ( $W_c$ ) is not always with same pace. The first region occurs at  $W_c \leq W_t$ , where the real part of the dielectric constant increases slowly with volumetric water content ( $W_c$ ). In the second region at  $W_c > W_t$ , the real part of the dielectric constant increases steeply with  $W_c$ . Transition moisture ( $W_t$ ) is not a constant value and depending on soil texture is larger for clay soils versus sandy soils (Newton 1977; Wang et al. 1978). A linear regression was found for calculation of transition moisture ( $W_t$ ) using soil data used in Wang model (Wang and Schmugge 1980) as follows,

$$W_t = 0.49WP + 0.165 \quad (3.13)$$

The expressions for the dielectric constants of a soil-water mixture for these two different regions are given by (Wang and Schmugge 1980)

$$\varepsilon = W_c \varepsilon_x + (P - W_c) \varepsilon_a + (1 - P) \varepsilon_r, \quad W_c \leq W_t \quad (3.14)$$

with

$$\varepsilon_x = \varepsilon_i + (\varepsilon_w - \varepsilon_i) \frac{W_c}{W_t} \cdot \gamma \quad (3.15)$$

and

$$\varepsilon = W_t \varepsilon_x + (W_c - W_t) \varepsilon_w + (P - W_c) \varepsilon_a + (1 - P) \varepsilon_r, \quad W_c > W_t \quad (3.16)$$

with

$$\varepsilon_x = \varepsilon_i + (\varepsilon_w - \varepsilon_i) \gamma. \quad (3.17)$$

Here  $P$  is the porosity of the dry soil and is assumed to be  $2.65 \text{ g/cm}^3$  for simplification as explained in (Wang and Schmugge 1980).  $\varepsilon_a$ ,  $\varepsilon_w$ ,  $\varepsilon_r$ , and  $\varepsilon_i$  are the dielectric constants of air, water, rock, and ice, respectively and are equal to 1, 80, 5.5, and 3.2 in sequential order.  $\varepsilon_x$  is the dielectric constant of the initially absorbed water.  $\gamma$  is a parameter which can be chosen to best fit Equations (3.15) and (3.17) to experimental data. Wang and Schmugge (1980) applied a linear regression and suggested following equation for calculation of  $\gamma$  (see Equation 3.18)

$$\gamma = -0.57 WP + 0.481. \quad (3.18)$$

### 3.2.2 Semi-empirical

Dobson and Mironov are the semi-empirical models used in L-band satellite missions. The Dobson model is a semi-empirical mixed dielectric constant model based on the soil permittivity experimental data at conditions of five soil types, and in frequency ranges of 0.3-1.3 GHz and 1.4-18 GHz, and temperature of  $22^\circ\text{C}$  (Dobson et al. 1985; Peplinski et al. 1995). The Soil experimental data used in Dobson model development is the same as



Hallikainen model, (see Table 3.1). The model requires frequency, soil moisture, soil temperature, sand fraction, clay fraction, and bulk density as input parameters. The Dobson model requires the following input data: the soil moisture ( $\text{m}^3/\text{m}^3$ ), the soil effective temperature, the clay percentage and the sand percentage content, the dry soil bulk density ( $\text{g}/\text{cm}^3$ ), and the solid particle density.

Table 3.2. Experimental data used in Wang model development (Wang and Schmutge 1980).

<b>Designation</b>	<b>Sand %</b>	<b>Silt %</b>	<b>Clay %</b>
<b>Field 1</b>	88.0	7.3	4.7
<b>Field 2</b>	56.0	26.7	17.3
<b>Field 3</b>	19.3	46.0	34.7
<b>Field 4</b>	2.0	37.0	61.0
<b>Field 5</b>	100.0	0	0
<b>Field 6</b>	90.0	7.0	3.0
<b>Field 7</b>	82.0	14.0	4.0
<b>Field 8</b>	70.0	22.0	8.0
<b>Field 9</b>	22.0	70.0	8.0
<b>Field 10</b>	58.0	28.0	14.0
<b>Field 11</b>	48.0	36.0	16.0
<b>Field 12</b>	45.0	39.0	16.0
<b>Field 13</b>	26.0	56.0	18.0
<b>Field 14</b>	22.0	56.0	22.0
<b>Field 15</b>	16.0	56.0	28.0
<b>Field 16</b>	6.0	54.0	40.0
<b>Field 17</b>	86.0	7.0	7.0
<b>Field 18</b>	40.0	26.0	34.0
<b>Field 19</b>	36.0	29.0	35.0
<b>Field 20</b>	52.0	9.0	39.0
<b>Field 21</b>	44.0	12.0	44.0
<b>Field 22</b>	3.0	35.0	62.0

Dobson considers that the soil dielectric constant is composed of four parts: soil solid particle dielectric constant, soil air dielectric constant, soil free water and soil combined water permittivity. However, since it is difficult to get the dielectric constant of the bound

water, the model combines the combined water and free water contribution into a free water contribution parameter by introducing an empirical coefficient.

Dobson model also known as semiempirical mixing dielectric model (SMDM) has the following form (Dobson et al. 1985; Mironov et al. 2009):

$$\varepsilon'_m = \left[ 1 + \frac{\rho_b}{\rho_s} (\varepsilon_s'^\alpha - 1) + m_v^{\beta'} \varepsilon_{fw}'^\alpha - m_v \right]^{1/\alpha} \quad (3.19)$$

$$\varepsilon''_m = \left[ m_v^{\beta''} \varepsilon_{fw}''^\alpha \right]^{1/\alpha} \quad (3.20)$$

where  $\varepsilon'_m$  and  $\varepsilon''_m$  are the moist soil dielectric constant (DC) and loss factor (LF), respectively,  $\varepsilon'_s$  is a composite DC of the soil mineral contents,  $m_v$  is the volumetric moisture content,  $\rho_b$  is the bulk density in grams per cubic centimetre,  $\rho_s$  is the specific gravity of the solid soil particles,  $\alpha = 0.65$  is an empirically determined constant, and  $\beta'$  and  $\beta''$  are the empirically determined soil-type-dependent constants given by

$$\beta' = 1.2748 - 0.00519 S - 0.00152 C \quad (3.21)$$

$$\beta'' = 1.33797 - 0.00603 S - 0.00166 C \quad (3.22)$$

where  $S$  and  $C$  are sand and clay percentages of soil, respectively.

The relationship between the specific gravity,  $\rho_s$ , and DC of soil mineral contents,  $\varepsilon'_s$ , is represented as equation (3.23), and the relationship between the soil bulk density,

$\rho_b$ , and dry-soil dielectric constant,  $\varepsilon'_{m0}$ , is expressed as Equation (3.24) as follow (Mironov et al. 2009),

$$\varepsilon'_s = (1.01 + 0.44\rho_s)^2 - 0.062 \quad (3.23)$$

$$\rho_b = [(\varepsilon'_{m0})^\alpha - 1]\rho_s/(\varepsilon'^\alpha_s - 1) \quad (3.24)$$

The quantities  $\varepsilon'_{fw}$  and  $\varepsilon''_{fw}$  are the DC and LF of free water, respectively, given by the Debye equations, with the latter being modified to include a term that accounts for the effective conductivity of the free-soil water (FSW)

$$\varepsilon'_{fw} = \varepsilon_{\omega\infty} + \frac{\varepsilon_{\omega0} - \varepsilon_{\omega\infty}}{1 + (2\pi f \tau_w)^2} \quad (3.25)$$

$$\varepsilon''_{fw} = \frac{2\pi f \tau_w (\varepsilon_{\omega0} - \varepsilon_{\omega\infty})}{1 + (2\pi f \tau_w)^2} + \frac{\sigma_{eff}}{2\pi \varepsilon_0 f} \frac{(\rho_s - \rho_b)}{\rho_s m_v} \quad (3.26)$$

Where  $\varepsilon_0$  is dielectric constant of free space and equal to  $8.854 \times 10^{-12} F/m$ ,  $\tau_w$  is the relaxation time for free water,  $f$  is the frequency in hertz,  $\varepsilon_{\omega0}$  is the low-frequency limit of dielectric constant for water,  $\varepsilon_{\omega\infty} = 4.9$  is the high-frequency limit of  $\varepsilon'_{fw}$ . Relaxation Time ( $\tau_w$ ) is the time it takes for a system to return to equilibrium after a change or disturbance. Expressions for  $\tau_w$  and  $\varepsilon_{\omega0}$  are given as functions of temperature by Ulaby et al. (1986). At room temperature (20°C),  $2\pi\tau_w = 0.58 \times 10^{-10} s$  and  $\varepsilon_{\omega0} = 80.1$ . The effective conductivity  $\sigma_{eff}$  was determined in the following form,

$$\sigma_{eff} = 0.0467 + 0.2204\rho_b - 0.004111S - 0.00614C . \quad (3.27)$$

All parameters which were empirically determined through regressions analysis to field data are  $\sigma_{eff}$ ,  $\varepsilon'_s$ ,  $\alpha$ ,  $\beta'$  and  $\beta'$  (Mironov et al. 2009). Then, the dielectric constant and loss factor of free water, are calculated with the analytical Debye relaxation formulas (See Equations 3.25 and 3.26). The Mironov model is a physical semi-empirical mixed dielectric constant model based on experimental data of wide frequency soil permittivity of various soil types, a wider range with comparison to Dobson, which has been improved several times (Mironov et al. 2004; Mironov et al. 2008; Mironov et al. 2009; Mironov and Fomin 2009).

The Mironov model discriminates the dielectric properties of the bound water and free water in the soil. The simulated soil dielectric constant in Mironov model is composed of four contribution parts of the soil solid particle, the soil air, the free water, and the combined water. Mironov was developed from a wider range of soil types and requires fewer input parameters in comparison with Dobson model. The Mironov model, also known as the Generalized Refractive Mixing Dielectric Model (GRMDM), uses the total mixed complex refractive index by calculating the weighted refractive index of these different components, and translating into the dielectric of the soil constant (Mironov et al. 2012). The Mironov model applies to a wider range of soil types, but also requires fewer input parameters – with clay percentage as the only soil input parameter.

Mironov and Sergey (2009) proposed the physically based generalized refraction mixing dielectric model, to consider the effect of the soil temperature allowing model to

distinguish between bound and free water (Mironov et al. 2012). The Mironov model uses soil moisture, soil effective temperature, and clay as input, thus avoiding uncertainties introduced by the computation of the global bulk density map compared with the Dobson model. Mironov was developed and validated on entire samples and texture ranges of soil including samples used in Dobson. Generally, the Mironov model tends to retrieve higher soil moisture values than the Dobson model.

In the Mironov model, the dielectric constant as functions of soil moisture ( $m_v$ ) is expressed as follow (Mironov et al. 2004),

$$\varepsilon'_m = n_m^2 - k_m^2, \quad \varepsilon''_m = 2n_mk_m \quad (3.28)$$

$$n_m = \begin{cases} n_d + (n_b - 1)m_v, & m_v \leq m_{vt} \\ n_d + (n_b - 1)m_{vt} + (n_u - 1)(m_v - m_{vt}), & m_v \geq m_{vt} \end{cases} \quad (3.29)$$

$$k_m = \begin{cases} k_d + (k_b)m_v, & m_v \leq m_{vt} \\ k_d + (k_b)m_{vt} + (k_u)(m_v - m_{vt}), & m_v \geq m_{vt} \end{cases} \quad (3.30)$$

where  $n_m, n_d, n_b, n_u$ , and  $k_m, k_d, k_b, k_u$  are the refractive index and normalized attenuation coefficients. The subscripts  $m, d, b$ , and  $u$  in Equations (3.28-3.30) stand for moist, dry soil, bound soil water (BSW) and free soil water (FSW), respectively. Using both BSW and FSW coefficients are representative of the fact that the Mironov model differentiates between bound and free water in the modeling unlike the Dobson model. Parameter  $m_{vt}$  is a value of the maximum bound water fraction (MBWF) that depends on the soil mineral contents (Mironov 2004). The values of refractive indexes and normalized attenuation

coefficients can be computed using their dielectric constant ( $\varepsilon'$ ) and loss factor ( $\varepsilon''$ ) through following Equations (3.31-3.32)

$$n_{d,b,u}\sqrt{2} = \sqrt{\sqrt{(\varepsilon'_{d,b,u})^2 + (\varepsilon''_{d,b,u})^2} + \varepsilon'_{d,b,u}} \quad (3.31)$$

$$k_{d,b,u}\sqrt{2} = \sqrt{\sqrt{(\varepsilon'_{d,b,u})^2 + (\varepsilon''_{d,b,u})^2} - \varepsilon'_{d,b,u}} \quad (3.32)$$

The dielectric constant and the loss factor for bound and free water components are as follows with the Debye relaxation equations (Mironov et al. 2009)

$$\varepsilon'_{b,u} = \varepsilon_{\infty} + \frac{\varepsilon_{0b,0u} - \varepsilon_{\infty}}{1 + (2\pi f \tau_{b,u})^2}, \quad (3.33)$$

$$\varepsilon''_{b,u} = \frac{\varepsilon_{0b,0u} - \varepsilon_{\infty}}{1 + (2\pi f \tau_{b,u})^2} 2\pi f \tau_{b,u} + \frac{\sigma_{b,u}}{2\pi f \varepsilon_0} \quad (3.34)$$

where  $f$  is the wave frequency,  $\sigma_{b,u}$ ,  $\tau_{b,u}$ , and  $\varepsilon_{0b,0u}$  are the conductivities, relaxation times and low-frequency limit of dielectric constants (BSW or FSW components), respectively. The value  $\varepsilon_0$  is the dielectric constant for free space, while  $\varepsilon_{\infty}$  represents the dielectric constant in the high-frequency limit equal to 4.9 for both bound and free soil water.

Data used in Mironov model development were measured in the temperature range from 20°C to 22°C with clay contents varying from close to 0% to 54%. Therefore,

spectroscopic parameters were correlated to clay percentages and correlation analysis resulted following regression equations (Mironov et al. 2008)

$$n_d = 1.634 - 0.539 \times 10^{-2}C + 0.2748 \times 10^{-4}C^2 \quad (3.35)$$

$$k_d = 0.03952 - 0.04038 \times 10^{-2}C \quad (3.36)$$

$$\sigma_u = 0.3631 + 1.217 \times 10^{-2}C \quad (3.37)$$

$$\sigma_b = 0.3112 + 0.467 \times 10^{-2}C \quad (3.38)$$

$$\tau_u = 8.5 \times 10^{-2}C \quad (3.39)$$

$$\tau_b = 1.062 \times 10^{-11} + 3.450 \times 10^{-12} \times 10^{-2}C \quad (3.40)$$

$$\varepsilon_{0,b} = 100 \quad (3.41)$$

$$\varepsilon_{0,b} = 79.8 - 85.4 \times 10^{-2}C + 32.7 \times 10^{-4}C^2 \quad (3.42)$$

Finally, for each type of soil all these parameters can be derived using these equations. However, although the Mironov and Dobson models have fewer empirical parameters, still they suffer from empirical model errors.

## **4. PHYSICS-INFORMED NEURAL NETWORK STRUCTURE**

In this research, a physics-informed neural network (PINN) algorithm is discussed to model the relationship between soil moisture and reflectivity. The problem of GNSS-R soil moisture retrieval can be improved by PINN as a new emerging technique. In PINN, the objective is to minimize the loss function or mean squared error of both NN regression and physics equations. PINN methodology is a multi-objective learning framework in which a NN must fit observed data while reducing physics equation loss functions.

### **4.1 PINN applications**

Physics-informed neural networks (PINNs) have already been used for a variety of applications in the field of computational physics and engineering, including i) solving partial differential equations (PDEs) in fluid dynamics (Cai et al. 2021) and heat transfer (Zobeiry and Humfeld 2021), ii) inverse problems to determine the underlying physical parameters or variables connecting a set of observations in fields such as image processing (Wu et al. 2022), tomography (Guo et al. 2023; Saba et al. 2022) and remote sensing (Chen and Negro 2022), iii) data-driven models of physical systems that can be useful for predicting and generating new insights to existing physics model, for example in weather forecasting and climate modelling (Kashinath et al. 2021; Chattopadhyay et al. 2022), and iv) uncertainty quantifications associated with predictions made by physical models to find the potential source of error or uncertainty in an engineering system (Daw et al. 2020). The main advantage of PINN is the convergence of results with less data and more understanding of physics.



To implement PINNs efficiently, new algorithms can be built with current machine learning libraries, such as TensorFlow<sup>150</sup>, PyTorch<sup>151</sup>, Keras<sup>152</sup>, and JAX<sup>153</sup>. Several software libraries have been specifically designed for physics-informed machine learning. Table 4.1 includes information on software libraries developed for physics-informed machine learning. At the present time, some of the actively developed libraries include SciANN<sup>159</sup> (Haghighat and Juanes 2021), DeepXDE<sup>154</sup> (Lu et al. 2021), SimNet<sup>155</sup> (Hennigh et al. 2021), PyDEns<sup>156</sup> (Koryagin et al. 2019), NeuroDiffEq<sup>157</sup> (Chen et al. 2020), NeuralPDE<sup>158</sup> (Rackauckas and Nie 2017) and ADCME<sup>160</sup> (Xu and Darve 2020). Because Python is the dominant programming language for ML, it is more convenient to use Python for physics-informed ML, and thus most of these libraries are written in Python, except NeuralPDE<sup>158</sup> and ADCME<sup>160</sup>, which are written in Julia.

Table 4.1. Software libraries specifically designed for physics-informed machine learning.

Software name	Language	Backend	Reference
<b>SciANN</b>	Python	TensorFlow	(Haghighat and Juanes 2021)
<b>DeepXDE</b>	Python	TensorFlow	(Lu et al. 2021)
<b>SimNet</b>	Python	TensorFlow	(Hennigh et al. 2021)
<b>PyDEns</b>	Python	TensorFlow	(Koryagin, Khudorozkov and Tsimfer 2019)
<b>NeuroDiffEq</b>	Python	PyTorch	(Chen et al. 2020)
<b>GPyTorch</b>	Python	PyTorch	(Gardner et al. 2018)
<b>NeuralPDE</b>	Julia	Julia	(Rackauckas and Nie 2017)
<b>ADCME</b>	Julia	TensorFlow	(Xu and Darve 2020)

## 4.2 PINN design for GNSS-R

In PINN, the objective is to minimize the loss function, which is the mean squared error of both NN regression and mathematical physics models at the same time. Therefore,

PINN structure includes both NN regression and physics-based models. In this research, the input layer includes CYGNSS data (reflectivity, incidence angle), vegetation data (NDVI, VWC), surface roughness data (h-parameter), soil temperature, soil texture data (clay, sand percentage and bulk density) and land cover related data (b-parameter and stem factor) (see Figures 4.1 and 4.2). All these geophysical parameters and their relationships with soil moisture and reflectivity are explained in Section 3.1. The input matrix has a dimension of  $11 \times S$ , when 11 is the number of inputs and S is the number of data samples. In a fully connected neural network, all neurons in one layer are connected to all neurons in the next layer. Each layer has a weigh array that is updated by forward and backward propagation. Updating of the weight array for each layer in NN is referred to as training or learning. In other words, NN learns the non-linear relationship between input and output parameters by updating these weigh-arrays in training. The result of weight multiplication is added to a bias vector which also updates in each step of training (Hinton 1990). Weight-arrays and bias are updated so that the loss function ( $L$ ) is minimized. Here, the loss function is the summation of two mean squared errors: First, the mean squared error between the calculated soil moisture from NN model and reference ISMN soil moisture data ( $MSE_{SM}$ ); and secondly, the mean squared error between inversely calculated reflectivity from mathematical equations and reflectivity data from CYGNSS ( $MSE_r$ ) (see Figures 4.1 and 4.2). The minimization of loss function ( $L$ ) is done through a gradient based optimizer. In the PINN model, the Adam solver first-order gradient-based optimizer with learning rate of 0.001 has been used due to its best performance for stochastic objective functions (Kingma and Ba 2014). ISMN soil moisture values were used as output layer for both training and validation.

The inverse reflectivity calculations are embedded in the PINN as a physics model. In each step of PINN training, first soil moisture is calculated through NN regression using a set of input data (X) including GNSS-R and surface properties data, explained in Section 3.1. Second, NN calculated soil moisture ( $SM_{NN}$ ) is used to compute reflectivity ( $\Gamma_{SM}$ ) inversely using Equations 3.10 to 3.4, described in Section 3.1. Since initial soil moisture data is available, Equations 3.4 to 3.10 are included in PINN model inversely to calculate reflectivity. The PINN goal is to minimize NN regression and physics model loss function simultaneously. NN regression or soil moisture loss function ( $MSE_{SM}$ ) is the difference between calculated soil moisture from NN ( $SM_{NN}$ ) and soil moisture data ( $SM_{data}$ ) (see Figure 4.2). Physics model loss function ( $MSE_{\Gamma}$ ) is the difference between inversely calculated reflectivity ( $\Gamma_{SM}$ ) and CYGNSS reflectivity data ( $\Gamma_{CYGNSS}$ ). Figure 4.1 demonstrates the process of PINN loss function calculations for N number of samples. The total loss function ( $MSE$ ) is the summation of both soil moisture loss function ( $MSE_{SM}$ ) and reflectivity loss function ( $MSE_{\Gamma}$ ) (see Figure 4.1).

In the inverse calculation of reflectivity, first dielectric constant ( $\epsilon$ ) is obtained from NN computed soil moisture ( $SM_{NN}$ ) using the soil dielectric constant model (see Figure 4.2). Next, reflectivity ( $\Gamma_{SM}$ ) will be calculated inversely from dielectric constant ( $\epsilon$ ) using Equations (3.10 to 3.4), while reflectivity data ( $\Gamma_{CYGNSS}$ ) comes from CYGNSS data. Equations 3.4 to 3.10 are embedded in the PINN model inversely since these Equations are not directly mathematically solvable due to non-linearity. Inverse calculated reflectivity ( $\Gamma_{SM}$ ) is then compared with reflectivity data ( $\Gamma_{CYGNSS}$ ) from CYGNSS to compute reflectivity mean squared error. Reflectivity mean squared error ( $MSE_{\Gamma}$ ) should be minimized in each step of training, as well as mean squared error of NN regression for

soil moisture ( $MSE_{SM}$ ). Therefore, as soil moisture values are updated in each step of NN training, the inverse calculated reflectivity changes as well. Figure 4.2 summarizes the PINN processing methodology and computations with all soil surface parameters and GNSS-R data. PINN structure allows the inclusion of physical models in the training process. Here, inverse calculation of reflectivity is added to NN in the PINN modelling as a new methodology to retrieve soil moisture using GNSS-R signals.

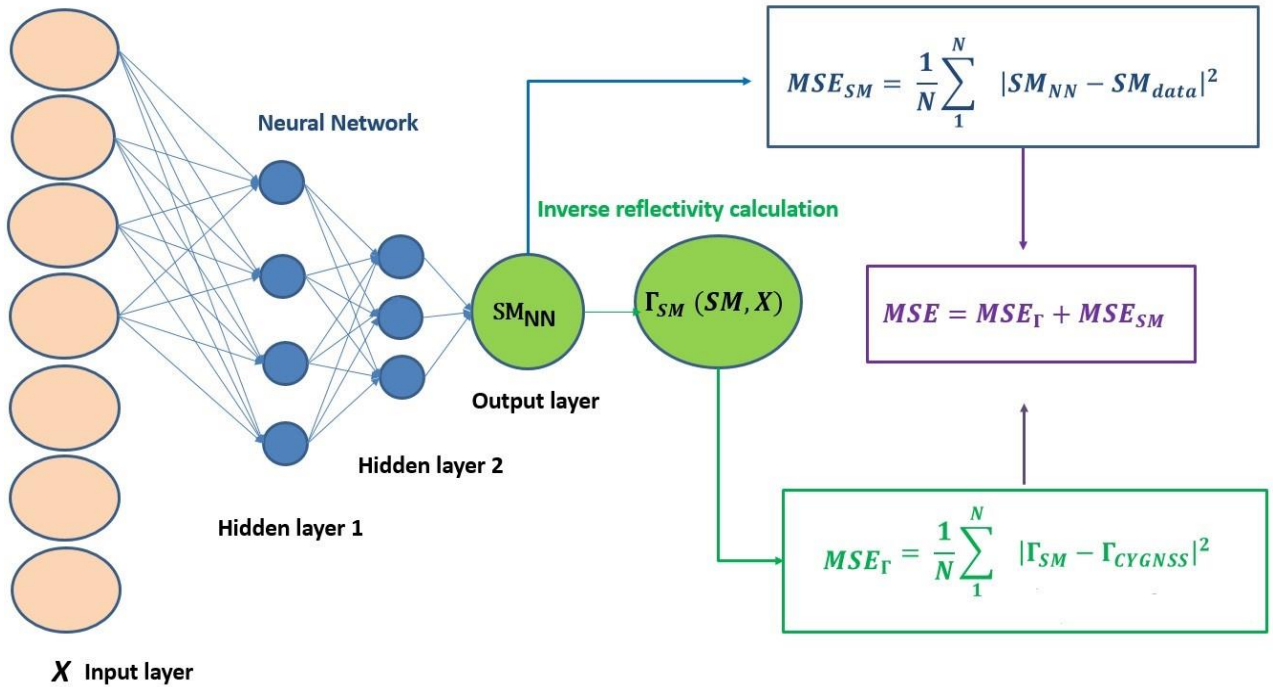


Figure 4.1. Physics-informed neural network loss function calculations. PINN aim is to minimize two loss functions: i)  $MSE_{SM}$  as NN regression loss function which is the difference between NN calculated soil moisture ( $SM_{NN}$ ) and soil moisture data ( $SM_{data}$ ), and ii)  $MSE_{\Gamma}$  as physics model loss function which is the difference between inversely calculated reflectivity ( $\Gamma_{SM}$ ) from  $SM_{NN}$  and reflectivity data from CYGNSS ( $\Gamma_{CYGNSS}$ ).

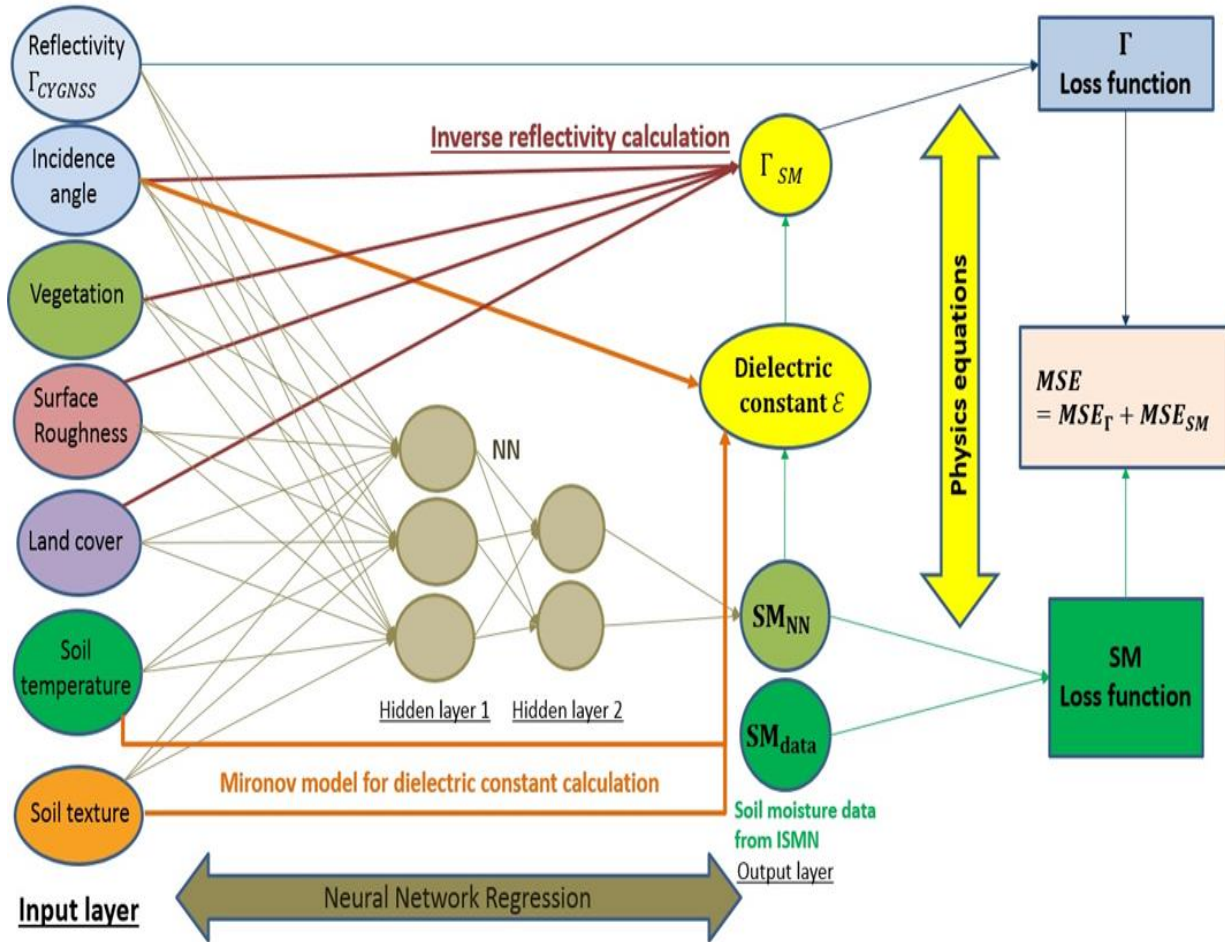


Figure 4.2. Physics-informed Neural Network (PINN) structure for CYGNSS soil moisture estimation. PINN consists of two parts: a NN regression and mathematical physics models. In PINN, the aim is to minimize two loss functions related to NN (SM loss function,  $MSE_{SM}$ ) and physics ( $\Gamma$  loss function,  $MSE_{\Gamma}$ ), together. Then, the summation of both mean squared errors is minimized in each step of training.

Neural network performance error is highly dependent on optimal hyper-parameters. Hyper-parameters are parameters defining the network structure, for example number of hidden layers and neurons, variables determining how the network is trained like learning rate, batch size, epoch, activation function and optimization algorithms. Hyper-parameters are set before training and updating the weights and bias. Choice of the best hyper-parameter requires expansive trial and error rather than science practices. The

process of finding a suitable set of hyper-parameters can be time consuming and ultimately not necessarily effective (Smith 2018). However, research has suggested practical ways to set hyper-parameters (Bengio 2012).

**Hyper-parameters in both NN and PINN with and without physics are as follows:**

- Learning rate: Learning rate is the most important hyper-parameter in NN. As suggested by (Bengio 2012; Goodfellow et al. 2016, 429-430) if one has time to optimize only one hyper-parameter, learning rate should be tuned as it is the most substantial hyper-parameter. Learning rate is simply the rate at which a NN learns. Learning rate specifies the step size of each step of the optimization problem which is stochastic gradient descent. Learning rate value determines how fast or slow a NN learns. So that, with a small learning rate, the model will converge slowly, and the computation time is high (see Figure 4.3). With a large learning rate, the descent steps are high, and the model will miss the global minimum (see Figure 4.3). To start with a prior value, Bengio (2012) suggests a default value of 0.01, which works well for most multi-layer neural networks. Learning rate is practically speaking a value between  $10^{-6}$  to 1 (Bengio 2012). The best way is to start with a prior 0.01 learning rate and check which learning rate works best for a specific problem. However, instead of working with a constant value for learning rate, some researchers have recommended varying learning rates. For example, Smith (2017) suggests a method in which the learning rate cyclically varies between reasonable boundary values. Another good practice is to use a decaying learning rate – that is one that starts out large and changes to a small value over time.

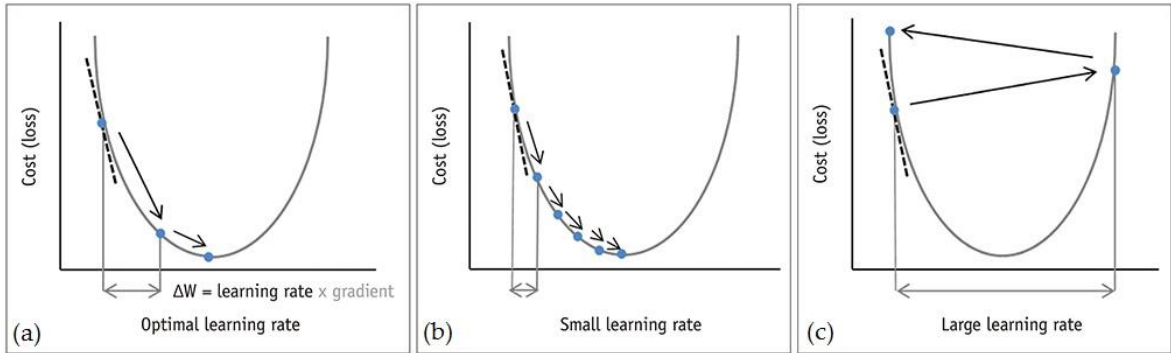


Figure 4.3. Different learning rate occurrences in a weight and cost plot. How different learning rates react in a similar weight and cost situation. a) optimal learning rate. b) small learning rate with slow convergence that need many steps to converge to minimum. c) large learning rate that jumps too much (Do et al. 2020).

- **Epoch and iterations:** One epoch is when the entire dataset is passed forward and backward through propagation of the neural network once and updates weights and biases parameters. Since one epoch is too big to feed to the computer at once it is divided into several smaller batches. The number of iterations is the total number of passes (Figure 4.4). For example, for a dataset of 6 images (6 training samples), three states are considered to understand the concept of epoch, batch size and iterations (Figure 4.4). A) In this case, batch size, epoch, and iteration are 3, 1, and 2, respectively. When batch size is 3 and epoch is 1, only 2 passes (iterations) are required to go through the entire dataset (6 images) once, since epoch is one (Figure 4.4). B) When batch size is 2, and epoch is 1, then 3 passes are required to move through entire datasets once. C) If two epochs are considered, it is required to pass through the entire dataset twice. Therefore, for batch size of 2, 6 passes (3 for each epoch, entire dataset) are needed to go through entire dataset twice. Therefore, number of iterations can be calculated as:

$$iteration = number\ of\ training\ samples \times \left( \frac{epoch}{batch\ size} \right) \quad (4.1)$$

Higher numbers of epochs are used to allow neural networks to go through the process of training. The right number of epochs depends on the complexity and size of your dataset. Best practice is trial and check the maximum epoch that the neural network loss changes significantly.

- **Batch size:** Batch size defines the number of samples divided as batches or groups in one epoch to train a neural network. Batch size usually ranges from 1 to 128. The higher the batch size, the more memory space will be needed. Some authors suggest that when multiplying batch size by  $k$ , the learning rate should also be multiplied by  $\sqrt{k}$  to keep the variance in the gradient expectation constant. More commonly, a simple linear scaling rule is used. For learning rates of 0.0001 and 0.001 as common learning rates used in GNSS-R research (Eroglu et al. 2019; Roberts et al. 2022), batch sizes of 32, 64, and 128 can be used. In this research, batch size of 64 with learning rate of 0.001 was the most efficient combination.
- **Activation functions:** Each neural network has two operations including a linear function and a non-linear function. Activation function helps the network to learn the non-linear properties of the data set. The Rectified Linear Unit (ReLU) is usually the most popular choice for NN hidden layers. ReLU activation function is used in PINN in this research too due its common appearance in most published GNSS-R research (Eroglu et al. 2019; Roberts et al. 2022).



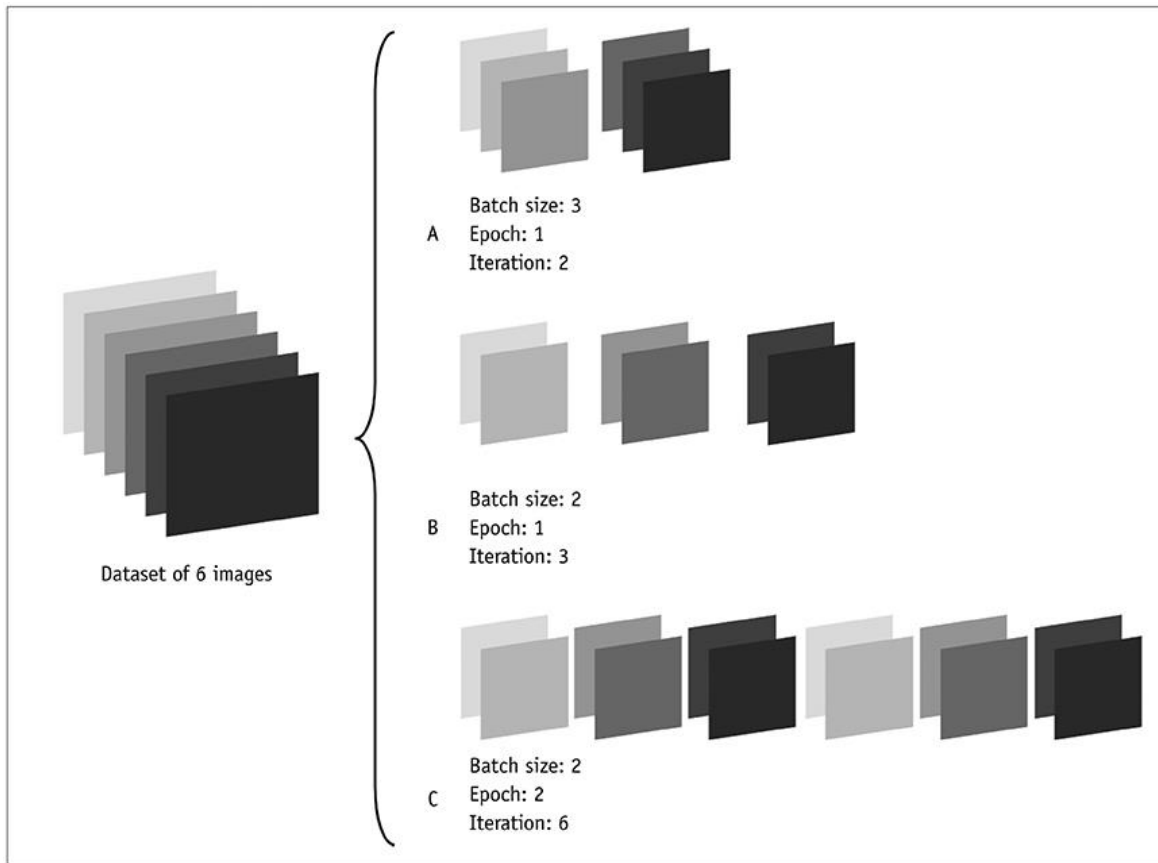


Figure 4.4. Dataset of 6 images for three different states of epoch, batch, and iteration. A) With batch size of 3, 2 iterations are needed for one epoch, B) Batch size is 2 and algorithm is set to 1 epoch, 3 iterations are needed for one epoch to cover all 6 images C) batch size is 2, and algorithm is set to run for 2 epochs. Therefore, in each epoch, there are 3 batches. Each batch gets passed through algorithm, so there are 3 iterations per epoch. Since 2 epochs were specified, there are total of 6 iterations ( $3 \times 2 = 6$ ) for training (Do et al. 2020).

- Number of hidden layers and neurons: To determine the number of neurons in the hidden layers, you can use common rule-of-thumb methods, such as: i) The number of hidden neurons should be between the size of the input layer and the size of the output layer. ii) The number of hidden neurons should be  $\frac{2}{3}$  the size of the input layer, plus the size of the output layer. iii) The number of hidden

neurons should be less than twice the size of the input layer. Number of neurons and layers can be selected with cross validation techniques too. High number of neurons or layers can result in overfitting while low number of neurons and layers can result in underfitting. PINN models usually have higher numbers of neurons and layers in comparison to NN, e.g., number of neurons as large as 10, 20, and 100 (Cai et al. 2021; Haghighat and Juanes 2021). In this research, the minimum number of neurons and layers that guaranteed acceptable results without underfitting were 12 and 20. Therefore, PINN models with the same data as NN models can have higher number of neurons and layers without experiencing overfitting. This matter is explained in more details in Section 6.2.

## 5. DATASETS

The performance of a neural network depends highly on both data quality and quantity. In this research, different datasets have been used to fulfill the physics theoretical dependencies explained in Sections 3.1 and 3.2. For this reason, different datasets such as CYGNSS, ISMN soil moisture and temperature, SMAP soil texture (Clay, sand percentage and bulk density), and MODIS land cover data have been used.

### 5.1 CYGNSS data

The Level 1 version 2.1 of CYGNSS data was used for a time duration of two years, from January 1st of 2020 to the end of 2021. Level 1 version 2.1 of CYGNSS data is available at the Physical Oceanography Distributed Active Archive Center (PODAAC; [https://podaac.jpl.nasa.gov/dataset/CYGNSS\\_L1\\_V2.1](https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_V2.1)). For each day, PODAAC provides a set of 8 NetCDF files—one for each CYGNSS satellite. Since the CYGNSS satellites are capable of recording 4 simultaneous reflections per second, each file includes 4 Delay Doppler Maps (DDMs) of analog scatter power for 1 second long for each day.

CYGNSS observables used in this research are as follows: i) reflectivity, ii) specular incidence angle, iii) specular point latitude and iv) specular point longitude. Reflectivity is the primary CYGNSS observable that must be input to the PINN algorithm to investigate soil moisture changes. The surface reflectivity can be derived from the CYGNSS data from four approaches: (i) reflectivity can be calculated by substituting the DDM SNR ( $ddm\_snr$ ) into  $P_{RL}^{coh}$  in Equation (3.3) and calibrating for the instrumental and geometric

parameters (Mathur 2021). DDM SNR ( $ddm\_snr$ ) is in dB and equal to  $10\log(Smax/Navg)$ , where  $Smax$  is the maximum value (in raw counts) in a single DDM bin and  $Navg$  is the the average per-bin raw noise counts. (ii) reflectivity can also be generated from the peak value of the analog power DDM ( $power\_analog$ ) into  $P_{RL}^{coh}$  in Equation (3.3) instead of DDM SNR (Chew and Small 2018; Clarizia et al. 2019). For cases where error level in the DDM noise floor is high, reflectivity of peak could provide increased correspondence to SM (Eroglu et al. 2019), iii) as explained in (Rodriguez-Alvarez et al. 2019; Eroglu et al. 2019), bistatic radar cross section (BRCS) explained in Section 2.2 can also be used for reflectivity calculations by correcting the incoherency assumption and compensating the path loss and  $4\pi$  term, iv) Reflectivity can also be defined and computed as the ratio of the reflected and direct SNRs ( $ddm\_snr$  and  $direct\_snr$ , respectively), which are first calibrated by the range terms, as previously practiced (Carreno-Luengo et al. 2018).

In this research, reflectivity is calculated from the peak of analog power. Peak of DDM cross-correlation is commonly used for soil moisture applications since the peak value of each DDM is controlled by surface properties such as surface roughness and dielectric constant (Chew and Small 2018; Clarizia et al. 2019). CYGNSS DDM is represented as “ $power\_analog$ ”, which is a 17 x 11 element array of calibrated power (in Watts) from a reflecting surface, where each array value represents the power at a specific time delay and Doppler shift.

Therefore, reflectivity was calculated from Equation (3.3) by substituting the peak of the  $power\_analog$  data from Level 1 CYGNSS products into  $P_{RL}^{coh}$  in Equation (3.3) in Section

3.1. Calculated reflectivity is calibrated for antenna gain and instrumental parameters in Equation (5.1) as explained by (Chew and Small 2018; Clarizia et al. 2019)

$$\Gamma_{RL}(dB) = 10 \log P_{RL}^{coh} - 10 \log N - 10 \log G^r - 10 \log G^t - 10 \log P_r^t + 20 \log(R_{ts} + R_{sr}) \quad (5.1)$$

where  $P_r^t$  is the transmitted RHCP power,  $G^t$  is the gain of the transmitting antenna,  $R_{ts}$  is the distance between the transmitter and the specular reflection point,  $R_{sr}$  is the distance between the specular reflection point and the receiver,  $G^r$  is the gain of the receiving antenna,  $\lambda$  is the GPS wavelength, and  $\Gamma_{RL}$  is the reflectivity of the surface in dB. Incidence angle was also added as an input to model since calculations of Fresnel reflection coefficient and reflectivity are dependent on the incidence angle (see Equations 3.4, 3.8-3.10) in Section 3.1.

Since CYGNSS satellites provide reflections over the ocean rather than land, the reflections overland was chosen using quality flags. In CYGNSS level-1 data, a unique quality flag directly differentiates land surface observations from other observations. Common data quality control in the land applications is also used (Mathur 2021). The DDM SNR lower than 2dB, receiver antenna gains at the specular point direction lower than 0 dB, and specular incidence angles over  $60^\circ$  are filtered.

In summary, the CYGNSS products needed for calculations here are (see Table 5.1): (i) peak of the *power\_analog* for  $P_{RL}^{coh}$  (ii) *sp\_lat*, *sp\_lon* and *sp\_inc\_angle* for latitude, longitude and incidence angle of specular points, (iii) *gps\_eirp* which is GPS equivalent isotropically radiated power calculated from  $P_t G_t$ , (iv) *rx\_to\_sp\_range* and *tx\_to\_sp\_range* as distances between the receiver and transmitter to specular point, respectively, and (v) *sp\_rx\_gain* as receiver antenna gain.

Table 5.1. CYGNSS Level 1 data used in this study.

<b>CYGNSS data</b>	<b>Description</b>
<i>power_analog</i>	17 × 11 array of DDM bin analog power, Watts. analog_power is the true power that would have been measured by an ideal (analog) power sensor. power_digital is the power measured by the actual 2-bit sensor, which includes quantization effects. power_analog has been corrected for quantization effects.
<i>sp_lat</i>	Specular point latitude, in degrees North
<i>sp_lon</i>	Specular point longitude, in degrees East.
<i>sp_inc_angle</i>	Specular point incidence angle, in degrees. This is the angle between the line normal to the Earth's surface at the specular point and the line extending from the specular point to the spacecraft.
<i>rx_to_sp_range</i>	The distance between the CYGNSS spacecraft and the specular point, in meters, at ddm_timestamp_utc.
<i>tx_to_sp_range</i>	The distance between the GNSS spacecraft and the specular point, in meters, at ddm_timestamp_utc.
<i>sp_rx_gain</i>	Specular point Rx antenna gain. The receive antenna gain in the direction of the specular point, in dBi.
<i>gps_tx_power_db_w</i>	GPS transmit power. Power input to transmitter antenna, in dBw.
<i>gps_ant_gain_db_i</i>	GPS transmit antenna gain. Antenna gain in the direction of the specular point, in dBi.
<i>ddm_noise_floor</i>	For non-black-body DDMs: Is equal to the average bin raw counts in the first 45 delay rows of the uncompressed 20 x 128 DDM, in counts, at ddm_timestamp_utc.  For black body DDMs: Is equal to the average bin raw counts in all 128 delay rows of the uncompressed 20 x 128 DDM, in counts, at ddm_timestamp_utc.

## 5.2 Soil moisture data

To provide global access to in situ soil moisture measurements, the International Soil Moisture Network (ISMN) was launched in 2009 as a community effort, funded by the European Space Agency. ISMN serves as a central repository for in situ soil moisture

measurements collected and freely shared by numerous organizations, harmonizes them by units, sampling rates, and applies advanced quality controls (Dorigo et al. 2011). Users can freely access the data from this database through an online web portal (<https://ismn.earth/>). The number of networks and stations covered by the ISMN is still growing. In this research, we exploited Soil Climate Analysis network (SCAN) stations from ISMN networks. SCAN began as a project of the Natural Resources Conservation Service in 1991 to measure soil moisture and soil temperature. SCAN network includes 200 stations over agricultural areas of U.S. SCAN station site monitors soil moisture content at several depths, air temperature, relative humidity, solar radiation, wind speed and direction, liquid precipitation, and barometric pressure. Table 5.2 shows SCAN data provided in detail.

Figure 5.1 displays dielectric constant measuring device probes which are in three different depths (2, 4, and 8 inches). However, the standard depths are at 2, 4, 8, 20 and 40 inches in most cases. Figure 5.2 also presents all data collectors and sensors needed in a SCAN station to obtain data explained in Table 5.2.

In this research, soil moisture and soil temperature are used from SCAN ISMN networks. Soil temperature is an important parameter in dielectric constant calculations from soil moisture data in experimental soil-dielectric constant models (see Section 3.2). Soil temperature and soil moisture data of 2 inches should be used in GNSS reflectometry since GNSS-R works with the first 5 cm depth of soil.

### **5.3 Ancillary data**

Since GNSS-R signals are affected by different parameters other than soil moisture as described in theoretical Section 2.1, the effect of these parameters must be included in the PINN model. These parameters include vegetation, soil texture, soil temperature, surface roughness and land cover related parameters.

Table 5.2. Data measured at SCAN stations.

<b>Data</b>	<b>Sensor</b>	<b>Accuracy and units</b>
<b>Soil moisture</b>	Collected by a dielectric constant measuring device; typical measurements are at 2, 4, 8, 20, and 40 inch	$\pm 0.03 \text{ m}^3 \text{ m}^{-3}$ <i>Volumetric soil moisture</i>
<b>Soil temperature</b>	Collected by an encapsulated thermistor; typical measurements are at 2, 4, 8, 20, and 40 inch.	$\pm 0.6^\circ \text{C}$
<b>Dielectric constant</b>	Measured by a dielectric probe	$\pm 0.2$ ( <i>unitless</i> )
<b>Precipitation</b>	Collected by storage gauge or tipping-bucket	$\pm 0.01 \text{ in}$
<b>Air temperature</b>	Collected by a shielded thermistor	$\pm 2^\circ \text{C}$
<b>Relative humidity</b>	Collected by a thin-film capacitance-type sensor	$\pm 2 \%$
<b>Wind speed and direction</b>	Collected by a propeller-type anemometer	$\pm 3 \text{ mph and } 355^\circ$
<b>Solar radiation</b>	Collected by a pyranometer	<i>5% of signal (<math>\text{W m}^{-2}</math>)</i>
<b>Barometric pressure</b>	Measured by a silicon capacitive pressure sensor	<i>Inch of Hg</i>



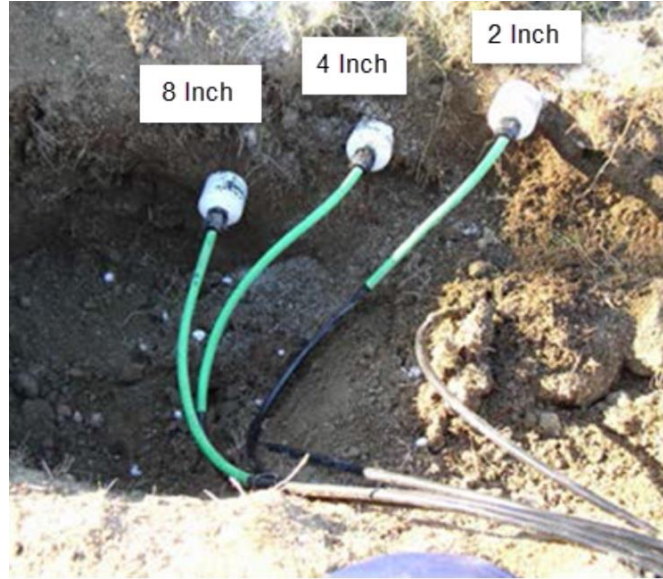


Figure 5.1. Soil moisture and temperature sensors at a SCAN site with different depths. <https://www.nrcs.usda.gov/wps/portal/wcc/home/aboutUs/monitoringPrograms/automatedSoilClimateMonitoring>

### 5.3.1 Vegetation

For vegetation, NDVI and VWC were included as input data to the PINN model. Normalized difference vegetation index (NDVI) has been used widely by the remote sensing community to determine vegetation cover (Gandhi et al. 2015; Meroni et al. 2019). NDVI parameter can be calculated from the near-infrared (NIR) and red (RED) bands of reflectance data (see Equation 5.2).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (5.2)$$

For NDVI data, the reflectance data were obtained from MODIS Aqua Surface Reflectance Daily Global 500m data set available at ([https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mod09ga\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09ga_v006)).

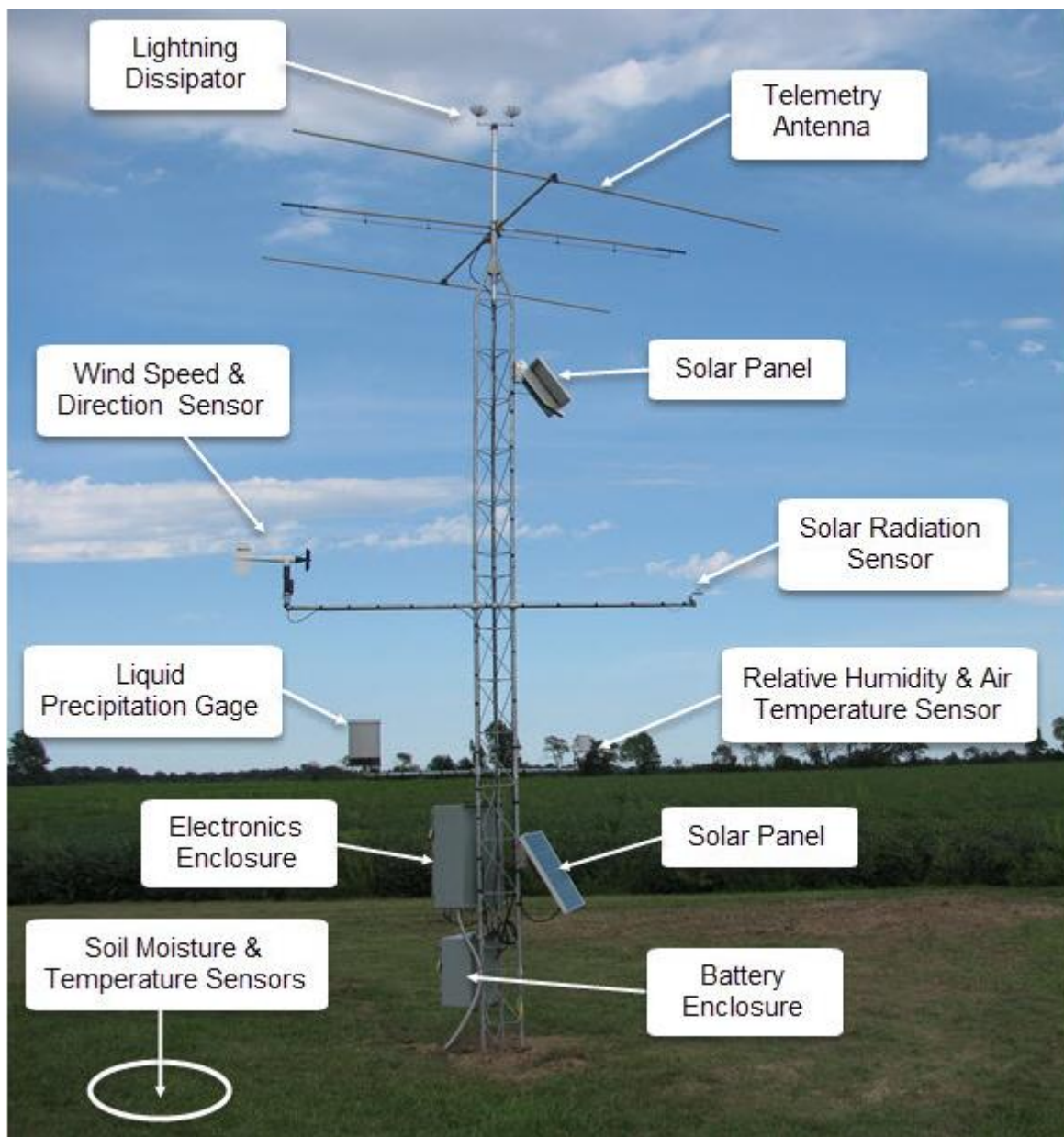


Figure 5.2. SCAN station sensors and data collectors.

<https://www.nrcs.usda.gov/wps/portal/wcc/home/aboutUs/monitoringPrograms/automatedSoilClimateMonitoring>

Then, NDVI values were calculated for each 500 m grid and averaged for a 4 km grid around the CYGNSS specular point to be respectful to previous research approaches (Eroglu et al. 2019). NDVI data usually suffers from clouds because it is generated by optical instruments such as in the MODIS mission. To deal with this problem, a cloud mask approach was applied to avoid cloudy pixels. Although MODIS satellites provide NDVI data directly, reflectance data was used since NDVI data of MODIS has a temporal resolution of 16 days. For example, in Senyurek et al. (2020), MODIS NDVI data with temporal resolution of 16 days were added to NN training. However, daily temporal resolution is preferred in this research for better resolution. To process MODIS data, Google Earth Engine (GEE) was used which is explained in detail in Section 5.4. GEE includes MODIS images. MODIS acquires images in 36 spectral bands, the spatial resolutions of which vary from 250 m to 1 km. MODIS time series are available in the GEE Data Catalog from 2000 to the present, facilitating temporal analysis over the globe. Vegetation water content (VWC) as another vegetation parameter was calculated from NDVI and stem factor data with Equation (3.7) in Section 3.1 and added to PINN inputs. To derive VWC values from NDVI, minimum and maximum values of NDVI are needed. Two different approaches can be used for calculation of VWC through NDVI: i) maximum and minimum of NDVI data ( $NDVI_{max}$ ,  $NDVI_{min}$ ) are computed from data, and ii) the current NDVI is considered as  $NDVI_{max}$ , and a global constant value of 0.1 is used for  $NDVI_{min}$ , as recommended in SMAP's VWC report (Chan et al. 2013). The second approach is only suggested for croplands and grasslands landcovers. Since data used in this research only includes grasslands and croplands, we also have used the second approach suggested by SMAP VWC report. In addition, Eroglu et al. (2019) have shown

that the second approach outperforms the first approach in CYGNSS soil moisture retrieval using NN.

### **5.3.2 Land cover and surface roughness**

To include land cover parameters effects in PINN, stem factor and b-parameter were used. Stem factor and b-parameter were obtained from SMAP look-up-table (LUT) for each station (Chan et al. 2013). First, land cover type should be determined by data coming from MODIS Terra and Aqua combined MCD12Q1 products. Then, stem factor and b-parameter data can be specified using land cover data and LUT (Chan et al. 2013). International geosphere-biosphere program (IGBP) land cover classification data provides open access 500-m spatial resolution data available at <https://lpdaac.usgs.gov/products/mcd12q1v006/> by NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center. MODIS time series are available in GEE Data Catalog from 2000 to present (Gorelick et al. 2017).

The stem factor and b-parameter values for the 4-km grid of each CYGNSS observation in this study were calculated as a weighted sum of LUT stem factors and b-parameters based on the land cover percentages, respectively.

Following the fact that only homogenous ISMN land cover stations have been selected, ISMN stations used in this research include croplands and grasslands land covers. For croplands and grasslands as suggested in (Chan et al. 2013), the instant NDVI value for  $NDVI_{max}$  and universal constant value of 0.1 for  $NDVI_{min}$  were used.

For surface roughness, h-parameter which is dependent on land cover was used. The h-parameter was also found in LUT of SMAP data like stem factor and b-parameter (Chan

et al. 2013). The same averaging computation as stem factor and b-parameter was done for h-parameter over the 4 km grids.

The SMAP L2\_SM\_P team calibrates, and updates values related to h-parameter, b-parameter, and stem factor as effective parameters in soil moisture calculations. The recent set of SMAP parameters are available for download from NSIDC website (<https://nsidc.org/data/smap/tools>). The values of h-parameter, b-parameter, and stem factor are provided as a land cover-based look up table in O'Neil et al. (2021), released in October 2021 and have been used in this research.

### **5.3.3 Soil texture**

Soil texture is defined as the relative proportions of each class (clay, silt, and sand). Sand gives material strength while clay binds it together and silt fulfils a less clear intermediate function. Soil texture data such as soil sand, clay and silt percentages have been used widely for soil classification. International soil classifications rely on relative fractions of soil particles of different sizes to establish soil textural class boundaries. These traditional classifications are convenient because grain size distributions can be measured relatively easily and can be estimated quickly and accurately in the field. Currently, soil texture classification is used commonly within agricultural, geotechnical, hydrological, and other related disciplines.

Soil classification with soil texture ratio was first appeared with Atterberg (1905) and the United States Department of Agriculture (USDA) (Whitney 1911). The international and USDA systems were accepted formally and led to the soil classification systems in wide use today (Davis 1927; International Society of Soil Science 1927). Figure 5.3 shows soil

classification with respect to clay, sand, and silt percentages in the USDA system (Groenendyk et al. 2015). To read soil texture classes based on clay, sand, and silt percentages, Figure 5.4. can be used to obtain soil texture classes with only two percentages, and one other can be calculated.

To retrieve soil moisture from dielectric constant, in addition to soil temperature, soil texture data including soil bulk density, clay and sand percentage are required.

Soil bulk density, clay and sand percentage were obtained from SMAP ancillary data available at <http://smap.jpl.nasa.gov/science/dataproducts/ATBD/>. The soil texture data are provided globally at 36 km, 9 km, 3 km and 1 km grid resolutions named the SMAP L2\_SM\_P, L2\_SM\_P\_E, and L2\_SM\_SP, respectively. It is assumed that clay, sand percentage and bulk density are constant over the 9 km x 9 km grid, allowing the use of ancillary data provided with “L2\_SM\_P\_E”.

## **5.4 Bulk data processing**

Neural networks require high numbers of data to be trained. However, processing different datasets of GNSS-R and remote sensing is time consuming and needs strong computer systems. So that even downloading data without processing it may take hours and days. Remote sensing systems have been collecting massive volumes of datasets for decades, managing and analyzing of which are not practical using common software packages and desktop computing resources. In this regard, Google has developed a cloud computing platform, called Google Earth Engine (GEE), to effectively address the challenges of big data analysis. This platform facilitates processing large amounts of geo data over large areas and monitoring the environment for long periods of time.



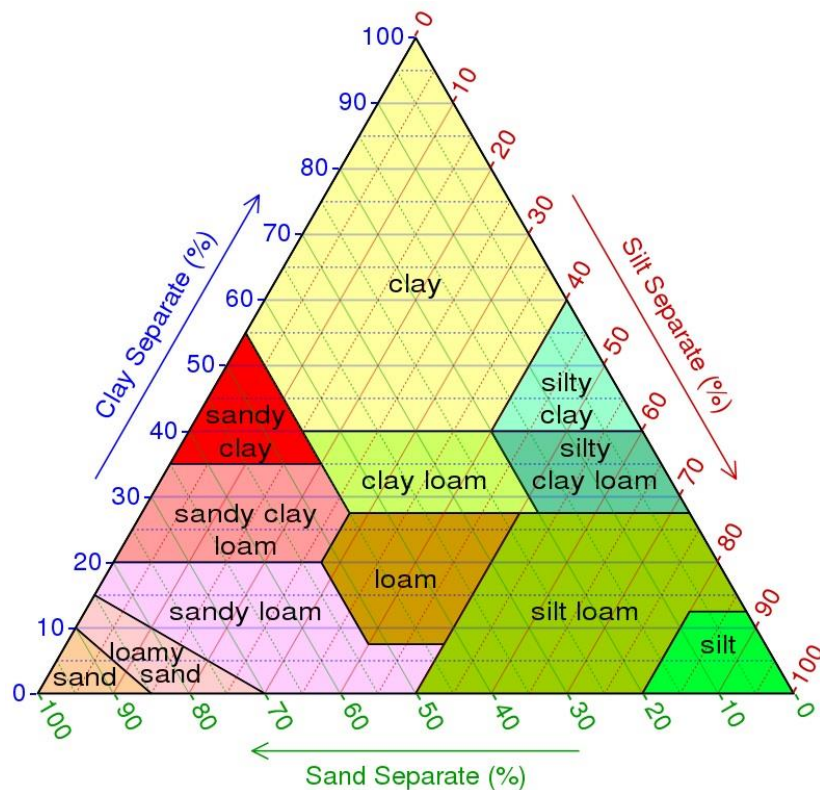


Figure 5.3. USDA Soil texture triangle (Groenendyk et al. 2015).

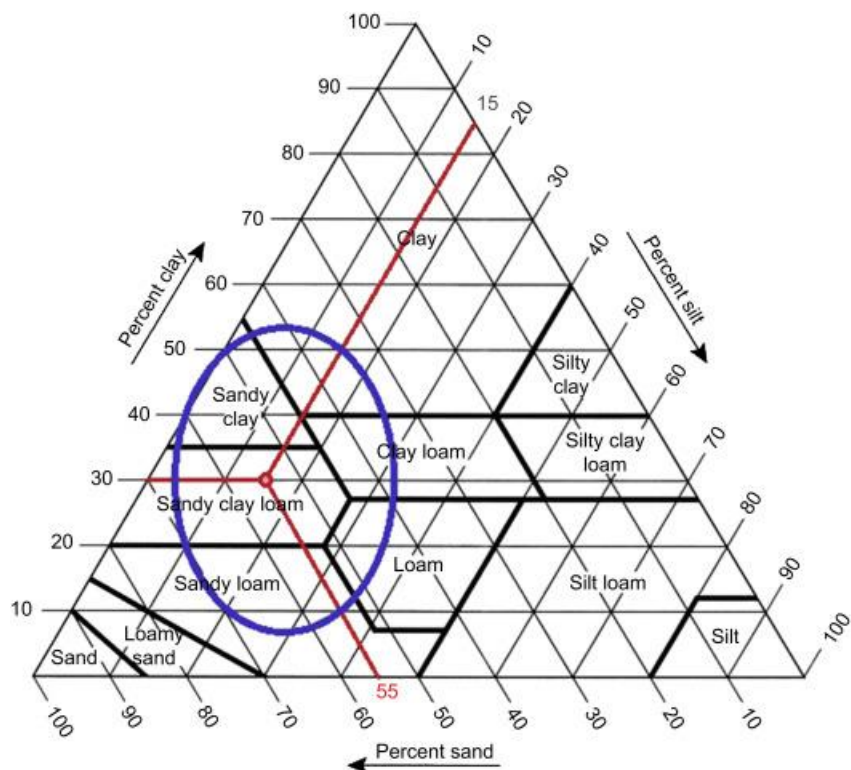


Figure 5.4. USDA Soil texture triangle visual assess (Harries and Sharma 2019).

Cloud computing platforms in general are efficient ways of storing, accessing, and analyzing datasets on very powerful servers, which virtualize supercomputers for the user. These systems provide infrastructure, platform, storage services, and software packages in a variety of ways for the customers (Chi et al. 2016; Ma et al. 2015). Several cloud computing platforms have so far been developed (Amani et al. 2020).

First, Amazon Web Services (AWS) which is a pay-as-you-go platform, where users pay based on the hours that they use the services (Tamiminia et al. 2020). AWS has a cloud Earth Observation (EO) platform called “Earth on AWS” as part of its public dataset program open data from several satellites such as Landsat-8, Sentinel-1, Sentinel-2, China–Brazil Earth Resources Satellite program, National Oceanographic, and Atmospheric Administration Advanced (NOAA) image datasets, as well as global model outputs.

Second, Azure is another cloud computing platform hosted by Microsoft. This platform has established the Artificial Intelligence (AI) for earth initiative to facilitate the use of its AI tools for addressing environmental challenges in four main areas of climate, agriculture, biodiversity, and water. Azure only contains Landsat and Sentinel-2 products for North America, since 2013, as well as moderate resolution imaging spectroradiometer (MODIS) imagery. Azure is also a pay-as-you-go platform which provides virtual systems for the users (Wilder 2012).

Google Earth Engine (GEE) is another cloud computing platform launched by Google in 2010. GEE uses Google’s computational infrastructure and available open access remote sensing datasets (Gorelick et al. 2017). GEE is the most popular big geo data processing platform due to following facts: i) GEE facilitates scientific processes by providing users



with free access to datasets (Tamiminia et al. 2020). Users can freely access GEE via an internet-based Application Programming Interface (API) and a web-based Interactive Development Environment (Tamiminia et al. 2020; Gorelick et al. 2017); ii) Users do not need to have expertise in web programming or HyperText Markup Language to use GEE; and iii) GEE has the features of an automatic parallel processing and fast computational platform to effectively deal with the challenges of big data processing (Hird et al. 2017). GEE was used in this research since downloading different datasets was time consuming. MODIS NDVI and land cover data are available on GEE platform for post processing. Using GEE, instead of downloading data and applying post processing tasks on them, MODIS data was available on GEE and no software was needed to do further post processes as various built-in analysis functions were already available on GEE. As a result, coding and processing all were done on GEE itself.

GEE main part is (Earth Engine) EE code editor. The Earth Engine (EE) Code Editor available at [code.earthengine.google.com](https://code.earthengine.google.com) is a web-based integrated development environment (IDE) for the Earth Engine JavaScript API. Code Editor features are designed to make developing complex geospatial workflows fast and easy. The Code Editor has the following elements (illustrated in Figure 5.5):

- JavaScript code editor
- Map display for visualizing geospatial datasets.
- API reference documentation (Docs tab).
- Git-based Script Manager (Scripts tab).
- Console output (Console tab).
- Task Manager (Tasks tab) to handle long-running queries

- Interactive map query (Inspector tab)
- Search of the data archive or saved scripts
- Geometry drawing tools.

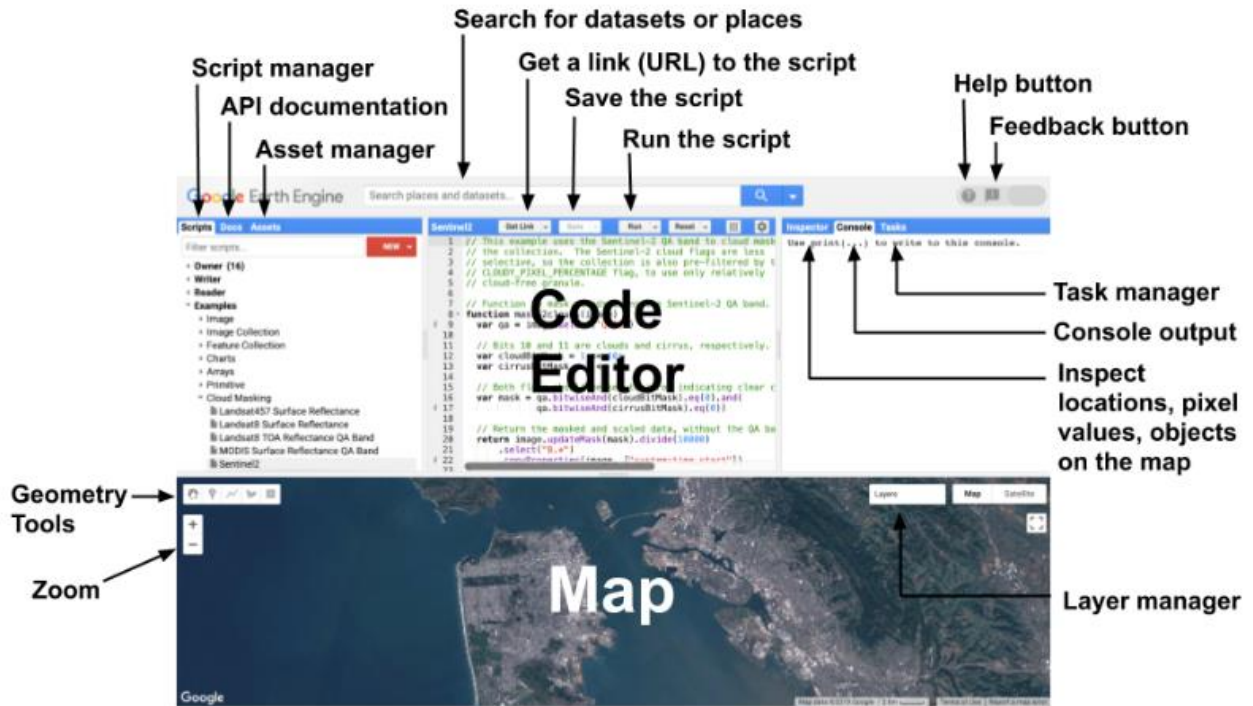


Figure 5.5. Diagram of components of the Earth Engine Code Editor from <https://developers.google.com/earth-engine/guides/playground>

The Code Editor has a variety of features to help you take advantage of the Earth Engine API. You can view example scripts or save your own scripts on the Scripts tab. Query objects are placed on the map with the Inspector tab. Google Visualization API can be used to display and chart numeric results. To share a unique URL to your script with collaborators and friends, the Get Link button can be used.

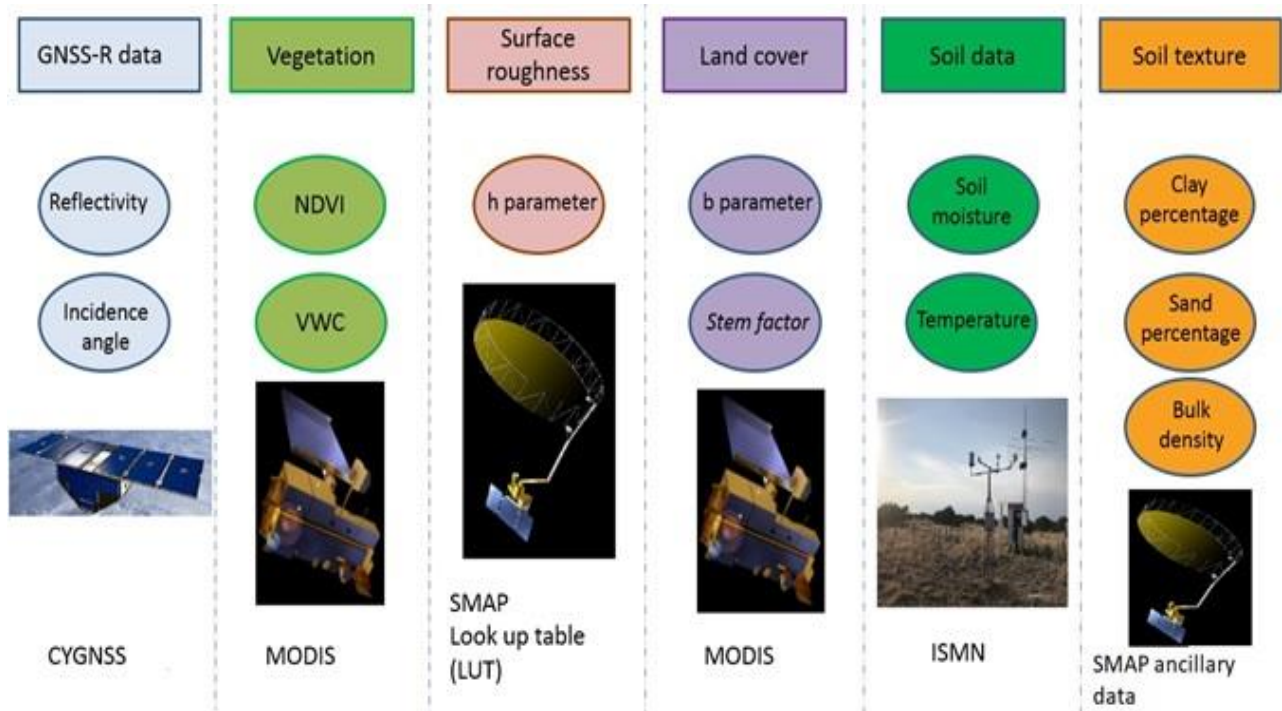


Figure 5.6. Different data sources used in PINN. Various data utilized in PINN are differentiated by different colours and summarized. GNSS-R data is obtained from CYGNSS mission explained in section 5.1, Vegetation, land cover and surface roughness, and soil texture data are explained in subsections 5.3.1, 5.3.2, and 5.3.3, respectively. Soil moisture and soil temperature data are also discussed in Section 5.2.

Scripts you develop in the Code Editor are sent to Google for processing and the generated map tiles and/or messages are sent back for display in the Map and/or Console tab. All you need to run the Code Editor is a web browser (use Google Chrome for best results) and an Internet connection. To summarize Chapter 5 and datasets used in PINN, a complete reference of data sources is listed in Figure 5.6. The following chapter describes the results of PINN using datasets explained in this Chapter.

## 6. GNSS-R SOIL MOISTURE RETRIEVAL BASED ON PINN APPROACH

The PINN model approach is examined in this chapter for GNSS-R soil moisture retrieval. Performance of PINN is expressed with RMSE and Pearson correlation coefficient metrics. The PINN model considers physics-based relationships despite NN. Therefore, parameters involved in physics have been used in this research, which leads to promising RMSE and Pearson correlation values.

### 6.1 PINN model performance analysis

As discussed in Section 4.2, the PINN model aims to minimize two loss functions together ( $MSE_{SM}$ ,  $MSE_{\Gamma}$ ) (see Figures 4.1 and 4.2). Figures 6.1 and 6.2 display how these two loss functions in PINN have changed during training. Figure 6.1 is a representative of the soil moisture loss function changes during training epochs. Figure 6.2 indicates reflectivity loss function variations. Soil moisture values are updated to satisfy both the physics model and NN regression. Soil moisture is computed in each step of training and then used in the inverse reflectivity calculation. As the soil moisture updates, the reflectivity loss function also changes as inverse calculated reflectivity ( $\Gamma_{SM}$ ) changes, while CYGNSS reflectivity data ( $\Gamma_{CYGNSS}$ ) is fixed.

The PINN algorithm is trained in order to minimize both soil moisture and reflectivity loss functions. In each step of training, soil moisture mean squared error values are updated (see Figure 6.1) as the PINN is learning. However, the reflectivity loss function ( $MSE_{\Gamma}$ ) has a slower learning process and remains within a specific range due to the following

reasons (see Figure 6.2): i) since the physics model equations are steady and the reflectivity data are constant, the reflectivity loss function ( $MSE_r$ ) changes are not significant as physics laws avoid sudden or large changes, iii) reflectivity is not the output of PINN training here; therefore, reflectivity variations are only because of soil moisture variations, and iii) although, the inverse calculated reflectivity values change due to soil moisture changes, the scale of reflectivity and soil moisture changes are not in the same range. For example, a change with magnitude of 0.2 in soil moisture can only cause a small change around 0.005 in reflectivity. Scale imbalances between two loss functions is a common issue in PINN. Conflicts can arise from different scales in loss function and can complicate convergence. Here, the soil moisture loss function has a higher scale than reflectivity values; therefore, the training was more affected by soil moisture loss function. Since soil moisture is the final output of the PINN, train and test dataset only apply to the soil moisture loss function (Figure 6.1).

The total data set is split into 20% test data and 80% train data. To find the best numbers of layers and neurons, an 8-fold cross-validation technique was applied to test the NN structure's overfitting and predictive ability. The train data set is divided randomly into eight portions. One portion is used as validation, and the other seven portions are used as training in each iteration. Therefore 10% of the data is used as validation, and 70% as training data in each iteration. After eight iterations, the mean squared error (MSE) for each combination is calculated. The NN structure with the best average MSE is chosen. Here, cross-validation is applied to NN and not PINN. Because the PINN model is more complex than NN, the PINN model with numbers of neurons as high as 20 is still not overfitted while NN is overfitted with 20 neurons. Since a comparison between NN and

PINN is worthwhile and is done in section 6.1.1, the number of layers and neurons is chosen to satisfy both NN and PINN. After choosing the best NN model using cross-validation, the best model with three layers with 12, 8, and 6 neurons in each layer, respectively, and the best model for two layers with 12, and 8 neurons in each layer are tested on test data.

As explained in Section 4.2, the PINN structure includes both NN regression and physics-based models. Therefore, the performance of PINN is dependent on both the NN model and the physics relationship. Hence, for PINN models still a strong NN is needed. To examine the impact of NN, an experiment was performed using different neural network sizes for PINN (number of layers and neurons per layer). In Section 6.1.1, different NN structures were used in the soil moisture retrieval PINN model to test NN size impact on PINN. Furthermore, Section 6.1.1 includes a comparison between NN and PINN performance before and after the addition of physics model loss function. To test the PINN sensibility, the PINN performance was then studied for different land covers. Pearson's correlation coefficient ( $R$ ) and RMSE performance metrics have been used to validate results of PINN. Pearson's coefficient is a standard method of estimating the degree to which two series are correlated. On the other hand, RMSE quantifies the exact value of differences between the soil moisture reference data and the predicted soil moisture for the model. In Section 6.1.2, sensitivity of PINN model to various land covers were surveyed.

### 6.1.1 PINN results for different NN structures

To investigate the effect of NN design on PINN results, different NN structures have been tested. First, two different NN structures have been used in the PINN model: i) two-layer NN (with 12 neurons on first layer and 8 on second layer), and ii) three-layer NN (with 12, 8 and 6 neurons on first, second and third layer, respectively). Physics-based relationships are fixed for both models. The number of layers and neurons have been chosen through experience with train-test learning curves to check over fitting. ISMN stations' information used in this processing is shown in Table 6.1.

Table 6.1 ISMN stations used in PINN.

<b>ISMN sites (SCAN)</b>	<b>Location</b>	<b>Dominant Soil texture</b>	<b>Latitude</b>	<b>Longitude</b>	<b>Dominant Land Cover (IGBP)</b>
<b>1. Adams Ranch</b>	New Mexico	Loam	34.25	-105.42	Grasslands
<b>2. Alkali Mesa</b>	Utah	Sand	37.67	-109.36	Grasslands
<b>3. Donkey Reservoir</b>	Utah	Clay Loam	38.21	-111.47	Grasslands
<b>4. Fort Reno</b>	Oklahoma	Clay	35.55	-98.01	Croplands, Grassland
<b>5. Manderfield</b>	Utah	Sand	38.37	-112.65	Grasslands
<b>6. Navajo Whiskey CK</b>	New Mexico	Loam	36.18	-108.95	Grasslands
<b>7. Panguitch</b>	Utah	Loam	37.87	-112.43	Grasslands
<b>8. Reese Center</b>	Texas	Loam	33.62	-102.03	Croplands, Grassland
<b>9. Uapb Dewitt</b>	Arkansas	Loam	34.28	-91.35	Croplands
<b>10. Vermillion</b>	Utah	Loam	37.18	-112.2	Grasslands
<b>11. West Summit</b>	Utah	Sandy Loam	38.02	-109.13	Grasslands

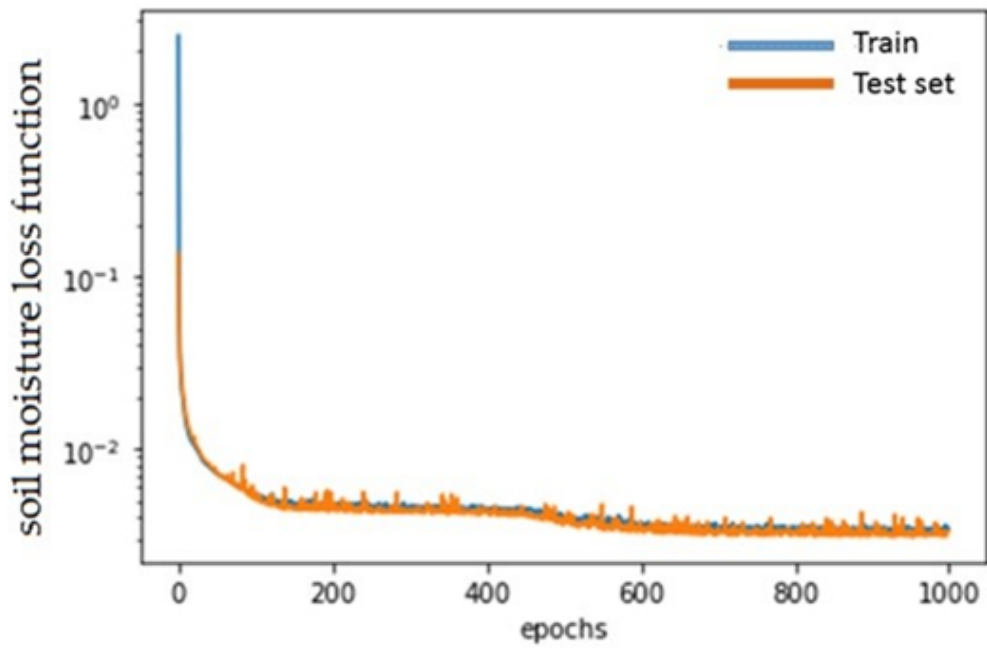


Figure 6.1. Soil moisture loss function ( $MSE_{SM}$ ) changes over 1000 epochs.

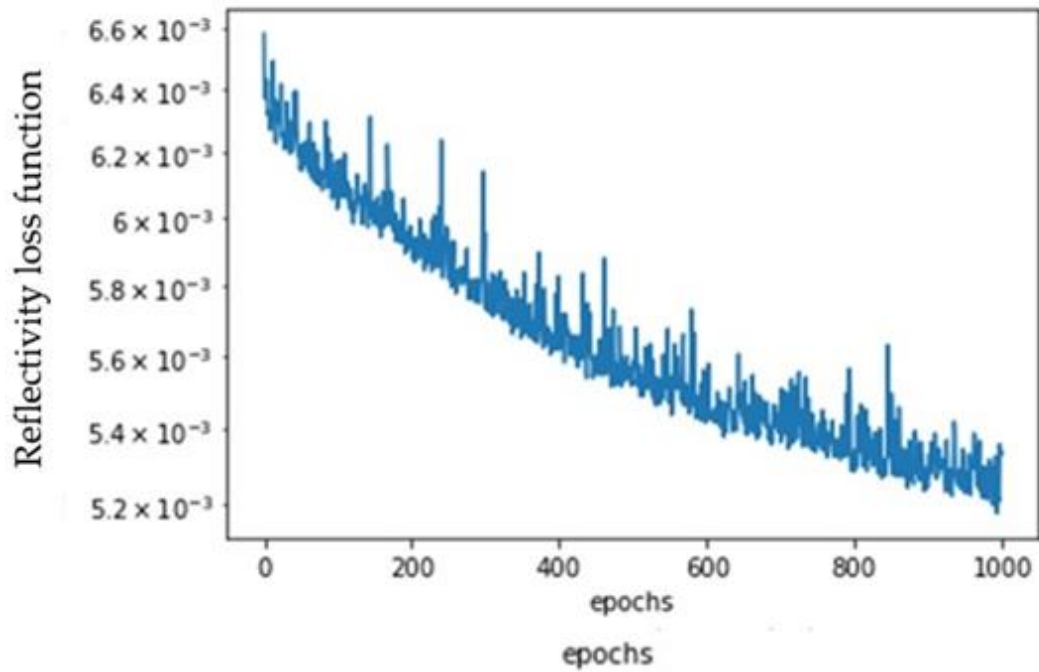


Figure 6.2. Reflectivity loss function ( $MSE_r$ ) changes over 1000 epochs.



Figure 6.3 shows the correlation scatter plots of the ISMN soil moisture data and CYGNSS-retrieved soil moisture values from PINN. For both PINN models, high overall agreement between the PINN soil moisture predictions and the reference ISMN data has been indicated. For example, for both two-layer and three-layer structures, the PINN model is able to generate soil moisture values with 0.82 and 0.85 Pearson's correlation coefficient (R) and RMSE of 0.0547 and 0.0500, respectively (see Table 6.2). While the number of layers is different for the two PINN models, the correlation coefficients have remained stable. The RMSE has almost the same value as well. For a two-layer PINN, the Pearson correlation coefficient improves from 0.78 for NN to 0.82 for PINN. On the other hand, for a three-layer PINN, the Pearson correlation value reduces from a 0.91 for NN to 0.85 (see Table 6.2). Although for a three-layer NN correlation coefficient was higher than PINN, for both PINN models RMSE have improved. For two-layer NN, RMSE improved from 0.0707 to 0.0547  $cm^3 / cm^3$ . For three-layer NN, the RMSE improved from 0.0774 to 0.0500  $cm^3 / cm^3$ . Table 6.2 illustrates how with changes of layers, the Pearson coefficient changes significantly for NN, but remains in the same range for the PINN.

Table 6.2. PINN performance metric values for different NN and PINN designs

<b>Model</b>	<b>RMSE</b>	<b>Pearson's coefficient</b>
<b>Two layer NN</b>	0.0707	0.78
<b>Three layer NN</b>	0.0774	0.91
<b>Two layer PINN</b>	0.0547	0.82
<b>Three layer PINN</b>	0.0500	0.85

Second, two different test-train ratios were applied to the PINN model. At first, the train / test ratio was 80% / 20% (see Figure 6.4 (i)). Second, a train / test ratio of 70% / 30%

was applied (see Figure 6.4 (ii)). For the three-layer NN, using the train / test ratio of 80% / 20%, the RMSE and correlation coefficient are 0.0519 and 0.82, respectively. For the train / test ratio of 70% / 30%, the RMSE and correlation coefficient values are 0.0469 and 0.83. Table 6.3 shows the comparison between two different train / test ratio options. With changes of train / test ratio, results of PINN remain within a solid range for both RMSE and Pearson coefficient (R).

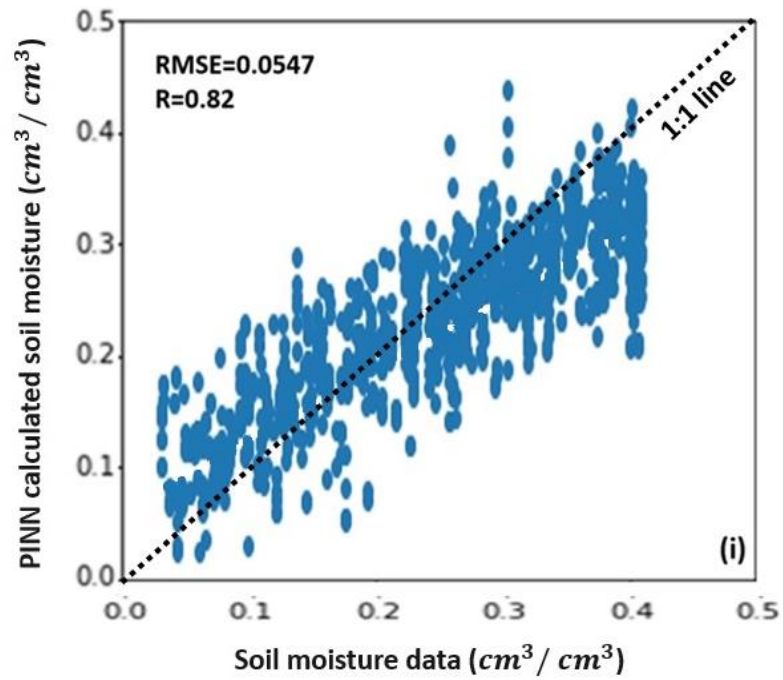
Table 6.3. PINN performance metric values for different train test ratios.

<b>Train-Test split ratio</b>	<b>RMSE</b>	<b>Pearson coefficient</b>
<b>80% - 20%</b>	0.0519	0.82
<b>70% - 30%</b>	0.0469	0.83

### 6.1.2 PINN sensitivity to different land covers

As discussed in Chapter 5, homogeneous land cover stations (see Table 6.1) with relatively less vegetation cover are used, since physics relations explained in Section 3.1 are valid for coherent reflections. Therefore, three categories (grasslands, mixed of grasslands and croplands, and croplands) have been studied to analyze PINN sensitivity to different land cover with different NDVI ranges. To study the impact of PINN on land covers only 7 stations were used to prevent addition of too many data points with grassland land cover as most of the stations in Table 6.1 have grassland land cover. Stations 3,6,7, and 11 from Table 6.1 were removed to have a fair distribution of data for each of three different land cover categories used.

### PINN soil moisture retrievals with a two-layer PINN



### PINN soil moisture retrievals with a three-layer PINN

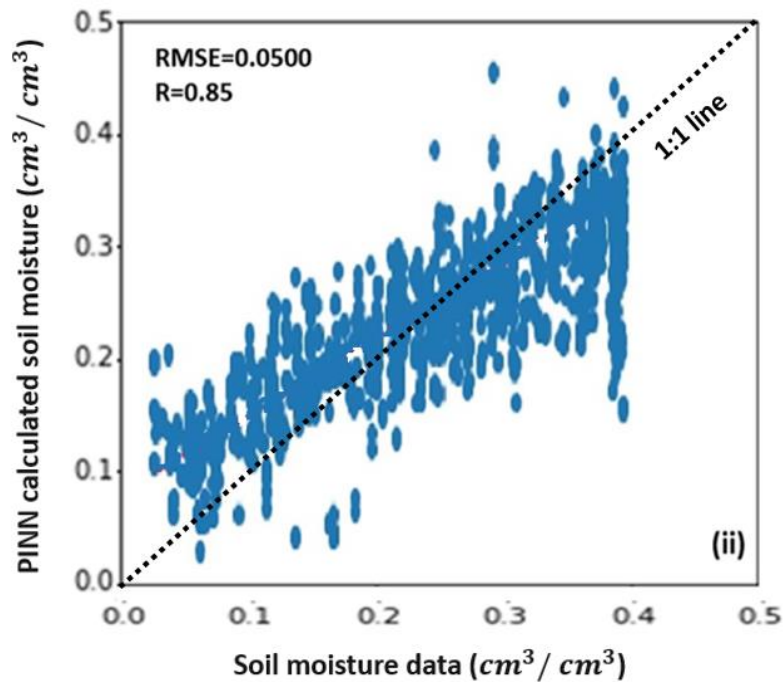


Figure 6.3. Correlation scatterplots of soil moisture retrievals from PINN with two different NN structures: (i) two-layer and (ii) three-layer.

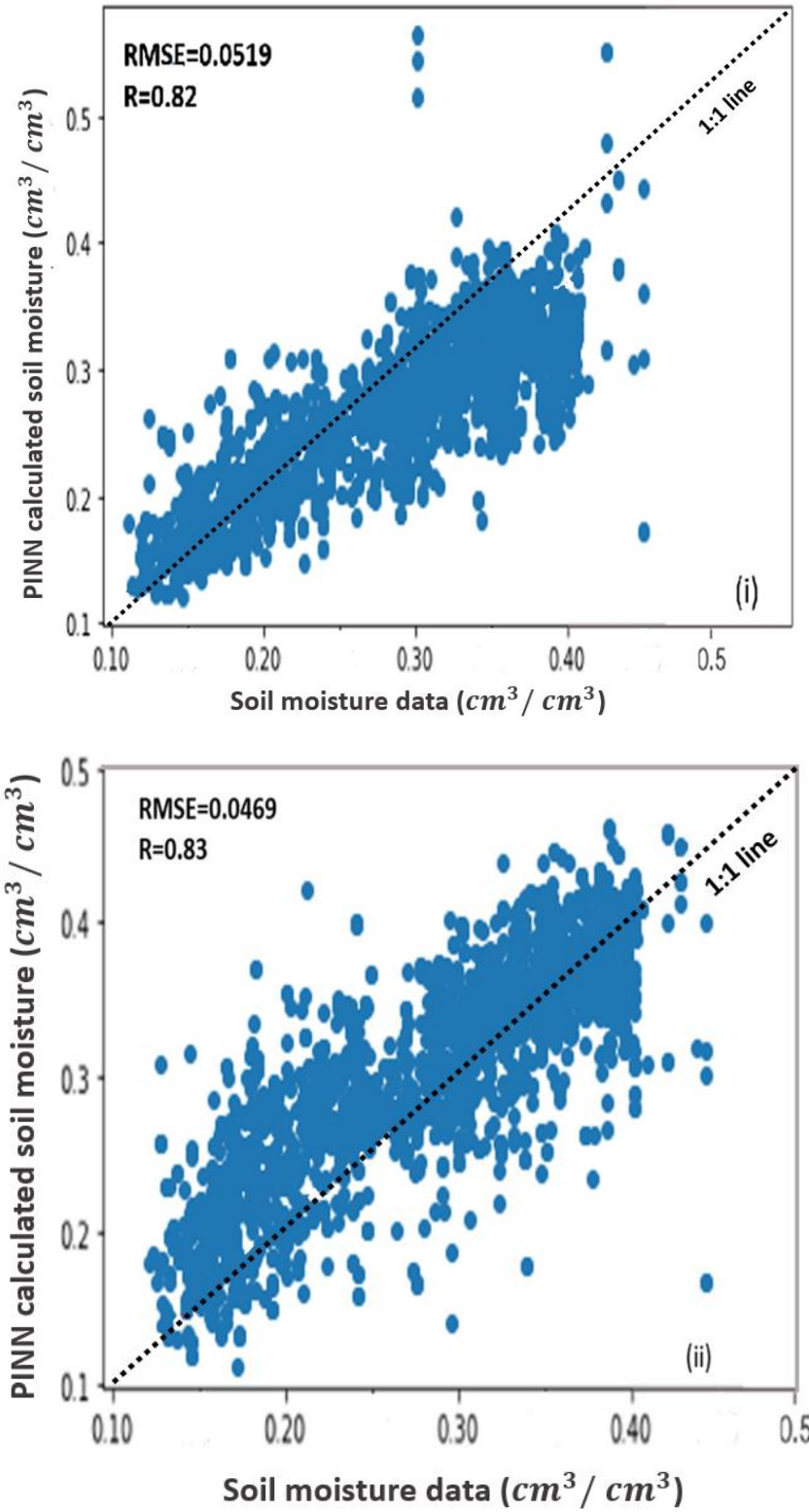


Figure 6.4. Correlation scatterplots of soil moisture retrievals from PINN for two different train/ test ratios: (i) 80% train and 20% test dataset (ii) 70% train and 30% test dataset.

The correlation of PINN calculated soil moisture values with ISMN soil moisture data are also examined for these three different land cover categories (see Figure 6.5). The Pearson correlation coefficients for grassland, mixed of grassland and croplands are 90%, 85% and 79% for three-layer NN (see Table 6.5) and 87%, 82% and 75% for two-layer NN (see Table 6.4), respectively (see Figure 6.5). These results show that PINN works better for less vegetated land covers, since physics equations explained in Section 3.1 are accurate for coherent reflections. As incoherent characteristics increase in datasets, the performance of PINN worsens. The sensitivity of performance metrics (RMSE and Pearson correlation coefficient) to different land covers are shown in Figures (6.5) and (6.6).

Table 6.4. PINN performance metric values for different land cover for two-layer PINN.

<b>Land cover</b>	<b>RMSE</b>	<b>Pearson's coefficient</b>
<b>Grassland</b>	0.0212	0.87
<b>Mix of grassland and cropland</b>	0.0331	0.82
<b>Croplands</b>	0.0435	0.75

Table 6.5. PINN performance metric values for different land cover for three-layer PINN.

<b>Land cover</b>	<b>RMSE</b>	<b>Pearson's coefficient</b>
<b>Grassland</b>	0.0187	0.90
<b>Mix of grassland and cropland</b>	0.0316	0.85
<b>Croplands</b>	0.0374	0.79

The same pattern exists for the RMSE values for various land covers. For instance, as the NDVI value ranges increase from grassland to croplands, the Pearson correlation coefficient (R) increases and RMSE decreases. These results indicate that the PINN is sensitive to physical parameters. For grassland, mixed of grasslands and croplands

RMSE values are 0.0212, 0.0331 and 0.0435, respectively, for a two-layer NN (see Table 6.4) and 0.0187, 0.0316 and 0.0375, respectively, for a three-layer NN (see Table 6.5). This result indicates increase in RMSE as land cover gets more vegetated (see Figure 6.6). The highest correlation value and lowest RMSE, which shows best performance, is related to grassland land cover with less vegetation. The lowest correlation value and highest RMSE is for croplands with higher vegetation and NDVI values. Two different NN structures are employed to ensure that same pattern of sensitivity to different land cover still exists for different neural network structures. As a result, after changing of the number of layers, grasslands still have the best result and croplands have the worst (see Tables 6.4 and 6.5, and Figures 6.5 and 6.6).

Figure 6.7 demonstrates correlation plots of PINN soil moisture retrievals and ISMN soil moisture data for different land covers. For grasslands having 90% of correlation shows PINN works better for less vegetated areas due to physics of the problem. Sample numbers for different land covers are 892, 1029, and 1264 for grasslands, mixed of grassland and cropland and croplands, respectively.

## **6.2 Discussion**

The PINN algorithm was able to estimate soil moisture values using a set of GNSS-R data and surface land surface parameters. Due to the fact that some physics understanding is provided in PINN, PINN was able to be trained on relatively less data due to integrating physics models with NN. This research was the first practice of PINN in CYGNSS soil moisture estimation, to the author's knowledge.

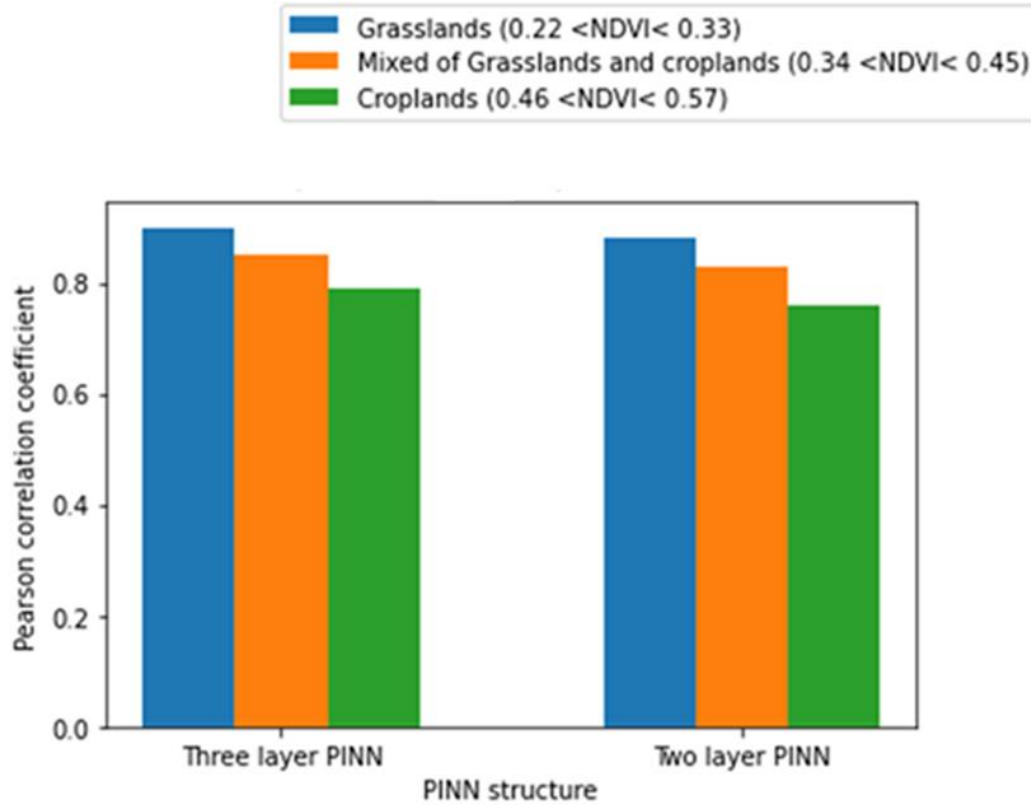


Figure 6.5 Sensitivity of PINN Pearson correlation coefficient metric to different land covers with two PINN structures (two-layer and three-layer PINN). Best performance (highest correlation) is related to grasslands with less vegetation and worst performance of PINN (lowest correlation) is for croplands with higher vegetation.

To confirm PINN capability in CYGNSS-derived soil moisture retrieval, PINN is compared with NN for the same data sets (see Table 6.2). Performance of PINN was expressed with RMSE and Pearson correlation coefficient metrics. The PINN model considers physics-based relationships in the modelling despite NN. Therefore, parameters involved in physics have been used in this research, which leads to promising RMSE and Pearson correlation values.

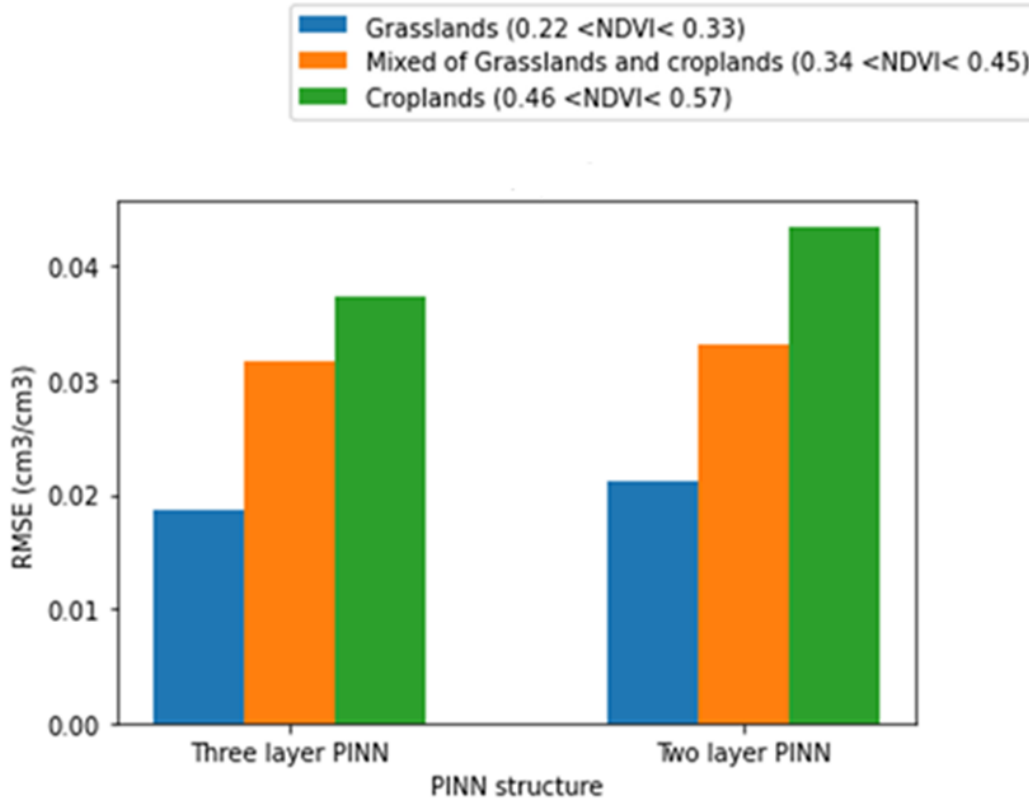
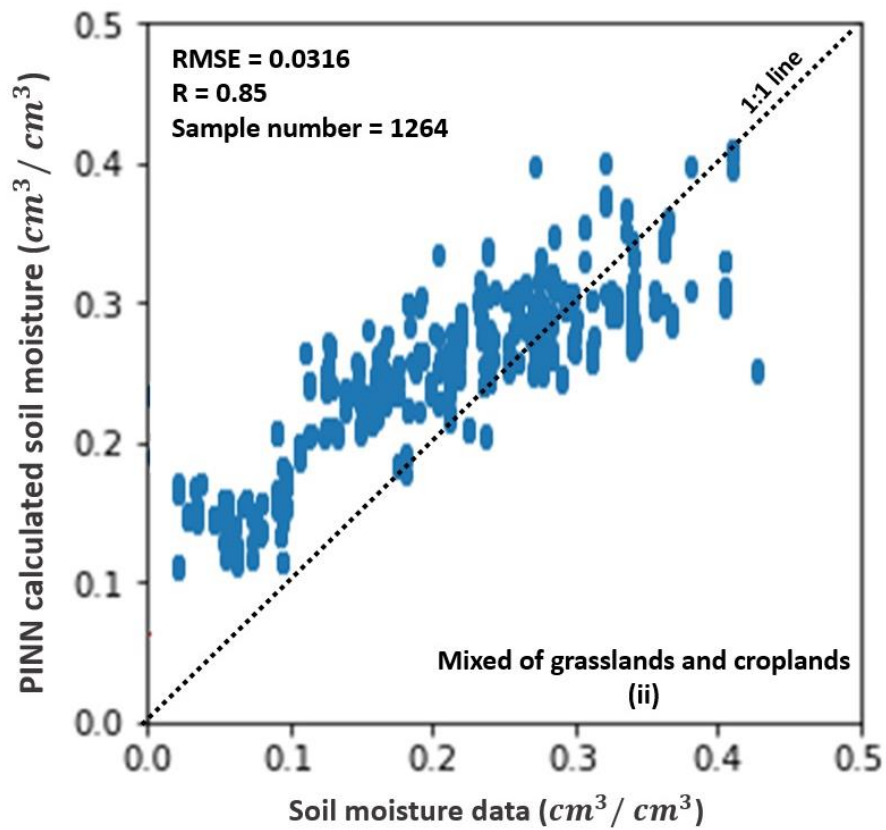
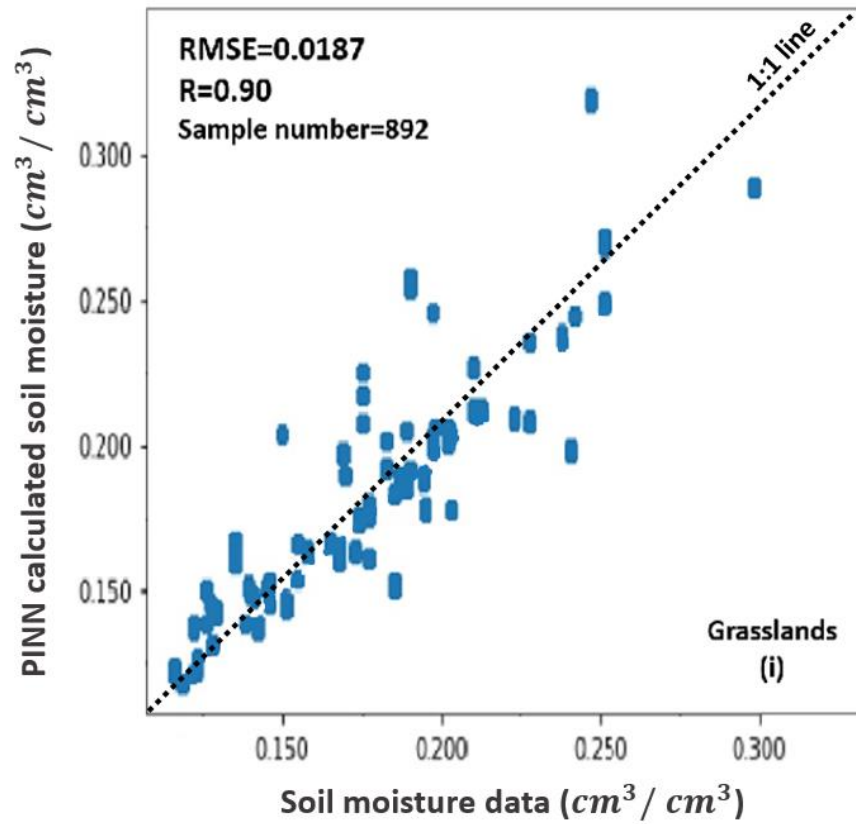


Figure 6.6. Sensitivity of PINN RMSE metric to different land covers with two PINN structures (two-layer and three-layer PINN). Best performance (lowest RMSE) is for grasslands with less vegetation and worst performance of PINN (highest RMSE) is related to croplands with higher vegetation.

The PINN model is capable of obtaining soil moisture values with an RMSE of 0.05 and a Pearson correlation coefficient of 0.85 for a three-layer PINN and an RMSE of 0.0547 and a Pearson correlation coefficient of 0.82 for a two-layer PINN. The RMSE difference between a two-layer NN and PINN is 0.016 and for a three-layer NN and PINN is 0.0274. PINN has a lower RMSE due to the addition of physics knowledge in modelling. NN has RMSEs around 0.07 showing its less capability over an approximately low number of stations in comparison to PINN.





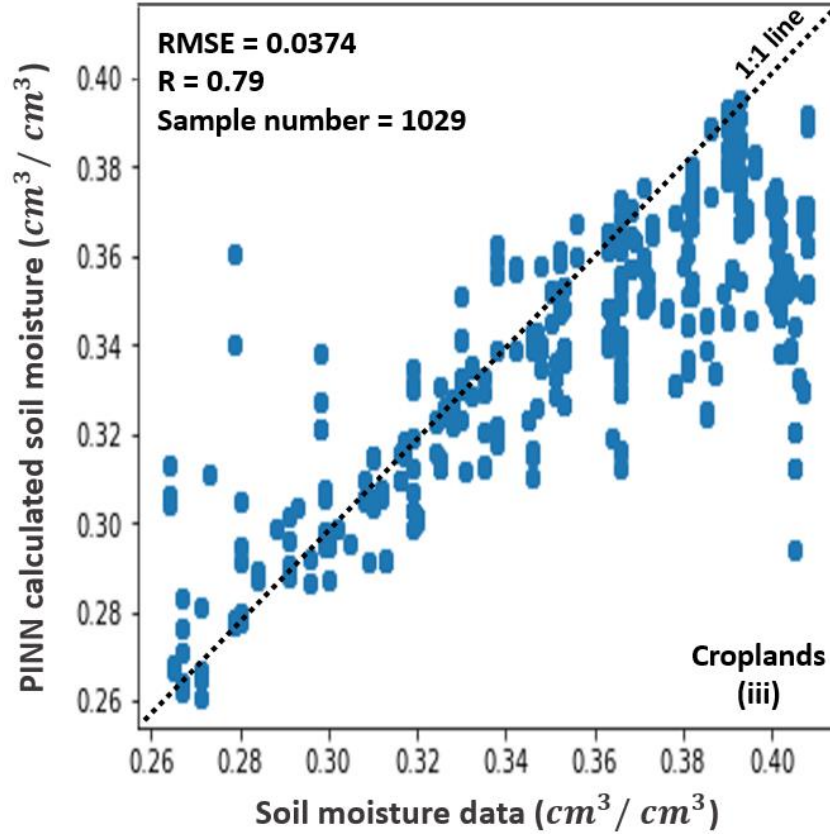


Figure 6.7. Correlation plots of PINN soil moisture retrievals and ISMN soil moisture data. (i) For grasslands with 90% Pearson's correlation. (ii) For mixed of grasslands and croplands with 85% Pearson's correlation. (iii) For Croplands with 79% Pearson's correlation.

As suggested by (Colliander et al. 2017; O'Neill et al. 2021), RMSEs lower than around 0.055 are required for soil moisture studies. Therefore, here PINN has acceptable results with RMSEs around 0.05 while NN has worse results with RMSEs around 0.07. Although the number of stations were relatively low for both NN and PINN cases, PINN has shown better results than NN in less data regime due to the addition of physical modelling. In addition, changes in NN design, such as number of layers and neurons, and train/test ratio have affected NN results; however, PINN results have remained consistent (see Tables 6.2 and 6.3). While NN structure can affect the outcome of the PINN model – as

a consequence of adding physics relationships to NN, results are in a relatively sustainable range.

Since this research is the first attempt of PINN in GNSS-R soil moisture retrieval, comparison of results with NN research was not straightforward. Validation of PINN results with comparison with previous published works using NN was challengeable owing to two issues:

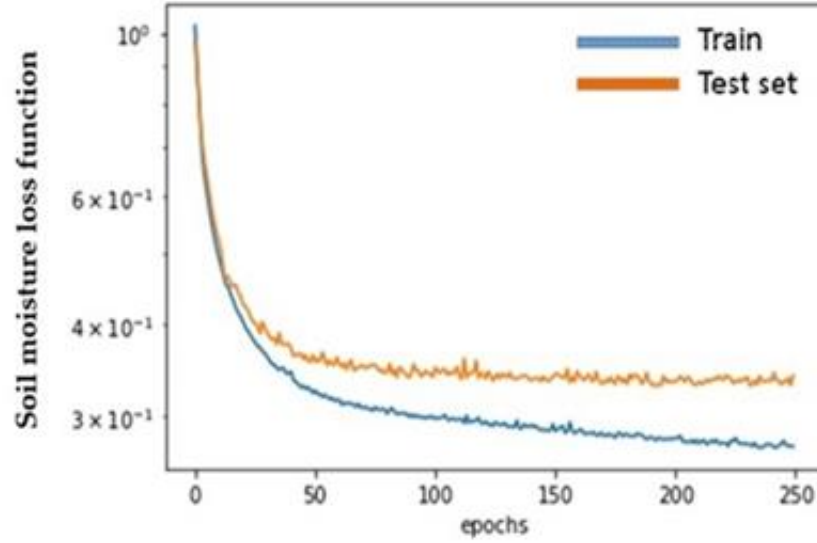
- i) It is not easy to choose the best NN model and compare it with PINN. As there are several differences between the previous publications based on NN algorithms including time span, number of observations, number of inputs in NN, reference sources for soil moisture data, ancillary datasets, and pre-processing and labeling filters and different resolutions. Each NN model includes different datasets and number of inputs. For example, in (Eroglu et al. 2019) researchers used elevation, h-parameter and slope for surface roughness; however, in (Senyurek et al. 2020) researchers used only elevation for surface roughness. The current study uses only h-parameter for surface roughness, since it was the only parameter affecting the physics of the problem. In both (Eroglu et al. 2019) and (Senyurek et al. 2020), for example, TES (Trailing Edge Slope) is added as a coherency/incoherency indicator. The current study only focuses on coherent reflections and TES is not involved in physics model. Therefore, TES is not used in PINN modeling. Soil moisture reference data for validation and testing can also come from different sources which makes comparison complicated. For example, in (Senyurek et al. 2020),

- authors used SMAP data for reference while in (Eroglu et al. 2019) authors used ISMN data for soil moisture reference. Finally, due to the difference in the data used, comparison of PINN with previous NN models is futile and difficult.
- ii) Although PINN algorithms are physically and scientifically more consistent predictive models, the PINN model suffers from the complexity of physics models. By adding physics to NN, new parameters and their uncertainties are added to the model. For example, by inclusion of soil moisture-dielectric constant model, more parameters are added to the problem. Semi-empirical models like Mironov and Dobson were developed for different purposes other than GNSS-R applications. These empirical parameters need to be calibrated or modified for GNSS-R applications. Uncertainty in soil moisture-dielectric constant models is added to the PINN as physics is added in the model. As a result, a more complex NN was needed to train datasets due to additional parameters. Although the number of inputs remains the same, PINN needed higher numbers of neurons and layers to learn. For example, a high parameter NN like three-layer NN with 20 neurons in each layer has been overfitted (see Figure 6.8 (i)). But PINN with same three layers NN with 20 neurons in each layer has shown a good fitting for our dataset (see Figure 6.8 (ii)). Because both train and test data set were decreasing, overfitting occurs when the NN does not respond well to test dataset with same number of layers and neurons. Consequently, a PINN requires higher numbers of neurons and layers in comparison to NN models, due to complexity and uncertainties of involving physics parameters, as the same number of layers and neurons may cause

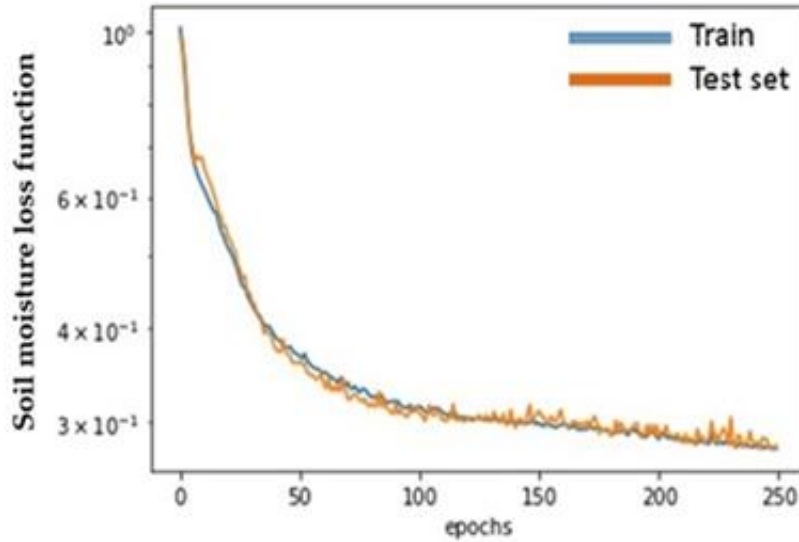
overfitting for NN (see Figure 6.8 (i)), but still be a good choice for PINN structure (see Figure 6.8 (ii)). Figure 6.8 shows that the same number of layers, neurons and hyper parameter choice seemed to be over-fitted for NN and good fit for PINN. This complicates the comparison of NN with PINN.

To check PINN sensibility to physical parameters such as land cover and vegetation, the dataset was separated into three categories (see Subsection 6.1.2). After separation of datasets by land cover categories, RMSE values decreased from RMSE value for entire data set (see Tables 6.4 and 6.5 and compare with Tables 6.2 and 6.3). Because after separation of the data to same land cover categories, the data points have higher similarities in each land cover category. Therefore, it is easier for PINN to learn. PINN was sensitive to different land cover categories and since physical equations involved in PINN obey coherent reflection assumption, as incoherency characteristics like vegetation increased in land cover categories, e.g., with comparison of croplands to grasslands, grasslands have better results due to higher possibility of coherent reflections because of less vegetation.

While PINN has been sensitive to different land covers, more diverse stations are needed to validate this consequence. Development of electromagnetic relations for more incoherent characteristics has been challenging in GNSS-R community. As physics understanding and models between soil moisture and dielectric constant improves, the PINN model can become stronger and more ubiquitous. This study only focused on homogenous flat stations with less vegetation. Development of PINN algorithms for diverse stations deserve future research.



(i)



(ii)

Figure 6.8. Learning curve plots for a three-layer NN and PINN comparison with 20 neurons in each layer. (i) Soil moisture loss function ( $MSE_{SM}$ ) changes for NN. Three-layer NN with 20 neurons in each layer shows over fitting because as test loss increases as epoch increases. It shows NN was not capable of learning new test dataset because of high number of neurons. (ii) Three-layer PINN with 20 neurons have shown a good fit indicating that PINN was able to learn test new dataset. Three layers with 20 neurons was a good fit for PINN while it caused overfitting for simple NN.

## **7. COMPARISON OF DIFFERENT SOIL DIELECTRIC CONSTANT MODELS FOR PINN SOIL MOISTURE ESTIMATES**

As explained in Section 3.2, different soil dielectric constant models exist that connect soil dielectric constant to soil moisture content. Since soil dielectric constant models play an important part in physical relationships between reflected signal power and soil moisture as discussed in Sections 3.1 and 3.2, the effect of these models on soil moisture predictions needs to be studied. Data used in these models also differ slightly; therefore, input data in PINN differs for different soil dielectric constant models. Concerning which soil dielectric constant model is used, physics models part in PINN changes resulting changes in PINN structure. In this chapter, the impact of four soil dielectric constant models on PINN results is investigated. Using different soil dielectric constant models in PINN changes physics models added to PINN. As a result, the performance of PINN changes as the soil dielectric constant model changes in PINN model.

Soil dielectric constant models differ from each other in two aspects: i) Soil dielectric constant model parameters were generated based on dissimilar base datasets with different soil texture values. For example, the Dobson and Hallikainen models used data sets from 5 soil fields with sand percentages less than 50% and clay percentages less than 47% (see Table 3.1). However, Wang and Mironov models were studied on a wider range of sand and clay percentages, ii) These four models, Hallikainen, Wang, Dobson, and Mironov, use different parameters in their modelling (see Table 7.1).

In Table 7.1, data used in PINN algorithm for different soil dielectric constant models is shown. For all four models land cover, surface roughness, CYGNSS GNSS-R data, and vegetation data are the same. The PINN model for these four soil-dielectric models varies

only in soil texture and soil temperature data. According to Table 7.1, Mironov has fewer physical features. In terms of soil texture data, Mironov only considers clay percentage. However, all other models (Hallikainen, Wang, and Dobson) contain sand percentage as well. Dobson is the only model that uses bulk density data. Wang and Dobson models consider temperature in their modeling while Hallikainen and Mironov models ignore temperature data. Consequently, Dobson has the highest numbers of data features used while Mironov has the lowest numbers of data features. In addition, Wang has the second highest numbers of features. The performance of four different soil dielectric constant models has been studied in Subsections 7.1 and 7.2.

Table 7.1. Comparison of soil dielectric models used in PINN algorithm in terms of required data.

<b>Data category</b>	<b>Hallikainen</b>	<b>Wang</b>	<b>Dobson</b>	<b>Mironov</b>
<b>Land cover</b>	b-parameter, stem factor	b-parameter, stem factor	b-parameter, stem factor	b-parameter, stem factor
<b>Surface roughness</b>	h-parameter	h-parameter	h-parameter	h-parameter
<b>CYGNSS GNSS-R data</b>	Reflectivity, Incidence angle	Reflectivity, Incidence angle	Reflectivity, Incidence angle	Reflectivity, Incidence angle
<b>Vegetation</b>	NDVI, VWC	NDVI, VWC	NDVI, VWC	NDVI, VWC
<b>Soil texture</b>	Clay%, Sand%	Clay%, Sand%	Clay%, Sand%, Bulk density	Clay%
<b>Soil temperature</b>	Not included	Included	Included	Not included
<b>Total numbers of data in PINN</b>	9 inputs	10 inputs	11 inputs	8 inputs



To investigate the effect of soil dielectric constant models in the PINN algorithm performance, two years of data, from January 1st of 2020 to the end of 2021, as explained in Chapter 5 were used. However, more stations were added to include more diverse soil texture classifications for further studies and comparison of models. Eight stations were added to the stations used in Chapter 6 (see Table 6.1 and Table 7.2 for comparison). Details of the stations used in processing different soil dielectric models are shown in Table 7.2. All of these stations come from SCAN ISMN networks explained in Section 5.2.

## **7.1 Impact of different soil dielectric models on PINN performance**

Four different soil dielectric models were used to understand their effect on PINN results. PINN produced soil moisture values highly correlated with initial soil moisture data. Figure 7.1 demonstrates four correlation plots related to PINN soil moisture retrievals for all four soil dielectric models studied in this research.

The PINN model is able to generate soil moisture values with four different soil dielectric models with following metrics (see Table 7.3 and Figures 7.2 and 7.3): i) Hallikainen model with Pearson's correlation coefficient ( $R$ ) of 0.94, RMSE of 0.0478 and ubRMSE of 0.0473, ii) Wang model with Pearson's correlation coefficient ( $R$ ) of 0.92, RMSE of 0.0509 and ubRMSE of 0.0499, iii) Dobson model with Pearson's correlation coefficient ( $R$ ) of 0.86, RMSE of 0.0568 and ubRMSE of 0.0560, and iv) Mironov model with Pearson's correlation coefficient ( $R$ ) of 0.81, RMSE of 0.0687 and ubRMSE of 0.0685. Table 7.3 compares Pearson's correlation coefficient, RMSE and ubRMSE values of four different soil dielectric models.

Table 7.2. ISMN stations used in comparison of soil dielectric models' performance in PINN algorithm.

<b>ISMN sites (SCAN)</b>	<b>Location</b>	<b>Dominant Soil texture</b>	<b>Latitude</b>	<b>Longitude</b>	<b>Dominant Land Cover (IGBP)</b>
1. Adams Ranch	New Mexico	Loam	34.25	-105.42	Grasslands
2. Alcalde	New Mexico	Sandy Loam	36.08	-106.05	Grasslands
3. Alkali Mesa	Utah	Sand	37.67	-109.36	Grasslands
4. Beasley Lake	Mississippi	Sandy Loam	33.38	-90.65	Croplands
5. Cochora Ranch	California	Loam	35.12	-119.60	Shrublands, Grassland
6. Donkey Reservoir	Utah	Clay Loam	38.21	-111.47	Grasslands
7. Fort Reno	Oklahoma	Clay	35.55	-98.01	Croplands, Grassland
8. Jornada Exp Range	New Mexico	Sandy Clay Loam	32.55	-106.7	Shrublands
9. Manderfield	Utah	Sand	38.37	-112.65	Grasslands
10. Mcracken Mesa	Utah	Clay Loam	37.45	-109.33	Shrublands, Grassland
11. Navajo Whiskey CK	New Mexico	Loam	36.18	-108.95	Grasslands
12. Panguitch	Utah	Loam	37.87	-112.43	Grasslands
13. Reese Center	Texas	Loam	33.62	-102.03	Croplands, Grassland
14. San Angelo	Texas	Loam	31.55	-100.50	Grasslands
15. Sevilleta	New Mexico	Sandy Clay Loam	34.35	-106.68	Shrublands, Grassland
16. Shadow Mtns	California	Loam	35.47	-115.72	Shrublands
17. Uapb Dewitt	Arkansas	Loam	34.28	-91.35	Croplands
18. Vermillion	Utah	Loam	37.18	-112.2	Grasslands
19. West Summit	Utah	Sandy Loam	38.02	-109.13	Grasslands

Hallikainen model has the best result due to the lowest RMSE and ubRMSE values, and highest correlation coefficient (see Figures 7.2 and 7.3, and Table 7.3). The second and third best model results are for Wang and Dobson models, respectively. Finally, the

Mironov model has the worst results due to the highest RMSE and ubRMSE values and lowest correlation coefficient. Figure 7.2 compares PINNRMSE and ubRMSE of four different soil dielectric constant models and Figure 7.3 shows the correlation coefficient values for different models. Figures 7.2 and 7.3 demonstrate that Mironov has the worst results and Hallikainen has the best results with respect to RMSE, ubRMSE and correlation coefficient values. RMSE and ubRMSE values range change are around 0.0209 and 0.0212 from the best model (Hallikainen) and worst model (Mironov). Pearson's correlation coefficient changes from 0.94 to 0.81 decreasing 0.13 when changing soil dielectric models from Hallikainen model to Mironov model.

Table 7.3. Comparison of four soil dielectric models' performance used in PINN.

<b>Soil dielectric constant model</b>	<b>RMSE</b>	<b>ubRMSE</b>	<b>Pearson's correlation coefficient</b>
<b>Dobson</b>	0.0568	0.0560	0.86
<b>Hallikainen</b>	0.0478	0.0473	0.94
<b>Mironov</b>	0.0687	0.0685	0.81
<b>Wang</b>	0.0509	0.0499	0.92

To understand the differences of the four soil dielectric models better, dielectric constant values were calculated from PINN retrieved soil moisture values and compared with soil moisture ISMN reference data as shown in Figures 7.4 and 7.5. Figures 7.4 and 7.5 display computed dielectric constant values from PINN soil moisture retrievals with respect to soil moisture data for 2020 and 2021, respectively. The dielectric constant calculated using Dobson model generally has a separation gap with other soil dielectric models (Liu and Liu 2020; Zhang et al. 2020; Guo et al. 2016) due to lack of physical consideration of the dielectric properties bound and free water as explained in Section 3.2. Figures 7.4 and 7.5 demonstrate rigidity of Dobson and Mironov model with

comparison to Wang and Hallikainen models. The differences among models cannot be neglected at medium and high moisture contents which increases with the sand content. Differences between Dobson model and other models are due to the lack of distinction between bound water and bulk water in the Dobson model.

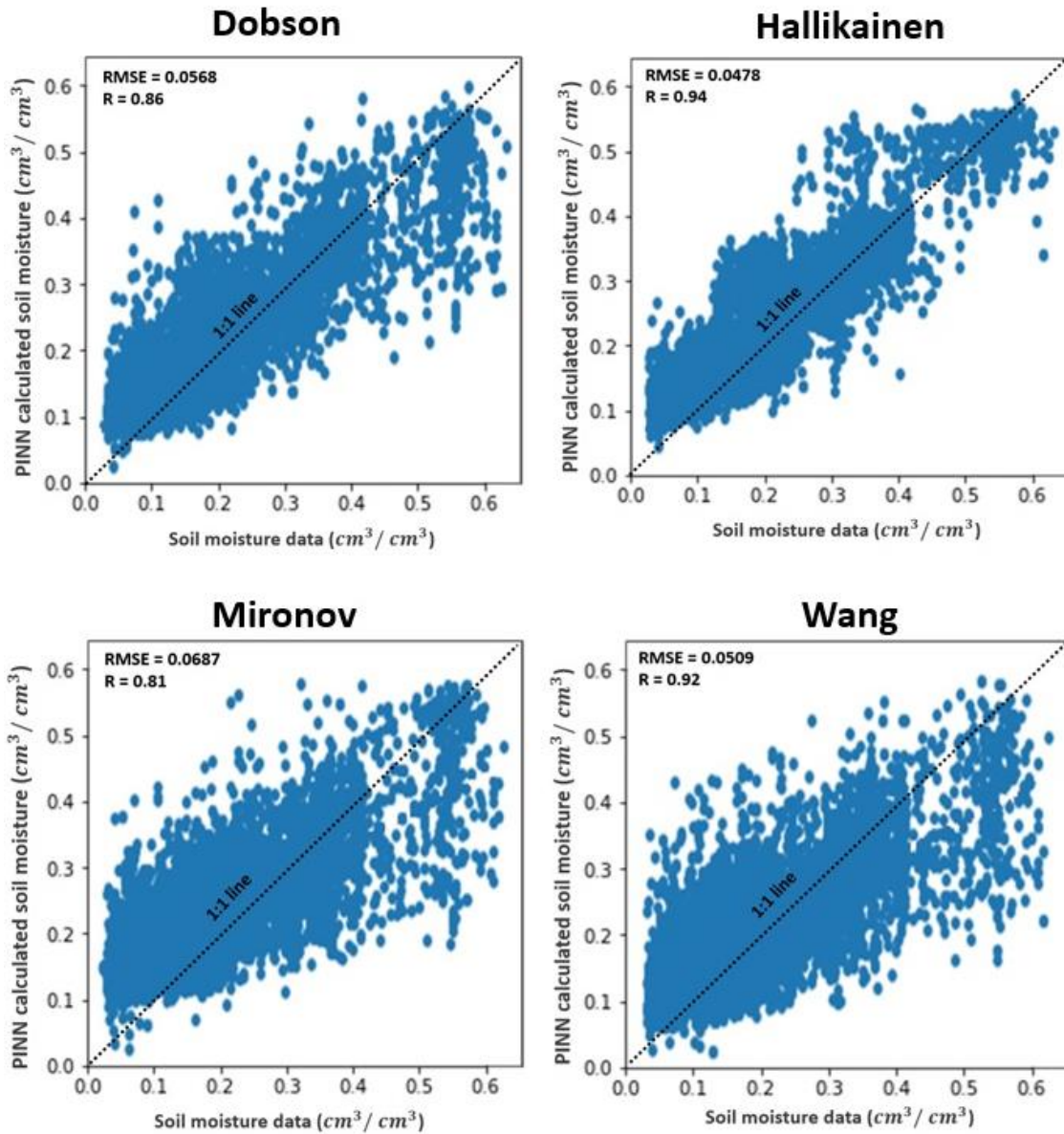


Figure 7.1. Correlation plots of PINN soil moisture retrievals for different soil dielectric models. Pearson's correlation coefficients for Dobson, Hallikainen, Mironov, and Wang models are 0.92, 0.94, 0.81, and 0.88, respectively. Hallikainen model resulted in highest correlation and Mironov resulted in the lowest correlation.

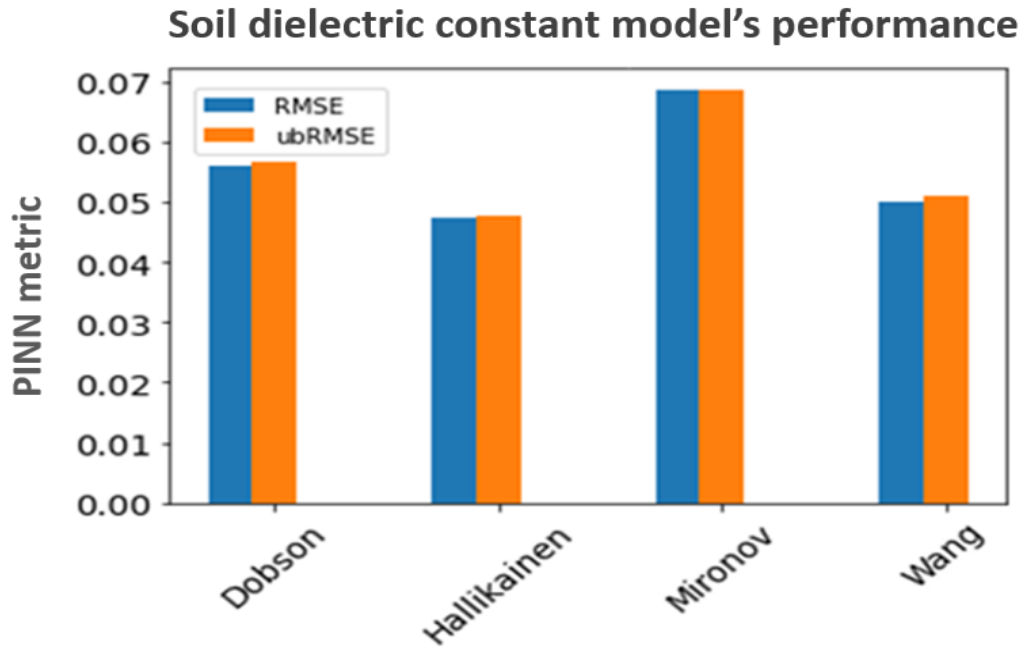


Figure 7.2. RMSE and ubRMSE comparison of four soil dielectric constant models used in PINN algorithm.

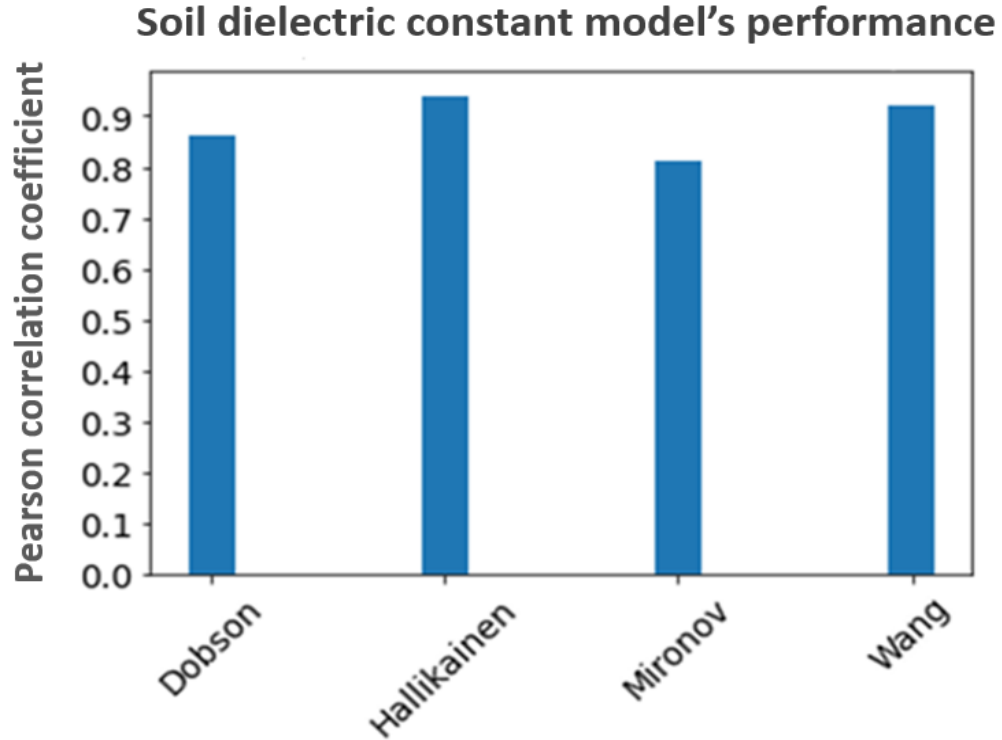


Figure 7.3. Pearson's correlation comparison of four soil dielectric constant models used in PINN algorithm.

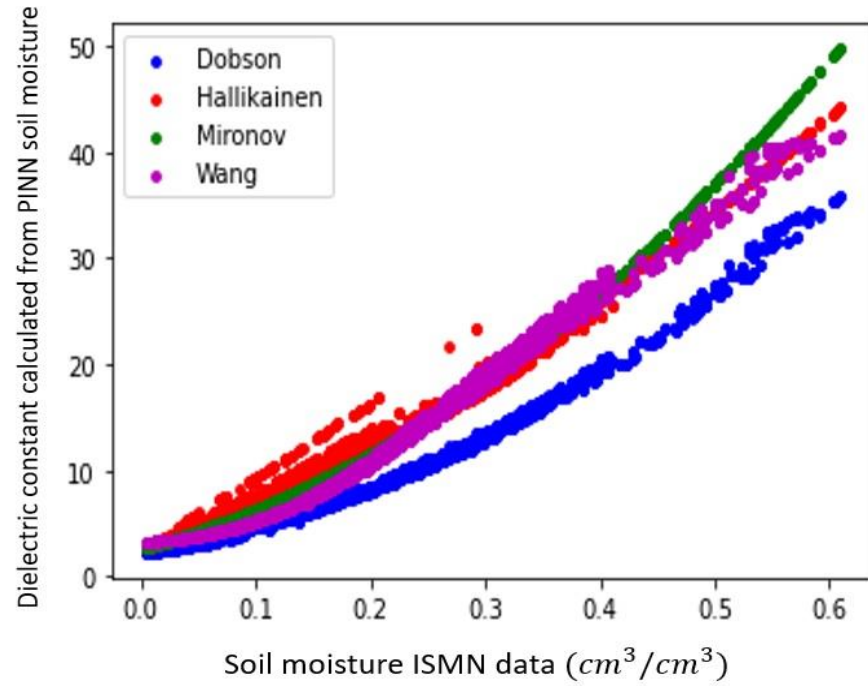


Figure 7.4. Dielectric constant calculated from soil moisture retrievals from PINN relationship with soil moisture reference ISMN data for 2020.

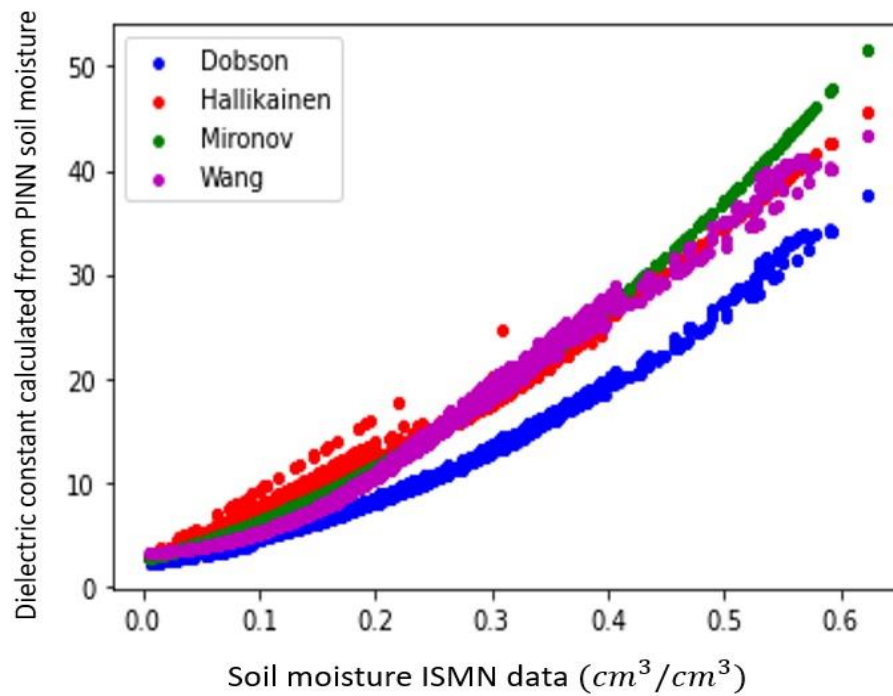


Figure 7.5. Dielectric constant calculated from soil moisture retrievals from PINN relationship with soil moisture reference ISMN data for 2021.

## **7.2 Investigation of different soil dielectric models' performance for different soil texture classes**

When using the PINN algorithm, soil texture data used is different depending on which soil dielectric constant model is applied. Due to the difference in soil texture data used in soil dielectric constant models, studying models' performance under different soil texture scenarios is worthwhile. For this reason, 4 soil texture classes were defined to examine performance of four soil dielectric constant models particularly (see Figure 7.6). Soil texture classes defined are as follows (see Table 7.4), i) Class 1 with sand percentage lower than 50% and clay percentage lower than 20% with 2174 data points, ii) Class 2 with sand percentage lower than 50% and clay percentage higher than 20% with 6469 data points, iii) Class 3 with sand percentage higher than 50% and clay percentage lower than 20% with 5882 data points, iv) Class 4 with sand percentage higher than 50% and clay percentage higher than 20% with 2280 data points. This classification enables us to differentiate the Dobson and Hallikainen model performance, which were developed on data with sand percentages lower than 50% with Mironov and Wang model. Figure 7.6 shows 4 different soil texture classes regarding the USDA soil texture classification triangle. Table 7.5 shows clay and sand percentages for 18 stations used in the processing and their assigned soil texture classes with respect to soil texture classification. Clay percentage of data ranges from 0.05 to 0.46 and sand percentage of data ranges from 0.29 to 0.89. Four classes are defined to investigate soil dielectric models' performance for different soil texture classes. Table 7.5 demonstrates which stations are assigned with each soil texture class.

Table 7.4. Classes defined for soil dielectric constant model's assessment.

Category	Sand percentage	Clay percentage	Sample number
<b>Class 1</b>	Lower than 50%	Lower than 20%	2174
<b>Class 2</b>	Lower than 50%	Higher than 20%	6469
<b>Class 3</b>	Higher than 50%	Lower than 20%	5882
<b>Class 4</b>	Higher than 50%	Higher than 20%	2280

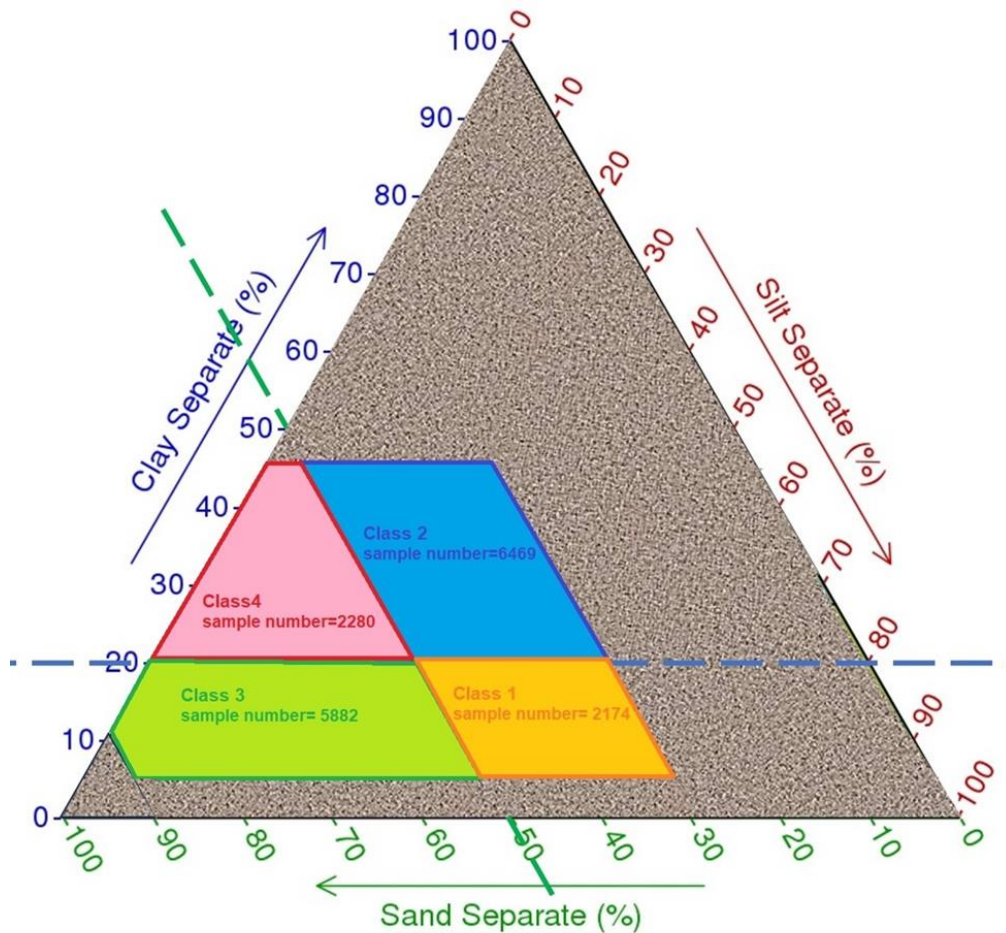


Figure 7.6. Four soil texture classes are defined to investigate soil dielectric constant model's performance. These four classes are i) Class 1 with sand percentage lower than 50% and clay percentage lower than 20% with 2174 data points (shown in orange), ii) Class 2 with sand percentage lower than 50% and clay percentage higher than 20% with 6469 data points (shown in blue), iii) Class 3 with sand percentage higher than 50% and clay percentage lower than 20% with 5882 data points (shown in green), iv) Class 4 with sand percentage higher than 50% and clay percentage higher than 20% with 2280 data points (shown in red).



Table 7.5. ISMN stations used in comparison of soil dielectric models' performance in PINN algorithm.

<b>ISMN sites (SCAN)</b>	<b>Clay %</b>	<b>Sand %</b>	<b>Bulk density</b>	<b>Soil texture</b>	<b>Classification</b>
<b>1. Adams Ranch</b>	0.104569	0.517212	1.460556	Loam	Class 3
<b>2. Alcalde</b>	0.130993	0.618444	1.464444	Sandy Loam	Class 3
<b>3. Alkali Mesa</b>	0.058642	0.873457	1.38	Sand	Class 2
<b>4. Beasley Lake</b>	0.0998611	0.508195	1.3631524	Sandy Loam	Class 2
<b>5. Cochora Ranch</b>	0.216815	0.378059	1.47962964	Loam	Class 3
<b>6. Donkey Reservoir</b>	0.272009	0.392611	1.380988	Clay Loam	Class 2
<b>7. Fort Reno</b>	0.46	0.29	1.386579	Clay	Class 2
<b>8. Jornada Exp Range</b>	0.2647654	0.54390	1.0466667	Sandy Clay Loam	Class 1
<b>9. Manderfield</b>	0.05	0.89	1.5	Sand	Class 1
<b>10. Mccracken Mesa</b>	0.2983334	0.368889	1.4738271	Clay Loam	Class 3
<b>11. Navajo Whiskey CK</b>	0.21	0.39	1.31	Loam	Class 3
<b>12. Panguitch</b>	0.21	0.38	1.4	Loam	Class 3
<b>13. Reese Center</b>	0.153333	0.42	1.35	Loam	Class 2
<b>14. San Angelo</b>	0.228333	0.427333	1.35622227	Loam	Class 4
<b>15. Sevilleta</b>	0.314814	0.527833	1.4033333	Sandy Clay Loam	Class 2
<b>16. Shadow Mtns</b>	0.2300000	0.43000	1.35	Loam	Class 2
<b>17. Uapb Dewitt</b>	0.143457	0.470093	1.35	Loam	Class 4
<b>18. West Summit</b>	0.06	0.56	1.35	Sandy Loam	Class 2

The PINN model for four soil texture classes generated soil moisture values with RMSE, ubRMSE and correlation values as follows (see Tables 7.6 to 7.8 and Figures 7.7 to 7.9 and correlation plots can be found in Figures 7.10 to 7.13).

- a) Cass 1 resulted in following metrics for i) Dobson model with Pearson's correlation coefficient (R) of 0.90, RMSE of 0.056 and ubRMSE of 0.053, ii) Hallikainen model with R of 0.95, RMSE of 0.040 and ubRMSE of 0.040, iii) Mironov model with R of 0.75, RMSE of 0.079 and ubRMSE of 0.079, and iv) Wang model with R equal to 0.87, RMSE of 0.060 and ubRMSE of 0.059.

- b) Class 2 resulted in following performance metrics for i) Dobson model with Pearson's correlation coefficient (R) of 0.88, RMSE of 0.057 and ubRMSE of 0.057, ii) Hallikainen model with R of 0.84, RMSE of 0.063 and ubRMSE of 0.063, iii) Mironov model with R of 0.71, RMSE of 0.080 and ubRMSE of 0.080, and iv) Wang model with R of 0.81, RMSE of 0.069 and ubRMSE of 0.068.
- c) Class 3 resulted in following performance metrics for i) Dobson model with Pearson's correlation coefficient (R) of 0.94, RMSE of 0.048 and ubRMSE of 0.048, ii) Hallikainen model with R of 0.96, RMSE of 0.041 and ubRMSE of 0.040, iii) Mironov model with R of 0.85, RMSE of 0.069 and ubRMSE of 0.069, and iv) Wang model with R of 0.95, RMSE of 0.042 and ubRMSE of 0.041.
- d) Class 4 resulted in following performance metrics for i) Dobson model with Pearson's correlation coefficient (R) of 0.88, RMSE of 0.047 and ubRMSE of 0.047, ii) Hallikainen model with R of 0.96, RMSE of 0.045 and ubRMSE of 0.045, iii) Mironov model with R of 0.82, RMSE of 0.049 and ubRMSE of 0.048, and iv) Wang model with R of 0.94, RMSE of 0.046 and ubRMSE of 0.046.

Table 7.6. RMSE results of different soil dielectric constant models for four different soil texture classes.

RMSE	class1	class2	class3	class4
<b>Dobson</b>	0.056	0.057	0.048	0.047
<b>Hallikainen</b>	0.040	0.063	0.041	0.045
<b>Mironov</b>	0.079	0.080	0.069	0.049
<b>Wang</b>	0.060	0.069	0.042	0.046

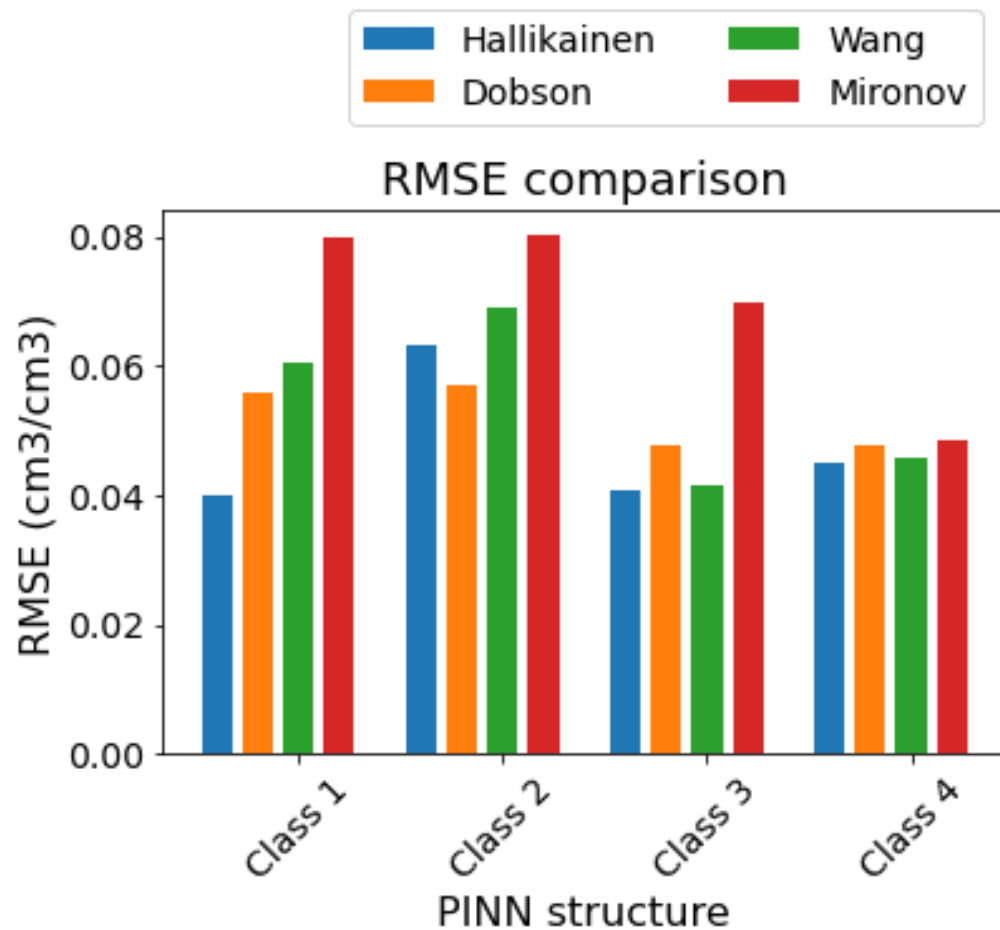


Figure 7.7. RMSE comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes.

Table 7.7. ubRMSE results of different soil dielectric constant models for four different soil texture classes.

ubRMSE	class1	class2	class3	class4
<b>Dobson</b>	0.053	0.057	0.048	0.047
<b>Hallikainen</b>	0.040	0.063	0.040	0.045
<b>Mironov</b>	0.079	0.080	0.069	0.048
<b>Wang</b>	0.059	0.068	0.041	0.046

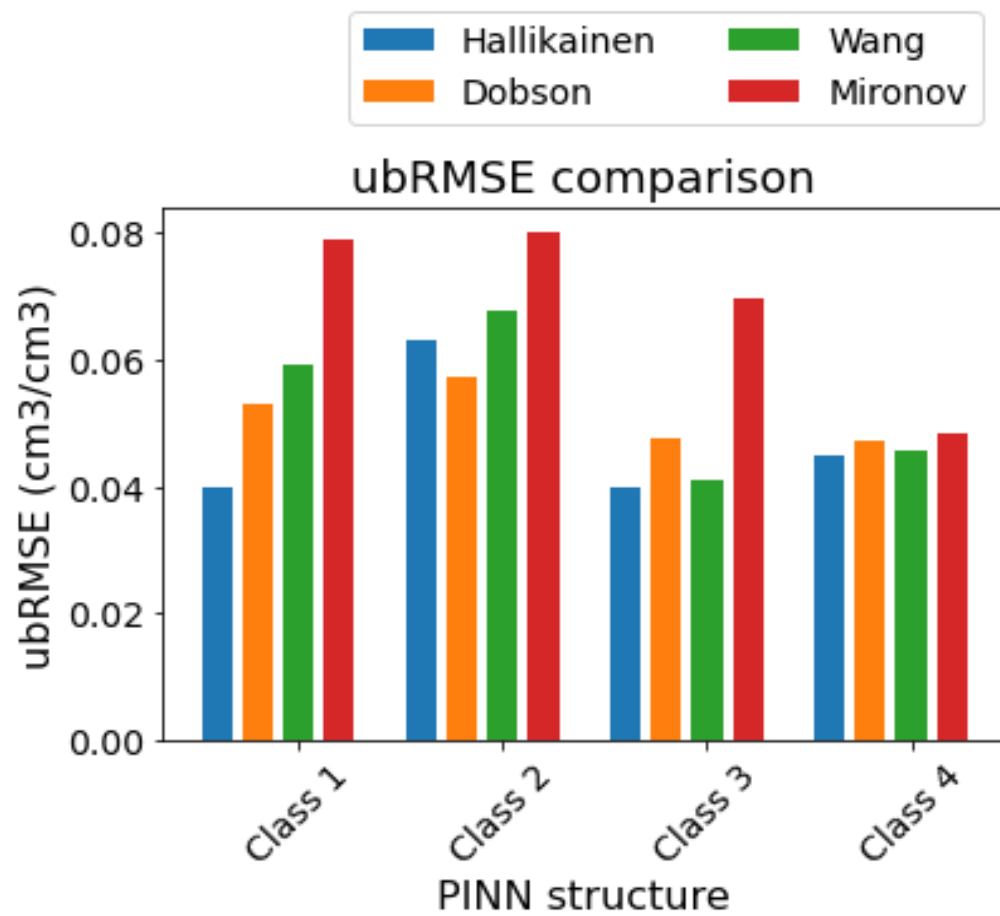


Figure 7.8. ubRMSE comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes.

Table 7.8. Pearson correlation coefficient results of different soil dielectric constant models for four different soil texture classes.

Correlation Coefficients	class1	class2	class3	class4
Dobson	0.90	0.88	0.94	0.88
Hallikainen	0.95	0.84	0.96	0.96
Mironov	0.75	0.71	0.85	0.82
Wang	0.87	0.81	0.95	0.94

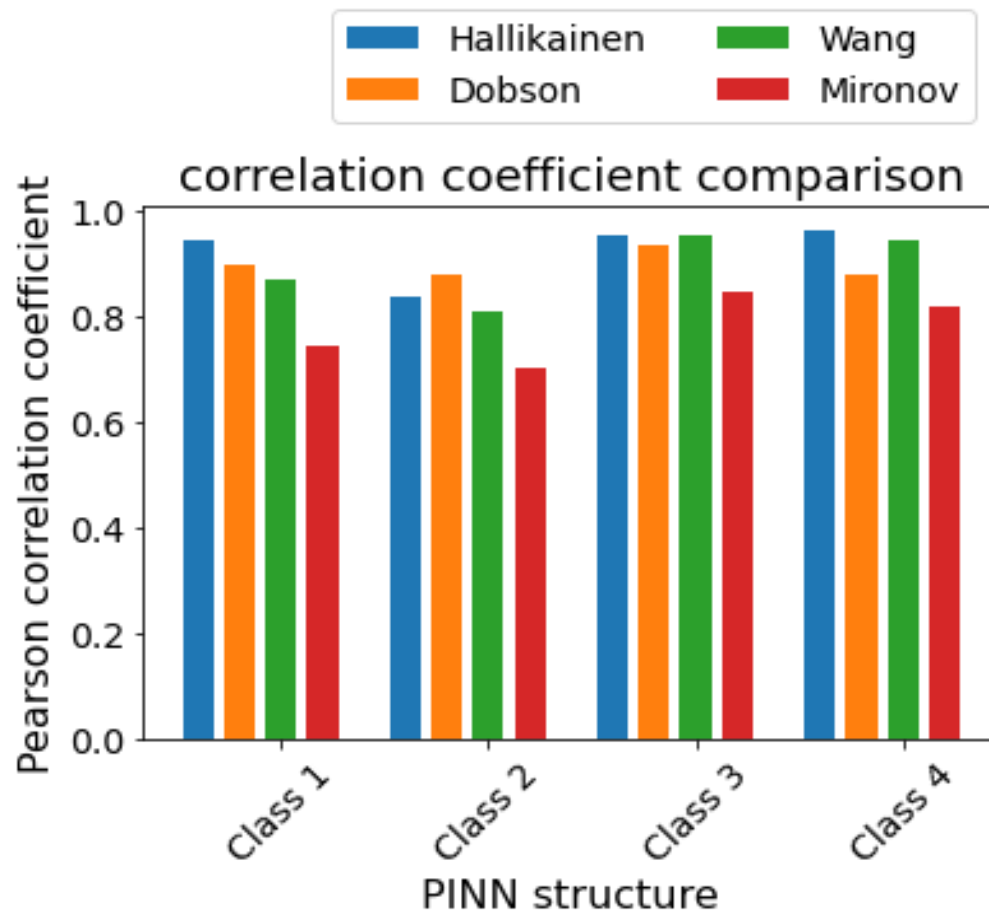


Figure 7.9. Pearson's correlation coefficients comparison of four soil dielectric constant models used in PINN algorithm for four defined soil texture classes.

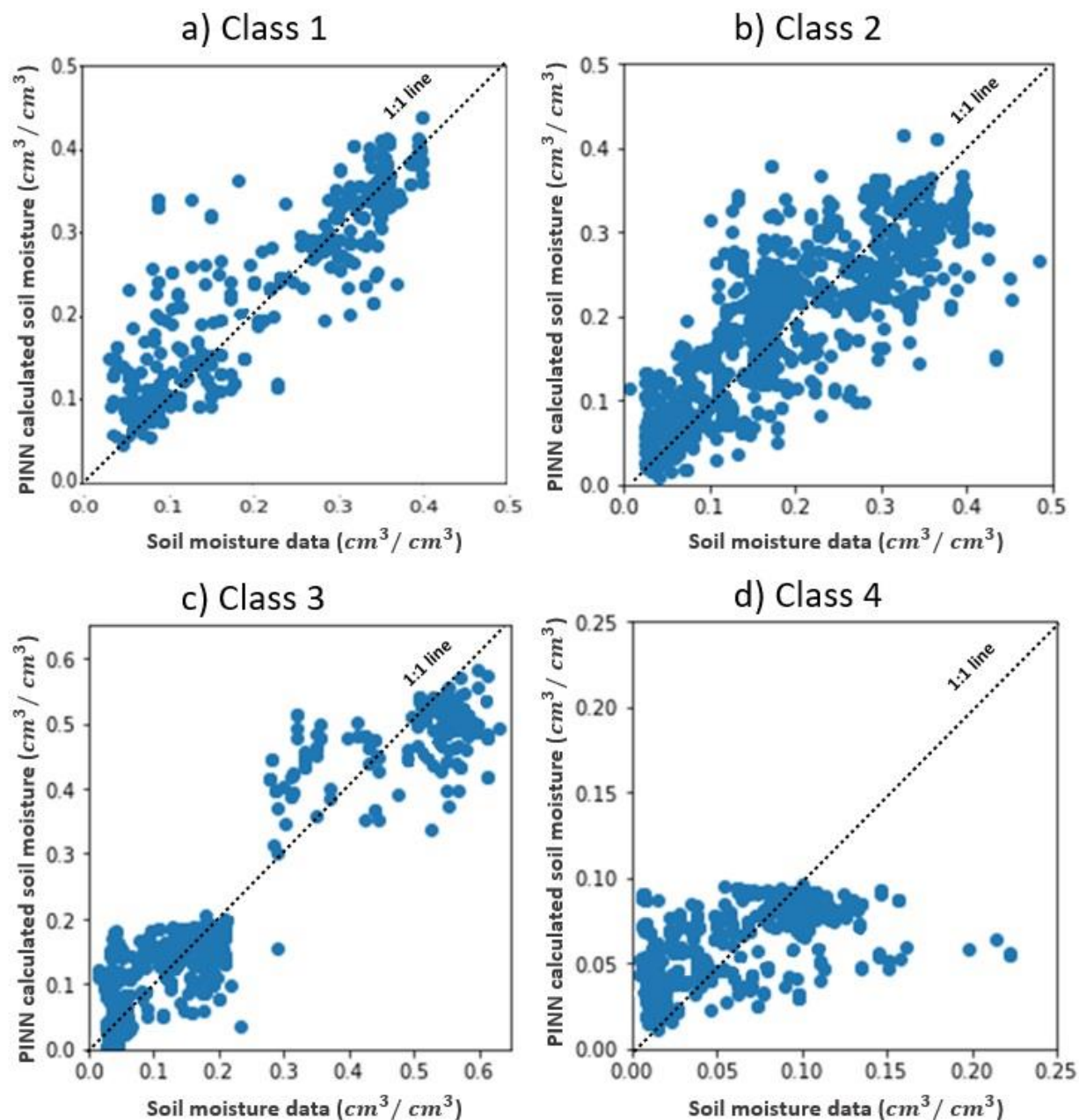


Figure 7.10. Correlation plots of PINN soil moisture retrievals with Dobson soil dielectric constant model for four different soil texture classes. Pearson's correlation coefficient values for class1, class2, class3 and class4 while using Dobson model are 0.90, 0.88, 0.94, and 0.88, respectively.

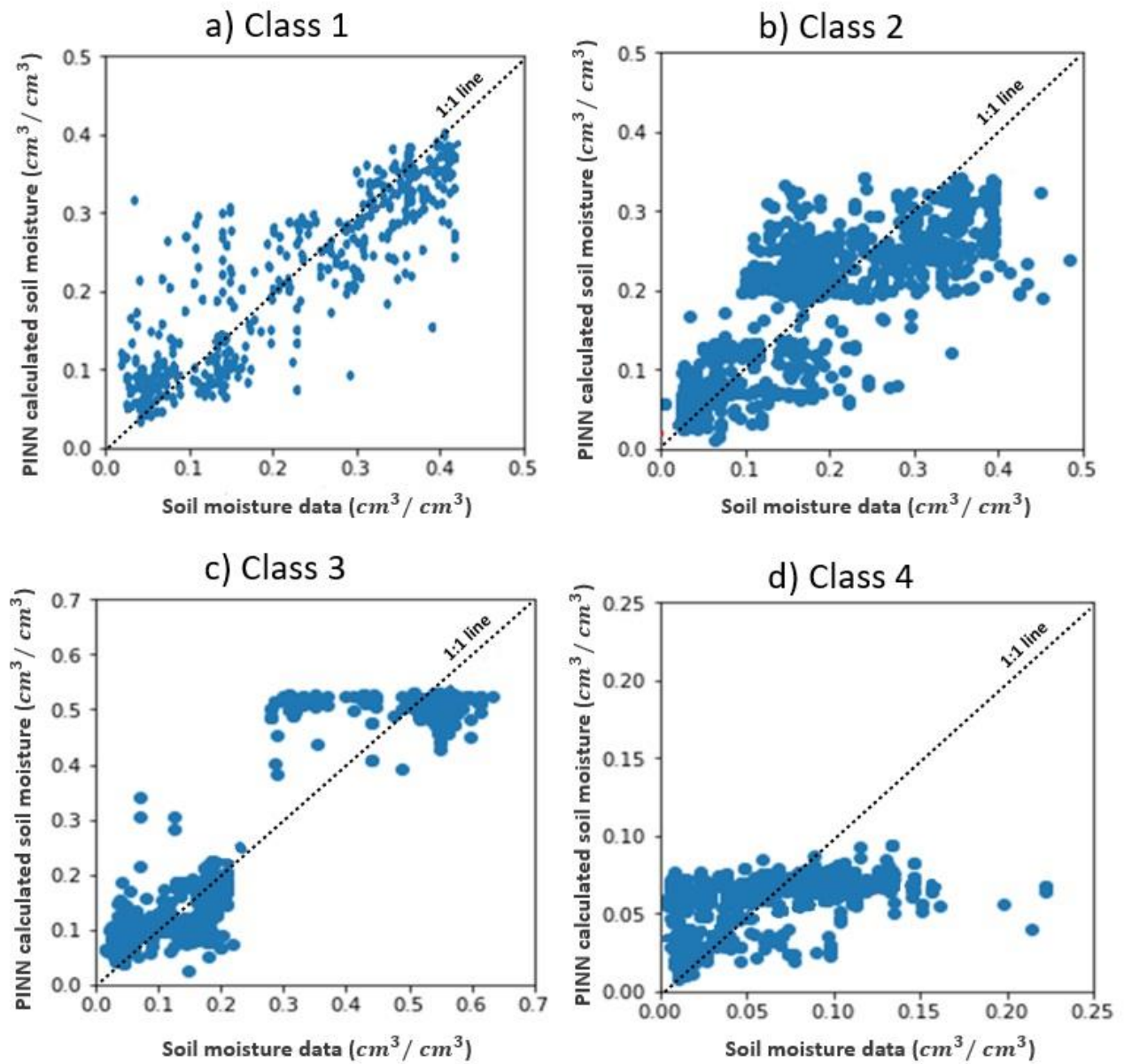


Figure 7.11. Correlation plots of PINN soil moisture retrievals with Hallikainen soil dielectric constant model for four different soil texture classes. Pearson's correlation coefficient values for class1, class2, class3 and class4 while using Hallikainen model are 0.95, 0.84, 0.96, and 0.96, respectively.

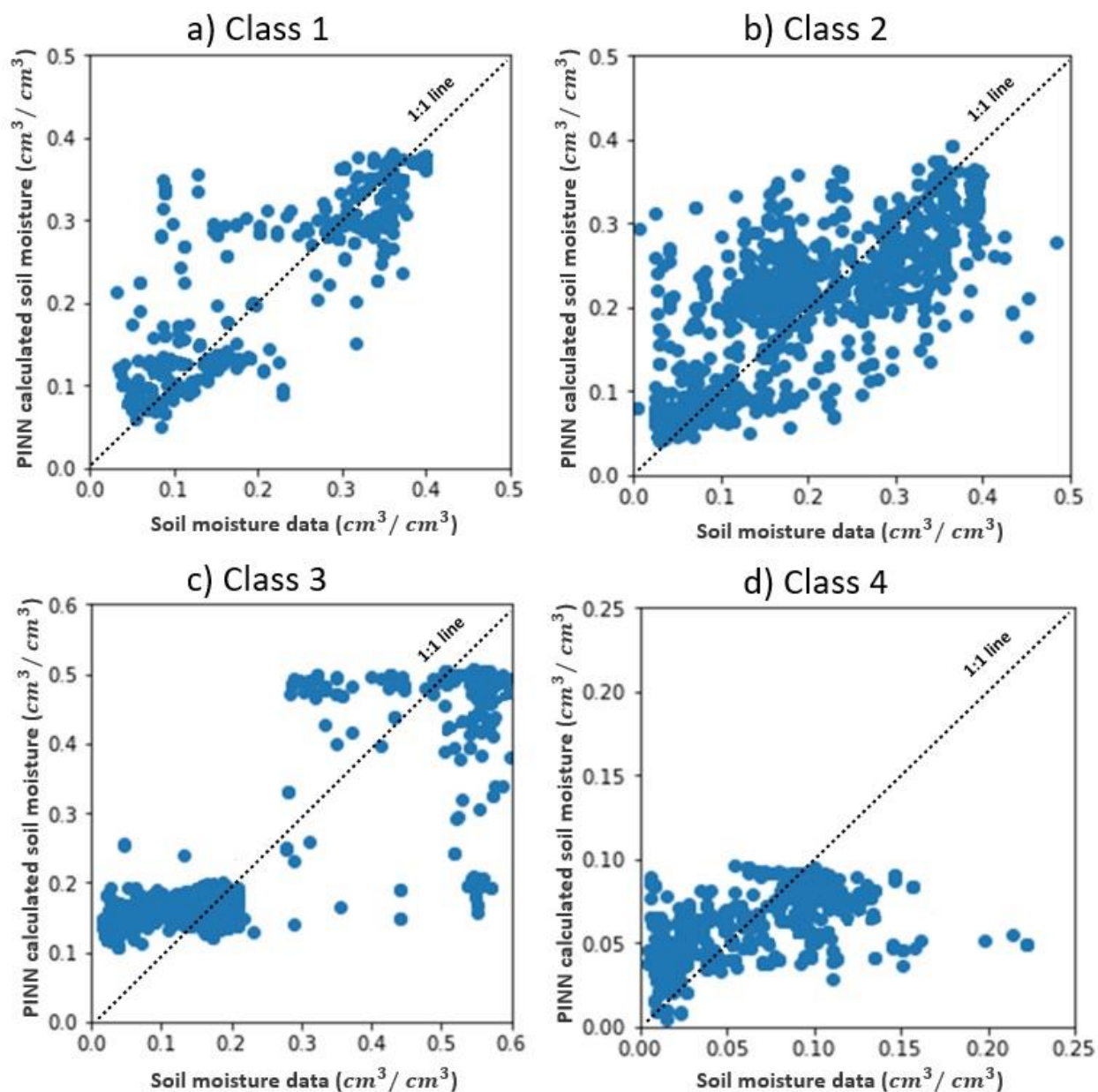


Figure 7.12. Correlation plots of PINN soil moisture retrievals with Wang soil dielectric constant model for four different soil texture classes. Pearson's correlation coefficient values for class1, class2, class3 and class4 while using Wang model are 0.87, 0.81, 0.95, and 0.94, respectively.



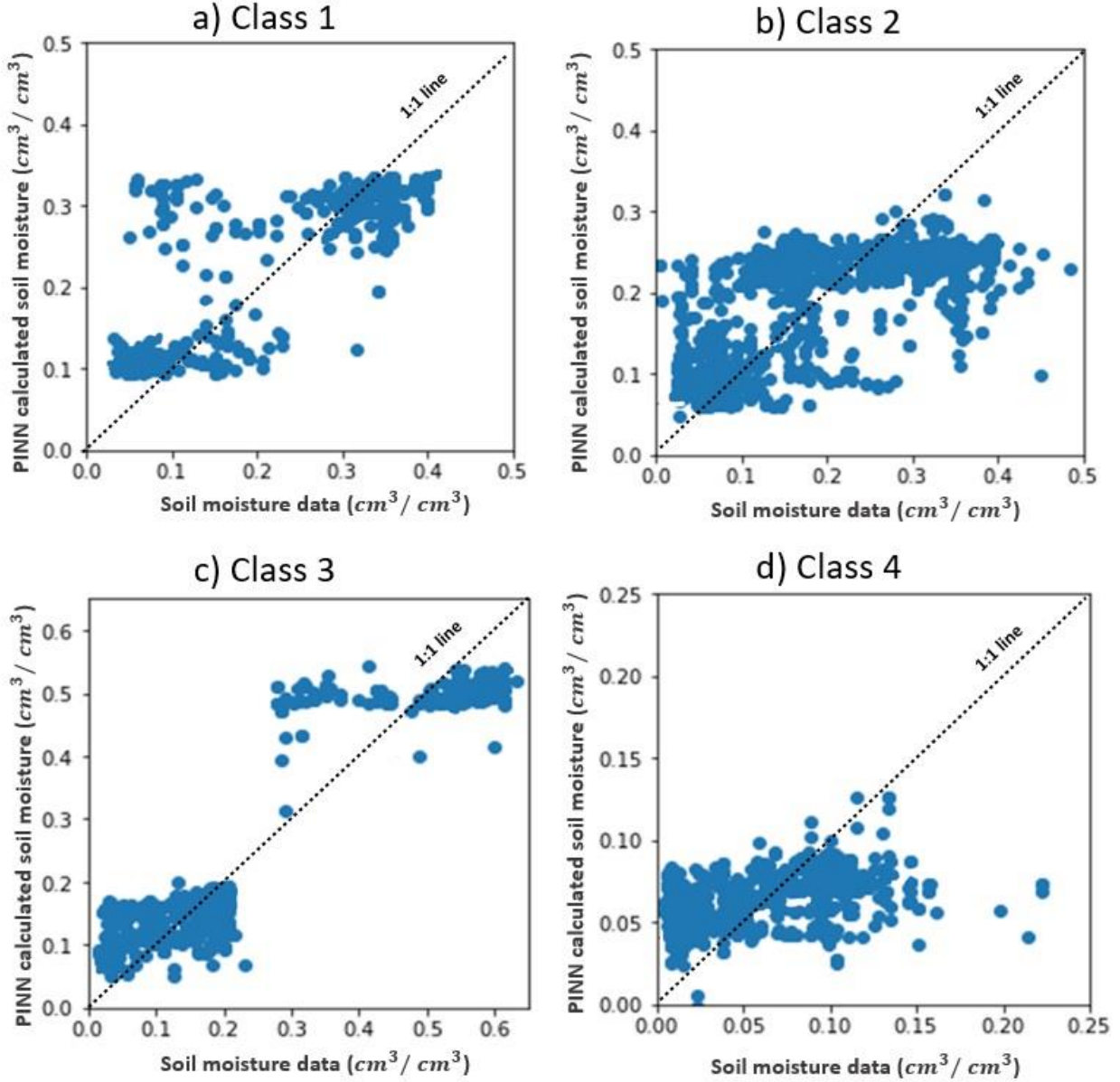


Figure 7.13. Correlation plots of PINN soil moisture retrievals with Mironov soil dielectric constant model for four different soil texture classes. Pearson's correlation coefficient values for class1, class2, class3 and class4 while using Mironov model are 0.75, 0.71, 0.85, and 0.82, respectively.

## 7.3 Discussion

Different soil dielectric constant models affected results of the PINN in soil moisture retrieval; therefore, the soil dielectric constant models play an important role in the physical relationship between soil moisture and reflected signal power. The RMSE distinction and correlation coefficient difference of the best model (Hallikainen) and worst model (Mironov) is around 0.0209 and 0.13, respectively (see Table 7.3). These differences can affect the results of soil moisture retrieval in some cases. For instance, SMAP mission requirement for the 36-km and 9-km soil moisture products is ubRMSE and RMSE of a maximum of 0.04, and 0.052, respectively (Colliander et al. 2017; O'Neill et al. 2021). Therefore, working with PINN algorithm, the Mironov model with RMSE of 0.0687 and ubRMSE of 0.0685 cannot be used while the Hallikainen model with RMSE of 0.0478 and ubRMSE of 0.0473 is sufficient to be used. For GNSS-R missions, higher values of RMSE and ubRMSE have been accepted (Eroglu et al. 2019; Senyurek et al. 2020; Jia et al. 2021) since both CYGNSS and GNSS missions were not designed for soil moisture retrieval initially and calibration of data for soil moisture studies is required to improve results. Still, ubRMSE of CYGNSS-derived soil moisture products compared with SMAP soil moisture is around  $0.045 \text{ cm}^3/\text{cm}^3$  after calibration of CYGNSS reflectivity data (Chew and Small 2020). Consequently, choice of the soil dielectric constant model impacts the overall results of PINN and in case of low RMSE and ubRMSE mission requirements (around 0.05) the Mironov model with PINN algorithm is not suggested while other models can be used with PINN.

The choice of a suitable soil dielectric constant model in PINN is therefore recommended for better results. Since all soil dielectric constant models were developed on in-situ

experimental data, understanding the reasons for differences in the models' performance can be helpful in development of new soil dielectric constant models for satellite remote sensing data especially GNSS-R applications. As shown in Table 7.1, the only difference between different soil dielectric models are soil texture data and soil temperature data. However, models were developed on different reference data as well. It was shown that these dissimilarities in models can affect the overall result of PINN depending on which soil dielectric model is used.

The Mironov model has presented the least capability in predicting soil moisture values due to the following reasons: i) Mironov model used in the PINN algorithm only considers clay percentage data for soil texture and ignores the effect of sand percentage, bulk density, and soil temperature. Therefore, the effect of sand percentage as well as the clay percentage must be included in the soil dielectric constant modeling as removing it from Mironov model worsens the results, ii) The Mironov model is a less empirical model and is the strictest model in comparison to other three models. The integration of physical relations into the architecture of neural networks in PINN must be applied with care since strict physical analytical relations may overly constrain the neural network training in the PINN on condition that the weight of physical relationships is very high. Mironov model is a semi-empirical model in which most of the model includes rigid analytical expressions that narrow training process more compared with other models, and iii) Different Mironov coefficients (explained in Subsection 3.2.2) such as refractive indexes  $n_m, n_d, n_b, n_u$ , and normalize attenuation coefficients  $k_m, k_d, k_b, k_u$  are all calculated from clay percentage data which results in calculation cumulative errors as soil moisture values get updated in

each step of training. In each step of training these coefficients must be calculated which increases loss function.

The Hallikainen model has shown the best results in soil moisture predictions due to following facts: i) The model is the simplest amongst other models and PINNs training is not limited by rigidity of the physics model, and ii) Less calculation cumulative errors are added in each step of training as Hallikainen model directly calculates dielectric constant from soil moisture and soil texture data without introducing secondary coefficients in model. The other two soil dielectric constant models (Dobson and Wang), both include calculation of secondary coefficients which with comparison to Hallikainen increases the cumulative error in each step of training. After calculation of dielectric constant values with PINN, all models have shown consistent behaviour; however, the Dobson model has the largest soil moisture estimate deviations, especially in medium to high soil moisture values due to lack of distinction between bound and free water in Dobson modelling.

In Section 7.2, a more detailed investigation for different soil dielectric constant models has been studied with definition of four soil texture classes. For class 1 and class 2 with sand percentages lower than 50%, Hallikainen and Dobson models have the best results. Because Hallikainen and Dobson models were developed on datasets with sand percentages lower than 50%. In addition, for class 3 and class 4 with sand percentages higher than 50%, Wang model has better results with comparison to Dobson model since Wang model was originated on datasets with higher ranges than Dobson and not only sand percentages lower than 50%. This shows the importance of suitable reference databases on soil dielectric constant modeling development. Since these models originated from limited numbers of soil data field, development of a new soil dielectric

constant model for GNSS-R applications is worthwhile. Still, the Mironov model has the worst results for all classes due to limited soil texture data and more strict physics relationships.

## **8. CONCLUSIONS AND FUTURE WORK**

This chapter provides the summary of the dissertation, as well as future work that can be continued in this research, as the proposed method for soil moisture retrieval still needs further improvements.

### **8.1 Conclusions**

Soil moisture is an essential variable for understanding natural processes such as hydrology and meteorology. As reflected GNSS signals can be used as a remote sensing tool for soil moisture studies, the development of a model connecting soil moisture to reflected signal using PINN algorithm is proposed in this research. A sufficient GNSS-R soil moisture retrieval model should be capable of accounting for effective soil surface parameters such as vegetation, soil texture, surface roughness, and soil temperature. This dissertation proposed using a learning algorithm called the physics-informed neural network (PINN) for GNSS-R soil moisture retrieval considering all surface parameters impacting reflected signal.

The use of reflected GNSS signals as sources of opportunity has been added to traditional remote sensing systems as a free, ubiquitous, and available technique. Reflected GNSS signals are related to more soil parameters than soil moisture alone. The relations between land surface parameters, soil moisture, and reflectivity are modelled through a PINN model. This study finds that PINN is capable of soil moisture estimation using GNSS-reflected signals in less data regimes due to additional physics knowledge. Therefore, the application of PINN is studied in this research as a new emerging approach

to model the relationship between soil moisture and GNSS-R signals offering: i) the addition of physics knowledge in the training process of NN, ii) convergence with less data, and iii) more investigations on different physical models impacts on soil moisture retrieval.

The PINN algorithm has been designed to predict soil moisture estimates using both NN regression and inclusion of physical dependencies of soil and land surface parameters affecting the reflected signals. PINNs are useful for a wide range of tasks in computational physics and engineering (Raissi et al. 2019; Mao et al. 2020; Cai et al. 2021; Guo et al. 2023), where the goal is to model and understand physical systems using both data and knowledge of the underlying physics. They have the advantage of being flexible and expressive, while also ensuring that the solution is physically meaningful and consistent with known laws of physics. The PINN model still uses NN regression as part of the algorithm; thus, relevant datasets to fulfill both physical laws and NN were used. PINN uses two loss functions, one related to soil moisture calculated from NN regression and the other related to inversely calculated reflectivity calculated from physical relations.

The main challenge in data preparation for PINN was bulk data processing of different datasets. For this reason, a free clouding platform called Google Earth engine was used. One of the main advantages of PINN over NN is its ability to converge over small data regime. In PINN, the addition of physics-based laws can improve generalization or regularize training. As suggested by previous works (Raissi et al. 2017; Karniadakis et al 2021; Muther et al. 2022), PINN can be trained on less available data due to additional information acquired by imposing physical constraints. Here, PINN also has shown promising results with quite a smaller number of stations compared to some previously

published works that only used NN. In spite of the fact that the number of stations were low, PINN has been able to generate soil moisture estimates in an acceptable RMSE range with correlation coefficients higher than 80. PINN has been able to generate daily soil moisture estimates with a root mean squared error (RMSE) of  $0.0547 \text{ cm}^3 / \text{cm}^3$ , and Pearson's correlation coefficient of 0.82 which is an improvement from RMSE of  $0.0707 \text{ cm}^3 / \text{cm}^3$  and Pearson's correlation coefficient of 0.78 for the underlying NN model due to adding physical models. Imposed physical relations in PINN have improved RMSE from  $0.0707 \text{ cm}^3 / \text{cm}^3$  to  $0.0547 \text{ cm}^3 / \text{cm}^3$ . The performance of PINN is compared with NN for the same data sets (see Table 6.2) in Section 6.1. As a result of the addition of physical modeling, PINN has improved upon NN despite relatively low number of stations. PINN has been able to provide promising soil moisture estimates with CYGNSS data with a more physically robust model in less data regime. Since RMSEs of not higher than around 0.052 are required for L-band microwave satellite-based soil moisture products (O'Neill et al. 2021), the 2% difference between NN's RMSE and PINN's RMSE can be substantial.

PINN results also remain stable while changing NN structure. The RMSE and correlation coefficient variations after switching from a two-layer to a three-layer NN are around 0.0047 and 0.03 for PINN, respectively. PINN metric variations are lower than NN metric variations when changing NN layers. For instance, the correlation coefficient of NN has changed around 0.13 after switching from a two-layer NN to three-layer NN which is almost four times higher than PINN correlation coefficient changes. In addition, the RMSE of NN has changed around 0.067 after switching from a two-layer NN to a three-layer NN which is almost 1.4 times higher than PINN RMSE changes. Changing the number of



layers affects NN more than PINN since the PINN model is more robust in comparison with NN. In addition, PINN results remain steady after changing train test ratios. The addition of physics relations to NN strengthens PINN design to have more stable results. Subsection 6.1.1 contains PINN results compared to NN for different structures. PINN results remain stable after changes in number of layers and NN structure.

PINN is also sensitive to different land covers and physical characteristics. Three land cover categories including grasslands, mixed grassland and cropland, and croplands were used to inspect PINN performance depending on land cover properties. Because physical models used in PINN follow coherent reflection conditions, as incoherency characteristics such as vegetation increase, PINN results worsen. Therefore, the best results are for the grasslands that have less vegetation, and the worst results are related to croplands with the highest vegetation.

As a consequence, by incorporating physical principles and governing laws into neural networks in this research, PINN offers following:

- Physically and scientifically consistent GNSS-R soil moisture retrieval models.
- Increase data efficiency as models can be trained with less data compared to neural networks owing to inclusion of physical relationship.
- Improve steadiness of models as PINN remains cohesive even after changing NN structures.
- Enhance interpretability, for example relating results to physical properties of soil such as land cover.

One of the most important parts of physical dependencies between soil moisture and signal power is soil dielectric constant models. Soil dielectric constant models reflect the

relationship between soil moisture and soil permittivity. Multiple soil dielectric constant models exist as discussed in section 3.2. Four soil dielectric constant models used in this research are Dobson, Hallikainen, Mironov, and Wang models. The PINN model was tested with all these four soil moisture dielectric constant models. The Hallikainen model has the best results with Pearson's correlation coefficient (R) of 0.94, RMSE of 0.0478 and ubRMSE of 0.0473. The Mironov model has the worst results with Pearson's correlation coefficient (R) of 0.81, RMSE of 0.0687 and ubRMSE of 0.0685. Wang and Dobson models have the second and third place, respectively. The Wang model has Pearson's correlation coefficient (R) of 0.92, RMSE of 0.0509 and ubRMSE of 0.0499 and the Dobson model has Pearson's correlation coefficient (R) of 0.86, RMSE of 0.0568 and ubRMSE of 0.0560. The RMSE and correlation coefficient variance between the best model (Hallikainen) and the worst model (Mironov) are around 0.0209 and 0.13, respectively (see Table 7.3). These differences while applying different soil moisture dielectric constant models confirm the importance of choosing an appropriate soil moisture dielectric constant model in PINN modelling. For example, when using PINN algorithm, the Mironov model with RMSE of 0.0687 and ubRMSE of 0.0685 is not recommended, while Hallikainen model with RMSE of 0.0478 and ubRMSE of 0.0473 is sufficient to be used. This is due to the accepted ranges for RMSE and ubRMSE for soil moisture retrieval application (Colliander et al. 2017). As an example, SMAP mission requirements for 36-km and 9-km soil moisture products require ubRMSE and RMSE of no more than 0.04, and 0.052, respectively (O'Neill et al. 2021). For GNSS-R mission, there has been some discussion regarding the acceptable values of RMSE and ubRMSE (Eroglu et al. 2020; Senyurek et al. 2020; Jia et al. 2021) which accepts higher values of

RMSE and ubRMSE. In CYGNSS and GNSS missions, soil moisture retrieval was not planned initially, so calibration of data for soil moisture studies is required in order to improve results and RMSEs as low as 0.04 may not be met. Therefore, RMSEs higher than 0.04 have also been accepted (Eroglu et al. 2020; Senyurek et al. 2020).

Section 7.1 explains that Hallikainen model surpasses other models due to less rigidity and less cumulative errors. The Hallikainen model is the simplest model and in contrast to other models does not include secondary defined coefficients. For models like Mironov, many coefficients are included in the model. These refractive coefficients need to be calculated in each step of training leading to cumulative errors. Also, the Mironov model only considers clay percentage data and ignores sand percentage and soil temperature. Consequently, sand percentage and soil temperature should be included in modelling.

In Section 7.2, different soil texture classes are defined to understand different soil dielectric constant model's behaviour in PINN algorithm and check PINN's sensitivity to soil texture as a physical property. These classes of soil texture are as follows: 1) Class 1 consists of sands at or below 50% and clay at or below 20%, 2) Class 2 consists of sands at or below 50% and clays at or above 20%, iii) Class 3 has a sand percentage of over 50% and a clay percentage of over 20%, and iv) Class 4 has a sand percentage of over 50% and a clay percentage of over 20%. Still, the Hallikainen model has the best results and Mironov has the worst results. However, for class 1 and 2 Dobson model excels Wang model. The Dobson model was developed on datasets with sand percentages lower than 50%. Owing to the fact that class 1 and class 2 soil texture classes also have soil sand percentages lower than 50%, the Dobson model has a better

outcome than Wang model for these two classes. However, for class 3 and class 4 Wang model has better results with comparison to Dobson models.

Although PINN has been able to generate daily soil moisture estimates in an incomplete model (Uncertainties in semi-empirical models), imperfect data and small data regime, enhancement of PINN model needs further studies in near future. Consequently, the proposed PINN algorithm should be further developed as understanding between physics relations of parameters improves. Subsequently, modifying soil moisture dielectric constant models to improve uncertainties in physics of problem can be contemplated for future work. Some future works suggestions can be found in the following section.

## **8.2 Future work**

PINN algorithm has the following advantages in soil moisture studies using CYGNSS data: (i) PINN model has been able to successfully estimate soil moisture using CYGNSS-based reflectometry data with less data than previous research owing to physics dependencies of parameters. (ii) The PINN model is sensitive to physical parameters such as land cover and vegetation, soil dielectric constant models, and their soil texture parameters. (iii) The PINN model is a scientifically consistent GNSS-R soil moisture retrieval model due to embedding physical constraints between parameters. And iv) PINN models' performance remains sturdy with different structures of NN due to additional physics.

Despite the fact that PINN has presented promising results, the PINN algorithm suffers from the following disadvantages: (i) The PINN is a combination of both NN and physics models; hence, a strong NN structure is still needed. Thus, all practices related to NN

optimization such as optimization function, loss function choice and more still should be done. These practices can be time-consuming, and the final choice is not always the best suggestion. (ii) The PINN is dependent on initial soil moisture data like NN algorithms. High numbers of soil moisture data are still needed as reference data. (iii) The proposed method was only applicable for coherent reflections, as physics models respectful to coherent reflections have been used. The PINN algorithm deserves further improvements for more diverse stations in terms of land cover and vegetation. And (iv) The physics model for inverse calculation of reflectivity is dependent on experimental soil moisture-dielectric constant models which were originated based on the in-situ measurements of microwave frequency waves rather than a satellite remote sensing technique. Because soil dielectric constant models are not customized for GNSS-R, the process consistently contains unexpected errors in conversion. Empirical parameters of soil moisture dielectric constant models should be modified in GNSS-R applications, especially in PINN algorithm.

With consideration of PINN algorithm advantages and disadvantages, future work is required to improve the current PINN model's performance in the following directions.

- i) The PINN algorithm has a good performance for ISMN stations, while its performance for larger areas is unclear. The generalization ability of PINN model needs to be further investigated for larger number of stations and global scale. The proposed methodology is only applicable to coherent reflection conditions due to the coherent reflection assumption. The effects of other error sources are reduced, e.g., resolution error in stations with non-homogenous land cover over 4km x 4km grid and complexity of electromagnetic physics models in incoherent

reflections in our PINN methodology with the coherent reflection assumption. This simplification enables PINN accuracy assessment without concern for additional possible error sources especially in incoherent reflection cases. In addition, as suggested by (Chew et al. 2016; Nghiem et al. 2017) most of reflections from land are coherent. However, the application of PINN should be further studied for reflections with incoherent characteristics. Since our physics model obeys coherent reflection rules, stations with high vegetation and roughness that have more incoherent characteristics in their reflections were not added and studied in this research.

- ii) The PINN algorithm consists of both NN regression and physical models. As physics relationships in GNSS-R modeling develop, the PINN model can be improved in further studies. Development of electromagnetic relations for more incoherent characteristics has been challenging in GNSS-R community. As physics understanding and models between soil moisture and dielectric constant improves, the PINN model can become stronger and more ubiquitous. Development of PINN algorithms after improvement of physical relationships can be further studied as knowledge in physical relationship between soil moisture and GNSS-R signal enhances.
- iii) The main benefit of PINN is the addition of physics laws because neural networks neglect analytical/empirical physics relationships. PINN reduces the search space and convergence to an answer faster due to additional physics. However, convergence is not always improved by addition of physics relations in PINN. Inclusion of physics in PINN may restrict the training process. While addition of

physics analytical equations decreases the space of acceptable solutions, it also narrows the training of the network naturally (Rohrhofer et al. 2023). Training a physics-informed neural network is a multi-objective optimization problem and conflicting objectives is one of the main issues complicating training (Rohrhofer et al. 2023; Maddu et al. 2022). Scale imbalances between different loss functions can cause conflicts and worsen convergence. Different methods have been proposed to solve convergence and training issues in PINN research. For example, weighting loss functions either manually or with adaptive weights, which adjust weights for different loss functions during training (Cai et al. 2021; Jin et al. 2021; Maddu et al. 2022) or working with more smooth activation functions such as self-scalable activation function for scale imbalances (Gnanasambandam et al. 2022). Choosing suitable weights for different loss function objectives can control stiffness of physics involved and improve PINN performance. Since this research was the first practice of PINN in CYGNSS soil moisture estimation, to the author's knowledge, the application of PINN was explored and tested without including extra improvement strategies such as weighted sum of loss functions. Here, reflectivity loss function coming from physical dependencies of parameters inherently has a lower weight than soil moisture loss function coming from NN regression balancing the stiffness of physical model; however, it is possible to investigate different weights for physics related loss functions and explore their results.

- iv) Soil moisture dielectric constant models were developed from a small number of soil fields with in-situ measurements. In Chapter 7, it is shown that Hallikainen and

Dobson models have the best results for datasets with sand percentages lower than 50% while Wang model has better results for datasets with sand percentages higher than 50% in comparison with Dobson model. This illustrates the importance of appropriate reference databases when developing soil dielectric constant models. Because Dobson model was developed on soil data with sand percentages lower than 50% and Wang model was developed on a larger dataset. GNSS-R soil moisture studies applications can benefit from a new soil dielectric constant model developed from larger satellite remote sensing data rather than few in-situ soil data fields.

- v) Both empirical and semi-empirical soil moisture dielectric constant models were originated for different purposes other than GNSS-R applications. Uncertainty and errors in these models' empirical parameters are constantly added to PINN algorithm. Soil moisture dielectric constant models need to be modified by modification of empirical parameters with GNSS-R data for reflectometry applications. Soil moisture products from GNSS-R data can be considered as a new source to calibrate and modify soil moisture dielectric constant models.



## REFERENCES

Ali, Iftikhar, Felix Greifeneder, Jelena Stamenkovic, Maxim Neumann, and Claudia Notarnicola. "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data." *Remote Sensing* 7, no. 12 (2015): 16398-16421.

Al-Khaldi, Mohammad M., Joel T. Johnson, Andrew J. O'Brien, Anna Balenzano, and Francesco Mattia. "Time-series retrieval of soil moisture using CYGNSS." *IEEE Transactions on Geoscience and Remote Sensing* 57, no. 7 (2019): 4322-4331.

Amani, Meisam, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, S. Mohammad Mirmazloumi, Sayyed Hamed Alizadeh Moghaddam et al. "Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020): 5326-5350.

Asgarimehr, Milad, Caroline Arnold, Tobias Weigel, Chris Ruf, and Jens Wickert. "GNSS reflectometry global ocean wind speed using deep learning: Development and assessment of CyGNSSnet." *Remote Sensing of Environment* 269 (2022): 112801.

Atterberg, Albert. *Die rationelle klassifikation der sande und kiese*. 1905.

Aubert, Maëlle, N. Baghdadi, Mehrez Zribi, A. Douaoui, C. Loumagne, Frédéric Baup, Mohammad El Hajj, and Sébastien Garrigues. "Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust." *Remote Sensing of Environment* 115, no. 8 (2011): 1801-1810.

Baker, J. M., and R. R. Allmaras. "System for automating and multiplexing soil moisture measurement by time-domain reflectometry." *Soil Science Society of America Journal* 54, no. 1 (1990): 1-6.

Bass, Fridrikh Gershonovich, and Iosif Moiseevich Fuks. *Wave Scattering from Statistically Rough Surfaces: International Series in Natural Philosophy*. Vol. 93. Elsevier, 2013.

Baver, Leonard David, Walter H. Gardner, and Wilford R. Gardner. *Soil physics*. No. 04; S592. 3, B38 1972. New York: Wiley, 1972.

Beckmann, Petr, and Andre Spizzichino. "The scattering of electromagnetic waves from rough surfaces." *Norwood* (1987).

Bengio, Yoshua. "Practical recommendations for gradient-based training of deep architectures." *Neural Networks: Tricks of the Trade: Second Edition* (2012): 437-478.

Brocca, Luca, Luca Ciabatta, Christian Massari, Stefania Camici, and Angelica Tarpanelli. "Soil moisture for hydrological applications: Open questions and new opportunities." *Water* 9, no. 2 (2017): 140.

Brocca, Luca, R. Morbidelli, F. Melone, and Tommaso Moramarco. "Soil moisture spatial variability in experimental areas of central Italy." *Journal of Hydrology* 333, no. 2-4 (2007): 356-373.

Cai, Shengze, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. "Physics-informed neural networks (PINNs) for fluid mechanics: A review." *Acta Mechanica Sinica* 37, no. 12 (2021): 1727-1738.

Cai, Shengze, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. "Physics-informed neural networks for heat transfer problems." *Journal of Heat Transfer* 143, no. 6 (2021).

Camps, Adriano, Hyuk Park, Miriam Pablos, Giuseppe Foti, Christine P. Gommenginger, Pang-Wei Liu, and Jasmeet Judge. "Sensitivity of GNSS-R spaceborne observations to soil moisture and vegetation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, no. 10 (2016): 4730-4742.

Camps, Adriano, and Joan Francesc Munoz-Martin. "Analytical Computation of the Spatial Resolution in GNSS-R and Experimental Validation at L1 and L5." *Remote Sensing* 12, no. 23 (2020): 3910.

Camps, Adriano, Mercedes Vall-Iloserra, Hyuk Park, Gerard Portal, and Luciana Rossato. "Sensitivity of TDS-1 GNSS-R reflectivity to soil moisture: Global and regional differences and impact of different spatial scales." *Remote Sensing* 10, no. 11 (2018): 1856.

Carreno-Luengo, Hugo, Guido Luzi, and Michele Crosetto. "Sensitivity of CyGNSS bistatic reflectivity and SMAP microwave radiometry brightness temperature to geophysical parameters over land surfaces." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12, no. 1 (2018): 107-122.

Cervellera, Francisco José Fabra. "GNSS-R as a source of opportunity for remote sensing of the cryosphere." PhD diss., Universitat Politècnica de Catalunya (UPC), 2013.

Chan, Gary. "Design and Testing of a GNSS Reflectometry CubeSat Payload for Monitoring Climate Change." Master's thesis, University of Waterloo, 2017.

Chan, Steven, Rajat Bindlish, Raymond Hunt, Tom Jackson, and John Kimball. "Vegetation water content." Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA (2013).

Chan, Steven K., Rajat Bindlish, Peggy E. O'Neill, Eni Njoku, Tom Jackson, Andreas Colliander, Fan Chen et al. "Assessment of the SMAP passive soil moisture product." *IEEE Transactions on Geoscience and Remote Sensing* 54, no. 8 (2016): 4994-5007.

Chattopadhyay, Ashesh, Mustafa Mustafa, Pedram Hassanzadeh, Eviatar Bach, and Karthik Kashinath. "Towards physics-inspired data-driven weather forecasting: integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5." *Geoscientific Model Development* 15, no. 5 (2022): 2221-2237.

Chen, Renpeng, Vincent P. Drnevich, Xiong Yu, Robert L. Nowack, and Yunmin Chen. "Time domain reflectometry surface reflections for dielectric constant in highly conductive soils." *Journal of geotechnical and geoenvironmental engineering* 133, no. 12 (2007): 1597-1608.

Chen, Yuyao, and Luca Dal Negro. "Physics-informed neural networks for imaging and parameter retrieval of photonic nanostructures from near-field data." *APL Photonics* 7, no. 1 (2022): 010802.

Chew, Clara. "Soil moisture remote sensing using GPS-interferometric reflectometry." PhD diss., University of Colorado at Boulder, 2015.

Chew, Clara, and Eric Small. "Description of the UCAR/CU soil moisture product." *Remote Sensing* 12, no. 10 (2020): 1558.

Chew, Clara C., Eric E. Small, Kristine M. Larson, and Valery U. Zavorotny. "Effects of near-surface soil moisture on GPS SNR data: Development of a retrieval algorithm for soil moisture." *IEEE Transactions on Geoscience and Remote Sensing* 52, no. 1 (2013): 537-543.

Chew, C. C., and E. E. Small. "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture." *Geophysical Research Letters* 45, no. 9 (2018): 4049-4057.

Chew, Clara, Rashmi Shah, Cinzia Zuffada, George Hajj, Dallas Masters, and Anthony J. Mannucci. "Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission." *Geophysical Research Letters* 43, no. 7 (2016): 3317-3324.

Chi, Mingmin, Antonio Plaza, Jon Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. "Big data for remote sensing: Challenges and opportunities." *Proceedings of the IEEE* 104, no. 11 (2016): 2207-2219.

Choudhury, B. J., Thomas J. Schmugge, A. Chang, and R. W. Newton. "Effect of surface roughness on the microwave emission from soils." *Journal of Geophysical Research: Oceans* 84, no. C9 (1979): 5699-5706.

Clarizia, M. P., C. P. Gommenginger, S. T. Gleason, M. A. Srokosz, C. Galdi, and M. Di Bisceglie. "Analysis of GNSS-R delay-Doppler maps from the UK-DMC satellite over the ocean." *Geophysical Research Letters* 36, no. 2 (2009).

Clarizia, Maria Paola, Christopher S. Ruf, Philip Jales, and Christine Gommenginger. "Spaceborne GNSS-R minimum variance wind speed estimator." *IEEE transactions on geoscience and remote sensing* 52, no. 11 (2014): 6829-6843.

Clarizia, Maria Paola, Nazzareno Pierdicca, Fabiano Costantini, and Nicolas Floury. "Analysis of CYGNSS data for soil moisture retrieval." *IEEE journal of selected topics in applied earth observations and remote sensing* 12, no. 7 (2019): 2227-2235.

Colliander, Andreas, Thomas J. Jackson, Rajat Bindlish, S. Chan, N. Das, S. B. Kim, M. H. Cosh et al. "Validation of SMAP surface soil moisture products with core validation sites." *Remote sensing of environment* 191 (2017): 215-231.

Davis, Royall Oscar Eugene. *Grouping of soils on the basis of mechanical analysis*. Vol. 419. US Department of Agriculture, 1927.

Daw, Arka, R. Quinn Thomas, Cayelan C. Carey, Jordan S. Read, Alison P. Appling, and Anuj Karpatne. "Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling." In *Proceedings of the 2020 siam international conference on data mining*, pp. 532-540. Society for Industrial and Applied Mathematics, 2020.

De Jeu, Richard AM, W. Wagner, T. R. H. Holmes, A. J. Dolman, N. C. Van De Giesen, and J. Friesen. "Global soil moisture patterns observed by space borne microwave radiometers and scatterometers." *Surveys in Geophysics* 29 (2008): 399-420.

De Roo, Roger D., and Fawwaz T. Ulaby. "Bistatic specular scattering from rough dielectric surfaces." *IEEE Transactions on Antennas and Propagation* 42, no. 2 (1994): 220-231.

De Roo, Roger D., Yang Du, Fawwaz T. Ulaby, and M. Craig Dobson. "A semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion." *IEEE Transactions on Geoscience and Remote Sensing* 39, no. 4 (2001): 864-872.

Dirksen, Christiaan. *Soil physics measurements*. Catena Verlag, 1999.

Do, Synho, Kyoung Doo Song, and Joo Won Chung. "Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning." *Korean journal of radiology* 21, no. 1 (2020): 33-41.

Dobson, Myron C., Fawwaz T. Ulaby, Martti T. Hallikainen, and Mohamed A. El-Rayes. "Microwave dielectric behavior of wet soil-Part II: Dielectric mixing models." *IEEE Transactions on geoscience and remote sensing* 1 (1985): 35-46.

Dong, Zhounan, and Shuanggen Jin. "Evaluation of the land GNSS-Reflected DDM coherence on soil moisture estimation from CYGNSS data." *Remote Sensing* 13, no. 4 (2021): 570.

Dorigo, Wouter, Peter Van Oevelen, Wolfgang Wagner, Matthias Drusch, Susanne Mecklenburg, Alan Robock, and Thomas Jackson. "A new international network for in situ soil moisture data." *Eos, Transactions American Geophysical Union* 92, no. 17 (2011): 141-142.

El-Rabbany, Ahmed. *Introduction to GPS: the global positioning system*. Artech house, 2002.

Entekhabi, Dara, Simon Yueh, Peggy E. O'Neill, Kent H. Kellogg, Angela Allen, Rajat Bindlish, Molly Brown et al. "SMAP handbook—soil moisture active passive: Mapping soil moisture and freeze/thaw from space." (2014).

Eroglu, Orhan, Mehmet Kurum, Dylan Boyd, and Ali Cafer Gurbuz. "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks." *Remote sensing* 11, no. 19 (2019): 2272.

Escorihuela, Maria Jose, Yann H. Kerr, Patricia de Rosnay, Jean-Pierre Wigneron, Jean-Christophe Calvet, and Francois Lemaitre. "A simple model of the bare soil microwave emission at L-band." *IEEE Transactions on Geoscience and Remote Sensing* 45, no. 7 (2007): 1978-1987.

Fabra, Fran, Estel Cardellach, Antonio Rius, Serni Ribo, Santi Oliveras, Oleguer Nogués-Correig, Maria Belmonte Rivas, Maximilian Semmling, and Salvatore D'Addio. "Phase altimetry with dual polarization GNSS-R over sea ice." *IEEE Transactions on Geoscience and Remote Sensing* 50, no. 6 (2011): 2112-2121.

Fontanet, Mireia, Daniel Fernández-Garcia, and Francesc Ferrer. "The value of satellite remote sensing soil moisture data and the DISPATCH algorithm in irrigation fields." *Hydrology and Earth System Sciences* 22, no. 11 (2018): 5889-5900.

Foti, Giuseppe, Christine Gommenginger, Philip Jales, Martin Unwin, Andrew Shaw, Colette Robertson, and Josep Rosello. "Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission." *Geophysical Research Letters* 42, no. 13 (2015): 5435-5441.

Franco, Giovanna. "The Digital Design and Synthesis of Delay Doppler Maps in GNSS Remote Sensing Receivers." PhD diss., Concordia University, 2013.

Gandhi, G. Meera, B. S. Parthiban, Nagaraj Thummalu, and A. Christy. "Ndvi: Vegetation change detection using remote sensing and gis—A case study of Vellore District." *Procedia computer science* 57 (2015): 1199-1210.

Gao, Zhouzheng, Hongping Zhang, Qile Zhao, Zhigang Hu, and Wenbin Shen. "Analyzing the impact of satellite clock-TGD coupled error on BDS positioning accuracy." In *China Satellite Navigation Conference (CSNC) 2014 Proceedings: Volume I*, pp. 267-278. Springer Berlin Heidelberg, 2014.

Gardner, Jacob, Geoff Pleiss, Kilian Q. Weinberger, David Bindel, and Andrew G. Wilson. "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration." *Advances in neural information processing systems* 31 (2018).

Garrison, James L., Stephen J. Katzberg, and Michael I. Hill. "Effect of sea roughness on bistatically scattered range coded signals from the Global Positioning System." *Geophysical research letters* 25, no. 13 (1998): 2257-2260.

Garrison, James L., Attila Komjathy, Valery U. Zavorotny, and Stephen J. Katzberg. "Wind speed measurement using forward scattered GPS signals." *IEEE Transactions on Geoscience and Remote Sensing* 40, no. 1 (2002): 50-65.

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

Ghiasi, Yusof, Claude R. Duguay, Justin Murfitt, Joost J. van der Sanden, Aaron Thompson, Hugo Drouin, and Christian Prévost. "Application of GNSS interferometric reflectometry for the estimation of lake ice thickness." *Remote Sensing* 12, no. 17 (2020): 2721.

Gleason, Scott. "Remote sensing of ocean, ice and land surfaces using bistatically scattered GNSS signals from low earth orbit." PhD diss., University of Surrey, 2006.

Gleason, Scott. "Towards sea ice remote sensing with space detected GPS signals: Demonstration of technical feasibility and initial consistency check using low resolution sea ice information." *Remote Sensing* 2, no. 8 (2010): 2017-2039.

Gleason, Scott, Demoz Gebre-Egziabher. "GNSS applications and methods." (2009).

Gleason, Scott, Mounir Adjrad, and Martin Unwin. "Sensing ocean, ice and land reflected signals from space: results from the UK-DMC GPS reflectometry experiment." In *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)*, pp. 1679-1685. 2005.

Gleason, Scott, Stephen Hodgart, Yiping Sun, Christine Gommenginger, Stephen Mackin, Mounir Adjrad, and Martin Unwin. "Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing." *IEEE Transactions on Geoscience and Remote Sensing* 43, no. 6 (2005): 1229-1241.

Gleason, Scott, and Valery Zavorotny. "Bistatic radar cross section measurements of ocean scattered GPS signals from low earth orbit." In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pp. 1308-1311. IEEE, 2006.

Gnanasambandam, Raghav, Bo Shen, Jihoon Chung, and Xubo Yue. "Self-scalable tanh (stan): Faster convergence and better generalization in physics-informed neural networks." arXiv preprint arXiv:2204.12589 (2022).

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote sensing of Environment* 202 (2017): 18-27.

Groenendyk, Derek G., Ty PA Ferre, Kelly R. Thorp, and Amy K. Rice. "Hydrologic-process-based soil texture classifications for improved visualization of landscape function." *PloS one* 10, no. 6 (2015): e0131299.

Guan, Dongliang, Hyuk Park, Adriano Camps, Yong Wang, Raul Onrubia, Jorge Querol, and Daniel Pascual. "Wind direction signatures in GNSS-R observables from space." *Remote Sensing* 10, no. 2 (2018): 198.

Guo, Peng, Jiancheng Shi, Bo Gao, and Hong Wan. "Evaluation of errors induced by soil dielectric models for soil moisture retrieval at L-band." In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1679-1682. IEEE, 2016.

Guo, Quan, Yue Zhao, Chunhui Lu, and Jian Luo. "High-dimensional inverse modeling of hydraulic tomography by physics informed neural network (HT-PINN)." *Journal of Hydrology* 616 (2023): 128828.

Haghighat, Ehsan, and Ruben Juanes. "SciANN: A Keras/TensorFlow wrapper for scientific computations and physics-informed deep learning using artificial neural networks." *Computer Methods in Applied Mechanics and Engineering* 373 (2021): 113552.

Hall, C. D., and R. A. Cordey. "Multistatic scatterometry." In *International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Moving Toward the 21st Century'*, vol. 1, pp. 561-562. IEEE, 1988.

Hallikainen, Martti T., Fawwaz T. Ulaby, Myron C. Dobson, Mohamed A. El-Rayes, and Lil-Kun Wu. "Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations." *IEEE Transactions on Geoscience and Remote Sensing* 1 (1985): 25-34.



Harries, Kent A., and Bhavna Sharma, eds. *Nonconventional and vernacular construction materials: Characterisation, properties and applications*. Woodhead Publishing, 2019.

Hennigh, Oliver, Susheela Narasimhan, Mohammad Amin Nabian, Akshay Subramaniam, Kaustubh Tangsali, Zhiwei Fang, Max Rietmann, Wonmin Byeon, and Sanjay Choudhry. "NVIDIA SimNet™: An AI-accelerated multi-physics simulation framework." In *Computational Science–ICCS 2021: 21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part V*, pp. 447-461. Cham: Springer International Publishing, 2021.

Hinton, Geoffrey E. "Connectionist learning procedures." In *Machine learning*, pp. 555-610. Morgan Kaufmann, 1990.

Hird, Jennifer N., Evan R. DeLancey, Gregory J. McDermid, and Jahan Kariyeva. "Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping." *Remote sensing* 9, no. 12 (2017): 1315.

Hofmann-Wellenhof, Bernhard, Herbert Lichtenegger, and Elmar Wasle. *GNSS—global navigation satellite systems: GPS, GLONASS, Galileo, and more*. Springer Science & Business Media, 2007.

Hu, Changjiang, Craig Benson, Chris Rizos, and Li Qiao. "Single-pass sub-meter space-based GNSS-R ice altimetry: Results from TDS-1." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10, no. 8 (2017): 3782-3788.

Hu, Changjiang, Craig Benson, Hyuk Park, Adriano Camps, Li Qiao, and Chris Rizos. "Detecting targets above the Earth's surface using GNSS-R delay Doppler maps: Results from TDS-1." *Remote Sensing* 11, no. 19 (2019): 2327.

Huang, Feixiong, James L. Garrison, S. Mark Leidner, Giuseppe Grieco, Ad Stoffelen, Bachir Annane, and Ross N. Hoffman. "Assimilation of GNSS reflectometry delay-Doppler maps with a two-dimensional variational analysis of global ocean surface winds." *Quarterly Journal of the Royal Meteorological Society* 147, no. 737 (2021): 2469-2489.

Imam, Rayan, Marco Pini, Gianluca Marucco, Fabrizio Dominici, and Fabio Dosis. "UAV-based GNSS-R for water detection as a support to flood monitoring operations: A feasibility study." *Applied Sciences* 10, no. 1 (2019): 210.

International Society of Soil Science. First Commission. *Transactions of the First Commission of the International Society of Soil Science*. 1927.

Jales, Philip. *Spaceborne receiver design for scatterometric GNSS reflectometry*. University of Surrey (United Kingdom), 2012.

Jia, Yan, Patrizia Savi, Yuekun Pei, and Riccardo Notarpietro. "GNSS reflectometry for remote sensing of soil moisture." In *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pp. 498-501. IEEE, 2015.

Jia, Yan, Shuanggen Jin, Patrizia Savi, Yun Gao, Jing Tang, Yixiang Chen, and Wenmei Li. "GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation." *Remote sensing* 11, no. 14 (2019): 1655.

Jia, Yan, Qingyun Yan, Shuanggen Jin, and Patrizia Savi. "Cygness Soil Moisture Estimation Using Machine Learning Regression." In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 6323-6326. IEEE, 2021.

Jin, Mengjie, Xingming Zheng, Tao Jiang, Xiaofeng Li, Xiaojie Li, and Kai Zhao. "Evaluation and improvement of SMOS and SMAP soil moisture products for soils with high organic matter over a forested area in Northeast China." *Remote Sensing* 9, no. 4 (2017): 387.

Jin, Shuanggen, Estel Cardellach, and Feiqin Xie. *GNSS remote sensing*. Vol. 16. Dordrecht: Springer, 2014.

Jin, Shuanggen, and Attila Komjathy. "GNSS reflectometry and remote sensing: New objectives and results." *Advances in Space Research* 46, no. 2 (2010): 111-117.

Jin, Xiaowei, Shengze Cai, Hui Li, and George Em Karniadakis. "NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations." *Journal of Computational Physics* 426 (2021): 109951.

Jones, Scott B., Jon M. Wraith, and Dani Or. "Time domain reflectometry measurement principles and applications." *Hydrological processes* 16, no. 1 (2002): 141-153.

Jonnalagadda, Sreeram. "Resistivity and Time Domain Reflectometry Sensors for Assessing in Situ Moisture Content in a Bioreactor Landfill." PhD diss., University of Florida, 2004.

Juang, Jyh-Ching. "On the Determination of the Specular Reflection Point in GNSS Reflectometry." In *2021 IEEE Specialist Meeting on Reflectometry using GNSS and other Signals of Opportunity (GNSS+ R)*, pp. 86-89. IEEE, 2021.

Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. "Physics-informed machine learning." *Nature Reviews Physics* 3, no. 6 (2021): 422-440.

Kashinath, Karthik, M. Mustafa, Adrian Albert, J. L. Wu, C. Jiang, Soheil Esmaeilzadeh, Kamyar Azizzadenesheli et al. "Physics-informed machine learning: case studies for weather and climate modelling." *Philosophical Transactions of the Royal Society A* 379, no. 2194 (2021): 20200093.

Katzberg, Stephen J., Omar Torres, Michael S. Grant, and Dallas Masters. "Utilizing calibrated GPS reflected signals to estimate soil reflectivity and dielectric constant: Results from SMEX02." *Remote sensing of environment* 100, no. 1 (2006): 17-28.

Kerr, Yann H. "Soil moisture from space: Where are we?." *Hydrogeology journal* 15, no. 1 (2007): 117-120.

Kerr, Yann H., Philippe Waldteufel, J-P. Wigneron, J. A. M. J. Martinuzzi, Jordi Font, and Michael Berger. "Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission." *IEEE transactions on Geoscience and remote sensing* 39, no. 8 (2001): 1729-1735.

Kerr, Yann H., Philippe Waldteufel, Jean-Pierre Wigneron, Steven Delwart, François Cabot, Jacqueline Boutin, Maria-José Escorihuela et al. "The SMOS mission: New tool for monitoring key elements of the global water cycle." *Proceedings of the IEEE* 98, no. 5 (2010): 666-687.

Kerr, Yann H., Philippe Waldteufel, Philippe Richaume, Jean Pierre Wigneron, Paolo Ferrazzoli, Ali Mahmoodi, Ahmad Al Bitar et al. "The SMOS soil moisture retrieval algorithm." *IEEE transactions on geoscience and remote sensing* 50, no. 5 (2012): 1384-1403.

Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).

Klute, Arnold. "Methods of soil analysis 2d ed., pt. 1; physical and mineralogical methods." *Soil Science* 146, no. 2 (1988): 138.

Komjathy, Attila, Michael Armatys, Dallas Masters, Penina Axelrad, Valery Zavorotny, and Steven Katzberg. "Retrieval of ocean surface wind speed and wind direction using reflected GPS signals." *Journal of Atmospheric and Oceanic Technology* 21, no. 3 (2004): 515-526.

Koryagin, Alexander, Roman Khudorozkov, and Sergey Tsimfer. "PyDEns: A python framework for solving differential equations with neural networks." *arXiv preprint arXiv:1909.11544* (2019).

Kuncham, Venkata Sriharsha, and N. V. Rao. "Sensors for managing water resources in agriculture." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* 9, no. 2 (2014): 145-163.

Larson, Kristine M. "GPS interferometric reflectometry: applications to surface soil moisture, snow depth, and vegetation water content in the western United States." *Wiley Interdisciplinary Reviews: Water* 3, no. 6 (2016): 775-787.

Larson, Kristine M., Eric E. Small, Ethan Gutmann, Andria Bilich, Penina Axelrad, and John Braun. "Using GPS multipath to measure soil moisture fluctuations: initial results." *GPS solutions* 12 (2008 a): 173-177.

Larson, Kristine M., Eric E. Small, Ethan D. Gutmann, Andria L. Bilich, John J. Braun, and Valery U. Zavorotny. "Use of GPS receivers as a soil moisture network for water cycle studies." *Geophysical Research Letters* 35, no. 24 (2008 b).

Larson, Kristine M., and Felipe G. Nievinski. "GPS snow sensing: results from the EarthScope Plate Boundary Observatory." *GPS solutions* 17 (2013): 41-52.

Lawrence, Heather, Jean-Pierre Wigneron, Francois Demontoux, Arnaud Mialon, and Yann H. Kerr. "Evaluating the semiempirical H-Q model used to calculate the L-band emissivity of a rough bare soil." *IEEE Transactions on Geoscience and Remote Sensing* 51, no. 7 (2013): 4075-4084.

Li, Weiqiang, Estel Cardellach, Fran Fabra, Serni Ribo, and Antonio Rius. "Lake level and surface topography measured with spaceborne GNSS-reflectometry from CYGNSS mission: Example for the lake Qinghai." *Geophysical Research Letters* 45, no. 24 (2018): 13-332.

Liu, Jing, and Qinhua Liu. "Soil moisture estimate uncertainties from the effect of soil texture on dielectric semiempirical models." *Remote Sensing* 12, no. 14 (2020): 2343.

Lowe, Stephen T., Clara Chew, Jesal Shah, and Michael Kilzer. "An aircraft wetland inundation experiment using GNSS reflectometry." *Remote Sensing* 12, no. 3 (2020): 512.

Lowe, Stephen T., John L. LaBrecque, Cinzia Zuffada, Larry J. Romans, Larry E. Young, and George A. Hajj. "First spaceborne observation of an Earth-reflected GPS signal." *Radio Science* 37, no. 1 (2002): 1-28.

Lu, Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. "DeepXDE: A deep learning library for solving differential equations." *SIAM review* 63, no. 1 (2021): 208-228.

Ma, Yan, Haiping Wu, Lizhe Wang, Bormin Huang, Rajiv Ranjan, Albert Zomaya, and Wei Jie. "Remote sensing big data computing: Challenges and opportunities." *Future Generation Computer Systems* 51 (2015): 47-60.

Maddu, Suryanarayana, Dominik Sturm, Christian L. Müller, and Ivo F. Sbalzarini. "Inverse Dirichlet weighting enables reliable training of physics informed neural networks." *Machine Learning: Science and Technology* 3, no. 1 (2022): 015026.

Mao, Zhiping, Ameya D. Jagtap, and George Em Karniadakis. "Physics-informed neural networks for high-speed flows." *Computer Methods in Applied Mechanics and Engineering* 360 (2020): 112789.

Marchan-Hernandez, Juan Fernando, Adriano Camps, Nereida Rodriguez-Alvarez, Xavier Bosch-Lluis, Isaac Ramos-Perez, and Enric Valencia. "PAU/GNSS-R: Implementation, performance and first results of a real-time Delay-Doppler map reflectometer using global navigation satellite system signals." *Sensors* 8, no. 5 (2008): 3005-3019.

Martin-Neira, Manuel. "A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry." *ESA journal* 17, no. 4 (1993): 331-355.

Mashburn, Jake, Penina Axelrad, Stephen T. Lowe, and Kristine M. Larson. "Global ocean altimetry with GNSS reflections from TechDemoSat-1." *IEEE Transactions on Geoscience and Remote Sensing* 56, no. 7 (2018): 4088-4097.

Masters, D. S. *Surface remote sensing applications of GNSS bistatic radar: Soil moisture and aircraft altimetry*. University of Colorado at Boulder, 2004.

Mathur, Shray. "Investigating CYGNSS-derived surface reflectivity for estimating soil moisture in Texas." PhD diss., BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, 2021.

Meroni, Michele, Dominique Fusbender, Felix Rembold, Clement Atzberger, and Anja Klisch. "Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options." *Remote sensing of environment* 221 (2019): 508-521.

Mironov, Valery L. "Spectral dielectric properties of moist soils in the microwave band." In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, vol. 5, pp. 3474-3477. IEEE, 2004.

Mironov, Valery L., M. Craig Dobson, Verne H. Kaupp, Sergey A. Komarov, and Vladimir N. Kleshchenko. "Generalized refractive mixing dielectric model for moist soils." *IEEE transactions on Geoscience and Remote sensing* 42, no. 4 (2004): 773-785.

Mironov, Valery L., Lyudmila G. Kosolapova, and Sergey V. Fomin. "Soil dielectric model accounting for contribution of bound water spectra through clay content." *PIERS online* 4, no. 1 (2008): 31-35.

Mironov, Valery L., Lyudmila G. Kosolapova, and Sergej V. Fomin. "Physically and mineralogically based spectroscopic dielectric model for moist soils." *IEEE Transactions on Geoscience and Remote Sensing* 47, no. 7 (2009): 2059-2070.

Mironov, Valery L., and Sergey V. Fomin. "Temperature and mineralogy dependable model for microwave dielectric spectra of moist soils." *Piers online* 5, no. 5 (2009): 411-415.

Mironov, Valery, Yann Kerr, Jean-Pierre Wigneron, Liudmila Kosolapova, and François Demontoux. "Temperature-and texture-dependent dielectric model for moist soils at 1.4 GHz." *IEEE Geoscience and Remote Sensing Letters* 10, no. 3 (2012): 419-423.

Moreira, Alberto, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. "A tutorial on synthetic aperture radar." *IEEE Geoscience and remote sensing magazine* 1, no. 1 (2013): 6-43.

Muñoz-Carpena, R. "Field Devices for Monitoring Soil Water Content. Extension Bul. 343 of the Dept. of Agr. and Bio. Engineering, University of Florida." (2005).

Muther, Temoor, Amirmasoud Kalantari Dahaghi, Fahad Iqbal Syed, and Vuong Van Pham. "Physical laws meet machine intelligence: current developments and future directions." *Artificial Intelligence Review* (2022): 1-67.

Newton, Richard Wayne. *MICROWAVE REMOTE SENSING AND ITS APPLICATION TO SOIL MOISTURE DETECTION.(VOLUMES I AND II)*. Texas A&M University, 1977.

Nghiem, Son V., Cinzia Zuffada, Rashmi Shah, Clara Chew, Stephen T. Lowe, Anthony J. Mannucci, Estel Cardellach, G. Robert Brakenridge, Gary Geller, and Ake Rosenqvist. "Wetland monitoring with global navigation satellite system reflectometry." *Earth and Space Science* 4, no. 1 (2017): 16-39.

Njoku, Eni G., and Dara Entekhabi. "Passive microwave remote sensing of soil moisture." *Journal of hydrology* 184, no. 1-2 (1996): 101-129.

Njoku, Eni G., and Li Li. "Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz." *IEEE Transactions on Geoscience and Remote Sensing* 37, no. 1 (1999): 79-93.

O'Neill, Peggy, Rajat Bindlish, Steven Chan, Eni Njoku, and Tom Jackson. "Algorithm theoretical basis document. Level 2 & 3 soil moisture (passive) data products." (2021).

Panciera, Rocco, Jeffrey P. Walker, Thomas J. Jackson, Douglas A. Gray, Mihai A. Tanase, Dongryeol Ryu, Alessandra Monerris et al. "The soil moisture active passive experiments (SMAPEX): Toward soil moisture retrieval from the SMAP mission." *IEEE transactions on geoscience and remote sensing* 52, no. 1 (2013): 490-507.

Parkinson, Bradford W. "Global Positioning System: Theory and Applications II." *Progress in Astronautics and Aeronautics* 164 (1996): 3-50.

Parkinson, Bradford W., and Stephen W. Gilbert. "NAVSTAR: Global positioning system—Ten years later." *Proceedings of the IEEE* 71, no. 10 (1983): 1177-1186.

Peplinski, Neil R., Fawwaz T. Ulaby, and Myron C. Dobson. "Dielectric properties of soils in the 0.3-1.3-GHz range." *IEEE transactions on Geoscience and Remote sensing* 33, no. 3 (1995): 803-807.

Pierdicca, Nazzareno, Davide Comite, Adriano Camps, Hugo Carreno-Luengo, Luca Cenci, Maria Paola Clarizia, Fabiano Costantini et al. "The potential of spaceborne GNSS reflectometry for soil moisture, biomass, and freeze–thaw monitoring: Summary of a

European space agency-funded study." *IEEE geoscience and remote sensing magazine* 10, no. 2 (2021): 8-38.

Portal, Gerard, Thomas Jagdhuber, Mercè Vall-Ilossera, Adriano Camps, Miriam Pablos, Dara Entekhabi, and Maria Piles. "Assessment of multi-scale SMOS and SMAP soil moisture products across the Iberian Peninsula." *Remote Sensing* 12, no. 3 (2020): 570.

Prakash, Shikha, Animesh Sharma, and Sitanshu Shekhar Sahu. "Soil moisture prediction using machine learning." In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 1-6. IEEE, 2018.

Rackauckas, Christopher, and Qing Nie. "Differential equations. jl—a performant and feature-rich ecosystem for solving differential equations in julia." *Journal of open research software* 5, no. 1 (2017).

Raissi, Maziar, Paris Perdikaris, and George Em Karniadakis. "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations." arXiv preprint arXiv:1711.10561 (2017).

Raissi, Maziar, Paris Perdikaris, and George E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational physics* 378 (2019): 686-707.

Rasheed, Muhammad Waseem, Jialiang Tang, Abid Sarwar, Suraj Shah, Naeem Saddique, Muhammad Usman Khan, Muhammad Imran Khan et al. "Soil moisture measuring techniques and factors affecting the moisture dynamics: A comprehensive review." *Sustainability* 14, no. 18 (2022): 11538.

Ray, Ram L., Ali Fares, Yiping He, and Marouane Temimi. "Evaluation and inter-comparison of satellite soil moisture products using in situ observations over Texas, US." *Water* 9, no. 6 (2017): 372.

Rius, Antonio, Estel Cardellach, Fran Fabra, Weiqiang Li, Serni Ribó, and Manuel Hernández-Pajares. "Feasibility of GNSS-R ice sheet altimetry in Greenland using TDS-1." *Remote Sensing* 9, no. 7 (2017): 742.



Rius, Antonio, Fran Fabra, Serni Ribo, Juan Carlos Arco, Santi Oliveras, Estel Cardellach, Adriano Camps et al. "PARIS interferometric technique proof of concept: Sea surface altimetry measurements." In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7067-7070. IEEE, 2012.

Rius, Antonio, Oleguer Nogués-Correig, Serni Ribó, Estel Cardellach, Santi Oliveras, Enric Valencia, Hyuk Park et al. "Altimetry with GNSS-R interferometry: First proof of concept experiment." *GPS solutions* 16 (2012): 231-241.

Rivas, Maria Belmonte. "Bistatic scattering of global positioning system signals from Arctic sea ice." PhD diss., University of Colorado at Boulder, 2007.

Roberts, Thomas Maximillian, Ian Colwell, Clara Chew, Stephen Lowe, and Rashmi Shah. "A deep-learning approach to soil moisture estimation with GNSS-R." *Remote Sensing* 14, no. 14 (2022): 3299.

Rodriguez-Alvarez, Nereida, Dennis M. Akos, Valery U. Zavorotny, Jeffrey A. Smith, Adriano Camps, and Christopher W. Fairall. "Airborne GNSS-R wind retrievals using delay–Doppler maps." *IEEE transactions on geoscience and remote sensing* 51, no. 1 (2012): 626-641.

Rodriguez-Alvarez, Nereida, Erika Podest, Katherine Jensen, and Kyle C. McDonald. "Classifying inundation in a tropical wetlands complex with GNSS-R." *Remote Sensing* 11, no. 9 (2019): 1053.

Rodriguez-Alvarez, Nereida, Joan Francesc Munoz-Martin, and Mary Morris. "Latest Advances in the Global Navigation Satellite System—Reflectometry (GNSS-R) Field." *Remote Sensing* 15, no. 8 (2023): 2157.

Rodriguez-Fernandez, Nemesio J., Filipe Aires, Philippe Richaume, Yann H. Kerr, Catherine Prigent, Jana Kolassa, Francois Cabot, Carlos Jimenez, Ali Mahmoodi, and Matthias Drusch. "Soil moisture retrieval using neural networks: Application to SMOS." *IEEE Transactions on Geoscience and Remote Sensing* 53, no. 11 (2015): 5991-6007.

Rohrhofer, Franz M., Stefan Posch, Clemens Gößnitzer, and Bernhard C. Geiger. "On the Role of Fixed Points of Dynamical Systems in Training Physics-Informed Neural Networks." *Transactions on Machine Learning Research* (2023).

Rose, Randy, Will Wells, Jillian Redfern, Debi Rose, John Dickinson, Chris Ruf, Aaron Ridley, and Kyle Nave. "NASA's Cyclone Global Navigation Satellite System (CYGNSS) Mission—Temporal Resolution of a Constellation Enabled by Micro-Satellite Technology." (2013).

Roussel, N., F. Frappart, G. Ramillien, C. Desjardins, P. Gegout, F. Pérosanz, and R. Biancale. "Simulations of direct and reflected waves trajectories for in situ GNSS-R experiments." *Geoscientific Model Development Discussions* 7, no. 1 (2014).

Ruf, Christopher S., Scott Gleason, Zorana Jelenak, Stephen Katzberg, Aaron Ridley, Randall Rose, John Scherrer, and Valery Zavorotny. "The CYGNSS nanosatellite constellation hurricane mission." In 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 214-216. IEEE, 2012.

Ruf, Christopher S., Robert Atlas, Paul S. Chang, Maria Paola Clarizia, James L. Garrison, Scott Gleason, Stephen J. Katzberg et al. "New ocean winds satellite mission to probe hurricanes and tropical convection." *Bulletin of the American Meteorological Society* 97, no. 3 (2016): 385-395.

Ruffini, N. G., F. Soulat, M. Caparrini, O. Germain, and M. Martín-Neira. "The Eddy Experiment: Accurate GNSS-R ocean altimetry from low altitude aircraft." *Geophysical research letters* 31, no. 12 (2004).

Saba, Amirhossein, Carlo Gigli, Ahmed B. Ayoub, and Demetri Psaltis. "Physics-informed neural networks for diffraction tomography." *Advanced Photonics* 4, no. 6 (2022): 066001.

Santerre, Rock, Lin Pan, Changsheng Cai, and Jianjun Zhu. "Single point positioning using GPS, GLONASS and BeiDou satellites." *Positioning* 5, no. 04 (2014): 107.

Santi, Emanuele, Simonetta Paloscia, Simone Pettinato, and Giacomo Fontanelli. "Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors." *International journal of applied earth observation and geoinformation* 48 (2016): 61-73.

Senyurek, Volkan, Fangni Lei, Dylan Boyd, Ali Cafer Gurbuz, Mehmet Kurum, and Robert Moorhead. "Evaluations of machine learning-based CYGNSS soil moisture estimates against SMAP observations." *Remote Sensing* 12, no. 21 (2020): 3503.

Senyurek, Volkan, Fangni Lei, Dylan Boyd, Mehmet Kurum, Ali Cafer Gurbuz, and Robert Moorhead. "Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS." *Remote Sensing* 12, no. 7 (2020): 1168.

Senyurek, V., M. Farhad, A. C. Gurbuz, M. Kurum, and R. Moorhead. "SoilMoistureMapper: A GNSS-R approach for soil moisture retrieval on UAV." *AI for Agriculture and Food Systems; Association for the Advancement of Artificial Intelligence: Menlo Park, CA, USA* (2021).

Shi, Chuang, and Na Wei. "Satellite navigation for digital earth." *Manual of Digital Earth* (2020): 125-160.

Shi, Jiancheng, Kun-Shan Chen, Qin Li, Thomas J. Jackson, Peggy E. O'Neill, and Leung Tsang. "A parameterized surface reflectivity model and estimation of bare-surface soil moisture with L-band radiometer." *IEEE Transactions on Geoscience and Remote Sensing* 40, no. 12 (2002): 2674-2686.

Shi, Yajie, Chao Ren, Zhiheng Yan, and Jianmin Lai. "Improving soil moisture retrieval from GNSS-interferometric reflectometry: parameters optimization and data fusion via neural network." *International Journal of Remote Sensing* 42, no. 23 (2021): 9085-9108.

Shin, Yeonjong, Jerome Darbon, and George Em Karniadakis. "On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs." *arXiv preprint arXiv:2004.01806* (2020).

Schmugge, Thomas, P. Gloersen, Tom Wilheit, and F. Geiger. "Remote sensing of soil moisture with microwave radiometers." *Journal of Geophysical Research* 79, no. 2 (1974): 317-323.

Schmugge, T. J., T. J. Jackson, and H. L. McKim. "Survey of methods for soil moisture determination." *Water Resources Research* 16, no. 6 (1980): 961-979.

Schmugge, Thomas, Peggy E. O'Neill, and James R. Wang. "Passive microwave soil moisture research." *IEEE Transactions on Geoscience and Remote Sensing* 1 (1986): 12-22.

Singh, Abhilash, Kumar Gaurav, Gaurav Kailash Sonkar, and Cheng-Chi Lee. "Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: review, bibliometric analysis, applications, research findings, and future directions." *IEEE Access* (2023).

Smith, Leslie N. "Cyclical learning rates for training neural networks." In 2017 IEEE winter conference on applications of computer vision (WACV), pp. 464-472. IEEE, 2017.

Soulat, Francois. "Sea surface remote sensing with gnss and sunlighth reflections." PhD diss., Universitat Politècnica de Catalunya (UPC), 2004.

Space Studies Board, and National Research Council. *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: Part I. Science and Design*. Vol. 1. National Academies Press, 2000.

Spilker Jr, James J., Penina Axelrad, Bradford W. Parkinson, and Per Enge, eds. *Global positioning system: theory and applications, volume I*. American Institute of Aeronautics and Astronautics, 1996.

Tabibi, Sajad, Renaldo Sauveur, Kelly Guerrier, Gerard Metayer, and Olivier Francis. "SNR-Based GNSS-R for Coastal Sea-Level Altimetry." *Geosciences* 11, no. 9 (2021): 391.

Tamiminia, Haifa, Bahram Salehi, Masoud Mahdianpari, Lindi Quackenbush, Sarina Adeli, and Brian Brisco. "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review." *ISPRS Journal of Photogrammetry and Remote Sensing* 164 (2020): 152-170.

Tartakovsky, Alexandre M., C. Ortiz Marrero, Paris Perdikaris, Guzel D. Tartakovsky, and David Barajas-Solano. "Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems." *Water Resources Research* 56, no. 5 (2020): e2019WR026731.

Topp, G. Clarke, J. L. Davis, and Aa P. Annan. "Electromagnetic determination of soil water content: Measurements in coaxial transmission lines." *Water resources research* 16, no. 3 (1980): 574-582.

Ulaby, Fawwaz. T., Richard K. Moore, and Adrian K. Fung. "Microwave remote sensing: Active and passive. Volume 2-Radar remote sensing and surface scattering and emission theory." (1982).

Ulaby, Fawwaz T., Richard K. Moore, and Adrian K. Fung. "Microwave remote sensing: Active and passive. Volume 3-From theory to applications." (1986).

Unwin, Martin, Philip Jales, Jason Tye, Christine Gommenginger, Giuseppe Foti, and Josep Rosello. "Spaceborne GNSS-reflectometry on TechDemoSat-1: Early mission

operations and exploitation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, no. 10 (2016): 4525-4539.

Vermote, E.; Wolfe, R. MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1 km and 500 m. SIN Grid V006. NASA EOSDIS Land Processes DAAC. 2015. Available online:[https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mod09ga\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09ga_v006) (accessed on 16 October 2016).

Voronovich, Alexander G., and Valery U. Zavorotny. "Bistatic radar equation for signals of opportunity revisited." *IEEE Transactions on Geoscience and Remote Sensing* 56, no. 4 (2017): 1959-1968.

Wang, James R., and Thomas J. Schmugge. "An empirical model for the complex dielectric permittivity of soils as a function of water content." *IEEE Transactions on Geoscience and Remote Sensing* 4 (1980): 288-295.

Wang, James, T. Schmugge, and Donald Williams. *Dielectric constants of soils at microwave frequencies-2*. No. NASA-TP-1238. 1978.

Wigneron, Jean-Pierre, André Chanzy, Yann H. Kerr, Heather Lawrence, Jiancheng Shi, Maria Jose Escorihuela, Valery Mironov et al. "Evaluating an improved parameterization of the soil emission in L-MEB." *IEEE Transactions on Geoscience and Remote Sensing* 49, no. 4 (2010): 1177-1189.

Wigneron, Jean-Pierre, Mike Schwank, Ernesto Lopez Baeza, Yann Kerr, Nathalie Novello, Cristina Millan, Christophe Moisy et al. "First evaluation of the simultaneous SMOS and ELBARA-II observations in the Mediterranean region." *Remote Sensing of Environment* 124 (2012): 26-37.

Whitney, Milton. "use of soils east of the Great Plains region." (1911).

Wilder, B. (2012). *Cloud architecture patterns: using microsoft azure*. " O'Reilly Media, Inc."

Wolf, Emil. *Introduction to the Theory of Coherence and Polarization of Light*. Cambridge university press, 2007.

Wu, Fan, Wei Zheng, Zongqiang Liu, and Xuezhi Sun. "Improving the Specular Point Positioning Accuracy of Ship-Borne GNSS-R Observations in China Seas Based on Comprehensive Geophysical Correction." *Frontiers in Earth Science* 9 (2021): 720470.

Wu, Xuerui, Ying Li, and Jin Xu. "Theoretical study on GNSS-R vegetation biomass." In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6380-6383. IEEE, 2012.

Wu, Xiaofeng, Ziling Wu, Sibi Chakravarthy Shanmugavel, Z. Yu Hang, and Yunhui Zhu. "Physics-informed neural network for phase imaging based on transport of intensity equation." *Optics Express* 30, no. 24 (2022): 43398-43416.

Xu, Kailai, and Eric Darve. "ADCME: Learning spatially-varying physical fields using deep neural networks." *arXiv preprint arXiv:2011.11955* (2020).

Yang, Liu, Dongkun Zhang, and George Em Karniadakis. "Physics-informed generative adversarial networks for stochastic differential equations." *arXiv preprint arXiv:1811.02033* (2018).

Yu, Kegen, Chris Rizos, Derek Burrage, Andrew G. Dempster, Kefei Zhang, and Markus Markgraf. "An overview of GNSS remote sensing." *EURASIP Journal on Advances in Signal Processing* 2014 (2014): 1-14.

Zavorotny, Valery U., Scott Gleason, Estel Cardellach, and Adriano Camps. "Tutorial on remote sensing using GNSS bistatic radar of opportunity." *IEEE Geoscience and Remote Sensing Magazine* 2, no. 4 (2014): 8-45.

Zavorotny, Valery U., and Alexander G. Voronovich. "Bistatic GPS signal reflections at various polarizations from rough land surface with moisture content." In *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*, vol. 7, pp. 2852-2854. IEEE, 2000.

Zavorotny, Valery U., and Alexander G. Voronovich. "Scattering of GPS signals from the ocean with wind remote sensing application." *IEEE Transactions on Geoscience and Remote Sensing* 38, no. 2 (2000): 951-964.

Zhang, Linlin, Qingyan Meng, Die Hu, Ying Zhang, Shun Yao, and Xu Chen. "Comparison of different soil dielectric models for microwave soil moisture retrievals." *International Journal of Remote Sensing* 41, no. 8 (2020): 3054-3069.

Zhang, Sibio, Nicolas Roussel, Karen Boniface, Minh Cuong Ha, Frédéric Frappart, José Darrozes, Frédéric Baup, and Jean-Christophe Calvet. "Use of reflected GNSS SNR

data to retrieve either soil moisture or vegetation height from a wheat crop." *Hydrology and Earth System Sciences* 21, no. 9 (2017): 4767-4784.

Zobeiry, Navid, and Keith D. Humfeld. "A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications." *Engineering Applications of Artificial Intelligence* 101 (2021): 104232.