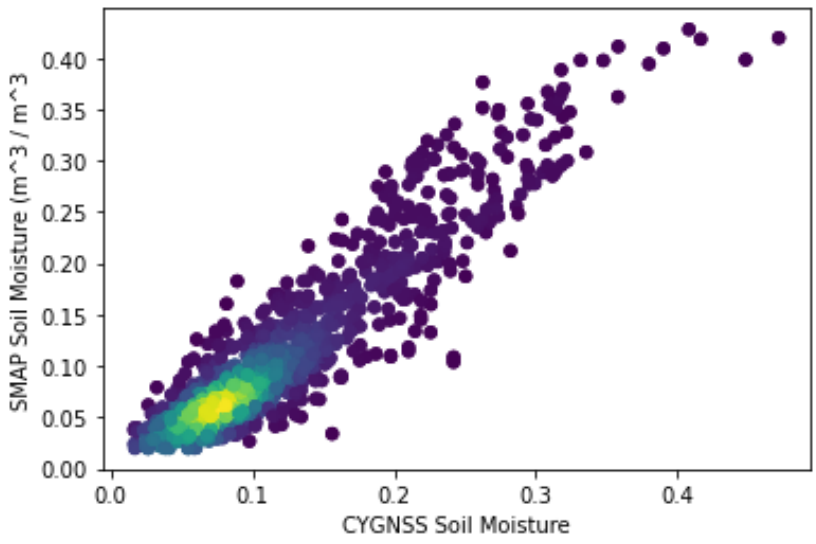
A COMPARISON OF GNSS REFLECTOMETRY SOIL MOISTURE DETECTION TECHNIQUES

Samuel Jasper John Christelow (ppysjch)

Supervisor: Dr Paul Blunt

20th August 2021



Submitted in partial fulfilment of the requirements for the degree of Master of Geospatial Data Science in the Faculty of Engineering

Nottingham Geospatial Institute
The University of Nottingham
Triumph Road
Nottingham
NG7 2TU

Declaration:

I hereby certify that this work is my own, except where otherwise acknowledged, and that it has not been submitted previously for a degree at this, or any other university.

Samuel Christelow

Acknowledgements:

I would like to thank my supervisor, Dr Paul Blunt, for his support throughout this project. I would also like to thank my Geospatial Systems CDT colleagues for helping answer many small questions and for the general camaraderie.

Abstract

This report covers an evaluation of the effectiveness of soil moisture retrievals made by the TechDemoSat-1 (TDS1) and Cyclone Global Navigation Satellite System (CYGNSS) missions by comparing them to radiometry data from the Soil Moisture Active Passive (SMAP) mission. The TDS1 and CYGNSS soil moisture data is derived from GNSS Reflectometry (GNSS-R) measurements made over land. Over four 24 hour time frames it is found that the CYGNSS data shows a much stronger correlation with SMAP data, with the Pearson correlation coefficient between the two never dropping below 0.9. In comparison, the TDS1 data shows results not exceeding 0.6. In addition, a short investigation into effective filtering of the TDS1 data is presented.

Key Words: GNSS, Reflectometry, Soil Moisture, GNSS-R, TechDemoSat-1, CYGNSS, SMAP

Contents

1 Introduction **6**

2 Aims & Objectives **10**

3 Literature Review **11**

4 Methodology **21**

5 Results & Discussion **25**
5.1 TDS1 Filtering 25
5.2 Retrievals 26
5.2.1 6th April 2017 26
5.2.2 14th April 2017 28
5.2.3 22nd April 2017 29
5.2.4 30th April 2017 30
5.3 Comparisons 31

6 Conclusion **32**

7 Future Work **33**

References **33**

List of Figures

From Literature:
Figure 1. Working principle of GNSS-R (page 6)
Figure 2. Typical DDM (page 9)
Figure 3. TDS1 Reflectivity vs Soil Moisture Content (page 12)
Figure 4. TDS1 derived wind speed vs ASCAT training data (page 13)
Figure 5. TDS1 derived wind speed vs ASCAT validation data (page 14)
Figure 6. Areas of low soil moisture accuracy (page 15)
Figure 7. Soil moisture accuracy over different NDVIs (page 16)
Figure 8. Soil moisture accuracy over different NDVIs (page 17)
Figure 9. CYGNSS reflectivity vs soil moisture in India (page 18)
Figure 10. CYGNSS vs SMAP time series (page 19)

Original Figures:
Figure 11. Plot of filtered TDS1 data (page 26)
Figure 12. TDS1 vs CYGNSS 06-04-2017 (page 28)
Figure 13. TDS1 vs CYGNSS 14-04-2017 (page 29)
Figure 14. TDS1 vs CYGNSS 22-04-2017 (page 30)
Figure 15. TDS1 vs CYGNSS 30-04-2017 (page 30)

List of Tables

Table 1. Results from filtered TDS1 data (page 27)

1 Introduction

Global Navigation Satellite System Reflectometry (GNSS-R) is a passive sensing technique that uses reflected L- band radio signals to determine properties of an area of reflection. This is an emerging Earth observation technique with its full potential yet to be reached. The behaviour of GNSS signals after reflecting off the Earth depends on many variables, with the most significant being the roughness and dielectric constant of the reflecting surface.

The bulk of the data used in GNSS-R investigations is taken from GNSS satellites themselves, with data from other sources often used in combination. It is common to see GNSS-R studies incorporate remote sensing Normalised Difference Vegetation Index (NDVI) data or data from in - situ sensors. GNSS data is plentiful as there are a large number of GNSS constellations in operation above the Earth. The USA's GPS, Russia's GLONASS ,the EU's Galileo and China's BeiDou system all provide continuous global coverage, while India's IRNSS and Japan's QZSS provide regional coverage. GNSS data is used in a huge variety of applications, including navigation systems, defence technologies and scientific research.

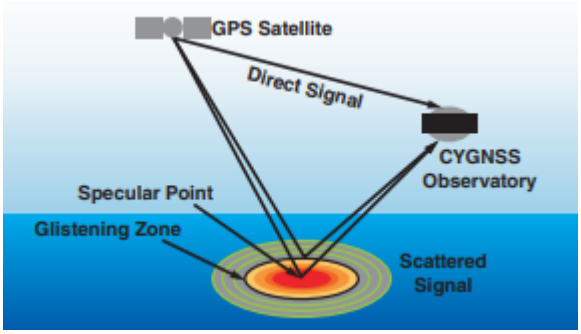


Figure 1: The working principle of GNSS-R. Reproduced from (Marsik et al., 2017).

GNSS-R is a passive sensing technique, meaning the sensor does not emit any electromagnetic radiation but only collects signals. A GNSS-R satellite will be placed in a Low Earth Orbit (LEO) at a lower altitude than GNSS satellites. When a GNSS satellite broadcasts a signal, the direct signal can

be picked up by the GNSS-R satellite. This signal then reflects off the Earth at an area referred to as the specular point and, a short time later, is picked up again by the GNSS-R satellite. The two signals can then be compared and any differences used to infer properties of the reflecting surface. This working principle is demonstrated in Figure 1. As the distances separating the transmitter, receiver and specular point are of similar scales, the GNSS-R technique is a form of bi-static radar.

There have been two main space - borne GNSS-R missions: TechDemoSat1 (TDS1) launched in 2014 and was designed by Surrey Satellite Technology (SSTL) and the Cyclone GNSS (CYGNSS) mission launched in 2016, operated by NASA.

TDS1 is a low power, low mass satellite which launched with eight experimental payloads on board, with measurement time to be split between them. As the reflectometry payload was only one of these experiments, GNSS-R data from the mission is only available for approximately one in every eight days. The satellite had an anetnna with 13.3 dBi peak gain and 3 dB half - beamwidth of 15 °. The orbital characteristics of TDS1 mean that it has a revisit time (time between subsequent observations of the same point on Earth) of a few days. TDS1 was eventually taken out of orbit through deployment of a drag sail in 2019.

The design of the CYGNSS mission built upon lessons learnt from TDS1. The mission was designed with a focus towards making wind speed retrievals using GNSS-R, and carried an identical reflectometry payload as the TDS1 mission before it. However, one crucial improvement was made. Rather than a single satellite, CYGNSS consists of a constellation of eight satellites, each in different orbital positions. This means that although the raw reflectometry data from CYGNSS is of approximately the same quality as TDS1, the data has a much higher temporal precision, with an average revisit time of only a few hours. Despite the similarity of the radio equipment itself, there were still some differences in the planned data processing and the increased temporal resolution opened up new scientific possibilities. As the mission's focus was on cyclone monitoring at the smallest possible temporal scale, and the vast majority of extreme weather events forming in the tropics, the CYGNSS satellites were set on an orbital course that does not run over the Earth's poles. Instead, CYGNSS orbits in a region of ± 37 ° latitude. This does limit the applications of CYGNSS data somewhat, as will be discussed later

in this report.

A full description of the working principles of GNSS is not required to understand the work presented here, although there is a great deal of literature available for the interested reader such as the European Space Agency's (ESA) book series (Subirana, Miguel, & Herna[Pleaseinsertintopreamble]ndez, 2013).

The Delay - Doppler Map (DDM) is a key concept that must be grasped to understand the process of GNSS-R. When the measured reflected signal is cross correlated with the PRN code of the transmitting satellite, it is done so over a period of time referred to as the coherent integration time. Note that the reflected signal actually reflects off an area around the specular point, due to variations in surface roughness on the Earth - this whole area is sometimes called the "glistening zone" when diffuse scattering is occuring due to surface roughness changes being larger than the wavelength of the incoming signal.

Fully diffuse scattering is commonly observed in GNSS-R measurements and in these scenarios the following bi-static radar equation holds, according to (Zavorotny, Gleason, Cardellach, & Camps, 2014) :

$$\left\langle |Y(\tau, f)|^2 \right\rangle = \frac{\lambda^2 T_i^2}{(4\pi)^3} P_t G_t \int\int \frac{G_r}{R_t^2 R_r^2} \chi^2(\tau, f) \sigma_0 dS \tag{1}$$

where T_i is the coherent integration time, $P_t G_t$ is the transmitter's Effective Isotropic Radiated Power, G_r is the receiver antenna gain pattern, R_t is the distance between the specular point and the transmitter, R_r is the distance between the specular point and the receiver, χ is the Woodward Ambiguity Function describing the range and Doppler selectivity of the coherent radar, σ_0 is the normalised bi-static radar cross section (BRCS) of the rough surface causing the diffuse scattering. This equation only holds in the case of entirely diffuse scattering, although corrections are available to take into account coherent reflections.

The DDM for a an observed signal emerges from Equation 1 in conjunction with models of the surface behaviour and electromagnetic scattering. Here the surface roughness must be taken into account along with the other dominant factor in the reflected signal's behaviour - the dielectric constant

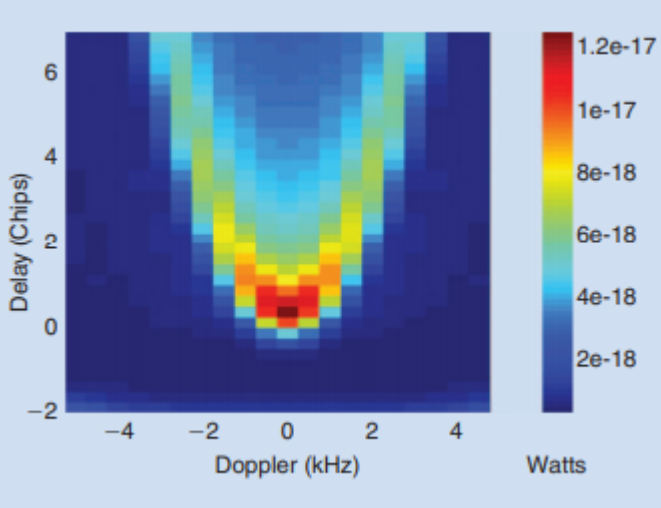


Figure 2: Typical DDM. Reproduced from Figure 9 in

of the surface. The dielectric constant will be different depending on several factors such as the water content of the surface, if it is covered in ice, how much vegetation cover there is etc. A typical DDM is shown in Figure 2.

A range of variables can be extracted from the DDM, with the most common variables used in GNSS-R studies being the Signal to Noise Ratio (SNR), the DDM peak value, or the variance in the DDM at its peak value.

Like with any remote sensing investigation, when evaluating the effectiveness of a GNSS-R retrieval a "ground truth" data set must be used as a comparison. In this work, data from NASA's Soil Moisture Active Passive (SMAP) mission was used to evaluate the GNSS-R data's accuracy. SMAP is a satellite designed specifically to measure soil moisture using radiometry and has a revisit time of about four days. Utilising both active and passive L - band radar measurements, SMAP is able to observe soil moisture as well as the freeze - thaw state of the ground.

Spaceborne GNSS-R is a fairly new technique and far from fully refined, especially in regard to the DDM processing and calibration. Interest in the field remains high, perhaps due to the advantages posed by GNSS-R remote sensing if the technique is able to be fully utilised and understood. For in-

stance, as GNSS-R is a passive sensing technique the satellite payloads can be light and small which reduces production and launch costs. Furthermore, due to being placed in Low Earth Orbits (LEO) to better collect direct GNSS signals, the satellites have very fast revisit times. This opens up a range of potential time - sensitive applications which traditional remote sensing techniques would not be able to pick up.

Funding interest in GNSS-R remains high, with the European Space Agency (ESA) announcing 30 million Euros of funding to be put towards the upcoming HydroGNSS mission (?, ?). This will be a 40kg satellite with the primary aim of studying hydrological variables, an area which GNSS-R has seen great success in previously.

2 Aims & Objectives

This project aims to evaluate the efficacy of two GNSS-R data sources when performing soil moisture retrievals. The evaluation of the performance of these two satellites (TDS1 and CYGNSS) will be made by comparison to data from the SMAP radiometry mission. This project aims to determine which of the two sources gives the most accurate soil moisture retrievals and what causes this difference.

The author aimed to get to grips with entirely new data sets and data formats by undertaking this project, and to hopefully come out with a useful application of the data.

With an eye towards the impact of this work, this research could influence the design of the next generation of GNSS-R satellites. If one variable is found to retrieve soil moisture more efficiently than another, this will improve the next generation of soil moisture retrieval algorithms.

If soil moisture detection is possible to the same accuracy as SMAP currently has, this work could pave the way for a shift from more costly radiometry satellites to cheaper constellations of GNSS-R based satellites used for the remote sensing of soil moisture.

3 Literature Review

GNSS-R was first demonstrated using a ground based receiver by (Martin-Neira, Caparrini, Font-Rossello, Lannelongue, & Vallmitjana, 2001) as a novel technique for performing ocean altimetry in the Netherlands. Even this first result was promising, with the ocean altimetry measurements having an RMS height error of only 30 cm over a five second measurement time. Ground - based GNSS-R sensors are still used for some near - field applications, most commonly still for ocean altimetry, such as in the work published by (Peng, Hill, Li, Switzer, & Larson, 2019). Here it was found that even a single ground GNSS-R altimetry station can do the job of several conventional mechanical tide gauges when detecting upcoming storm surges.

A new technique, known at the Interference Pattern Technique (IPT) was introduced by (Rodriguez-Alvarez et al., 2009) as a way of detecting soil moisture from GNSS-R signals. This early study made use of a ground - based reflectometer called a Soil Moisture Interference - pattern GNSS Obsevations at L - band (SMIGOL) reflectometer.

Following earlier work establishing the viability of the IPT, (Rodriguez-Alvarez et al., 2011) found the geophysical parameters of topography, soil moisture and vegetation height can be retrieved from vegetation - covered soils using GPS based GNSS-R measurements. As in (Rodriguez-Alvarez et al., 2009), the SMIGOL reflectometer was used to detect direct and reflected signals. The footprint of the SMIGOL reflectometer is defined as the "glistening zone" - the area around the specular point where the normalised scattering coefficient ≥ 0.63. This area contains scattering signals which have a significant contribution to the received power.

Three experiments were performed in two different regions of Spain to test the SMIGOL reflectometer's accuracy when measuring the soil moisture of vegetation - covered soils. The soil moisture was measured with a percentage error of 14.4%, 13.9% and 22% at the first, second and third experiments respectively. These errors are moderate, and expected considering the small glistening zones covered by the SMIGOL reflectometer. If averaged over a larger area, it is likely these errors would be lower.

After these initial tests, the feasibility of GNSS-R for detecting soil moisture was further investigated using aircraft. in 2014 (Egido et al., 2014a)

published the results of their study using GNSS-R instruments mounted on aircraft. Over two different sites in Italy, it was found that the GNSS signal's reflectivity polarisation ratio was strongly correlated with soil moisture content (SMC), with a reported Pearson correlation coefficient of 0.93 and a 0.2 dB / SMC sensitivity. This optimum correlation was recorded over areas with moderate vegetation cover and low surface roughness. This study also looked into using GNSS-R for above - ground biomass (AGB) measurement and found a similarly strong correlation between AGB and the copolarised reflection coefficient. This study was an important landmark towards space-

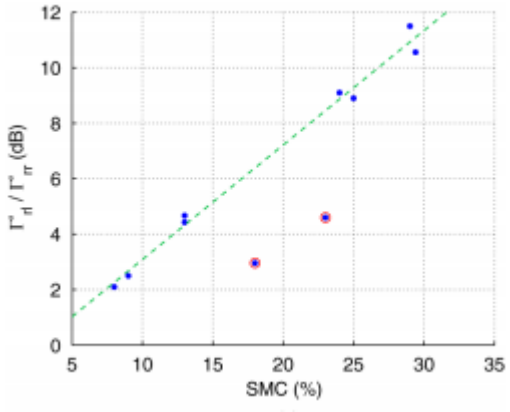


Figure 3: Relationship between reflectivity polarisation ratio and SMC with two outliers highlighted in red. Reproduced from Figure 9 in (Egido et al. 2014)

borne study of GNSS-R's capability, but was rather limited in scope. In particular, and as can be seen in Figure 3, very few data points were used to calculate the Pearson correlation coefficients. Also, two of the ten total data points have been excluded as outliers. The reason given for this exclusion is that they lie outside the main distribution of the data due to being measurements from a sunflower - covered field and a recently ploughed field with a resultingly high surface roughness.

A fantastic overview of the underlying principles of GNSS-R, as well as a comprehensive literature review (at time of publication) is provided by (Zavorotny et al., 2014). This is a seminal publication within the field of GNSS-R.

Following the successful launch of the TDS1 mission, the first published results focused on using the new spaceborne GNSS-R instrument to detect wind speeds over the ocean (Foti et al., 2015a). A wind speed retrieval algorithm using the bistatic radar equation and the signal - to - noise (SNR) ratio is developed and presented, with comparison to data from the NOAA Advanced Scatterometer (ASCAT) instrument. The key result from this paper is that when only SNRs above 3 dB are considered, wind speeds between 3 and 18 m/s are measured with a 2.2 ms precision. The authors acknowledge the algorithm is weaker when fed lower SNR data, and list a number of other factors which must be known to still achieve good retrieval. Fig-

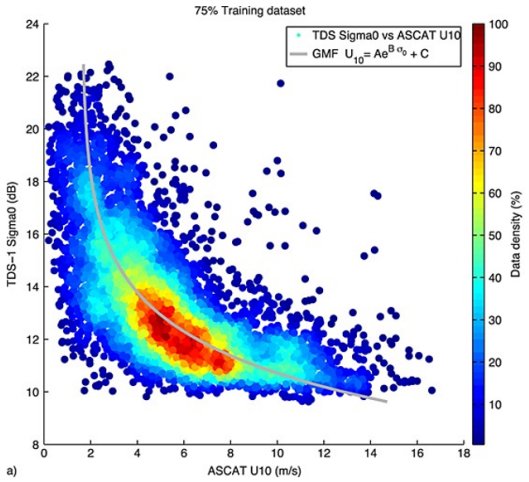


Figure 4: TDS1 derived wind speed vs ASCAT data, training set. Reproduced from Figure 4a in (Foti et al. 2015)

ure 4 shows the training dataset, encompassing 75% of the total data, used to train the wind speed retrieval algorithm. Figure 5 shows the validation dataset, encompassing 25% of the total data, used to assess the performance of the wind speed retrieval model. Analysis of the performance shows the algorithm can retrieve ocean wind speed with an RMSE value of 2.213 ms^{-1} between the TDS1 retrieval and ASCAT values. These first results from the TDS1 mission were a very promising start and began to show the potential of spaceborne GNSS-R data.

The first work to explore how TDS1 data can be used to detect soil moisture

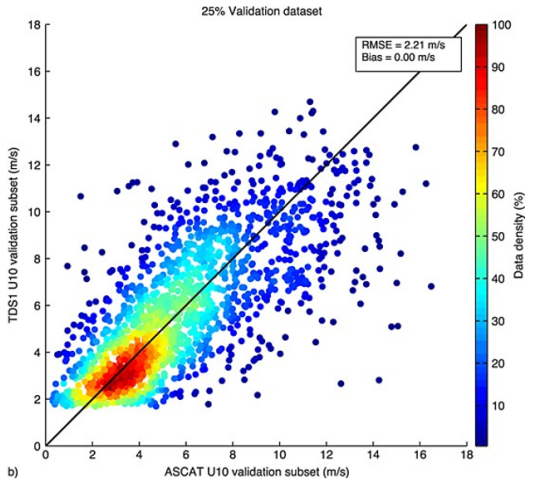


Figure 5: TDS1 derived wind speed vs ASCAT data, validation set. Reproduced from Figure 4b in (Foti et al. 2015)

was published in 2016, showing that retrieval is possible with good accuracy (Camps et al., 2016). This study was carefully designed to examine the soil moisture retrieval performance over a range of surfaces; different vegetation cover levels and different ranges of soil moisture. The soil moisture retrieval is validated against soil moisture data from ESA's Soil Moisture and Ocean Salinity (SMOS) mission, which uses microwave L - band measurements to measure soil moisture.

Since the GNSS-R instrument onboard TDS1 measures only reflected GNSS signals, the work presented in this study uses signals which are uncalibrated. To get around this issue, the soil moisture retrieval process is based on the ratio of the signal power at the peak of the DDM and the noise power in an area of the DDM with no signal present - this is the signal - to - noise ratio (SNR). To improve accuracy of the measurements, only data with an antenna gain $\geq$ 10 dB are used. This corresponds to reflections which are incident close to the nadir direction of the antenna, with incidence angle θ_i $\leq 15°$.

Two other datasets are used in the data analysis, namely NDVI data from NASA's MODIS mission and soil moisture data from SMOS as described previously.

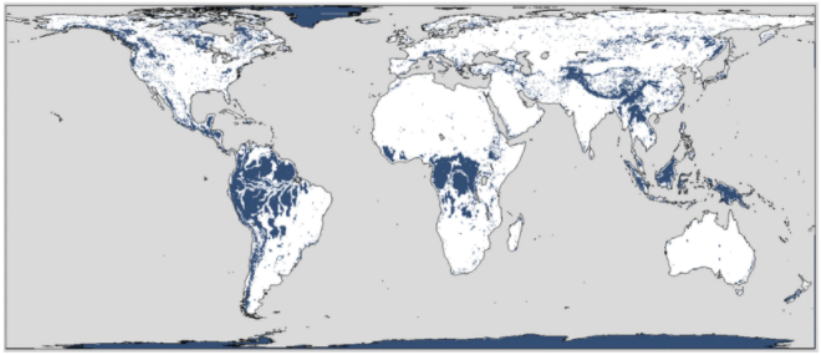


Figure 6: Areas of low soil moisture accuracy. Reproduced from Figure 5 in (Camps et al. 2016)

The authors set an accuracy threshold of 0.04 m^3m^{-3}, data points with errors higher than this would be discarded as unreliable measurements. A map of areas where this requirement is not consistently met is supplied by the authors, and reproduced here in Figure 6. This map provided a useful rough guide for which areas were appropriate for testing out any soil moisture retrievals during the work presented later in this report.

Correlations between the TDS1 SNR (measured in dB) and SMOS soil moisture for a range of vegetation types are shown in Figure 7 and Figure 8. Each plot includes a line of linear robust fit showing the relationship between the two data sets.

Results are shown for Evergreen Needleaf Forest, Evergreen Broadleaf Forest, Decideous Broadleaf Forest, Mixed Forest, Open Shrubland, Savanna, Woody Savanna and Grassland. From visual inspection, it is clear that forested areas show weak or no correlation between the TDS1 SNR and SMOS soil moisture. For nearly bare soils, it is found that TDS1 SNR has high sensitivity to soil moisture of approximately 38 dB/m^3m^{-3} with a Pearson correlation coefficient of 0.63. This is found to decrease with increasing opacity of the vegetation covering the soil. Interestingly, when data points with small NDVIs (meaning low vegetation cover) are excluded there is an increase of the sensitivity with increasing NDVI. The authors suggest this may be due to areas with higher vegetation cover tending to be in more humid regions, how-

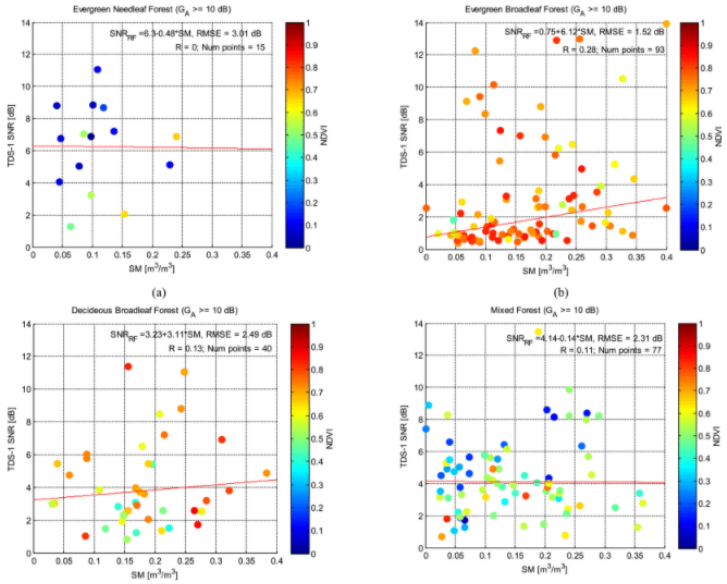


Figure 7: SNR - SMAP correlation for range of vegetation covers. Reproduced from Figure 12 in (Camps et al. 2016)

ever this relationship is not fully explored. Further work is required to fully understand this relationship.

Work published in 2018 by (Chew & Small, 2018) focused on establishing links between CYGNSS reflected power and soil moisture measurements from NASA's SMAP mission. For the comparison dataset from SMAP, the Level 3 Soil Moisture Product was used in 36 km grids. This data is available in a 9 km gridded format too, but the authors chose a coarser resolution as it was found to increase the number of matches between the two data sets by a factor of 16. It was found that there is a positive linear relationship between changes in the CYGNSS - derived reflectivity and changes in SMAP - derived soil moisture. However, there is no linear relationship between the absolute values of CYGNSS - derived reflectivity and SMAP - derived soil moisture.

The CYGNSS variable used is the forward scattered signal power, denoted $P_{r,eff}$. The authors somewhat justify the coarse grid choice by arguing that although a GNSS signal has a theoretical footprint of only about 0.5 km x 0.5 km, on rough surfaces this is much larger on the order of 10s of km square.

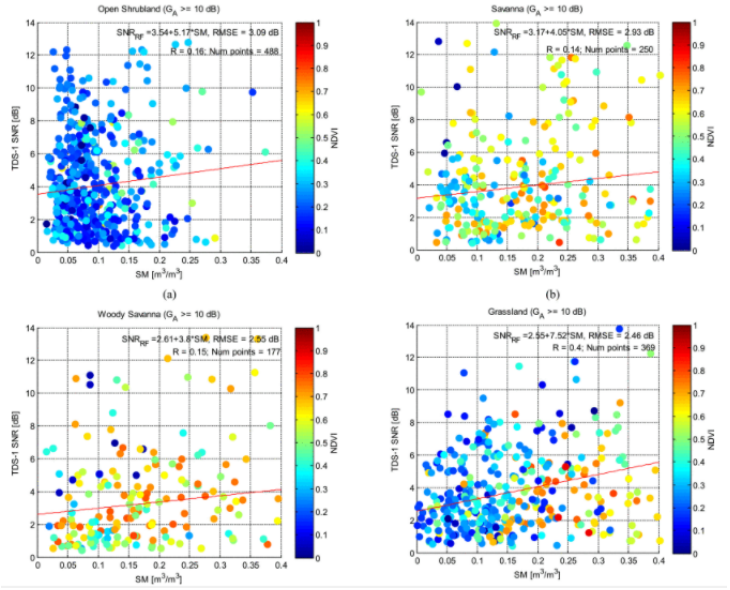


Figure 8: SNR - SMAP correlation for range of vegetation covers. Reproduced from Figure 13 in (Camps et al. 2016)

Some interesting seasonal variations in $P_{r,eff}$ were noted in tests on the Indian subcontinent. Between the months of May and August it "increased significantly in several regions" (Chew & Small, 2018). This seasonal change in soil moisture content is expected and would be observed in data from the TDS1 mission as well. Various processes could drive this change, such as changing evaporation due to higher temperatures and winds or changing levels of precipitation. Either way, this highlights an important facet of soil moisture data - it cannot be assumed to stay constant over time. For this reason, any comparison between GNSS-R derived soil moisture and ground truth measurements must be done within a short period of time as well as within close spatial proximity. Despite the large volumes of data input on both sides, this severely restricts how much of the data is directly comparable.

Many grid cells show the correlation between the change in $P_{r,eff}$ and change in soil moisture as being ≥ 0.7, with the highest results being recorded in wetter regions. Contrastingly, the regions with the lowest measured soil mois-

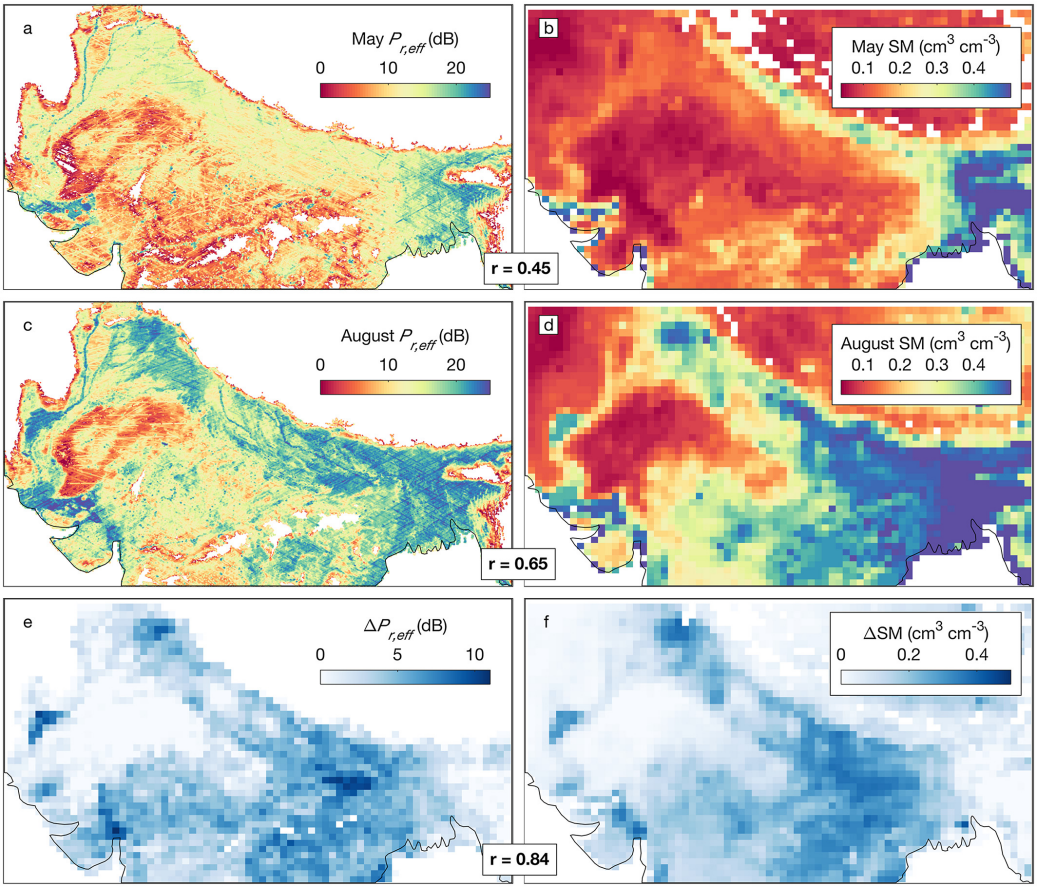


Figure 9: Observations of $P_{r,eff}$ and mean soil moisture between May and August 2017 over northern India. Reproduced from Figure 2 in (Chew & Small, 2018)

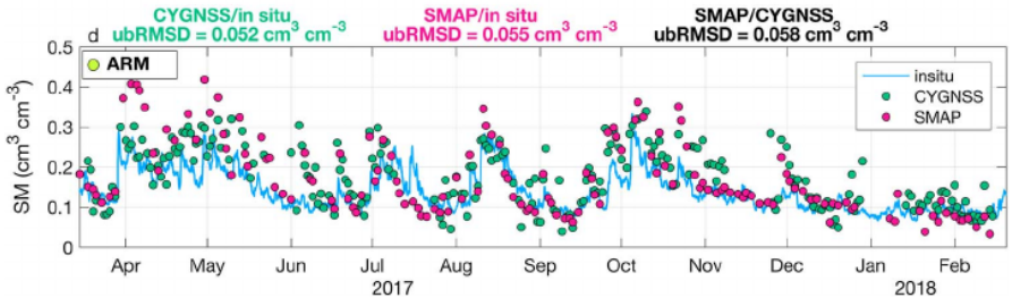


Figure 10: Time series CYGNSS reflectivity against SMAP soil moisture. Reproduced from Figure 5 in (Chew & Small, 2018)

tures showed the lowest Pearson correlation coefficients, with this weak result being driven by a lack of variability over time in the soil moisture of those regions. This is expected to give a lower correlation as the signal noise is more dominant in the reflectometry measurements.

A time series analysis of CYGNSS, SMAP and in - situ results for an example region are shown in Figure 10. The similarities between the three data sets are clear but it also clear that although CYGNSS and SMAP data points tend to agree well, this is not always in line with the in - situ measurements.

However, the authors argue for these results still having significant value. By converting the $P_{r,eff}$ values to direct soil moisture values via linear regression with SMAP data then comparing these to the mean SMAP soil moisture value at each grid region, the CYGNSS data is mapped onto an absolute scale. This allows the calculation of an unbiased RMS difference (ubRMSD) which showed CYGNSS data fares better than many other space - based sensors. This is a promising result that implies CYGNSS $P_{r,eff}$ data could be used in a soil moisture retrieval algorithm.

A study was published later that same year by (Kim & Lakshmi, 2018) utilising CYGNSS data to estimate soil moisture with comparison to SMAP soil moisture data and investigating if the two could be combined into a single soil moisture product. This is a similar approach to the previous study, however here a relative SNR (rSNR) variable is used for comparison against the SMAP data. The authors discuss the benefits of using data from CYGNSS

over TDS1, including the much faster revisit times of 2.8 (median) and 7.2 (mean) hours and that CYGNSS is online constantly, rather than TDS1 that only collects GNSS-R data for two out of every eight days.

The rSNR is calculated as

$$rSNR(x_i, y_i, t) = \frac{nSNR(x_i, y_i, t) - \mu nSNR_{dry}(x_i, y_i)}{nSNR_{wet}(x_i, y_i) - nSNR_{dry}(x_i, y_i)}\phi \tag{2}$$

where x_i and y_i are positions of the SMAP grid centroid, t is the time the measurement was taken and ϕ is the soil porosity at that point. The nSNR is the normalised SNR with maximum SNR being labelled "wet" and minimum being "dry".

When discussing data validation, the authors argue against using just the Pearson correlation coefficient to validate satellite data against in - situ observations. Four reasons are given, namely that the depth of the measurements do not exactly match, the uncertainty in the in - situ measurements are not taken into account, the in - situ data points are spatially sparse and the two data sets are at different spatial resolutions.

The main takeaway from this study is that CYGNSS soil moisture retrievals correlate well with SMAP data, even with the new rSNR variable. It also shows there is potential for a combined soil moisture product, with the high temporal resolution data from CYGNSS being used to fill in the gaps in soil moisture retrievals from SMAP.

In addition to soil moisture, a substantial number of studies using GNSS-R techniques to study other geophysical phenomena have been published over the past 12 years. These bear mentioning in this report but will not be examined in as detail as those relating directly to this project are.

GNSS-R has been successfully used to examine sea-ice concentration (Zhu et al., 2021), above-ground biomass (Egido et al., 2014b), ocean salinity (Liu, Wan, & Hong, 2021), ocean wind speed (Foti et al., 2015b) (Li, Yang, Yang, & Zheng, 2019) and sea surface roughness using both ground based (Marchan-Hernandez et al., 2010) and aircraft based (Germain et al., 2004) sensors. The utilisation of GNSS-R data from both TDS1 and CYGNSS for such varied applications shows another powerful side of the technique. It is

possible there are other parameters which may be retrievable using GNSS-R data, although further work would be required to determine this.

4 Methodology

The first work package in this project involved pre-processing and transformation of the three data sets from the TDS1, CYGNSS and SMAP missions.

Data from the TDS1 mission is downloadable via the FTP server from MERRByS. Surrey Satellite Technology publish a product manual with format descriptions and a data dictionary which is kept up to date when new data products are released. The document is now in its 7th iteration.

The variables used for soil moisture retrievals are based on the L1B data product. In the MERRByS release, these products are split into two parts: the L1B DDM file and L1B Metadata file. Both are in netcdf4 file format. All the variables needed for soil moisture retrieval here are taken from the metadata file and not from the ddm file, so the structure of the ddm file is not outlined here.

The metadata file has a two - tiered netcdf structure. The file contains attributes which describe the source of the data, such as the file ID and if the data was processed onboard or on the ground. It also has an array of variables which have information on the configuration of the receiver such as its clock status, position of the signal transmitter and the satellite's attitude (roll, pitch and yaw) when the reflected signal was received. The data is then subdivided into tracks, each with a corresponding netcdf group. Each track represents one specular point as it moves across the Earth's surface and TDS1 is capable of tracking several at once. These groups are named with a six digit number starting at 000000 and incrementing each time an integration is made. Each track group also has attributes describing configuration information about the recorded signal, such as the antenna gain, integration time and PRN code. The groups also have variables which contain the majority of information used in this work. Specifically the x, y, z positions of the specular point and transmitter, as well as the variable "DDMSNRAtPeakSingleDDM". This variable provides the SNR of the peak maximum DDM pixel.

The noise here is calculated in one of several ways depending on the data quality. The first choice is using the variable "MeanNoiseBox", whereby a noise box is defined within the DDM in a region above any reflections. If there are not enough rows in the DDM that contain no reflections then a different variable "MeanHighDoppler" is used. This is the mean pixel value at the first delay row and the first and last Doppler cells. As this is based on only those two pixels, it tends to have a high variance which makes it the less desirable of the two options.

Before analysis can begin, some work must be done in order to get this data into a more user friendly form. Using Python 3.8 scripts made available by SSTL under a permissive MIT license, data covering a user - defined time range can be downloaded from the MERRByS FTP server. This data can be opened using the h5py Python package before looping through the available track groups and extracting the desired variables of specular point latitude, specular point longitude, SNR, the integration mid - point time, the antenna gain and the signal incidence angle. These data points are converted into numpy arrays and serially concatenated until there are no further tracks. These are combined into one GeoDataFrame object from the GeoPandas library, with a geometry column generated from the x, y coordinates of the specular point. This process is repeated four times, once for each 6 hour portion of data in the selected day, with the final output being one GeoDataFrame which covers the entire 24 hour period.

The data is then filtered, based on suggestions found in literature by (Camps et al., 2016). Some of the suggestions here were not able to be implemented by the author with the resources available during this project, but three of the recommendations were implemented.

Firstly, the data is filtered such that measurements taken when the satellite was being eclipsed from the Sun by the Earth are removed. TDS1 has onboard sun sensors which are used to keep track of the satellite's attitude. Clearly, when the sun is obscured these sensors cannot function and this is known to have an adverse effect on the data. Note here that the SSTL data documentation is unclear which marker is used to denote the data was recorded when in an eclipse. The value takes either 1 or 0 but it is not explicitly stated what the values refer to - this is investigated later in the report. The variable containing this information is called *InEclipse* and is

not a track variable, meaning it must be linked to each track data point by comparing the times they were both recorded at.

Secondly, only signals where the antenna's gain towards the specular point was $\geq$ 10 dB. The variable containing this information is a track variable called *AntennaGainTowardsSpecularPoint*, which makes it much simpler to check against as it is already aligned with the rest of the dataset.

Thirdly, only signals with an incidence angle $\theta_i \leq 15$ ° are accepted, by removing data at larger reflecting angles the amount of scattering in the received signals is reduced. The variable here is also a track variable and is called *SPIncidenceAngle*. The effects of all three of these filtering techniques in different combinations were investigated as part of this project.

For ease of notation, results from the filtering tests are assigned a code containing any of the following four parts: "E0" filtered to only have data with corresponding *InEclipse* values of 0, "E1" filtered to only have data with corresponding *InEclipse* values of 1, "G" if filtered for antenna gain and "θ" if filtered for incidence angle. These can be combined together, for example data marked "E1G" has been filtered for $InEclipse = 1$ and antenna gain $\geq$ 10 dB but has no limit on the incidence angles.

Data from the CYGNSS mission is available through NASA's Earthdata portal and comes in netcdf4 format. The specific data product used is the UCAR/CU CYGNSS Soil Moisture Product, developed by Clara Chew and Eric Small as noted in (Chew & Small, 2018). Using the xarray Python package, this data is opened as a Dataset object before being trasformed to a DataFrame.

Some basic cleaning is done, such as removing any rows from the DataFrame which contain NaN values. The soil moisture measurement, longitude and latitude are all extracted before being merged into a GeoPandas GeoDataFrame with geometry generate from the longitude and latitude. This GeoDataFrame is now converted from EPSG 4326 to EPSG 3857 to be in units of m. This data is now in the same format as the TDS1 data, but with a much shorter process. This is largely due to the CYGNSS data source being a bespoke soil moisture product, whereas the TDS1 files are designed to be general and cover a range of use cases.

SMAP data is also available through NASA's Earthdata service, allowing dynamic selection of regions and timeframes to download over. The data is supplied in .h5 format which is an hdf5 format. The data under the key "Soil_Moisture_Retrieval_Data" is extracted into an h5py Dataset object. This object supports slicing via Numpy style indexing, so the soil moisture, longitude and latitude are easily extracted and appropriately renamed. Once again, the data is converted to the GeoPandas GeoDataFrame format with geometry which is generated from the latitude and longitude data.

The data is cleaned, with all points with negative values of soil moisture being removed. SMAP data points where errors occurred in the data processing (data of insufficient quality, high noise levels *etc.*) are given values of -9999, so a simple filter for only positive values is sufficient to screen these out. This data is then converted from EPSG 4326 to EPSG 3857 (units of m). This process is designed to combine an entire days worth of SMAP data into one GeoDataFrame.

With data from all three sources now cleaned and transformed into GeoDataFrame format, the next step is to match up data points which are in the same geographic location. This is done by using the GeoPandas *total_bounds* method to extract minimum and maximum x, y coordinates of the SMAP data for that day. A custom algorithm then generates polygons of user defined length to fill the space between these minimum and maximum points. Generally, these polygons are given dimensions of 36000 m by 36000 m as this corresponds to the size of the SMAP soil moisture pixels. These grid cells are geometries and are stored in another GeoDataFrame, each given a unique index.

A spatial left join is then performed between a GNSS-R data source (either TDS1 or CYGNSS) and the grid, assigning the matching grid index to each data point. The same spatial join is performed with the SMAP data and matching grid indices are assigned. It is simple to then match up the GNSS-R data to the SMAP data and compare the two.

The python file is set up to take an input of month and day (both as two digit strings) from the user. After the data was downloaded it was organised into a file structure, allowing the month and day strings to be formed into

filepaths that will automatically read in the TDS1, CYGNSS and SMAP data from that day. There are a few important things to note here. Firstly that although the data is all from the same temporal period, the TDS1 and CYGNSS datasets will not necessarily geographically match for that day. Both are geolocated with SMAP data for comparison directly to SMAP of course, but these comparisons may be taking place at very different locations. Furthermore, although many of the raw TDS1 and CYGNSS pixels would be located over the oceans (making soil moisture retrieval an impossibility), and these data sets are not directly filtered to be on land only, none of these data points make it through to the final comparison as no SMAP data exists over ocean pixels. This is because SMAP automatically removes data which it deems to have failed to retrieve soil moisture measurements - so there is nothing for the TDS1 or CYGNSS data to be compared to at these locations. There will, of course, be some edge pixels which consist partly of ocean, of pixels with lakes and rivers, but the overall effect of these is very small.

The work presented here contains data points taken from all across the Earth. When using temporally rich data such as GNSS-R, it makes little sense to confine an investigation to one geographical area.

5 Results & Discussion

This section of the report will discuss the two investigations conducted as part of this body of work. Firstly into filtering of the TDS1 SNR data to see if this affects the strength of its correlation with SMAP data, and secondly a comparison of the accuracy of TDS1 and CYGNSS soil moisture retrievals over a one month period of April 2017.

5.1 TDS1 Filtering

The results from applying ten different data filters to the TDS1 SNR data are shown in Figure 11. The meanings of these codes are explained in Section 5 of this report. The number of data points retained and the Pearson correlation coefficient between the TDS1 SNR and SMAP soil moisture for each different filter are tabulated in Table 1. The colourmaps on each figure

represent the data density - with lighter, more yellow colours representing areas of the highest data density while darker and more purple areas represent areas with the lowest data density. This was done to help appreciate the distribution of the data properly, as some of the data points lie very close to each other and can be hard to distinguish without a colour map such as this one.

It can be seen in Figure 11 that, generally, the more filters that are applied to the data the more strongly correlated the two variables become and the distribution appears more narrow and linear. This visual inspection result is confirmed by the statistical analysis, with the Pearson correlation coefficient of the unfiltered data ("None") $p = 0.293974$ showing a weakly positive relationship while the Pearson correlation coefficient of the most filtered data ("E1Gθ") is 0.586591 showing a moderately positive relationship. The effectiveness of these filters is clear. This was an expected result as shown in the literature and this small investigation provides further confirmation.

As discussed in the methodology section, there was some uncertainty about the correct marker to use to determine if the data was taken when the satellite was in eclipse from the Sun. It is clear from this investigation that keeping data points with the variable $InEclipse = 1$ is the correct choice. In all cases, only using data with the *InEclipse* marker $= 0$ results in worse performance than using $InEclipse = 1$. In fact, just using this filter ("E0") results in a weaker correlation than using no filter at all. This clears up a point of ambiguity from the TDS1 documentation and ensured that data was properly filtered for the comparison against CYGNSS.

5.2 Retrievals

5.2.1 6th April 2017

The first side by side comparison comes from 06/04/2017, where the TDS1 data had a Pearson correlation coefficient p of 0.597 and the CYGNSS data had a Pearson correlation coefficient p of 0.904. The plot of this data is shown in Figure 12. A visual inspection here shows results in - line with what the correlation suggests. The CYGNSS data is very tightly grouped while the

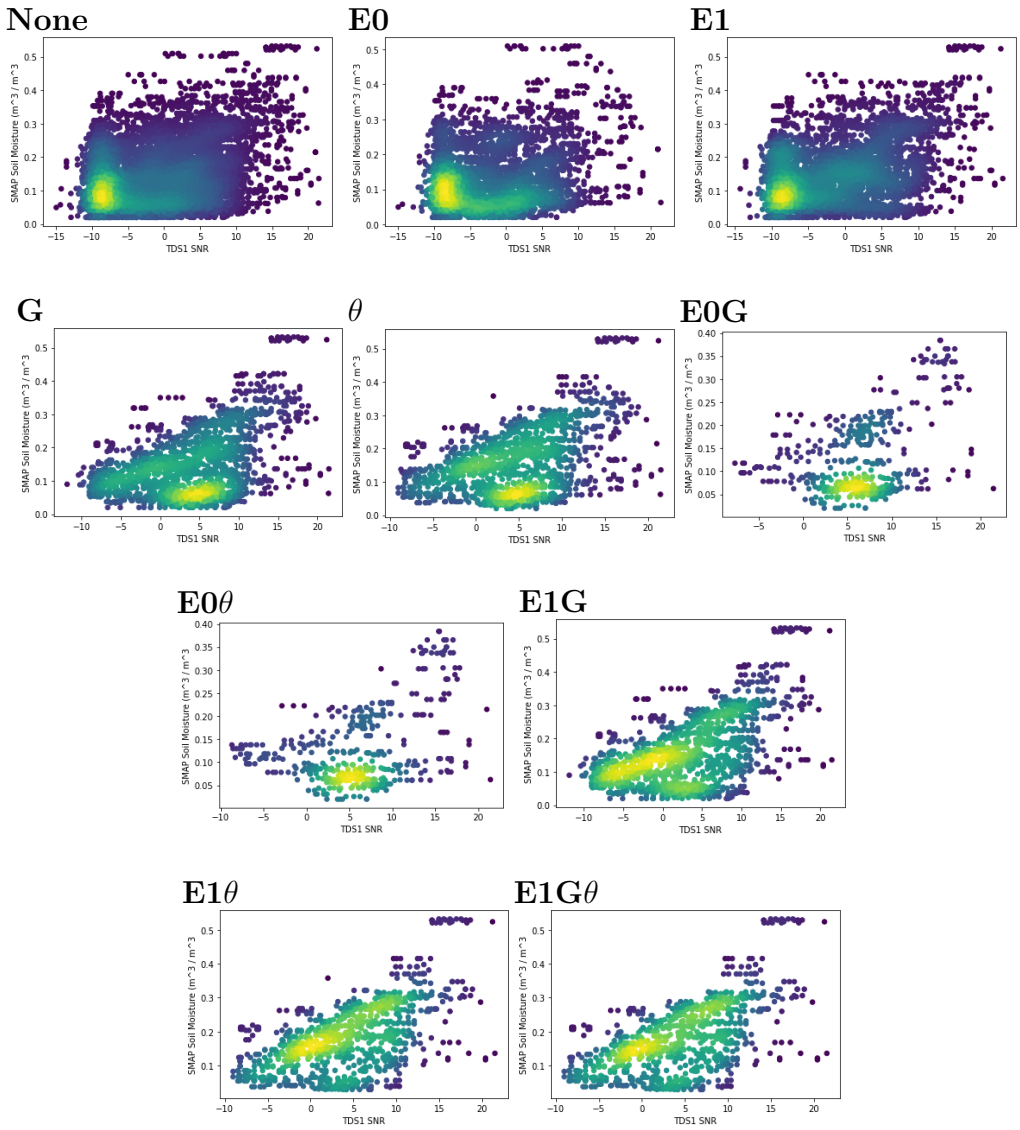


Figure 11: Figures showing SMAP soil moisture data (Y axis) against TDS1 soil moisture retrievals (x axis) for a variety of filters.

TDS1 data, although moderately positively correlated with the SMAP data, has a weaker relationship than the CYGNSS data.

Code	Points Retained	Pearson Coefficient
None	327513	0.293974
E0	207836	0.189162
E1	119677	0.361259
G	44699	0.482361
θ	27986	0.477919
E0G	23556	0.433105
E0θ	15110	0.407598
E1G	21143	0.560046
E1θ	12876	0.575283
E1Gθ	11367	0.586591

Table 1: Results from different filter combinations on TDS1 data from 14/04/2017. Codes are described in Section 6.

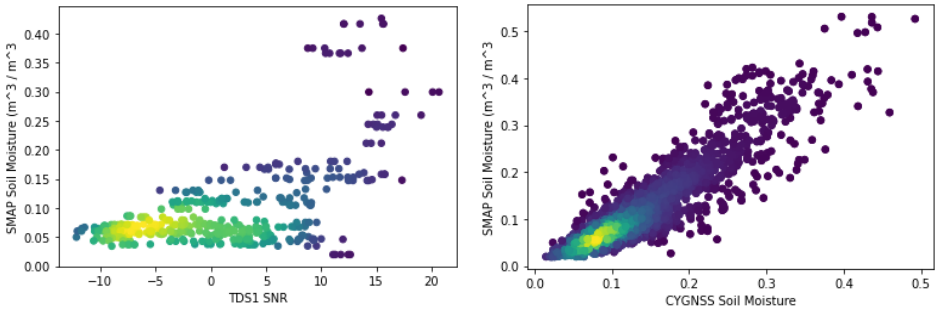


Figure 12: TDS1 (left) and CYGNSS (right) soil moisture retrievals compared to SMAP soil moisture. Data from 06/04/2017

5.2.2 14th April 2017

Moving to the next window of TDS1 GNSS-R operation, results from the 14/04/2017 are shown in Figure 13. The TDS1 data had $p = 0.586592$ while CYGNSS showed $p = 0.927375$. The behaviours of both graphs are similar to the data from 06/04/2017, although there is a much larger volume of

TDS1 data available on this day comparde to 06/04/2017. This could be due to a variety of factors including the path of the specular points tracking over smoother terrain resulting in fewer discarded results, inclement weather affecting the data on 06/04/2017 *etc.*

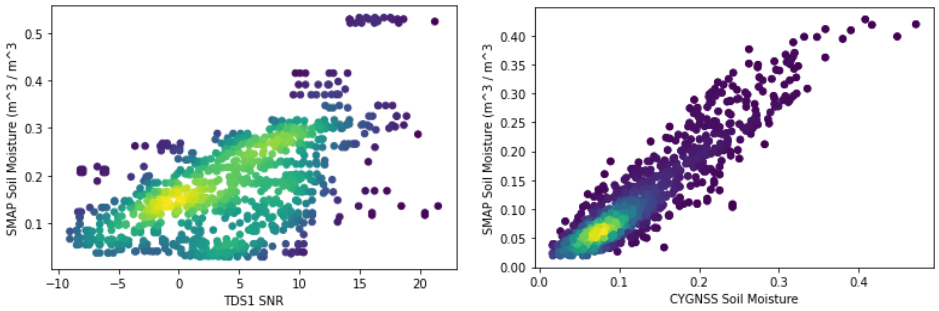


Figure 13: TDS1 (left) and CYGNSS (right) soil moisture retrievals compared to SMAP soil moisture. Data from 14/04/2017.

5.2.3 22nd April 2017

Data from 22/04/2017 is shown in Figure 14. The CYGNSS results show a similar grouping to the last set with a very comparable p value of 0.923107. The outstanding result here is the TDS1 data which has lost all trace of a positive correlation. With a Pearson correlation coefficient of 0.057063, there is no discernible correlation between the TDS1 retrieved soil moisture and the SMAP soil moisture. Such an extreme change in the pattern is quite common in the very noisy GNSS-R data sets, and even though the data has been somewhat filtered it is still not enough to get a passable result here. It is quite possible external factors affected the data collection on this day, such as high levels of radio frequency jamming, extremely rough terrain or a surface with dense vegetation cover.

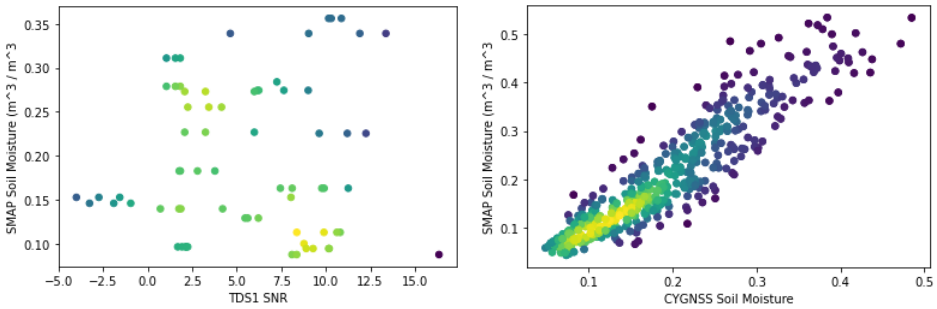


Figure 14: TDS1 (left) and CYGNSS (right) soil moisture retrievals compared to SMAP soil moisture. Data from 22/04/2017.

5.2.4 30th April 2017

Results from 30/04/2017 are shown in Figure 15. The TDS1 data had $p = 0.430685$ while the CYGNSS data had $p = 0.929571$. Once again, the CYGNSS data shows a much stronger positive correlation while TDS1 shows a moderate positive correlation with the SMAP soil moisture. Whatever occurred in the previous data set had clearly been resolved by the time these readings were taken as the TDS1 result is broadly in line with the other recorded results.

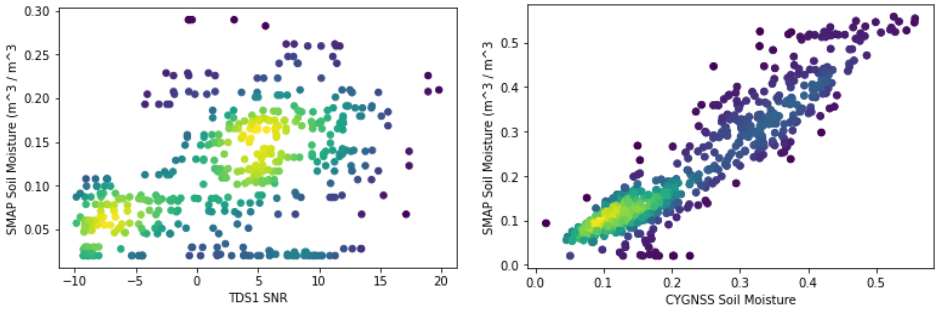


Figure 15: TDS1 (left) and CYGNSS (right) soil moisture retrievals compared to SMAP soil moisture. Data from 30/04/2017.

5.3 Comparisons

It is clear from the direct comparisons that the CYGNSS soil moisture retrieval is far superior to the TDS1 soil moisture retrieval, at least in the way it was implemented in this study.

For each of the four selected 24 hour time periods the Pearson correlation coefficient between the retrievals and SMAP was much higher for CYGNSS than TDS1. The data from CYGNSS never showed a p value lower then 0.9, while the p values for the TDS1 data never exceeded 0.6.

This result may appear disappointing for TDS1, but this value is in line with values taken from literature. In the first published soil moisture investigation using TDS1, (Camps et al., 2016) found that the TDS1 SNR has a Pearson correlation coefficient of 0.63 over nearly bare soils, which is close to the maximum 0.59 recorded in this invetigation. Note that the published result is a correlation with data from the SMOS mission, not SMAP, and the investigation is also split to group data into vegetation cover types. These factors make the results not directly comparable but a similar result would still be expected - it is reassuring that this was observed.

The performance of the CYGNSS data exceeds published results, such as those discussed earlier in this report. (Chew & Small, 2018) state that the CYGNSS to SMAP correlation exceeded 0.7 in the majority of grid cells. Both results do show a strong positive correlation between the two variables which is good, however the small performance improvement noted in this work here is troubling. It is not immediately clear what the cause of this improved correlation is but there are several possibilities. The CYGNSS data used in this study is the same data set used by (Chew & Small, 2018) which means the difference is either a factor of the areas being studied or due to additional processing. Due to the limited scope of this project and the fact that the CYGNSS data is explicitly a soil moisture product, a thorough investigation into the effects of filtering the data was not undertaken in the same way that it was for the TDS1 data. Aside from some common sense checks and removing NaN values, no further filtering was applied to the CYGNSS data. Further work would be required to fully investigate the cause of this discrepancy.

With all these values it is worth noting there are several error sources present in this data, from atmospheric effects to assumptions about the ground cover of measured pixels. These are hard to quantify exactly during a project such as this and would require further work to properly establish.

6 Conclusion

The main conclusion to draw from this work is clear; data from the CYGNSS soil moisture product has a stronger correlation with SMAP soil moisture measurements than TDS1 SNR data does. This was expected from literature, and has been confirmed again here, with TDS1 giving a maximum Pearson correlation coefficient value of 0.597 while the maximum Pearson correlation coefficient value for the CYGNSS mission was 0.929. This value slightly exceeds previously published data. This could be due to a greater degree of averaging due to CYGNSS's higher temporal resolution or, most likely, it is due to the more advanced processing put into the CYGNSS soil moisture product. This is based on reflectivity and filtered in myriad ways, while the TDS1 data used here uses an SNR variable and had to be manually filtered. There is no publicly available TDS1 soil moisture product to allow direct comparison.

Confirming another literature result, the filtering of TDS1 SNR data for points which are not in solar eclipses, have high antenna gain levels and small incidence angles dramatically improves the effectiveness of the soil moisture retrievals. With no filtering, a maximum Pearson correlation coefficient value of 0.294 was achieved, whereas with all three filters applied the Pearson correlation coefficient increased to 0.587. This final maximum value is in line with previously published literature.

7 Future Work

Several areas of further study identified themselves during this work. Simply extending the time periods correlations were examined over would provide data to a higher degree of accuracy.

In specific regard to filtering the TDS1 data, one thing noted by (Camps et al., 2016) that was not implemented in this study is accounting for different areas of the world having different levels of radio background noise and different levels of radio jamming activity. In particular, active warzones feature heavy radio band interference which will have a large effect on any data. A thorough investigation into the effects of man - made interference would improve data quality.

As touched on in the methodology section, conventional TDS1 soil moisture studies tend to use a reflectivity variable for soil moisture retrievals, rather than the SNR variable used in this study. The SNR was discounted by (Camps et al., 2016) as it has a higher uncertainty than the reflectivity when compared to SMOS soil moisture data. It would be interesting for this same work to be repeated with reflectivity data compared to SMAP soil moisture, to allow an even more direct comparison of TDS1 to CYGNSS.

Further work could also be done on the comparison dataset. The SMAP data used in this work is set to 36 km square grids. These were used as they conform nicely with the 36 km square grids the CYGNSS data comes in, but less coarse SMAP data is available. Specifically, the SMAP soil moisture dataset is available in a 9km square gridded format. These smaller pixel sizes would lower error in the SMAP dataset and may have a particular improvement on the TDS1 SNR data which is not gridded - allowing better assumptions to be made about the surface pixel.

References

Camps, A., Park, H., Pablos, M., Foti, G., Gommenginger, C. P., Liu, P.-W., & Judge, J. (2016). Sensitivity of gnss-r spaceborne observations

to soil moisture and vegetation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(10), 4730–4742. doi: 10.1109/jstars.2016.2588467

Chew, C. C., & Small, E. E. (2018). Soil moisture sensing using spaceborne gnss reflections: Comparison of cygnss reflectivity to smap soil moisture. *Geophysical Research Letters*, *45*(9), 4049–4057. doi: 10.1029/2018gl077905

Egido, A., Paloscia, S., Motte, E., Guerriero, L., Pierdicca, N., Caparrini, M., ... Floury, N. (2014a). Airborne gnss-r polarimetric measurements for soil moisture and above-ground biomass estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(5), 1522–1532. doi: 10.1109/jstars.2014.2322854

Egido, A., Paloscia, S., Motte, E., Guerriero, L., Pierdicca, N., Caparrini, M., ... Floury, N. (2014b). Airborne gnss-r polarimetric measurements for soil moisture and above-ground biomass estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(5), 1522–1532. doi: 10.1109/jstars.2014.2322854

Foti, G., Gommenginger, C., Jales, P., Unwin, M., Shaw, A., Robertson, C., & Roselló, J. (2015a). Spaceborne gnss reflectometry for ocean winds: First results from the uk techdemosat-1 mission. *Geophysical Research Letters*, *42*(13), 5435–5441. doi: 10.1002/2015gl064204

Foti, G., Gommenginger, C., Jales, P., Unwin, M., Shaw, A., Robertson, C., & Roselló, J. (2015b). Spaceborne gnss reflectometry for ocean winds: First results from the uk techdemosat-1 mission. *Geophysical Research Letters*, *42*(13), 5435–5441. doi: 10.1002/2015gl064204

Germain, O., Ruffini, G., Soulat, F., Caparrini, M., Chapron, B., & Silvestrin, P. (2004). The eddy experiment: Gnss-r speculometry for directional sea-roughness retrieval from low altitude aircraft. *Geophysical Research Letters*, *31*(21). doi: 10.1029/2004gl020991

Kim, H., & Lakshmi, V. (2018). Use of cyclone global navigation satellite system (cygnss) observations for estimation of soil moisture. *Geophysical Research Letters*, *45*(16), 8272–8282. doi: 10.1029/2018gl078923

Li, X., Yang, D., Yang, J., & Zheng, G. (2019). Assessment of cygnss coastal wind speed retrieval. *2019 Photonics amp; Electromagnetics Research Symposium - Fall (PIERS - Fall)*. doi: 10.1109/piers-fall48861.2019.9021587

Liu, B., Wan, W., & Hong, Y. (2021). Can the accuracy of sea surface salinity measurement be improved by incorporating spaceborne gnss-

reflectometry? *IEEE Geoscience and Remote Sensing Letters*, *18*(1), 3–7. doi: 10.1109/lgrs.2020.2967472

Marchan-Hernandez, J. F., Valencia, E., Rodriguez-Alvarez, N., Ramos-Perez, I., Bosch-Lluis, X., Camps, A., ... Marcello, J. (2010). Sea-state determination using gnss-r data. *IEEE Geoscience and Remote Sensing Letters*, *7*(4), 621–625. doi: 10.1109/lgrs.2010.2043213

Marsik, F., Hanson, H., & Ward, A. (2017). *Cygnss cyclone global navigation satellite system mission brochure.* NASA. Retrieved from `https://clasp-research.engin.umich.edu/missions/cygnss/docs/cygnss`$_b$*rochure*$_1$216

Martin-Neira, M., Caparrini, M., Font-Rossello, J., Lannelongue, S., & Vallmitjana, C. (2001, Jan). The paris concept: an experimental demonstration of sea surface altimetry using gps reflected signals. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(1), 142–150. doi: 10.1109/36.898676

Peng, D., Hill, E. M., Li, L., Switzer, A. D., & Larson, K. M. (2019). Application of gnss interferometric reflectometry for detecting storm surges. *GPS Solutions*, *23*(2). doi: 10.1007/s10291-019-0838-y

Rodriguez-Alvarez, N., Bosch-Lluis, X., Camps, A., Vall-llossera, M., Valencia, E., Marchan-Hernandez, J., & Ramos-Perez, I. (2009). Soil moisture retrieval using gnss-r techniques: Experimental results over a bare soil field. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(11), 3616–3624. doi: 10.1109/tgrs.2009.2030672

Rodriguez-Alvarez, N., Camps, A., Vall-llossera, M., Bosch-Lluis, X., Monerris, A., Ramos-Perez, I., ... et al. (2011). Land geophysical parameters retrieval using the interference pattern gnss-r technique. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(1), 71–84. doi: 10.1109/tgrs.2010.2049023

Subirana, J. S., Miguel, J. Z. J., & Herna´ndez, P. M. (2013). *Gnss data processing.* ESA Publications Division.

Zavorotny, V. U., Gleason, S., Cardellach, E., & Camps, A. (2014). Tutorial on remote sensing using gnss bistatic radar of opportunity. *IEEE Geoscience and Remote Sensing Magazine*, *2*(4), 8–45. doi: 10.1109/mgrs.2014.2374220

Zhu, Y., Tao, T., Zou, J., Yu, K., Wickert, J., & Semmling, M. (2021). Spaceborne gnss reflectometry for retrieving sea ice concentration using tds-1 data. *IEEE Geoscience and Remote Sensing Letters*, *18*(4), 612–616. doi: 10.1109/lgrs.2020.2982959

APPENDIX

Project management timeline and Data Management Plan (DMP) for the project are included as an appendix to this document.

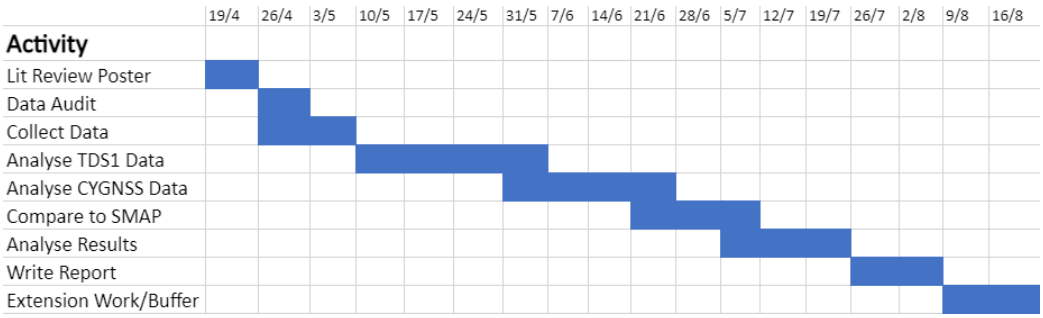


Figure 16: Project management Gantt chart.