Masters Program in Geospatial Technologies

HIGH-RESOLUTION SOIL MOISTURE RETRIEVAL USING SENTINEL-1 DATA FOR MONITORING REGENERATIVE AGRICULTURAL PRACTICES

A feasibility study from Alentejo, Portugal.

Syver Jahren Petersen

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

HIGH-RESOLUTION SOIL MOISTURE RETRIEVAL USING SENTINEL-1 DATA FOR MONITORING REGENERATIVE AGRICULTURAL PRACTICES

A feasibility study from Alentejo, Portugal.

Supervised by:

Joaquín Torres-Sospedra Ph.D.

Prof. Dr. Steffen Kuntz

Prof. Dr. Hanna Meyer

March 2022

DECLARATION OF ORIGINALITY

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or not published) are referenced.

This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere.

Lisbon, 24.02.2022

Syver Jahren Petersen

ACKNOWLEDGEMENTS

I would like to extend my gratitude to everyone that helped me and supported me throughout the process of this thesis. Of course, a big thank you to my supervisors Steffen Kuntz, Hanna Meyer, and especially Joaquín Torres Sospedra for all your time and the exceptional close guidance you have given me throughout. Thank you Prof. Maria da Conceição Gonçalves for lending me your TDR Trace, and thank you Prof. Marco Painho for sharing your contacts. Thank you David De Brito for showing me the wonderful work you are doing and allowing me access to collect data at Terramay. Last, but not least, thank you Fabio Volkmann at Climate Farmers for all the time, guidance, and enthusiasm you shared with me from the early stages of the project.

HIGH-RESOLUTION SOIL MOISTURE RETRIEVAL USING SENTINEL-1 DATA FOR MONITORING REGENERATIVE AGRICULTURAL PRACTICES

A feasibility study from Alentejo, Portugal.

ABSTRACT

Timely, reliable, and cost-efficient information about soil moisture is important for supporting agricultural practitioners in monitoring the impact of alternative agricultural practices. Regenerative agriculture is increasingly gaining traction; however, farmers lack easy access to information on key agricultural parameters such as soil moisture. Therefore, this study seeks to explore the feasibility of soil moisture estimation at high-resolution (around 10 m) using Sentinel-1 remote sensing radar data. A machine learning model was developed using a random forest regression algorithm with a combination of SAR-based, topography and Seninel-2 optical-based data as inputs. Through a k-fold cross-validation of the model, an average r-squared (R^2) of 0.17, a root mean squared error (RMSE) of 3.51 (% VMC), and an mean absolute percentage error (MAPE) of 83.34, was achieved.

LIST OF ACRONYMS

DEM - Digital elevation model

MAE - Mean absolute error

MAPE - Mean absolute percentage error

VMC - Volumetric moisture content

QGIS - Quantum GIS

R^2 - R (correlation) squared

RMSE - Root mean squared error

SAR - Synthetic aperture radar

SNAP - Sentinel Application Platform

TDR - Time domain reflectometry

TF - Terrain Flattened

VV - Vertical – Vertical

VH - Vertical – Horizontal

INDEX OF THE TEXT

INDEX OF TABLES

INDEX OF FIGURES

1. Introduction

1.1 Thesis organisation

The thesis is organised into four chapters. In the first chapter, I introduce the background and motivation behind this thesis, the problem I'm trying to solve and the objectives of the work. I also present the research questions, a brief introduction to the methodology and approach, and elaborate on what the expected contributions of the research will be to the broader literature. In the second chapter I perform a review of relevant literature and discuss previous findings on the topics of soil moisture retrieval using SAR data, and based on this, I stake out the path of this study, and how I seek to build on what has been done and contribute to filling gaps in the present knowledge. In the third chapter I elaborate on the choices of data and machine learning model. I further present the how I process my SAR data and how optical-based indices and topographic data are implemented in the model, as well as model training, testing and evaluation procedures. In the fourth chapter I present and discuss the results and evaluate the performance of the model. While in the final fifth chapter I discuss the overall contributions of the study, attempt to compare it to other similar studies, discuss the potential applicability of the model to practitioners, potential points of improvement, as well as elaborate paths for future research.

1.2 Background

This study is written in collaboration with Climate Farmers, an industry leading company, working with farmers to increase adoption of regenerative agricultural practices (Climate Farmers n.d a). The goal of the study is to explore the feasibility of monitoring soil moisture using open access remotely sensed SAR data, at a high enough resolution for it to be a useful tool for Climate Farmers as an organisation and the collective of farmers they are trying to help.

In the following sections, I will further introduce the motivation behind the study, the objectives, and what I hope to contribute to the research literature.

In recent years, regenerative agriculture is increasingly being applied as a strategy for improving the health of farm soil and surrounding ecosystems. According to Climate Farmers regenerative agriculture can be defined as "agricultural practices which enhance

and improve soils' capacity to deliver the above soil functions and increase the long-term resilience of its functions" (Climate Farmers, n.d b, para. 6). Soil functions refers to the ability of the soil to produce food and provide ecosystem services, while ecosystem services refer to general ecological processes or functions of value to people and society at large, such as such providing food, clean air and water (IPPC, 2019).

Especially in the context of the looming global climate crisis, the ability of soil to sequester carbon is increasingly coming into focus for policy makers, researchers, and the agriculture industry. Largely this is because effective methods for removing carbon from the atmosphere are challenging to find (IPPC, 2019). This has also resulted in a recent explosion of interest in and active promotion of regenerative agriculture by civil society and NGOs as well many of the key multi-national commercial actors (Giller et al. 2021).

Soil moisture as a factor in regenerative agriculture.

The Global Climate Observing System (GCOS), one of the major actors in climate monitoring, names soil moisture as one of their 54 Essential Climate Variables (ECVs) (GCOS, n.d). The Intergovernmental Panel on Climate Change (IPCC) (2019), highlights that "precipitation, by affecting soil moisture content, is considered to be the principal determinant of the capacity of drylands to sequester carbon" (p.271)

Half of the total 119 Gt carbon emitted into the Earth's atmosphere from the terrestrial ecosystem is attributed to soil microbial respiration (IPCC, 2019), meaning the CO2 produced by biological activities of soil organisms. To put things into perspective, this process makes up 10 times more of the CO2 in the global carbon cycle than fossil fuel combustion annually (Phillips & Nickerson, 2015).

Decomposition of so-called soil organic content, meaning the organic content of the soil, which is the part of the soil that contains carbon, is a major factor in global net CO2 emissions (IPCC, 2019). On global scale industrial agriculture is a major contributor to soil organic content decomposition due, among others, the use of pesticides that kill microorganisms in the soil and tillage that destabilise the soil structure and leaves the soil more exposed to the sun and other weather, which in turn kills of organic organisms in the soil (Hes & Rose, 2019). Regenerative agriculture largely focuses on practises that increase soil organic content and reverse the loss cycles often associated with modern industrial farming.

Soil moisture plays a key role in the decomposition of soil organic content as microbial processes rely on water to survive. However, studies have shown that some land restoration projects, with the intention of increasing ecosystem health, has had negatively impacted soil moisture rather than increased it. For example, Deng et al. (2016), studying 83 sites in eight provinces in the north of China found that "changes in land use for restoration of ecosystems led to severe depletion in soil moisture levels" (p.1).

SAR remote sensing and agriculture

Remote sensing based on satellite data for monitoring agriculture has long been a useful tool for the agricultural industry, for monitoring crops and other agricultural targets (Lui et al. 2019).

Soil moisture, as a key feature of agricultural production, has received a great degree of research attention as something possible to monitor using satellite-based data. Remote sensing can provide an opportunity for characterizing the spatial and temporal structure and dynamics of soil moisture but is somewhat limited in terms of providing a high degree of detail (Ma et al., 2020).

Most models resulting from this research only enable retrieval of soil moisture values at a very coarse spatial resolution, mostly around 25-50 km (Peng et al., 2021). For a farmer however, data at this scale has limitations for being practically applied to agricultural monitoring and planning. There is therefore a need to explore how to produce data at higher resolutions, preferably down to the field scale and below. Among satellite-based remote sensing approaches, synthetic aperture radar (SAR) data have shown the most potential in providing soil moisture estimates at a high enough resolution to be useful for small scale planning and monitoring purposes (Ma et al., 2020). This is largely due to the inherent properties of SAR data not being influenced by weather conditions, sensitivity to geometrical structures, as well as dielectric properties of objects and surfaces. Furthermore, in the case of some agricultural targets, SAR has a degree of penetration ability, for example, through vegetation (Lui et al., 2019).

1.3 Problem statement and motivation to do this work

The European Space Agency (ESA) Sentinel-1 C-band SAR mission grants free access to high-resolution data (around 10 m) every six days (ESA, n.d a). Due to the limited accessibility of

commercial high-resolution SAR images and the mostly shorter wavelengths of these, with less ability to penetrate vegetation canopies especially, Sentinel-1 data has largely been the focus of studies on high-resolution soil moisture retrieval. The simple fact that Senitinel-1 data is freely accessible, makes research on the use of this data for modelling phenomena such as soil moisture more relevant to broader society than commercial SAR data. Considering this openly accessible stream of data, if soil moisture retrieval is possible using Sentinel-1 data, this could be an affordable and powerful way of monitoring soil moisture developments over time at field level and in turn using this for evaluation of the effects of regenerative agriculture.

Thus, the motivation and rationale behind the study is that if remote sensing-based soil moisture retrieval is possible, it could potentially lower the cost of acquiring information about soil moisture in the regenerative farming industry. An improvement in information access that hopefully could play a small role in improving agricultural practices that in ultimately, in extension, would mitigate against climate change.

1.4 Research question and objectives

The main objective of the study is to statistically explore the feasibility of using Sentinel-1 data for monitoring soil moisture at around 10 meters resolution. In more concrete terms doing a case study of soil moisture retrieval in two small fields at a farm in Alentejo, Portugal, over a limited time-period of one month. The feasibility of soil moisture retrieval is to be evaluated through the development of a site-specific machine learning-based model for soil moisture prediction, leveraging Sentinel-1 data in combination with other topographic data and Sentinel-2 optical-based data.

Research questions

The current study seeks to answer the following questions:

1) Is it feasible to extract soil moisture values from agricultural fields at a resolution of around 10 meters using Sentinel-1 data?
2) What are the key limitations and technical challenges in performing high-resolution soil moisture estimation in agricultural fields?

Methodology

To answer the above research questions, the development of a machine learning-based model using a random forest regressor algorithm is proposed. As discussed, the model will be based on Sentinel-1 SAR backscatter data for prediction, but also other relevant auxiliary data to control for key influences on the SAR backscatter. The auxiliary data used are water and vegetation indices calculated from Sentinel-2 optical satellite images and digital elevation model derived topographic features.

1.5 Expected contributions

The expected contribution is first of all a better understanding of the challenges and limitations in SAR-based soil moisture estimation at this scale. In addition, using machine learning and different relevant auxiliary data, the goal is to achieve results that show improved prediction values compared to a more simple predictor baseline.

2.Litterature Review

In this section, set the scene discussing the key challenges identified the literature when it comes to estimating high-resolution soil moisture. After setting the scene I review the main SAR-based soil moisture estimation models found in the literature: empirically based models, physically based models and machine learning based models.

2.1 Key challenges

In a review paper on the use of SAR data for soil moisture estimation by Kornelsen & Coulibaly (2013), they highlight that it is a well-established truth that that SAR radar waves, generally, give less backscatter the more water or moisture is present in the surface it scatters off of, such as soil in agricultural fields. However, as they also emphasize that moisture is only one of many physical properties of scattering surfaces that influences the radar backscatter, and that a key challenge remains: to isolate the effects of the water/moisture from especially the effects of surface roughness and vegetation. These two properties are also the main reason SAR-based soil moisture retrieval at a high-resolution is difficult (Ma et al., 2020). As I will discuss further below, in most research on trying to estimate soil moisture using SAR data, the focus is on disentangling these two properties from the backscatter.

Vegetation

A study recent by Bousbih et al. (2017) confirms that Sentinel-1 backscatter sensitivity to soil moisture decreases with increase in surface vegetation. In theory, the radar waves of Sentinel-1, with a wavelength of around 5.6 cm, has some ability to penetrate sparse vegetation. However, tree canopies and plants scatter the SAR signal if the size of the leaves or branches are equal or above the size of 5.6 centimiters and has an orientation parallel to the polarisation of incoming signal (Alemohammad et al., 2019). This means that density and size of vegetation matters. In addition to the scattering effects of the vegetation the SAR microwaves are also affected by water content in the vegetation. This also means that there is a threshold of how much vegetation can be present before soil moisture is practically impossible to estimate from SAR data.

Surface roughness

The surface roughness property creates unpredictable scattering as the angles on the surface that the radar waves bounce off of vary over space beyond what can be controlled for using terrain correction data processing methods, which relies on satellite derived digital elevation models (DEMs) with insufficient detail to control for smaller variations such as tramlines in agricultural fields. Studies have shown that this is a major limiting factor in estimation of soil moisture values (Sahebi et al., 2002; Schuler et al., 2002).

Other influences

Other influences are linked to topography, land cover, local incidence angle, as well as radiometric noise or speckle, inherent to SAR data (Massonnet & Souyris 2008). Some of these influences can be controlled for through image processing techniques such as terrain correction in the case of incidence angle and topography, and speckle filtering in the case of radiometric noise. Another issue is that the influence of vegetation and surface roughness, are also affected by the frequency, polarization, and incidence angle of the SAR satellites (Bousbih et al., 2017; Sahebi et al., 2002; Schuler et al., 2002).

2.2 Three approaches

In the literature, the main ways in which researchers have approached the problem of extracting soil moisture using SAR data can be divided into four groups: physically based models, empirically based models, change detection-based models, and machine learning based models (Barrett et al., 2009).

Physically based models

Examples of physical based models found in the literature are: Water Cloud Model, Kirchhoff Approximation model, and Integral Equation Model (Baghdadi et al., 2017; Gu et al., 2019; Paloscia et al., 2013). These models are all trying to establish the concrete sensor-specific relationship between to scattering surfaces at the ground and backscatter values. The advantage is that they are non-site specific and therefore can in theory be universally applicable, the challenge is that the lack of vegetation parameters makes it not usable in when vegetation is present.

Empirically backscatter models

In the case of empirically backscatter models, they investigate the interaction of microwaves with site-specific surface characteristics and estimate soil moisture based on this empirically observed interaction. These models rely on high-quality vegetation parameters and surface roughness that are measured in-situ. They can achieve high accuracy in predictions; however, as they are limited to the specific location being measured and are very labour intensive, they have a limited potential for scalability (Barrett et al., 2009).

Change Detection Approach

A third approach worth mentioning is soil moisture retrieval using change detection. This approach relies on the assumption that if surface roughness and vegetation are constant, the variations in over time in the SAR backscatter can be accredited to variations in moisture. Gao et al. (2017), for example, are fairly successful in extracting soil moisture at a 100-meter scale by leveraging time series of SAR images.

Machine learning based models.

Machine-learning based approaches have the advantage of taking in a variety of data and extracting linear or non-linear relationships between mainly satellite-based parameters and in-situ soil moisture reference data (Chakrabarti et. al., 2015). These models are also the most promising in terms of being able to estimate soil moisture values at a high-resolution and being less labour intensive than the empirical-based models. In a recent paper by Schönbrodt-Stitt et al. (2021), the authors developed a model based on Sentinel-1 data and a Random Forest (RF) machine learning algorithm to estimate soil moisture values of an agroforestry area in Central Italy. As this SAR model shows great promise in terms of accuracy of estimation (RMSE 0.028 m3, mean absolute error of 0.022 m−3, and R^2 0.86),

the current study will build on this model's SAR parameters. The model specifically uses a series of mathematical band combinations of multiple polarisations of the SAR data, Vertical and Vertical (VV), and Vertical Horizontal (VH), that papers such as Omar et al., (2017), and Ahmadian et al., (2019) have found to counteract radiometric instability and as well as vegetation moisture variations introduced by original non-synthetic band polarization.

Summary and presentation of this approach used in this study

Although a series of attempts have been made by researchers to extract soil moisture at a resolution close to or around 10 meters, as is the project of the current study, they are mostly site-specific and therefore not necessarily relevant or valid in the geographic context of this study. This study seeks to further build on what has been done, using insights from different studies on the topic of interest, to experiment with building an improved model for soil moisture estimation.

Thus, based on the promising results of studies using machine learning to entangle relationships between SAR backscatter and soil moisture this is also the proposed approach for the current study. In terms of SAR data preparation and feature engineering the methodology used in this study builds on the study by Schönbrodt-Stitt et al. (2021).

In terms of choice of machine learning algorithms this study will use, as in the Schönbrodt-Stitt et al. (2021) study, a random forest regressor algorithm. Other algorithms identified in the literature as having a high performance in soil moisture estimation such as Support Vector Machine (SVM) and Artificial Neural Networks (ANNs) (Ayehu et al. 2020; Ezzahar et al. 2020; Hajdu et al. 2018; Schonbrodt-Stitt et al. 2021) were also considered. However, due to time constraints only random forrest regressor was chosen.

Inspired by studies such as Xu et al. (2020) and Ayehu et al. (2020), which applies vegetation and water indices derived from Sentinel-2 data to control for vegetation, a set of optical based indices will also be used to improve the machine learning model's ability to control for vegetation and water content in the vegetation.

Furthermore, the geographical distribution of soil moisture over a given terrain will logically, simply due to gravity, to some extent follow topographic patterns such as water gathering in local depressions. Based on this rational a series of topographic parameters were developed based on a DEM, to capture these effects in the model. The parameters

chosen are slope (degrees), elevation (meters), Topographic Position Index (TPI) which is designed to represent topographic slope positions, i.e. ridge top, valley bottom, mid-slope, etc. (De Reu et al., 2013), and finally Terrain Ruggedness Index (TRI) is used to try to control for the local ruggedness of the terrain. In the case of the latter, as the DEM has a resolution of 30 meters, this will presumably not help control for the previously discussed small-scale surface roughness effects on the SAR backscatter, however, it might support the algorithm in separating other topographic effects on soil moisture or SAR backscatter.

Having discussed available research literature on the use of SAR data for soil moisture extraction, as well as presenting the approach this research seeks to take based on insights from previous studies, I will in the following chapter elaborate further on the methodology and data preparation process used.

3.Data and Methodology

The methodology is divided into two main steps, 1) the collection of in-situ soil moisture data, and 2) based on the in-situ data, train a machine learning model for estimation of soil moisture values. I will start elaborating the data collection step.

3.1 In-situ data collection

In order to train and validate the machine learning model, in-situ reference data for soil moisture is needed. The following subsections I elaborate on the process of collecting these reference data, the sites of collection and sampling strategy; and I present an exploratory analysis of the reference data.

About Terramay

The data was collected at a farm in the Alentejo region of Portugal by the name of Terramay (Terramay**,** n.d). Climatically, the area is semi-arid and suffers from periods of little rain and water access during summer. Soil moisture is especially important to ecosystem sustainability in the context of more arid regions of the world where water is a more scares resource (IPCC, 2019). Since 2018, the owners of Terramay have applied various targeted farming techniques in hopes of regenerating soil health and stopping desertification on the farm. This makes Terramay an interesting case for potentially analysing the effects of regenerative practises on soil moisture, on top of the core mission of collecting soil data for training the model.

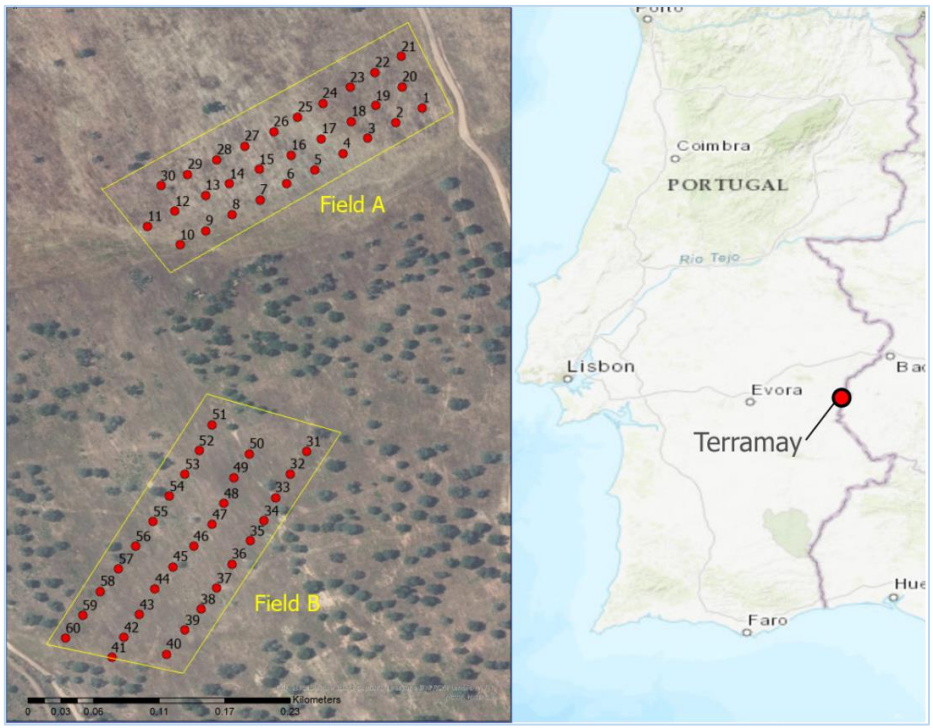

Figure 1. Overview map of data collection site (Terramay) and in reference data points.

Description of data collection sites

At Terramay two different fields or sites, were selected for the soil moisture sampling, based on having their own unique properties (see figure 1 and 2). The first one is a field without significant vegetation, mainly consisting of bare soil, thus having as few disturbances on the SAR backscatter signal as possible. The second is an agroforestry area which has more disturbing elements regarding the SAR backscatter such as some scattered vegetation and the occasional tree. Furthermore, tramlines have been constructed across the second field to preserve water in the soil and prevent runoff. Both sites have a degree of topographic variations across the fields in terms of soft slopes and slight differences in elevation.

Figure 2. Pictures taken by author of field A (left) and field B (right).

Sampling

In order for the in-situ data to be as relevant as possible sampling was coordinated with the overpass of the two Sentinel-1 satellites (A and B). The initial plan was to collect soil moisture data for each of the sites on five consecutive Sentinel-1 overpasses between 7th of November and 1st of December 2021, specifically, 7th, 13th, 19th, 25th of November and 1st of December (**table 1)**.

Collection dates	Nr. of points
07.11	35
13.11	60
19.11	60
25.11	0
01.12	60

Table 1. Overview of reference data collection. Colour scheme to highlight missed date Nov. 25th and less than planned number of points on Nov. 7th.

Each of the Sentinel-1 A and B satellites, at the time, had an overpass every 6 days alternating between one of them overpassing in the morning around 06:30 and the other in the evening around 18:30 on the same day (table 2).

Overpass times	Satellite	Relative orbit/ Track	Incidence angle
07.11.2021, 18:30	S1A	147	39.39
07.11.2021, 06:30	S1B	52	36.6

13.11.2021, 06:30	S1A	52	36.6
13.11.2021, 18:30	S1B	147	39.39
19.11.2021, 18:30	S1A	147	39.39
19.11.2021, 06:30	S1B	52	36.6
01.12.2021, 06:30	S1A	52	36.6
01.12.2021, 18:30	S1B	147	39.39

Table 2. Overview of Sentinel-1 scenes used for in study.

The reference data was collected across each of the days, between the two acquisition times. As sampling controls of the same extract points in different times of the day showed little to no variations across the day, effects of sun and heat exposure are considered to be negligible.

However, as there was it was raining on the 25th of November, data could not be collected on this day. Collecting data on a rainy day would ruin the reliability of the reference data as variations would be dependent on the time of the day they were sampled, and thus not be representative for the whole day and both overpasses.

Soil moisture measurements were taken with a portable time domain reflectometry (TDR) system (TRASE SYSTEM 1) (ICT International, n.d). The system measures volumetric moisture content (VMC) with a measuring range of 0 – 100 percent, and a precision of +/- 2 VMC percent or better. When I throughout this study refer to soil moisture, reference data, in-situ data, or target data; the unit is VMC, expressed in percentage of the soils volume that is water.

VMC can be expressed as:

$$\mathrm{VMC} = \frac{volume\ of\ water\ (cm3)}{volume\ of\ soil\ (cm3)} x\ 100$$

For each of the sites, samples were planned to be taken at 30 sample points (with some variation on the 7th of November). The samples were collected along three parallel transects of roughly 200-250 meters with roughly 20-30 meter spacing between the samples taken along the transect lines using a handheld GPS to estimate the location and distances. Each of the three lines were positioned roughly 20-30 meters apart. In order to avoid errors in single samples taken at each of the sampling points, the soil moisture was measured three times at each point along the line, with only a few centimetres apart (see figure 3).

Figure 3. Picture from field sampling taken by author.

As mentioned, as samples were taken across the day, the effects of heat and sunshine during could have had a small effect on the sampling accuracy. However, to control for this effect additional samples were taken at the same points at varying times of the day, showing no significant changes in the moisture content.

In table 3, the results of the sampling can be observed. The first day of sampling, November 7th was done following a short period of heavy rain which came after a long period of no rain. Thus, the soil was very dry before this initial rainfall in the beginning of November. The rest of November there was little rain, only sporadic short periods of light rain throughout the month. This weather pattern observable in terms of the mean of the first and second day being higher than the last two days where the soil had had more time to dry following the aforementioned heavy rain.

Date	**07.11**	13.**11**	19.**11**	**01.12**
Max	27.5	30.7	26.5	27.4
Min	14.5	10.3	10	9.2
Range	13	20.4	16.5	18.2
Stdev	3.1	4.3	3.2	3.8
Mean	19.7	17.1	15.3	15.68

Table 3. Sampled data statistics aggregated per day.

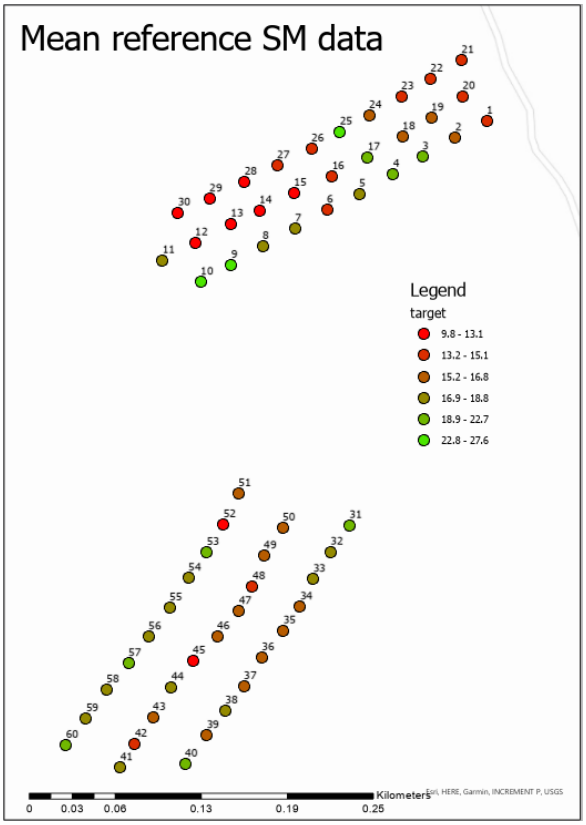

Figure 4. All soil moisture reference data (across all days aggregated to each point).

3.2 Preparation of model features/parameters for the machine learning model

In the following section I focus on how the different features or parameters used for prediction in the machine learning model are developed.

In a machine learning model, you need to establish what features or parameters you want to use as inputs to do training and prediction with. The choice of these is usually based on a rationale or hypothesis that they might have some explanatory power of the target variable of interest. Depending on what type of machine learning algorithm one is using, these features can be either categorical or continuous, meaning either referring to categories of for example landcover, or a continuous range of elevation values. As this study aims to perform a prediction of a value, meaning a regression type prediction, as opposed to classification, all the features will be continuous numerical values.

As discussed, SAR backscatter data has shown to be sensitive to soil moisture, however, to control for the key impacts of surface roughness and vegetation effects, as well as basic topographic effects on soil moisture distribution over terrain, a series of additional parameters are developed based on 10 meters resolution Sentinel-2 optical data, and

digital elevation model (DEM) The Shuttle Radar Topography Mission (SRTM) digital elevation model which has a cell resolution of 30 meters (NASA, n.d) was chosen.

In the following subsections I elaborate on how each of the features are designed. For the sake of clarity, they are organised into three groups: 1) SAR-based, 2) optical-based, and 3) topography-based parameters.

SAR-based parameters

As discussed, this study leans on the approach of Schönbrodt-Stitt et al. (2021). Following this, synthetic bands based on combinations of the two Sentinel-1 polarisations VV and VH, and the three SAR backscatter conventions Beta0, Gamma0 and Sigma0 are developed. The data used are a series of Interferometric Wide (IW) level-1 Ground Range Detected (GRD) (ESA n.d b) Sentinel-1 scenes from both satellites A and B (see table 2). All SAR processing was done with the ESA's SAR processing software, Sentinel Application Platform (SNAP), mainly using the SNAP Python API.

The three radar backscatter conventions or "products" are produced by three different radiometric calibration approaches.

Radiometric calibration is the conversion of digital numbers recorded by the SAR sensor into physical units, or the conversion of the raw registered intensity for each pixel, into pixel values that can be directly related to the radar backscatter of the image. The calibration happens using a specified function applied to the raw SAR data, and each of the three products are derived using different functions (ESA, n.d c). Beta0, is the most basic calibration of a SAR product, also referred to as the "radar brightness". By using an "internal calibration constant" derived from the metadata. Sigma0, considers both the internal calibration constant and the incidence angle (also, most of the time derived from metadata). Gamma0, takes into account not just the incidence angle of the radar wave, but also uses the "local incidence angle", meaning the angle at which the radar wave is hitting the Earth's surface, which can be calculated using a DEM (Massonnet & Souyris 2008).

Thus, these products were calculated from the raw SAR data in addition a fourth Terrain-Flattened Gamma0 (TF Gamma0) was calculated, using a series of processing steps (figure 5).

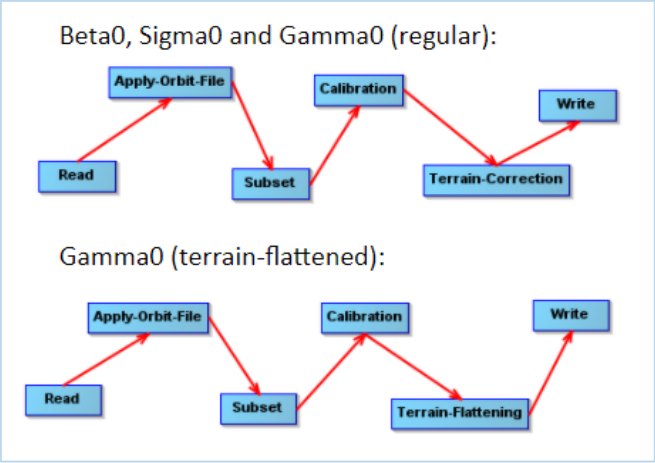

Figure 5. SNAP processing graphs for Beto0, Sigma0, Gamma0 (top) and TF Gamma0 (bottom)

The Beta0, Sigma0 and regular Gamma0, were produced using the following processing steps: 1) application of orbit file, 2) subsetting to the region of interest, 3) radiometric calibration (Beta0, Sigma0 and Gamma0), and 4) Doppler terrain correction using GLO-30 (Copernicus DEM with 30 m resolution).

While the terrain flattened gamma0 was produced using the following processing steps: 1) application of orbit file, 2) subsetting to the region of interest, 3) radiometric calibration (Gamma0), and 4) gamma terrain flattening using GLO-30 (Copernicus DEM with 30 m resolution); outputting the final Terrain Flattened Gamma0.

The orbit file is applied to attach the most precise available information about the satellites positioning during acquisition, while the Doppler terrain correction and the terrain flattening are methods for calibrating and controlling for local topographic effects on the SAR backscatter (Massonnet & Souyris 2008).

After the above SAR processing steps, a series of band maths operations were performed on the four products to create the final synthetic bands used in the machine learning model.

Synthetic bands were created for each of the acquisitions, two per day. As the two acquisitions on each day had different relative orbits/tracks (147 and 52 (see table 2), each were processed as separate stacks. The reason being that the different tracks have different incidence angles (39.4 and 26.6 degrees, respectively), and as the incidence angles had not been normalised, keeping them separately was necessary to achieve optimal radiometric and geometric correction results.

Thus, for each of the tracks per day, synthetic bands were calculated for Beta0, Sigma0, Gamma0 and TF Gamma0 using combinations of the two polarisations, in terms of addition (VH+VV), subtracting (VH-VV, VV-VH), division (VH/VV, VV/VH), and multiplication (VH∗VV) with the other (table 4)

Parameter name	Band maths operations	Parameter name	Band maths operations
Beta0_VH_min_VV	Beta0 (VH) - Beta0 (VV)	**Sigma0_VH_min_VV**	Sigma0 (VH) - Sigma0 (VV)
Beta0_VV_min_VH	Beta0 (VV) - Beta0 (VH)	**Sigma0_VV_min_VH**	Sigma0 (VV) - Sigma0 (VH)
Beta0_VH_plus_VV	Beta0 (VH) + Beta0 (VV)	**Sigma0_VH_plus_VV**	Sigma0 (VH) + Sigma0 (VV)
Beta0_VH_div_VV	Beta0 (VH)/ Beta0 (VV)	**Sigma0_VH_div_VV**	Sigma0 (VH)/ Sigma0 (VV)
Beta0_VV_div_VH	Beta0 (VV)/ Beta0 (VH)	**Sigma0_VV_div_VH**	Sigma0 (VV)/ Sigma0 (VH)
Beta0_VH_multi_VV	Beta0 (VH) * Beta0 (VV)	**Sigma0_VH_multi_VV**	Sigma0 (VH) * Sigma0 (VV)
Gamma0_VH_min_VV	Gamma0 (VH) - Gamma0 (VV)	**TFGamma0_VH_min_VV**	TFGamma0 (VH)-TFGamma0 (VV)
Gamma0_VV_min_VH	Gamma0 (VV) - Gamma0 (VH)	**TFGamma0_VV_min_VH**	TFGamma0 (VV)-TFGamma0 (VH)
Gamma0_VH_plus_VV	Gamma0 (VH) + Gamma0 (VV)	**TFGamma0_VH_plus_VV**	TFGamma0 (VH)+TFGamma0 (VV)
Gamma0_VH_div_VV	Gamma0 (VH)/ Gamma0 (VV)	**TFGamma0_VH_div_VV**	TFGamma0 (VH)/TFGamma0 (VV)
Gamma0_VV_div_VH	Gamma0 (VV)/ Gamma0 (VH)	**TFGamma0_VV_div_VH**	TFGamma0 (VV)/TFGamma0 (VH)
Gamma0_VH_multi_VV	Gamma0 (VH) * Gamma0 (VV)	**TFGamma0_VH_multi_VV**	TFGamma0 (VH)*TFGamma0 (VV)

Table 4. Overview of band math operations for producing synthetic bands.

This process resulted in 48 synthetic bands per day, 24 for each track. After having computed the synthetic bands, values were extracted at the geographic coordinates of the in-situ soil moisture sampling. Values were extracted using a mean of a 3x3 window centred on each of the collection points. The result being a total of 48 SAR based features.

Optical based parameters

Three optical based indices are used in order to control for effects of vegetation on the backscatter.

In order to control for vegetation water content (VWC), a normalized difference water index (NDWI) is used, as well as the similar index normalized difference infrared index (NDII) is calculated. Furthermore, normalised difference vegetation index (NDVI), is used to further identify green vegetation in the fields.

Thus, using Quantum GIS (QGIS), the following three vegetation indices were calculated using combinations of Sentinel-2 band data:

1. NDVI (Normalized Difference Vegetation Index), 10 meters resolution.
2. NDWI (Normalized Difference Water Index), 20 meters resolution.

3. NDII (Normalized Difference Infrared Index), 20 meters resolution.

NDVI was calculated with:

$$NDWI = \frac{B8 - B4}{B8 + B4}$$

NDWI was calculated with:

$$NDWI = \frac{B8 - B12}{B8 + B12}$$

NDII was calculated with:

$$NDII = \frac{B8 - B11}{B8 + B11}$$

Sentinel-2 data was gathered from dates as close as possible to the Sentinel-1 acquisition and reference data sampling dates (see table 5). For each of the three surfaces standard deviation, mean and median values were extracted using a 3x3 window, centred at the points where reference data was collected, resulting in a total of nine features used in the machine learning model.

Date of acquisition
07.11.2021
12.11.2021
17.11.2021
02.12.2021

Table 5. Overview of Sentinel-2 data acquisitions.

DEM based parameters

As discussed, to simulate topographical effects on soil moisture distribution in space, a series of DEM-based features were developed. With QGIS, using the SRTM 30m DEM (NASA, n.d) five features were calculated and extracted:

1. aspect
2. slope
3. elevation
4. Topographic Positioning Index (TPI)
5. Terrain Ruggedness Index (TRI)

Based on these five DEM based surfaces, values were extracted only at the points of the in-situ data collection, as opposed to the two previously described feature categories where a 3x3 window was used. Thus, I used a total of five DEM based machine learning features.

3.3 Machine learning

In the section I explain the implementation and design of the random forest regression model including hyperparameters and the training, testing and evaluation process.

As mentioned, the random forest regression implementation in the Scikit-learn Python package was used to build the machine learning model. According to Scikit-learn, "a random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting" (Scikit-learn n.d, para. 2). The advantage of the Random Forrest algorithm is that it is easy to use and performs well in most use cases. The key hyperparameters of the algorithm are 1) the "criterion" for predictive accuracy, 2) "maximum depth" of the tree in terms of vertical nodes, 3) the "minimum samples split", meaning the minimum number of samples required to split an internal node, and 4) the "number of trees" in the forest.

My model was build using the following fixed parameters: "criterion" = squared error, "maximum depth" = None, "minimum samples split" = 2, and "number of trees" = 5000.

Training and testing the machine learning model

The Random Forest model was trained and tested using a K-fold cross-validation process. As the data covers four days, I chose to perform a 4-fold cross-validation splitting the data into one fold for each day. Thus, four different permutations were made with three days being used for training and one day used for testing in each permutation (table 6).

Permutation nr.	Training folds (dates)	Testing folds (dates)
1	13.11, 19.11, 1.12	7.11
2	7.11, 19.11, 1.12	13.11
3	7.11, 13.11, 1.12	19.11
4	7.11, 13.11, 19.11	1.12

Table 6. Permutations and fold splits.

Each of the folds/days had 60 records, except from the first day, 7th of November that had only 35 records for training and testing (table 7), add

Fold/days	Number of records
7.11	35
13.11	60
19.11	60
1.12	60
Total	**215**

Table 7. Number of records per. fold.

Evaluation of machine learning model

To check if the stability and the robustness of the model, this process was done seven times using the standard deviation within each run and between all the runs as an evaluation metric.

In terms of the success or accuracy of the model, the model is evaluated using root mean squared error (RMSE), r-squared (R^2), and mean absolute percent error (MAPE). The choice of evaluation metrics is based on what is commonly used in similar studies (for example, Ma et al., 2020; Schönbrodt-Stitt et al., 2021). The RMSE metric gives an idea of the level of error involved in predictions but weighting individual large prediction errors more heavily than typical absolute errors metrics. While the MAPE just gives an idea of the overall accuracy without penalising outlier predictions like the RMSE. The R^2 metric gives an idea of how the prediction and target values are correlated with each other. In the context of this study, R^2 is to what extent the variance of the target explains the variance of the predictions.

Having explained the methodology of the study, I will in the following chapter present results and analysis, as well as evaluate the random forest regressor model's performance and robustness.

4.Analysis and Results

In this chapter, I present and evaluate the results of the Random Forest model that was trained using all the 62 features discussed in the previous chapter. Furthermore, I will

present an analysis of the errors as well discuss the potential causes of errors and potential for improvement of the model.

4.1 Summary of results

After a doing the 4-fold cross-validation process seven times, I got the following results (see table 8): an average root mean squared error (RMSE) of 3.51, meaning the average of all the seven runs' RMSE averages, with a standard deviation (all 28 permutations) of 0.07; and an average R^2 of 0.17, meaning the average of all the seven runs' R^2 averages, with a standard deviation of 0.22. Furthermore, the average MAPE (of the seven runs) was 83.34%, with a standard deviation of 2.16%.

To contextualise the results, a baseline RMSE was also calculated for each of the 28 permutations using the mean of all the training data target values of each respective permutation. Comparing with this baseline, the model has an average of 16.1 % relative improvement from the baseline RMSE, with an average standard deviation of 5.3 % between all 28 permutations. This metric is useful as it accounts for the diversity of data ranges, and data variation of target values in each fold (see table 8).

Run nr.	r-squared **(R^2)**	**MAPE (%)**	**RMSE predictions**	**RMSE baseline**	**Relative improvement from RMSE vs. baseline (%)**	**Folds**
1	0.079	86.33	3.580	4.789	25.25	07.11
1	0.201	84.26	3.816	4.302	11.30	13.11
1	0.146	81.92	3.223	3.709	13.09	19.11
1	0.255	80.60	3.439	4.015	14.33	01.12
mean of 1:	0.17	83.28	3.514	4.204	15.99	
stdev of 1:	0.075	2.53	0.248	0.459	6.293	
2	0.072	86.33	3.593	4.789	24.98	07.11
2	0.195	84.26	3.830	4.302	10.97	13.11
2	0.145	81.91	3.227	3.709	12.99	19.11
2	0.249	80.62	3.442	4.015	14.26	01.12
mean of 2:	0.165	83.28	3.523	4.204	15.80	
stdev of 2:	0.065	2.188	0.219	0.397	5.42	
3	0.071	86.29	3.609	4.789	24.62	07.11
3	0.199	84.27	3.821	4.302	11.18	13.11
3	0.154	82.07	3.202	3.709	13.66	19.11
3	0.257	80.71	3.423	4.015	14.73	01.12
mean of 3:	0.170	83.34	3.514	4.204	16.05	
stdev of 3:	0.068	2.124	0.228	0.397	5.11	
4	0.082	86.36	3.572	4.789	25.40	07.11
4	0.198	84.18	3.822	4.302	11.16	13.11
4	0.148	81.99	3.217	3.709	13.25	19.11
4	0.249	80.66	3.433	4.015	14.48	01.12
mean of 4:	0.169	83.30	3.518	4.204	15.94	
stdev of 4:	0.062	2.17	0.219	0.397	5.51	

5	0.07	86.31	3.603	4.789	24.76	07.11
5	0.199	84.22	3.821	4.302	11.18	13.11
5	0.154	82.05	3.208	3.709	13.51	19.11
5	0.252	80.61	3.440	4.015	14.30	01.12
mean of 5:	0.171	83.34	3.51	4.204	16.14	
stdev of 5:	0.067	2.163	0.224	0.397	5.22	
6	0.077	86.3	3.585	4.789	25.14	07.11
6	0.199	84.32	3.821	4.302	11.18	13.11
6	0.154	82.15	3.195	3.709	13.86	19.11
6	0.254	80.60	3.438	4.015	14.37	01.12
mean of 6:	0.174	83.41	3.499	4.204	16.41	
stdev of 6:	0.065	2.156	0.227	0.397	5.33	
7	0.076	86.33	3.590	4.789	25.03	7.11
7	0.204	84.34	3.809	4.302	11.45	13.11
7	0.157	82.11	3.192	3.709	13.95	19.11
7	0.259	80.85	3.405	4.015	15.19	01.12
mean of 7:	0.174	83.41	3.499	4.204	16.41	
stdev of 7:	0.067	2.09	0.228	0.397	5.15	
mean of all runs:	**0.171**	**83.34**	**3.511**	**4.204**	**16.10**	
stdev of all 28 records:	**0.065**	**2.15**	**0.223**	**0.397**	**5.321**	

Table 8. Overview of all results from each run of the seven runs and 28 permutations.

In terms of individual folds (see table *9*), the November 7th fold, meaning the permutation using data from this day for testing, is an outlier in terms of having an 25% RMSE improvement from the baseline (vs. 11.2%, 13.5% and 14.5%, respectively) and the highest MAPE of all the folds (86.32%). However, interestingly, it has the lowest R^2 (0.075).

Folds	r-squared (R^2)	MAPE (%)	RMSE predictions	RMSE baseline	Improvement from RMSE baseline (%)
07.11	0.075	86.324	3.591	4.71	25.03
13.11	0.199	84.269	3.82	4.303	11.211
19.11	0.151	82.033	3.21	3.71	13.479
01.12	0.254	80.672	3.432	4.016	14.527

Table 9. Evaluation metric averages of each fold across all seven runs.

Another interesting case is the December 1st fold that has a significantly higher R^2 than the other folds (0.254), meaning that the prediction variable explains 25.4 % of the variance of the target variable, while having the worst MAPE (80.7 %).

Having presented the key aggregated results of the model training and testing process, I will in the following section explore error distributions and variations between folds in more detail.

4.2 Exploration of results and errors

In this section I first present first a statistical exploration of errors focusing on over and under prediction of low and high target values respectively; second, I analyse and discuss the errors from a geographical perspective; and third, I discuss the cumulative distribution of errors using empirical cumulative distribution functions.

Statistical exploration of errors

In figure 6 one can see the distribution of the all the 215 target values and all the average predicted values by the model per point. This shows that the predictions are clustering around the middle values of the target range, and thus, consistently underpredicting the highest target values and overpredicting lowest target values.

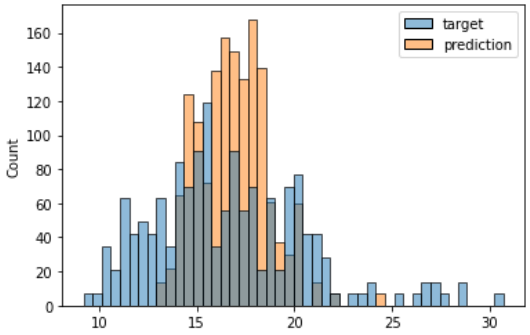

Figure 6. Histogram of targets vs. predictions (% VMC).

Figure 7 provides a clear perspective on the aforementioned dynamic of under and over prediction. As one can clearly observe that the model has some degree of sensitivity of the high peaks and low points as the prediction values follow the general pattern of the target values, although not merely to the same extent when observing for example the highest peak (around point ID 9).

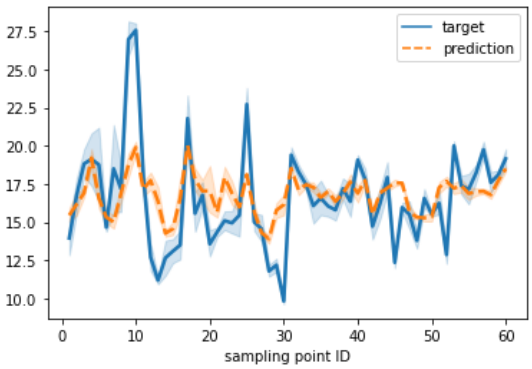

Figure 7. Line graph of target and prediction values per sampling ID point.

Furthermore, if you look at figure 8, where the absolute errors are plotted against the target values, there is a clear pattern between higher and lower errors and higher and lower target values.

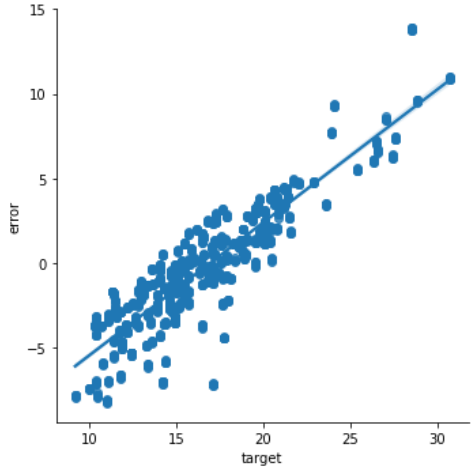

Figure 8. Scatterplot of absolute errors and target values with line of best fit.

Geographical exploration of results

In figure *9,* the geographical distribution of the observed soil moisture data and the predicted soil moisture data is clearly showing a similar pattern. Thus, the model seems to be able to, with its predictions, to capture the geographical patterns of the fields fairly well. In terms of the geographical distribution of errors, they are, as discussed, clustered around the areas of the fields with the highest and the lowest observed soil moisture values.

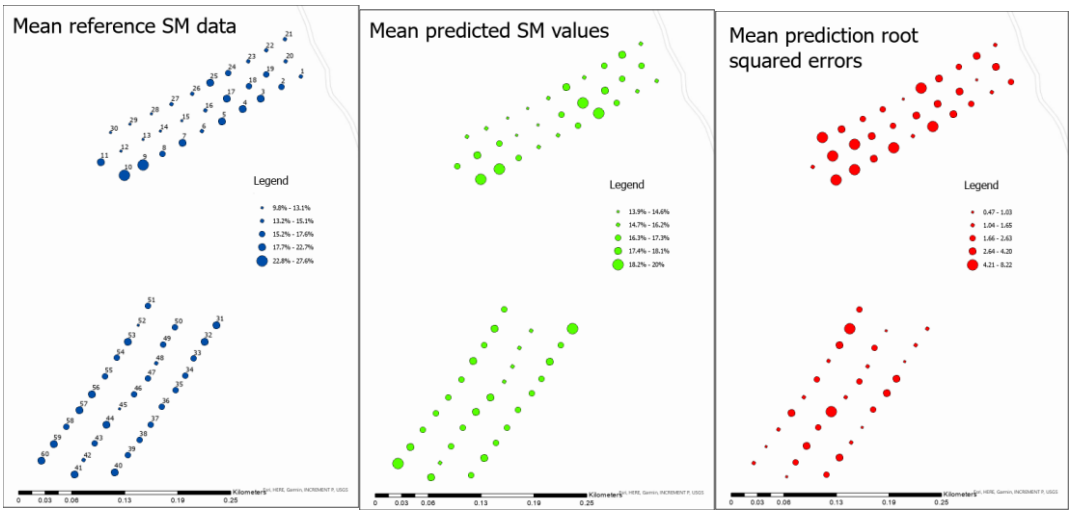

***Figure 9. Geographic distribution of reference soil moisture data(left), predictions (middle) and errors(right)*. Labels on left figure are sampling point IDs.**

Cumulative distribution of errors

The cumulative distribution of errors, that can be observed in figure 10, clearly show that the model overall has less errors than the calculated baseline errors. The function shows that around 80 % of the prediction errors are below 4, 40 % are under 2, and only 20 % are under 1. Furthermore, around 85 % of the prediction errors are significantly lower than the baseline errors, while the top 15 %, the prediction and baseline are more similar. This could potentially mean that if the records with the top 20% errors in predictions could be identified and improved upon or excluded, the model could see significant improvements.

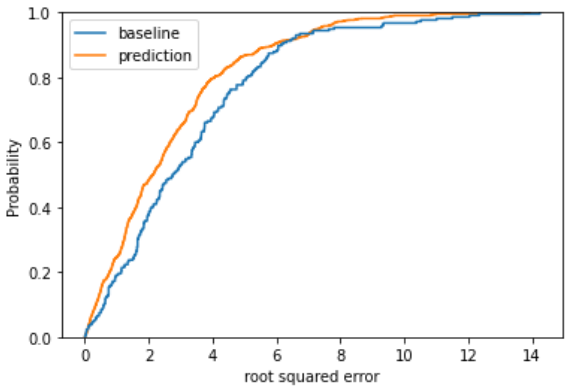

Figure 10. Empirical cumulative distribution function of all baseline and prediction root squared errors.

In figure 11 the cumulative error distribution of each of the folds (data from all runs) can be observed. Interestingly one fold, November 13th, sticks out in the top 10% of the distribution.

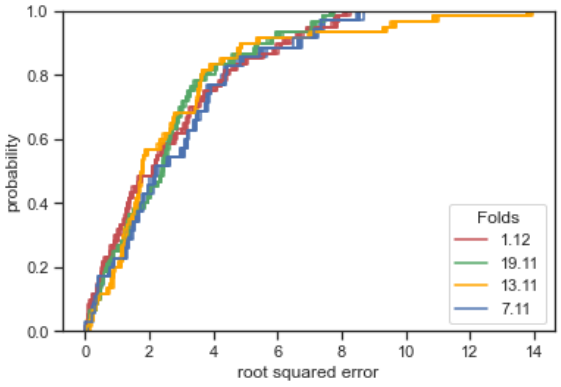

Figure 11. Empirical cumulative distribution function for the four folds' prediction errors (all 28 permutations).

When plotting the cummulative error distribution of the two fields seperatelly (see figure 12 and 13), it is clear that the outlier pattern of the November 13th fold stems from the field A predictions where it has lower errors than average initially, until around the top 10% where it has very high errors of 8-14.

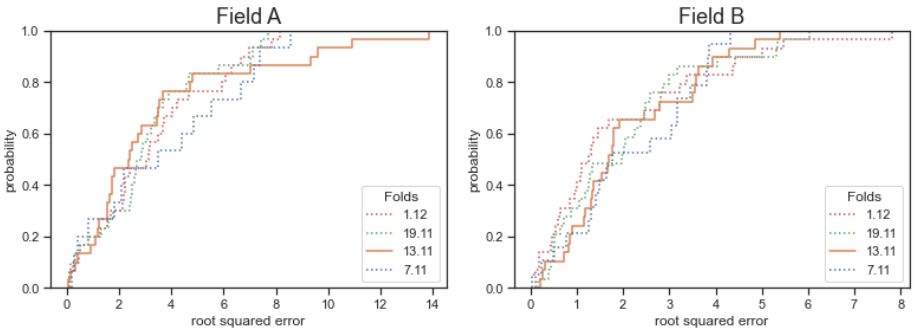

Figure 12 and 13. Empirical cumulative distribution function for the four folds' prediction errors, by field A (left) and field B (right) (all 28 permutations) (November 13th highlighted).

In table 10, we can see that the standard deviation of the reference soil moisture data is higher than average on the November 13th, and especially at field a where it has a significantly higher variation than the rest.

Dates	07.11	13.11	19.11	01.12
Stdev field A	3.16	5.39	3.82	4.24
Stdev field B	2.59	2.71	2.3	2.72
Stdev all data	3.1	4.28	3.24	2.82

Table 10. Overview of standard deviations of collected in-situ soil moisture data (November 13th highlighted).

As can be observed in figure 14, the distribution of the November 13th fold in the first 10 sampling points has a more seemingly more random extreme variation than the other folds.

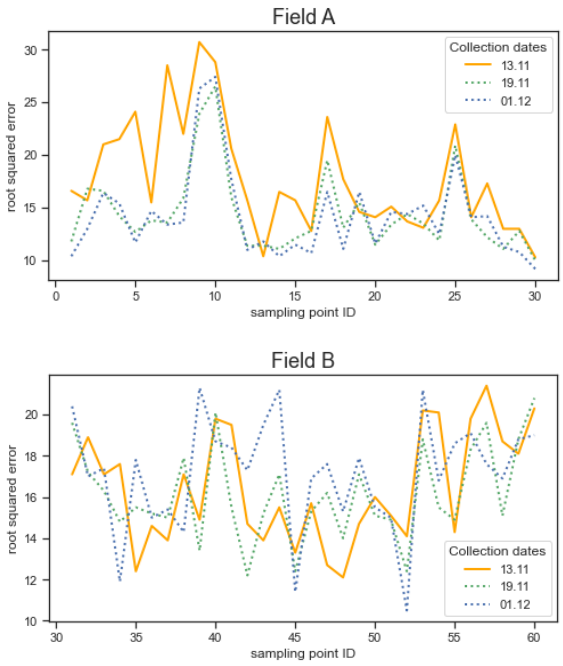

Figure 14. Linegraphs of collected soil moisture variations highlighting November 13th. November 7th excluded due to lack of matching temporal coverage.

However, this can also be a due to inaccuracies in the reference data introduced in the sampling process. The outlier pattern of the November 13th fold is something to explore further as perhaps the presense of some extreme highs and lows in this particular fold could be the reason. Altough further exploration and experimentation is needed to make definate conclusions.

5.Discussion and conclusion

In this feasibility study I have presented a remotes sensing-based site-specific machine learning model to explore the potential of open access Sentinel-1 SAR data to estimate and monitor soil moisture on a high-resolution (down to 10 m). Overall, the model performs significantly better than the calculated baseline and manages to recreate a similar geographic distribution pattern to the observed soil moisture with the model's predicted soil moisture. The model does perhaps not display a high enough accuracy to estimate single day soil moisture values at this scale, however, for the purposes of monitoring very rough changes over longer periods of time it might be sufficient.

In terms of contributions to the research literature, the study has shown that there is a potential for monitoring soil moisture at this resolution, with a significant error involved.

This is just one of many steps needed to achieve reliable soil moisture estimates at a very high resolution. Thus, a key contribution achieved is the carving out some insights and better understanding of what combinations of data might work when building models, and the challenges and limitations involved in using SAR in combination with other data sources for monitoring soil moisture.

Compared to other studies

If compared to other similar studies trying to estimate soil moisture based on SAR satellite data, the results from this model seem equivalent or perhaps slightly worse. For example, Schönbrodt-Stitt et al. (2021), achieved an R^2 of 0.025, an RMSE of 2 (% VMC) and an MAPE of 89%. Other studies using in-situ surface roughness inputs to calibrate the model achieves similar results, for example Ma et al. (2020) with R^2 of 0.472 to 0.665 and RMSE from 7.8 to 3.9 (% VMC). Other approaches such as Goa et al. (2017), using a timeseries of SAR images to extract a minimum and maximum wetness/soil moisture across an area based on SAR backscatter over a larger area (assuming a direct relationship on a larger scale), to estimate soil moisture at 100-meter resolution, achieved an approximate RMSE of 5.9. Although for example the RMSE result of this study is similar (3.51), these studies are not necessarily directly comparable due to a lack of baselines to contextualise the error results especially, as the ranges and distributions of soil moisture values, or target values for the models, are not the same. Therefore, evaluating the true performance of this study's model through direct comparison with these other studies, has clear limitations.

Applicability to practitioners

Also, for this to be practically useful for a farmer both the fact that the model is not the most accurate and that it is bad at capturing extreme low and high values, will likely mean that the model is in practice not robust enough for the intended application.

With this major caveat, the approach does show some degree of promise in terms of initial results. However, to further test the results one would need to collect substantially more reference data over more time and with more geographical diversity. One major challenge with the explored approach is that collecting the soil moisture reference samples requires a lot of time and planning, as well as expensive specialised measurement equipment. Unless done using research funding, for most intended beneficiaries this would be a great barrier to further develop satellite-based soil moisture retrieval models.

Potential points of improvement

The key pattern discussed is that when it comes to the highest and the lowest soil moisture values the model is not working particularly well. This might be due to the simple fact that there are fewer records of very high and low values than the ones closer to average that provides less training data to the model containing these particular characteristics. The most extreme values might also be concentrated in one or two folds which could cause some of the permutations to be ran more or less void of these values in the training data. This could potentially speak against the approach of using each day as a separate fold as opposed to for example randomly splitting the folds each run, that might better distribute the extreme values between folds.

Generally, a key issue with the use of SAR-based data at this scale is radiometric speckle effects, that must be offset by using smoothing filters. Unfortunately, the speckle still constitutes a major disturbance factor with or without this filtering, and it is clear that in this study high and low speckles in the SAR scenes used in the model greatly disturb the reliability of this data. Furthermore, as the speckle seemingly is random noise, and does not represent actual physical properties of surface objects, thus representing relevant information, the model is "poisoned" by these non-relevant variations which ultimately gives worst prediction results. This problem could however potentially be overcome by having more data in terms of Sentinel-1 scenes and in-situ measurement points. Presumably, with more data available the model would be able to disregard these random instabilities in the SAR data.

Future research

In terms of recommendations for future research, other machine learning algorithms should be tested with the same or similar datasets. Especially deep learning algorithms, perhaps with components of logistical regression to better capture non-linear relationships in the data. This is especially relevant in the context highlighted problem of higher prediction errors connected to extreme high and low soil moisture values. As discussed, although the model does not predict the extremes well, predictions are sensitive to the extremes in terms of predicting higher than average and lower than average values at these points, just not the extent that it matches the observed values. Taking this into consideration, the model might just have a scaling problem when it comes to the challenge of extreme values. Apart from focusing on improving the machine learning approach, most

of the time in machine learning improved or more data is often a more efficient way of improving prediction results. Thus, collecting more and more diverse data for the model to train with could be another possibility for future studies to focus on. In summary, the model developed in this study, although falling short of being useful for the intended application, has shown some promise that merits further exploration.

Bibliographic References

Ahmadian, N., Ullmann, T., Verrelst, J., Borg, E., Zölitz, R., & Conrad, C. (2019). Biomass Assessment of Agricultural Crops Using Multi-temporal Dual-Polarimetric TerraSAR-X Data. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 87(4), 159–175. https://doi.org/10.1007/s41064-019-00076-x

Alemohammad, S. H., Jagdhuber, T., Moghaddam, M., & Entekhabi, D. (2019). Soil and vegetation scattering contributions in L-band and P-band polarimetric SAR observations. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 8417-8429.

Ayehu, G., Tadesse, T., Gessesse, B., Yigrem, Y., & M. Melesse, A. (2020). Combined Use of Sentinel-1 SAR and Landsat Sensors Products for Residual Soil Moisture Retrieval over Agricultural Fields in the Upper Blue Nile Basin, Ethiopia. Sensors, 20(11), 3282. https://doi.org/10.3390/s20113282

Baghdadi, N., El Hajj, M., Zribi, M., & Bousbih, S. (2017). Calibration of the water cloud model at C-band for winter crop fields and grasslands. Remote Sensing, 9(9), 969.

Barrett, B. W., Dwyer, E., & Whelan, P. (2009). Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. Remote Sensing, 1(3), 210-242.

Bousbih, S., Zribi, M., Lili-Chabaane, Z., Baghdadi, N., El Hajj, M., Gao, Q., & Mougenot, B. (2017). Potential of Sentinel-1 radar data for the assessment of soil and cereal cover parameters. Sensors, 17(11), 2617.

Chakrabarti, S., Bongiovanni, T., Judge, J., Nagarajan, K., & Principe, J. C. (2014). Downscaling satellite-based soil moisture in heterogeneous regions using high-resolution remote sensing products and information theory: A synthetic study. IEEE Transactions on Geoscience and Remote Sensing, 53(1), 85-101.

Climate Farmers. (No date a). 'About us' [website]. Retrieved 31 January 2022 (https://www.climatefarmers.org/about-us/).

Climate Farmers. (No date b). 'DEFINITION OF REGENERATIVE AGRICULTURE: An approach to defining regenerative agriculture based on outcomes' [website]. Retrieved 18 January 2022 (https://www.climatefarmers.org/definition-of-regenerative-agriculture/).

Deng, L., Yan, W., Zhang, Y., & Shangguan, Z. (2016). Severe depletion of soil moisture following land-use changes for ecological restoration: evidence from northern China. Forest Ecology and Management, 366, 1-10.

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., ... & Crombé, P. (2013). Application of the topographic position index to heterogeneous landscapes. Geomorphology, 186, 39-49.

ESA (The European Space Agency). (No date a). 'Sentinel-1' [website]. Retrieved 19 January 2022 (https://sentinel.esa.int/web/sentinel/missions/sentinel-1).

ESA (The European Space Agency). (No date b). 'Level-1 GRD Products' [website]. Retrieved 31 January 2022 (https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-algorithms/ground-range-detected#:~:text=Level%2D1%20Ground%20Range%20Detected,using%20an%20Earth%20ellipsoid%20model.&text=Ground%20range%20coordinates%20are%20the,Pixel%20values%20represent%20detected%20magnitude).

ESA (The European Space Agency). (No date c). 'Level-1 Radiometric Calibration [website]. Retrieved 31 January 2022 (https://sentinels.copernicus.eu/web/sentinel/radiometric-calibration-of-level-1-products).

Ezzahar, J., Ouaadi, N., Zribi, M., Elfarkh, J., Aouade, G., Khabba, S., ... & Jarlan, L. (2020). Evaluation of backscattering models and support vector machine for the retrieval of bare soil moisture from Sentinel-1 data. Remote Sensing, 12(1), 72.

GCOS (Global Climate Observing System). (No date). 'Essential Climate Variables' [website]. Retrieved 1 February 2022 (https://gcos.wmo.int/en/essential-climate-variables).

Giller, K. E., Hijbeek, R., Andersson, J. A., & Sumberg, J. (2021). Regenerative Agriculture: An agronomic perspective. Outlook on Agriculture, 50(1), 13-25.

Gao, Q., Zribi, M., Escorihuela, M. J., & Baghdadi, N. (2017). Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. Sensors, 17(9), 1966.

Gu, W., Xu, H., & Tsang, L. (2019). A numerical kirchhoff simulator for GNSS-R land applications. Progress in electromagnetics research, 164, 119-133.

Hajdu, I., Yule, I., & Dehghan-Shear, M. H. (2018, July). Modelling of near-surface soil moisture using machine learning and multi-temporal sentinel 1 images in New Zealand. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1422-1425). IEEE.

Hes, D., & Rose, N. (2019). Shifting from farming to tending the earth: A discussion paper. Journal of Organics, 6(1), 3-21.

ICT International (No date). 'TDR TRASE System 1' [website]. Retrieved 1 November 2021 https://www.soilmoisture.com/TRASE-1/

Kornelsen, K. C., & Coulibaly, P. (2013). Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. Journal of Hydrology, 476, 460-489.

Liu, C. A., Chen, Z. X., Yun, Shoa., Chien, J. S., Hasi, T., & Pan, Hai-zhu. (2019). Research advances of SAR remote sensing for agriculture applications: A review. Journal of integrative agriculture, 18(3), 506-525.

IPCC (Intergovernmental Panel on Climate Change). (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.

Ma, C., Li, X., & McCabe, M. F. (2020). Retrieval of High-Resolution Soil Moisture through Combination of Sentinel-1 and Sentinel-2 Data. Remote Sensing, 12(14), 2303. https://doi.org/10.3390/rs12142303

Massonnet, D., & Souyris, J. C. (2008). Imaging with synthetic aperture radar. EPFL press. https://doi.org/10.1201/9781439808139

NASA (National Aeronautics and Space Administration). (No date). 'The Shuttle Radar Topography Mission' [website]. Retrieved 19 January 2022 (https://www2.jpl.nasa.gov/srtm/).

Omar, H., Misman, M. A., & Kassim, A. R. (2017). Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia. Applied Sciences, 7(7), 675. https://doi.org/10.3390/app7070675

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., & Reppucci, A. J. R. S. O. E. (2013). Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. Remote Sensing of Environment, 134, 234-248.

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., & Loew, A. (2021). A roadmap for high-resolution satellite soil moisture applications–confronting product characteristics with user requirements. Remote Sensing of Environment, 252, 112162.

Phillips, C. L., & Nickerson, N. (2015). Soil Respiration. In Reference Module in Earth Systems and Environmental Sciences. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09442-2

Sahebi, M. R., Angles, J., & Bonn, F. (2002). A comparison of multi-polarization and multi-angular approaches for estimating bare soil surface roughness from spaceborne radar data. Canadian Journal of Remote Sensing, 28(5), 641-652.

Schuler, D. L., Lee, J. S., Kasilingam, D., & Nesti, G. (2002). Surface roughness and slope measurements using polarimetric SAR data. IEEE Transactions on Geoscience and Remote Sensing, 40(3), 687-698.

Schönbrodt-Stitt, S., Ahmadian, N., Kurtenbach, M., Conrad, C., Romano, N., Bogena, H. R., Vereecken, H., & Nasta, P. (2021). Statistical Exploration of SENTINEL-1 Data, Terrain Parameters, and in-situ Data for Estimating the Near-Surface Soil Moisture in a Mediterranean Agroecosystem. Frontiers in Water, 3, 655837. https://doi.org/10.3389/frwa.2021.655837

Scikit-learn (No date). 'sklearn.ensemble.RandomForestRegressor' [website]. Retrieved 19 January 2022 (https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html)

Terramay (No date). 'Our Story' [website]. Retrieved 14 January 2022 (https://www.terramay.com/our-story).

Xu, C., Qu, J. J., Hao, X., & Wu, D. (2020). Monitoring Surface Soil Moisture Content over the Vegetated Area by Integrating Optical and SAR Satellite Observations in the Permafrost

Region of Tibetan Plateau. Remote Sensing, 12(1), 183.
https://doi.org/10.3390/rs12010183