Masters Program in Geospatial Technologies

SOIL MOISTURE ESTIMATION OF EUCALYPTUS FORESTS IN PORTUGAL WITH L-BAND SAR USING POLARIMETRIC DECOMPOSITIONS AND MACHINE LEARNING

Daniel Erickson

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

SOIL MOISTURE ESTIMATION OF EUCALYPTUS FORESTS IN PORTUGAL WITH *L*-BAND SAR USING POLARIMETRIC DECOMPOSITIONS AND MACHINE LEARNING

Dissertation supervised by:

Joel Dinis Baptista Ferreira da Silva

PhD, Universidade Nova de Lisboa, IMS

Hugo Alexandre Gomes da Costa

PhD, Universidade Nova de Lisboa, IMS

Carlos Granell Canut

PhD, Universitat Jaume I, INIT

February 2023

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.

Daniel Erickson

24.02.2023

Lisbon

The signed original has been archived by NOVA IMS.

ACKNOWLEDGMENTS

To my parents, who have supported and guided me to get the best education possible,

to my partner, who is my daily emotional support,

to my advisors and professors, who helped me to reach this great achievement,

to the other students of the GeoTech Master program, who have been true friends,

to my friends and other family members, who have encouraged me,

to Dr. Marco Painho, who helped innumerable times throughout the program,

and to 8 year-old me, who only dreamt of reaching this.

SOIL MOISTURE ESTIMATION OF EUCALYPTUS FORESTS IN PORTUGAL WITH *L*-BAND SAR USING POLARIMETRIC DECOMPOSITIONS AND MACHINE LEARNING

ABSTRACT

Soil moisture is a critical ecological parameter because it is a primary input for all processes that involve the complex interaction between land surface and the atmosphere. Remote sensing, especially using microwaves, has shown great promise in measuring soil moisture- with several operating satellites focused on its continuous estimation and monitoring on a global scale. Portugal is predominantly characterized by Mediterranean and semi-arid climates that feature low and sporadic precipitation. Over 10% of Portugal's land area has been planted with *Eucalyptus globulus*- a non-native, fast-growing tree primarily planted for industrial use. Some studies have demonstrated that eucalyptus plantations adversely affect water availability, but overall results have been inconclusive as there are numerous other confounding variables. The goals of this study were to determine, using fully polarimetric *L*-band SAR and machine learning, if soil moisture could be accurately predicted in eucalyptus forests, and if there is a significant difference in soil moisture inside eucalyptus forests relative to other forests. Vegetated surfaces complicate the estimation of soil moisture because their structure and water content contribute significantly to backscatter of the radar signal. Thus, four polarimetric decompositions were compared to separate vegetative versus surface backscatter. The inputs from those decompositions, as well as several additional radar indices and polarizations from the microwave images, were used as feature inputs into two different machine learning models. After a feature selection process, the soil moisture estimations were retrieved and compared using cross-validation. The best overall soil moisture retrieval for Eucalyptus forests came from Random Forest with a RMSE of 0.021, a MAE of 0.017, and a MBE of 0.001. Through a statistical *t*-test, predicted soil moisture values in eucalyptus forests did not differ significantly as compared to other forest types in the study area.

KEYWORDS

Soil Moisture

Polarimetry

Eucalyptus

L-Band SAR

Feature Selection

Machine Learning

ACRONYMS

ALOS – Advanced Land Observing Satellite (Japan)

ASCAT – Advanced Scatterometer (European Space Agency)

ASI – Agenzia Spaziale Italiana (Italian Space Agency)

BIOMASS – Biomass Monitoring Mission for Carbon Assessment (European Space Agency)

[C] – Covariance Matrix

CONAE – Comisión Nacional de Actividades Espaciales (Argentinian Space Agency)

COSMO-SkyMed – Constellation of Small Satellites for Mediterranean basin Observation (Italy)

CSI – Canopy Structure Index

CSV – Comma Separated Values File

dB – Decibel

DEM – Digital Elevation Model

ECMWF – European Centre for Medium-Range Weather Forecasts

EDO – European Drought Observatory

EFAS – European Flood Awareness System

ESA – European Space Agency

ESRI – Environmental Systems Research Institute

GHz – Gigahertz

GRNN – Generalized regression neural network

GRVI – Generalized Radar Vegetation Index

GSD – Ground Sample Distance

HH – Horizontal Transmit, Horizontal Receive

HV – Horizontal Transmit, Vertical Receive

ICNF – Instituto da Conservação da Natureza e das Florestas (Portuguese Institute for the Conservation of Nature and Forests)

IEM – Integral Equation Model

JAXA – Japanese Aerospace Exploration Agency

JRC – Joint Research Centre (European Union)

L – Correlation Length

LAI – Leaf Area Index

LULC – Land Use Land Cover

MAE – Mean Absolute Error

MBE – Mean Bias Error

NASA – National Aeronautics and Space Administration (United States)

NOVA IMS – Information Management School, Universidade Nova de Lisboa

NDVI – Normalized Difference Vegetation Index

PRISM – Panchromatic Remote-sensing Instrument for Stereo Mapping

R^2 – Coefficient of Determination

RCS – Radar Cross Section

RF – Random Forest

RMS – Root Mean Square (height)

RMSE – Root Mean Square Error

RVI – Radar Vegetation Index

[S] – Scattering Matrix

SAOCOM – Satélite Argentino de Observación con Microondas (Argentinian Microwave Observation Satellite)

SAR – Synthetic Aperture Radar

SIASGE – Sistema Italo Argentino de Satélites para la Gestión de Emergencias (Italian-Argentinian Satellite System for Emergency Management)

SLC – Single Look Complex

SMI – Soil Moist Index

SNAP – Sentinel Application Platform

SWI – Soil Water Index

[T] – Coherency Matrix

TU-Wien – Technische Universität Wien

UJI – Universitat Jaume I

VH –Vertical Transmit, Horizontal Receive

VSI – Volume Scattering Index

VV – Vertical Transmit, Vertical Receive

WCM – Water Cloud Model

WWU – Westfälische Wilhelms Universität

INDEX OF THE TEXT

Page

INDEX OF TABLES

INDEX OF FIGURES

1. Introduction

This project was conducted at Universidade Nova de Lisboa, at the Information Management School (NOVA IMS), in Lisbon, Portugal. The development of this project was made possible under the Master of Science in Geospatial Technologies, supported by the European Union's ERASMUS program, and the two other member schools of the Master's degree- Universitat Jaume I (UJI) in Castellón, Spain, and Westfälische Wilhelms Universität (WWU) in Münster, Germany.

1.1 Motivation

Soil moisture is a key physical variable that varies greatly both in time and space. It can and has been defined in many ways, but a simple definition is that it represents the amount of water stored in the unsaturated soil zone [1]. It both affects and is affected by a range of environmental processes in a nonlinear fashion [2]. Soil moisture is an important indicator for both drought and flood risk, and can also be a useful predictor for these hazards [3]. Soil moisture connects land surface and atmospheric phenomena, including the water, carbon, and energy cycles [4]. As a result, it is an important indicator for hydrological, meteorological, climatological, geological, and ecological studies. With the increasing global threat of climate change, its utility has greatly increased, for both monitoring and predictive purposes. Soil moisture directly affects runoff generation, groundwater recharge, and the rate of evaporation, which are crucial inputs for plant growth, water availability, and risk of soil erosion and flooding [5]. Land ecosystems, including forests and agricultural areas, heavily depend on soil moisture- in the form of soil water content and surface soil moisture- to sustain all forms of life. Studies have shown that Mediterranean countries in Europe are already prone to droughts and forest fires and that it is likely that they will experience a significant decline in overall water availability due to warming temperatures, decreased precipitation, and changes in evapotranspiration [6]. Already, rainfall is becoming more unreliable and temperature averages are trending higher. Droughts and heatwaves are becoming more common, and they are increasing in both duration and intensity. Reservoirs in recent years have been operating at increasingly lower

levels. This diminishing amount of water resources, in concert with the relatively large human populations of Portugal and Spain, has put a huge strain on availability of freshwater for both human and ecological use [7].

Forests cover just under a third of the Earth's land surface [8]. Therefore, for global monitoring of soil moisture, it is essential to develop inversion approaches to estimate soil moisture in areas under forest tree canopies. One recent study showed that satellite-based sensors are able to determine soil moisture levels in a temperate forest using *L*-band Synthetic Aperture Radar (SAR), and in high agreement with ground measurements [9]. In one recent study, ALOS-2 and Sentinel-1 was shown to provide good results in estimating soil moisture in agroforestry areas [10].

Portugal is the Southern European country with the most wildfire events and the second-most total burned land area [11]. Portugal has a propensity to be affected by heat waves and droughts- which is likely to accelerate due to the warming global climate. In addition, Portugal has been identified as a "climate hotspot", with particular vulnerabilities identified such as an increase in extreme rainfall events, a lower overall yearly rainfall amount, higher temperatures, and longer dry spells [12]. This worrying mix of factors leads to a potentially bleak climatic outlook for Portugal and highlights the need for continuous monitoring of soil moisture. Regardless, Portugal is already hotter and drier than anytime in recorded history, and that trend is likely to continue [13]. Precipitation has decreased on average by 90mm every decade, as of 2017 [7].

The study area in Central Portugal is dominated by Eucalyptus (*Eucalyptus globulus*) and Maritime Pine (*Pinus pinaster*) forests [11]. Approximately 75% of the forests in Portugal are represented by just 3 species: eucalyptus, maritime pine, and cork oak [14]. Eucalyptus plantations are typically planted as monocultures and used primarily for paper pulp production. They are harvested every 7-12 years. Understory vegetation is sparse in these forests and populated by a narrow range of species [15]. Approximately 3% of Portugal's forests burn every year, and lower soil moisture as a result of warming temperatures and decreased precipitation is a significant risk factor [7].

Reforestation in the Western Iberian Peninsula by fast-growing, commercially-planted tree species - especially eucalyptus, but also pine- has been one of the most significant

land-use changes in the region's history [14]. Replacing traditional pastures and native trees with exotic trees has been identified in previous research as having the potential to change soil characteristics. These soil characteristics with regard to eucalyptus include, but are not limited to: water repellency, porosity, and infiltration rate and hydrological variables include evapotranspiration, water table depth, and streamflow rate [16]. Eucalyptus plantations represent approximately 26% of the total forested area and about 10% of the total land area of Portugal [17]. This conversion of land use has important hydrological and environmental implications. A range of environmentally deleterious effects have been theorized for ecosystems in which eucalyptus is introduced; these effects include being especially susceptible to soil depletion and erosion, decreased biodiversity levels, and contributing to hard-to-control and increasingly frequent wildfires [14]. In particular, eucalyptus plantations may have higher water use and lower overall water use efficiency compared to pine plantations, pasture, or native forests [18]. However, forest plantations in general potentially provide important ecological services by acting as a net carbon-store, cleaning the air, protecting soils from erosion, and providing wood, paper, and pulp products. In Portugal, this wood-products industry is a significant contributor to the economy. Eucalyptus has been widely planted in Portugal because it is fast-growing, comparatively productive, and well-adapted to the climate. Most remote sensing studies on eucalyptus have focused on carbon sequestration, biomass, and productivity. Little focus has been placed on the hydrological effects of eucalyptus species when introduced into non-native areas [19]. Some studies have shown that eucalyptus has negative effects on water resources and can affect hydrological processes in other ways; for example, some studies have investigated whether eucalyptus affects soil water repellency, groundwater, or streamflow [20]. Generalizations about eucalyptus water use are hard to make because of varying local conditions and results remain inconclusive [16]. Nevertheless, this warrants investigation- especially because in Portugal water availability is a major concern.

Microwave electromagnetic waves are sensitive to the dielectric constant of soils, and thus their water content [21]. *In situ* measurements are the traditional method to measure soil moisture, but they can be costly, laborious to set up and maintain, and even ecologically destructive. Remote sensing allows soil moisture to be estimated on a much larger scale and on a continual basis as compared to point measurements [2].

Additionally, hard-to-reach, isolated, or ecologically sensitive areas can be served by remote sensing data, which may not be possible with *in situ* measurements. However, *in situ* measurements will likely remain useful well into the future because they complement and validate remote sensing products. For example, typically satellites only will take an image of the same area on a weekly or biweekly basis. Nevertheless, new satellite systems, in which a constellation of identical satellites can cover the entire earth's land surface in a matter of hours is a fast approaching new reality. Even though, validating and calibrating these remote observations using *in situ* measurements will be beneficial- at least in the areas where it is feasible.

Soil moisture measurements are usually indirect and only apply to a specific definition of soil moisture. Soil moisture estimation usually depends on the volume of soil measured, but may take other forms. This must be taken into account when comparing methods of soil moisture estimation [1]. Most areas of Portugal have been altered and cultivated by humans for centuries and even millenia, so forests typically exist in patches. Because of this, a high spatial resolution is important to retrieve an accurate characterization of soil moisture in forested areas on small scales. *L*-band images around 1.26 GHz are considered ideal for estimating soil moisture in forested areas because trade-offs are minimized between the ability to penetrate vegetation, spatial resolution, and scattering effects [22]. A majority of previous studies have largely been concerned with estimating soil moisture on large scales [3]. SAR's viability has been proven to be effective for a wide variety of applications; nevertheless, downscaling the imagery for analysis, or upscaling point data for validation, has been a major focus [23]. On smaller scales, agriculture is a primary application of remote sensing for this purpose, but limited progress has been made in estimating soil moisture in forests and other heavily-vegetated areas [24]. Drone-based SAR is a relatively new direction of the field; in one recent study, *P* and *C*-band drone-borne SAR approach was shown to provide good results in estimating several biophysical parameters of the forest floor [25]. One study compared *L*-band emission data from a passive microwave sensor to both a theoretical model and a regression model and found that they were able to determine soil moisture with good accuracy in a moderately-dense Eucalyptus forest [26].

Several parameters of the earth's surface in the study area need to be taken into account in order to accurately estimate soil moisture from microwave images; these

include surface roughness, vegetation cover, and topography. Machine learning has been demonstrated to be useful in this regard because of the complexity of the geophysical processes and because estimating soil moisture involves an "inverse problem" (indirect measures are taken); for passive sensors, brightness values are used as inputs, and for active sensors, backscatter coefficients are used [27].

As of yet, data from *L*-Band sensors is less available; data from *C*-band and *X*-band sensors are much more common- but that is likely to change in the near future with the launch of new satellite systems. Most studies before recently used dual-polarized images for polarimetric studies; as fully-polarimetric satellite data has become more available, polarimetric studies have become more common [28]. Point measurements of soil moisture in forested areas are sparse to non-existent (unless placed on a one-off basis for a particular study). Whereas bare surface areas and, increasingly, agricultural areas, have been extensively studied with SAR images to estimate soil moisture, studies are lacking for using SAR to estimate soil moisture in forested areas [22]. When they have taken place, it has been of tropical or boreal forests, and not of temperate forests in the mid-latitudes. More studies are needed for *L*-band emissions in a wider variety of forests and environmental conditions. Some have argued that soil emissions are masked in forests where vegetation is dense, but the degree to which this is the case, and whether ground litter or canopy structure may cause it, is disputed [26]. In Portugal, most forest types are relatively sparse and more needs to be understood about the behavior of long-wavelength radars in areas such as these. Many studies of soil moisture have been conducted using *C*-band data (such as Sentinel-1, which is freely-available), but very few with *L*-band sensors [10]. *C*-band microwaves are not appropriate for forested areas, as it is a shorter wavelength with limited ability to penetrate vegetation cover and reach the soil. *X*-band sensors have shown some preliminary ability to detect soil moisture, but they are not ideal for soil moisture retrieval in general for the same reason as with *C*-band sensors [29]. Using microwave-based vegetation descriptors such as the Radar Vegetation Index (RVI) (as opposed to ground measurements or indices extracted from optical imagery) to use in backscattering models is a relatively unexplored method in the field, and warrants investigation- especially in densely-vegetated areas such as a forest.

1.2 Objectives

The aim of this project is to estimate surface soil moisture in a region of central Portugal with a high proportion of land use characterized by eucalyptus plantations. Eucalyptus has been theorized to negatively affect water availability in areas that it is introduced to. Eucalyptus is not native to Portugal, and a significant proportion of the total land area has been converted to this land use. Moreover, water resources are a major concern because the predominant Mediterranean climate is typified by sporadic and unreliable precipitation patterns. Thus, a secondary but no less important aim is to discern whether surface soil moisture varies significantly between eucalyptus plantations and other forests in the study area. *L*-band SAR satellite images will be used because of their promising balance of trade-offs; they are comparatively high-resolution, can penetrate vegetation cover in forested areas, and have a proven track record of accurately estimating soil moisture in a wide variety of areas. In order to separate backscatter into different components, several polarimetric decompositions will be used. In this study, four incoherent target decompositions will be evaluated to determine their effectiveness: the Freeman-Durden decomposition, which is a model-based decomposition approach, and three eigenvector-eigenvalue decomposition methods- which are the Van Zyl, H/A/α, and the Model-Free Three Component decompositions. In addition, a Digital Elevation Model (DEM) is necessary for terrain correction and for georeferencing the SAR imagery. In a following step, two machine learning algorithms, Generalized Regression Neural Network (GRNN) and Random Forest (RF), will be used for feature selection, to see if the algorithms can produce better results with fewer variables. The algorithms will then be used again to estimate soil moisture values and a cross validation will be conducted to assess model performance. An accuracy assessment for the predicted soil moisture values will be conducted by using the reference data to compare. Finally, forest inventory data will be used together with a statistical *t*-test to see if there is a statistically significant difference between soil moisture values in Eucalyptus forests versus other forests in the study area.

1.3 Research Questions

This study seeks to answer the following questions:

1. Can *L*-band SAR be used to accurately measure soil moisture in eucalyptus

forests?

2. Can features extracted from polarimetric decompositions be used with machine learning to minimize vegetation effects effectively in densely-vegetated environments?
3. Are predicted mean soil moisture values in areas populated by *Eucalyptus globulus* significantly different than predicted mean soil moisture values in the other forested areas in the study area?

1.4 Study Area and Timeframe

Figure 1. Study area location.

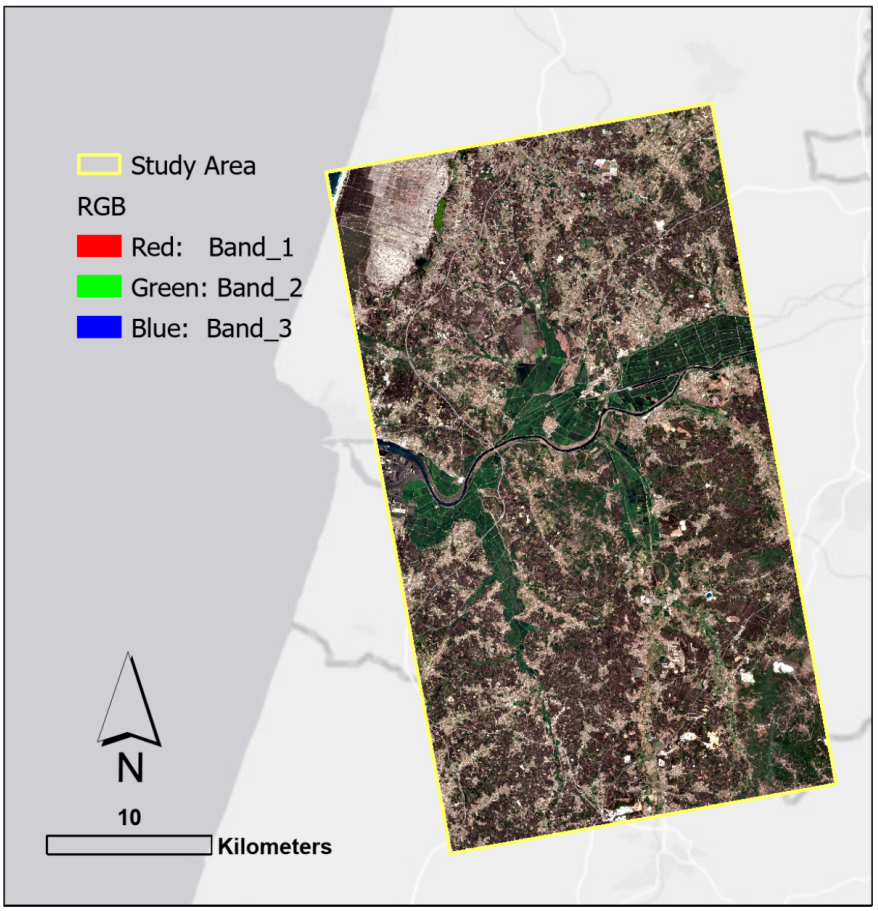

Figure 2. Sentinel-2 True Color satellite image of the study area taken on July 23, 2022 for reference.

The study area consists of 1000 km^2 and lies in Central mainland Portugal, as seen in Figures 1 and 2. It is roughly bounded by the towns of Pombal in the south and Cantanhede in the north, respectively. To the west lies the Atlantic Ocean and the city of Figueira da Foz, and to the East lies the Coimbra urban area. It is a rectangle measuring approximately 24 kilometers wide and 42 kilometers long that was chosen because of the high proportion of Eucalyptus forests as seen in Figures 3 and 4, and also to ensure that all SAR images used in the study overlapped. The first SAOCOM image used in the study was taken on June 21, 2022, and two following images were taken on September 1, 2022 and on October 11, 2022.

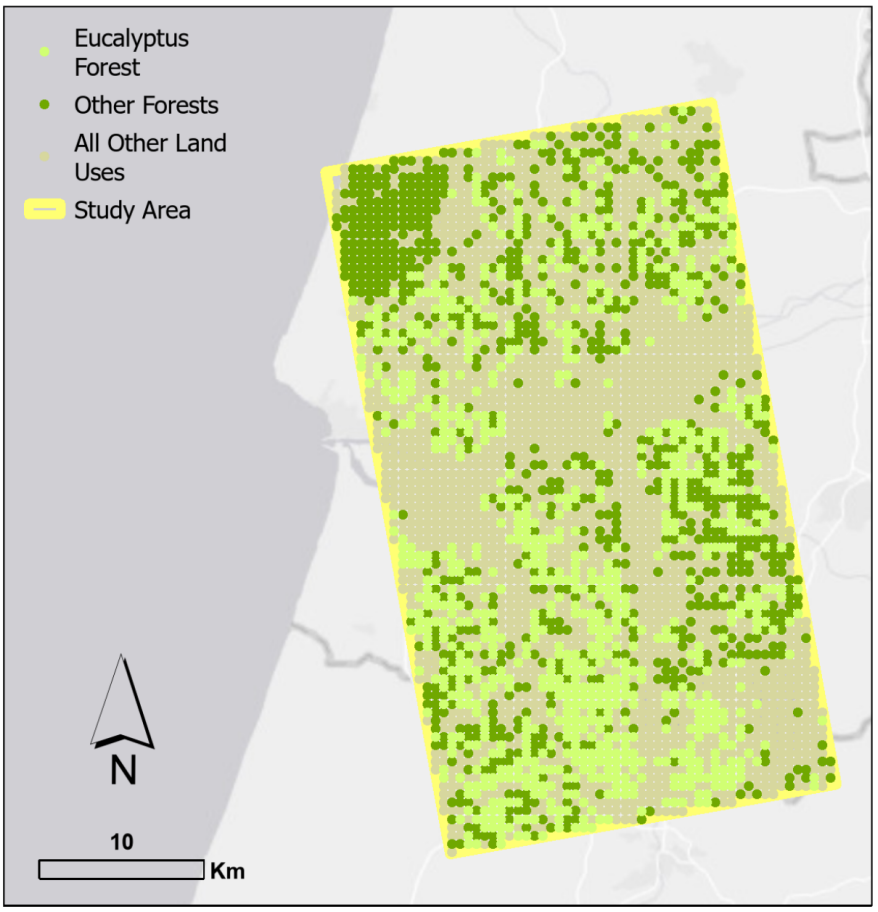

Figure 3. Sample points in the study area divided into three categories for the purpose of the study: Eucalyptus Forests, Other Forests, and All Other Land Uses.

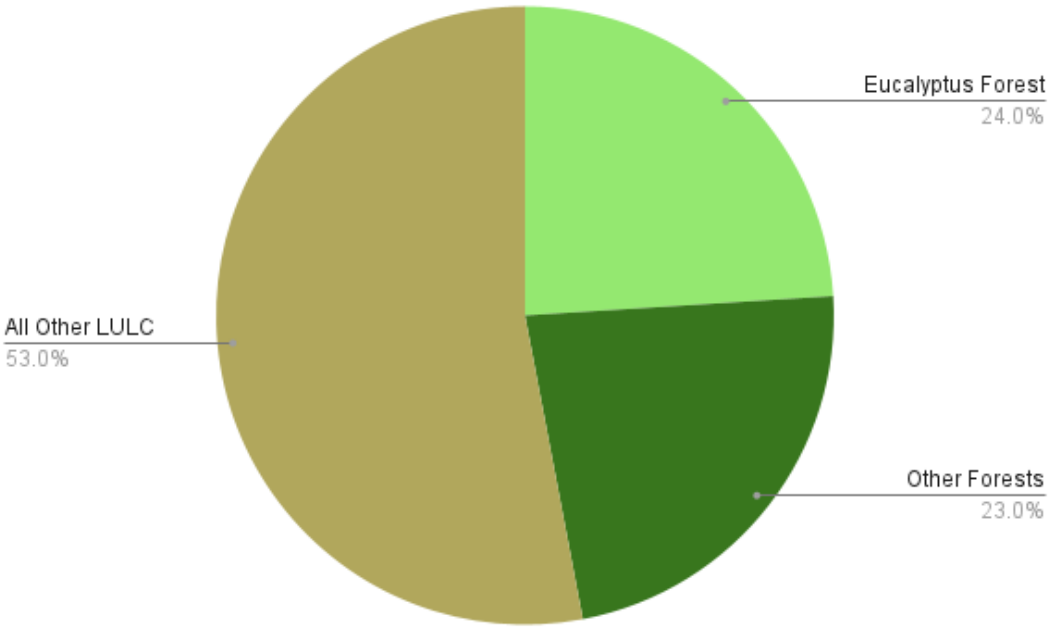

Figure 4. Pie chart showing the percentages of the sample points by the three categories of LULC used in the study.

The climate of the area is Mediterranean, with warm, dry summers and mild and moderately rainy winters. The dominant vegetation types are eucalyptus, pine, rice, olives, grapes, and maize [30], as seen in Figure 5. The Mondego River roughly divides the study area in half, and along its banks lie a large agricultural area [29]. Topographically, the area is mainly flat along the river, but hilly beyond it, with variation between sea level to 400m in elevation. In general, there is heterogeneous land use in the study area, with forests and agriculture being the primary LULC, and urban development (buildings, roadways etc.) representing a significant minority LULC [31].

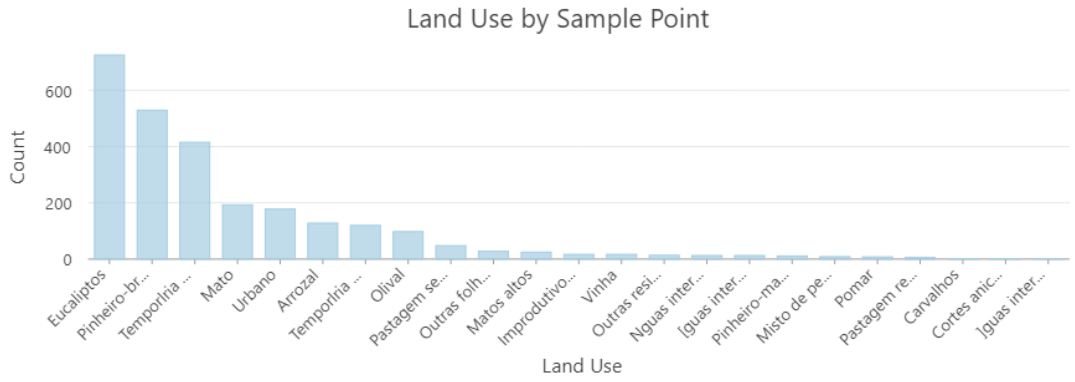

Figure 5. LULC by sample point within the study area using all categories of the ICNF 6th National Forest Inventory for primary land use.

2. Background

Microwave remote sensing exploits the antipodean dielectric constant of water-whether it is located in solid objects or air. This dielectric signature can be interpreted based on the emissivity and backscattering of the microwave signal as it interacts with the Earth's surface, along with other parameters [32]. Passive microwave sensors record the brightness temperature or intensity of land surface electromagnetic emissivity. Passive microwave satellite-based sensors have been used for soil moisture retrieval since 1978 [33].They are characterized by a coarse spatial resolution and are therefore primarily used for soil moisture estimation on regional and global scales. SAR sensors capture a much higher resolution image and instead function by using an active signal that is transmitted by the sensor itself. They work by using "backscatter" values that are reflected by the earth's surface [32]. Backscatter coefficients are defined as the difference in energy of emitted and reflected radiation from the Earth's surface [34]. Mathematically, these backscatter coefficients (σ) are given by the Radar Cross Section (RCS), also called the Radar Signature, which is defined as the ratio of received versus incident radar signal intensities, as shown in Figure 6 [35].

$$\sigma = \frac{I_{received}}{I_{incident}} 4\pi r^2 \left[m^2\right]$$

Figure 6. Radar Cross Section equation, where backscatter coefficients are σ, I is signal intensity, $4\pi r^2$ is the surface of a sphere, and m^2 is the surface area of the target surface.

These backscatter coefficients are functions of both the physical and electrical properties of the soil surface, as well as characteristics of the radar sensor itself [32]. The dielectric constant of the surface, the electrical property, is used as a proxy for soil moisture. The other primary physical characteristics are vegetation and terrain roughness- measured either as correlation length (L), root-mean-square height (rms), or autocorrelation functions [36]. Characteristics of the radar sensor are wavelength, polarization, and incidence angle. Soil moisture retrieval algorithms, or "models" as they are often referred to as, attempt to compute the dielectric constant of the soil surface using backscatter coefficients as inputs [37]. One popular model that accounts for a vegetated surface is called the "Water Cloud Model" or WCM. The WCM is simple to implement and requires few input parameters, so it has been widely used for soil moisture retrieval in vegetated areas [10]. It assumes that the vegetation canopy can be represented by a uniform cloud layer of water droplets and ignores the contribution of higher-order scatterers [38]. Selecting vegetation descriptors that accurately characterize the vegetation canopy in a given study area is a critical step to use the WCM. In the past, most of these descriptors have come from either hand measurements on the ground- which are time-consuming to collect, costly, and laborious- or from indices from optical imagery
such as the Leaf Area Index (LAI) or Normalized Difference Vegetation Index (NDVI). In practice, however, it is unclear how spatially accurate these hand measurements can be, and for optical imagery, it can be difficult to get images at the same resolution and same time as the radar images used. Therefore, a relatively newer technique that has shown promising results is to use microwave-based vegetation descriptors instead such as the Radar Vegetation Index (RVI), or other polarization ratios.

In the beginning years of using microwave remote sensing for soil moisture estimation, the signal was typically analyzed in study areas characterized by bare soils. Moreover, the sensors in use during that time typically only had a single polarization. This led to poor results for soil moisture estimation, however, because with this method, both terrain roughness and soil moisture are contained in one signal [39]. With the arrival of dual-polarized sensors and fully polarimetric ones, these factors could now be separated for a much more accurate soil moisture retrieval using

an inversion approach [40]. These inversion models can be grouped into four broad categories: empirical, semi-empirical, physical/theoretical, and change detection methods [3]. In practice, most models are either semi-empirical or physical, however, because a purely empirical model is very difficult to construct, is only narrowly applicable, and requires very precise data on a number of parameters in order to be accurate. For physical models, the Integral Equation Model (IEM) is perhaps the best-known. In this model, backscattering coefficients are produced as a function of sensor configuration, frequency, polarization, Root Mean Square Height, Correlation Length, autocorrelation, and, of course, the dielectric constant, from which soil moisture can be derived [41]. Two well-known semi-empirical models are the Dubois and Oh models. The Dubois model generates co-polarized backscatter coefficients by using radar sensor parameters, RMS, and the dielectric constant. The Oh model simulates backscatter coefficients by using the ratio of co-polarized to cross-polarized backscatter, along with RMS and the incidence angle of the sensor [34]. Notably however, IEM, as well as the Dubois and Oh models, are only effective in bare soil conditions. Another popular semi-empirical model was developed in response to this, known as the Water Cloud Model, which was mentioned previously. The Change Detection method also deserves mention because, although it cannot be used to absolute soil moisture values, it can be used to estimate the relative change over a period of time [34]. The TU-Wien change detection algorithm is one popular example, which uses observations from the wettest and driest days of the year. Because surface roughness and vegetation structure are assumed to remain the same under this method, no auxiliary information about them is required, which of course can be very helpful in varied use cases.

The microwave signal mainly responds to three factors: soil surface roughness, vegetation canopy, and the dielectric constant- which is primarily related to the surface soil moisture [42]. Backscattering coefficients are affected to a large degree by surface roughness and vegetation. Soil moisture affects them as well, but to a lesser degree [43]. Vegetation has a large effect on the scattering of microwave radiation both because of the water present inside of it as well its structure. SAR microwaves do have the ability, however, to penetrate through vegetation, especially with a longer wavelength- as with *L*-band. The longer the wavelength of the microwave signal, the greater its ability to penetrate through vegetation [44], as shown in Figure 7. *L*-band is

used in this study because it can both penetrate vegetation and the soil to a few centimeters to ascertain soil moisture. *P*-band, an ever longer wavelength band, has even more penetrative ability, which may be useful for determining root-zone soil moisture. However, it also has a higher signal-to-noise ratio and is thus more prone to ionospheric effects and radar frequency interference [45]. Moreover, no *P*-band satellite sensors are currently in orbit.

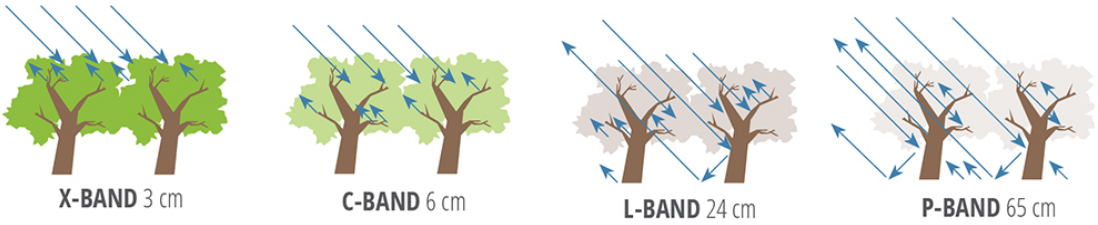

Figure 7. Penetration of vegetation by microwave band wavelength [35].

The presence of vegetation on a study surface greatly complicates estimating soil moisture because both the structure of the vegetation as well as the water present inside of it acts as a scatterer as well as an attenuator of the active microwave signal [46]. There are two primary ways to account for vegetation backscatter: polarization decompositions and vegetation backscattering models, such as the WCM [47]. Polarimetric decompositions are not possible with single-polarized data, and only partially feasible with dual-polarized data [48].

Target decompositions use several different matrices derived from the SAR imagery, such as the scattering matrix [S], the second-order coherency matrix [T], and the covariance matrix [C]. These matrices can be in turn used to derive polarimetric features. These matrices provide scattering descriptors that can be interpreted in relation to the target features on the surface of the Earth [49]. Most natural features are considered partially polarized because they backscatter the transmitted SAR signal in various directions. Based on the way the signal scatters, it can be characterized one of three ways: it may hit the target and bounce once which is called single or surface scattering, it may be reflected twice and is called a dihedral or double-bounce scatterer, or it may reflect more than this and is called a multiple or volume scatterer (all displayed in Figure 8). The elements of these scattering events, both intensity and phase, compose the information in the scattering matrices [37]. Model-based

decompositions, such as the Freeman-Durden decomposition, use these three types of scattering events and their respective magnitudes to single out vegetation scattering. On the other hand, eigenvalue-eigenvector decompositions are solely based on mathematical theory and do not rely on normal statistical assumptions or physical models [50]. The other decompositions used in this study fall under this latter category.

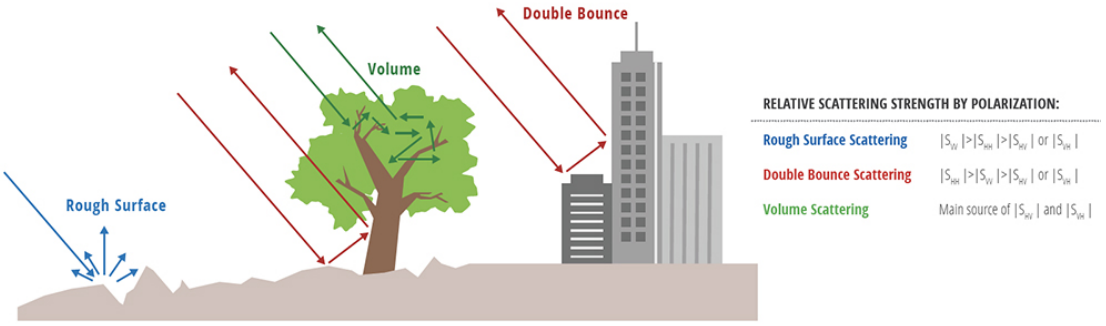

Figure 8. Scattering mechanisms by type and their relationship to scattering strength by polarization [52].

Single-look, non-filtered radar images, such as the ones in this study, have a salt-and-pepper effect, referred to as "speckle" arising from the coherence image formation [45]. Removing speckle does not remove any useful information, as it is simply an image artifact of microwave images; therefore, it should be removed before any polarimetric analysis is conducted. As a result, multi-looking or speckle filters should be used, which reduce speckle.

Polarimetry is the science of acquiring, processing, and analyzing the polarization state of an electromagnetic field [51]. Fully polarimetric SAR refers to both the transmitted waves and received waves as two orthogonal wave polarizations. These polarized electromagnetic waves have two components- a horizontal one and a vertical one, as shown in Figure 9- and are associated together by a well-defined mathematical relationship. When the electric field of a wave is unpredictable, it is considered to be "non-polarized" and random. Conversely, when a wave is considered to be "polarized," its electric field (voltage or strength) is predictable [52]. Radar systems are usually designed to transmit both horizontally and vertically-polarized microwave radiation, which results in wave backscatter being received in a wide

variety of polarizations. The analysis and measurement of these transmitted and received waves is what defines polarimetry and it is useful because of the other information it can reveal about the target surface [53]. A fully polarimetric sensor transmits alternating horizontal and vertical waveforms and records the intensity of both these horizontal and vertical polarizations as well as the phase difference between them [54]. These Quad polarized SAR sensors can therefore capture a complete picture of the scattering behavior of their target and can discriminate between surface features of that target [37].

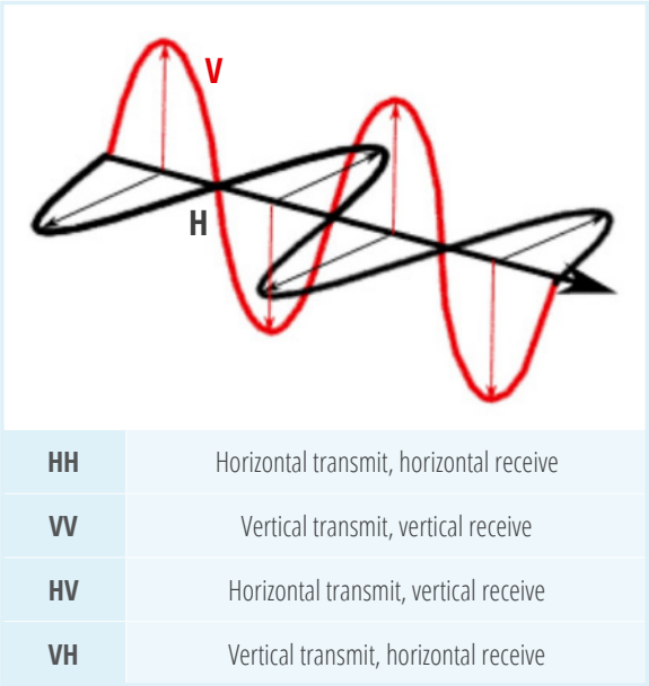

Figure 9. Plane of polarizations for transmitted and received radar signals [35].

Polarimetric target decomposition techniques developed from Dr. J.R. Huynen's research in the 70's. These methods use polarimetric matrices which are "decomposed" into an aggregate of scattering patterns. Using this technique, the dominant scattering form can be determined and, importantly, physical parameters can also be retrieved. Since the scattering matrix is affected by speckle, incoherent analysis using second-order functions from the coherency and covariance matrices are used [55].

Machine learning is highly beneficial for estimating soil moisture based on polarimetric decompositions of SAR imagery for three reasons: (1) theoretical backscatter models are exceedingly complex because they have a large number of inputs that relate to each other in ways that are difficult to discern; (2) they are not able to be applied in widely-varying conditions, because decompositions only prove useful with the specific biological and geological parameters of an individual study area; and (3) there exists a non-linear relationship between SAR image features and natural targets such as soil and vegetation. Machine learning is a useful approach when inputs and outputs relate in a non-linear fashion in complex and varying ways [49]. In the context of environmental remote sensing, data is primarily used to retrieve ecological parameters based on the output of physical models [56]. Machine learning models have demonstrated an ability to accurately approximate the complex, non-linear relationships between ecological parameters due to multilayer learning [57]. Machine learning has been shown to be effective for establishing statistical relationships between *in situ* ecological parameters and remotely sensed data because of its ability to determine complicated and interwoven relationships between many variables [26]. Studies that have used machine learning to retrieve soil moisture can be grouped into three categories based on the type of data that they use to train their models: model-generated training data (from physical or semi-empirical algorithms), *in situ* point measurements, or land surface model simulations. Studies have shown some benefit for comparing satellite products for validation purposes in soil moisture research because it may lead to more accurate results [58].

3. Methodology

In this section, the data used will first be presented and described, followed by a sub-section describing the methods used in the study.

3.1 Data

The SAR images were produced by the Argentinean Microwave Observation Satellite (SAOCOM, Satélite Argentino de Observación Con Microondas) twin satellite constellation referred to as SAOCOM-1A (launched on 8 October 2018) and SAOCOM-1B (launched 30 August 2020) [57]. They are managed by the Argentinean space agency, the National Commission for Space Activities (CONAE,

Comisión Nacional de Actividades Espaciales) and jointly operated with the Italian Space Agency's (ASI, Agenzia Spaziale Italiana) COSMO-SkyMed's satellites. All together, they form the Italian-Argentinian Satellite System for Emergency Management (SIASGE, Sistema Italo-Argentino de Satélites para la Gestión de Emergencias). One of the primary objectives of the SAOCOM satellites is to "generate operative soil moisture maps." The SAOCOM images used in this project are fully polarimetric (also called Quad-Polarized) "Stripmap" images, with a swath width of between 13km and 30km, a spatial resolution of 10m (ground range) by 6m (azimuth), and incidence angles between 18° and 35 ° [56]. They feature the lowest level of processing available, called "L1A" Single Look Complex (SLC) which are "complex data in slant range, radiometrically calibrated with no geometric corrections" [58]. These satellites carry identical active sensors that are 1.275 GHz L-Band SAR sensors. When only one satellite was operating, it covered a given spot on the Earth's surface every 16 days; now that there are two, that has been cut in half to every 8 days. Although the data is not open-access, ASI (which operates jointly with ESA- the European Space Agency), accepts data requests from the global scientific community for non-commercial, research purposes in study areas that fall within the ASI's "Zone of Exclusivity" under international agreements- which roughly corresponds to the European continent [59].

The ALOS World 3D DEM produced by the Japanese Space Agency (JAXA) was used for terrain correction and georeferencing of the SAR images, as displayed in Figure 10. The data is collected by the Panchromatic Remote-sensing Instrument for Stereo Mapping sensor (PRISM) aboard the Advanced Land Observing Satellite (ALOS) which captures panchromatic images at a Ground Sample Distance (GSD) of roughly 1 arc-second (which corresponds roughly to 30m) [60].

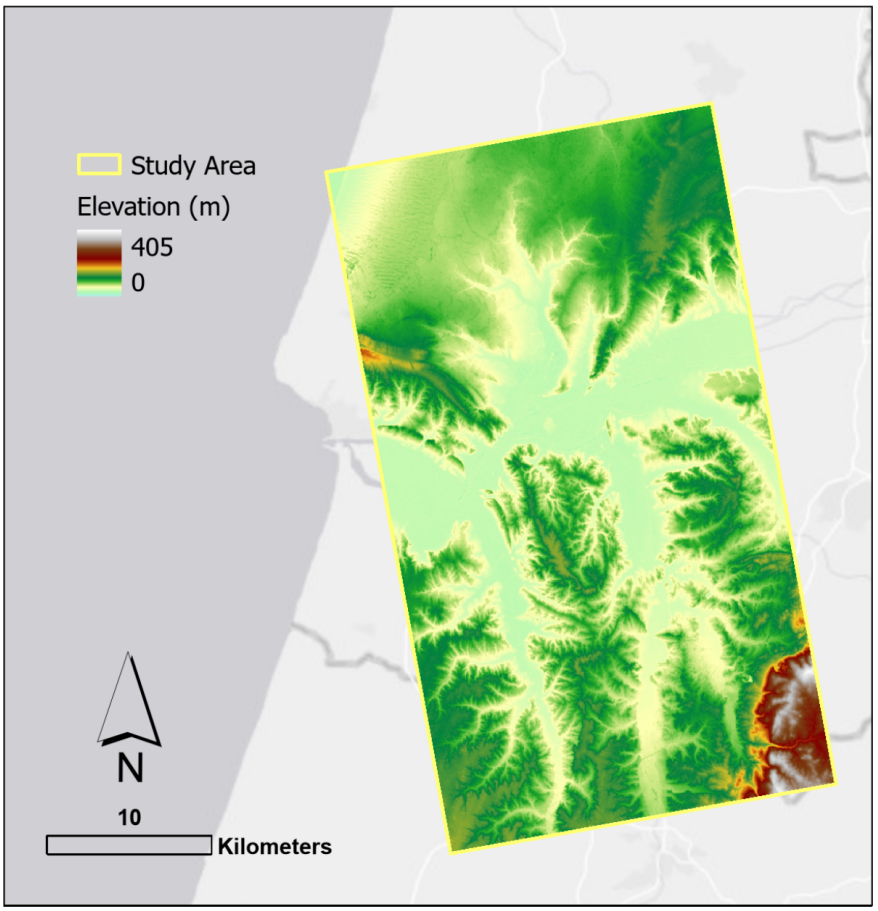

Figure 10. DEM of study area.

The 6th National Forest Inventory of Portugal created by The Institute for Nature Conservation and Forests (ICNF) was used for Land Use/Land Cover (LULC) data about the study area, with a focus on forest types and distributions. The inventory covers all of mainland Portugal using a grid of sample points spaced evenly every 500m and is derived from both aerial images and field evaluations of vegetation [61]. Importantly for this study, forests are characterized by the species of trees that make up the majority of individuals and the inventory includes major species such as eucalyptus and pine. In the study area, the subset of the inventory dataset included 3.999 points.

The soil moisture information originates from two different rasters, as seen in Figure 11. The first is the Soil Moisture Index (SMI) in a raster format at 5-km spatial

resolution and a temporal frequency of 10 days. It is produced by the European Centre for Medium-Range Weather Forecast (ECMWF), and is a component of both the European Flood Awareness System (EFAS) and the European and Global Drought Observatories (EDO)[62]. It is derived from the LISFLOOD hydrological land surface model projections which are produced every 6 hours, and which can determine soil moisture in the top two soil layers. LISFLOOD has been in use since 1996 and is managed by the Joint Research Centre of the European Commission (JRC) [63].

The Soil Water Index (SWI) is also in raster format at a 1km spatial resolution and a temporal frequency of 1 day. It is produced by Copernicus Global Land Service and is based on data fusion of surface soil moisture observations from both Sentinel-1 and Advanced Scatterometer (ASCAT) satellite sensors, as well as a hydrological land surface model. It estimates soil moisture at 8 soil depths and has been disseminated since 2015 [64].

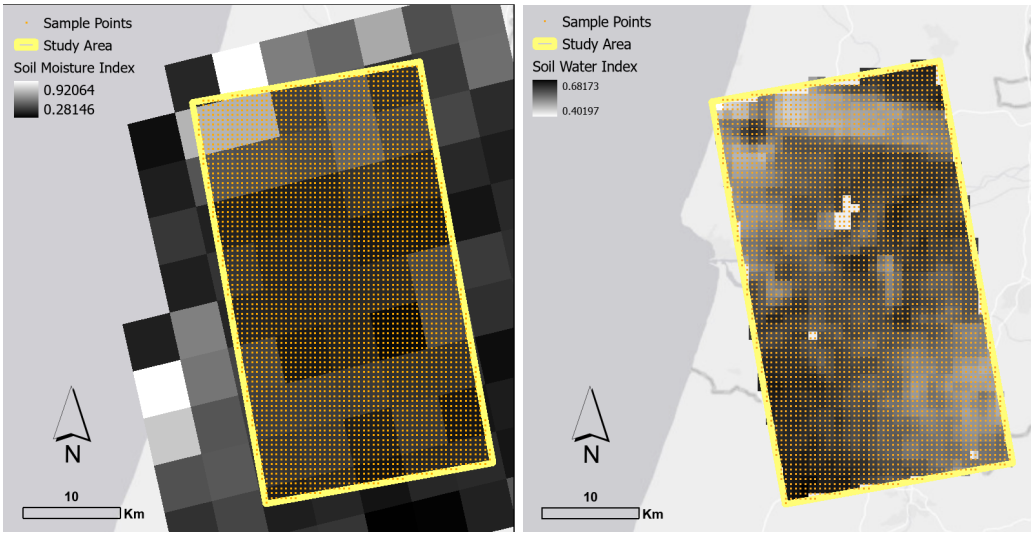

Figure 11. Sample point grid laid from ICNF over the SMI soil moisture reference raster (on the left) and the SWI soil moisture reference raster (on the right) from June 21.

3.2 Methods

The open-source SNAP (Sentinel Application Platform) software, developed by the European Space Agency (ESA), was used for all pre-processing of the

fully-polarimetric SAR images. First, the environment was set in the software and the images were then subsetted; the subset corresponded to the area where the three images overlapped. Because the images were already pre-calibrated, no radiometric calibration was applied. Radar indices (which use band ratios) were extracted directly from the subset to be used as features for inputs; these included RVI, the Generalized Radar Vegetation Index (GRVI), the Volume Scattering Index (VSI), and the Canopy Structure Index (CSI). Then, the scattering matrix [S] was also processed based on the subset. Using the scattering matrix, the coherence [T] and covariance [C] matrices were extracted. To reduce speckle, the coherence and covariance matrices were passed through a speckle filter, called the Lee Refined Filter- using a 5 by 5 moving window (based on the number of pixels). From the covariance matrix, the diagonal elements were also taken as features; they are C11, C22, and C33 (which correspond to HH, HV, and VV polarizations). These elements were also converted from a linear to a decibel (dB) scale. Next, the coherence matrix was used for the polarimetric decompositions, also using a 5 by 5 pixel window; these decompositions were the Freeman-Durden, H/A/α, Van Zyl, and Model Free Three Component decompositions. Altogether, there were 23 features used to predict soil moisture values, as shown in Table 2.

In the following step, the DEM was used for orthoreferencing and range doppler terrain correction using bilinear interpolation. This orthorectified and terrain corrected image was then exported to ArcGIS Pro, created by ESRI, along with the other two datasets used in this study- the forest inventory and soil moisture raster- so that all layers could be overlaid. The point features of the forest inventory were used to extract the information present in the rasters (using bilinear interpolation), which now included soil moisture reference data, along with all of the features used in a later step for the predicted soil moisture values. Thus, several datasets were produced from each of the three dates: the points located in Eucalyptus forests, the points located in other forests, and the points located in all non-forest land use classes. Subsequently, the point data was exported as Comma Separated Value (CSV) tables to Jupyter Notebooks, an open-source software, using Anaconda, for further analysis using Python. Because the data were using several different ranges, all fields were then normalized to be between 0 and 1. Next, to clean up the data, rows with missing values were removed from the datasets, as shown in Table 1. Additionally, outliers

were removed from all features using the metric of interquartile range (IQR); the IQR is multiplied by 1.5 and the resultant product is both subtracted from the 1st quartile and added to the 3rd quartile. Any sample value outside of this resultant range was considered a statistical outlier and thus removed, as seen in Figure 12. Correlation matrices were produced for all of the features using the *pandas* [65] library for Python, as shown in Figure 13. The features were subsequently fed into a feature selection process using two different machine learning methods- Random Forest and GRNN.

In both processes, several levels of feature selection were used: one set included only a few features, another less-restrictive level included several more features, and a final set ignored feature selection altogether by using all features. With the Random Forest algorithm, this feature selection process consists of the mean impurity decrease within each tree- specifically, on each node [66]. The features for internal nodes are selected by variance reduction. For each feature, average decrease in impurity for all trees in the forest is the measure of the feature importance. Some results of this process are shown in Figure 14. For GRNN, the same criteria of mean impurity decrease was used- except using the GRNN algorithm to determine it instead of Random Forest. The three methods of feature selection for Random Forest were Top 3 features (by mean feature importance), all features greater or equal to 0.5 (resulting in around 10 features), and no feature selection. The three methods of feature selection for GRNN were backward, forward, and no feature selection. Backward feature selection starts with all features and removes features one at a time, whereas forward selection starts with no features and then adds features one at a time [67]. Typically, backward feature selection ends with the selection of a greater number of features than forward feature selection.

Then, all three levels were tested in the aforementioned machine learning algorithms to retrieve the predicted soil moisture values using the *scikit-learn* [68] and *pygrnn* [67] libraries for Python. The algorithms then underwent a cross-validation process to tune the parameters and assess accuracy, before all models were given a general accuracy assessment with the best predictions achieved from each of the algorithms. The accuracy of retrieved soil moisture estimations was tested by comparing the best predicted values to the soil moisture reference values using Root Mean Square Error

(RMSE), the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Mean Bias Error (MBE).

Finally, a statistical assessment using a *t*-test and the *scipy* [69] library for Python was conducted to see if there was a statistically meaningful difference between the predicted soil moisture values in Eucalyptus forests as compared to other forest types. For an overall picture of the methodology used in this study, see the flowchart in Figure 15.

	Original Number	**After Missing Values Removed**	**After Outliers Removed**
June 21	3999	3957	2589
September 1	3999	3951	2740
October 11	3999	3961	2761

Table 1. Size of the datasets by date after removal of missing values and outliers.

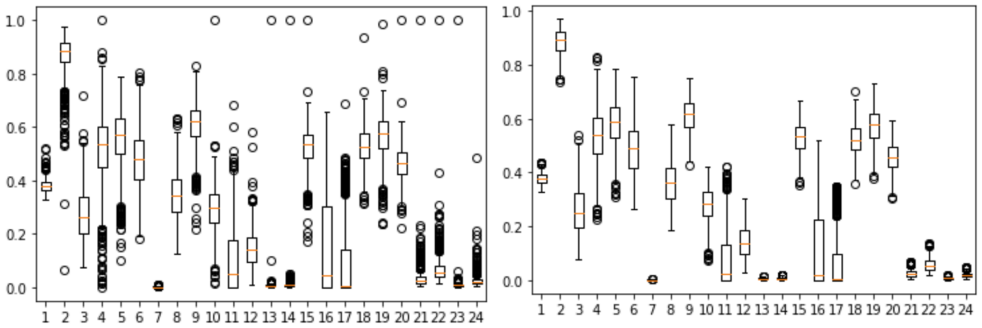

Figure 12. Box plots showing the June 21 Eucalyptus Forest dataset before outlier removal (on the left) and after (on the right). The x-axis are the features and the y-axis are the values of each feature.

Two different machine learning algorithms were used- GRNN and Random Forest (RF). In this way, their capacity to predict soil moisture values could be compared. GRNN is an improved technique for radial basis neural networks that has been shown to be able to work with sparse data, is relatively computationally quick as compared to other neural network models, and functions well when the assumption of linearity cannot be justified (i.e. a non-linear regression problems) [70]. GRNN is a forward-propagation network that executes non-parametric estimation and forms network output by the principle of maximum probability [71]. It contains four layers:

input, pattern, summation, and output. Random Forest, like GRNN, is a machine learning algorithm well-suited to non-parametric estimation, is relatively quick, and works effectively in cases in which the relationships between independent and dependent variables are non-linear [72]. RF uses a technique referred to as bootstrap aggregation to take a sample from the original data to train an individual tree and this procedure is repeated many successive times on other trees (the "forest") until a "bagging" prediction can be formulated from the average of all of the individual predictions of each tree [73]. Thus, it is classified as an ensemble method. It can be used for both classification or regression.

Four different polarimetric decomposition techniques were used in this study: Freeman-Durden, H/A/α (or Cloude-Pottier), Van Zyl, and the Model-Free Three component decomposition. Freeman-Durden is a model-based decomposition that expands the coherency matrix into three sub-matrices each based on non-correlated scattering behaviors of the electromagnetic radiation. First, surface scattering is related to the dielectric constant of the soil, the incidence angle of the sensor itself, and terrain roughness. Second, dihedral, or double-bounce scattering, is a function of the dielectric constant in both vegetation and the soil, and also the incidence angle of the sensor relative to the ground. Third, volume scattering is based on the vegetation canopy which is modeled as a cloud of randomly scattered dipoles [74]. H/A/α also categorizes the backscatter into three different components, but unlike the Freeman-Durden method, H/A/α is a decomposition based on eigenvectors and eigenvalues. First, entropy is related to the probability of effective scattering and ranges between 0 and 1; with 0 representing completely predictable scattering and 1 representing completely random scattering. Second, anisotropy is related to the noise level of a scatterer and is derived from secondary and tertiary eigenvalues. It also ranges between 0 and 1, and needs to be interpreted together with entropy. When it is below 0.7, it largely measures noise, but when it approaches or exceeds 0.7, it can be used to distinguish object scattering type. Third, alpha contains information about the type of scattering that occurs. When it is close or equal to 0, it represents surface, or odd scattering; when it is near $\pi/2$, it represents dihedral scattering; finally, when it approaches $\pi/4$, it represents volume scattering [75]. Van Zyl, like H/A/α, is an eigenvector and eigenvalue approach that was designed for naturally distributed

targets in which the radar cross section would have non-negative values. The radar cross section is the ratio of intensity of backscattered energy to the sensor as compared to the intensity of energy absorbed or otherwise scattered by the target [49]. Like Freeman-Durden, the decomposition is organized into the same three components: surface, dihedral, and volume scattering. The Model-Free Three component decomposition uses three scattering types- odd bounce, even bounce, and volume scattering- and is equivalent to the α (alpha) parameter in the Cloude-Pottier decomposition. Each power component is non-negative (as in the eigenvalue and eigenvector approaches), but it is independent of the antenna receiver configuration allowing total power to be conserved. This technique incorporates other data present in the total power of the signal that cannot be derived from the individual elements of the coherency matrix, hopefully allowing more useful information about the scattering targets to be extracted [76].

The three diagonal elements- C11, C22, and C33- from the 3 by 3 covariance matrix were used to provide information about the target scatterers. C11 represents the HH co-polarization, C22 represents the HV cross-polarization, and C33 represents the VV co-polarization [81]. These single polarimetric intensities can also provide information about target objects. HH is a horizontal polarization that is sensitive to returns from the surface. The cross-polarized HV is sensitive to volume scattering in vegetated areas. VV is a vertical polarization that is sensitive to dihedral scattering. Span, or the sum of the diagonal elements of the covariance matrix, was also included as an input feature. It represents the total power [77]. In addition, band ratios were used as input features; these include the co-polarization ratio of HH/VV, and the cross-polarization ratios of HH/HV and VV/VH.

Four different radar band indices were used as feature inputs for the machine learning algorithms: RVI , GRVI, VSI , and CSI. These are useful for the analysis in this study because they can identify varying vegetation characteristics in forested areas using the backscatter signature of the radar signal. RVI attempts to measure the effects of random scattering from vegetation [78]. It uses scattering intensities in a ratio from cross-polarization to the total power from all polarization bands (span). It varies from between 0 and 1, with a 0 representing a bare surface, and a 1 representing a very dense forest. GRVI is a modified version of the RVI, which incorporates the

generalized volume scattering model [79]. It is based on the geodesic distances between Kennaugh scattering matrices. VSI primarily measures the thickness or density of the tree canopy [80]. It is a ratio of the cross-polarized backscatter to the like-polarized average backscatter. It ranges from 0 to 1, in a similar manner to the RVI and GRVI. CSI primarily measures the relative importance of horizontal versus vertical vegetation structure(e.g. trunks and branches) [80]. It is the normalized ratio difference of co-polarization powers. It ranges from -1 to 1, with a -1 representing completely vertical vegetation scatters only, and a 1 representing completely horizontal scatterers.

Polarimetric Decomposition Features	**Bands and Band Ratios**
1. Entropy - H/A/α	13. HH
2. Anisotropy - H/A/α	14. HV
3. Alpha - H/A/α	15. VV
4. Surface Scattering - Freeman-Durden	16. Span
5. Dihedral Scattering - Freeman-Durden	17. HH/VV
6. Volume Scattering - Freeman-Durden	18. VV/VH
7. Surface Scattering - Van Zyl	19. HH/HV
8. Dihedral Scattering - Van Zyl	20. RVI
9. Volume Scattering - Van Zyl	21. GRVI
10. Odd-Bounce Scattering - Model-Free Three Component	22. VSI
11. Even-Bounce Scattering - Model-Free Three Component	23. CSI
12. Volume Scattering - Model-Free Three Component	

Table 2. All features used in the study; on the left side, the features that were produced from polarimetric decompositions, and on the right, features from single bands and band indices.

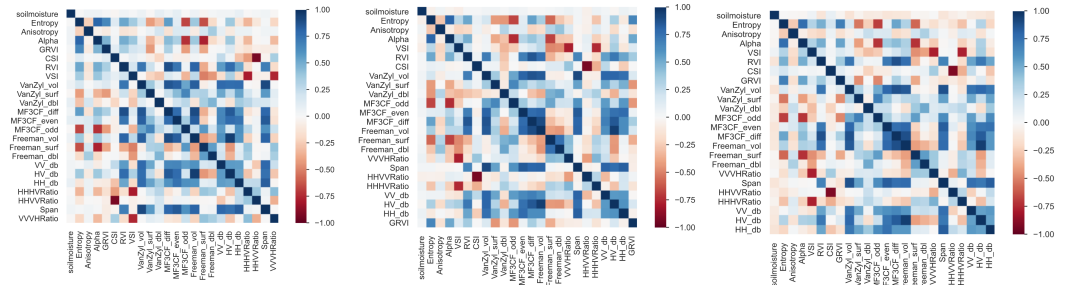

Figure 13. Correlation matrices from the Eucalyptus Forest datasets on all three dates using the SMI raster: June 21, September 1, and October 11 (from left to right).

From the correlation matrices, it is difficult to discern any global patterns. This may reinforce the results of feature selection, in which it was determined that, overall, there are not any variables that have higher predictive ability than the rest. Moreover, it highlights the value of machine learning; with so many features combined with large datasets, it becomes arduous, or even impossible, to find useful patterns and make accurate predictions without the help of algorithms.

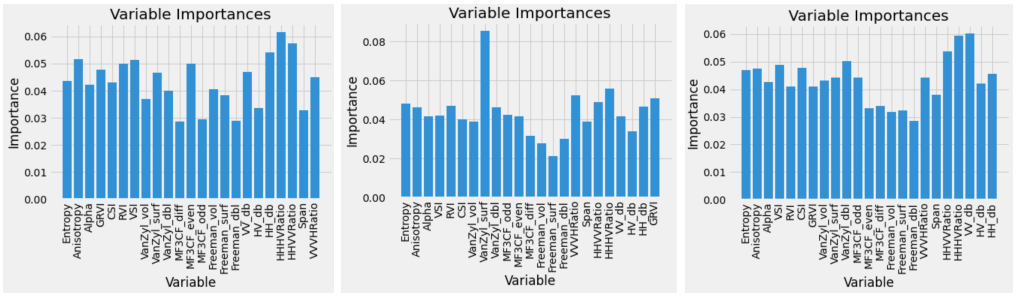

Figure 14. Bar charts showing relative importances of all features using the *scikit-learn* Python library and how effective each feature is individually at predicting the target variable; these are produced from the Eucalyptus forest datasets on all three dates (sequentially from left to right) using the SMI raster..

From the variable importances, as a result of feature selection using GRNN, we see that the band ratios (HHHV, HHVV) are relatively more important during all three dates. In addition, two single bands (HH for June, VV for October) are also among the most important variables. Interestingly for September, we see that the Van Zyl surface scattering is rated much higher than all of the rest. Nevertheless, all features fall within a small range of importance.

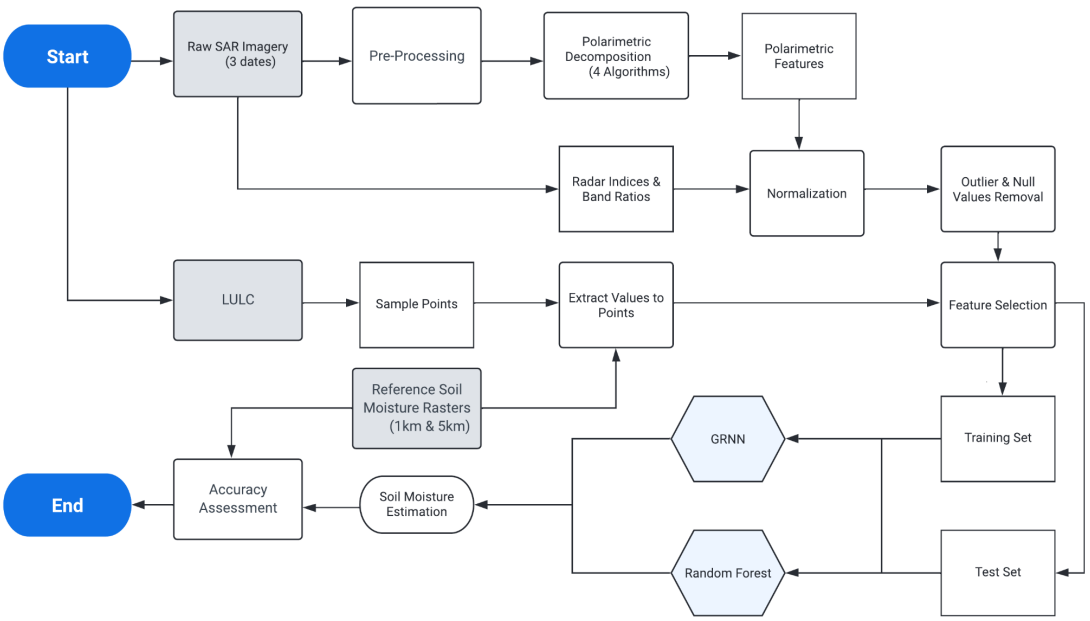

Figure 15. Flowchart of methodology used in this study.

4. Results

In this section, the soil moisture estimations from both machine learning algorithms, GRNN and Random Forest, as well as from both reference rasters, SMI and SWI, are presented. These results are divided by the three dates the SAR images were taken- June 21, September 1, and October 11- and also by LULC type- primarily focusing on eucalyptus forests. The feature selection process is evaluated to see if it led to more accurate estimations and whether it varied by date or LULC. The results of an accuracy assessment are presented to determine how precise soil moisture estimations were in forested areas and if eucalyptus forest, as a LULC, has soil moisture values that are statistically different from soil moisture values in other forest types.

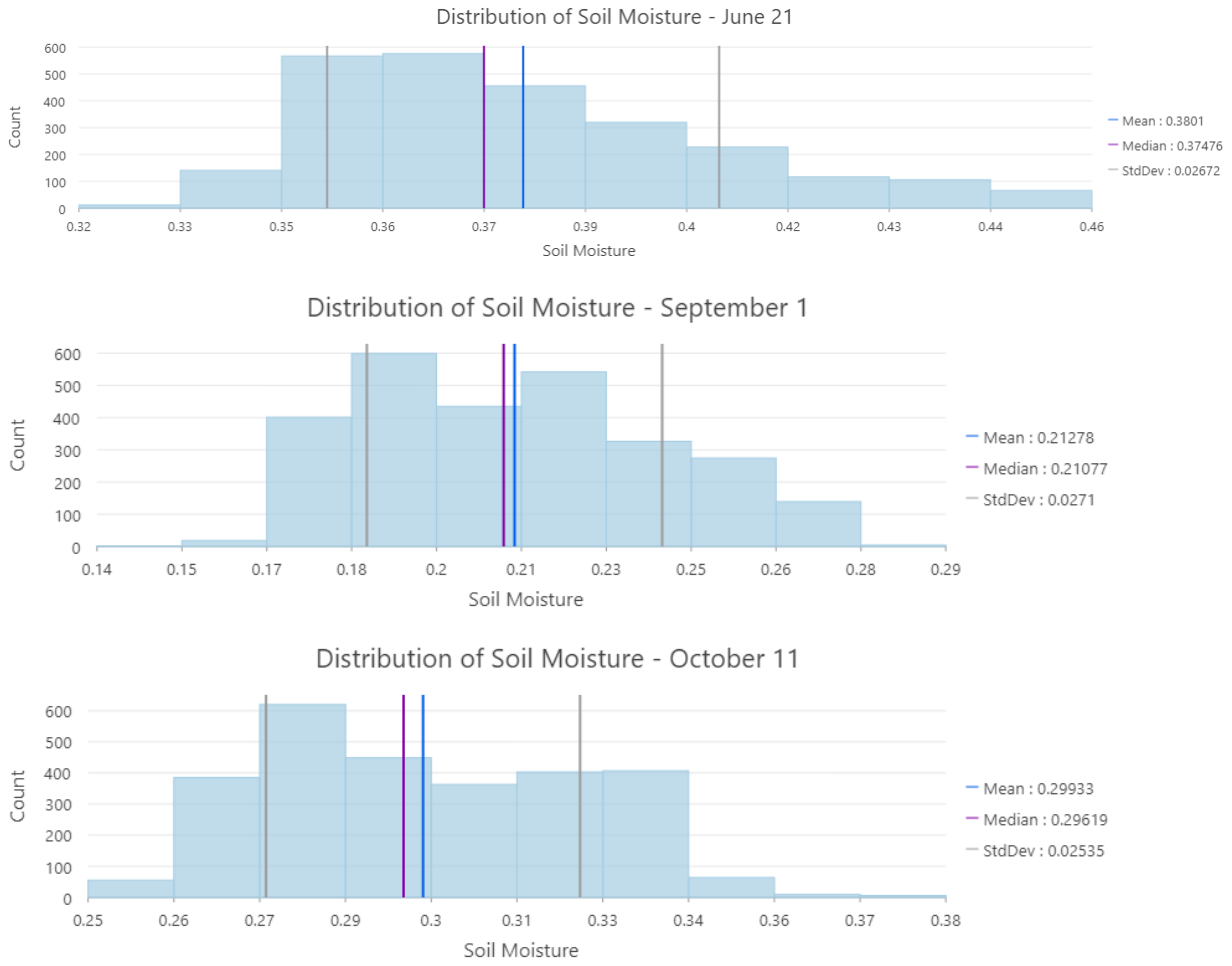

Figure 16. Soil moisture distribution by sample point and by date with descriptive statistics using SMI raster.

Several relevant factors can be interpreted for the soil moisture distribution of all samples in the study area by date, as displayed in Figure 16. When the first SAR image was taken, soil moisture values had a comparatively wider range (from 0.32 to 0.46) and higher average (0.38) than those of the following two images. This makes sense from a climactic perspective because the dry season had not yet taken full effect, and the soils still retained moisture from precipitation earlier in the year. By September 1, the height of the dry season, the range had narrowed and fell substantially (from 0.14 to 0.29) and the average also reflects that (0.21). By October 11, soil moisture had recovered to a small degree, but the distribution is still comparatively much more narrow than in June (from 0.25 to 0.38), and the average is still much lower (0.29). Because the mean is greater than the median in all three

instances, all three distributions are positively skewed, with most samples falling toward the lower values.

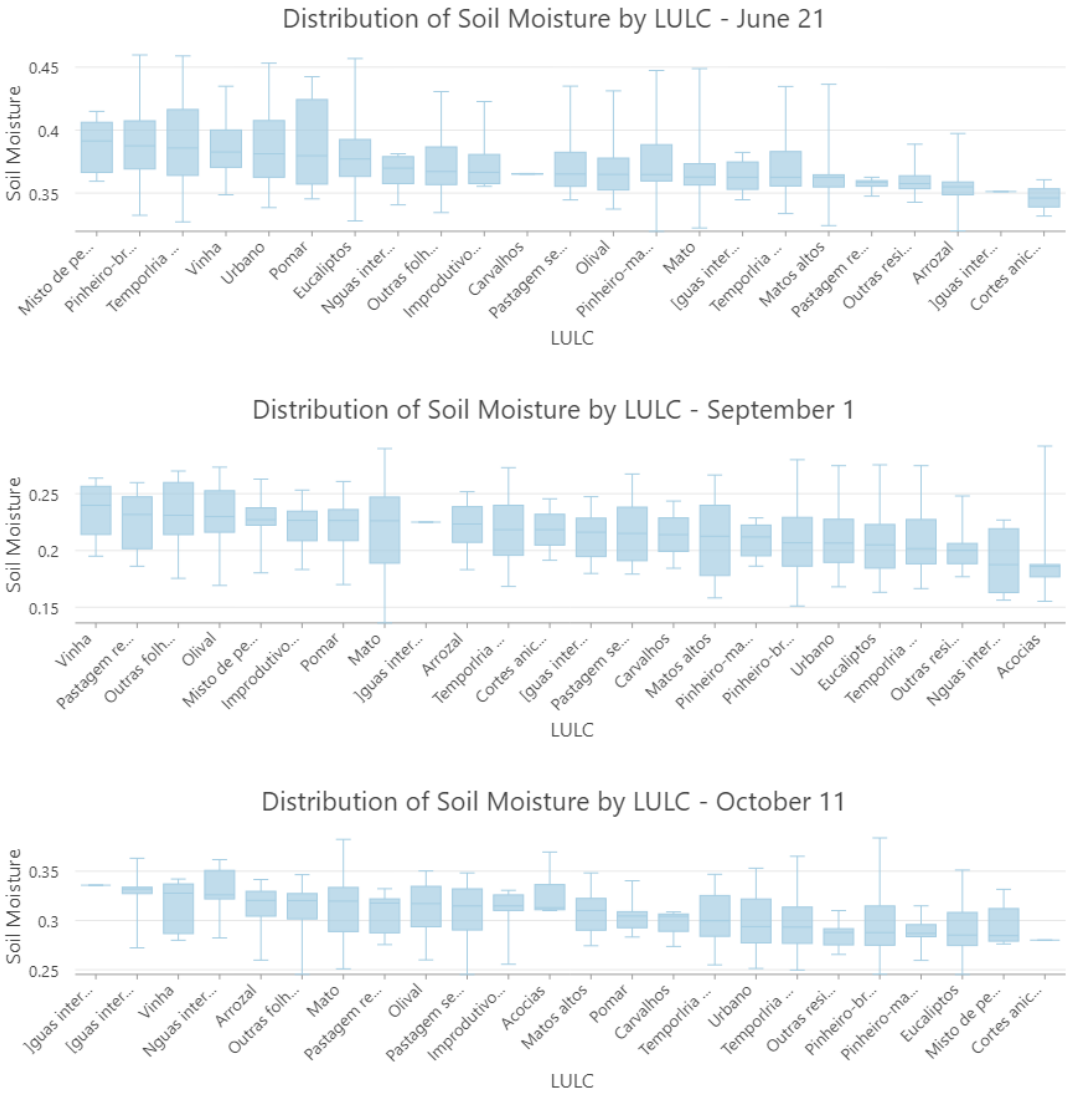

Figure 17. From left to right, decreasing medians of soil moisture by LULC within the study area using all categories of the ICNF 6th National Forest Inventory for primary land use and by date using SMI raster.

From the distributions of soil moisture by LULC (as seen in Figure 17), it is noticeable that Eucalyptus and Pine forests represent the LULC's with some of the highest soil moisture values in June, while they fall toward the bottom of all categories in September and October. This likely means that forests act as moisture sinks- holding onto greater amounts of water when there is more of it available in the wetter months, and gradually shedding it when the weather turns drier.

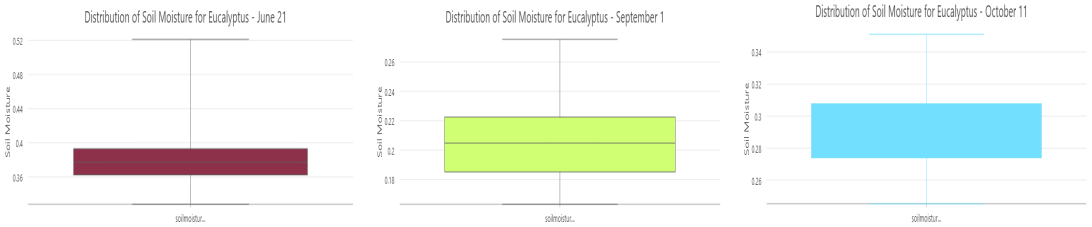

Figure 18. Box plots of soil moisture distributions of only the sample points classified as Eucalyptus Forest by date using SMI raster.

From the distributions of soil moisture for just Eucalyptus forests by date (as shown in Figure 18), we can see the same general trend as the overall distributions of soil moisture by sample point. In June, the distribution is heavily positively skewed; most of the soil moisture values fall toward the lower end, but the range is quite wide and the median is higher than in the following months. In June, the highest value is around 0.52 and the lowest is around 0.34; by September, the highest value is only 0.28 and the lowest value is 0.16- virtually half of the June values.

		R^2	RMSE	MAE	MBE
No Feature Selection	**June 21**	-0.01 (-0.01)	0.023 (0.021)	0.019 (0.017)	-0.002 (0.001)
	September 1	-0.07 (-0.23)	0.028 (0.029)	0.024 (0.024)	0.003 (0.004)
	October 11	-0.09 (-0.07)	0.023 (0.023)	0.020 (0.020)	0.002 (0.001)
Top 3	**June 21**	-0.11	0.024	0.020	-0.001
	September 1	-0.13	0.029	0.024	0.002
	October 11	-0.24	0.025	0.021	0.003
Threshold	**June 21**	-0.05	0.024	0.019	-0.002
	September 1	-0.07	0.028	0.023	0.002
	October 11	-0.16	0.024	0.020	0.002

Table 3. Soil Moisture Retrieval Accuracies in Eucalyptus forests using Random Forest for SMI soil moisture reference raster (results with 5-fold cross-validation in parentheses)

		R^2	RMSE	MAE	MBE
No Feature Selection	**June 21**	-0.10 (-0.08)	0.165 (0.164)	0.132 (0.131)	-0.013 (0.001)
	September 1	-0.11 (-0.07)	0.069 (0.065)	0.055 (0.052)	-0.003 (0.001)
	October 11	-0.11 (-0.05)	0.158 (0.159)	0.133 (0.132)	-0.007 (-0.007)
Top 3	**June 21**	-0.15	0.169	0.138	-0.016
	September 1	-0.19	0.072	0.058	-0.003
	October 11	-0.12	0.159	0.133	-0.003
Threshold	**June 21**	-0.14	0.168	0.132	-0.014
	September 1	-0.22	0.073	0.059	-0.002
	October 11	-0.07	0.156	0.130	-0.004

Table 4. Soil Moisture Retrieval Accuracies in Eucalyptus forests using Random Forest for SWI soil moisture reference raster (results with 5-fold cross-validation in parentheses)

		R^2	RMSE	MAE	MBE
No Feature Selection	**June 21**	-0.01	0.023	0.019	-0.002
	September 1	0.00	0.027	0.023	0.001
	October 11	-0.03	0.023	0.019	0.000

Forward	**June 21**	-0.02	0.023	0.019	-0.003
	September 1	0.00	0.027	0.023	0.001
	October 11	0.02	0.022	0.019	0.001
Backward	**June 21**	-0.01	0.023	0.019	-0.002
	September 1	0.00	0.027	0.023	0.001
	October 11	-0.03	0.023	0.019	0.000

Table 5. Soil Moisture Retrieval Accuracies in Eucalyptus forests using GRNN for SMI soil moisture reference raster

		R^2	**RMSE**	**MAE**	**MBE**
No Feature Selection	**June 21**	0.00	0.158	0.127	-0.004
	September 1	-0.11	0.070	0.056	-0.005
	October 11	-0.07	0.155	0.133	-0.015
Forward	**June 21**	0.00	0.158	0.128	-0.006
	September 1	-0.02	0.067	0.053	-0.004
	October 11	0.00	0.150	0.129	-0.010
Backward	**June 21**	0.00	0.158	0.127	-0.004
	September 1	-0.11	0.070	0.056	-0.005
	October 11	-0.06	0.155	0.133	-0.015

Table 6. Soil Moisture Retrieval Accuracies in Eucalyptus forests using GRNN for SWI soil moisture reference raster

Overall, both Random Forest and GRNN performed similarly well, as shown in Figures 3-6. Neither outperformed the other significantly, and somewhat surprisingly, feature selection did not notably improve results, as was expected. For both reference rasters, and for both machine learning algorithms, feature selection did not result in higher accuracies.

For the SWI raster, accuracies were noticeably lower than those from the SMI raster. Still, the accuracies were fairly high, and the same patterns were evident in the regression plots as with the SMI raster.

A noticeable pattern emerged for prediction accuracies: both algorithms predicted marginally better in June and October than it did in September for the SMI raster, while the opposite was true for the SWI raster. This is likely a result of the SWI raster being more detailed. For Random Forest, feature selection actually decreased accuracy (although not significantly), while for GRNN, feature selection helped to a similarly very small degree. Another significant finding was that predictive accuracies were actually higher in forested areas (at least Eucalyptus forests) than for non-forested areas or for other land uses, in most cases. This is surprising because the sample points in forests were smaller in number compared to the other land uses, and one would also expect soil moisture retrievals to be more difficult in areas with dense vegetation.

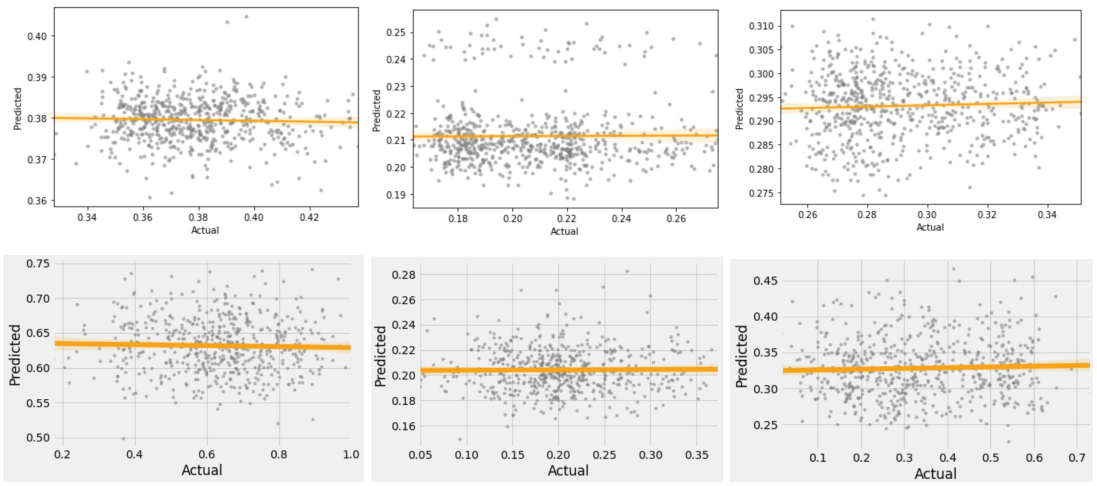

Figure 19. Regression plots of soil moisture reference values (on the x-axis) versus predicted values (on the y-axis) for eucalyptus forests and all three dates (sequentially from left to right) using Random Forest and 5-fold cross-validation for the SMI raster (on top) and the SWI raster (below).

Some interesting patterns emerge for the plots of predicted versus actual soil moisture values, as displayed in Figure 19. In general, the regression line seems to align moderately well with the overall trend. For June and October, generally the predictions are clustered on both sides of the regression line, with a slight tendency to over-predict. In September, however, there is a very apparent break in the values, with the majority of predictions clustered around the lower end of the range, while a separate cluster is apparent on the very top of the range (for the SMI raster).

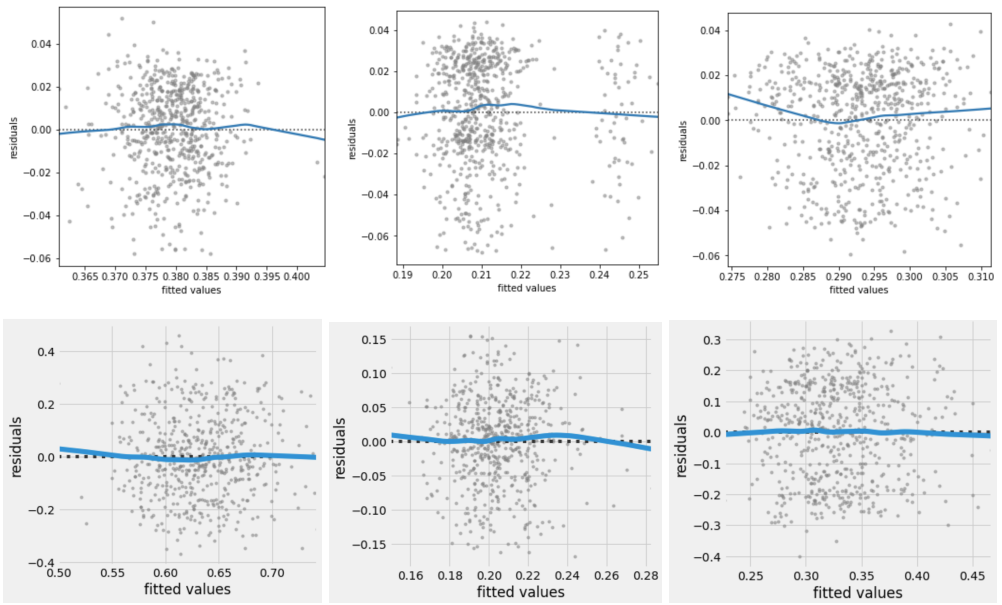

Figure 20. Residual plots of the soil moisture predicted values with fitted values (on the x-axis) and residuals (on the y-axis) for eucalyptus forests and all three dates (sequentially from left to right) for the SMI raster (on top) and the SWI raster (below).

The residual plots look robust overall (as shown in Figure 20), but the same general trends that can be noticed in the regression plots are even more noticeable here. In general, it seems there is a slight tendency to over-predict.

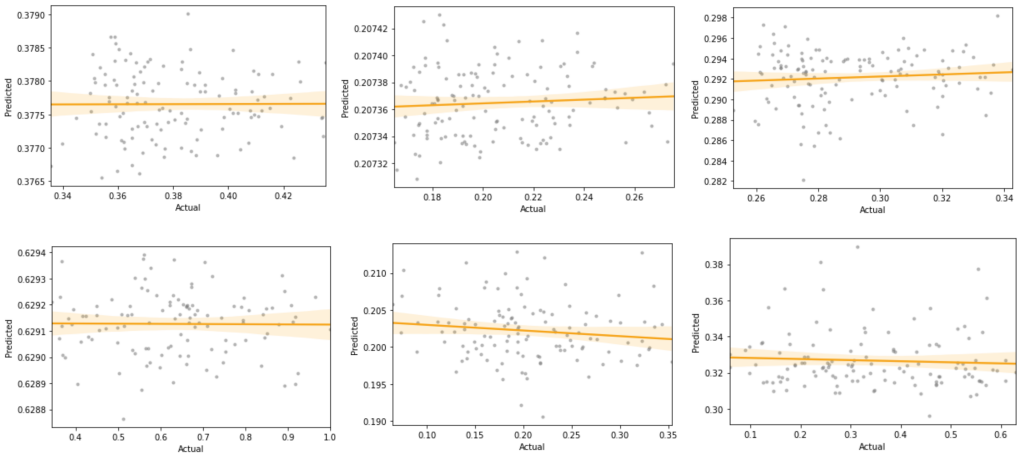

Figure 21. Regression plots of soil moisture reference values (on the x-axis) versus predicted values (on the y-axis) for Eucalyptus forests and all three dates (sequentially from left to right) using GRNN and an 80/20 training/test split for the SMI raster (on top) and the SWI raster (below).

The prediction plots for Eucalyptus from GRNN are also displayed for comparison (Figure 21). Prediction accuracies are comparable to those from Random Forest, and in general, we see the same regression patterns as with Random Forest. Notably, there is a wider error interval (95% confidence interval is shown by the orange shading surrounding the regression line) for the GRNN algorithm, because there are fewer sample points to create the regression as compared to the 5-folds used with Random Forest.

	June 21	**September 1**	**October 11**
Minimum	0.33 (0.36)	0.16 (0.19)	0.25 (0.27)
Maximum	0.44 (0.40)	0.28 (0.25)	0.35 (0.31)
Standard Deviation	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)
Mean	0.38 (0.38)	0.21 (0.21)	0.29 (0.29)
Median	0.38 (0.38)	0.20 (0.21)	0.29 (0.29)

Table 7. Comparative descriptive statistics of reference soil moisture values versus predicted values (in parentheses) using the SMI raster (5km) for eucalyptus forest datasets by date.

	June 21	September 1	October 11
Minimum	0.40 (0.50)	0.05 (0.15)	0.01 (0.23)
Maximum	0.68 (0.74)	0.37 (0.28)	0.73 (0.47)
Standard Deviation	0.05 (0.04)	0.06 (0.02)	0.15 (0.04)
Mean	0.58 (0.63)	0.20 (0.20)	0.34 (0.33)
Median	0.59 (0.63)	0.20 (0.20)	0.32 (0.32)

Table 8. Comparative descriptive statistics of reference soil moisture values versus predicted values (in parentheses) using the SWI raster (1km) for eucalyptus forest datasets by date.

As was mentioned before, studying the distributions of soil moisture for the reference sample points in Eucalyptus forests can help us possibly understand the patterns in the predictions. Although there is not as much variance in values present for the points with eucalyptus as compared to all of the sample points in the area shown at the beginning of this section, September can certainly be seen to be the driest month in comparison to the other two, which may have made predictions in that month more difficult. In general, the predicted values exhibit a smaller range than the reference values for both reference rasters; they have lower minimums, maximums, and standard deviations, while the mean and median values align closely with the reference values (as seen in Tables 7 and 8).

	June 21	September 1	October 11
***p*-value for SMI**	2.64	3.87	2.63
***p*-value for SWI**	1.57	4.26	1.86

Table 9. Statistical *t*-test showing *p*-values for both soil moisture reference rasters using Random Forest with 5 folds and comparing predicted soil moisture values in eucalyptus forests versus other forests.

The statistical *t*-test for eucalyptus forests on all three dates, and using both reference rasters, did not show a statistically significant difference between predicted soil moisture values in eucalyptus forests versus predicted values in other forest types (Table 9). A *p*-value equal to or lower than 0.05 would be required to show statistical

significance at a 95% confidence interval- which is not the case for any of the predicted mean values in this study for eucalyptus versus other forests.

5. Discussion

The results differed based on the reference soil moisture values used. In this study, two reference rasters were used, one at a 5km spatial resolution (SMI) and one with a 1km spatial resolution (SWI). Unsurprisingly, the accuracies were higher for the SMI raster because it is less detailed. Nevertheless, the higher resolution raster provided important context for the accuracies of the predicted soil moisture values. Even more granular soil moisture reference data than was used in this study would certainly lead to lower accuracies as well, but in the end would provide more robust results, as it would better reflect reality. As was mentioned before, because the forests in the study area are fragmented, it would have been much preferred if the spatial resolution of the reference data was higher to discern soil moisture differences in more detail. This is important because soil moisture is known to vary spatially to a high degree and it is difficult to evaluate differences between forested and non-forested areas at 5km and 1km resolutions. It is unclear whether these are sufficient resolutions to discern differences between soil moisture in different LULC's in the study area. Incorporating some *in situ* reference values in future studies could possibly alleviate some of these concerns, if they were available.

Also, although the terrain of the study area is not mountainous, rocky, or characterized by drastic variations, surface roughness is known to distort backscatter information, which makes it more difficult to estimate soil moisture. Therefore, including that information, if it were available or could be easily gathered, would be beneficial. Unfortunately, surface roughness measurements suffer from the same drawbacks as *in situ* soil moisture measurements- which is why they are usually gathered simultaneously. In particular, increasing incidence angles are known to exacerbate distortion from surface variation, as well as imagery that is taken at a very high spatial resolution. On the other hand, increasing incidence angles have been shown to provide more information about the surface in general (because less backscatter comes from vegetation structure). Balancing these tradeoffs is an important

consideration for this type of study. Because there was only one set of images available from SAOCOM in this particular study area, the effect of different incidence angles could not be evaluated; this will certainly change in the future as more *L* and *P*-band SAR data become available. Correlation length, root mean square height, and Gaussian or Exponential autocorrelation function are three metrics that are frequently cited in the literature as being most used for surface roughness reference measurements [36]. It may be useful to incorporate varied sensor incidence angles as well as ground surface roughness measurements in the future to bolster the quality of results.

Data availability was another major concern in this study. The LULC survey data for forests in Portugal only occurs every 10 years. Consequently, the data used in this study was from 2015. To be fair, in the Environmental Remote Sensing field in general, it can be quite difficult to ensure that satellite imagery exactly matches validation observations temporally on the Earth's surface. Nonetheless, it was a source of frustration that *in situ* soil moisture measurements on the ground did not coincide with the L-band imagery that was available. Moreover, those *in situ* measurements were not located in forested areas which, unfortunately for this type of research, is a common occurrence- not just in Portugal, but globally. As a consequence, a soil moisture raster was used instead; this is not necessarily problematic, as point measurements often need to be upscaled regardless using interpolation techniques in order to match the scales of satellite images. However, it is unclear whether the 5km and 1km spatial resolution of the validation data was ideal for the analysis at hand, given that forests in the study area are spatially fragmented. In addition, for polarimetric decomposition techniques to work effectively, quad-polarimetric imagery is required- but in Portugal only one set of these images were available through SAOCOM. ALOS and perhaps other providers will likely make more open data available in the future, but as of yet, there is very limited *L*-band data available in general. Luckily, this set of SAOCOM images that were available covered roughly the same area over a timespan of a half a year at regular intervals; but this essentially dictated the timeframe for this study, as there were no other options. To add to the limited nature of the data, one of the soil moisture rasters was corrupted by algorithm anomalies in the first month of the availability of the microwave imagery. Subsequently, the study could only cover a very limited time of the year- from

summer to early fall. Undoubtedly, it would have been far more preferable to have data from other times of the year, because in Portugal the summers are rather uniformly dry and there is much more variation in precipitation levels, and thus soil moisture, in the winter months. To sum up this concern, it certainly would have added to the robustness of the analysis if more high-quality data were available.

Speckle filtering is another concern in this study that could possibly be improved. A Lee Refined filter was used, which has been shown to lead to good results in other similar studies. Nevertheless, it may have been worthwhile to test using different parameters, or with other filters in comparison, to see if there was a measurable effect on the results. Specifically, speckle filters using a global mean or median for the study area have been shown to generally perform poorly. Testing other adaptive local filters using a moving kernel to maximize parameter retrieval is worth investigating [82].

The parameters of the machine learning algorithms could be tweaked further to possibly yield better results, and it may be additionally worthwhile to compare the results from more types of machine learning algorithms. Artificial Neural Networks (ANN), of which the GRNN is a variation, have successfully been used for over 30 years to retrieve soil moisture values using remotely sensed imagery [83]. Ensemble methods, such as Random Forest used in this study, are a relatively more recent machine learning technique for soil moisture retrieval. Several other methods gaining in popularity deserve mention, which have also been showing increasing promise, such as Deep Learning methods, like Convolutional Neural Networks (CNN), Support Vector methods like Support Vector Machines (SVM), and Gradient Boosting methods like Extreme Gradient Boosting (XGBoost). Incorporating more of these algorithms, and also using varied types of algorithms, to be used for comparative predictive ability in a similar use-case (i.e. soil moisture from a densely-vegetated area using SAR) would almost certainly be useful. It should be noted that some models are much more computationally demanding, so processing time and resources need to be additional considerations. Furthermore, machine learning models largely depend on clean datasets in order to function optimally; statistical outliers, missing or erroneous values, and zeroes (if not warranted) can and will distort predictions. In this study, as in many others, many sample points had to be removed because of this; unfortunately, this comes at the cost of fewer samples overall to train the models, and

marginally lower accuracies than what could theoretically be achieved with a larger dataset (of identical quality) covering the same study area.

6. Conclusion

The primary objectives of this study were: (1) to assess how effective *L*-band SAR is to estimate soil moisture in forested areas populated by *Eucalyptus globulus*, (2) if features derived from polarimetric decompositions and band indices, in concert with machine learning, could provide accurate predictions, and (3), if there is any statistically meaningful difference in the soil moisture values of areas with eucalyptus forests relative to other forested areas. To be successful, the predicted values, especially in forested areas, needed to be in close agreement with the reference values. A machine learning approach was proposed, using two popular algorithms- Random Forest and GRNN- along with feature selection and removal of extreme values. In total, 23 features- coming both from four polarimetric decompositions and band indices- were fed into the algorithms. Several feature selection methods were tried to train the machine learning models with the features with the best predictive ability and several different training-test split methods were experimented with to get the lowest predictive error. In the end, feature selection did not noticeably improve results, and both algorithms performed similarly. The best overall soil moisture retrieval for Eucalyptus forests with the SMI raster used Random Forest and the June 21 dataset, using 5-fold cross-validation and no feature selection with a RMSE of 0.021, a MAE of 0.017, and a MBE of 0.001. Additionally, with p-values of 2.64 and 1.57 for June 21, 3.87 and 4.26 for September 1, and 2.63 and 1.86 for October 11, there was no statistically significant difference found between predicted soil moisture values in eucalyptus forests versus other forest types. Nevertheless, it was shown that *L*-band SAR, combined with features extracted from polarimetric decompositions and machine learning algorithms, could provide good soil moisture predictions in a forested area populated with eucalyptus.

7. Future Studies

Using microwave signals to accurately retrieve soil moisture from vegetated areas, using passive and active sensors, and ground-based, airborne, and satellite platforms, has been a scientific goal for decades; and to a large degree, it has been successful. Using satellite-borne SAR for this purpose, a comparatively newer technology than passive systems, is a less well-developed niche in the research. Many more satellite sensors are coming online soon, and SAR has been one of the technologies with the highest investment, as its utility for diverse monitoring and research purposes has been universally acknowledged. Development of robust and reproducible algorithms for retrieving soil moisture will become even more essential in the future, especially as climate change inevitably causes unpredictable and widely varying effects. Specifically, how the increasing frequency and severity of heat waves, droughts, and conversely, extreme rainfall events, may be related to feedback loops with soil moisture is of special concern. In particular, *P*-band sensors hold unique promise for estimating soil moisture in densely vegetated areas, and several are to be launched in the upcoming years. The utility of SAR for estimating soil moisture, especially in agricultural areas, is at this point well established and proven to work effectively; it remains an important research gap to explore how SAR can be used to estimate soil moisture and other variables in other densely vegetated areas- especially forests. Moreover, fully polarimetric data will become more available with the new sensors and, importantly, much of it will now be free and openly-available; the most-discussed missions in the near future are NISAR (a joint project between NASA and the Indian space agency, ISRO), Tandem-L (funded by the German space agency, DLR), and BIOMASS (from ESA) [35]. Thus, comparing different polarimetric decompositions using varying study areas and sensors is a worthwhile subject for future inquiry. Likewise, deforestation (and land degradation, more broadly) is a major global concern, and as more land is converted from forests to other land uses, using remote sensing to study how soil moisture and other hydrological variables are affected will be of growing importance. For semi-empirical models, a large amount of experimental data is required to assess the importance of parameters and for calibration; as a result, they are often site-specific and therefore poorly reproducible,

and thus only valid under a narrow set of conditions. The extremely complex, interwoven relationships between parameters that can affect soil moisture are very difficult to accurately model, even by experts. Therefore, machine learning will undoubtedly play an important role in future studies, which can produce great results by identifying patterns in non-linear, dynamic, and complex data. More work needs to be done in this area, possibly with the assistance of experts, to avoid the "black box" problem- whereby the user has limited to no control over algorithms except to provide input data. Validation is a crucial consideration in this research area, as soil moisture can be defined in multiple ways and is only measured indirectly in the vast majority of cases. Having access to more *in situ* data points would clearly be preferable; however, it remains to be seen whether this is feasible or even advisable considering the high maintenance costs, difficulty in access, and possibly environmentally destructive nature of it in practice. Scaling soil moisture also remains a concern, as *in situ* point measurements are extremely sparse in most areas of the world, and validation becomes nebulous when trying to compare data on vastly different scales. The growing number of soil moisture "networks" and citizen/volunteer collected data projects will address this dearth of data to some degree, but there is no doubt that remote sensing will play an increasingly large role in estimating soil moisture on both global and local scales.

BIBLIOGRAPHIC REFERENCES

[1] S. I. Seneviratne *et al.*, “Investigating soil moisture–climate interactions in a changing climate: A review,” *Earth-Sci. Rev.*, vol. 99, no. 3, pp. 125–161, May 2010, doi: 10.1016/j.earscirev.2010.02.004.

[2] A. W. Western, R. B. Grayson, and G. Blöschl, “Scaling of Soil Moisture: A Hydrologic Perspective,” *Annu. Rev. Earth Planet. Sci.*, vol. 30, no. 1, pp. 149–180, 2002, doi: 10.1146/annurev.earth.30.091201.140434.

[3] E. Babaeian, M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller, “Ground, Proximal, and Satellite Remote Sensing of Soil Moisture,” *Rev. Geophys.*, vol. 57, no. 2, pp. 530–616, 2019, doi: 10.1029/2018RG000618.

[4] D. A. Robinson *et al.*, “Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review,” *Vadose Zone J.*, vol. 7, no. 1, pp. 358–389, 2008, doi: 10.2136/vzj2007.0143.

[5] H. Li, A. Robock, and M. Wild, “Evaluation of Intergovernmental Panel on Climate Change Fourth Assessment soil moisture simulations for the second half of the twentieth century,” *J. Geophys. Res.*, vol. 112, no. D6, p. D06106, Mar. 2007, doi: 10.1029/2006JD007455.

[6] P. M. M. Soares and D. C. A. Lima, “Water scarcity down to earth surface in a Mediterranean climate: The extreme future of soil moisture in Portugal,” *J. Hydrol.*, vol. 615, p. 128731, Dec. 2022, doi: 10.1016/j.jhydrol.2022.128731.

[7] C.-F. Schleussner, I. Menke, E. Theokritoff, N. van Maanen, and A. Lanson, “Climate Impacts in Portugal”.

[8] “Seeing Forests for the Trees and the Carbon: Mapping the World’s Forests in Three Dimensions,” Jan. 09, 2012. https://earthobservatory.nasa.gov/features/ForestCarbon (accessed Jan. 07, 2023).

[9] A. Colliander *et al.*, “SMAP Detects Soil Moisture Under Temperate Forest Canopies,” *Geophys. Res. Lett.*, vol. 47, no. 19, p. e2020GL089697, 2020, doi: 10.1029/2020GL089697.

[10] H. Cui *et al.*, “The Potential of ALOS-2 and Sentinel-1 Radar Data for Soil Moisture Retrieval With High Spatial Resolution Over Agroforestry Areas, China,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2021.3082805.

[11] J. Parente, A. Girona-García, A. R. Lopes, J. J. Keizer, and D. C. S. Vieira, “Prediction, validation, and uncertainties of a nation-wide post-fire soil erosion risk assessment in Portugal,” *Sci. Rep.*, vol. 12, no. 1, Art. no. 1, Feb. 2022, doi: 10.1038/s41598-022-07066-x.

[12] L. A. Espinosa, M. M. Portela, J. P. Matos, and S. Gharbia, “Climate Change Trends in a European Coastal Metropolitan Area: Rainfall, Temperature, and Extreme Events (1864–2021),” *Atmosphere*, vol. 13, no. 12, Art. no. 12, Dec. 2022, doi: 10.3390/atmos13121995.

[13] R. M. Cardoso, P. M. M. Soares, D. C. A. Lima, and P. M. A. Miranda, “Mean and extreme temperatures in a warming climate: EURO CORDEX and WRF regional climate high-resolution projections for Portugal,” *Clim. Dyn.*, vol. 52, no. 1, pp. 129–157, Jan. 2019, doi: 10.1007/s00382-018-4124-4.

[14] L. J. R. Nunes, C. I. R. Meireles, C. J. Pinto Gomes, and N. M. C. Almeida Ribeiro, “Historical Development of the Portuguese Forest: The Introduction of Invasive Species,” *Forests*, vol. 10, no. 11, p. 974, Nov. 2019, doi: 10.3390/f10110974.

[15] G. Leighton-Boyce *et al.*, “Temporal Dynamics of Water Repellency and Soil Moisture in Eucalypt Plantations, Portugal,” *Aust. J. Soil Res. - AUST J SOIL RES*, vol. 43, Jan. 2005, doi: 10.1071/SR04082.
[16] W. Liu *et al.*, “Estimations of evapotranspiration in an age sequence of Eucalyptus plantations in subtropical China,” *PLOS ONE*, vol. 12, no. 4, p. e0174208, Apr. 2017, doi: 10.1371/journal.pone.0174208.
[17] F. X. Catry, F. Moreira, E. Deus, J. S. Silva, and A. Águas, “Assessing the extent and the environmental drivers of Eucalyptus globulus wildling establishment in Portugal: results from a countrywide survey,” *Biol. Invasions*, vol. 17, no. 11, pp. 3163–3181, Nov. 2015, doi: 10.1007/s10530-015-0943-y.
[18] J. A. Stanturf, E. D. Vance, T. R. Fox, and M. Kirst, “Eucalyptus beyond Its Native Range: Environmental Issues in Exotic Bioenergy Plantations,” *Int. J. For. Res.*, vol. 2013, p. e463030, Apr. 2013, doi: 10.1155/2013/463030.
[19] “Water consumption of Eucalyptus urophylla plantations on the Leizhou Peninsula | Request PDF.” https://www.researchgate.net/publication/287516208_Water_consumption_of_Eucalyptus_urophylla_plantations_on_the_Leizhou_Peninsula (accessed Dec. 30, 2022).
[20] V. Engel, E. G. Jobbágy, M. Stieglitz, M. Williams, and R. B. Jackson, “Hydrological consequences of Eucalyptus afforestation in the Argentine Pampas,” *Water Resour. Res.*, vol. 41, no. 10, 2005, doi: 10.1029/2004WR003761.
[21] J. R. Wang, “The dielectric properties of soil-water mixtures at microwave frequencies,” *Radio Sci.*, vol. 15, no. 05, pp. 977–985, Sep. 1980, doi: 10.1029/RS015i005p00977.
[22] M. Kurum, S.-B. Kim, R. Akbar, and M. H. Cosh, “Surface Soil Moisture Retrievals Under Forest Canopy for L-Band SAR Observations Across a Wide Range of Incidence Angles by Inverting a Physical Scattering Model,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 1741–1753, 2021, doi: 10.1109/JSTARS.2020.3047883.
[23] S. Sabaghy *et al.*, “Comprehensive analysis of alternative downscaled soil moisture products,” *Remote Sens. Environ.*, vol. 239, p. 111586, Mar. 2020, doi: 10.1016/j.rse.2019.111586.
[24] Y. Bao, L. Lin, S. Wu, K. A. Kwal Deng, and G. P. Petropoulos, “Surface soil moisture retrievals over partially vegetated areas from the synergy of Sentinel-1 and Landsat 8 data using a modified water-cloud model,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 72, pp. 76–85, Oct. 2018, doi: 10.1016/j.jag.2018.05.026.
[25] L. Moreira *et al.*, “Drone-Borne P-band Single-Pass InSAR,” in *2020 IEEE Radar Conference (RadarConf20)*, Sep. 2020, pp. 1–6. doi: 10.1109/RadarConf2043947.2020.9266502.
[26] R. Rahmoune, P. Ferrazzoli, J. P. Walker, and J. P. Grant, “L-band emission from a Eucalyptus forest in various soil conditions during the nafe campaign,” in *2010 11th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment*, Washington, DC, USA, Mar. 2010, pp. 81–85. doi: 10.1109/MICRORAD.2010.5559584.
[27] Q. Yuan *et al.*, “Deep learning in environmental remote sensing: Achievements and challenges,” *Remote Sens. Environ.*, vol. 241, p. 111716, May 2020, doi: 10.1016/j.rse.2020.111716.
[28] S. Gharechelou, R. Tateishi, J. T. Sri Sumantyo, and B. A. Johnson, “Soil Moisture Retrieval Using Polarimetric SAR Data and Experimental Observations

in an Arid Environment," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 10, Art. no. 10, Oct. 2021, doi: 10.3390/ijgi10100711.
[29] M. El Hajj *et al.*, "Soil moisture retrieval over irrigated grassland using X-band SAR data," Jan. 2016.
[30] L. Dimuccio, R. Ferreira, L. Cunha, and A. Almeida, "Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment," vol. 20, pp. 776–791, May 2011.
[31] M. Cunha, J. Marques, J. Azevedo, and A. Castilho, "Understanding the Impact of a Major Hydro-Agricultural Project in Low Mondego Area (Portugal)," *Land*, vol. 10, no. 2, Art. no. 2, Feb. 2021, doi: 10.3390/land10020114.
[32] F. T. Ulaby, R. K. Moore, and A. K. Fung, "Microwave remote sensing: Active and passive. Volume 2 - Radar remote sensing and surface scattering and emission theory." Jan. 01, 1982. Accessed: Dec. 28, 2022. [Online]. Available: https://ntrs.nasa.gov/citations/19830032822
[33] X. Chen *et al.*, "A semi-empirical inversion model for assessing surface soil moisture using AMSR-E brightness temperatures," *J. Hydrol.*, vol. 456–457, pp. 1–11, Aug. 2012, doi: 10.1016/j.jhydrol.2012.05.022.
[34] L. Karthikeyan, M. Pan, N. Wanders, D. N. Kumar, and E. F. Wood, "Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms," *Adv. Water Resour.*, vol. 109, pp. 106–120, Nov. 2017, doi: 10.1016/j.advwatres.2017.09.006.
[35] "SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation," *SERVIR*. https://servirglobal.net/Global/Articles/Article/2674/SAR-Handbook-Comprehensive-Methodologies-for-Forest-Monitoring-and-Biomass-Esti (accessed Dec. 16, 2022).
[36] "DL15-NEWFOR_Roughness_state_of_the_art.pdf." Accessed: Jan. 27, 2023. [Online]. Available: https://www.newfor.net/wp-content/uploads/2015/02/DL15-NEWFOR_Roughness_state_of_the_art.pdf
[37] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996, doi: 10.1109/36.485127.
[38] E. P. W. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Sci.*, vol. 13, no. 2, pp. 357–364, 1978, doi: 10.1029/RS013i002p00357.
[39] R. Bernard, P. H. Martin, J. L. Thony, M. Vauclin, and D. Vidal-Madjar, "C-band radar for determining surface soil moisture," *Remote Sens. Environ.*, vol. 12, no. 3, pp. 189–200, Jul. 1982, doi: 10.1016/0034-4257(82)90052-9.
[40] I. Hajnsek, E. Pottier, and S. R. Cloude, "Inversion of surface parameters from polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 727–744, Apr. 2003, doi: 10.1109/TGRS.2003.810702.
[41] A. K. Fung and K. S. Chen, "An update on the IEM surface backscattering model," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 2, pp. 75–77, Apr. 2004, doi: 10.1109/LGRS.2004.826564.
[42] T. J. JACKSON, J. SCHMUGGE, and E. T. ENGMAN, "Remote sensing applications to hydrology: soil moisture," *Hydrol. Sci. J.*, vol. 41, no. 4, pp. 517–530, Aug. 1996, doi: 10.1080/02626669609491523.
[43] S.-B. Kim *et al.*, "Surface Soil Moisture Retrieval Using the L-Band Synthetic Aperture Radar Onboard the Soil Moisture Active–Passive Satellite and

Evaluation at Core Validation Sites," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1897–1914, Apr. 2017, doi: 10.1109/TGRS.2016.2631126.
[44] S. R. Cloude, D. G. Corr, and M. L. Williams, "Target detection beneath foliage using polarimetric synthetic aperture radar interferometry," *Waves Random Media*, vol. 14, no. 2, pp. S393–S414, Apr. 2004, doi: 10.1088/0959-7174/14/2/015.
[45] "Synergies for Soil Moisture Retrieval Across Scales From Airborne Polarimetric SAR, Cosmic Ray Neutron Roving, and an In Situ Sensor Network - Fersch - 2018 - Water Resources Research - Wiley Online Library." https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018WR023337 (accessed Dec. 29, 2022).
[46] I. Gherboudj, R. Magagi, A. A. Berg, and B. Toth, "Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 33–43, Jan. 2011, doi: 10.1016/j.rse.2010.07.011.
[47] Z. Wang, T. Zhao, J. Qiu, X. Zhao, R. Li, and S. Wang, "Microwave-based vegetation descriptors in the parameterization of water cloud model at L-band for soil moisture retrieval over croplands," *GIScience Remote Sens.*, vol. 58, no. 1, pp. 48–67, Jan. 2021, doi: 10.1080/15481603.2020.1857123.
[48] J. M. Lopez-Sanchez *et al.*, "Agriculture and Wetland Applications," in *Polarimetric Synthetic Aperture Radar: Principles and Application*, I. Hajnsek and Y.-L. Desnos, Eds. Cham: Springer International Publishing, 2021, pp. 119–178. doi: 10.1007/978-3-030-56504-6_3.
[49] Z. Akhavan, M. Hasanlou, M. Hosseini, and H. McNairn, "Decomposition-Based Soil Moisture Estimation Using UAVSAR Fully Polarimetric Images," *Agronomy*, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/agronomy11010145.
[50] "Polarimetric Radar Imaging | From Basics to Applications | Jong-Sen Le." https://www.taylorfrancis.com/books/mono/10.1201/9781420054989/polarimetric-radar-imaging-jong-sen-lee-eric-pottier (accessed Jan. 05, 2023).
[51] "ADA437484.pdf." Accessed: Jan. 04, 2023. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA437484.pdf
[52] "SECTION 3 - Carleton SAR Training - SAR Polarimetry - Final.pdf." Accessed: Jan. 04, 2023. [Online]. Available: https://dges.carleton.ca/courses/IntroSAR/Winter2019/SECTION%203%20-%20Carleton%20SAR%20Training%20-%20SAR%20Polarimetry%20%20-%20Final.pdf
[53] "ClosaTarresMarc_TFG.pdf." Accessed: Dec. 16, 2022. [Online]. Available: https://upcommons.upc.edu/bitstream/handle/2117/168497/ClosaTarresMarc_TFG.pdf?sequence=4&isAllowed=y
[54] D. Haldar, A. Das, S. Mohan, O. Pal, R. S. Hooda, and M. Chakraborty, "ASSESSMENT OF L-BAND SAR DATA AT DIFFERENT POLARIZATION COMBINATIONS FOR CROP AND OTHER LANDUSE CLASSIFICATION," *Prog. Electromagn. Res. B*, vol. 36, pp. 303–321, 2012, doi: 10.2528/PIERB11071106.
[55] S.-W. Chen, Y.-Z. Li, X.-S. Wang, S.-P. Xiao, and M. Sato, "Modeling and Interpretation of Scattering Mechanisms in Polarimetric Synthetic Aperture Radar: Advances and perspectives," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 79–89, Jul. 2014, doi: 10.1109/MSP.2014.2312099.
[56] "SAOCOM - Earth Online." https://earth.esa.int/eogateway/missions/saocom

(accessed Jan. 03, 2023).
[57] "SAOCOM data products - Earth Online." https://earth.esa.int/eogateway/catalog/saocom-data-products (accessed Jan. 03, 2023).
[58] "fb4628a8-ee31-2b31-4f59-ebb8a3f0e8b8.pdf." Accessed: Jan. 03, 2023. [Online]. Available: https://earth.esa.int/eogateway/documents/20142/37627/Technical+Note+on+Quality+Assessment+for+SAOCOM.pdf/fb4628a8-ee31-2b31-4f59-ebb8a3f0e8b8
[59] "SAOCOM," *ASI*. https://www.asi.it/en/earth-science/saocom/ (accessed Jan. 03, 2023).
[60] "Dataset | ALOS@EORC." https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm (accessed Jan. 21, 2023).
[61] J. S. Uva *et al.*, "Forestry Inventory 2015", doi: 10.15468/33hvm4.
[62] E. E. C. for M.-R. W. Forecast, E. E. F. A. System, and E. E. and G. D. Observatories, "EDO Soil Moisture Index (SMI) (version 2.1.1)," Apr. 2021, Accessed: Jan. 28, 2023. [Online]. Available: http://data.europa.eu/89h/a47c68f3-818f-4c41-92cc-1b94bdefcb6f
[63] "model LISFLOOD - LISFLOOD hydrological model." https://web.jrc.ec.europa.eu/policy-model-inventory/explore/models/model-lisflood (accessed Jan. 28, 2023).
[64] B. Bauer-Marschallinger *et al.*, "Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering," *Remote Sens.*, vol. 1019, no. 7, p. 1030, 2018, doi: 10.3390/rs10071030.
[65] "pandas - Python Data Analysis Library." https://pandas.pydata.org/ (accessed Jan. 29, 2023).
[66] "Feature importances with a forest of trees," *scikit-learn*. https://scikit-learn/stable/auto_examples/ensemble/plot_forest_importances.html (accessed Feb. 21, 2023).
[67] "What Is Backward Elimination Technique In Machine Learning? | Simplilearn," *Simplilearn.com*, Aug. 22, 2022. https://www.simplilearn.com/what-is-backward-elimination-technique-in-machine-learning-article (accessed Feb. 21, 2023).
[68] "scikit-learn: machine learning in Python — scikit-learn 1.2.1 documentation." https://scikit-learn.org/stable/index.html (accessed Jan. 29, 2023).
[69] "SciPy documentation — SciPy v1.10.0 Manual." https://docs.scipy.org/doc/scipy/index.html (accessed Jan. 30, 2023).
[70] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.*, vol. 2, no. 6, pp. 568–576, Nov. 1991, doi: 10.1109/72.97934.
[71] Q.-C. Cai, T.-H. Hsu, and J.-Y. Lin, "Using the General Regression Neural Network Method to Calibrate the Parameters of a Sub-Catchment," *Water*, vol. 13, no. 8, Art. no. 8, Jan. 2021, doi: 10.3390/w13081089.
[72] Y. Tramblay and P. Quintana Seguí, "Estimating soil moisture conditions for drought monitoring with random forests and a simple soil moisture accounting scheme," *Nat. Hazards Earth Syst. Sci.*, vol. 22, no. 4, pp. 1325–1334, Apr. 2022, doi: 10.5194/nhess-22-1325-2022.
[73] *Classification and Regression Trees*. Routledge, 2017. doi: 10.1201/9781315139470.
[74] H. Wang, R. Magagi, and K. Goita, "Comparison of different polarimetric decompositions for soil moisture retrieval over vegetation covered agricultural

area," *Remote Sens. Environ.*, vol. 199, pp. 120–136, Sep. 2017, doi: 10.1016/j.rse.2017.07.008.
[75] Q. Xie, Q. Meng, L. Zhang, C. Wang, Q. Wang, and S. Zhao, "Combining of the *H*/*A*/Alpha and Freeman–Durden Polarization Decomposition Methods for Soil Moisture Retrieval from Full-Polarization Radarsat-2 Data," *Adv. Meteorol.*, vol. 2018, p. e9436438, Aug. 2018, doi: 10.1155/2018/9436438.
[76] S. Dey, A. Bhattacharya, D. Ratha, D. Mandal, and A. C. Frery, "Target Characterization and Scattering Power Decomposition for Full and Compact Polarimetric SAR Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3981–3998, May 2021, doi: 10.1109/TGRS.2020.3010840.
[77] C. D. Richardson, "POLARIMETRIC SCATTERING MODEL FOR MULTILAYERED VEGETATION IN TROPICAL FOREST".
[78] Y. Kim and J. J. van Zyl, "A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2519–2527, Aug. 2009, doi: 10.1109/TGRS.2009.2014944.
[79] D. Ratha, D. Mandal, V. Kumar, H. Mcnairn, A. Bhattacharya, and A. C. Frery, "A Generalized Volume Scattering Model-Based Vegetation Index From Polarimetric SAR Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1791–1795, Nov. 2019, doi: 10.1109/LGRS.2019.2907703.
[80] K. O. Pope, J. M. Rey-Benayas, and J. F. Paris, "Radar remote sensing of forest and wetland ecosystems in the Central American tropics," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 205–219, May 1994, doi: 10.1016/0034-4257(94)90142-2.
[81] C. López-Martínez and E. Pottier, "Basic Principles of SAR Polarimetry," in *Polarimetric Synthetic Aperture Radar: Principles and Application*, I. Hajnsek and Y.-L. Desnos, Eds. Cham: Springer International Publishing, 2021, pp. 1–58. doi: 10.1007/978-3-030-56504-6_1.
[82] F. Zagolski, S. Foucher, and C. Gaillard, "Sensitivity Of SAR Speckle Filtering On The Assessment Of Surface Roughness And Soil Moisture Content," Mar. 2000.
[83] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data," *Remote Sens.*, vol. 7, no. 12, Art. no. 12, Dec. 2015, doi: 10.3390/rs71215841.

APPENDIX

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-0.01 (-0.10)	0.028 (0.028)	0.023 (0.023)	0.000 (0.004)
	Eucalyptus Forest	-0.01 (-0.01)	0.023 (0.021)	0.019 (0.017)	-0.002 (0.001)
	Other Forest	-0.04 (-0.03)	0.031 (0.033)	0.024 (0.026)	0.003 (0.002)
	All Other Land Uses	-0.02 (-0.04)	0.028 (0.027)	0.022 (0.022)	-0.002 (0.003)
September 1	**Entire Area**	-0.06 (-0.04)	0.028 (0.028)	0.024 (0.023)	0.003 (0.001)
	Eucalyptus Forest	-0.07 (-0.23)	0.028 (0.029)	0.024 (0.024)	0.003 (0.004)
	Other Forest	-0.12 (-0.08)	0.029 (0.029)	0.024 (0.024)	0.000 (0.002)
	All Other Land Uses	-0.09 (-0.05)	0.027 (0.027)	0.023 (0.023)	0.001 (0.001)
October 11	**Entire Area**	-0.03 (-0.01)	0.026 (0.026)	0.022 (0.022)	-0.001 (0.001)
	Eucalyptus Forest	-0.09 (-0.07)	0.023 (0.023)	0.020 (0.020)	0.002 (0.001)
	Other Forest	-0.12 (-0.11)	0.028 (0.027)	0.022 (0.023)	0.000 (0.002)
	All Other Land Uses	-0.04 (-0.02)	0.026 (0.025)	0.022 (0.021)	-0.003 (-0.001)

Soil Moisture Retrieval Accuracies Without Feature Selection using Random Forest (results with 5-fold cross-validation in parentheses) - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.01	0.027	0.022	-0.001
	Eucalyptus Forest	-0.01	0.023	0.019	-0.002
	Other Forest	-0.10	0.032	0.025	0.001
	All Other Land Uses	-0.04	0.028	0.022	-0.004
September 1	**Entire Area**	-0.01	0.027	0.023	0.002
	Eucalyptus Forest	0.00	0.027	0.023	0.001
	Other Forest	0.00	0.027	0.023	0.000
	All Other Land Uses	-0.03	0.026	0.022	0.000
October 11	**Entire Area**	0.02	0.025	0.021	-0.001
	Eucalyptus Forest	-0.03	0.023	0.019	0.000
	Other Forest	0.00	0.026	0.021	-0.002
	All Other Land Uses	-0.01	0.025	0.022	-0.003

Soil Moisture Retrieval Accuracies Without Feature Selection using GRNN - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-0.09	0.029	0.023	-0.001
	Eucalyptus Forest	-0.11	0.024	0.020	-0.001
	Other Forest	-0.19	0.033	0.026	0.002
	All Other Land Uses	-0.04	0.028	0.022	-0.002
September 1	**Entire Area**	-0.11	0.029	0.024	0.001
	Eucalyptus Forest	-0.13	0.029	0.024	0.002
	Other Forest	-0.24	0.031	0.025	0.002
	All Other Land Uses	-0.20	0.028	0.024	0.001
October 11	**Entire Area**	-0.17	0.027	0.023	0.000
	Eucalyptus Forest	-0.24	0.025	0.021	0.003
	Other Forest	-0.17	0.028	0.023	0.000
	All Other Land Uses	-0.18	0.027	0.023	-0.004

Random Forest: Soil Moisture Retrieval Accuracies With Top 3 Feature Selection - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-0.04	0.028	0.023	-0.001
	Eucalyptus Forest	-0.05	0.024	0.019	-0.002
	Other Forest	-0.05	0.031	0.024	0.002
	All Other Land Uses	-0.04	0.028	0.022	-0.003
September 1	**Entire Area**	-0.06	0.028	0.024	0.002
	Eucalyptus Forest	-0.07	0.028	0.023	0.002
	Other Forest	-0.10	0.029	0.023	0.002
	All Other Land Uses	-0.10	0.027	0.023	0.001
October 11	**Entire Area**	-0.04	0.026	0.022	-0.001
	Eucalyptus Forest	-0.16	0.024	0.020	0.002
	Other Forest	-0.12	0.027	0.022	0.000
	All Other Land Uses	-0.07	0.026	0.022	-0.003

Random Forest: Soil Moisture Retrieval Accuracies With Threshold (= 0.5) Feature Selection - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.01	0.027	0.022	-0.001
	Eucalyptus Forest	-0.02	0.023	0.019	-0.003
	Other Forest	-0.03	0.031	0.023	-0.001
	All Other Land Uses	-0.02	0.028	0.022	-0.004
September 1	**Entire Area**	0.00	0.027	0.023	0.001
	Eucalyptus Forest	0.00	0.027	0.023	0.001
	Other Forest	0.00	0.028	0.023	0.000
	All Other Land Uses	-0.04	0.026	0.022	0.001
October 11	**Entire Area**	0.02	0.025	0.021	-0.001
	Eucalyptus Forest	0.02	0.022	0.019	0.001
	Other Forest	0.00	0.026	0.021	-0.002
	All Other Land Uses	0.00	0.025	0.022	-0.003

GRNN: Soil Moisture Retrieval Accuracies With Forward Feature Selection - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.01	0.027	0.022	-0.001
	Eucalyptus Forest	-0.01	0.023	0.019	-0.002
	Other Forest	-0.10	0.032	0.025	0.001
	All Other Land Uses	-0.04	0.028	0.022	-0.004
September 1	**Entire Area**	-0.01	0.027	0.023	0.002
	Eucalyptus Forest	0.00	0.027	0.023	0.001
	Other Forest	0.00	0.027	0.023	0.000
	All Other Land Uses	-0.03	0.026	0.022	0.000
October 11	**Entire Area**	0.02	0.025	0.021	-0.001
	Eucalyptus Forest	-0.03	0.023	0.019	0.000
	Other Forest	0.00	0.026	0.021	-0.002
	All Other Land Uses	-0.01	0.025	0.022	-0.003

GRNN: Soil Moisture Retrieval Accuracies With Backward Feature Selection - SMI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.01 (-0.05)	0.179 (0.183)	0.148 (0.151)	-0.003 (-0.012)
	Eucalyptus Forest	-0.10 (-0.08)	0.165 (0.164)	0.132 (0.131)	-0.013 (0.001)
	Other Forest	0.08 (-0.05)	0.186 (0.202)	0.153 (0.168)	0.004 (-0.013)
	All Other Land Uses	-0.01(-0.09)	0.175(0.184)	0.151(0.155)	-0.034 (-0.008)
September 1	**Entire Area**	-0.02 (-0.01)	0.092 (0.093)	0.071 (0.072)	0.006 (0.002)
	Eucalyptus Forest	-0.11 (-0.07)	0.069 (0.065)	0.055 (0.052)	-0.003 (0.001)
	Other Forest	-0.14 (-0.08)	0.074 (0.078)	0.059 (0.060)	0.010 (0.003)
	All Other Land Uses	0.07 (0.01)	0.120 (0.119)	0.095 (0.096)	0.003 (0.013)
October 11	**Entire Area**	-0.03 (-0.02)	0.184 (0.184)	0.146 (0.150)	-0.001 (0.012)
	Eucalyptus Forest	-0.11 (-0.05)	0.158 (0.159)	0.133 (0.132)	-0.007 (-0.007)
	Other Forest	-0.04 (-0.07)	0.157 (0.156)	0.129 (0.127)	-0.012 (0.005)
	All Other Land Uses	0.04 (-0.06)	0.229 (0.221)	0.189 (0.185)	-0.028 (0.028)

Random Forest: Soil Moisture Retrieval Accuracies Without Feature Selection - SWI Raster
(Results with 5-fold cross-validation in parentheses) - SWI

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-0.18	0.195	0.160	-0.005
	Eucalyptus Forest	-0.15	0.169	0.138	-0.016
	Other Forest	-0.07	0.201	0.152	0.019
	All Other Land Uses	-0.02	0.177	0.154	-0.030
September 1	**Entire Area**	-0.08	0.095	0.073	0.005
	Eucalyptus Forest	-0.19	0.072	0.058	-0.003
	Other Forest	-0.27	0.079	0.061	0.006
	All Other Land Uses	-0.09	0.130	0.102	-0.001
October 11	**Entire Area**	-0.11	0.191	0.153	-0.008
	Eucalyptus Forest	-0.12	0.159	0.133	-0.003
	Other Forest	-0.25	0.173	0.145	-0.006
	All Other Land Uses	0.00	0.233	0.190	-0.030

Random Forest: Soil Moisture Retrieval Accuracies With Top 3 Feature Selection - SWI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.01	0.179	0.148	0.000
	Eucalyptus Forest	-0.14	0.168	0.132	-0.014
	Other Forest	0.09	0.185	0.150	0.009
	All Other Land Uses	0.03	0.172	0.150	-0.030
September 1	**Entire Area**	-0.04	0.093	0.072	0.005
	Eucalyptus Forest	-0.22	0.073	0.059	-0.002
	Other Forest	-0.16	0.075	0.060	0.010
	All Other Land Uses	-0.01	0.125	0.099	-0.001
October 11	**Entire Area**	-0.05	0.185	0.147	-0.004
	Eucalyptus Forest	-0.07	0.156	0.130	-0.004
	Other Forest	-0.06	0.159	0.130	-0.008
	All Other Land Uses	0.01	0.232	0.192	-0.026

Random Forest: Soil Moisture Retrieval Accuracies With Threshold (= 0.5) Feature Selection - SWI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-7.00	0.180	0.149	0.002
	Eucalyptus Forest	0.00	0.158	0.127	-0.004
	Other Forest	-0.01	0.195	0.155	0.016
	All Other Land Uses	-0.02	0.176	0.154	-0.032
September 1	**Entire Area**	0.01	0.090	0.069	0.002
	Eucalyptus Forest	-0.11	0.070	0.056	-0.005
	Other Forest	0.00	0.070	0.054	0.002
	All Other Land Uses	0.06	0.121	0.095	-0.006
October 11	**Entire Area**	0.01	0.180	0.143	-0.010
	Eucalyptus Forest	-0.07	0.155	0.133	-0.015
	Other Forest	-0.01	0.155	0.127	-0.012
	All Other Land Uses	0.00	0.234	0.191	-0.036

GRNN: Soil Moisture Retrieval Accuracies Without Feature Selection - SWI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	0.00	0.180	0.149	0.002
	Eucalyptus Forest	0.00	0.158	0.128	-0.006
	Other Forest	0.01	0.195	0.155	0.015
	All Other Land Uses	-0.02	0.176	0.154	-0.031
September 1	**Entire Area**	0.00	0.091	0.070	0.004
	Eucalyptus Forest	-0.02	0.067	0.053	-0.004
	Other Forest	0.01	0.069	0.053	0.004
	All Other Land Uses	0.07	0.120	0.093	-0.008
October 11	**Entire Area**	0.00	0.181	0.144	-0.008
	Eucalyptus Forest	0.00	0.150	0.129	-0.010
	Other Forest	-0.02	0.155	0.126	-0.016
	All Other Land Uses	0.00	0.233	0.192	-0.035

GRNN: Soil Moisture Retrieval Accuracies With Forward Feature Selection - SWI Raster

		R^2	RMSE	MAE	MBE
June 21	**Entire Area**	-7.00	0.180	0.149	0.002
	Eucalyptus Forest	0.00	0.158	0.127	-0.004
	Other Forest	-0.01	0.195	0.155	0.016
	All Other Land Uses	-0.02	0.176	0.154	-0.032
September 1	**Entire Area**	0.01	0.090	0.069	0.002
	Eucalyptus Forest	-0.11	0.070	0.056	-0.005
	Other Forest	0.00	0.070	0.053	0.002
	All Other Land Uses	0.06	0.121	0.095	-0.006
October 11	**Entire Area**	0.01	0.180	0.143	-0.010
	Eucalyptus Forest	-0.06	0.155	0.133	-0.015
	Other Forest	-0.01	0.155	0.127	-0.012
	All Other Land Uses	0.00	0.234	0.191	-0.036

GRNN: Soil Moisture Retrieval Accuracies With Backward Feature Selection - SWI