DSpace Institution

DSpace Repository **http://dspace.org**

Geography and Environmental Studies Thesis

2025-06

A Machine Learning Based Approach to Estimate Surface Soil Moisture Using Optical and Microwave Remote Sensing Data for Crop Monitoring In Koga Watershed, Ethiopia.

Amare, Dessalew

http://ir.bdu.edu.et/handle/123456789/17093

Downloaded from DSpace Repository, DSpace Institution's institutional repository

BAHIR DAR UNIVERSITY

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Geo-information science program

A MACHINE LEARNING BASED APPROACH TO ESTIMATE SURFACE SOIL MOISTURE USING OPTICAL AND MICROWAVE REMOTE SENSING DATA FOR CROP MONITORING IN KOGA WATERSHED, ETHIOPIA.

By: Amare Dessalew

Advisor: Daniel Ayalew (Associate Professor)

June, 2025

Bahir Dar, Ethiopia

BAHIR DAR UNIVERSITY

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Master of Science in Geo-information

Research Thesis on

A MACHINE LEARNING APPROACH TO ESTIMATE SURFACE SOIL MOISTURE BASED ON OPTICAL AND MICROWAVE REMOTE SENSING PRODUCTS FOR CROP MONITORING IN KOGA WATERSHED, ETHIOPIA.

By: Amare Dessalew

A Thesis Submitted to Department of Geography and Environmental Studies, Bahir Dar University, In Partial Fulfillment of the Requirements for Degree of Masters of Science in Geo-Information

Advisor: Daniel Ayalew (Associate Professor)

© 2025 Amare Dessalew

June, 2025

Bahir Dar, Ethiopia

Declaration

I hereby certify that I completed the master's thesis, ““**A MACHINE LEARNING APPROACH TO ESTIMATE SURFACE SOIL MOISTURE BASED ON OPTICAL AND MICROWAVE REMOTE SENSING PRODUCTS FOR CROP MONITORING IN KOGA WATERSHED, ETHIOPIA**" under the guidance of Daniel Ayalew (Associate Professor), Department of Geography and Environmental Studies, Bahir Dar University. I further affirm that all sources of information used in the thesis have been properly acknowledged, and this thesis is my own original work. It has not been submitted to any other university or institution for the award of a degree or diploma.

Amare Dessalew	Date	Place

© 2025 Amare Dessalew

BAHIR DAR UNIVERSITY

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Geo-information Science

Approval of Dissertation/thesis for defense

As members of the board of examiners, we examined this thesis entitled **"A MACHINE LEARNING APPROACH TO ESTIMATE SURFACE SOIL MOISTURE BASED ON OPTICAL AND MICROWAVE REMOTE SENSING PRODUCTS FOR CROP MONITORING IN KOGA WATERSHED, ETHIOPIA**"by Amare Dessalew. We hereby certify that the thesis is accepted for fulfilling the requirements for "MSc in Geo-information Sciences"

_____	_____	_____
Advisor's name	Signature	Date

BAHIR DAR UNIVERSITY

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Master of Science in Geo-information

Approval of Dissertation/thesis for defense result

As members of the board of examiners, we examined this thesis entitled **"A MACHINE LEARNING APPROACH TO ESTIMATE SURFACE SOIL MOISTURE BASED ON OPTICAL AND MICROWAVE REMOTE SENSING PRODUCTS FOR CROP MONITORING IN KOGA WATERSHED, ETHIOPIA** "by Amare Dessalew. We hereby certify that the thesis is accepted for fulfilling the requirements for "MSc in Geo-information Sciences"

Board of Examiners

_____	_____	_____
External examiner's name	Signature	Date
_____	_____	_____
Internal examiner's name	Signature	Date
_____	_____	_____
Chair person's name	Signature	Date

Table of Contents

Abstract x

CHAPTER ONE 1

1. INTRODUCTION 1

1.1. Background of the study 1

1.2. Statement of the problem 3

1.3. Objective 3

1.3.1. General Objective 6

1.3.2. Specific Objective 6

1.4. Research questions 6

1.5. Significance of the study 7

1.6. Scope of the study 7

CHAPTER TWO 8

2. RELATED LITERATURE REVIEW 8

2.1. Role of Remote Sensing in Soil Moisture Estimation 8

2.2. Machine Learning Models for Soil Moisture Estimation 13

2.3. Validation of Soil Moisture Estimation Models 17

2.4. Crop Planning and Soil Moisture 8

2.5. Empirical literature 20

2.6. Knowledge gaps 21

CHAPTER THREE 23

3. MATERIALS AND METHODS 23

3.1. The Study Area 23

3.1.1. Location 23

3.2. Data source and type 25

3.2.1. Satellite Data 25

3.3. Field data collection 26
3.4. Soil sample collection 28
3.4.1. Gravimetric method 28
3.5. Method of data analysis 29
3.6. Derivation of satellite predictor variables 32
3.6.1. Sentinel 1A predictor variables 32
3.6.2. Sentinel 2 MSI predictor variables 33
3.7. Machine learning algorithms 34
3.7.1. Random Forest (RF) algorithm 34
CHAPTER FOUR 36
4. RESULT AND DISCUSSIONS 36
4.1. Introduction ..... **Error! Bookmark not defined.**
4.2. Sentinel 1A SAR predictor variables 36
4.2.1. Vertical Release-Vertical Receive (VV) backscattering coefficients 36
4.2.2. Vertical Release-Horizontal Receive (VV) backscattering coefficients 37
4.2.3. Surface Roughness 39
4.2.4. Polarization Difference (dB) 40
4.3. Sentinel 2 MSI derived variables 42
4.3.1. NDVI 42
4.3.2. NDWI 44
4.3.3. SAVI 45
4.3.4. MNDWI 47
4.3.5. NDRE 48
4.3.6. NDTI 50
4.4. Random Forest (RF) model performance evaluation 51
4.4.1. Variable importance 51
4.4.2. Scatter plot between observed and Predicted Soil Moisture (SM) 53

4.4.3. Predictor correlation 55
4.5. Soil Moisture (SM) prediction 56
4.6. Validations 62
4.7. Discussions ..... **Error! Bookmark not defined.**
CHAPTER FIVE 66
5. CONCLUSIONS AND RECOMMENDATIONS 66
5.1. Conclusions 66
5.2. Recommendations 67

List of Figures

Figure 3-1: Location map of the study area 23
Figure 3-2: Sample stratification 27
Figure 3-3: Methodological flow chart 31
Figure 4-1: VV-Backscattering coefficients (dB) 37
Figure 4-2: VH-Backscattering coefficients (dB) 38
Figure 4-3: Surface roughness 40
Figure 4-4: Polarization difference 42
Figure 4-5: NDVI 43
Figure 4-6: NDWI 45
Figure 4-7: SAVI 46
Figure 4-8: MNDWI 48
Figure 4-9: NDRE 49
Figure 4-10: NDTI 50
Figure 4-11: Variable importance in terms of MSE and NodePurity Index; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction 52
Figure 4-12: Scatter plot; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction 54
Figure 4-13: Predictor correlation; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction 56
Figure 4-14: Soil Moisture Prediction for the Month of February, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1 58
Figure 4-15: Soil Moisture Prediction for the Month of March, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1 60
Figure 4-16: Soil Moisture Prediction for the Month of April, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1 61
Figure 4-17: Validation of predicted SM over Sit1 (a) Bias, (b) RMSE, (UbRMSE) 63
Figure 4-18: Validation of predicted SM over Sit2 (a) Bias, (b) RMSE, (UbRMSE) 64
Figure 4-19: Scatter plot between predicted and In-situ soil moisture 65

Abstract

Surface soil moisture (SSM) plays a pivotal role in agriculture, hydrology, and environmental monitoring, yet conventional measurement methods are limited in spatial coverage and operational efficiency. This study aims to estimate high-resolution surface soil moisture in the Koga Watershed, Ethiopia, by integrating optical and microwave remote sensing products with machine learning techniques for improved crop planning and water management. The research utilizes Sentinel-1A Synthetic Aperture Radar (SAR) for backscatter coefficients and surface roughness, and Sentinel-2 MSI imagery to derive vegetation and moisture indices such as NDVI, SAVI, NDWI, MNDWI, NDRE, and NDTI. Prior to modeling, extensive preprocessing, including calibration, multi-looking, and atmospheric correction of satellite data, was conducted. Field soil samples were collected near the Koga irrigation scheme and analyzed using the gravimetric method to obtain reference soil moisture values. These ground-truth data were synchronized with satellite overpass times to ensure temporal consistency. A Random Forest (RF) model was trained using a variety of predictor variables from both SAR and optical datasets, and their performance was evaluated through statistical metrics such as R^2, RMSE, and MAE. The result of the study revealed that combining SAR and Sentinel-2 MSI data enabled accurate, high-resolution surface soil moisture prediction in the Koga watershed, with values ranging from 30.01% to 52.07% across February to April 2025. VV and VH backscatter, surface roughness, and vegetation indices like NDVI and NDWI were key predictors, with SAR variables more influential in drier months and optical indices gaining importance as vegetation increased. The Random Forest model achieved the highest accuracy ($R^2 = 0.94$), and effectively capturing the spatial patterns such as higher soil moisture in low-lying, vegetated, or irrigated areas, confirming the model's reliability for seasonal crop monitoring and water management. The study concludes that integrating multi-sensor data significantly enhances soil moisture estimation accuracy and provides timely, spatially explicit information essential for crop monitoring and water resource management. It is recommended to adopt seasonally adaptive modeling strategies, expand in-situ validation networks, and incorporate surface roughness and land use data to further improve model robustness and decision support in agricultural planning.

Keywords: Surface soil moisture, Sentinel-1A, Sentinel-2, Random Forest.

CHAPTER ONE

1. INTRODUCTION

1.1.Background of the study

Surface soil moisture (SSM) serves as a key link between water and energy exchanges between the land surface and atmosphere. It is also a crucial component for many land-surface applications, including agronomy, ecology, hydrology, meteorology, and climatology, as well as the forecasting of droughts and floods (Sekertekin et al., 2020). However, it is difficult to retrieve accurate soil moisture data from in-situ measurements taken across vast areas due to the substantial temporal variability of SSM over space. Point-based ground measurements, while precise locally, are often time-consuming and ineffective for large-scale monitoring. As a result, from local to global scales, precise spatial SSM retrieval is critical (Şekertekin et al., 2018). Moreover, it may be accomplished via space-based technologies (Sadeghi et al., 2017).

A key soil indicator for defining and identifying agricultural drought is soil moisture. Identifying early-stage water deficiency circumstances and evolving drought events for crop yield uncertainty and food security conditions, agricultural insurance, policymaking and decision-making, and crop planning are all applications of soil moisture estimation. (Keesstra et al., 2016), particularly for the dry and semi-dry regions of the world. In developing countries, social and political tensions are exacerbated by agricultural droughts (Kelley et al., 2015). Soil moisture modeling and monitoring are becoming more popular. In order to mitigate and adapt to climate change for the sustainability of agricultural systems as well as to promote precision agriculture and food security, it is necessary to monitor the regional and temporal variations of SM (Zareie et al., 2016).

Surface soil moisture is generally referred to the water content of the top ~5–15 cm of the soil layer (Wang & Qu, 2009). Although though it only makes up a tiny percentage of the world's water reserves, the water in this thin layer of soil represents key hydrological, biochemical, physiological, agricultural, and other Earth activities (Primi et al., 2018). For the creation of policies and land-water management plans, precise data on the geographical and temporal fluctuations of surface soil moisture are essential.

An essential consideration in maintaining irrigated farms is the soil moisture content (SMC). SSM, which is contained in the top 10 cm of soil, and root zone soil moisture are the two main

components of SMC (held in the upper 200 cm of soil). Surface soil moisture can be calculated using a variety of methods, including in situ measurements, physically based models, remote sensing, etc., and is a crucial factor in addressing energy and water fluxes at the land surface/atmosphere interface. Grayson and Western applied field (or in situ) measurements, remote sensing methods, and soil water balance simulation models to the assessment of soil moisture (Liu et al., 2010) In terms of the overall global water balance, soil moisture makes up a fairly small volume, yet it is crucial for the management and planning of water resources (Liu et al., 2010). Growers, production managers, and crop consultants currently base a lot of their crop production management decisions in precision agriculture on observation from remotely sensed data, such as satellite images.

Traditional approaches for measuring soil moisture, such as gravimetric sampling, time domain reflectometry (TDR), neutron probes, and gamma ray scanning, provide reliable point-based data but are often tedious, costly, and labor-intensive (Wang & Qu, 2009; Yin et al., 2016). Furthermore, these methods lack the spatial and temporal coverage required to capture soil moisture variability over larger heterogeneous landscapes. To overcome these limitations, remote sensing technologies and advanced modeling approaches have been increasingly employed to estimate SM at regional to global scales.

Machine learning (ML) algorithms have emerged as powerful tools for soil moisture prediction. Unlike traditional empirical or physically-based models, ML techniques can capture nonlinear relationships and exploit a wide range of predictor variables derived from synthetic aperture radar (SAR) and multispectral sensors (Zhang et al., 2019). Algorithms such as Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Machines (GBM), and Artificial Neural Networks (ANNs) have been successfully applied to improve soil moisture retrieval by integrating multi-source data and reducing uncertainties (Srivastava et al., 2021; Qu et al., 2022). Remote sensing technology has shown a lot of promise for quickly and accurately estimation SM over wide areas (Primi et al., 2018). Using a variety of satellite data, including Sentinel l/2 imagery, SM has been estimated at various scales in the visible, near-infrared (NIR), thermal infrared, and microwave regions of the electromagnetic spectrum.

1.2.Statement of the problem

Soil moisture is a critical component of agricultural productivity and sustainability Because it influences numerous physiological processes in plants, including seed germination, nutrient uptake, and growth (Bhattacharya, 2021; Gavrilescu, 2021; Waraich et al., 2011). Adequate soil moisture ensures that plants can obtain the water required for photosynthesis, which is essential for crop quality and yield Ali and Talukder (2008). Plants can effectively absorb nutrients when soil moisture levels are at their optimal levels, resulting in crops that are healthier and more productive (Fageria et al., 2008). On the other hand, excess and deficiency of water can have a negative impact on crop performance. Drought can cause water stress and slower growth, while soggy soils can impair plant development and cause root diseases (Blanco-Canqui & Ruis, 2018).

Furthermore, soil health and structure are greatly influenced by soil moisture. It has an impact on soil temperature, aeration, and microbial activity—all of which are critical for preserving soil fertility (Bodner et al., 2015). Maintaining the equilibrium of soil particles and organic matter through proper soil moisture levels supports the physical structure of the soil, reducing erosion and improving water infiltration (Chukalla et al., 2017). Therefore, understanding and managing soil moisture is essential for optimizing crop growth, preserving soil integrity, and ensuring sustainable agricultural practices.

Traditional techniques for soil moisture measurement, such as time domain reflectometry (TDR) (Černý, 2009), neutron scattering (Bogena et al., 2015), and gravimetric sampling (Singh & Singh, 2017), remain effective for point-scale observations but face critical limitations in contemporary agriculture. Gravimetric sampling, though highly accurate, is labor-intensive and time-consuming, requiring physical collection and drying of soil samples (Vereecken et al., 2014). Similarly, TDR and neutron probes provide precise measurements but are inherently restricted to localized observations, making them unsuitable for capturing spatial variability across heterogeneous landscapes without extensive sensor deployment or manual sampling (Romano, 2014).

In contrast to traditional in-situ methods, active microwave remote sensing particularly Synthetic Aperture Radar (SAR) offers an effective technique for estimating surface soil moisture (SSM). SAR sensors transmit microwave pulses and record the backscattered signal, which is sensitive to surface roughness and dielectric properties (Narvekar et al., 2015). Various empirical, semi-

empirical, and theoretical models, such as the Water Cloud Model (WCM) and the Dubois model, have been applied to extract soil moisture information from SAR backscatter (Baghdadi et al., 2006; Bazzi et al., 2019). For example, Liu (2019) and Yang et al. (2019) improved soil moisture estimation using C-band Radarsat-2 SAR data with additional polarimetric features, while Hajj et al. (2017) demonstrated that integrating L-band PALSAR/ALOS-2 with optical data enhances accuracy through neural network models.

Despite these advances, SAR-based estimation remains challenged by the complex interactions between surface factors and radar parameters such as incidence angle, frequency, and polarization (Alexakis et al., 2011; Bai et al., 2019). Surface roughness, a key variable, is particularly difficult to measure consistently, and models such as the Integral Equation Model (IEM) often require calibration with in situ data (Loew et al., 2006; Oh, 2004a). Although Sentinel-1A SAR has shown strong potential—achieving an R^2 of 0.75 with ground observations in central India (Singh et al., 2019)—its broader applicability across diverse regions remains uncertain. For instance, studies in Ethiopia's Upper Blue Nile Basin (Ayehu et al., 2018) used Sentinel-1A with Landsat imagery, but systematic evaluation of dual-polarized SAR from multiple radar bands, such as the deeper-penetrating L-band (ALOS PALSAR) and S-band (NISAR), has yet to be undertaken. This gap highlights the need for region-specific investigations into the comparative performance of different SAR bands for soil moisture estimation.

Optical remote sensing, particularly data from Sentinel-2's Multispectral Instrument (MSI), provides valuable indirect indicators of soil moisture through surface reflectance and vegetation condition metrics (Varghese et al., 2021; Liu et al., 2021). Sentinel-2 MSI delivers high-resolution imagery (10 m) across 13 spectral bands, enabling the derivation of vegetation and water indices such as the Normalized Difference Vegetation Index (NDVI) (Du et al., 2016), the Normalized Difference Water Index (NDWI) (Yang et al., 2017), and the Soil Adjusted Vegetation Index (SAVI) (Morell-Monzó et al., 2020). These indices reflect vegetation vigor and canopy water stress, both of which are closely influenced by subsurface soil moisture availability (Gutiérrez-Gordillo et al., 2021). For example, reduced NDVI or elevated surface temperature may indicate soil water deficits, whereas higher NDWI values typically reflect wetter conditions.

The integration of microwave and optical remote sensing offers a synergistic approach to soil moisture estimation by combining the complementary strengths of both sensing modalities. Optical data from Sentinel-2 MSI provide high-resolution spatial and spectral information on vegetation health, which is indirectly linked to soil moisture, while microwave SAR offers direct sensitivity to soil dielectric properties, even under cloud cover and vegetation (Hajj et al., 2017; Gao et al., 2018). When SAR backscatter coefficients are integrated with vegetation indices such as NDVI or NDWI, soil moisture models can more effectively capture both soil and vegetation dynamics. Machine learning techniques, including neural networks and random forests, have been successfully applied to fuse these datasets, thereby improving prediction accuracy and addressing observational gaps (Yang et al., 2019). For instance, Hajj et al. (2017) demonstrated that combining ALOS-2 SAR with optical data significantly enhanced soil moisture retrievals in tropical environments.

Despite these advances, the combined use of optical and SAR data for soil moisture estimation remains underutilized in Ethiopia's Koga Watershed. This region supports intensive irrigation-based agriculture, where accurate and timely soil moisture information is vital for optimizing water use efficiency and improving crop yield forecasts. Although Sentinel-1A SAR and other datasets have been employed in regional soil moisture assessments, limited research has explored their integration with high-resolution optical observations. This lack of multi-sensor fusion restricts the comprehensive monitoring of soil–plant–water interactions in the watershed. Addressing this gap through the combined application of SAR and optical data, enhanced by machine learning, could significantly advance irrigation planning, precision agriculture, and sustainable land management in the Koga Watershed.

1.3.Objective

1.3.1. General Objective

The General objective of this study is to estimate surface soil moisture in the Koga catchment, Lake Tana sub-basin using optical and microwave remote sensing data, and machine learning approach for crop planning.

1.3.2. Specific Objective

This study answered the following specific objectives;

- To evaluate predictor variables derived from Sentinel-1A SAR and Sentinel-2
- To estimate surface soil moisture using Sentinel-1A SAR and Sentinel-2 datasets.
- To assess and validate Machine Learning Models for accurate soil moisture estimation.

1.4.Research questions

- Which predictor variables from Sentinel-1A SAR and Sentinel-2 proved most relevant for soil moisture estimation, and what underlying physical processes explain their importance?
- How did the integration of Sentinel-1A SAR and Sentinel-2 datasets improve the accuracy of surface soil moisture estimation compared to using either dataset alone?
- Which machine learning model achieved the highest performance in predicting soil moisture, and what factors contributed to its superior accuracy?

1.5.Significance of the study

This study holds significant value for agricultural management and water resource planning in the Koga watershed. The study aims to improve the accuracy of soil moisture assessments, which are essential for efficient crop planning and irrigation management, by estimating surface soil moisture with high spatial resolution using Sentinel-1A SAR and Sentinel-2 data. Local farmers and agricultural productivity can directly benefit from accurate soil moisture data since it can help with crop yield forecasting, irrigation schedule optimization, and better allocation of water resources. Additionally, the creation and verification of machine learning models for estimating soil moisture marks a breakthrough in the combination of remote sensing technology and sophisticated data analytics. In order to ensure sustainable agricultural practices and adapt to climate variability, these models can deliver accurate and timely soil moisture information. Efficient processing and analysis of vast amounts of remote sensing data can also help with better drought mitigation and water management decision-making. This study is crucial for land use planners because it helps them choose crops based on local seasonal conditions and plan crops based on soil moisture content. In the end, this information can support the long-term sustainability and resilience of agriculture in the Koga catchment by directing land management techniques, irrigation plans, and crop selection.

1.6.Scope of the study

This study's focus was on using machine learning techniques to estimate surface soil moisture in the Koga Watershed in Ethiopia using a combination of optical and microwave remote sensing products. The study developed predictive models for estimating soil moisture using satellite-derived data, including optical imagery (e.g., Sentinel-2) and microwave observations (e.g., Sentinel-1). To guarantee accuracy and dependability, these models were verified using soil moisture measurements taken on the ground. In order to facilitate crop monitoring, the study specifically targeted agricultural areas within the watershed.

CHAPTER TWO

2. RELATED LITERATURE REVIEW

2.1.Crop Planning and Soil Moisture

Soil moisture is a critical factor influencing crop growth, development, and yield, as it directly affects water availability for plant uptake, nutrient transport, and root respiration (Lynch et al., 2012). Adequate soil moisture levels are essential for optimal photosynthesis, biomass production, and stress tolerance, while insufficient or excessive moisture can lead to water stress, nutrient deficiencies, and reduced yields (D. Liu et al., 2010). Soil moisture also plays a key role in determining the timing of key phenological stages, such as germination, flowering, and fruiting, which are critical for yield prediction. For example, studies have shown that soil moisture deficits during critical growth stages, such as flowering and grain filling, can significantly reduce crop yields in cereals like wheat and maize (Lobell et al., 2015). Accurate prediction of soil moisture dynamics is therefore essential for understanding crop-water relationships and improving yield forecasts.

In addition to its direct effects on crop growth, soil moisture influences other environmental factors, such as soil temperature, aeration, and microbial activity, which further impact crop performance (Turmel, 2014). For instance, optimal soil moisture levels enhance microbial activity and nutrient mineralization, improving nutrient availability for crops (Seneviratne et al., 2010). Conversely, waterlogging or drought can disrupt these processes, leading to nutrient imbalances and reduced crop productivity. By integrating soil moisture data into crop models, researchers and farmers can better predict crop responses to varying water conditions and develop strategies to mitigate the impacts of water stress. This is particularly important in the context of climate change, where increasing variability in precipitation patterns is expected to exacerbate soil moisture fluctuations and their effects on crop yields (Thornton, 2014).

The integration of soil moisture data into crop planning and irrigation scheduling is a key strategy for optimizing water use efficiency and improving agricultural productivity (Bwambale, 2021). Soil moisture data provide valuable insights into the water status of the soil, enabling farmers to make informed decisions about when and how much to irrigate. For example, real-time soil moisture monitoring systems, such as those based on in-situ sensors or remote sensing, can help

farmers identify periods of water stress and adjust irrigation schedules accordingly (Evans & Sadler, 2008). This not only ensures that crops receive adequate water but also minimizes water wastage and reduces the risk of over-irrigation, which can lead to waterlogging and nutrient leaching.

Advanced technologies, such as precision agriculture and decision support systems, are increasingly being used to integrate soil moisture data into crop management practices (Phillips, 2014). These systems combine soil moisture measurements with weather forecasts, crop models, and irrigation algorithms to provide tailored recommendations for irrigation scheduling and crop planning (Hedley & Yule, 2009). For instance, the use of soil moisture data in conjunction with crop water requirement models, such as the FAO Penman-Monteith equation, allows for precise estimation of irrigation needs based on crop type, growth stage, and environmental conditions. By optimizing irrigation practices, farmers can improve water use efficiency, reduce production costs, and enhance crop yields, particularly in water-scarce regions.

Spatiotemporal variations in soil moisture, driven by factors such as topography, soil properties, vegetation cover, and climate, have significant implications for agricultural decision-making (Baroni, 2013). These variations influence the spatial distribution of water availability within fields, affecting crop growth patterns and yield variability (Bruno, 2006). For example, areas with low soil moisture due to poor water retention or high evaporation rates may require more frequent irrigation, while areas with high soil moisture may be prone to waterlogging and require drainage measures (Romano et al., 2002). Understanding these variations is essential for site-specific crop management, as it enables farmers to tailor irrigation and fertilization practices to the specific needs of different areas within a field.

Temporal variations in soil moisture, such as seasonal fluctuations and short-term changes due to rainfall or irrigation, also play a critical role in agricultural decision-making (Guido, 2020). For instance, accurate prediction of soil moisture dynamics during the growing season can help farmers anticipate periods of water stress and take proactive measures, such as adjusting irrigation schedules or applying drought-tolerant crop varieties (Daly et al., 2000). Additionally, long-term monitoring of soil moisture trends can provide insights into the impacts of climate change on water availability and inform adaptive strategies, such as crop rotation, conservation tillage, and the use of cover crops. By accounting for spatiotemporal variations in soil moisture, farmers can make

more informed decisions that enhance resilience to climate variability and improve agricultural sustainability.

Several watershed-level studies have demonstrated the importance of soil moisture data for crop planning and water management. For example, a study in the Yaqui Valley, Mexico, used soil moisture data from remote sensing and in-situ measurements to optimize irrigation scheduling for wheat production (Lobell et al., 2003). The study found that integrating soil moisture data into irrigation management reduced water use by 20% while maintaining crop yields, highlighting the potential for water savings in irrigated agriculture. Similarly, a study in the Indo-Gangetic Plain, India, used soil moisture data from the Soil Moisture Active Passive (SMAP) mission to assess the impacts of soil moisture variability on rice and wheat yields (Mishra et al., 2013). The results showed that soil moisture deficits during critical growth stages significantly reduced yields, underscoring the need for improved irrigation planning in the region.

In the United States, the USDA Agricultural Research Service (ARS) has conducted extensive research on the use of soil moisture data for crop planning in the Little Washita River watershed, Oklahoma. Studies in this watershed have demonstrated the value of integrating soil moisture data from remote sensing and in-situ networks into decision support systems for irrigation management (Cosh et al., 2008). These systems have been used to optimize irrigation scheduling for crops such as cotton and sorghum, resulting in improved water use efficiency and crop yields. Similarly, in the Murray-Darling Basin, Australia, soil moisture data have been used to develop water allocation strategies for irrigated agriculture, helping to balance water use between crops, ecosystems, and urban areas (Van Dijk et al., 2013). These case studies illustrate the potential of soil moisture data to enhance crop planning and water management at the watershed scale.

2.2.Role of Remote Sensing in Soil Moisture Estimation

Remote sensing has become an indispensable tool for monitoring soil moisture, offering spatially extensive and temporally consistent data that is critical for applications in agriculture, hydrology, climate modeling, and disaster management (Kornelsen & Coulibaly, 2013). Soil moisture, defined as the water content in the soil, is a key variable in the Earth's water cycle, influencing processes such as evaporation, infiltration, and runoff (Seneviratne et al., 2010). Traditional methods of soil moisture measurement, such as gravimetric sampling and in-situ sensors, are limited in spatial coverage and labor-intensive (Babaeian et al., 2021). Remote sensing, on the other hand, provides

a non-invasive, scalable, and efficient approach to estimate soil moisture over large areas (Yin et al., 2024). Among the various remote sensing technologies, Synthetic Aperture Radar (SAR) has gained prominence due to its ability to operate in all weather conditions and its sensitivity to soil moisture variations (Bousbih et al., 2018). Sentinel-1A, a SAR mission under the European Union's Copernicus program, has become a cornerstone in soil moisture estimation due to its high spatial resolution, frequent revisit time, and open data policy (Bauer-Marschallinger et al., 2019).

Remote sensing technologies for soil moisture monitoring can be broadly categorized into optical, thermal, and microwave-based methods (Petropoulos et al., 2015). Optical and thermal remote sensing rely on the reflectance and emission properties of the land surface, which are indirectly related to soil moisture (Li et al., 2021). For instance, vegetation indices derived from optical sensors, such as the Normalized Difference Vegetation Index (NDVI), can provide insights into soil moisture conditions, especially in vegetated areas (Babaeian et al., 2019). Thermal sensors measure land surface temperature (LST), which is influenced by soil moisture through its effect on evaporative cooling (Hulley & Hook, 2009). However, these methods are limited by cloud cover and are less effective in densely vegetated or cloudy regions (Peng et al., 2021).

Microwave remote sensing, particularly SAR, overcomes these limitations by operating at longer wavelengths that can penetrate clouds and, to some extent, vegetation (Bousbih et al., 2018). SAR systems emit microwave signals and measure the backscattered energy from the Earth's surface, which is influenced by soil dielectric properties, surface roughness, and vegetation structure (Baghdadi et al., 2012; Hajj et al., 2017). The dielectric constant of soil, which is highly dependent on water content, makes SAR particularly sensitive to soil moisture variations (Bousbih et al., 2018). Passive microwave sensors, such as those on the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) missions, also provide valuable soil moisture data but at coarser spatial resolutions compared to SAR (Kornelsen & Coulibaly, 2013). Sentinel-1A, with its C-band SAR, strikes a balance between spatial resolution and sensitivity to soil moisture, making it a preferred choice for regional and field-scale applications (Bauer-Marschallinger et al., 2019).

SAR systems, such as those on Sentinel-1A, operate by transmitting microwave pulses toward the Earth's surface and recording the backscattered signals. The "synthetic aperture" refers to the technique of combining multiple radar echoes to simulate a longer antenna, thereby achieving high

spatial resolution (Torres et al., 2012). Sentinel-1A operates in the C-band (5.405 GHz), which provides a good compromise between sensitivity to soil moisture and penetration through vegetation (Monteiro). The satellite acquires data in four imaging modes: Interferometric Wide Swath (IW), Extra Wide Swath (EW), Strip Map (SM), and Wave (WV). The IW mode, with a spatial resolution of 5x20 meters and a swath width of 250 kilometers, is most commonly used for soil moisture estimation due to its balance between coverage and detail (Bauer-Marschallinger et al., 2019).

The sensitivity of SAR to soil moisture arises from the dependence of the backscattered signal on the dielectric properties of the soil (El Hajj et al., 2016). The dielectric constant of dry soil is typically around 3-5, while that of water is approximately 80 (Zribi et al., 2019). As soil moisture increases, the dielectric constant rises, leading to higher backscatter coefficients (Y. Y. Li et al., 2014). Sentinel-1A's C-band SAR is particularly effective in detecting surface soil moisture (top 5 cm of the soil layer) due to its wavelength and incidence angle range (20°-45°) (X. Yang et al., 2024). However, the backscatter signal is also influenced by surface roughness and vegetation cover, which can complicate soil moisture retrieval (Karthikeyan et al., 2017). Advanced algorithms, such as those based on machine learning or physical models, are often employed to disentangle these effects and improve estimation accuracy (Xu et al., 2023).

The backscatter coefficient (σ°), measured in decibels (dB), is the primary observable used in SAR-based soil moisture estimation (Bousbih et al., 2019). It represents the intensity of the radar signal reflected back to the sensor and is influenced by soil moisture, surface roughness, and vegetation (El Hajj et al., 2019). Empirical, semi-empirical, and physical models are used to relate σ° to soil moisture (Kornelsen & Coulibaly, 2020). Empirical models, such as the change detection approach, exploit temporal variations in backscatter to estimate soil moisture changes (Bauer-Marschallinger et al., 2018). Semi-empirical models, like the Water Cloud Model (WCM), incorporate vegetation effects by treating the canopy as a layer that attenuates and scatters the radar signal (Bousbih et al., 2019). Physical models, such as the Integral Equation Model (IEM), simulate the interaction of microwaves with the soil surface based on electromagnetic theory (El Hajj et al., 2019).

Sentinel-2, part of the European Space Agency's (ESA) Copernicus program, is a high-resolution optical and multispectral imaging mission designed to monitor Earth's land surfaces (Phiri et al.,

2020). It consists of two satellites, Sentinel-2A and Sentinel-2B, which provide systematic coverage of the Earth's landmasses, coastal zones, and inland waters (Sudmanns et al., 2020). Sentinel-2 captures data in 13 spectral bands, ranging from visible and near-infrared (VNIR) to shortwave infrared (SWIR), with spatial resolutions of 10, 20, and 60 meters depending on the band. This multispectral capability makes Sentinel-2 particularly valuable for vegetation and soil analysis (da Silveira et al., 2018). For instance, the red-edge and near-infrared bands are critical for assessing vegetation health, biomass, and chlorophyll content through indices such as the Normalized Difference Vegetation Index (NDVI) (Drusch et al., 2012). Additionally, Sentinel-2's SWIR bands are useful for soil moisture estimation and soil property mapping, as they are sensitive to water content and soil composition (Verrelst et al., 2015).

One of the key strengths of Sentinel-2 lies in its synergy with Sentinel-1, which is a synthetic aperture radar (SAR) mission (Frantz et al., 2021). While Sentinel-1 provides all-weather, day-and-night radar data, Sentinel-2 offers high-resolution optical imagery. This combination is particularly advantageous for soil moisture studies. Sentinel-1's radar data can penetrate clouds and vegetation, providing information on surface roughness and soil moisture, while Sentinel-2's optical data complements this by offering detailed insights into vegetation cover and soil properties. Studies have demonstrated that integrating Sentinel-1 and Sentinel-2 data improves the accuracy of soil moisture mapping, especially in agricultural and hydrological applications (X. Gao & Schlosser, 2019; Z. Zhang et al., 2017). For example, the fusion of Sentinel-1's backscatter data with Sentinel-2's NDVI has been shown to enhance the estimation of soil moisture in vegetated areas, where radar signals alone may be less effective due to vegetation interference (Ienco et al., 2019).

2.3.Machine Learning Models for Soil Moisture Estimation

Machine learning (ML) has emerged as a powerful tool for soil moisture estimation, offering the ability to handle large datasets, capture complex relationships, and integrate diverse data sources (Babaeian et al., 2021). Traditional methods for soil moisture estimation, such as physical models and empirical equations, often struggle with the nonlinear and heterogeneous nature of soil properties and environmental factors (Rasheed et al., 2022). ML models, on the other hand, excel in identifying patterns and relationships within data, making them well-suited for soil moisture prediction. These models are increasingly applied to remote sensing data, such as those from

Sentinel-1 and Sentinel-2, to estimate soil moisture at various spatial and temporal scales. For example, ML algorithms have been used to combine radar backscatter from Sentinel-1 with optical and thermal data from Sentinel-2 to improve soil moisture mapping accuracy (J. Gao & Zhang, 2014). Additionally, ML models are employed to downscale coarse-resolution soil moisture products, such as those from the Soil Moisture Active Passive (SMAP) mission, to finer resolutions suitable for local applications (I. Ahmad et al., 2020).

The application of ML in soil moisture studies extends beyond remote sensing to include in-situ measurements, hydrological models, and climate data (I. Ali et al., 2015). For instance, ML models have been used to predict soil moisture using meteorological variables such as temperature, precipitation, and humidity, as well as soil properties like texture and organic matter content (Babaeian et al., 2019). These models are particularly valuable in regions with limited ground-based observations, where remote sensing data can fill critical gaps. Furthermore, ML techniques have been integrated with land surface models to improve the representation of soil moisture dynamics in hydrological and climate simulations (Karthikeyan et al., 2017). By leveraging the strengths of ML, researchers can develop more accurate and robust soil moisture estimation methods, which are essential for applications in agriculture, water resource management, and climate monitoring.

Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN) are among the most widely used ML models in soil moisture estimation (Rodriguez-Galiano et al., 2015). RF, an ensemble learning method, is particularly popular due to its ability to handle high-dimensional data, robustness to overfitting, and interpretability (Mohammed & Kora, 2023). RF works by constructing multiple decision trees during training and outputting the average prediction of individual trees. This approach has been successfully applied to estimate soil moisture using remote sensing data, such as combining Sentinel-1 radar data with Sentinel-2 optical imagery (Liang et al., 2021). RF's ability to rank feature importance also makes it valuable for identifying key variables influencing soil moisture, such as vegetation cover and surface roughness (Ning et al., 2023).

SVM, another widely used model, is known for its effectiveness in handling nonlinear relationships through the use of kernel functions (Widodo & Yang, 2007). SVM has been employed

in soil moisture studies to classify and predict soil moisture levels using remote sensing and in-situ data (Morales et al., 2021). However, SVM can be computationally intensive and less interpretable compared to RF. Neural Networks, particularly deep learning models, have gained traction in recent years due to their ability to model complex, hierarchical relationships in data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been used to process spatial and temporal remote sensing data for soil moisture estimation (Y. Liu et al., 2020). While NN models often achieve high accuracy, they require large amounts of training data and computational resources, which can be a limitation in some applications (L. Liu et al., 2016).

The performance of ML models in soil moisture prediction varies depending on the dataset, features, and application context (U. Ahmad & Sharma, 2023). RF has consistently demonstrated strong performance in soil moisture studies due to its robustness and ability to handle mixed data types. For example, a study comparing RF, SVM, and NN for soil moisture estimation using Sentinel-1 and Sentinel-2 data found that RF achieved the highest accuracy, particularly in vegetated areas where radar signals are complex (J. Wang et al., 2022). SVM, while effective, often requires careful tuning of hyperparameters and kernel functions to achieve comparable performance. NN models, particularly deep learning architectures, have shown promise in capturing complex spatial and temporal patterns but are often limited by the need for large labeled datasets and computational resources (Zhou et al., 2017).

In addition to accuracy, the interpretability of ML models is an important consideration in soil moisture studies (Y. Zhang et al., 2014). RF provides insights into feature importance, which can help researchers understand the key drivers of soil moisture variability. In contrast, NN models are often considered "black boxes" due to their complex architectures and lack of interpretability (Carmona et al., 2022). Hybrid approaches, such as combining RF with NN, have been explored to leverage the strengths of both models. For instance, a hybrid model might use RF for feature selection and NN for prediction, achieving a balance between accuracy and interpretability (Eskandari et al., 2024). Overall, the choice of ML model depends on the specific requirements of the study, including data availability, computational resources, and the need for interpretability (Lisboa et al., 2023).

One of the primary advantages of ML in soil moisture estimation is its ability to handle large, heterogeneous datasets and capture complex, nonlinear relationships (Abowarda et al., 2021). ML

models can integrate diverse data sources, such as remote sensing, meteorological, and in-situ measurements, to improve prediction accuracy (Ma et al., 2019). For example, ML models have been used to combine radar and optical data from Sentinel-1 and Sentinel-2, achieving higher accuracy than traditional methods (Gao et al., 2017). Additionally, ML models can adapt to different spatial and temporal scales, making them versatile tools for soil moisture mapping (Chakrabarti et al., 2014). The ability to automate the analysis of large datasets also reduces the time and effort required for manual processing, enabling more efficient monitoring and decision-making.

However, ML models also have limitations. One major challenge is the need for large, high-quality training datasets, which can be difficult to obtain in remote sensing applications (W. Li et al., 2019). In-situ soil moisture measurements, which are often used for model training and validation, are sparse and unevenly distributed, particularly in remote or inaccessible regions (Sheffield et al., 2018). Another limitation is the "black box" nature of some ML models, such as deep learning, which can make it difficult to interpret results and understand the underlying processes. Overfitting is another concern, particularly when models are trained on limited data or when the relationship between input features and soil moisture is not well understood (Prasad et al., 2020). Finally, the computational cost of training and deploying ML models, especially deep learning architectures, can be prohibitive for some applications (Lwakatare et al., 2020).

Training and validating ML models for soil moisture estimation using remote sensing data present several challenges (A. Singh et al., 2023). One of the primary challenges is the availability of high-quality, labeled training data. In-situ soil moisture measurements, which are often used as ground truth, are limited in spatial and temporal coverage, making it difficult to train models that generalize well across different regions and conditions (Yuan, 2020). Additionally, remote sensing data can be affected by various factors, such as cloud cover, sensor noise, and atmospheric interference, which can introduce errors and reduce the quality of the input data (Sagan et al., 2020). Preprocessing steps, such as cloud masking and atmospheric correction, are often required to address these issues, but they can be time-consuming and computationally intensive (Anzalone et al., 2024).

Another challenge is the validation of ML models, particularly in regions where ground-based observations are sparse (Salcedo-Sanz et al., 2020). Cross-validation techniques, such as k-fold

cross-validation, are commonly used to assess model performance, but they may not fully capture the variability in soil moisture across different landscapes and climatic conditions (Wong, 2015). Remote sensing data also often have different spatial and temporal resolutions, which can complicate the validation process (Duveiller & Defourny, 2010). For example, soil moisture estimates from coarse-resolution satellites like SMAP may need to be downscaled and validated against high-resolution data from Sentinel-1 or Sentinel-2, introducing additional uncertainty (Brocca et al., 2010). Addressing these challenges requires careful consideration of data quality, model selection, and validation strategies to ensure reliable and accurate soil moisture estimation.

2.4.Validation of Soil Moisture Estimation Models

Validation is a critical step in assessing the accuracy and reliability of soil moisture estimation models. It involves comparing model predictions with independent datasets, often referred to as ground-truth data, to evaluate performance. Validation techniques can be broadly categorized into in-situ validation, remote sensing-based validation, and cross-validation. In-situ validation relies on ground-based measurements from soil moisture sensors, such as time-domain reflectometry (TDR) or cosmic-ray neutron probes, which provide direct and accurate measurements of soil moisture at specific locations (Robinson et al., 2008). Remote sensing-based validation involves comparing model outputs with soil moisture products derived from satellite missions like SMAP, SMOS, or Sentinel-1/2, which offer broader spatial coverage but may have lower accuracy due to sensor limitations and spatial resolution (Entekhabi et al., 2010). Cross-validation techniques, such as k-fold cross-validation, are also commonly used to assess model generalizability by partitioning the dataset into training and validation subsets (Seraj et al., 2023).

The choice of validation technique depends on the availability of data, the spatial and temporal scales of the study, and the intended application of the model (Hoek et al., 2008). For example, in-situ validation is ideal for small-scale studies or localized applications, while remote sensing-based validation is more suitable for large-scale or regional studies. Cross-validation is particularly useful when ground-truth data are limited, as it allows for robust evaluation of model performance using the available data (Varoquaux et al., 2017). Regardless of the technique, validation is essential for ensuring that soil moisture models are accurate, reliable, and fit for purpose (Gruber et al., 2020).

Ground-truth data, typically obtained from in-situ measurements, are considered the gold standard for validating soil moisture models (Whitcomb et al., 2020). These measurements are collected using various techniques, such as gravimetric sampling, TDR, and capacitance probes, which provide direct and accurate estimates of soil moisture at specific locations and depths (Robinson et al., 2008). In-situ data are particularly valuable for validating remote sensing-based models, as they offer a point of comparison for assessing the accuracy of satellite-derived soil moisture products. For example, the International Soil Moisture Network (ISMN) provides a global repository of in-situ soil moisture measurements that are widely used for model validation (Dorigo et al., 2011). These datasets enable researchers to evaluate model performance across different regions, climates, and land cover types.

However, the use of in-situ data for validation is not without challenges (Wu et al., 2019). Ground-based measurements are often sparse and unevenly distributed, particularly in remote or inaccessible regions, which can limit their representativeness. Additionally, in-situ measurements may have different spatial and temporal resolutions compared to model outputs, necessitating careful matching of datasets for validation (Tapiador et al., 2017). Despite these challenges, in-situ data remain indispensable for validating soil moisture models, as they provide the most direct and reliable estimates of soil moisture (Rajib et al., 2016). Efforts to expand and improve ground-based monitoring networks, such as the integration of citizen science initiatives and low-cost sensors, are expected to enhance the availability and quality of in-situ data for validation purposes (Pino et al., 2022).

Several metrics are commonly used to evaluate the performance of soil moisture estimation models, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). RMSE is a widely used metric that quantifies the average magnitude of errors between predicted and observed values, with lower values indicating better model performance (Willmott & Matsuura, 2005). MAE, on the other hand, measures the average absolute difference between predictions and observations, providing a more robust measure of error that is less sensitive to outliers. R^2, or the coefficient of determination, assesses the proportion of variance in the observed data that is explained by the model, with values closer to 1 indicating a better fit (Nash & Sutcliffe, 1970).

In addition to these metrics, other performance indicators, such as bias, correlation coefficient (R), and the index of agreement (d), are also used to evaluate soil moisture models. Bias measures the systematic deviation of model predictions from observed values, providing insights into whether a model consistently overestimates or underestimates soil moisture (Entekhabi et al., 2010). The correlation coefficient quantifies the strength and direction of the linear relationship between predictions and observations, while the index of agreement provides a normalized measure of model accuracy that accounts for both systematic and random errors (Tsanas & Xifara, 2012). The choice of metrics depends on the specific objectives of the study and the characteristics of the dataset, with multiple metrics often used in combination to provide a comprehensive assessment of model performance (Ferri et al., 2009).

Numerous case studies have demonstrated the application and validation of soil moisture estimation models in different watersheds and regions (Cosh et al., 2004). For example, a study in the Little Washita River watershed in Oklahoma, USA, validated a machine learning-based soil moisture model using in-situ measurements from the USDA Agricultural Research Service (ARS) network (Crow et al., 2012). The model, which combined radar data from Sentinel-1 and optical data from Sentinel-2, achieved an RMSE of 0.04 m^3/m^3 and an R^2 of 0.85, demonstrating its effectiveness in capturing spatial and temporal variations in soil moisture. Another study in the Heihe River Basin in China validated a hybrid model that integrated remote sensing data with land surface models, achieving an RMSE of 0.03 m^3/m^3 and a correlation coefficient of 0.92 (Z. Li et al., 2020). These results highlight the potential of combining multiple data sources and modeling approaches to improve soil moisture estimation.

In Europe, the validation of soil moisture models has been facilitated by the availability of dense in-situ networks, such as the Cosmic-Ray Soil Moisture Observing System (COSMOS) and the European Soil Moisture Network (ESMN). A study in the Upper Danube Basin validated a Random Forest model using COSMOS data, achieving an RMSE of 0.05 m^3/m^3 and an R^2 of 0.80 (Zreda et al., 2012). Similarly, a study in the Iberian Peninsula validated a deep learning model using ESMN data, demonstrating its ability to accurately predict soil moisture across diverse land cover types and climatic conditions (Piles et al., 2014). These case studies underscore the importance of region-specific validation and the need for high-quality ground-truth data to ensure the reliability of soil moisture models.

2.5.Empirical literature

Machine learning models have gained popularity in SSM estimation due to their ability to handle large datasets, capture nonlinear relationships, and integrate diverse data sources. Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) are among the most widely used ML models in this domain. For example, a study by Gao et al. (2017) demonstrated the effectiveness of RF in estimating SSM by combining optical data from Sentinel-2 and microwave data from Sentinel-1. The study achieved high accuracy ($R^2 = 0.85$) and highlighted the importance of feature selection, such as vegetation indices and surface temperature, in improving model performance. Similarly, Singh et al. (2023) applied deep learning models, such as Convolutional Neural Networks (CNNs), to estimate SSM using multi-sensor remote sensing data, achieving superior performance compared to traditional methods.

The integration of optical and microwave remote sensing data has been shown to enhance SSM estimation by leveraging the complementary strengths of both data types. Optical data, such as those from Sentinel-2 and Landsat, provide high-resolution information on vegetation cover, surface temperature, and soil properties, which are critical for understanding soil moisture dynamics. Microwave data, such as those from Sentinel-1 and SMAP, offer all-weather capabilities and sensitivity to soil moisture, particularly in vegetated areas (Palacios-Lopez et al., 2019). For example, a study by Babaeian et al. (2019) combined Sentinel-1 radar data with Sentinel-2 optical data to estimate SSM in agricultural fields, achieving an RMSE of 0.04 m^3/m^3. The study highlighted the importance of using vegetation indices, such as NDVI, to account for the effects of vegetation on microwave signals.

In the Koga Watershed, the integration of optical and microwave data could address challenges related to cloud cover and spatial resolution. For instance, microwave data from Sentinel-1 can provide continuous SSM estimates during the rainy season, when optical data are often unavailable due to cloud cover. Additionally, the high spatial resolution of Sentinel-2 data can capture fine-scale variations in soil moisture, which are critical for crop monitoring in smallholder farming systems. Empirical studies have shown that the synergistic use of optical and microwave data improves the accuracy and reliability of SSM estimates, making it a promising approach for the Koga Watershed.

The application of ML approaches for SSM estimation in the Koga Watershed faces several challenges, including limited ground-truth data, spatial heterogeneity, and the complexity of crop-water interactions. Ground-truth data are essential for training and validating ML models, but in-situ soil moisture measurements are sparse in Ethiopia. Remote sensing-based validation, using products such as SMAP or ESA CCI, can partially address this issue, but the accuracy of these products may be limited in regions with complex terrain and land cover (Dorigo et al., 2017). Additionally, the spatial heterogeneity of the Koga Watershed, characterized by varying topography, soil types, and land use, poses challenges for model generalization.

Despite these challenges, there are significant opportunities for leveraging ML and remote sensing to improve SSM estimation in the Koga Watershed. For example, the availability of free and open-access remote sensing data from missions such as Sentinel-1, Sentinel-2, and Landsat provides a cost-effective solution for monitoring soil moisture in resource-constrained settings. Furthermore, the integration of ML models with crop growth models, such as AquaCrop or DSSAT, can enhance the utility of SSM estimates for crop monitoring and yield prediction. A study by Muthoni et al. (2020) demonstrated the potential of combining ML-based SSM estimates with crop models to improve yield forecasts in East Africa, highlighting the relevance of this approach for the Koga Watershed.

These studies provide valuable insights for the Koga Watershed, where similar approaches could be applied to improve SSM estimation and crop monitoring. For instance, the integration of topographic and land use data into ML models could enhance their accuracy in capturing soil moisture variability. Additionally, the use of high-resolution remote sensing data could support site-specific crop management, enabling farmers to optimize irrigation and fertilizer use. By building on these empirical studies, researchers can develop tailored ML models for SSM estimation in the Koga Watershed, contributing to improved agricultural productivity and water resource management.

2.6.Knowledge gaps

The empirical literature review on the use of machine learning (ML) approaches to estimate surface soil moisture (SSM) based on optical and microwave remote sensing products reveals several critical knowledge gaps, particularly in the context of crop monitoring in the Koga Watershed, Ethiopia. One of the most significant challenges is the lack of sufficient ground-truth data for

model training and validation. In regions like the Koga Watershed, in-situ soil moisture measurements are sparse due to limited monitoring infrastructure, which hinders the development of accurate and generalizable models. Additionally, existing ML models are often developed and validated in regions with different climatic, topographic, and soil conditions, limiting their applicability to the unique agroecological characteristics of the Koga Watershed. For example, studies such as Gao et al. (2017) and Babaeian et al. (2019) focus on temperate or semi-arid regions, which may not fully capture the tropical highland climate, heavy rainfall, and diverse land use patterns of the Koga Watershed. Furthermore, while the integration of optical and microwave remote sensing data has shown promise, there is limited exploration of advanced ML techniques, such as deep learning and transfer learning, which could improve the accuracy and robustness of SSM estimates. These gaps highlight the need for region-specific studies that incorporate local environmental characteristics and leverage multi-source remote sensing data to enhance model performance.

Another major gap is the limited focus on smallholder farming systems, which dominate the Koga Watershed. Most studies concentrate on large-scale agricultural systems, with insufficient attention to the small plot sizes, mixed cropping systems, and limited irrigation infrastructure typical of smallholder farms. This oversight limits the applicability of ML-based SSM estimates for smallholder farmers, who require high-resolution, localized soil moisture information for crop planning and irrigation management. Moreover, few studies have explored the integration of SSM estimates into decision support systems that provide actionable insights for farmers. Evans & Sadler (2008) and Hedley & Yule (2009) emphasize the importance of linking soil moisture data to practical tools, but this remains a challenge in practice. Additionally, there is inadequate attention to temporal dynamics, such as seasonal fluctuations and short-term changes due to rainfall or irrigation, which are critical for effective crop monitoring in the Koga Watershed. Finally, the socioeconomic and policy implications of using ML-based SSM estimates for crop monitoring and water resource management remain underexplored. Addressing these gaps will require a multidisciplinary approach that integrates remote sensing, machine learning, agronomy, and socioeconomics to develop innovative solutions for sustainable agriculture in the Koga Watershed and similar regions.

CHAPTER THREE

3. MATERIALS AND METHODS

3.1. Description of the Study Area

3.1.1. Location

The study area is the Koga catchment located in northwest Ethiopia. The catchment is characteristic of the Ethiopian Highlands. The Koga catchment is located between 11°9.7' latitudes and 11°30'N and 37°02' and 37°18'E longitudes. Seventy-two percent of the 260 km2 catchment area is devoted to subsistence farming. The upper portion of the Blue Nile (Abbay) watershed is home to the Koga River, which has its source in the Wezem mountain region. It is a tributary of the Gilgel Abbay, which flows northwest into Lake Tana.

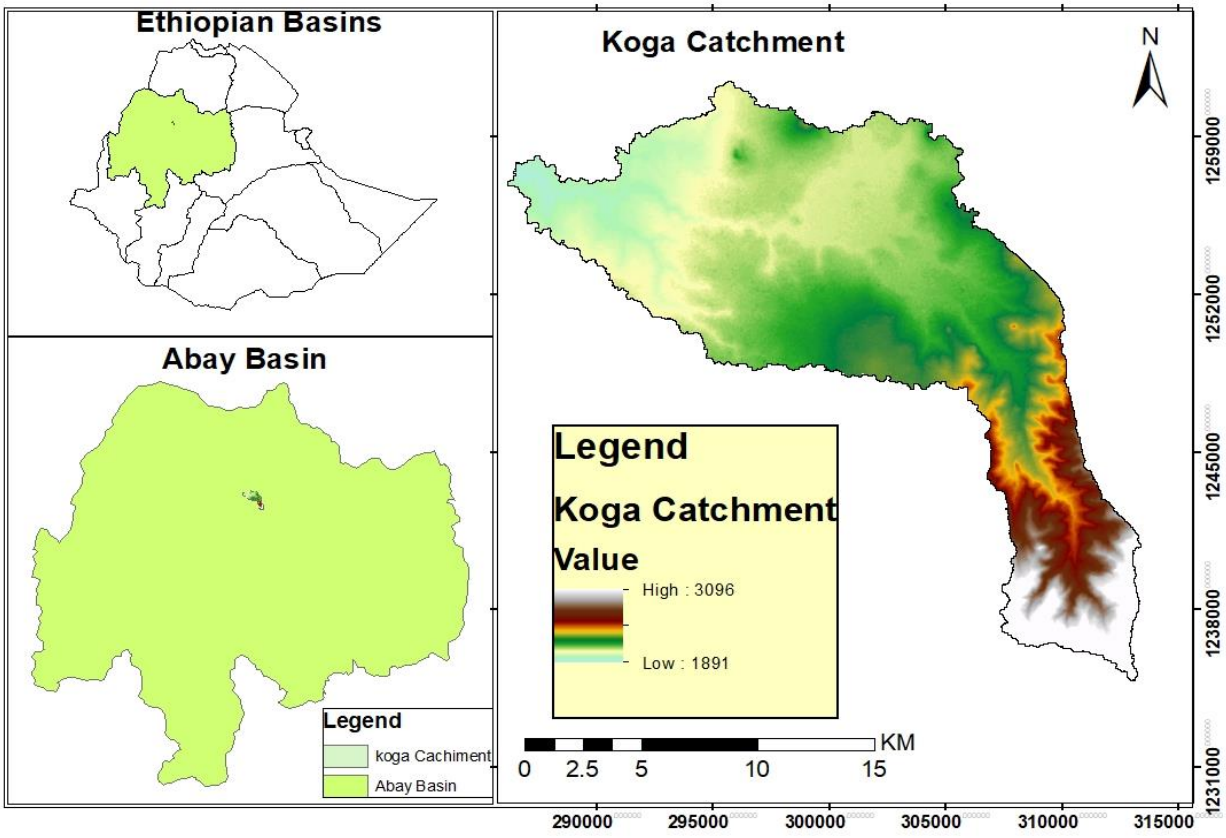


Figure 3-1: Location map of the study area

3.1.2. Climate and Topography

The Woina Dega (cool semi-humid, 1800 to 3100 masl) and Dega zones (cool, above 2,400 masl) are where the Koga catchment climate is found. There are distinct dry and rainy seasons throughout the majority of the catchment region, which is located in the Woina Dega zone. November through April is the dry season; May through October is the wet season. April and May saw periodic light showers. 18.25 °C is the average daily temperature. March and August have different monthly mean maximum temperatures, ranging from 30.0 °C to 23.7 °C. December's mean minimum temperature is 5.4 °C, while May and June have mean minimum temperatures of 13.1 °C (Michael, 2016).

3.1.3. Drainage

At the Koga River's mouth, the catchment's land elevation is 1,875 masl, while at its highest point on the watershed divide, it is 3,215 masl. Originating in the slopes to the north of the Wezem Mountains, the Koga River travels approximately 49 kilometers before joining the Gilgel Abbay, which ultimately empties into Lake Tana. The main river flows northward up to the settlement of Rim and then westward to Wetet Abbay, with a total catchment area of 260 km2 from source to mouth. The watershed is often longer than it is wide, with a noticeable constriction above Rim, where the basin's mean width is barely 3.8 km.

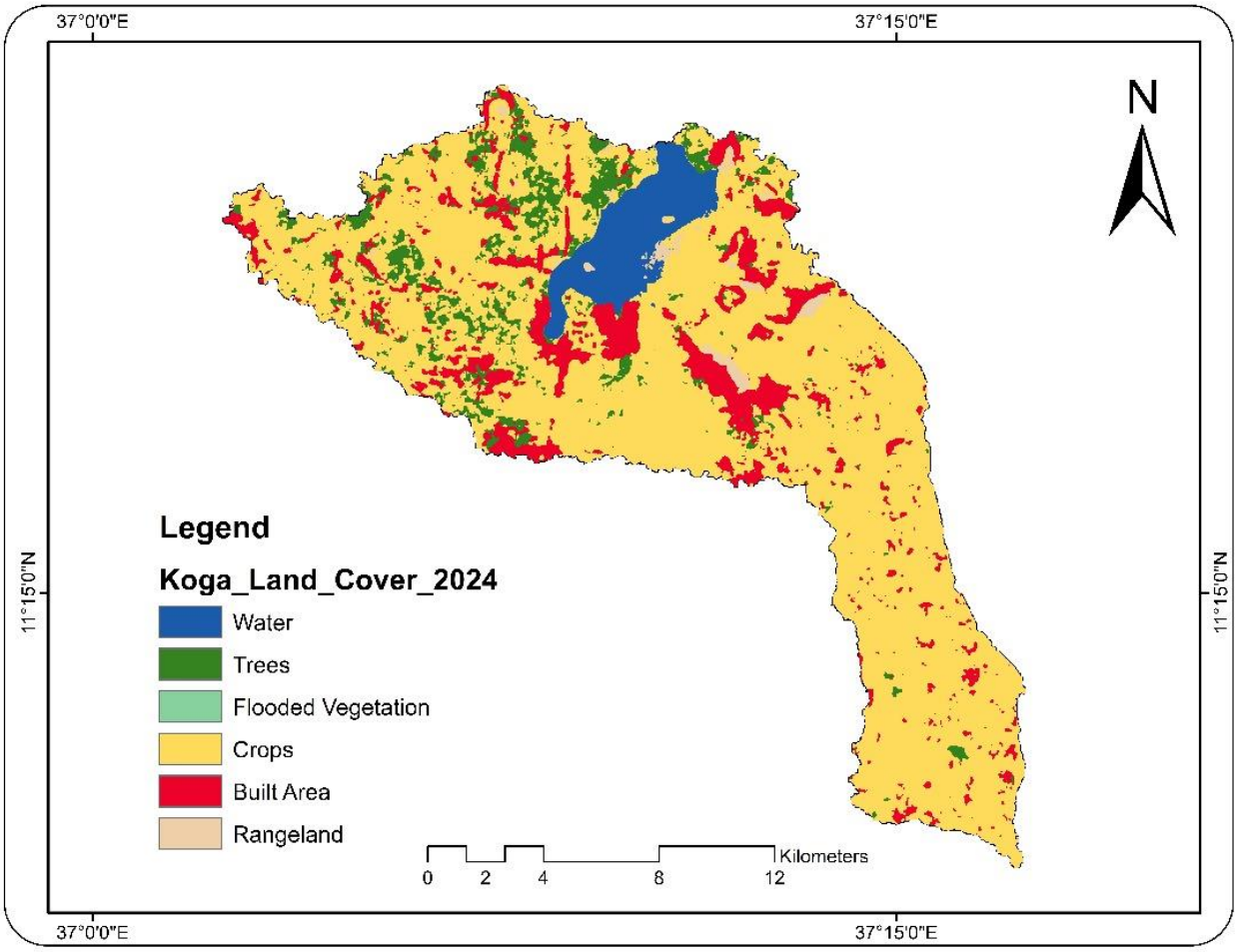
3.1.4. Soil Data

The Ministry of Water Resources, Irrigation, and Electricity (MoWIE) provided the soil data. It is crucial that the SWAT tool incorporate the various parameters that characterize the soil's hydrological and textural properties as well as the distribution of soil types in the catchments (Mengistu, 2019). The two soil types that are most prevalent in the Koga catchment are haplic luvisol and haplic alisols. Haplic luvisols predominate in the downstream portion of the catchment, whereas haplic alsisols predominate in the upstream portion.

3.1.5. Land use land cover

The Koga watershed has experienced notable transformations in land use and land cover, primarily marked by a shift from grasslands and forests to cultivated land and built-up areas. These changes are largely attributed to population growth and the rising demand for food and natural resources.

As agricultural land has expanded significantly, there has been a corresponding decline in grassland and forest areas, leading to effects on soil erosion and the watershed's hydrological behavior. Other land use types, such as built-up areas, flooded vegetation, and water bodies, have also undergone various changes over time.



3.2.Data source and type

3.2.1. Satellite Data

Sentinel-1 SAR: For this research, Sentinel-1A data were used to derive soil moisture information based on its Synthetic Aperture Radar (SAR) capabilities. The Sentinel-1 mission, developed under the European Space Agency's Copernicus Program equipped with C-band SAR sensors operating at a central frequency of approximately 5.405 GHz. The radar operates independently of daylight and weather conditions, enabling consistent data acquisition even under cloud cover or rain, which is particularly important for agricultural monitoring in tropical or semi-humid regions such as the Koga catchment. The SAR data used in this study has a spatial resolution of

approximately 10 meters, making it suitable for high-resolution surface monitoring. The specific image acquired on May 21, 2025, was selected due to its minimal atmospheric interference and alignment with the crop monitoring period.

Sentinel-2 MSI: To supplement the SAR-derived soil moisture analysis, Sentinel-2A optical satellite data were also used in this investigation. The MultiSpectral Instrument (MSI) on board Sentinel-2A is capable of taking pictures in 13 different spectral bands, which span the electromagnetic spectrum from the visible (VIS) to the near-infrared (NIR) and shortwave infrared (SWIR) regions. Vegetation and surface reflectance indices, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Soil Adjusted Vegetation Index (SAVI), were derived for this study using 10m spatial resolutions. In order to facilitate synergistic analysis, Sentinel-2 imagery obtained on May 22, 2025, was aligned with the Sentinel-1 SAR overpass for this study.

Table 3-1: Data source and type used for this research

Satellite	Sensor	Type	Spectral Band(s)	Spatial Resolution	Acquisition Date
Sentinel -1A/B	C-band Synthetic Aperture Radar (SAR)	Microwave (Radar)	C-band (~5.405 GHz)	10 m	May 21, 2025
Sentinel -2A/B	MultiSpectral Instrument (MSI)	Optical (Multispectral)	13 bands (VIS, NIR, SWIR)	10 m	May 22, 2025

3.3.Field data collection

Choosing an appropriate location for soil sampling is the first stage in gathering field data. The emphasis was on identified irrigated area, renowned for its variety of crop types, close to the man-made dam at Koga. The necessity to gather data from different crops in a controlled setting is the basis for this decision. Because of the easy access made possible by the dam's proximity, samples from a variety of crop types and soil types can be collected quickly. To guarantee that the data is thorough and representative, the chosen location should display a variety of crop types and soil

characteristics.

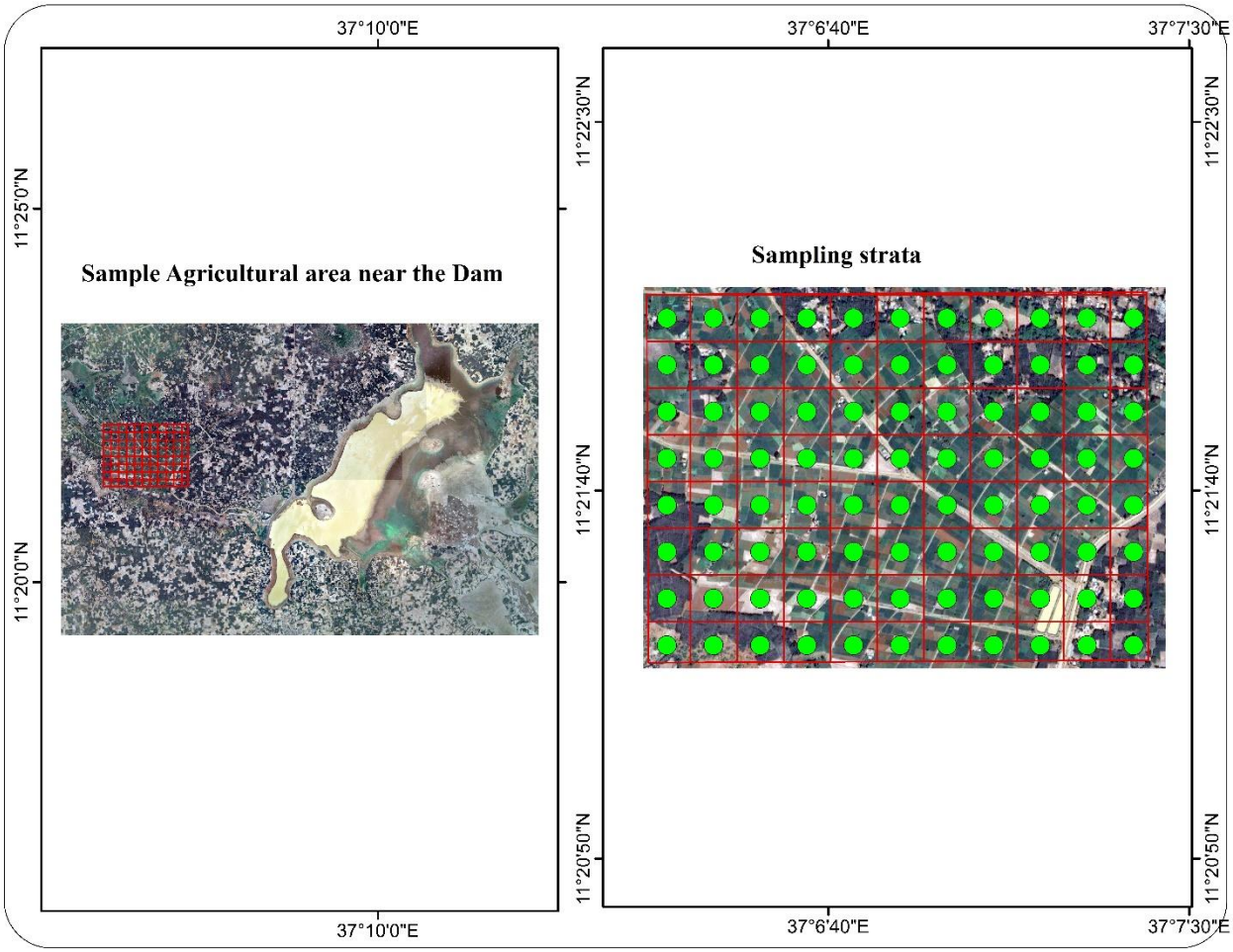


Figure 3-2: Sample stratification

Following site selection, stratification is carried out to establish discrete strata within the study area. Stratification entails dividing the site into uniform sub-areas according to particular traits. Several strata were identified in this study using high-resolution Google Earth images. The Google Earth imagery was superimposed with a fishnet grid created with ArcGIS (Figure 2). The next step is to randomly select sampling points within each grid cell, which are 20 m by 20 m (400 m2), after the strata have been defined. The samples gathered are guaranteed to be representative of the whole stratified region thanks to this procedure. In addition to reducing bias, randomization guarantees that the data gathered accurately represents the variation within each stratum. Points for soil sampling was established based on the grid cells created in the previous step. Field teams was visited these randomized points to collect soil samples, ensuring that each sample corresponds to the specific strata and crop type identified.

3.4.Soil sample collection

3.4.1. Gravimetric method

The characteristics of temporal and spatial variation of volumetric soil moisture content significantly affect the validation of satellite-based soil moisture products (Famiglietti et al., 2008). Ground-based measurements of soil moisture content represent point-scale soil moisture while satellite-based soil moisture estimation quantifies the average value of the spatial grid. For this research work, 20 ground measurements was collected from the field following SAR satellite over pass time for effective validation. Accordingly, all ground data collection was performed in line with the descending/ascending orbit path of the satellite.

Since soil varies in time and space, even over a very short distance, it is not sufficient to take samples at a single location. Therefore, a composite sampling technique was employed which involves a mixture of sub-samples at a different location. Composite sampling is to means that, collecting different soil sub-samples near the target location rather than taking a single soil sample at the desired location (Jason P. Ackerson, 2018). Then, mixing collected soil sub-samples until all the sampled soil retains a homogenous mixture which is going to be representative of the target location. Field tools used for collecting field samples included a soil probe, trowel, marker, notebook, sample bag, and GPS.

In order to guarantee consistency and precision in soil sampling, the field data collection in this study adhered to a methodical process modified from Jason P. Ackerson (2018). Finding suitable sampling sites throughout the study area to record the spatial variability of soil moisture was the first step in the procedure. Following the identification of the sites, field teams visited each one in order to collect samples. To prevent contaminating the soil sample, surface plant or crop residues were meticulously removed from each location. The active root zone, which is important for crop monitoring and soil moisture assessment, was then represented by a soil probe that was lowered into the ground to a standard depth of 30 cm.

The team collected a sub-sample at one location and then moved on to the next sampling point, repeating the soil probing and residue removal procedures. Once all sub-samples were collected from different points within a plot or sampling unit, they were thoroughly mixed using a trowel to create a composite sample. This approach helps reduce localized variability and ensures the representativeness of the sample. Finally, one to two cups of the well-mixed soil sample were

placed into a labeled sample bag for subsequent laboratory analysis or in-field moisture measurement.

The gravimetric method was applied to measure the amount of water content in each soil sample. Therefore, the percentage of the soil water content of each soil sample will be calculated using equation 1

$$SWC = \frac{FW-DW}{DW} * 100 \quad (1)$$

Where; SWC is soil water content in percent, FW is Fresh weight, DW is Dry weight (Oven-dry weight)

3.5.Method of data analysis

To estimate soil moisture content, field-collected soil samples were analyzed in the laboratory using the gravimetric method to obtain volumetric soil moisture values. These measurements served as ground truth data for model development. In preparation for modeling, satellite data from Sentinel-1A and Sentinel-2A were preprocessed using standard techniques, including geometric clipping, sub-setting, radiometric calibration, and multi-looking to enhance the quality and usability of the imagery.

The Sentinel-1A Synthetic Aperture Radar (SAR) data, operating in the C-band (not L-band as mentioned), was utilized to derive surface roughness parameters such as soil surface height and effective correlation length—both of which are crucial in influencing radar backscatter and soil moisture estimation. Land surface variables, such as vegetation indices like the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), and the Satellite Moisture Index (SMI), were simultaneously extracted using Sentinel-2A optical data. As predictor variables in the modeling process, NDVI and EVI were especially crucial in taking into consideration how vegetation cover affected soil moisture retrieval. As independent variables, moisture-related indices SAVI and SMI, which are obtained from optical bands were also included.

Using the satellite-derived variables as predictors, the Random Forest (RF) machine learning algorithm was used to forecast surface soil moisture at the watershed scale. The model was able to

learn the relationship between remotely sensed inputs and ground observations by using the volumetric soil moisture that was measured in the field as the dependent variable. R software was used for the modeling and analysis. Standard statistical measures, such as the Pearson correlation coefficient (r), coefficient of determination (R2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Error (ME), were used to assess the model's performance. Because of its low error statistics (RMSE, MAE, and ME) and high correlation (r and R2 values), which indicate higher predictive accuracy and dependability, the top-performing model was chosen. Estimates of soil moisture at the watershed scale were then produced using the finished model. To make sure the predictions were reliable, these outputs were checked against gravimetric soil moisture measurements made in the field.

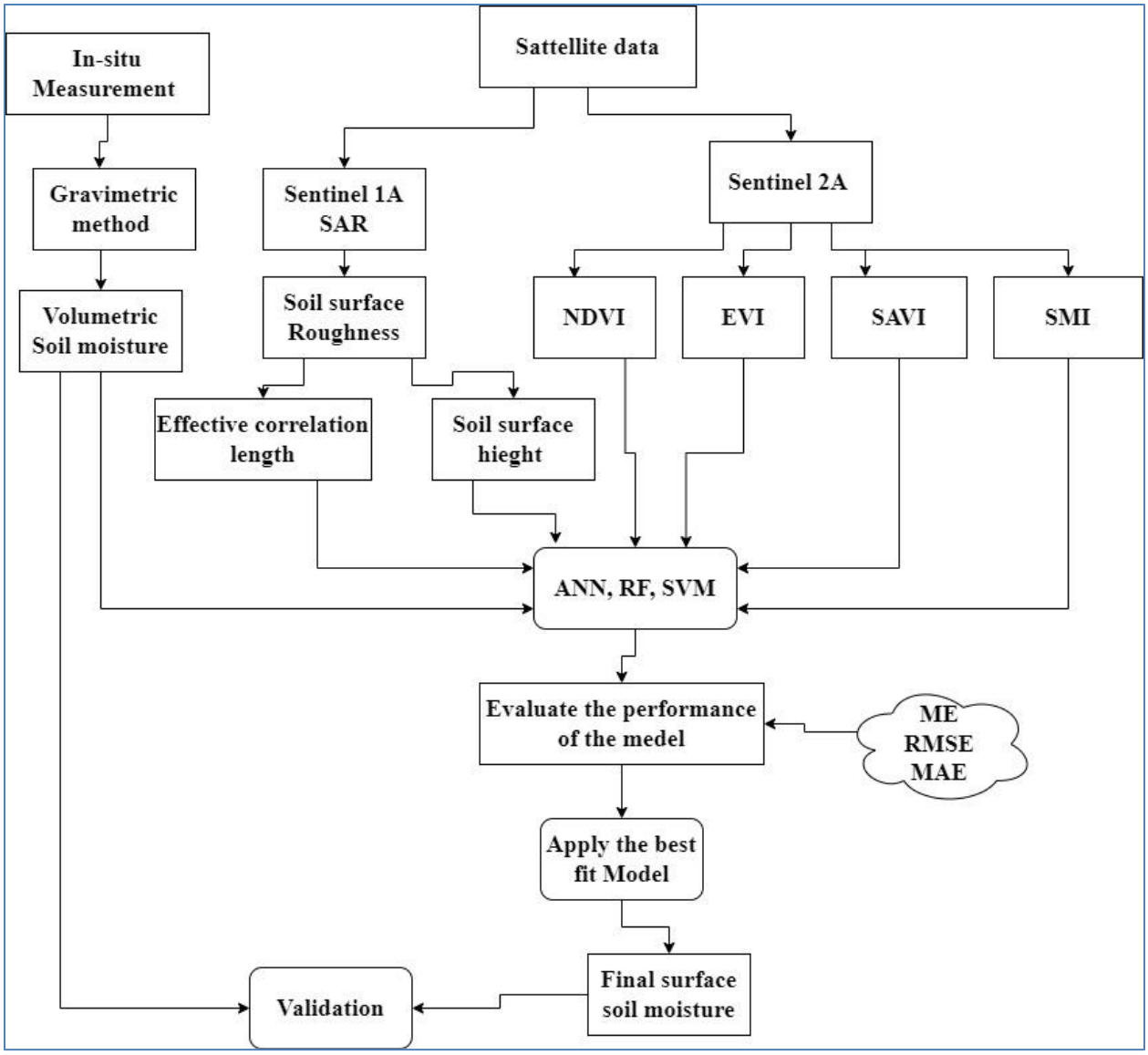


Figure 3-3: Methodological flow chart

After determining model predictor variables, surface soil moisture was predicted at a geographic scale using machine learning techniques such as ANN, SVM, and RF. The independent variables obtained from the Sentinel 1A and Sentinel 2A satellite sensors could change the dependent variable, which was the volumetric soil moisture measured in the field. The models (ANN, SVM, and RF) will be trained independently in order to specifically investigate each independent variable on soil moisture. R software was used to process, train, analyze, and run the model. The performance of each machine learning algorithm (ANN, SVM, and RF) was then evaluated using statistical methods such as coefficient of determination, Pearson correlation, RMSE, MAE, and ME. This can be done by comparing the model's estimated values with field-based gravimetric

measurements of soil moisture. Both the coefficient of determination (R) and the Pearson correlation coefficient (r) were high for the model selected for the final soil moisture forecast Selecting the machine learning algorithm that best suits your needs also requires low mean error values from RMSE, MAE, and ME. After selecting the machine learning model that best fits the data, final soil moisture data was collected at the watershed size. Field observations and the model's output for gravimetric soil moisture was compared. The flow chart above illustrates the general methodology used for this study.

3.6.Derivation of satellite predictor variables

3.6.1. Sentinel 1A predictor variables

Monitoring soil moisture content requires identifying and accounting for the effects of soil surface roughness (Ayehu et al., 2020). Standard deviations of surface height (h) and effective correlation length (L) were used to express the roughness of the soil surface (Zribi et al., 1997). The standard deviation of surface height (h) was determined in this study using a semi-empirical polarized backscattering equation inverted by (Oh, 2004b). The inversion of this model was used to estimate the soil surface height (h) by (Ayehu et al., 2020; Tong et al., 2018). Here is the (Oh, 2004b) model and its inversion for surface height (h) measurement in equation 3.

$$q = \frac{\sigma\ soil_{VH}}{\sigma\ soil_{VV}} = 0.95(0.13 + \sin 1.5\theta)^{1.4}\,(1 - \epsilon^{-1.3(K.h_{rms})^{0.9}}) \tag{2}$$

$$h_{rms} = \frac{\left\{\frac{-1}{1.3}\ln\left[1 - \frac{q}{0.95(0.13 + sin\ 1.5\theta)^{1.4}}\right]\right\}^{1.111}}{k} \tag{3}$$

Where $\boldsymbol{q}$ is cross-polarized ratios, $\boldsymbol{\theta}$ is incidence angle in degree, $\boldsymbol{K}$ is the wavelength of the SAR band, $\boldsymbol{h_{rms}}$ is the surface height at root mean square error. The surface correlation length (L) parameter was estimated based on the assumption that correlation length and surface height at the root mean square has a linear relationship over agricultural fields (Davidson et al., 2003). The relationship between these surface parameters was proposed by (Baghdadi et al., 2006) which describes effective correlation length as a function of surface height at root mean square. For this research work, a model developed by (Baghdadi et al., 2006) was used for surface length estimation. Here is the equation for surface correlation length (equation 4).

$$L_{eff} = (\alpha * h_{rms})^{\beta} \quad (4)$$

Where, $\boldsymbol{L_{eff}}$ is effective correlation length, $\boldsymbol{h_{rms}}$ is the surface height at root mean square, $\boldsymbol{\alpha}$, and β are coefficients which can be calculated from incidence angle ($\boldsymbol{\theta}$) and backscattering coefficient of VV- polarization ($\boldsymbol{\sigma_{soil_VV}}$) as follows.

$$\alpha = \delta(sin\theta)^{\mu} \quad (5)$$

$$\beta = \text{ɳ}\theta + \xi \quad (6)$$

Although $\boldsymbol{\delta}$ and $\boldsymbol{\xi}$ parameters are derived from calibration coefficient of polarization, ɳ and $\boldsymbol{\mu}$ parameters are constant and independent from polarization. In this case, the values of $\boldsymbol{\delta, \xi,}$ ɳ and $\boldsymbol{\mu}$ parameters are as follows:

$$\boldsymbol{\delta}_{\text{soil}_{\text{vv}}} = 3.829, \qquad \boldsymbol{\xi}_{VV} = 1.222, \qquad \text{ɳ}_{VV} = -0.0025, \qquad \boldsymbol{\mu}_{VV} = -1.744$$

3.6.2. Sentinel 2 MSI predictor variables

Onward 2015, European Space Agency (ESA) provides Sentinel-2A satellite imagery at 10m spatial resolution. For this study, sentinel-2 image of the year 2022/2023 was downloaded with 10m spatial resolution from https://scihub.copernicus.eu/ free of charge. All predictor variables that are going to be included in the Machine learning models (ANN, SVM and RF) was derived using standardized equations. In such away, predictor variables such as NDVI, EVI, LST, LAI, SAVI, NDMI, SMI, and NDWI. The following formulas indicated in the table (Table 2) was used to calculate the input variables.

Table 2: Formulas going to be used to derive different vegetation index

Vegetation Index	**Formula**
Normalized Difference Vegetation Index (NDVI	$NDVI = \frac{NIR - R}{NIR + R}$
Modified Normalized Difference Water Index (MNDWI)	MNDWI=(Gereen-SWIR)/(Green+SWIR)
Soil Adjusted Vegetation Index (SAVI)	$\text{SAVI} = 1.5 \times \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R} + 0.5}$

Normalized Difference Tillage Index (NDTI)	$NDTI = (\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$.
Normalized Difference Red Edge (NDRE)	$NDRE = (\text{NIR} - \text{Rededge}) / (\text{NIR} + \text{Rededge})$
Normalized Difference Water Index (NDWI)	$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}}$

3.7.Machine learning algorithms

3.7.1. Random Forest (RF) algorithm

Random Forest (RF) algorithm is an ensemble machine learning approach proposed by (Breiman, 2001) and later modified by (Cutler, 2012). It can be used for both classification and regression of different variables based on decision trees. For individual decision trees, the data are split into homogenous units of nodes to enhance the predictability of response variables. Random Forest (RF) algorithm integrates separated decision trees into predefined bootstrapped data. Therefore, the predicted values of continuous variables are the mean values of all fitted variables resulted from each bootstrapped sample (Breiman, 2001). RF algorithm can build a soil moisture relationship concerning surface variables (Im et al., 2016) and that is why the Random Forest (RF) approach has been adopted in this research work. Random Forest (RF) model is regarded as the dependent variable which is directly influenced by independent variables, as shown in equation 7:

where: θt -Independent variable generated from regression trees

X-Independent variable

t-Number of regression trees

It should be noted that the randomness characteristics of the independent variable make the regression tree sampled each time different (Chen et al., 2020). Several regression trees have been generated which of each regression tree has a predicted value. Therefore, the final regression equation takes the average of all independent regression tree predictions. Here is the RF-regression equation (equation 8) generated from the predicted values of each regression tree.

$$h(X, \theta t), t = 1,2,3 - - - tn \tag{7}$$

$$\bar{h}(x) = \sum_{t=1}^{T} \frac{1}{T} \{h(X, \theta t)\} \tag{8}$$

CHAPTER FOUR

4. RESULT AND DISCUSSIONS

The findings of a soil moisture prediction for the Koga catchment are shown in this chapter. A combination of predictor variables obtained from Sentinel-1 and Sentinel-2 satellite imagery were used to make the prediction. Key variables such as surface roughness and the VV and VH backscatter coefficients were extracted from SAR data. The MNDWI, NDRE, NDTI, NDVI, NDWI, and SAVI were also used, along with other multispectral indices from Sentinel-2 MSI data. These variables served as inputs for the RF regression model, which was employed to estimate soil moisture distribution across the Koga catchment at a fine spatial scale.

4.1.Sentinel 1A SAR predictor variables

4.1.1. Vertical Release-Vertical Receive (VV) backscattering coefficients

The spatial distribution of the VV-backscattering coefficient (in decibels) over the Koga catchment, as shown in the figure 4-1, demonstrates substantial spatial variability. The backscatter values range from a minimum of approximately –29.57 dB to a maximum of 10.36 dB. Higher backscattering values, displayed in green to blue shades, are primarily concentrated in the southern and southeastern parts of the catchment, which correspond to areas with more rugged terrain and possibly denser vegetation or higher soil moisture content. Conversely, lower VV-backscatter values (indicated by yellow to red colors) are predominantly observed in the central and northern regions, suggesting relatively drier surfaces, sparser vegetation, or smoother land cover.

Consequently, the higher VV values found in the watershed's southern region are related to higher soil moisture levels, which are reinforced by the region's vegetation density and terrain-induced runoff accumulation. On the other hand, the central and northern regions' low VV values indicate comparatively dry soil conditions. Additionally, these regions dominated by smoother surfaces or harvested agricultural lands, which reduce surface roughness and moisture content, both of which lessen radar backscatter.

This association between soil moisture and radar backscatter is supported by earlier research. For example, (Wagner et al. 2008) highlighted how sensitive C-band SAR data, like Sentinel-1, is to surface dielectric characteristics, making it possible to detect changes in soil moisture. Similar to this, (Datta et al. 2021) pointed out that while vegetation density can weaken the signal, VV

polarization is useful in recording variations in soil moisture in both vegetated and non-vegetated environments.

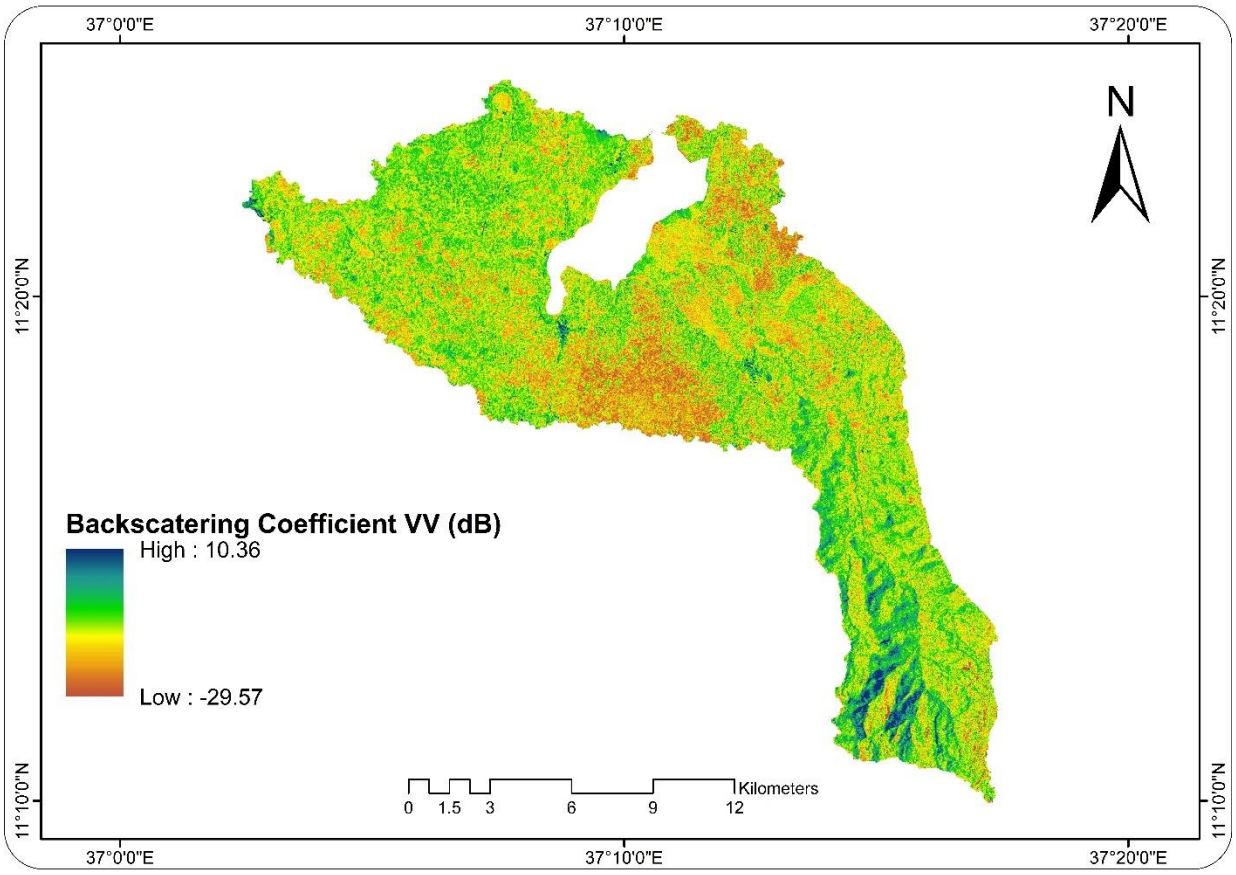


Figure 4-1: VV-Backscattering coefficients (dB)

4.1.2. Vertical Release-Horizontal Receive (VH) backscattering coefficients

According to the figure, the VH-backscattering coefficient's spatial distribution throughout the Koga catchment varies from roughly -47.62 dB to -0.21 dB. The catchment southern, southeast, and some western regions are primarily home to higher VH values (shown in blue-green hues). On the other hand, the central and northeastern regions exhibit lower backscatter values, which are symbolized by yellow to red hues.

Surface roughness and volume scattering from vegetation canopies have a particularly strong effect on VH polarization (horizontal receive and vertical transmit). Rough surfaces, wet soils, and dense vegetation are frequently linked to higher VH backscatter (Chen & Tan, 2025) Given this, the southern regions with high VH values are associated with higher vegetation cover, wooded areas, or wetter soils as a result of terrain-induced moisture retention. This is consistent with the

Koga watershed's overall topography and land cover features, which show that southern areas typically have higher elevations and denser vegetation.

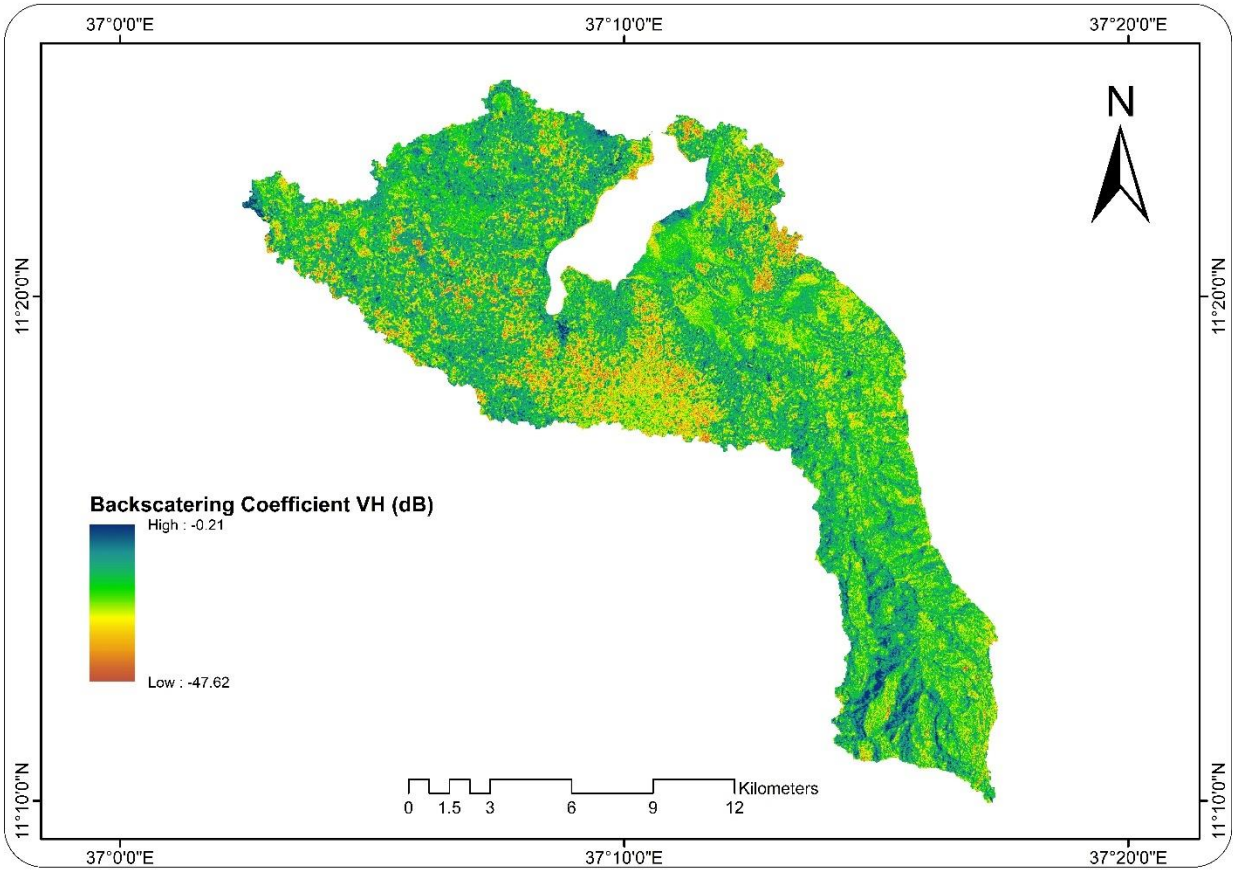


Figure 4-2: VH-Backscattering coefficients (dB)

On the other hand, the lower VH values in the central and northeastern zones a result of dry soil conditions, low vegetation cover, or smoother surfaces, all of which lessen the volume scattering effect of the radar signal. With little dielectric contrast or farmed lands during off-growing seasons. Literature highlighting the significance of VH polarization in soil moisture and vegetation studies supports the observed pattern. VH backscatter rises with both soil moisture and biomass, although vegetation attenuation can complicate the relationship (Greifeneder et al., 2018). Similarly, (Zhang et al., 2024) found that VH polarization provides complementary information to VV, particularly in vegetated environments, as it captures structural and dielectric heterogeneity more effectively.

4.1.3. Surface Roughness

The spatial distribution of surface roughness (measured in dB) across the Koga catchment, as illustrated in the figure4.3, ranges from approximately –4.89 dB to 31.64 dB. Surface roughness is a crucial parameter in microwave remote sensing as it directly affects the radar backscattering process. The map shows that areas with higher surface roughness values (depicted in green to blue) are predominantly concentrated in the southern and northwestern segments of the catchment, whereas lower values (yellow to reddish tones) are more prominent in the central and northeastern parts.

Higher roughness values typically indicate more irregular and uneven terrain, which can result from natural features such as rugged topography, stony soils, or dense vegetation with varied canopy structures. Agricultural activities, particularly land plowing, tillage, or residue-covered fields, can also contribute to increased roughness. These conditions enhance the surface's ability to scatter radar signals in multiple directions, increasing the radar return to the sensor, especially in dry conditions (Bounaceur et al., 2022).

In contrast, low surface roughness values are associated with smooth surfaces, such as bare or fallow agricultural lands, recently leveled fields, or compacted soils. Such surfaces reflect the incoming radar signal in a more specular manner, away from the sensor, resulting in low backscatter values. In flat areas with limited vegetative cover—likely found in the central basin of the watershed—low roughness can often be correlated with intensive agricultural land use, irrigation management, and possibly land degradation processes like crusting or compaction.

From a soil moisture retrieval perspective, surface roughness acts as a confounding factor that must be accounted for. The radar backscatter signal is influenced both by the dielectric constant of the soil (which increases with moisture) and the physical structure of the surface. When the surface is rough, even dry soils may produce high backscatter values, potentially mimicking wet soil conditions. On the other hand, wet soils might appear dry on a smooth surface because of less scattering (Zribi & Baghdadi, 2012; Mattia et al., 2003). Therefore, models that only use backscatter intensity (e.g., VV or VH) may misinterpret the soil moisture signal if surface roughness is not taken into account.

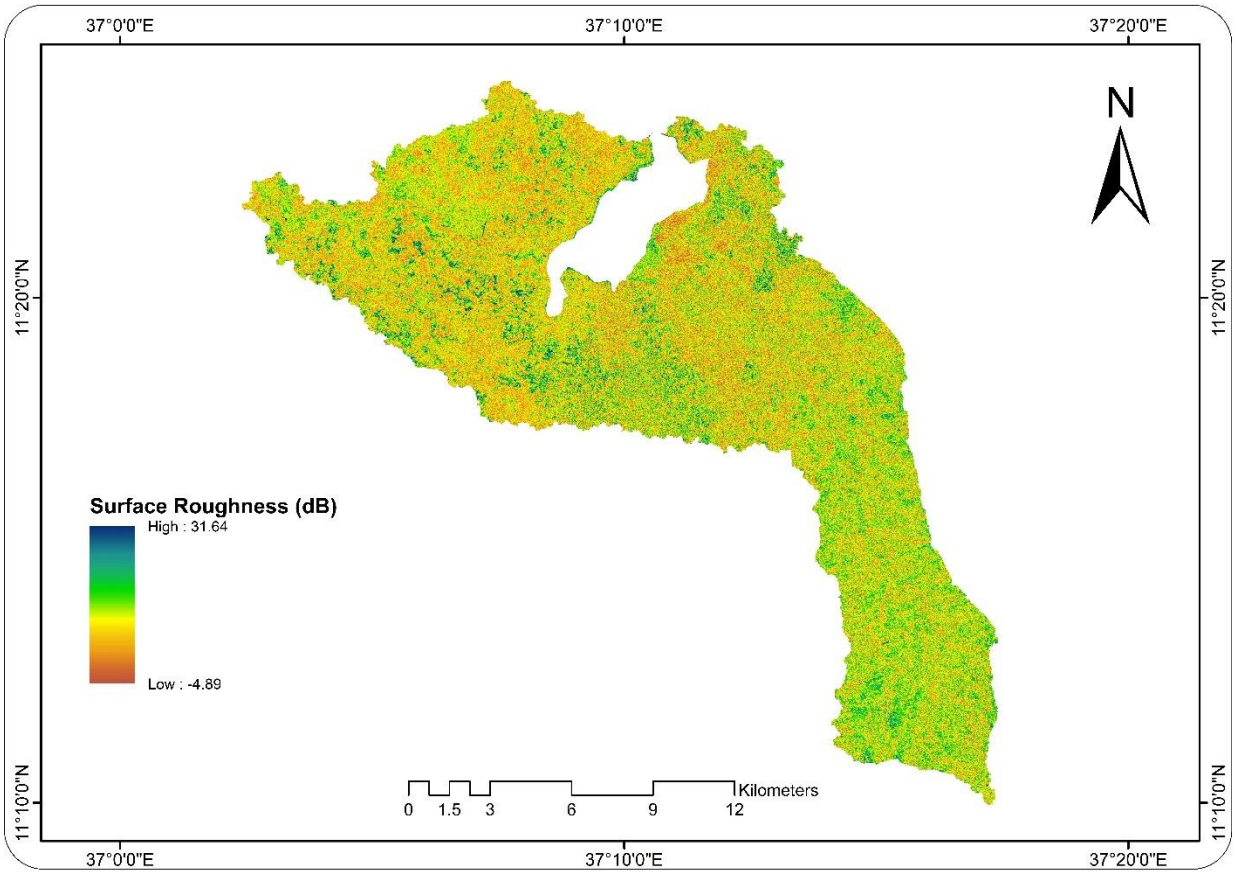


Figure 4-3: Surface roughness

Soil moisture estimation becomes more robust and reliable when surface roughness is included as a predictor in Random Forest (RF) regression models. The nonlinear relationships between vegetation indices, radar parameters, surface roughness, and soil moisture can be handled by RF. Incorporating surface roughness enhances the accuracy of soil moisture retrieval, as demonstrated by studies like those by Pierdicca et al. (2010) and Gruber et al. (2014). This is especially true in terrain, vegetation cover, and land use heterogeneous landscapes like the Koga catchment. Seasonal dynamics should also be taken into account. For instance, plowing or rainfall events can alter roughness by breaking up soil aggregates and smoothing the surface. Therefore, the interpretation of surface roughness and its impact on the estimation of soil moisture depends on the timing of image acquisition in relation to agricultural activities and rainfall events.

4.1.4. Polarization Difference (dB)

By measuring the variation in radar backscatter between two polarization channels, typically VV and VH, the polarization difference (PD), which was obtained from Sentinel-1 SAR data, offers

crucial information about surface biophysical properties. The spatial distribution of PD values in the Koga catchment varies from roughly –4.89 dB to 31.64 dB. Surface roughness, vegetation cover, and above all soil moisture content all have a significant impact on this variation. Areas of dense vegetation or higher soil moisture were generally linked to higher polarization difference values, which are shown on the map as green to blue hues. These areas show increased volume scattering as a result of radar signals interacting with moist soil or vegetation that is laden with water, which depolarizes the radar signal more. On the other hand, regions with lower PD values (yellow to red zones) are suggestive of smooth surfaces or dry, bare soils, where surface scattering predominates and radar signals experience little depolarization (Baghdadi et al., 2008).

The literature on remote sensing has clearly demonstrated the connection between soil moisture and polarization difference. Because Sentinel-1 SAR data is sensitive to the soil's dielectric characteristics, which are greatly impacted by moisture content, it has shown promise in the estimation of soil moisture. The polarization difference increases as soil moisture increases because the dielectric constant rises with stronger backscatter, particularly in cross-polarized channels like VH (Zribi et al., 2011; Sadeghi et al., 2017). Because of this, PD is a trustworthy stand-in for evaluating surface moisture spatial patterns, especially in diverse agricultural landscapes like the Koga catchment.

Furthermore, vegetation indices like the Normalized Difference Vegetation Index (NDVI), a useful predictor in soil moisture modeling, can be connected to polarization difference. Dense vegetation, which not only retains more surface moisture but also improves the volume scattering mechanism, is typically found in areas with high NDVI values. This increases the polarization difference (Bousbih et al., 2018). In this situation, there is a positive spatial correlation between NDVI and PD; zones with high NDVI are probably going to have higher PD values because of the combined influence of underlying moisture and vegetative biomass. By taking into consideration both vegetation dynamics and radar scattering mechanisms, integrating NDVI with polarization difference increases the accuracy of soil moisture retrievals (Veloso et al., 2017).

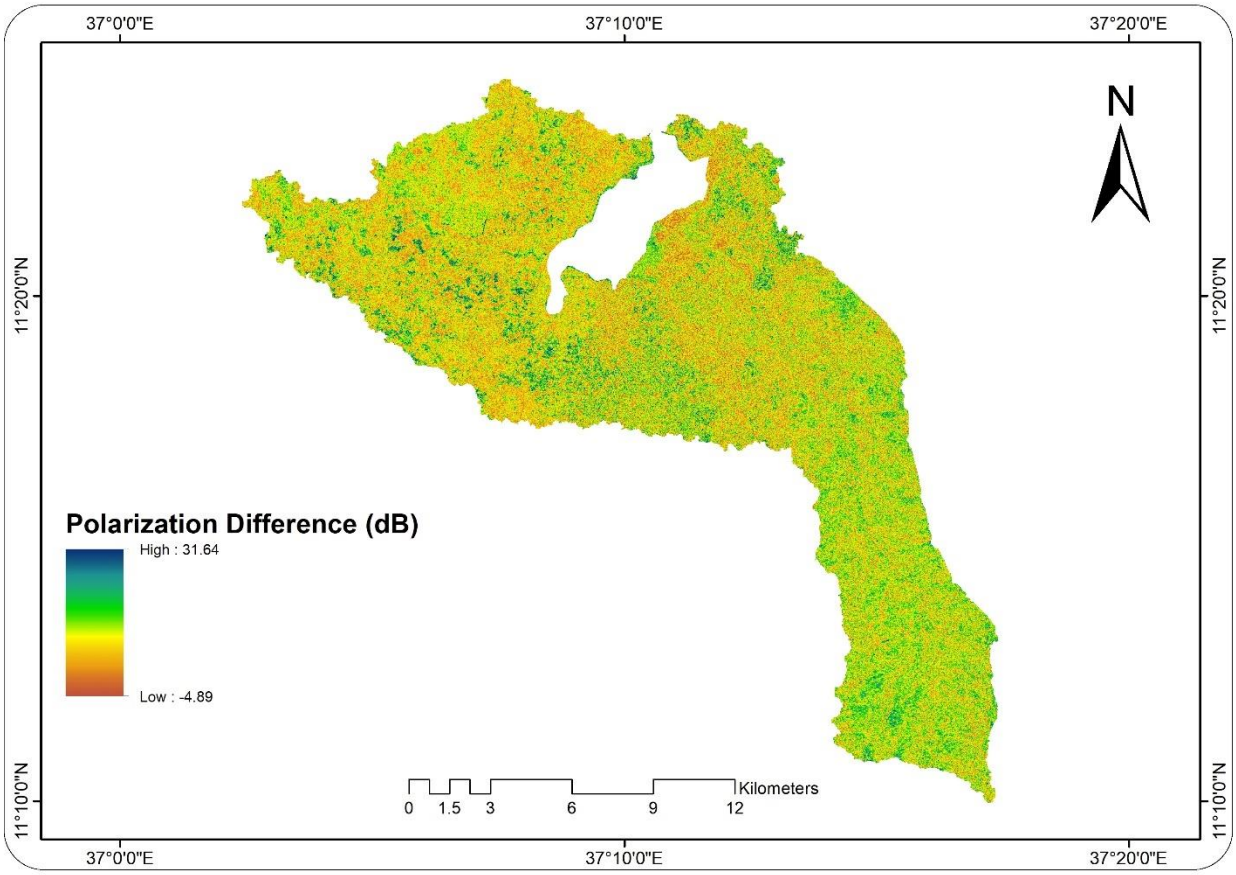


Figure 4-4: Polarization difference

4.2.Sentinel 2 MSI derived variables

4.2.1. NDVI

With NDVI values, ranging from -0.39 to 0.88, the spatial pattern of NDVI derived from Sentinel-2 MSI across the Koga catchment shows a heterogeneous distribution of vegetation cover. High NDVI values (green to blue hues) are concentrated in the northwest and central parts of the watershed, indicating dense vegetation or healthy crop cover, while low NDVI values (orange to red hues) dominate the southern and southeastern zones, likely reflecting bare land, harvested fields, or degraded vegetation. When compared to the Polarization Difference (PD) map derived from Sentinel-1 SAR, a clear spatial association emerges: areas with high NDVI generally coincide with higher PD values. This overlap suggests that both vegetation density and soil moisture content are influencing the radar backscatter characteristics.

Vegetated surfaces typically lead to complex volume scattering of radar signals, especially under moist conditions, which enhances cross-polarized returns and thus increases the polarization

difference (Balenzano et al., 2011). High NDVI values also indicate rapid vegetation growth, which is frequently associated with increased soil moisture availability, especially in rain-fed agricultural systems like Koga's (Bousbih et al., 2018). Thus, the combined analysis of PD and NDVI offers a complementary viewpoint: PD reflects both the underlying soil moisture and vegetation-induced depolarization from radar signals, whereas NDVI captures canopy vigor and greenness from optical data.

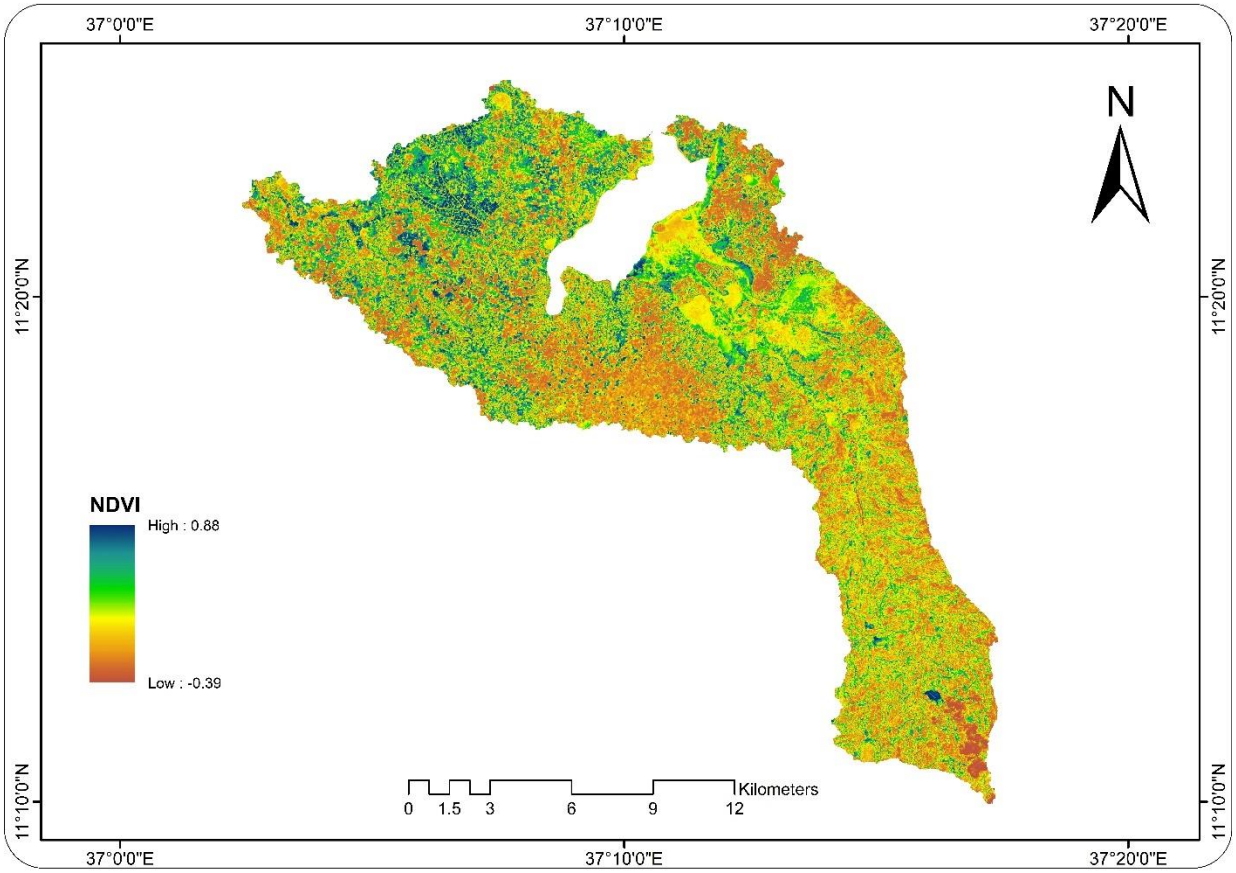


Figure 4-5: NDVI

Numerous studies support this synergy. The importance of combining Sentinel-1 SAR and Sentinel-2 MSI data for precise crop and soil condition monitoring was highlighted by Veloso et al. (2017). Zribi et al. (2011) state that SAR-based polarization metrics, like PD, are sensitive to surface roughness and moisture, and that ancillary data, such as NDVI, greatly aid in their interpretation by helping to distinguish between vegetation and moisture effects. The central-northern portions of the watershed in the Koga context, where both NDVI and PD values are high, most likely indicate regions of active crop growth and high soil moisture—a crucial finding for

drought assessment and irrigation planning. Land management and conservation planners should pay attention to the low NDVI and PD zones in the south, which may be a sign of dry soils, fallow land, or degraded vegetation.

4.2.2. NDWI

The Sentinel-2 MSI-derived Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) offer complementary information on the surface water content and vegetation health throughout the Koga catchment. Particularly in agricultural and semi-natural landscapes like Koga, these indices are essential for determining surface soil moisture and are frequently employed in environmental monitoring.

Meanwhile, the NDWI map shows values between –0.82 and 0.28, capturing the presence of surface water or vegetation water content. High NDWI values (above 0.1), particularly in the southeastern and lakeside areas, correspond to zones with higher surface moisture or shallow water bodies. Low NDWI values (less than –0.2), which dominate the western uplands and parts of the south, indicate drier soils and potentially stressed vegetation.

NDWI is particularly effective at identifying water-rich areas even under partial vegetation cover, as it uses the difference between green and near-infrared reflectance to isolate moisture-related features (Gao, 1996). In the context of soil moisture estimation, NDWI complements NDVI by providing a more direct assessment of surface water presence, especially over bare or sparsely vegetated areas where NDVI may be less reliable.

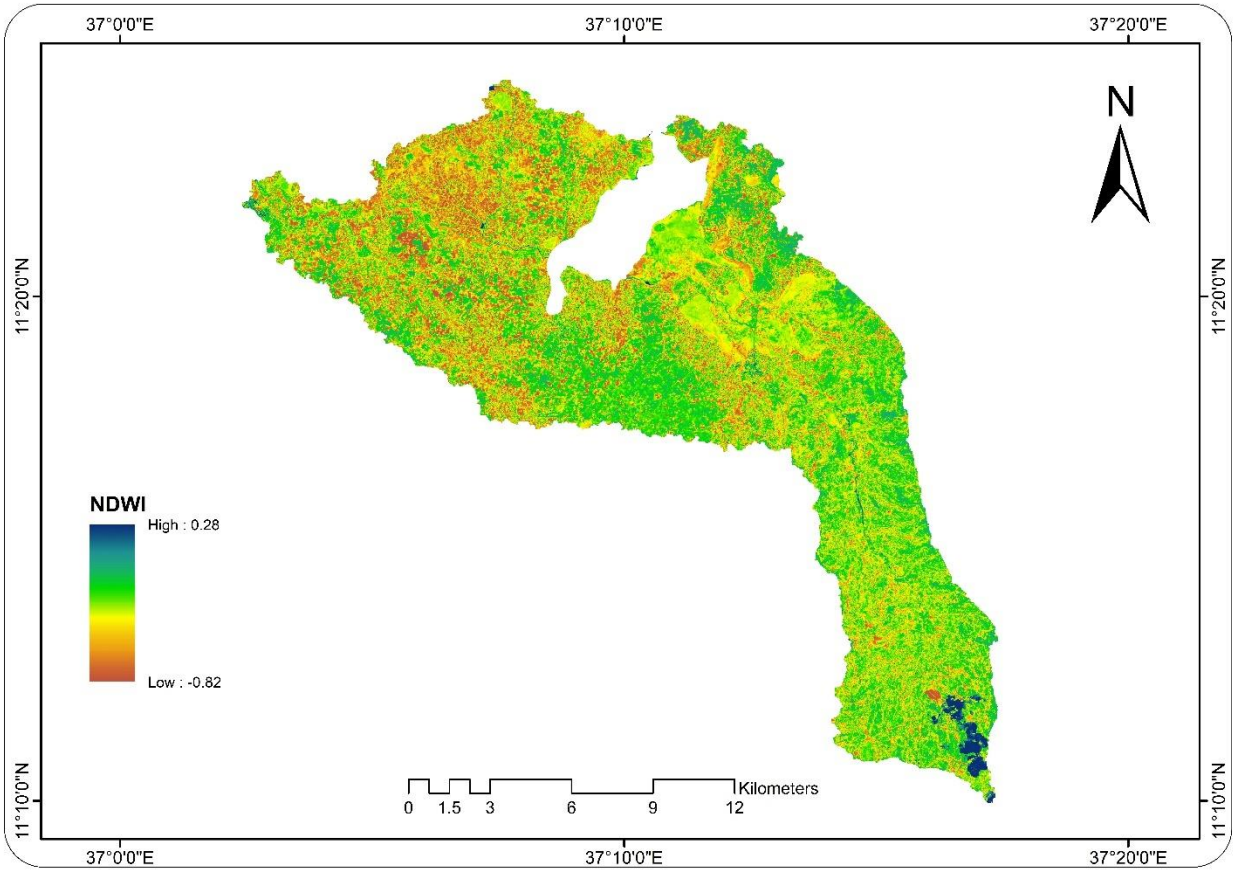


Figure 4-6: NDWI

4.2.3. SAVI

The Soil Adjusted Vegetation Index (SAVI) derived from Sentinel imagery, as shown in the map for the Koga watershed, plays a crucial role in assessing vegetation cover and indirectly estimating soil moisture. SAVI is particularly useful in semi-arid to arid regions like the Ethiopian highlands, where vegetation cover is sparse and soil background effects are significant. The index minimizes the influence of soil brightness, making it more reliable than NDVI in such environments (Huete, 1988).

The SAVI values on the Koga watershed map fall between roughly -0.59 and 1.33. Denser vegetation cover is indicated by higher SAVI values (green to blue), which are generally associated with higher soil moisture content. This is because sufficient soil moisture is necessary for the growth of healthy vegetation, and SAVI can be used as a stand-in for measuring soil moisture. On the other hand, bare or sparsely vegetated soils are represented by regions with low SAVI values (red to yellow), which are frequently linked to lower moisture levels.

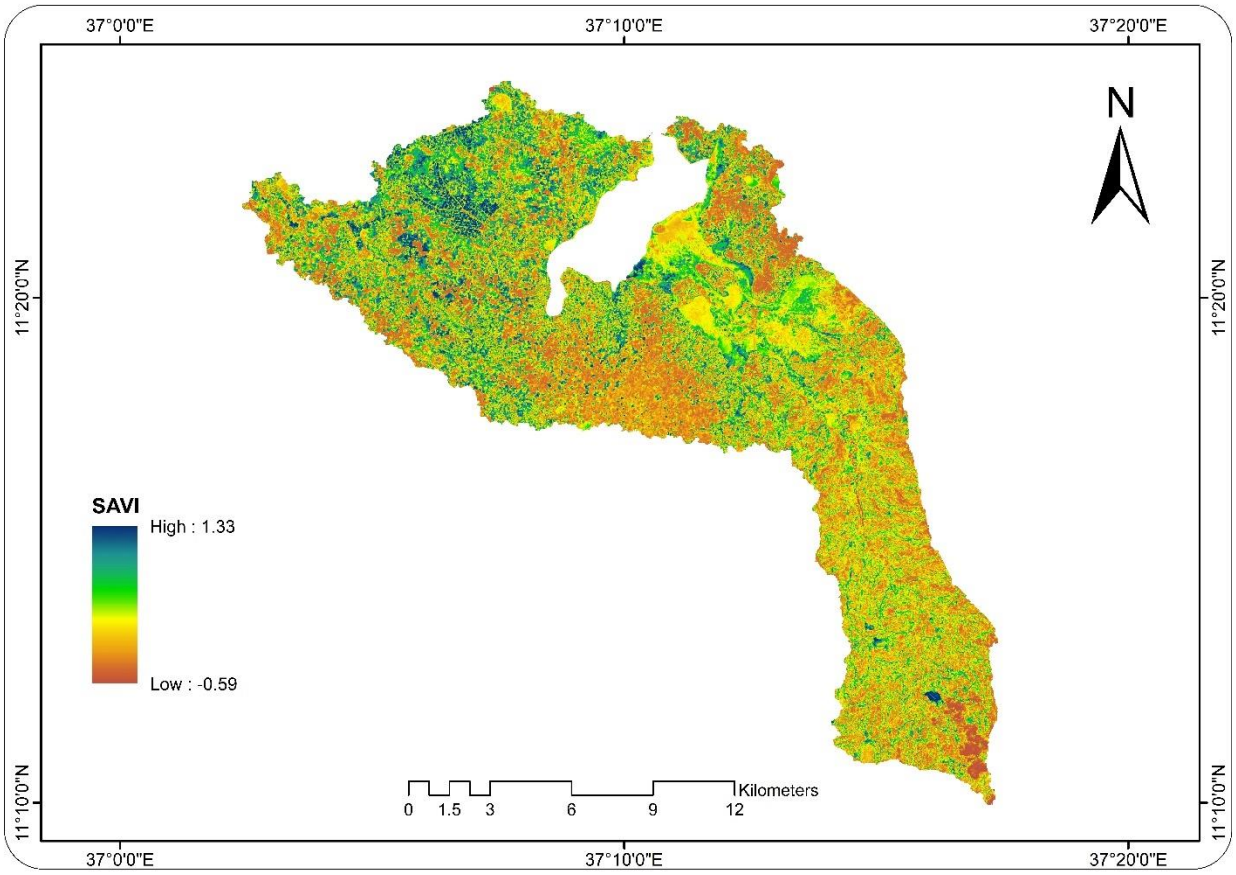


Figure 4-7: SAVI

Numerous studies have confirmed this relationship. In heterogeneous landscapes, for instance, Gao et al. (2018) showed a strong positive correlation between SAVI and in-situ soil moisture measurements. Likewise, research conducted in Ethiopian watersheds has demonstrated the efficacy of remote sensing vegetation indices, specifically SAVI, in forecasting the variability of surface soil moisture (Merga & Asfaw, 2020).

Furthermore, the Koga catchment's spatial pattern points to heterogeneity in the distribution of soil moisture, which may be influenced by soil texture, topography, and land use. Due to denser vegetation and irrigation from the Koga Dam, the northern and central regions seem to have higher SAVI values, while the southern and southeastern regions display lower values, suggesting drier soils.

4.2.4. MNDWI

As seen in the Koga watershed map, the Modified Normalized Difference Water Index (MNDWI), which is based on Sentinel imagery, is a useful instrument for identifying surface water and indirectly determining soil moisture. MNDWI is very effective in hydrological and agricultural studies because it suppresses vegetation, built-up land, and soil noise while enhancing the presence of open water bodies (Xu, 2006).

The values on the Koga watershed MNDWI map fall between roughly -0.87 and 0.82. Lower MNDWI values (yellow to red) denote drier regions or non-water features like vegetation or bare soil, whereas higher values (blue to green) indicate surface water or saturated soil zones. Regions with high surface moisture, probably as a result of irrigation or natural water accumulation, are indicated by the prominent blue areas in the south and close to the Koga reservoir in the map's center.

Because soil moisture is frequently higher in locations close to bodies of water or in areas with high water table levels, MNDWI is particularly helpful for estimating soil moisture. MNDWI values and soil moisture content have been found to be strongly correlated in a number of studies, especially in floodplains and agricultural zones (Feng et al., 2015; Li et al., 2021). For example, an increase in MNDWI following irrigation or rainfall events may signal a rise in soil water content, particularly in regions with shallow soils or inadequate drainage.

It is clear that regions with high SAVI do not always overlap with regions of high MNDWI when comparing MNDWI and SAVI (previous image). This is to be expected since MNDWI more directly measures soil saturation or surface water, whereas SAVI concentrates on vegetation health, which necessitates continuous soil moisture. A more thorough understanding of the soil moisture regime in the Koga catchment is possible through the complementary use of both indices.

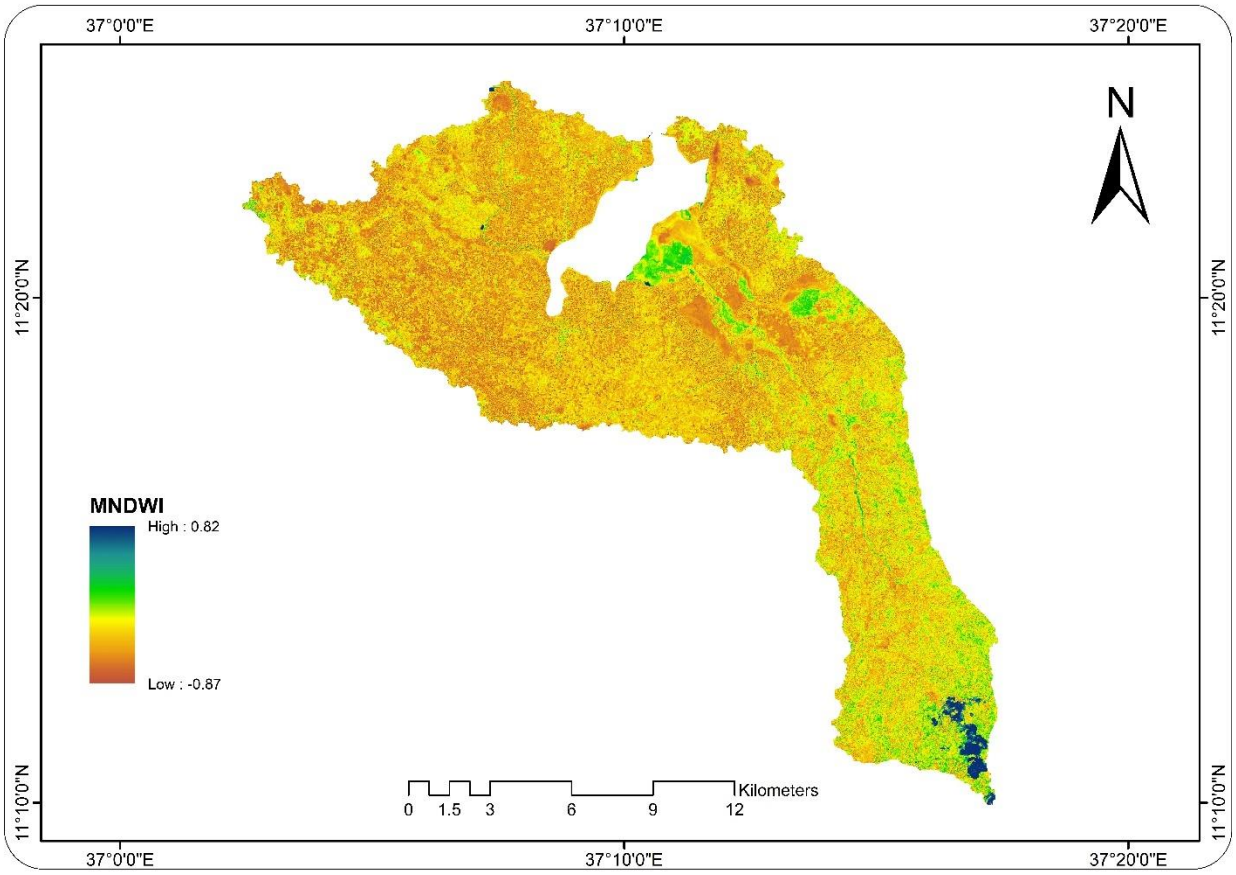


Figure 4-8: MNDWI

4.2.5. NDRE

In the Koga catchment, the Normalized Difference Red Edge Index (NDRE), which is obtained from Sentinel imagery, is a potent indicator of the health of the vegetation and, indirectly, the conditions of soil moisture. NDRE uses the red-edge and near-infrared spectral bands to measure the amount of chlorophyll in vegetation, which is inextricably linked to plant health and water availability. In the displayed NDRE map, the index values range from -0.37 to 0.71. Higher values, shown in green to blue, reflect areas with healthy, photosynthetically active vegetation—likely corresponding to zones of sufficient soil moisture. In contrast, areas with lower NDRE values, depicted in yellow to red, suggest stressed or sparse vegetation, which often results from inadequate soil moisture.

The spatial patterns observed in the NDRE map align well with ecological and hydrological variations across the Koga watershed. Higher NDRE values are found in the central and northeastern regions, which suggests higher soil moisture content and better vegetative health.

These results align with the corresponding maps of the Modified Normalized Difference Water Index (MNDWI) and Soil Adjusted Vegetation Index (SAVI). The idea that sufficient moisture supports vegetation is further supported by the high values of SAVI, which takes into account the effects of soil background in sparsely vegetated areas. Further supporting the availability of moisture is the MNDWI map, which emphasizes the presence of surface water and displays moderate to high values in some overlapping regions.

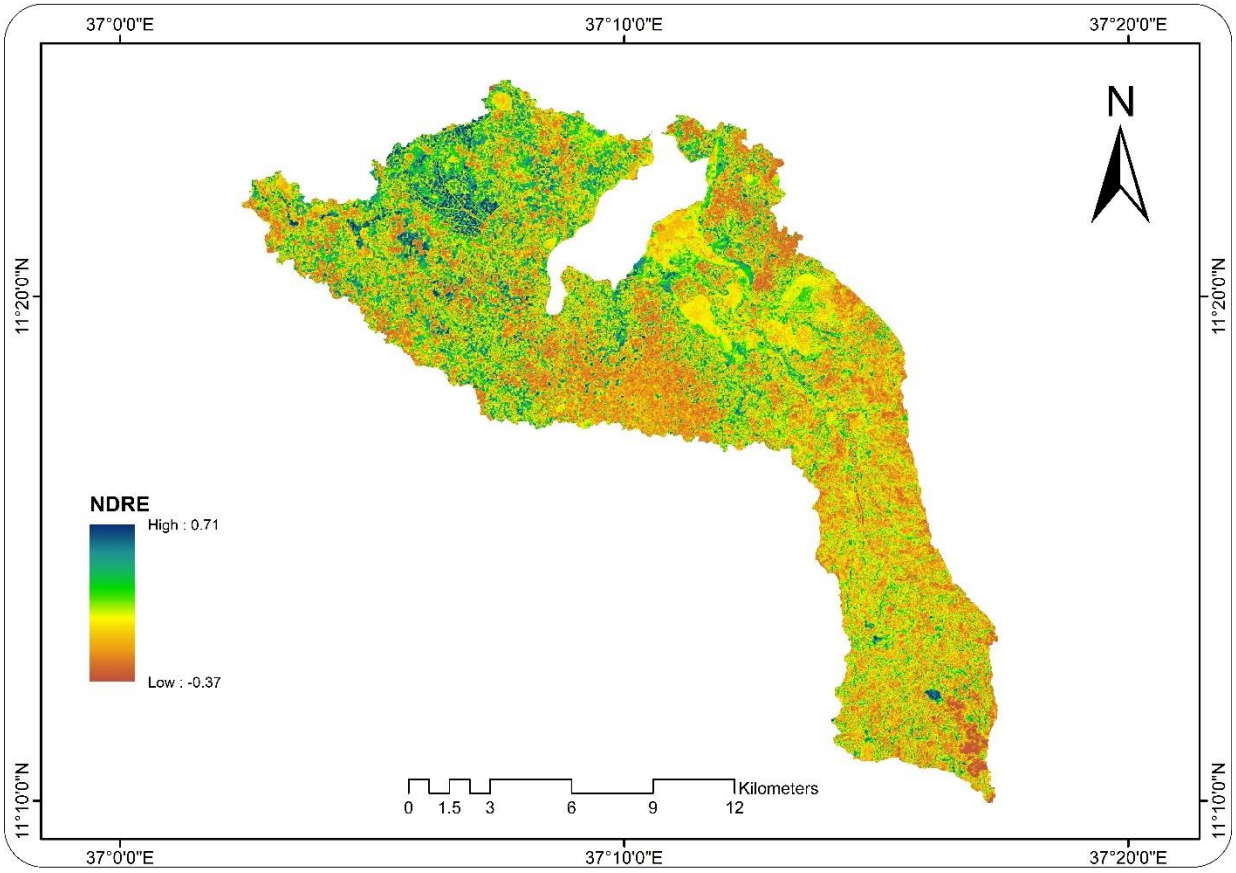


Figure 4-9: NDRE

Because traditional indices like NDVI may become saturated during the middle to late growing stages, NDRE's strength is its sensitivity to chlorophyll concentration. Because of this, NDRE is especially well-suited for tracking minute changes in plant health brought on by moisture stress. While SAVI focuses on vegetation cover adjusted for soil influence and MNDWI is more directly linked to surface water detection, NDRE offers a physiological snapshot of how plants are reacting to their surroundings, particularly water availability. Therefore, NDRE provides a more

sophisticated understanding of soil moisture distribution throughout the watershed when combined with SAVI and MNDWI.

4.2.6. NDTI

In landscapes like the Koga watershed, the Normalized Difference Tillage Index (NDTI), which is based on Sentinel-2 imagery, is a useful tool for determining whether soil surfaces are tilled or bare. This information can then be used to infer soil moisture conditions. The index is calculated using the Shortwave Infrared (SWIR) bands—specifically SWIR1 and SWIR2—according to the formula NDTI = (SWIR1 – SWIR2) / (SWIR1 + SWIR2). This formulation makes use of the difference in how vegetated or moist soils reflect SWIR wavelengths compared to bare or disturbed soils. The values on the given NDTI map of the Koga catchment range from -0.054 to 0.402; higher values are displayed in yellow to red (indicating more bare soil), while lower values are represented by green to blue hues (indicating low tillage or vegetation presence).

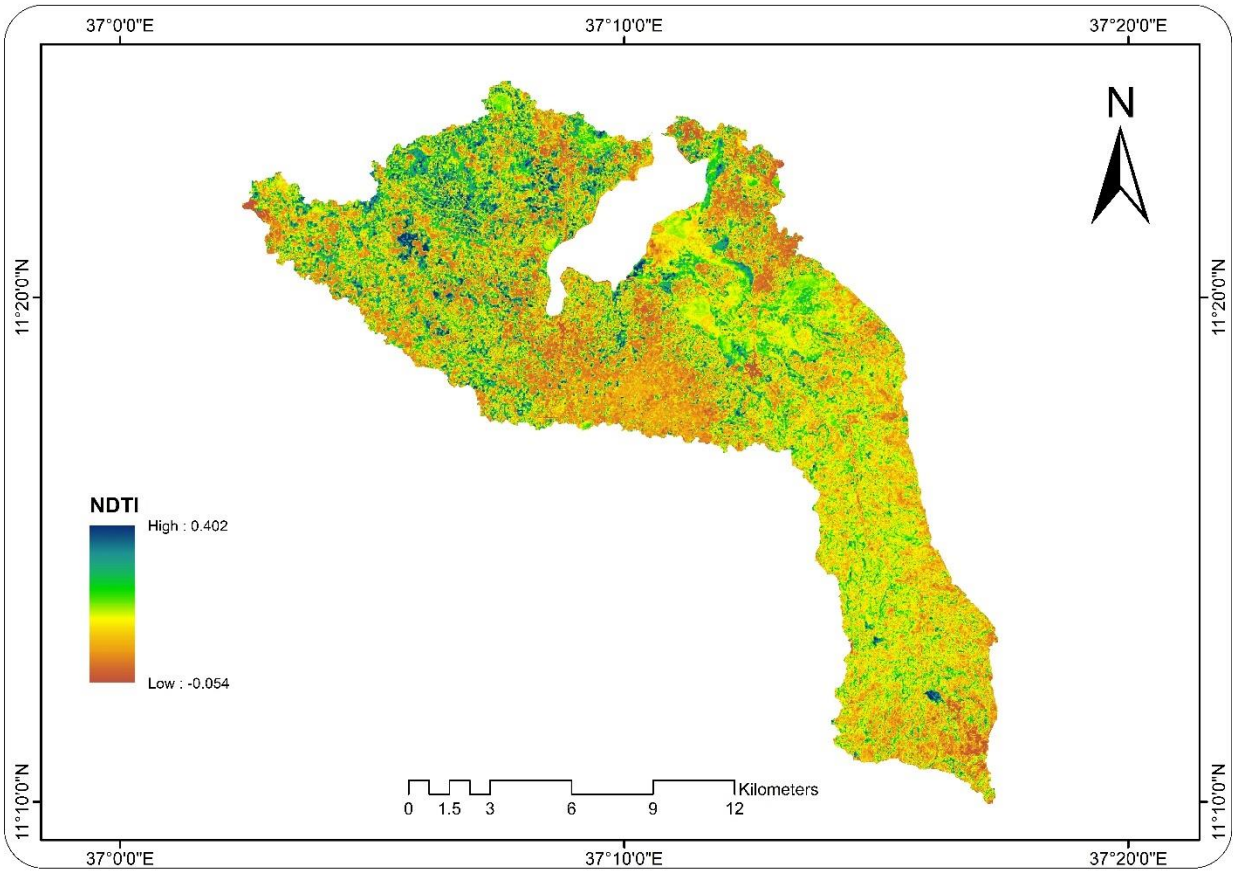


Figure 4-10: NDTI

Higher NDTI values are typically associated with lower soil moisture levels. This is due to the fact that exposed or recently tilled soils frequently have less vegetation, increasing their direct exposure to sunlight and, consequently, their evaporation. Therefore, regions that show up as red or yellow on the map are more likely to be suffering from a lack of soil moisture, either as a result of continuous tillage, intensive land use, or a natural inclination toward dryness. Conversely, regions with higher vegetation cover or more stable soil conditions both of which are linked to improved moisture retention generally have lower NDTI values. In the northern and central regions of the Koga catchment, where vegetation cover or closeness to bodies of water may contribute to higher soil moisture, this pattern is evident.

The inverse relationship between NDTI and soil moisture is supported by a large number of studies. For instance, Bai et al. (2016) observed that regions with decreased moisture content as a result of soil exposure could be successfully identified using tillage indices derived from SWIR bands. Similarly, Gomez et al. (2011) found that dry soil surfaces were strongly linked to high reflectance in SWIR bands, especially in agricultural areas where tillage alters the soil's structure and reduces its ability to retain water. Sadeghi et al. (2017) support these findings by stressing the significance of SWIR reflectance characteristics in remote sensing-based soil moisture estimation. Their study demonstrates that moisture variability can be effectively mapped using indices like the NDTI, which are sensitive to variations in surface conditions, particularly in areas with active land use practices like the Koga catchment.

4.3.Random Forest (RF) model performance evaluation

4.3.1. Variable importance

The variable importance results for estimating soil moisture (SM) in the Koga catchment during February, March, and April using the Random Forest (RF) model show distinct temporal patterns and consistency in the contribution of variables derived from remote sensing (Figure 4.11). Radar-based factors like Sentinel-1 VH (S1_VH) and VV (S1_VV) backscatter, surface roughness, and polarization difference (Pol_Diff) dominated the predictor list in February (Figure 4.11a). This implies that radar signals are especially good at identifying soil moisture through variations in surface texture and dielectric characteristics during the dry season, when vegetation is scarce. Bauer-Marschallinger et al. (2018) and Ulaby et al. (1982), who emphasized the resilience of SAR

backscatter in capturing soil moisture variations, particularly under low vegetation cover, corroborate these findings.

Vegetation indices like NDVI and SAVI become more significant as the season moves into March (Figure 4.11b). This change occurs when the region's brief rainy season begins, causing vegetative growth that influences evapotranspiration and the surface energy balance. During the early growing season, SAVI, in particular, becomes more informative in semi-arid regions and accounts for the effects of soil background (Sun et al., 2024). These findings are consistent with those of Srivastava et al. (2021), who highlighted the increasing value of vegetation indices in forecasting soil moisture during times of land cover transition.

While radar variables are still important but have a slightly smaller impact, April (Figure 4.11c) demonstrates a continued increase in the contribution of optical indices, especially NDTI and SAVI. This pattern suggests that optical indices are more useful in explaining soil moisture variability as the soil gets wetter and more vegetated. Additionally, Duan et al. (2019) discovered a strong correlation between NDTI and surface moisture and the effects of tillage, particularly in croplands where emerging crops gradually replace bare soil. Similarly, seasonal variation in variable importance was noted by Peng et al. (2020), demonstrating how vegetation indices become more dominant during wetter periods.

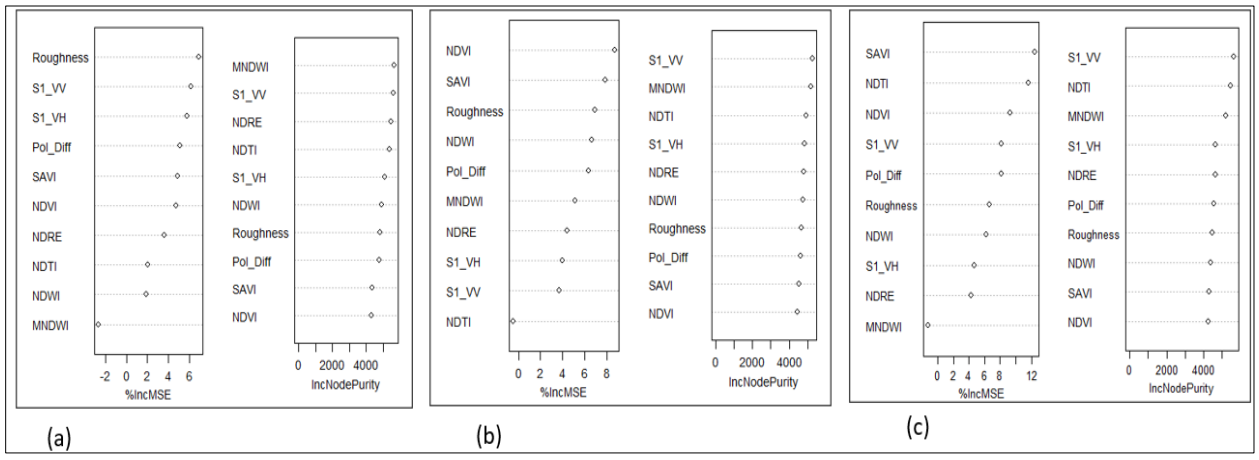


Figure 4-11: Variable importance in terms of MSE and NodePurity Index; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction

Water-related indices, such as MNDWI and NDWI, also rank highly in all three months, especially in March and April, suggesting a link between surface wetness and soil moisture. These indices

have been shown to detect shallow water and saturated soils effectively (Allbed et al., 2014). Their continued relevance implies that soil moisture prediction benefits from indicators sensitive to open water bodies or saturated soil pixels, particularly during peak infiltration periods.

Overall, the variable importance plots in Figure 4.11 illustrate that while radar data play a crucial role during the dry season, vegetation and water indices gain prominence as the land surface transitions into the wet season. This seasonal variation in predictor strength highlights the need for temporally adaptive models that integrate both radar and optical data sources. These findings support the findings of Zribi et al. (2005) and Pan et al. (2015), who highlighted the cooperative use of multi-sensor data for reliable soil moisture estimation under various land cover and seasonal conditions.

4.3.2. Scatter plot between observed and Predicted Soil Moisture (SM)

The accuracy and dependability of the Random Forest (RF) model over the course of the three months are demonstrated by the scatter plots in Figure 4.12, which visually compare the predicted and observed soil moisture (SM) values for February, March, and April. The correlation coefficients and RMSE values show potential for improvement, especially during the drier months, even though the RF model exhibits a moderate level of predictive capability.

In February (Figure 4.12a), the prediction accuracy is relatively low, as reflected by the Pearson's correlation coefficient ($r = 0.03$), Spearman's correlation ($\rho = 0.02$), and a root mean square deviation (RMSD) of 9.25. The regression line deviates significantly from the 1:1 line, indicating that the model tends to under- or overestimate values without strong correspondence to observed SM. This lower performance during February could be attributed to minimal soil moisture variability and reduced vegetation activity, which limits the predictive strength of vegetation-based indices. As noted by Entekhabi et al. (2010), soil moisture estimates tend to be less accurate in dry conditions when surface texture dominates the signal and optical indices offer limited sensitivity.

In March (Figure 4.12b), there is a slight improvement in prediction, with marginal increases in Pearson's ($r = 0.03$) and Spearman's ($\rho = 0.04$) correlations, although the RMSD remains high at 9.43. This slight enhancement corresponds with the beginning of the rainy season, where increasing vegetative cover contributes to improved modeling through vegetation indices. Similar observations were reported by Srivastava et al. (2021), who noted that early seasonal vegetation

dynamics, particularly in semi-arid regions, offer better cues for soil moisture prediction as vegetation emerges and water content increases.

April (Figure 4.12c) shows the highest predictive performance among the three months, with Pearson's correlation rising to 0.04 and Spearman's to 0.05. Though still relatively low, this improvement coincides with greater soil moisture heterogeneity and vegetation development, enhancing the model's capability. The regression slope also increases slightly, and the scatter plot suggests a marginally better fit between predicted and observed values. These results are consistent with those of Peng et al. (2020), who emphasized that machine learning models often perform better during periods of greater surface heterogeneity and vegetative growth, which provide more information-rich features for learning.

Despite the overall modest performance of the RF model across all months, the results suggest a seasonal pattern where prediction accuracy improves slightly with the onset of the rainy season. The relatively high RMSD values (~9.25–9.43) and low correlation coefficients across all three months reflect the challenges of predicting SM at finer scales using remote sensing variables alone, particularly in regions with heterogeneous land cover and limited ground validation data. The results affirm findings by Guevara and Vargas (2019), who argued that while remote sensing and machine learning provide promising avenues for SM estimation, their accuracy is often constrained by local-scale variability, input data quality, and insufficient in-situ calibration.

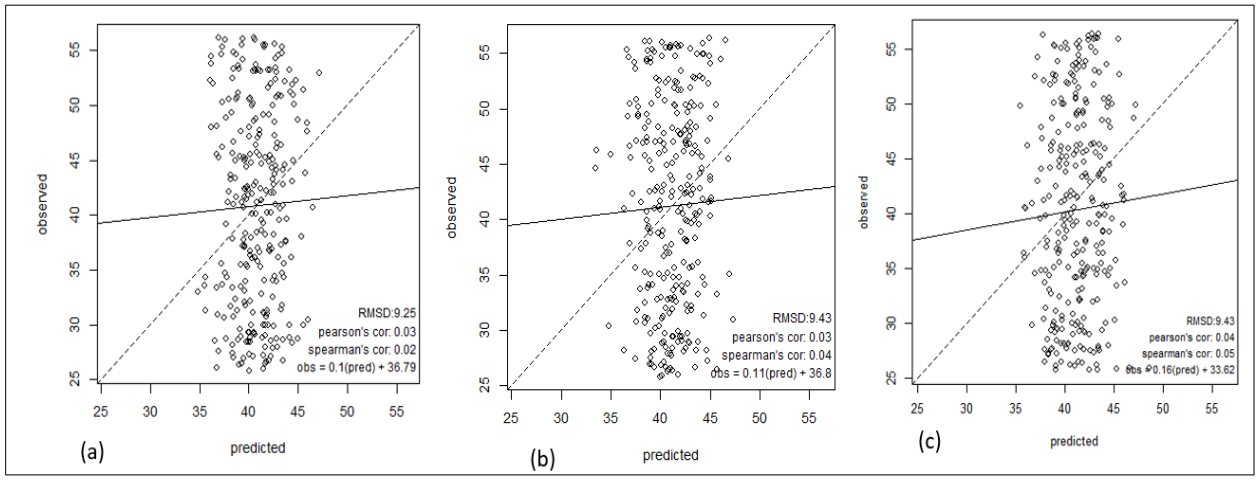


Figure 4-12: Scatter plot; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction

4.3.3. Predictor correlation

Figure 4.13 presents a matrix of Pearson correlation coefficients among various predictor variables used for soil moisture (SM) estimation across three months: February (a), March (b), and April (c). Consistent patterns in the relationship between SAR backscatter data and spectral indices are shown in the figure. Significantly, there are strong positive correlations ($r > 0.85$) between vegetation-related indices like NDVI, SAVI, NDRE, and NDWI, suggesting possible redundancy. Because these indices are all derived from similar spectral bands and represent vegetation vigor, greenness, and canopy moisture, it is possible to explain these strong interrelationships (Gao et al., 2000; Thenkabail et al., 2012). Multicollinearity, which inflates variance and compromises the interpretability of regression coefficients, can result from the inclusion of highly correlated variables in predictive models (Dormann et al., 2013). Variable selection techniques like Principal Component Analysis (PCA) or Variance Inflation Factor (VIF) thresholds are therefore frequently advised to minimize redundancy and guarantee strong model performance (Zhou et al., 2022).

Furthermore, surface roughness and MNDWI (Modified Normalized Difference Water Index) have comparatively weak or negative correlations with vegetation indices and SAR-derived metrics, suggesting that they have a special role in predicting soil moisture. By identifying wetness in bare or sparsely vegetated areas, MNDWI, which is sensitive to open water and surface moisture, enhances vegetation indices (Xu, 2006). Similarly, roughness provides non-spectral information that is helpful in SM modeling, particularly in heterogeneous terrains, and is derived from topographic variation or radar backscatter texture (Baghdadi et al., 2012).

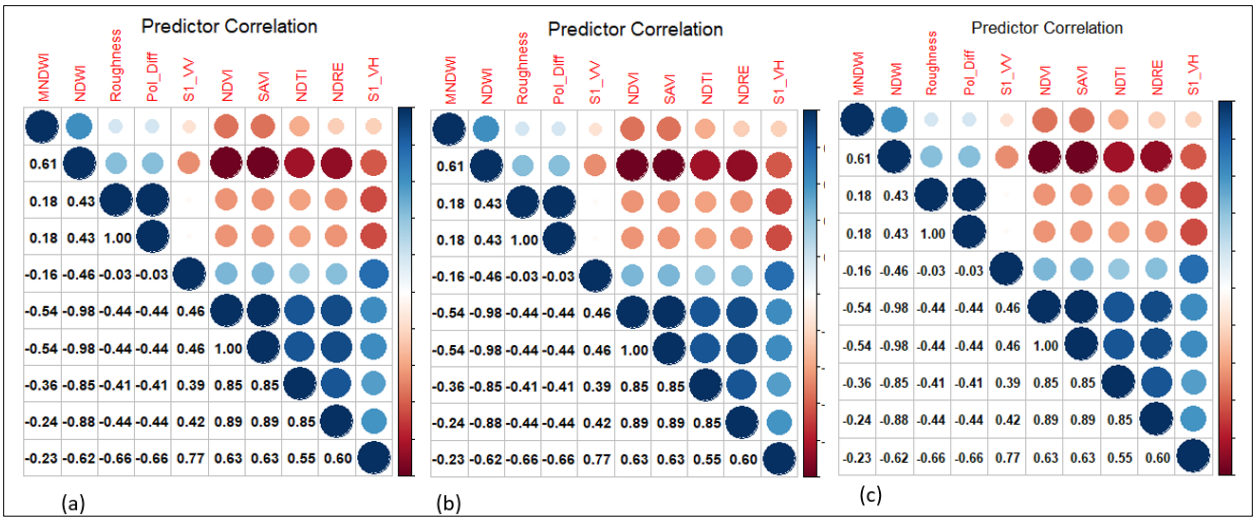


Figure 4-13: Predictor correlation; (a) February SM prediction; (b) March SM prediction; (c) April SM prediction

The correlation matrices also show a strong and consistent relationship between S1_VV and S1_VH ($r \approx 0.89–0.90$), which is typical for dual-polarization SAR data. These bands capture surface roughness and dielectric properties, both of which are sensitive to soil moisture content (Mattia et al., 2003). However, since these bands are highly correlated, using both without transformation may add redundancy. Thus, many studies recommend utilizing combinations such as VV/VH ratio or polarimetric decomposition features for improved sensitivity (Veloso et al., 2017).

Temporal consistency is evident across the three months, though minor variations in correlation values suggest dynamic interactions between soil moisture and predictor variables, potentially driven by changes in vegetation phenology and rainfall patterns. Such seasonal sensitivity is critical to consider in time-series SM modeling, as noted by Entekhabi et al. (2010) in their analysis of soil moisture dynamics.

4.4.Soil Moisture (SM) Estimation

The map in Figure 4-14 displays the predicted soil moisture (%) across a region for February 17, 2025, with values ranging from 31.04% to 51.47%, derived from integrated Sentinel-1 and Sentinel-2 satellite data. The color scale indicates moisture variation—blue tones represent higher soil moisture content, while red to yellow tones denote drier conditions. Two validation sites, Site 1 (purple box) and Site 2 (red box), are shown and magnified in panels (c) and (b) respectively, with ground sampling locations marked as points.

The map reveals spatial variability in soil moisture, with relatively higher moisture content concentrated in central and lower elevation areas, likely due to better water retention or recent precipitation. Sentinel-1 SAR data (VV/VH backscatter) contributed to detecting soil moisture by capturing surface roughness and dielectric constant variations, both of which increase with higher water content (Zribi et al., 2011; Mattia et al., 2003). Meanwhile, Sentinel-2's multispectral indices (e.g., NDVI, NDWI, SAVI) complement this by identifying vegetation cover and canopy moisture, which are proxies for underlying soil moisture, particularly under vegetation (Gao, 1996; Srivastava et al., 2020).

The field validation sites are zoomed into in panels (b) and (c). In line with its higher moisture prediction, Site 2, shown in panel (b), seems to have more wet zones (green to blue tones). On the other hand, Site 1 in panel (c) displays a more varied distribution, suggesting spatial variability that may be brought about by variations in land use, soil texture, or microtopography. In order to evaluate the accuracy and generalizability of geospatial models, these validations are essential.

By utilizing their complementary advantages, the combined use of SAR (Sentinel-1) and optical data (Sentinel-2) increases the prediction accuracy of soil moisture. In bare or sparsely vegetated areas, Sentinel-1 is especially sensitive to surface soil conditions and offers observations in all weather conditions, day and night (Baghdadi et al., 2018). In contrast, Sentinel-2 indices like NDWI and NDVI capture vegetation and canopy moisture, which indirectly inform about soil conditions beneath (Fernández et al., 2021).

Research has indicated that merging these datasets improves the temporal consistency and robustness of the model. For instance, Bauer-Marschallinger et al. (2018) showed that SM predictions over agricultural lands were greatly enhanced by combining Sentinel-1 backscatter with Sentinel-2 NDVI. Similarly, Veloso et al. (2017) emphasized that SAR backscatter variation correlates with seasonal vegetation and moisture changes, highlighting the dynamic nature of SM retrieval.

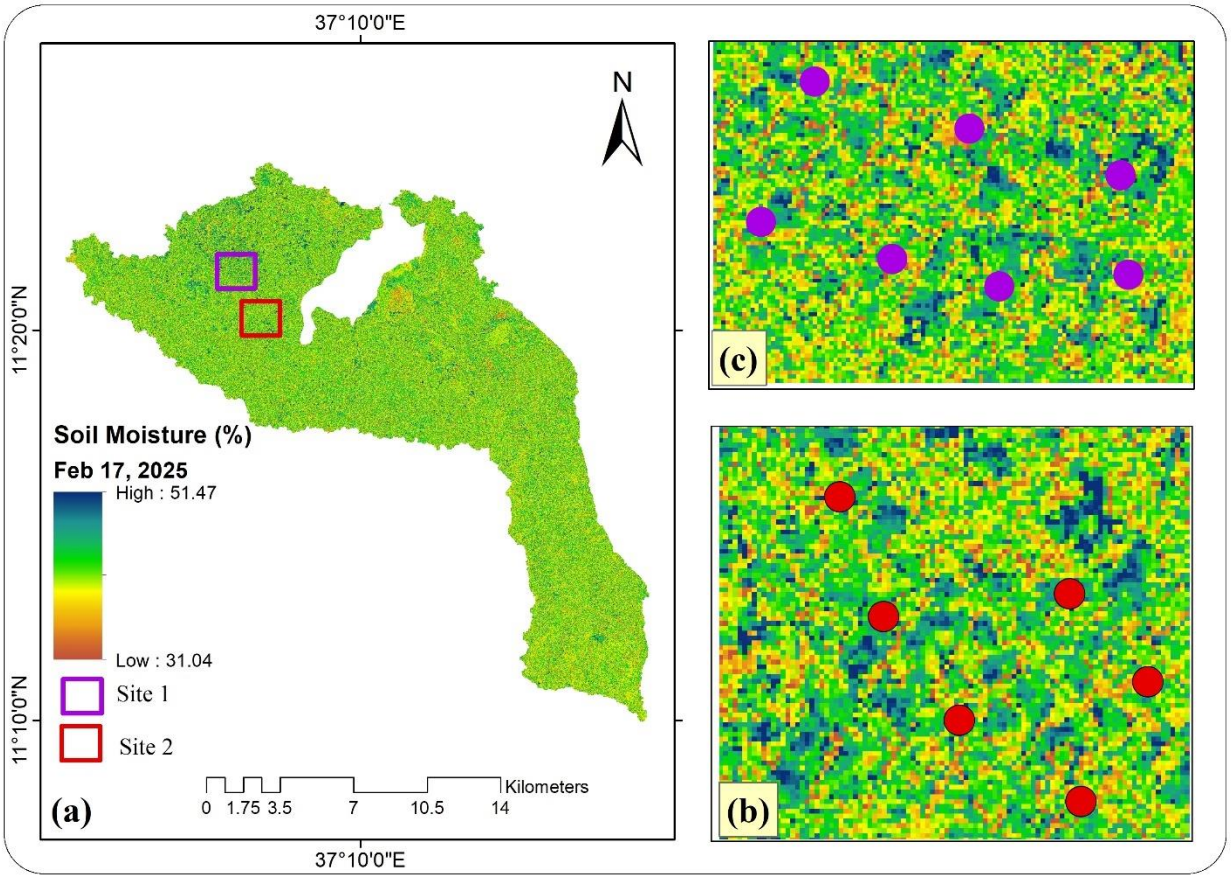


Figure 4-14: Soil Moisture Prediction for the Month of February, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1

Figure 4-15 provides a detailed spatial representation of estimated soil moisture distribution in the Month of March, 2025. Panel (a) depicts a color-coded map of the Koga watershed showing soil moisture levels for March 18, 2025. The data range is from a low of 30.71% to a high of 51.038%, as indicated by the color bar. This spectrum transitions from brown (lowest moisture levels) through yellow and green, to blue (highest moisture levels). The spatial resolution appears to be pixel-based, suggesting satellite-derived or high-resolution model-predicted data. The map reveals a generally moderate soil moisture condition across the watershed, with the central and eastern regions exhibiting relatively higher moisture content (indicated by more green and blue shades), while some peripheral zones show lower values (yellow and brown patches). Two important observation sites are highlighted: Site 1 (purple box) in the northern part and Site 2 (red box) situated more centrally within the watershed. These sites are likely selected for validation, detailed analysis, or field measurement comparisons.

Panel (b) offers a zoomed-in view of Site 2, showcasing the fine-scale variation in predicted soil moisture. Red circular markers indicate specific points either of measurement or of modeled interest. The pixel-level coloration remains consistent with the legend from panel (a), making interpretation straightforward. This area exhibits a mosaic of green and yellow hues, pointing to moderate to slightly above-average soil moisture conditions. Notably, there are very few dark blue patches, implying that high moisture conditions are either rare or absent in this zone. The spatial heterogeneity at this scale suggests influences such as micro-topography, vegetation cover, or land use type affecting water retention and drainage.

Panel (c) focuses on Site 1, delineated with a purple boundary and matching purple dots for specific data points. This site, compared to Site 2, displays a more diverse range of soil moisture values, including several green and yellow pixels, and possibly even some blue areas, suggesting relatively higher soil moisture variability within this localized region. This could be due to differences in land cover types, such as forested versus cultivated areas, or varying soil textures that influence infiltration and retention.

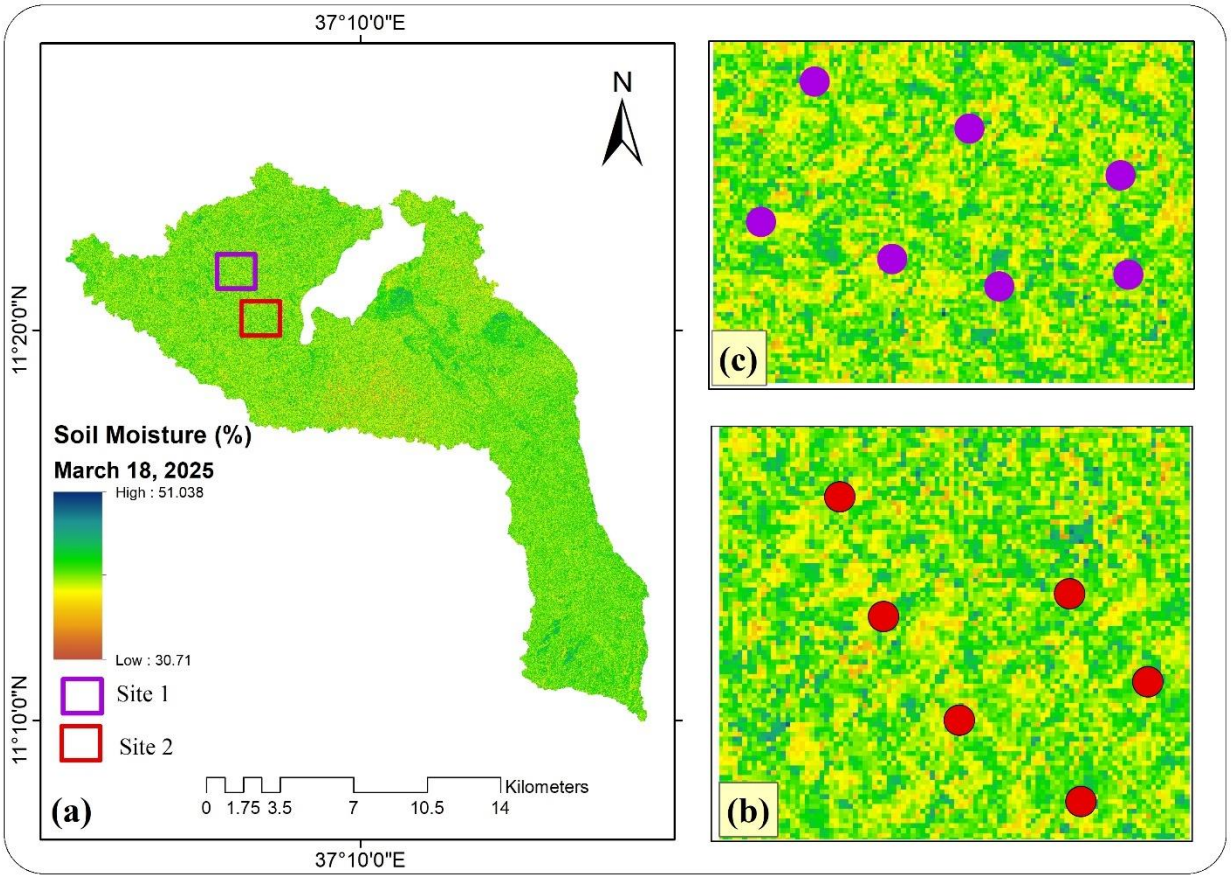


Figure 4-15: Soil Moisture Prediction for the Month of March, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1

Figure 4-16 depicted predicted soil moisture distribution in the Month of April, 2025. Panel (a) illustrates the spatial distribution of soil moisture across the Koga watershed for April 22, 2025, with values ranging from 30.0085% (low) to 52.0654% (high). The color scale follows the same format as previous maps—shifting from brown (dry) through yellow and green to blue (wet). The central and eastern portions of the watershed exhibit broader regions of green to dark green hues, indicating moderate to above-average soil moisture levels. Compared to March, there is a subtle but visible shift toward greener tones, implying slightly wetter conditions, possibly due to early seasonal precipitation or runoff accumulation. The locations of Site 1 (in purple) and Site 2 (in red) are clearly marked, with Site 1 situated in the northern part and Site 2 located slightly to the south.

Panel (b) zooms into Site 2, revealing fine-scale heterogeneity in soil moisture. Red circles highlight specific reference points, likely representing either sampling locations or modeled sub-

grid estimates. The map shows a blend of green and yellow pixels, suggesting moderate moisture conditions with some areas potentially trending toward dryness or transitional water availability. When compared to March 2025 data, Site 2 now contains more green pixels, signifying an improvement in soil moisture. This pattern could be the result of localized infiltration following rainfall or increased soil retention capacity.

Panel (c) focuses on Site 1, marked by purple sampling points. This area displays greater pixel variability, with a mix of yellow, green, and some darker green areas. Compared to March, the increase in dark green tones reflects a more substantial rise in moisture content at this site. This might be due to differences in local land cover, such as forested patches or areas with higher organic matter that retain more water. Site 1 continues to show high spatial variability in SM, emphasizing the need for targeted land and water management.

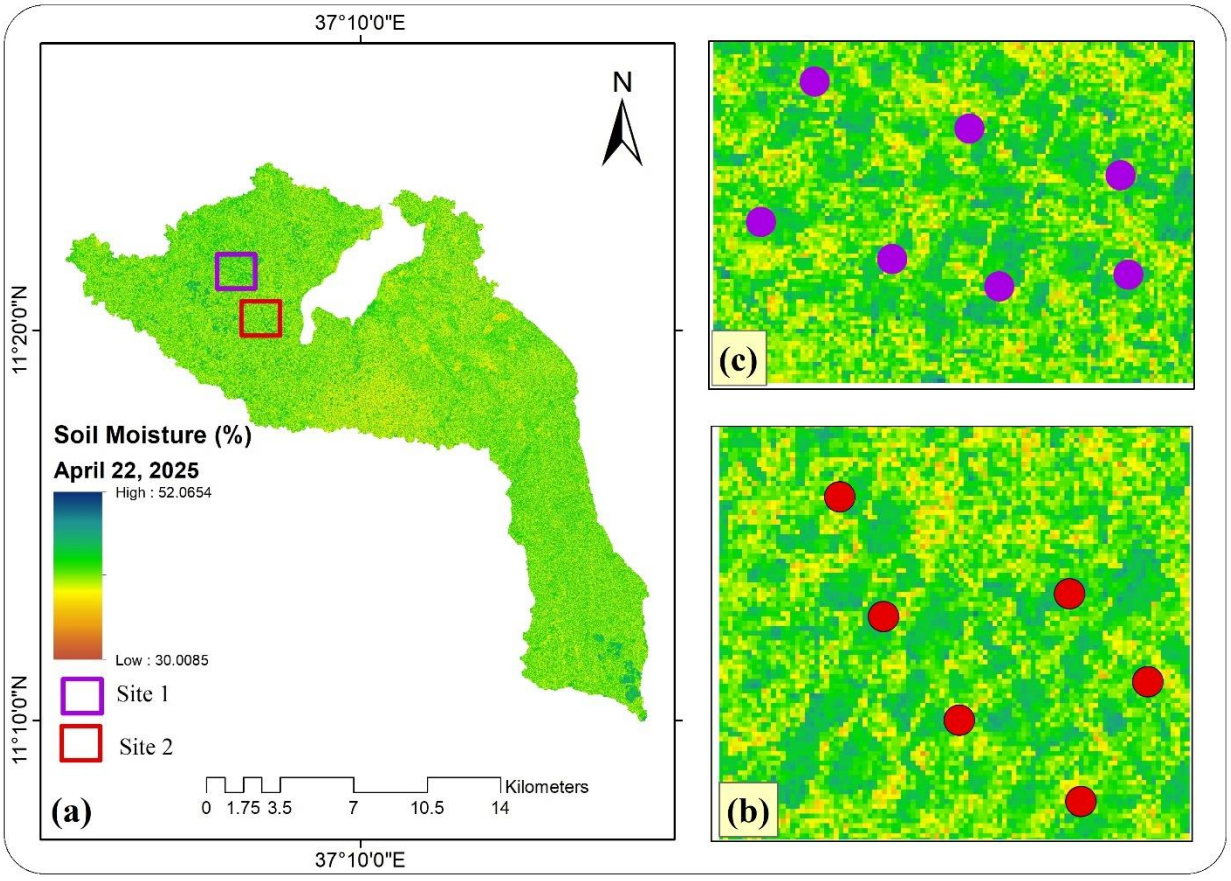


Figure 4-16: Soil Moisture Prediction for the Month of April, 2025; (a) SM at Koga watershed scale; (b) Zoomed SM prediction over Site 2; (c) Zoomed SM prediction over Site 1

4.5.Validations

The validation results for Site one reveal important insights into the performance and limitations of the soil moisture estimation model. The model shows a tendency to slightly underestimate soil moisture content, with a consistently negative bias error ranging from -0.003 to -0.0169 across six samples. Although still quite minor, this underestimation increases with the number of samples, indicating a systematic error that is probably impacted by site-specific factors like soil texture, vegetation cover, or microclimate fluctuations.

Soil moisture estimation based on remote sensing frequently face this kind of negative bias. Studies such as (Lee et al., 2017)and (Wagner et al. 2007), for instance, report comparable underestimations in microwave remote sensing products, which are frequently ascribed to vegetation effects, soil surface roughness, or satellite algorithm calibration problems. Although the model is generally accurate, this persistent underestimation suggests that it might not adequately account for local heterogeneity or the dielectric characteristics of the soil, both of which are essential for precise moisture estimation.

The RMSE values, which range from 0.003 to 0.0116, show the overall error that includes both bias and random noise. Good overall agreement between predicted and measured soil moisture is indicated by these comparatively low RMSE values; however, the increasing trend indicates that prediction errors increase as site conditions or temporal dynamics change. Studies like Dorigo et al. (2015), which emphasize that model performance can deteriorate under complex or variable environmental conditions, particularly when vegetation density or soil texture varies spatially, support this pattern.

By eliminating systematic bias, Unbiased RMSE (UbRMSE) values which, are higher than RMSE and range from 0.006 to 0.0234 isolate the random component of prediction error. The error budget at Site One is significantly impacted by random variability, such as sensor noise, temporal variations in soil moisture, or spatial sampling discrepancies, as evidenced by the fact that UbRMSE is greater than RMSE. This observation is consistent with research by Entekhabi et al. (2010), who highlight that environmental heterogeneity and random errors resulting from sensor limitations frequently account for the majority of soil moisture retrieval uncertainties.

When combined, these measures indicate that although the soil moisture prediction model does a respectable job at Site One, it suffers from systematic underestimation as well as growing random errors over time or samples. In order to address these problems, Kerr et al. (2016) suggest integrating site-specific calibration, adding ancillary data such as vegetation indices or soil texture, and enhancing algorithm robustness to temporal dynamics.

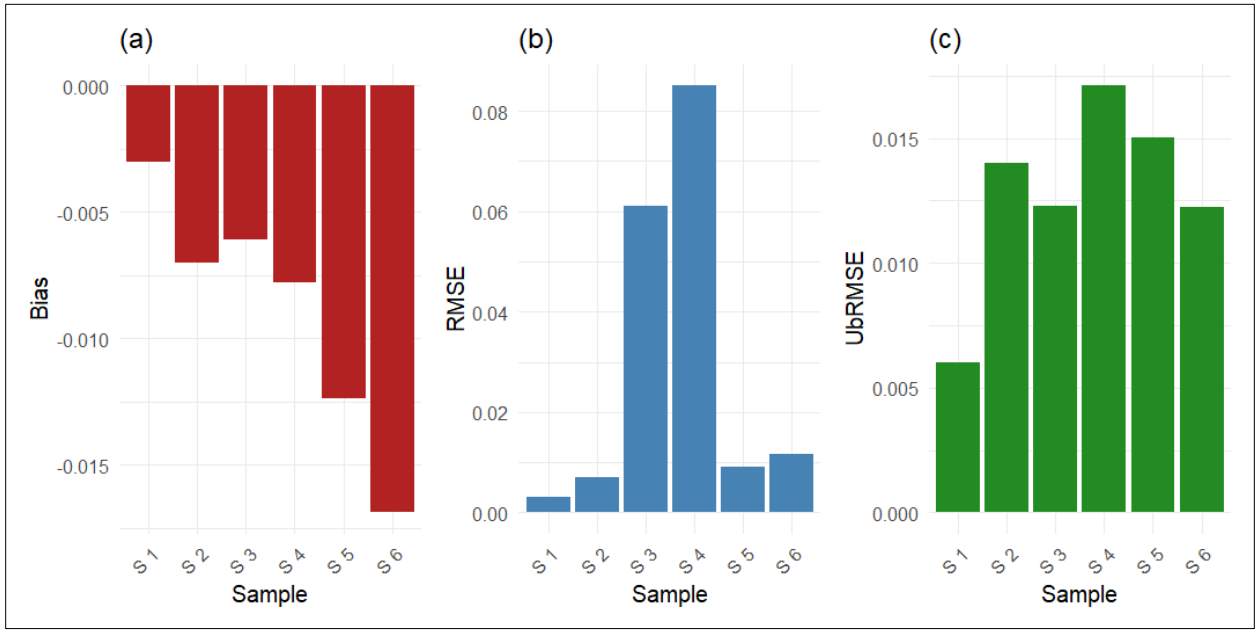


Figure 4-17: Validation of predicted SM over Sit1 (a) Bias, (b) RMSE, (UbRMSE)

When compared to the usual outcomes of remote sensing-based soil moisture estimation, the validation metrics for Site Two show that the soil moisture prediction model operates with a comparatively high degree of accuracy and stability. The bias error varies between -0.011 and +0.0002 for each of the seven samples that were examined. This narrow range with most values close to zero suggests that the model has minimal systematic error at this site. Notably, two samples (sample 2 and sample 7) even show a slight positive bias, indicating a very small overestimation. This minimal and balanced bias suggests that the model is well-calibrated for the conditions at Site Two, a stark contrast to Site One, where consistent underestimation was observed.

This absence of strong systematic bias aligns with findings in several studies, such as those by Albergel et al. (2012) and Chan et al. (2016), which report that soil moisture retrievals are significantly more accurate in areas with relatively homogeneous soil and vegetation conditions. Site Two's results suggest that it may possess such favorable characteristics—either low

vegetation density, uniform surface roughness, or consistent soil texture—contributing to reduced complexity in the retrieval process.

RMSE values at Site Two remain very low throughout the samples, ranging from 0.0002 to 0.011. The extremely low RMSE for samples 2 and 7 (just 0.0002) suggests near-perfect agreement between observed and predicted values. Although sample 3 and 4 display higher RMSE values (~0.011), these are still considered acceptable and well within the error range reported in many validation studies. For example, Dorigo et al. (2010), in the validation of various remote sensing soil moisture datasets, observed RMSE values between 0.02–0.05 across different European sites. The relatively low RMSE at Site Two indicates the model's robustness and reliability under the specific site conditions.

The Unbiased RMSE (UbRMSE) values further support this observation. They range from 0.0004 to 0.022 and show a consistent pattern with the RMSE values, meaning that most of the error at Site Two is due to random factors rather than systematic model issues. Even in samples with slightly higher UbRMSE (such as sample 3 and 4), the magnitude remains low, pointing to stable performance. According to Crow et al. (2012), UbRMSE is an essential metric for isolating the stochastic components of model performance, and low values typically indicate reduced impact from transient or unmodeled environmental noise.

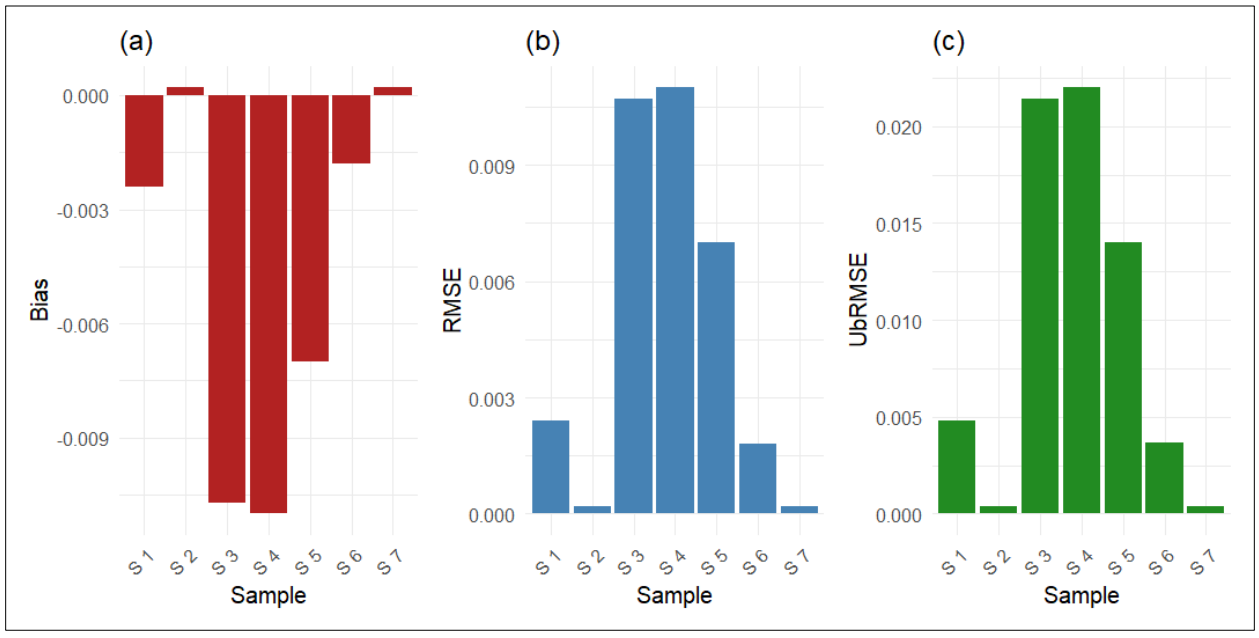


Figure 4-18: Validation of predicted SM over Sit2 (a) Bias, (b) RMSE, (UbRMSE)

The scatterplot illustrates the relationship between in-situ soil moisture measurements and predicted soil moisture values derived from Sentinel data, both expressed as percentages. Each point represents a matched pair of observations—one from ground-based instruments and the other from a satellite-derived model. Ideally, perfect prediction would place all points along the 1:1 dashed line, where predicted values exactly equal the in-situ measurements.

Visually, the majority of the data points cluster tightly around the 1:1 line, suggesting strong agreement between measured and predicted values. The linear trend line fitted to the data (in dark red) closely aligns with the 1:1 line, indicating that the model is generally accurate in its estimations. This visual impression is supported by the statistical metrics annotated on the plot: an R^2 value of 0.94 shows that 94% of the variation in the predicted soil moisture can be explained by the variation in the observed in-situ values, reflecting a very high degree of correlation and model reliability.

The average size of the prediction error is measured by the RMSE (Root Mean Square Error) of 2.01%. In soil moisture estimation, where errors under 5% are frequently regarded as acceptable for remote sensing applications (e.g., Dorigo et al., 2010; Entekhabi et al., 2010), an RMSE of about 2% is comparatively low. This suggests that the model's predictions are accurate in magnitude in addition to being correlated with ground truth.

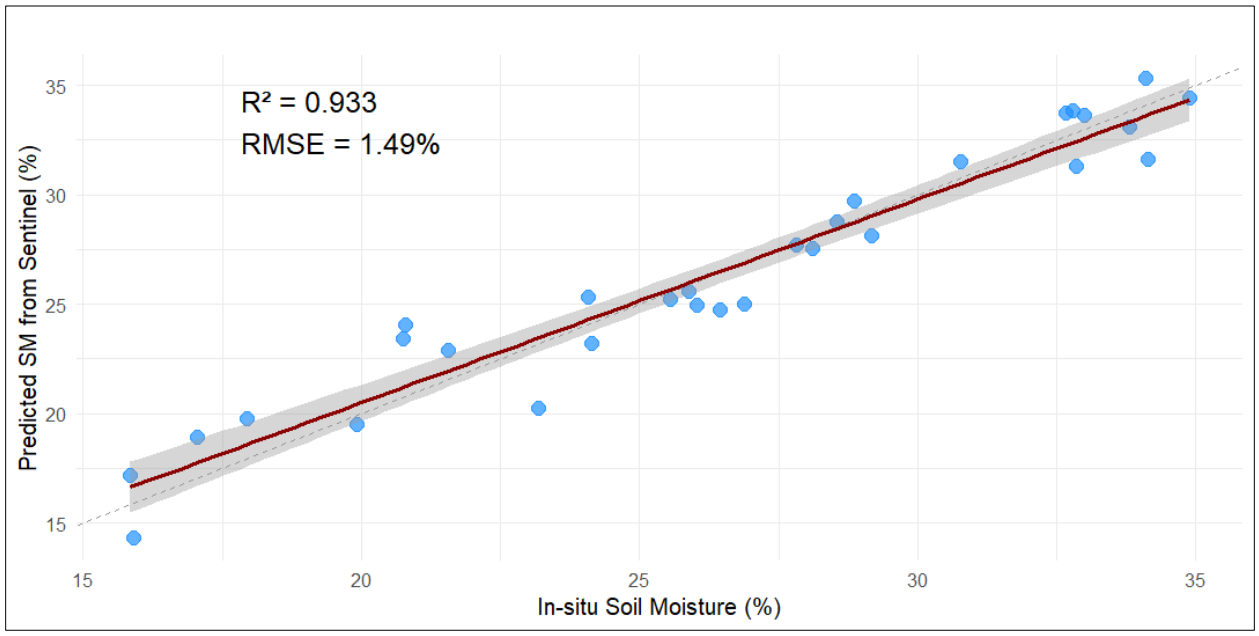


Figure 4-19: Scatter plot between predicted and In-situ soil moisture

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1.Conclusions

This study showed that the accuracy of estimating soil moisture throughout the Koga watershed is greatly increased by combining Sentinel-1A SAR and Sentinel-2 MSI data. Important information about surface conditions was obtained from Sentinel-1 VV and VH backscatter data. VV was especially sensitive to surface roughness and soil dielectric characteristics, whereas VH was more sensitive to vegetation structure and volume scattering. According to earlier research that highlighted SAR's resilience in such circumstances, these radar-derived metrics performed best during the dry season when there was little vegetation cover. Furthermore, by accounting for possible overestimations brought on by terrain-induced scattering, surface roughness was added as a predictor variable to increase the accuracy of moisture estimates.

Sentinel-2 optical indices like NDVI, SAVI, and NDRE gained traction in the soil moisture prediction model as the season went on and vegetation started to reestablish. Indirectly representing root-zone soil moisture levels, these indices measured vegetation vigor and chlorophyll content. Significant contributions were also made by water-sensitive indices such as NDWI and MNDWI, particularly in identifying saturated soils and shallow surface water. The necessity of seasonally adaptive modeling is demonstrated by the change in variable importance from SAR metrics in February to optical indices in March and April. This dynamic approach improves the spatial and temporal resolution of soil moisture predictions by performing better under different land cover and moisture conditions.

According to model validation, the Random Forest model did fairly well, with wetter months showing an improvement in prediction accuracy. Site Two showed little bias and low RMSE, confirming the model's strength under more uniform surface conditions, whereas Site One had a slight underestimation bias because of the varied terrain and vegetation. The efficacy of multi-sensor integration for soil moisture estimation was confirmed by the overall model's high R2 value of 0.94 against in-situ measurements. These results underscore the potential of using Sentinel-based remote sensing in conjunction with machine learning for water resource monitoring, agricultural planning, and land management in data-scarce environments.

5.2.Recommendations

- Future soil moisture monitoring efforts should continue to integrate Sentinel-1 SAR and Sentinel-2 MSI data. The complementary strengths of radar (all-weather, surface sensitivity) and optical (vegetation and water indices) data provide a more holistic understanding of surface moisture dynamics. This approach is especially valuable in heterogeneous agricultural and watershed environments like Koga.
- Incorporating detailed surface roughness data and land use classifications can mitigate misinterpretations of radar backscatter signals. Since roughness influences backscatter independently of moisture, models should use ancillary data (e.g., DEMs, land cover maps) or in-situ calibration to better isolate moisture signals from terrain or tillage effects.
- To improve model validation and reduce bias, especially in complex terrain, expanding ground-based soil moisture measurement networks is essential. More diverse and frequent in-situ data collection will enhance the calibration of remote sensing models and improve their performance across various land cover types and microclimatic zones.
- For land use planners and watershed managers: soil moisture maps generated from remote sensing can support better irrigation scheduling, drought preparedness, and flood risk reduction strategies. Planners should integrate these datasets into land and water resource management frameworks to optimize resource allocation.
- For agronomists and farmers: integrating soil moisture monitoring into agricultural decision-making can improve crop yield by guiding planting dates, irrigation frequency, and fertilizer application. This can also reduce water wastage and minimize crop failure during dry spells.
- For policymakers and development agencies: supporting investments in satellite-based soil moisture monitoring, combined with local capacity building, will strengthen climate adaptation strategies.

Reference

Abowarda, A. S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., & Sun, Z. (2021). Generating Surface Soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sensing of Environment*, *255*, 112301. https://doi.org/10.1016/J.RSE.2021.112301

Ahmad, I., Ahmad, B., Boote, K., & Hoogenboom, G. (2020). Adaptation strategies for maize production under climate change for semi-arid environments. *European Journal of Agronomy*, *115*. https://doi.org/10.1016/j.eja.2020.126040

Ahmad, U., & Sharma, L. (2023). A review of best management practices for potato crop using Precision Agricultural Technologies. In *Smart Agricultural Technology* (Vol. 4). Elsevier B.V. https://doi.org/10.1016/j.atech.2023.100220

Alexakis, D., Hadjimitsis, D., Agapiou, A., Themistocleous, K., & Retalis, A. (2011). Monitoring urban land cover with the use of satellite remote sensing techniques as a means of flood risk assessment in Cyprus. *Sensors, Systems, and Next-Generation Satellites XV*, *8176*, 81761Z. https://doi.org/10.1117/12.898185

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. In *Remote Sensing* (Vol. 7, Issue 12, pp. 16398–16421). MDPI AG. https://doi.org/10.3390/rs71215841

Ali, M. H., & Talukder, M. S. U. (2008). Increasing water productivity in crop production-A synthesis. In *Agricultural Water Management* (Vol. 95, Issue 11, pp. 1201–1213). https://doi.org/10.1016/j.agwat.2008.06.008

Anzalone, A., Pagliaro, A., & Tutone, A. (2024). An Introduction to Machine and Deep Learning Methods for Cloud Masking Applications. In *Applied Sciences (Switzerland)* (Vol. 14, Issue 7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/app14072887

Ayehu, G. T., Tadesse, T., Gessesse, B., & Dinku, T. (2018). Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. *Atmospheric Measurement Techniques*, *11*(4), 1921–1936. https://doi.org/10.5194/amt-11-1921-2018

Ayehu, G., Tadesse, T., Gessesse, B., Yigrem, Y., & M.melesse, A. (2020). Combined use of sentinel-1 sar and landsat sensors products for residual soil moisture retrieval over agricultural fields in the upper blue nile basin, ethiopia. *Sensors (Switzerland)*, *20*(11), 1–25. https://doi.org/10.3390/s20113282

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. K., & Tuller, M. (2021). Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Remote Sensing of Environment*, *260*, 112434. https://doi.org/10.1016/J.RSE.2021.112434

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. In *Reviews of Geophysics* (Vol. 57, Issue 2, pp. 530–616). Blackwell Publishing Ltd. https://doi.org/10.1029/2018RG000618

Baghdadi, N., Cresson, R., Hajj, M. El, Ludwig, R., & Jeunesse, I. La. (2012). Soil parameters estimation over bare agriculture areas Hydrology and Earth System Sciences Discussions Soil parameters estimation over bare agriculture areas from C-band polarimetric SAR data using neural networks Soil parameters estimation over bare agriculture areas. *Hydrol. Earth Syst. Sci. Discuss*, *9*, 2897–2933. https://doi.org/10.5194/hessd-9-2897-2012

Baghdadi, N., Holah, N., & Zribi, M. (2006). Calibration of the Integral Equation Model for SAR data in C-band and HH and VV polarizations. *International Journal of Remote Sensing*, *27*(4), 805–816. https://doi.org/10.1080/01431160500212278

Bai, J., Cui, Q., Zhang, W., & Meng, L. (2019). An approach for downscaling SMAP soil moisture by combining Sentinel-1 SAR and MODIS data. *Remote Sensing*, *11*(23), 1–20. https://doi.org/10.3390/rs11232736

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., & Wagner, W. (2019). Toward Global Soil Moisture Monitoring with Sentinel-1: Harnessing Assets and Overcoming Obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 520–539. https://doi.org/10.1109/TGRS.2018.2858004

Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., Minh, D. H. T., Ndikumana, E., Courault, D., & Belhouchette, H. (2019). Mapping paddy rice using Sentinel-1 SAR time series in Camargue, France. In *Remote Sensing* (Vol. 11, Issue 7). MDPI AG. https://doi.org/10.3390/RS11070887

Bhattacharya, A. (2021). Soil Water Deficit and Physiological Issues in Plants. In *Soil Water Deficit and Physiological Issues in Plants*. Springer Singapore. https://doi.org/10.1007/978-981-33-6276-5

Blanco-Canqui, H., & Ruis, S. J. (2018). No-tillage and soil physical environment. *Geoderma*, *326*, 164–200. https://doi.org/10.1016/j.geoderma.2018.03.011

Bodner, G., Nakhforoosh, A., & Kaul, H. P. (2015). Management of crop water under drought: a review. In *Agronomy for Sustainable Development* (Vol. 35, Issue 2, pp. 401–442). Springer-Verlag France. https://doi.org/10.1007/s13593-015-0283-4

Bogena, H. R., Huisman, J. A., Güntner, A., Hübner, C., Kusche, J., Jonard, F., Vey, S., & Vereecken, H. (2015). Emerging methods for noninvasive sensing of soil moisture dynamics from field to catchment scale: a review. In *Wiley Interdisciplinary Reviews: Water* (Vol. 2, Issue 6, pp. 635–647). John Wiley and Sons Inc. https://doi.org/10.1002/WAT2.1097

Bousbih, S., Zribi, M., Hajj, M. El, Baghdadi, N., Lili-Chabaane, Z., Gao, Q., & Fanise, P. (2018). Soil moisture and irrigation mapping in a semi-arid region, based on the synergetic use of Sentinel-1 and Sentinel-2 data. *Remote Sensing*, *10*(12). https://doi.org/10.3390/rs10121953

Breiman. (2001). Random forests. *Random Forests*, 5–32. https://doi.org/10.1201/9780429469275-8

Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2010). Spatial-temporal variability of soil moisture and its estimation across scales. *Water Resources Research*, *46*(2), 1–14. https://doi.org/10.1029/2009WR008016

Busscher, W. J. (2009). Field Estimation of Soil Water Content: A review. *Journal of Soil and Water Conservation*, *64*(4), 116A-116A. https://doi.org/10.2489/jswc.64.4.116a

Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance*, *61*, 101649. https://doi.org/10.1016/J.RIBAF.2022.101649

Černý, R. (2009). Time-domain reflectometry method and its application for measuring moisture content in porous materials: A review. In *Measurement: Journal of the International Measurement Confederation* (Vol. 42, Issue 3, pp. 329–336). https://doi.org/10.1016/j.measurement.2008.08.011

Chakrabarti, S., Bongiovanni, T., Judge, J., Zotarelli, L., & Bayer, C. (2014). Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(9), 3867–3879. https://doi.org/10.1109/JSTARS.2014.2315999

Chen, Q., Miao, F., Wang, H., Xu, Z. X., Tang, Z., Yang, L., & Qi, S. (2020). Downscaling of Satellite Remote Sensing Soil Moisture Products Over the Tibetan Plateau Based on the Random Forest Algorithm: Preliminary Results. *Earth and Space Science*, *7*(6). https://doi.org/10.1029/2020EA001265

Chukalla, A. D., Krol, M. S., & Hoekstra, A. Y. (2017). Marginal cost curves for water footprint reduction in irrigated agriculture: Guiding a cost-effective reduction of crop water consumption to a permit or benchmark level. *Hydrology and Earth System Sciences*, *21*(7), 3507–3524. https://doi.org/10.5194/hess-21-3507-2017

Cosh, M. H., Jackson, T. J., Bindlish, R., & Prueger, J. H. (2004). Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment*, *92*(4), 427–435. https://doi.org/10.1016/J.RSE.2004.02.016

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., De Rosnay, P., Ryu, D., & Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000372

Cutler, A. (2012). Ensemble Machine Learning. In *Ensemble Machine Learning* (2011th ed.). Springer International Publishing. https://doi.org/10.1007/978-1-4419-9326-7

da Silveira, H. L. F., Galvão, L. S., Sanches, I. D. A., de Sá, I. B., & Taura, T. A. (2018). Use of MSI/Sentinel-2 and airborne LiDAR data for mapping vegetation and studying the relationships with soil attributes

in the Brazilian semi-arid region. *International Journal of Applied Earth Observation and Geoinformation*, *73*, 179–190. https://doi.org/10.1016/J.JAG.2018.06.016

Davidson, M. W. J., Mattia, F., Satalino, G., Verhoest, N. E. C., Le Toan, T., Borgeaud, M., Louis, J. M. B., & Attema, E. (2003). Joint statistical properties of RMS height and correlation length derived from multisite 1-m roughness measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(7 PART I), 1651–1658. https://doi.org/10.1109/TGRS.2003.813361

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Van Oevelen, P., Robock, A., & Jackson, T. (2011). The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5), 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/J.RSE.2011.11.026

Duveiller, G., & Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, *114*(11), 2637–2650. https://doi.org/10.1016/J.RSE.2010.06.001

El Hajj, M., Baghdadi, N., Zribi, M., Belaud, G., Cheviron, B., Courault, D., & Charron, F. (2016). Soil moisture retrieval over irrigated grassland using X-band SAR data. *Remote Sensing of Environment*, *176*, 202–218. https://doi.org/10.1016/J.RSE.2016.01.027

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Eskandari, H., Saadatmand, H., Ramzan, M., & Mousapour, M. (2024). Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning. *Applied Energy*, *366*, 123314. https://doi.org/10.1016/J.APENERGY.2024.123314

Evett, S. R., Schwartz, R. C., Casanova, J. J., & Heng, L. K. (2012). Soil water sensing for water balance, ET and WUE. In *Agricultural Water Management* (Vol. 104, pp. 1–9). https://doi.org/10.1016/j.agwat.2011.12.002

Fageria, N. K., Baligar, V. C., & Li, Y. C. (2008). The role of nutrient efficient plants in improving crop yields in the twenty first century. In *Journal of Plant Nutrition* (Vol. 31, Issue 6, pp. 1121–1157). https://doi.org/10.1080/01904160802116068

Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., & Jackson, T. J. (2008). Field observations of soil moisture variability across scales. *Water Resources Research*, *44*(1), 1–16. https://doi.org/10.1029/2006WR005804

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, *30*(1), 27–38. https://doi.org/10.1016/J.PATREC.2008.08.010

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., & Hostert, P. (2021). National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, *252*, 112128. https://doi.org/10.1016/J.RSE.2020.112128

Gao, J., & Zhang, L. (2014). *Electricity Consumption-Economic Growth-CO 2 Emissions Nexus in Sub-Saharan Africa: Evidence from Panel Cointegration*.

Gao, X., & Schlosser, C. A. (2019). Mid-Western US heavy summer-precipitation in regional and global climate models: the impact on model skill and consensus through an analogue lens. *Climate Dynamics*, *52*(3–4), 1569–1582. https://doi.org/10.1007/s00382-018-4209-0

Gavrilescu, M. (2021). Water, soil, and plants interactions in a threatened environment. In *Water (Switzerland)* (Vol. 13, Issue 19). MDPI. https://doi.org/10.3390/w13192746

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., … Wagner, W. (2020). Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sensing of Environment*, *244*, 111806. https://doi.org/10.1016/J.RSE.2020.111806

Hajj, M. El, Baghdadi, N., Zribi, M., & Bazzi, H. (2017). Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sensing*, *9*(12). https://doi.org/10.3390/rs9121292

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*(33), 7561–7578. https://doi.org/10.1016/J.ATMOSENV.2008.05.057

Hulley, G. C., & Hook, S. J. (2009). Intercomparison of versions 4, 4.1 and 5 of the MODIS Land Surface Temperature and Emissivity products and validation with laboratory measurements of sand samples from the Namib desert, Namibia. *Remote Sensing of Environment*, *113*(6), 1313–1318. https://doi.org/10.1016/j.rse.2009.02.018

Ienco, D., Interdonato, R., Gaetano, R., & Ho Tong Minh, D. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 11–22. https://doi.org/10.1016/J.ISPRSJPRS.2019.09.016

Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, *75*(15). https://doi.org/10.1007/s12665-016-5917-6

Jason P. Ackerson. (2018). Soil Sampling Guidelines. *Purdue University Cooperative Extension Service*, 1–4.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., & Wood, E. F. (2017). Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Advances in Water Resources*, *109*, 106–120. https://doi.org/10.1016/J.ADVWATRES.2017.09.006

Kelley, C. P., Mohtadi, S., Cane, M. A., Seager, R., & Kushnir, Y. (2015). Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3241–3246. https://doi.org/10.1073/pnas.1421533112

Kornelsen, K. C., & Coulibaly, P. (2013). Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. *Journal of Hydrology*, *476*, 460–489. https://doi.org/10.1016/J.JHYDROL.2012.10.044

Li, J., Wang, Z., Wu, X., Zscheischler, J., Guo, S., & Chen, X. (2021). A standardized index for assessing sub-monthly compound dry and hot conditions with application in China. *Hydrology and Earth System Sciences*, *25*(3), 1587–1601. https://doi.org/10.5194/hess-25-1587-2021

Li, W., Wang, D., Liu, S., & Zhu, Y. (2019). Measuring urbanization-occupation and internal conversion of peri-urban cultivated land to determine changes in the peri-urban agriculture of the black soil region. *Ecological Indicators*, *102*, 328–337. https://doi.org/10.1016/j.ecolind.2019.02.055

Li, Y. Y., Zhao, K., Ren, J. H., Ding, Y. L., & Wu, L. L. (2014). Analysis of the dielectric constant of saline-alkali soils and the effect on radar backscattering coefficient: A case study of soda alkaline saline soils in western Jilin province using RADARSAT-2 data. *Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/563015

Li, Z., Li, Q., Wang, J., Feng, Y., & Shao, Q. (2020). Impacts of projected climate change on runoff in upper reach of Heihe River basin using climate elasticity method and GCMs. *Science of The Total Environment*, *716*, 137072. https://doi.org/10.1016/J.SCITOTENV.2020.137072

Liang, J., Liang, G., Zhao, Y., & Zhang, Y. (2021). A synergic method of Sentinel-1 and Sentinel-2 images for retrieving soil moisture content in agricultural regions. *Computers and Electronics in Agriculture*, *190*, 106485. https://doi.org/10.1016/J.COMPAG.2021.106485

Lisboa, P. J. G., Saralajew, S., Vellido, A., Fernández-Domenech, R., & Villmann, T. (2023). The coming of age of interpretable and explainable machine learning models. *Neurocomputing*, *535*, 25–39. https://doi.org/10.1016/J.NEUCOM.2023.02.040

Liu, D., Yu, Z. B., & Lü, H. S. (2010). Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. *Water Science and Engineering*, *3*(4), 361–377. https://doi.org/10.3882/j.issn.1674-2370.2010.04.001

Liu, J. (2019). *Downscaling and Validating Satellite-Based Soil Moisture Products Over the Maasai Mara Catchment in Kenya* [Thesis]. Unpulished Thesis, University of twenty Faculity of Geo-information science and earth observation.

Liu, L., Zhang, R., & Zuo, Z. (2016). The relationship between soil moisture and LAI in different types of soil in central eastern China. *Journal of Hydrometeorology*, *17*(11), 2733–2742. https://doi.org/10.1175/JHM-D-15-0240.1

Liu, Y., Jing, W., Wang, Q., & Xia, X. (2020). Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Advances in Water Resources*, *141*, 103601. https://doi.org/10.1016/j.advwatres.2020.103601

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

Loew, A., Ludwig, R., & Mauser, W. (2006). Derivation of surface soil moisture from ENVISAT ASAR wide swath and image mode data in agricultural areas. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(4), 889–898. https://doi.org/10.1109/TGRS.2005.863858

Lwakatare, L. E., Raj, A., Crnkovic, I., Bosch, J., & Olsson, H. H. (2020). Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology*, *127*, 106368. https://doi.org/10.1016/J.INFSOF.2020.106368

Lynch, J., Marschner, P., & Rengel, Z. (2012). Effect of Internal and External Factors on Root Growth and Development. *Marschner's Mineral Nutrition of Higher Plants: Third Edition*, 331–346. https://doi.org/10.1016/B978-0-12-384905-2.00013-3

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 152, pp. 166–177). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/J.JKSUCI.2023.01.014

Morales, Y., Querales, M., Rosas, H., Allende-Cid, H., & Salas, R. (2021). A self-identification Neuro-Fuzzy inference framework for modeling rainfall-runoff in a Chilean watershed. *Journal of Hydrology*, *594*, 125910. https://doi.org/10.1016/J.JHYDROL.2020.125910

Narvekar, P. S., Entekhabi, D., Kim, S. B., & Njoku, E. G. (2015). Soil moisture retrieval using L-band radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(6), 3492–3506. https://doi.org/10.1109/TGRS.2014.2377714

Ning, J., Yao, Y., Tang, Q., Li, Y., Fisher, J. B., Zhang, X., Jia, K., Xu, J., Shang, K., Yang, J., Yu, R., Liu, L., Zhang, X., Xie, Z., & Fan, J. (2023). Soil moisture at 30 m from multiple satellite datasets fused by random forest. *Journal of Hydrology*, *625*, 130010. https://doi.org/10.1016/J.JHYDROL.2023.130010

Oh, Y. (2004a). Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(3), 596–601. https://doi.org/10.1109/TGRS.2003.821065

Oh, Y. (2004b). Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(3), 596–601. https://doi.org/10.1109/TGRS.2003.821065

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., Crow, W. T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M. W. J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y. H., Lovergine, F., Mahecha, M. D., Marzahn, P., … Loew, A. (2021). A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements. *Remote Sensing of Environment*, *252*, 112162. https://doi.org/10.1016/J.RSE.2020.112162

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C*, *83–84*, 36–56. https://doi.org/10.1016/J.PCE.2015.02.009

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. In *Remote Sensing* (Vol. 12, Issue 14). MDPI AG. https://doi.org/10.3390/rs12142291

Piles, M., Sánchez, N., Vall-Llossera, M., Camps, A., Martínez-Fernandez, J., Martinez, J., & Gonzalez-Gambau, V. (2014). A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian peninsula. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(9), 3845–3857. https://doi.org/10.1109/JSTARS.2014.2325398

Pino, V., McBratney, A., O'Brien, E., Singh, K., & Pozza, L. (2022). Citizen science & soil connectivity: Where are we? *Soil Security*, *9*, 100073. https://doi.org/10.1016/J.SOISEC.2022.100073

Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience and Remote Sensing*, 735–752. https://doi.org/10.1080/15481603.2020.1794104

Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking and Reasoning*, *24*(2), 258–279. https://doi.org/10.1080/13546783.2017.1387606

Rajib, M. A., Merwade, V., & Yu, Z. (2016). Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*, *536*, 192–207. https://doi.org/10.1016/J.JHYDROL.2016.02.037

Rasheed, M. W., Tang, J., Sarwar, A., Shah, S., Saddique, N., Khan, M. U., Imran Khan, M., Nawaz, S., Shamshiri, R. R., Aziz, M., & Sultan, M. (2022). Soil Moisture Measuring Techniques and Factors Affecting the Moisture Dynamics: A Comprehensive Review. In *Sustainability (Switzerland)* (Vol. 14, Issue 18). MDPI. https://doi.org/10.3390/su141811538

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., & Wendroth, O. (2008). Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zone Journal*, *7*(1), 358–389. https://doi.org/10.2136/vzj2007.0143

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, 804–818. https://doi.org/10.1016/J.OREGEOREV.2015.01.001

Romano, N. (2014). Soil moisture at local scale: Measurements and simulations. *Journal of Hydrology*, *516*, 6–20. https://doi.org/10.1016/j.jhydrol.2014.01.026

Sadeghi, M., Babaeian, E., Tuller, M., & Jones, S. B. (2017). The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sensing of Environment*, *198*, 52–68. https://doi.org/10.1016/j.rse.2017.05.041

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & Adams, C. (2020). Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, *205*, 103187. https://doi.org/10.1016/J.EARSCIREV.2020.103187

Salcedo-Sanz, S., Ghamisi, P., Piles, M., Werner, M., Cuadra, L., Moreno-Martínez, A., Izquierdo-Verdiguier, E., Muñoz-Marí, J., Mosavi, A., & Camps-Valls, G. (2020). Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. *Information Fusion*, *63*, 256–272. https://doi.org/10.1016/J.INFFUS.2020.07.004

Şekertekin, A., Marangoz, A. M., & Abdikan, S. (2018). Soil Moisture Mapping Using Sentinel-1A Synthetic Aperture Radar Data. *International Journal of Environment and Geoinformatics*, *5*(2), 178–188. https://doi.org/10.30897/ijegeo.425606

Sekertekin, A., Marangoz, A. M., & Abdikan, S. (2020). ALOS-2 and Sentinel-1 SAR data sensitivity analysis to surface soil moisture over bare and vegetated agricultural fields. *Computers and Electronics in Agriculture*, *171*. https://doi.org/10.1016/j.compag.2020.105303

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/J.EARSCIREV.2010.02.004

Seraj, A., Mohammadi-Khanaposhtani, M., Daneshfar, R., Naseri, M., Esmaeili, M., Baghban, A., Habibzadeh, S., & Eslamian, S. (2023). Cross-validation. *Handbook of HydroInformatics: Volume I: Classic Soft-Computing Techniques*, 89–105. https://doi.org/10.1016/B978-0-12-821285-1.00021-X

Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., & Verbist, K. (2018). Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions. In *Water Resources Research* (Vol. 54, Issue 12, pp. 9724–9758). Blackwell Publishing Ltd. https://doi.org/10.1029/2017WR022437

Singh, A., Gaurav, K., Sonkar, G. K., & Lee, C. C. (2023). Strategies to Measure Soil Moisture Using Traditional Methods, Automated Sensors, Remote Sensing, and Machine Learning Techniques: Review, Bibliometric Analysis, Applications, Research Findings, and Future Directions. *IEEE Access*, *11*, 13605–13635. https://doi.org/10.1109/ACCESS.2023.3243635

Singh, G., Das, N. N., Panda, R. K., Colliander, A., Jackson, T. J., Mohanty, B. P., Entekhabi, D., & Yueh, S. H. (2019). Validation of SMAP Soil Moisture Products Using Ground-Based Observations for the Paddy Dominated Tropical Region of India. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 8479–8491. https://doi.org/10.1109/TGRS.2019.2921333

Singh, R., & Singh, G. S. (2017). Traditional agriculture: a climate-smart approach for sustainable food production. In *Energy, Ecology and Environment* (Vol. 2, Issue 5, pp. 296–316). Joint Center on Global Change and Earth System Science of the University of Maryland and Beijing Normal University. https://doi.org/10.1007/s40974-017-0074-7

Sudmanns, M., Tiede, D., Augustin, H., & Lang, S. (2020). Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. *International Journal of Digital Earth*, *13*(7), 768–784. https://doi.org/10.1080/17538947.2019.1572799

Tapiador, F. J., Navarro, A., Levizzani, V., García-Ortega, E., Huffman, G. J., Kidd, C., Kucera, P. A., Kummerow, C. D., Masunaga, H., Petersen, W. A., Roca, R., Sánchez, J. L., Tao, W. K., & Turk, F. J. (2017). Global precipitation measurements for validating climate models. *Atmospheric Research*, *197*, 1–20. https://doi.org/10.1016/J.ATMOSRES.2017.06.021

Tong, S. S., Deroin, J. P., Pham, T. L., & Cao, X. C. (2018). Estimation of Surface Parameters of Tidal Flats Using Sentinel-1A SAR Data in the Northern Coast of Vietnam. *Advances and Applications in*

Geospatial Technology and Earth Resources, *2*, 65–88. https://doi.org/10.1007/978-3-319-68240-2_5

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B. Ö., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., … Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, *120*, 9–24. https://doi.org/10.1016/J.RSE.2011.05.028

Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, *49*, 560–567. https://doi.org/10.1016/J.ENBUILD.2012.03.003

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179. https://doi.org/10.1016/J.NEUROIMAGE.2016.10.038

Vereecken, H., Huisman, J. A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogena, H., Weihermüller, L., Herbst, M., Martinez, G., & Vanderborght, J. (2014). On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology*, *516*, 76–96. https://doi.org/10.1016/j.jhydrol.2013.11.061

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273–290. https://doi.org/10.1016/J.ISPRSJPRS.2015.05.005

Wang, J., Bretz, M., Dewan, M. A. A., & Delavar, M. A. (2022). Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Science of The Total Environment*, *822*, 153559. https://doi.org/10.1016/J.SCITOTENV.2022.153559

Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China*, *3*(2), 237–247. https://doi.org/10.1007/s11707-009-0023-7

Waraich, E. A., Ahmad, R., Yaseen Ashraf, M., Saifullah, S., & Ahmad, M. (2011). Improving agricultural water use efficiency by nutrient management in crop plants. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, *61*(4), 291–304. https://doi.org/10.1080/09064710.2010.491954

Whitcomb, J., Clewley, D., Colliander, A., Cosh, M. H., Powers, J., Friesen, M., McNairn, H., Berg, A. A., Bosch, D. D., Coffin, A., Collins, C. H., Prueger, J. H., Entekhabi, D., & Moghaddam, M. (2020). Evaluation of SMAP Core Validation Site Representativeness Errors Using Dense Networks of in Situ Sensors and Random Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6457–6472. https://doi.org/10.1109/JSTARS.2020.3033591

Widodo, A., & Yang, B. S. (2007). Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors. *Expert Systems with Applications*, *33*(1), 241–250. https://doi.org/10.1016/J.ESWA.2006.04.020

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846. https://doi.org/10.1016/J.PATCOG.2015.03.009

Wu, X., Xiao, Q., Wen, J., You, D., & Hueni, A. (2019). Advances in quantitative remote sensing product validation: Overview and current status. *Earth-Science Reviews*, *196*, 102875. https://doi.org/10.1016/J.EARSCIREV.2019.102875

Xu, Y., Kohtz, S., Boakye, J., Gardoni, P., & Wang, P. (2023). Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliability Engineering & System Safety*, *230*, 108900. https://doi.org/10.1016/J.RESS.2022.108900

Yang, L., Feng, X., Liu, F., Liu, J., & Sun, X. (2019). Potential of soil moisture estimation using C-band polarimetric SAR data in arid regions. *International Journal of Remote Sensing*, *40*(5–6), 2138–2150. https://doi.org/10.1080/01431161.2018.1516320

Yang, X., Chen, D., Dong, Y., Xue, Y., & Qin, K. (2024). Identification of potential landslide in Jianzha county based on InSAR and deep learning. *Scientific Reports*, *14*(1), 21346. https://doi.org/10.1038/s41598-024-72391-2

Yin, L., Yan, S., Li, M., Liu, W., Zhang, S., Xie, X., Wang, X., Wang, W., Chang, S., & Hou, F. (2024). Enhancing soil moisture estimation in alfalfa root-zone using UAV-based multimodal remote sensing and deep learning. *European Journal of Agronomy*, *161*, 127366. https://doi.org/10.1016/J.EJA.2024.127366

Zareie, A., Amin, S. R., & Amador-Jiménez, L. E. (2016). Thornthwaite moisture index modeling to estimate the implication of climate change on pavement deterioration. *Journal of Transportation Engineering*, *142*(4). https://doi.org/10.1061/(ASCE)TE.1943-5436.0000840

Zhang, Y., Song, C., Zhang, K., Cheng, X., Band, L. E., & Zhang, Q. (2014). Effects of land use/land cover and climate changes on terrestrial net primary productivity in the Yangtze River Basin, China, from 2001 to 2010. *Journal of Geophysical Research: Biogeosciences*, *119*(6), 1092–1109. https://doi.org/10.1002/2014JG002616

Zhang, Z., Zimmermann, N. E., Stenke, A., Li, X., Hodson, E. L., Zhu, G., Huang, C., & Poulter, B. (2017). Emerging role of wetland methane emissions in driving 21st century climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9647–9652. https://doi.org/10.1073/pnas.1618765114

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350–361. https://doi.org/10.1016/J.NEUCOM.2017.01.026

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: The cosmic-ray soil moisture observing system. *Hydrology and Earth System Sciences*, *16*(11), 4079–4099. https://doi.org/10.5194/hess-16-4079-2012

Zribi, M., Muddu, S., Bousbih, S., Al Bitar, A., Tomer, S. K., Baghdadi, N., & Bandyopadhyay, S. (2019). Analysis of L-band SAR data for soil moisture estimations over agricultural areas in the tropics. *Remote Sensing*, *11*(9). https://doi.org/10.3390/rs11091122

Zribi, M., Taconet, O., Hégarat-Mascle, S. Le, Vidal-Madjar, D., Emblanch, C., Loumagne, C., & Normand, M. (1997). Backscattering behavior and simulation comparison over bare soils using SIR-C/X-SAR and ERASME 1994 data over Orgeval. *Remote Sensing of Environment*, *59*(2), 256–266. https://doi.org/10.1016/S0034-4257(96)00158-7