UNIVERSITY OF MISKOLC

Faculty of Earth Science and Engineering

Institute of Geography and GeoInformatics

Mikoviny Sámuel Doctoral School of Earth Sciences

Head of the Doctoral School: Dr. Péter Szűcs, professor

SOIL MOISTURE CHARACTERIZATION AND SOIL TEXTURE PREDICTION USING MACHINE LEARNING AND MULTI-SOURCE ENVIRONMENTAL DATA

PhD THESIS

By:

MOHAMED RAJHI

Scientific supervisor:

Prof. Endre Dobos

Miskolc, 2025

HUNGARY

ACKNOWLEDGEMENT

The successful completion of this Ph.D. journey would not have been possible without the unwavering support, invaluable guidance, and sincere encouragement I received from many individuals. I take this opportunity to extend my deepest gratitude to all who have supported me along the way.

First and foremost, I would like to express my profound appreciation to my supervisor, Prof. Endre Dobos, for his exceptional mentorship, continuous support, and insightful guidance throughout every stage of my doctoral studies at the University of Miskolc. His encouragement, critical feedback, and dedicated supervision were instrumental in shaping this work and bringing it to completion.

I am also thankful to all the esteemed professors at the University of Miskolc who provided me with a solid academic foundation during the initial years of the Ph.D. program. Their courses, advice, and academic excellence greatly enriched my knowledge and skills.

My sincere thanks go to the staff members of the Mikoviny Sámuel Doctoral School of Earth Sciences for their technical assistance, administrative support, and collaborative spirit throughout the duration of my research. Their professionalism and readiness to help create an ideal environment for learning and discovery.

I am deeply grateful to my fellow colleagues and collaborators for the countless hours of joint research, discussions, and shared experiences. Their knowledge, camaraderie, and teamwork made this journey both intellectually stimulating and personally fulfilling.

I also wish to acknowledge the Hungarian Government for awarding me a full scholarship, which enabled me to pursue this academic endeavor with financial stability and peace of mind.

To my close friends and colleagues, thank you for your constant encouragement, patience, and understanding through the highs and lows of this journey.

Lastly, I owe the deepest gratitude to my beloved family. Your unconditional love, endless sacrifices, and unwavering belief in me have been my greatest source of strength and motivation. This achievement is as much yours as it is mine.

Table of Contents

List of Figures

List of Tables

1. Introduction

The comprehension of soil moisture has evolved from historical agricultural knowledge to a pivotal focus in modern environmental and earth sciences. In the 3rd century BCE, agronomists like Mago of Carthage emphasized the importance of soil type, moisture retention, and irrigation scheduling in sustainable agriculture, principles subsequently reaffirmed by Roman scientists like Columella and Varro (Columella, 1941; Cato the Elder, 1934; Daniel J. Hillel., 1991; White, 1970). These early practices, grounded in empirical observation, demonstrate a fundamental understanding of the soil–water–plant relationship.

In the 19th century, scientists confirmed the significance of soil water in plant growth and yield, particularly via the research of German agronomist Ewald Wollny, who studied soil water retention and its impact on crops. In the early 20th century, advancements in physics and soil science led to the establishment of concepts like field capacity, permanent wilting point, and soil water potential, formalized by L.A. Richards in 1931, who introduced the tension table and defined the soil water retention curve (Richards, 1931).

The post-World War II era marked the development of soil physics as a quantitative science. The introduction of lysimeters, tensiometers, and neutron probes enabled more precise measurements of soil water content and dynamics (Daniel Hillel, 1980). During the period of 1970s and 1980s, soil moisture emerged as an essential component in agro-hydrological modeling and irrigation science. Nevertheless, field measurements continued to be challenging, costly, and significantly localized. This led in increase the interest in remote sensing methods that could provide a more comprehensive perspective on soil moisture monitoring.

The launch of microwave satellites, including NASA's Skylab in the 1970s and following by the passive sensors on SMMR (1978), AMSR-E (2002), SMOS (2009), and SMAP (2015), enabled the integration of soil moisture into the domain application of remote sensing (Njoku & Entekhabi, 1996). These missions made a significant advancement in the global monitoring of surface soil moisture and frequent revisits, however at low spatial resolutions and limited depth sensitivity (Kerr et al., 2010; Zribi et al., 2011). Simultaneously, active radar missions (ERS, Envisat, Sentinel-1) provided greater resolution datasets but are influenced by surface roughness, vegetation, and dielectric characteristics and require advanced algorithms to derive dependable moisture signals. Optical sensors (e.g., Sentinel-2) offer indirect vegetation-derived measures of moisture stress, such as NDVI and NDWI, which can enhance microwave observations but are constrained by cloud cover. Alongside advancements in satellite technology, the development of digital soil mapping and machine learning in the early 2000s created new opportunities for forecasting soil moisture and texture using multi-source environmental data (Minasny & McBratney, 2016; Grunwald, 2009).

Today, soil moisture is considered as a key climatic and hydrological parameter essential for comprehending land-atmosphere interactions, ecosystem functionality, and for water resource management (Seneviratne et al., 2010). It controls infiltration, runoff, and evapotranspiration, and functions as an essential component in land surface models, hydrological forecasts, and agricultural decision-making systems. Soil moisture acts as a memory variable in the climate system, influencing extreme weather events, including droughts, flood and heatwaves (Koster et al., 2004). Moreover, it is a fundamental variable recognized by the Global Climate Observing System (GCOS).

However, the spatial and temporal heterogeneity of soil moisture makes it difficult to observe comprehensively. Soil moisture is influenced by soil texture and structure, weather, vegetation cover, topography, irrigation, and land management practices. These vary not only in space but also across soil depth and time, making point measurements poorly representative of larger areas (Grayson & Western, 1998; Ochsner et al., 2013a). Although in-situ sensors like TDR, capacitance probes, and neutron moisture meters provide great precision, their application is limited to experimental sites and does not account for broad-scale variability.

Soil texture is a fundamental property of the soil, referring to the proportions of sand, silt, and clay. It controls key hydrological and agricultural processes, including water retention, infiltration rates, and the plant-available water capacity. Consequently, soil texture is crucial for calibrating soil moisture sensors in the field, as sensor performance is dependent upon the dielectric characteristics of the soil matrix (Topp et al., 1980; Robinson et al., 2008). Conventional methods to determine soil texture, utilizing laboratory analyses such hydrometer or laser diffraction techniques, are time-consuming, labor-intensive, and costly, particularly when applied over extensive areas (Minasny & McBratney, 2016). This constraint limits the spatial resolution especially in resource-constrained areas.

To address the limitations of traditional laboratory-based texture analysis, inverse modeling approaches have gained increasing attention. In this context, soil texture is inferred indirectly from the dynamics of soil moisture, rather than measured directly. Soil texture influences the rate and magnitude of soil moisture fluctuations following precipitation or irrigation, allowing for the prediction of textural classes or fractions through the analysis of the temporal dynamics of soil moisture profiles (Mohanty et al., 2017; Vereecken et al., 2016). Inversion-based techniques, whether integrated with sensor data or remote sensing soil moisture, present an attractive method for non-invasive, scalable texture mapping. This method is especially beneficial in agricultural monitoring systems where the expense or practicality of frequent ground sampling is constrained.

Moreover, soil moisture is affected not only by soil texture but also by various environmental variables, including vegetation cover, topography, land use, and climatic factors such as precipitation and evapotranspiration (Grunwald, 2009; Heung et al., 2014). These variables influence the surface energy balance and water flows, hence impacting soil moisture patterns even in homogeneous soils. Consequently, integrating multi-source environmental data into machine learning models may significantly improve the precision and resilience of soil texture forecasts (Wadoux et al., 2020). Integrating satellite data, ground-based sensors (e.g., Sentek EnviroSCAN), and environmental covariates enables the development of data-driven models that accurately map the texture at high spatial resolution, thereby enhancing water management and precision agriculture.

Research Rationale and the Relevance of the study

This research responds to the urgent need for integrated approaches that combine the strengths of remote sensing and ground-based measurements. It proposes a multi-source framework that merges Sentinel-1 SAR, Sentinel-2 optical indices, in-situ soil moisture from multi-depth Sentek EnviroSCAN sensors, and environmental variables (e.g., temperature, precipitation) using machine learning. This approach aims to enhance the accuracy, resolution, and scalability of soil moisture estimation in a cost-effective, transferable manner.

The rationale for focusing on two climatically distinct regions, Tunisia (semi-arid, water-stressed) and Hungary (temperate, data-rich), is to evaluate and characterize the soil moisture across heterogeneous landscapes. Tunisia’s Merguellil basin offers a relevant test site for assessing the potential of remote sensing in regions with limited monitoring infrastructure, whereas Hungary provides a reference environment with established datasets and field instrumentation. This dual-site strategy allows for rigorous model calibration, validation, and transferability testing under varying soil textures, land cover types, and climatic gradients.

The scientific contribution of this thesis lies in advancing digital soil mapping (DSM) techniques through a fusion of EO data, in-situ measurements, and machine learning. The relevance extends to precision agriculture, drought risk management, and climate adaptation. By providing spatially explicit soil moisture maps and texture predictions, the research supports data-driven decisions in water allocation, crop management, and land use planning. Furthermore, it aligns with global initiatives such as the FAO's digital agriculture agenda, SDG 2 (Zero Hunger) and SDG 13 (Climate Action).

In summary, this study addresses critical scientific and operational gaps in soil moisture estimation. It demonstrates that remote sensing and AI-based models, when calibrated with reliable in-situ data, can generate high-resolution outputs for sustainable land and water management, particularly valuable in regions where traditional monitoring is scarce or fragmented.

Aim and objectives

The primary aim of this doctoral research is to develop an integrated framework for accurate, non-invasive estimation of soil moisture using multi-source data, including Sentinel-1 SAR, Sentinel-2 optical imagery, in-situ sensor networks, and meteorological variables, combined through machine learning techniques. The study targets operational soil moisture monitoring at multiple spatial scales and in contrasting agroecological contexts, specifically the semi-arid Merguellil basin in Tunisia and temperate agricultural zones in Hungary.

To fulfil this aim, the research is structured around the following objectives:

1. To assess the role of soil physical properties, including texture, porosity, structure, and surface characteristics, in controlling soil moisture dynamics, using both laboratory-measured and remote-sensing-derived parameters.

2. To support the calibration and validation of the in-situ soil moisture measurements from Sentek EnviroSCAN capacitance probes across different depths (10 cm, 20 cm, 30 cm), ensuring reliable ground-truth data for model training and validation.
3. To extract and process Sentinel-1 backscatter coefficients (VV, VH) and Sentinel-2 vegetation indices (NDVI, NDWI) for monitoring vegetation–soil–water interactions at plot and watershed scales.
4. To construct predictive models using Random Forest regression algorithms for estimating soil moisture and soil texture classes (sand, silt, clay) from environmental and satellite-derived covariates.
5. To evaluate the spatio-temporal variability of soil moisture across the study area by integrating sensor data, topographic parameters, and climatic drivers (temperature, rainfall, evapotranspiration).
6. To contribute to the advancement of digital soil mapping (DSM) by demonstrating the effectiveness of combining remote sensing, in-situ data, and machine learning for soil moisture and soil texture estimation.

2. Literature review

2.1. Soil Moisture monitoring and measurements

Soil moisture refers to the amount of water held within the soil pores, primarily in the unsaturated zone between the surface and the groundwater table. It is typically expressed as volumetric water content or gravimetric water content, depending on the measurement context. In the context of hydrology, soil moisture governs key processes such as infiltration, evapotranspiration, and surface runoff. It affects the partitioning of rainfall into infiltration and excess water, shaping flood dynamics and groundwater recharge. From a climatological perspective, near-surface soil moisture modulates land–atmosphere interactions by controlling latent and sensible heat fluxes, thereby influencing boundary layer development and weather patterns (Seneviratne et al., 2010).

In agricultural systems, soil moisture determines water availability for plant uptake and is a major factor in crop yield variability. It controls seed germination, root development, nutrient transport, and susceptibility to drought stress. As such, accurate soil moisture information is essential for precision agriculture, irrigation scheduling, and drought early warning systems (Ochsner et al., 2013; Dorigo et al., 2017).

Despite its central role, continuous and large-scale monitoring of soil moisture remains a significant challenge. Traditional in-situ measurement methods, such as gravimetric sampling, time domain reflectometry (TDR), or capacitance sensors, provide accurate point-level observations but are limited in spatial coverage and scalability (Robock et al., 2000). In response, satellite-based remote sensing has emerged as a powerful tool to estimate surface soil moisture across broad geographic regions. Active microwave sensors, such as those on the Sentinel-1 platform, measure the backscatter coefficient which is sensitive to changes in surface roughness and dielectric properties related to soil moisture. Passive microwave missions like SMOS and SMAP offer global soil moisture products with coarse spatial but high temporal resolution (Entekhabi et al., 2010; Kerr et al., 2010).

In addition to radar systems, optical remote sensing provides indirect indicators of soil moisture via vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI). These indices capture vegetation health and canopy water content, both of which are influenced by underlying soil moisture conditions (Gao, 1996; Tucker, 1979).

To integrate these heterogeneous data sources and account for the nonlinear relationships between soil moisture and environmental variables, machine learning approaches, especially ensemble methods like Random Forest, have gained increasing attention. These models can combine satellite observations, topographic data, climatic inputs, and field measurements to generate high-resolution soil moisture estimates with improved accuracy and generalizability (Belgiu & Drăguţ, 2016)

Ultimately, soil moisture remains one of the most influential and yet challenging environmental variables to monitor and model. As demands for sustainable water management, food security, and climate resilience grow, the ability to accurately estimate soil moisture at scale has become a scientific and operational priority.

2.2. Soil Properties

Soil properties are the foundation of terrestrial ecosystem functioning and play a critical role in determining the capacity of soils to support vegetation, regulate water movement, store carbon, and influence energy exchanges between the land surface and the atmosphere. These properties are generally classified into physical, chemical, and biological categories. Among these, physical soil properties, notably texture, structure, and porosity, are central to understanding the distribution and dynamics of soil moisture, which is essential for hydrological modeling, irrigation planning, and climate-related studies (Daniel Hillel, 2004; Weil & Brady, 2017)

Soil physical properties influence key processes such as infiltration, percolation, evapotranspiration, root penetration, and aeration. They also govern the capacity of soil to store water and nutrients, buffer environmental changes, and support microbial communities. Understanding these parameters is therefore a prerequisite for accurately modeling soil water regimes, developing precision agriculture strategies, and designing nature-based solutions to address land degradation and climate vulnerability (Minasny & McBratney, 2016).

2.2.1. Soil Texture

Soil texture refers to the relative proportion of mineral particles in three size classes: sand (2.0–0.05 mm), silt (0.05–0.002 mm), and clay (<0.002 mm). These size fractions combine to determine a soil's textural class, as defined by the USDA soil texture triangle (Nrcs, 2017). Texture is a fundamental and relatively stable property that governs a wide range of hydraulic and ecological functions.

Textural composition directly influences soil moisture retention, infiltration rate, and drainage characteristics. Coarse-textured soils (e.g., sandy soils) have large pores that promote rapid drainage and high aeration but low water-holding capacity. In contrast, fine-textured soils (e.g., clay soils) have a predominance of micropores, which retain water effectively but limit air circulation and slow drainage. Loamy soils, often considered optimal for agriculture, strike a balance between these extremes by supporting both adequate water retention and good aeration (Saxton & Rawls, 2006).

Soil texture also affects the thermal and dielectric properties of the soil matrix, which in turn influences the retrieval of soil moisture from radar and passive microwave remote sensing (Topp et al., 1980). Texture has also been shown to modulate vegetation response to moisture, indirectly influencing vegetation indices such as NDVI, which are frequently used in remote sensing models for soil moisture prediction. Texture class is also important in soil moisture sensor calibration and transformation of raw data reading to soil moisture value. Different texture classes have often different calibration equations. Therefore, information on soil texture is a prerequisite of soil moisture monitoring and accurate estimation.

Traditional particle size analysis, such as sieve or hydrometer methods, though accurate, are labor-intensive, time consuming and not scalable. Consequently, recent advances in machine learning and proximal sensing technologies offer viable alternatives for large-scale texture prediction. Algorithms such as Random Forest have demonstrated high performance in predicting soil texture classes from multispectral satellite imagery, terrain attributes, and in-situ moisture readings (Heung et al., 2014; Wadoux et al., 2020).

2.2.2. Soil Structure

Soil structure describes the spatial arrangement of soil particles into aggregates or "peds" and the pore spaces between them. Unlike texture, which is determined by mineral composition, soil structure is a dynamic property influenced by organic matter content, microbial activity, wetting-drying cycles, and anthropogenic factors such as tillage (Bronick & Lal, 2005).

Different structural types, such as granular, blocky, platy, or prismatic, affect the geometry and continuity of pores in the soil. Well-structured soils have interconnected macropores that enhance infiltration and root penetration, while poorly structured or compacted soils exhibit low porosity and reduced water permeability. For instance, a platy structure may result in restricted vertical water movement, causing surface ponding and reduced gas exchange (Six et al., 2004).

The stability of soil aggregates also influences soil erodibility, bulk density, and hydraulic conductivity, all of which are crucial for sustainable land management. Aggregate stability is often used as an indicator of soil health, as it reflects the soil's ability to resist disintegration under rainfall and mechanical disturbance. Furthermore, structure significantly modulates the plant-available water capacity (PAWC) by influencing the balance between micropores (which hold water) and macropores (which facilitate drainage and aeration).

Because soil structure evolves over time and in response to management practices, it is a critical variable to monitor in long-term soil moisture studies and modeling. When integrated into hydrological or climate models, soil structure data can improve predictions of infiltration, runoff, and plant water availability (Reynolds et al., 2007).

2.2.3. Soil Porosity

Soil porosity refers to the proportion of the soil volume that consists of voids or pore spaces, and it is a critical property for determining the soil's capacity to store water and air. Total porosity is a function of particle size distribution, bulk density, and soil structure. It typically ranges from 30% to 60% in natural soils (Daniel Hillel, 2004a).

Pores are typically classified by their size:

- Macropores (>0.08 mm) enable rapid water movement and gas exchange.
- Mesopores (0.03–0.08 mm) retain plant-available water.

- Micropores (<0.03 mm) retain water tightly bound to particles and are less accessible to plants (Luxmoore, 1981).

High porosity enhances water infiltration, root aeration, and microbial activity, but excessive macroporosity may result in nutrient leaching and poor water retention, particularly in sandy soils. Conversely, while clay-rich soils exhibit high total porosity due to the abundance of micropores, much of this water is held at high matric tensions and is unavailable for plant uptake.

The spatial and vertical distribution of pores is also vital for understanding water movement through the soil profile. Capillary action in fine-textured soils supports water rise against gravity, whereas gravitational drainage dominates in coarse-textured soils. Porosity is often linked to hydraulic conductivity, which is a measure of the soil's ability to transmit water. Accurate estimation of porosity is essential in modeling soil water retention curves, unsaturated flow, and evapotranspiration dynamics.

In field conditions, soil porosity is typically inferred from the relationship between bulk density and particle density, or estimated through in-situ sensors that track soil moisture fluctuations over time. In recent years, geophysical and remote sensing techniques, such as ground-penetrating radar, electromagnetic induction, and satellite-based reflectance data, have been increasingly utilized to estimate porosity indirectly across large spatial scales, contributing to the advancement of digital soil mapping and hydrological modeling (Pradipta et al., 2022).

2.3. Soil Surface parameters

Soil surface parameters are critical physical descriptors that govern the interaction between the land surface and the atmosphere, influencing key processes such as evaporation, runoff, energy exchange, and soil moisture dynamics. These parameters, such as surface roughness, albedo, crust formation, temperature, and soil cover characteristics, are essential for accurately modeling surface hydrological responses and for interpreting remote sensing data in soil moisture estimation (Verhoef et al., 1997).

One of the most important surface parameters is surface roughness, which refers to the variability in micro-relief caused by tillage, erosion, aggregation, or vegetation residue. Surface roughness alters the aerodynamic properties of the soil-atmosphere interface and directly affects water infiltration, and the scattering of radar signals used in remote sensing. In radar remote sensing, particularly with synthetic aperture radar (SAR), surface roughness is a major control on backscatter intensity, and distinguishing its effects from those of soil moisture is a key challenge in soil moisture retrieval algorithms (Zribi & Dechambre, 2003; Baghdadi et al., 2008).

Albedo, the proportion of incident radiation reflected by the surface, is another vital surface parameter. Soils with lighter color, due to mineral composition or dryness, typically exhibit higher albedo, reducing net radiation absorption and influencing soil temperature. In contrast, darker, wetter soils absorb more radiation, increasing evaporation potential. Albedo also interacts with vegetation cover and surface crusting, altering the energy balance and soil thermal regime (Idso et al., 1975).

Crust formation at the soil surface, often resulting from the impact of raindrops or irrigation, can significantly reduce water infiltration, increase runoff, and inhibit seedling emergence. Crusted soils often exhibit higher reflectance and reduced surface roughness, which can be detected using high-resolution optical or radar sensors. Crusting is particularly prevalent in fine-textured soils under arid and semi-arid conditions and is influenced by land management and rainfall intensity (Valentin & Bresson, 1992).

Soil surface temperature, typically measured at depths of 2 to 5 cm or inferred from thermal remote sensing, is a critical parameter influencing evaporation and soil moisture distribution. It is dynamically affected by factors such as soil moisture content, solar radiation, and thermal conductivity. Wet soil generally exhibits lower surface temperatures due to evaporative cooling, whereas dry soils heat up more rapidly, especially under sparse vegetation cover. These temperature dynamics provide indirect indicators of surface soil moisture, making thermal infrared data valuable for soil moisture modeling. For instance, (D. Zhang et al., 2014) developed the Temperature Rising Rate Vegetation Dryness Index (TRRVDI), which utilizes the mid-morning land surface temperature rising rate and vegetation index to estimate regional soil water content, demonstrating the efficacy of thermal infrared remote sensing in soil moisture estimation.

Vegetation cover, whether as live biomass, crop residue, or mulching, modifies the surface energy balance, reduces direct solar radiation on the soil, and affects evaporation rates. It also mitigates wind and water erosion, enhances infiltration, and changes radar backscatter properties. Vegetation indices derived from remote sensing, such as the Normalized Difference Vegetation Index (NDVI), are commonly used as proxies for plant cover and can be linked to underlying soil moisture conditions (Carlson et al., 1995).

Together, soil surface parameters shape the interface between soil processes and environmental forcing, thereby determining the fate of precipitation, influencing plant-available water, and controlling erosion susceptibility. Their importance in soil moisture estimation models, especially those integrating radar and optical remote sensing, cannot be overstated. Improving the characterization of these surface parameters, whether through field measurement, modeling, or remote sensing, is therefore crucial for enhancing the accuracy of soil moisture predictions across scales.

2.4. Spatio-temporal variability of soil moisture

Soil moisture exhibits significant variability both in space and time due to a complex interplay of climatic, soil, topographic, land use, and vegetation factors. Understanding this spatio-temporal variability is fundamental to accurately modeling hydrological processes, optimizing irrigation, improving climate models, and interpreting remote sensing data. It also has practical implications for agricultural management, drought monitoring, and flood prediction (Western et al., 2002; Vereecken et al., 2008).

2.4.1. Spatial Variability

The spatial heterogeneity of soil moisture is driven primarily by intrinsic soil properties, such as texture, porosity, structure, and organic matter content, as well as topographic gradients, vegetation patterns, and microclimatic conditions. For example, coarse-textured soils drain faster and retain less water than fine-textured soils, leading to drier conditions under identical rainfall inputs (Brocca et al., 2009). Similarly, convex topographic features like ridges tend to be drier due to runoff and drainage, whereas concave depressions retain more moisture and often exhibit wetter conditions (Famiglietti & Wood, 1994).

Land use and land cover significantly influence spatial variability as well. Cultivated areas, forests, and bare soils show different infiltration and evaporation behaviors. Vegetation canopies intercept rainfall and modulate evaporation, while root systems affect water uptake and soil hydraulic properties, creating spatial patterns in moisture distribution at plot to watershed scales (Yang et al., 2012).

Moreover, spatial patterns may exhibit scale-dependent behavior. At fine scales (meters), variability is dominated by local soil characteristics and vegetation cover; at coarser scales (kilometers), climate gradients, soil types, and geomorphological units become the primary drivers (Grayson & Western, 1998). This scale dependency is critical in designing observation networks and modeling frameworks.

2.4.2. Temporal Variability

Temporally, soil moisture fluctuates in response to precipitation, evapotranspiration, irrigation, snowmelt, and drainage. These dynamics can occur at multiple time scales. from sub-daily diurnal cycles to seasonal and interannual trends (Entekhabi et al., 1996). Short-term changes are often driven by rainfall events and evapotranspiration cycles, while long-term variability is governed by climatic seasonality, land management, and changing vegetation cover.

Soil moisture memory, defined as the persistence of moisture anomalies over time, plays a key role in land–atmosphere interactions. For instance, wet soil conditions can sustain higher latent heat fluxes, influencing local atmospheric boundary layer development and precipitation feedback (Seneviratne et al., 2010). Conversely, dry soils can amplify drought severity by reducing evaporation and increasing sensible heat flux.

Soil depth also influences temporal variability. Surface layers (e.g., 0–10 cm) respond quickly to atmospheric conditions and rainfall, while deeper layers (e.g., >50 cm) change more slowly and reflect cumulative water balance over longer periods. This vertical gradient affects water availability to plants and the timing of recharge processes.

2.4.3. Implications for Remote Sensing and Modeling

Capturing the spatio-temporal variability of soil moisture is challenging, especially over large areas or in data-scarce regions. In-situ measurements offer high temporal resolution but are limited in spatial extent. In contrast, remote sensing methods, such as radar (e.g., Sentinel-1) and passive microwave sensors (e.g., SMAP, AMSR2), provide broader coverage but are sensitive to surface conditions and often limited to shallow depths (Mohanty et al., 2017).

To improve accuracy, data fusion approaches that integrate in-situ measurements, remote sensing, and model simulations (e.g., land surface models or machine learning models) are increasingly used to produce high-resolution soil moisture estimates with temporal continuity (Peng et al., 2017).

The study of spatio-temporal variability is not only theoretical but highly applied, supporting early warning systems, precision irrigation, land degradation monitoring, and climate adaptation strategies. Understanding and predicting soil moisture dynamics thus remains a central focus of earth system science and environmental engineering.

2.5. In-situ Soil Moisture Measurement Methods

In-situ soil moisture methods provide direct, point-based observations of volumetric or gravimetric water content within the soil profile. These methods can be broadly categorized into destructive (gravimetric) and non-destructive (sensor-based) techniques (Robinson et al., 2008).

2.5.1. Gravimetric Method

The gravimetric method is the most fundamental and direct approach for determining soil moisture content. It involves collecting undisturbed soil samples, weighing them before and after oven drying at 105°C for 24 hours, and calculating moisture content as the mass of lost water relative to the dry soil mass. Although accurate and reliable, this method is time-consuming, labor-intensive, and not suitable for real-time or repeated measurements (Gardner, 1986). It is commonly used as a reference for calibrating other methods.

2.5.2. Time Domain Reflectometry (TDR)

TDR is a widely used non-destructive technique that measures soil moisture by analyzing the travel time of an electromagnetic pulse along metal probes inserted into the soil. The propagation velocity of the pulse is related to the dielectric constant of the soil, which is strongly correlated with volumetric water content. TDR sensors are valued for their accuracy, repeatability, and minimal salinity sensitivity, although they can be expensive and require calibration for different soil types (Topp et al., 1980; Evett, 2003).

2.5.3. Capacitance and Frequency Domain Reflectometry (FDR)

Capacitance-based and FDR sensors measure the dielectric permittivity of soil using oscillating electrical fields. Unlike TDR, FDR uses a range of frequencies, and the measured signal is converted to soil moisture via calibration equations. These sensors are lower in cost, easier to install, and allow multi-depth profiling, but they are more sensitive to soil texture, temperature, and salinity (Dean et al., 1987;Bogena et al., 2007).

One example of such a system is the Sentek EnviroSCAN probe, which includes multiple capacitance sensors along a vertical access tube. This system enables real-time monitoring of vertical soil moisture profiles, making it particularly useful for precision agriculture and irrigation management.

2.5.4. Neutron Scattering

The neutron probe is a nuclear-based method that measures hydrogen atoms in the soil, which are primarily associated with water. It involves placing a radioactive source into the soil and measuring neutron moderation. Neutron probes offer deep profile measurements and good accuracy but require strict regulatory control, radiation safety procedures, and periodic calibration (Greacen, 1981; S. R. Evett & Steiner, 1995) .

2.5.5. Tensiometers and Gypsum Blocks

Tensiometers measure soil water potential by equilibrating the tension of water in a porous ceramic tip with the surrounding soil matrix. They are accurate in moist conditions (typically 0 to -80 kPa) but become ineffective in drier soil due to cavitation. Gypsum blocks, in contrast, estimate soil moisture tension by measuring the electrical resistance of the porous block as it equilibrates with soil water. While they operate over a wider moisture range, their accuracy is limited compared to tensiometers and they are prone to degradation over time (Cassel & Nielsen, 1986).

2.6. Remote sensing soil moisture estimation

Remote sensing has emerged as a pivotal tool for estimating soil moisture across large spatial extents and varying temporal resolutions. Unlike in-situ measurements, which provide accurate but point-based data, remote sensing techniques offer spatial continuity, enabling regular monitoring of soil moisture at regional to global scales. This has important implications for agricultural water management, hydrological forecasting, and climate modeling (Entekhabi et al., 2010; Kerr et al., 2010).

Remote sensing techniques for soil moisture estimation are generally categorized into active and passive microwave methods. Passive microwave sensors measure naturally emitted microwave radiation from the Earth's surface, with longer wavelengths (e.g., L-band at 1.4 GHz) being especially sensitive to surface soil moisture within the top 0–5 cm of the soil profile. Missions such as the Soil Moisture Active Passive (SMAP) and the Soil Moisture and Ocean Salinity (SMOS) satellites have provided valuable global soil moisture data products at moderate temporal resolution (2–3 days) and coarse spatial resolution (~40 km) (Kerr et al., 2010;Entekhabi et al., 2010). These passive systems are advantageous due to their sensitivity to moisture variations, but their utility is limited in regions with dense vegetation or heterogeneous land cover due to signal attenuation.

Active microwave remote sensing, particularly Synthetic Aperture Radar (SAR), is a valuable technique for soil moisture estimation due to its ability to operate in all weather conditions and during both day and night. SAR sensors, such as Sentinel-1 (C-band) and ALOS-2 PALSAR (L-band), transmit microwave pulses and record the backscattered signal, which is influenced by surface moisture, roughness, and vegetation cover. While SAR offers valuable spatial resolution (10–100 m), retrieval accuracy is often limited by the complex interactions among these surface factors. To overcome this, machine learning approaches like artificial neural networks have been successfully applied to enhance soil moisture prediction from SAR data, as demonstrated by Chung et al., (2022) and supported by broader reviews such as Kornelsen & Coulibaly (2013).

Beyond microwave sensing, optical and thermal remote sensing systems also contribute indirectly to soil moisture estimation. Indices such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI), derived from sensors like Sentinel-2 or Landsat, provide vegetation condition proxies that correlate with moisture availability. For instance, healthy and dense vegetation, as indicated by high NDVI values, typically reflects sufficient soil moisture, while reduced NDVI values and signs of vegetative stress may serve as proxies for underlying water deficits. (Jiang et al., 2006; Gao, 1996). Similarly, thermal infrared data can provide insights into surface temperature dynamics and evapotranspiration, which are influenced by soil moisture levels (Anderson et al., 2007).

Nevertheless, several limitations affect the accuracy of remotely sensed soil moisture data. The primary constraint is the depth sensitivity of microwave sensors, which only capture surface soil moisture and may not reflect the conditions of deeper root zones. Additionally, the presence of dense vegetation and rough terrain can obscure or distort microwave signals, leading to uncertainties in retrievals. Furthermore, passive sensors, while reliable, lack spatial detail, whereas active sensors demand complex processing algorithms and ancillary data for accurate estimation (Njoku & Entekhabi, 1996).

To overcome these limitations, current research trends emphasize the integration of remote sensing data with ancillary information such as in-situ sensor networks, digital elevation models (DEMs), land cover maps, and climate reanalysis data. Advanced modeling approaches, including machine learning algorithms have shown promise in improving soil moisture retrieval by fusing multi-source environmental datasets (Peng et al., 2017; Wang & Gao, 2023). These hybrid approaches not only enhance spatial resolution and temporal consistency but also provide better generalization across diverse landscapes.

In conclusion, remote sensing represents a non-invasive and scalable approach for soil moisture estimation, with significant advantages in terms of coverage and frequency. While no single sensor or technique can provide complete accuracy under all conditions, the combination of active and passive microwave data, optical indices, in-situ validation, and data-driven models offers a robust framework for soil moisture monitoring. These advancements are crucial for supporting sustainable land and water management, early warning systems, and climate adaptation strategies.

2.7. Soil Moisture Mapping

Soil moisture mapping refers to the spatial representation of water content within the soil, providing crucial information for applications in agriculture, hydrology, meteorology, and land management. The ability to generate accurate and timely maps of soil moisture is essential for monitoring drought, optimizing irrigation, assessing flood risks, and supporting climate models. It also plays a key role in understanding land-atmosphere interactions and energy fluxes at different spatial and temporal scales (Brocca et al., 2017; Vereecken et al., 2008).

Traditionally, soil moisture maps were generated using ground-based measurements from sparse in-situ sensor networks and gravimetric sampling. While these methods offer high accuracy, they are labor-intensive and insufficient for covering large areas. The spatial resolution is highly dependent on the number and distribution of monitoring stations, making it difficult to extrapolate measurements across heterogeneous landscapes (W. A. Dorigo et al., 2011).

The emergence of remote sensing technologies has significantly enhanced soil moisture mapping capabilities. Satellite-based sensors such as SMAP, SMOS, Sentinel-1, and MODIS provide repetitive, spatially continuous observations at regional to global scales. Passive microwave sensors (e.g., SMAP, AMSR2) offer near-daily global coverage but at coarse spatial resolutions (~36–50 km), while active radar systems (e.g., Sentinel-1 SAR) deliver finer spatial resolution (~10–100 m) with sensitivity to surface roughness and vegetation (Entekhabi et al., 2010; Kerr et al., 2010). These datasets can be fused using data assimilation and downscaling techniques to improve spatial detail while maintaining temporal frequency (Peng et al., 2017).

In recent years, the application of machine learning (ML) within digital soil mapping (DSM) frameworks has become increasingly important for soil moisture mapping. ML models such as Random Forest, Support Vector Machines, and Deep Neural Networks can learn complex, nonlinear relationships between soil moisture and environmental covariates including land cover, topography, climate variables, and vegetation indices (Gruber et al., 2019; Wadoux et al., 2020). These models, when trained on both remote sensing data and in-situ measurements, offer robust predictive maps of soil moisture at varying depths and spatial resolutions.

An effective soil moisture mapping framework typically integrates multiple data sources: (1) remote sensing imagery, (2) point-based sensor data, (3) climatic variables (e.g., precipitation, temperature), (4) terrain attributes (e.g., slope, aspect), and (5) soil properties (e.g., texture, porosity). The fusion of these datasets enables the creation of high-resolution maps that capture spatial heterogeneity and temporal dynamics in soil moisture. These maps are particularly valuable in semi-arid and data-scarce regions where decision-makers rely on spatial diagnostics for resource management and early warning systems (Kibaroglu, 2016).

The spatial extension of soil water regime parameters can be effectively achieved by integrating geomorphological variables with point-based soil moisture observations. Terrain attributes such as slope, elevation, and curvature play a significant role in controlling soil moisture distribution, and their incorporation into geostatistical models enhances the capacity to upscale local measurements across heterogeneous landscapes. This method addresses the challenges associated with sparse sensor networks in operational monitoring systems (Deák et al., 2024). In addition, sensor calibration remains a critical step for ensuring the accuracy of soil moisture measurements. Off-site calibration approaches, which incorporate site-independent parameters such as soil texture and bulk density, have been shown to improve the reliability of capacitance-based sensors like the Sentek EnviroScan probe, particularly in diverse field conditions where in-situ calibration is not feasible (Kibirige & Dobos, 2021).

In summary, soil moisture mapping has evolved from laborious field-based assessments to sophisticated remote sensing and modeling approaches. Advances in geospatial technologies, sensor networks, and artificial intelligence have enabled researchers to generate accurate, timely, and high-resolution maps of soil moisture. These developments contribute significantly to sustainable land use planning, agricultural productivity, climate resilience, and environmental monitoring on global and local scales.

3. Soil moisture characterization using optical and radar remote sensing data

3.1. Study area

The governorate of Kairouan, located in the central region of the country (Figure 1), occupies a strategic position at the regional and the national level. It extends over 658,000 ha and is in the form of a wide corridor of plains, which are limited to the west by mountainous areas and to the east by depressions made up of sebkhas. This natural environment is a made up of quite contrasting physical units (plains, hills, and mountains) offering climatic nuances and different resources, which necessarily generate specific uses and modes of occupation (Sarra Bel Haj Salem, 2013). The study area, corresponding to the Merguellil basin, is included in the Kairouan plain. It is one of the three largest river basins on the southern flank of the Tunisian ridge, flowing into the Kairouan plain (Zribi et al., 2020). It is a relatively large homogeneous valley, but is also very sensitive to erosion (Le Goulven et al., 2009). The basin consists of two geographically dissimilar parts. The upstream section, which corresponds to the El Haouareb Dams watershed, is mountainous. The Kairouan alluvial plain encompasses the downstream part (Hermassi et al., 2014).

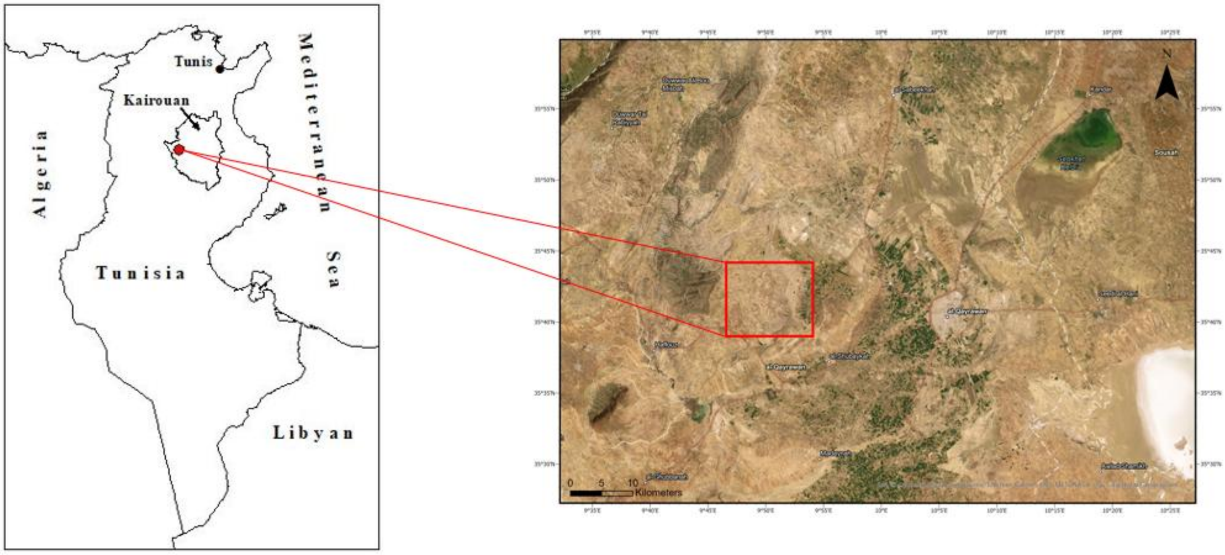


Figure 1. Study area Merguellil downstream basin of the Kairouan plain

The Merguellil upstream catchment (1200 km²) has a hilly topography (altitudes ranging from 200 to 1200 m, with a median elevation of 500 m) and diverse geology, morphology, vegetation, and land-use conditions. The Merguellil downstream watershed is part of the very large and flat Kairouan alluvial plain, which extends over about 3000 km² and with the altitude decreasing very gradually from 200 m to 80 m (Chulli, 2011). Our study area, the downstream part of the watershed, covered an area of 87 km² close to the dam, west of Kairouan city.

3.1.1. Geology

The upstream basin of Merguellil is composed of limestone, calcareous marl and marly sedimentary formations of the Cretaceous and Eocene (Figure 2). The oldest formation dates from the Triassic. Our study area, the Merguellil downstream basin of the Kairouan plain, corresponds to a collapsed basin, where the Plio-Quaternary continental detrital filling can exceed 700 m in thickness. A small Cretaceous limestone structure outcrops at the western part of the basin located in the El Haouareb mountain (Ammar et al., 2006).

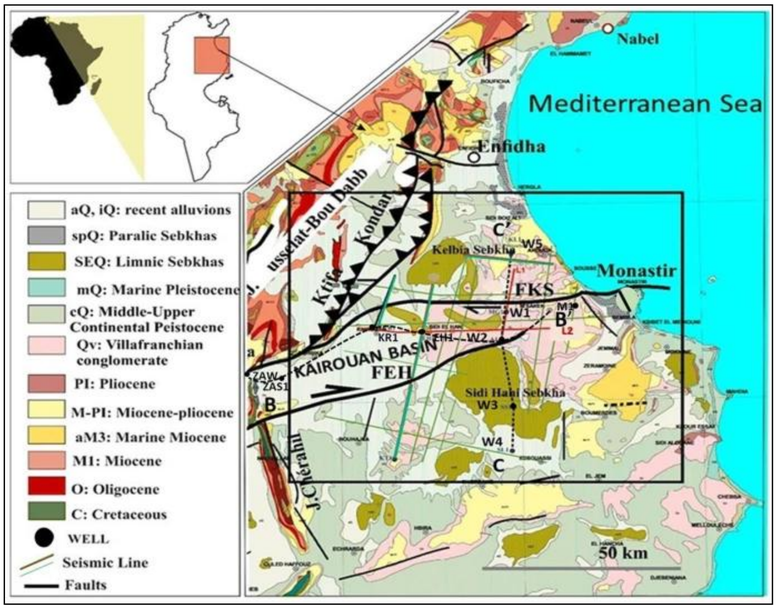


Figure 2. The geological map of the Kairouan Basin based on the 1/50,000 (Bédir et al., 2020)

3.1.2. Climate of the Kairouan Region

The climate of central Tunisia, particularly in the Kairouan region, is distinguished by significant year-to-year and intra-year variations in rainfall. This fluctuation, along with water scarcity and drought periods, has a negative impact on rain-fed agricultural productivity, resulting in unpredictable yields (Mougou et al., 2011).

The study of several parameters, namely precipitation and temperature, served to characterize the climate of the study area and to assess its degree of aridity. Indeed, these parameters have a significant impact on water availability (Sarra Bel Haj Salem, 2013).

The weather data were obtained for this study from the World Meteorological Organization (WMO). The meteorological station is located in Kairouan, Tunisia, at an elevation of 65 meters above sea level. The station operates in the Africa/Tunis time zone and is geographically positioned at 35.6781° N latitude and 10.0963° E longitude. This location provides valuable climatic data representative of central Tunisia's inland semi-arid conditions.

The annual precipitation in Kairouan shows considerable interannual variability, with total rainfall in 2023 remaining below 402 mm, characterized by irregular and concentrated events, particularly in May and November. Daily mean temperatures ranged from approximately 9 °C during the winter months to over 34 °C in the peak of summer, with July and August being the hottest periods, as reflected in the temperature trends. This thermal regime, combined with scarce and erratic rainfall, contributes to high atmospheric water demand. The annual potential evapotranspiration in the region is estimated at around 1460 mm (Hamdi et al., 2017), highlighting the strong water deficit typical of the semi-arid Mediterranean climate.

Building on this climatic context, Figure 3 illustrated the daily variation in maximum, average, and minimum air temperatures throughout 2023. Maximum daily temperatures peaked at around 47 °C in mid-summer, while winter minimum dropped to nearly 2 °C in January. Average daily temperatures fluctuated between approximately 9 °C in the coldest months and 34 °C during the hottest period. Clear seasonal patterns are evident, with pronounced peaks during the summer months (June to August) and lower values in the winter (December to February). A color-coded bar at the bottom enhances visualization of the temporal temperature gradient, offering insight into seasonal transitions.

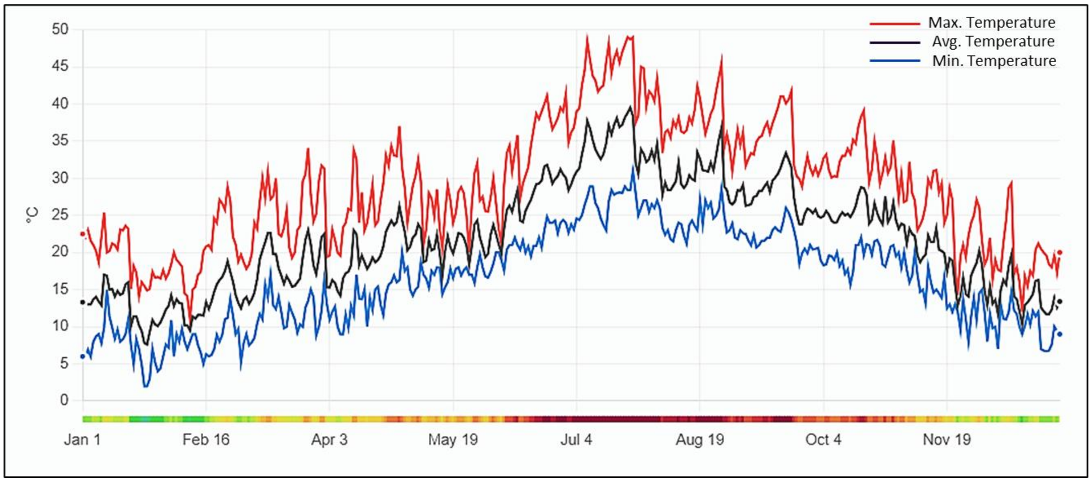


Figure 3. Daily maximum, average, and minimum air temperature trends for the year 2023. Data source: World Meteorological Organization (WMO).

Figure 4 presents daily precipitation patterns over the same period. Rainfall events were highly irregular, with intense peaks exceeding 65 mm/day in December and 30 mm/day in May. Most days recorded zero precipitation, especially between June and September, underscoring the prolonged dry season. These infrequent yet sometimes extreme rainfall events contribute to the uneven annual distribution of water resources in Kairouan.

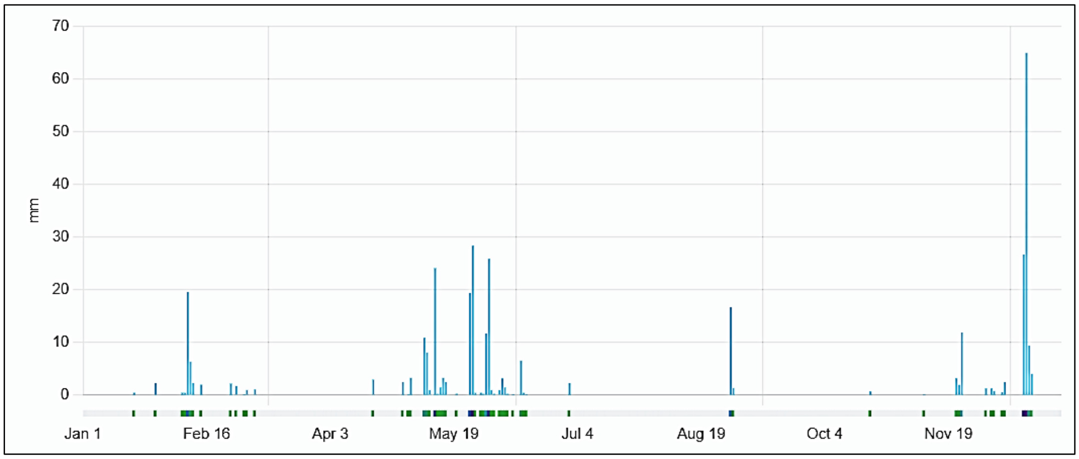


Figure 4. Daily precipitation recorded in 2023. The vertical bars represent rainfall amounts in millimeters. Data source: World Meteorological Organization (WMO).

Daily wind speed measurements in Kairouan during 2023 (Figure 5) show notable variability across the year, with peaks exceeding 16 km/h particularly evident in the winter and spring months. Wind speeds tend to be more moderate and stable during the summer season, although fluctuations persist. These patterns are influenced by seasonal atmospheric dynamics and the local topography, which modulate wind flow across the region. Monitoring wind speed is essential for understanding its impact on evapotranspiration, dust dispersion, and agricultural planning in semi-arid Mediterranean environments such as Kairouan.

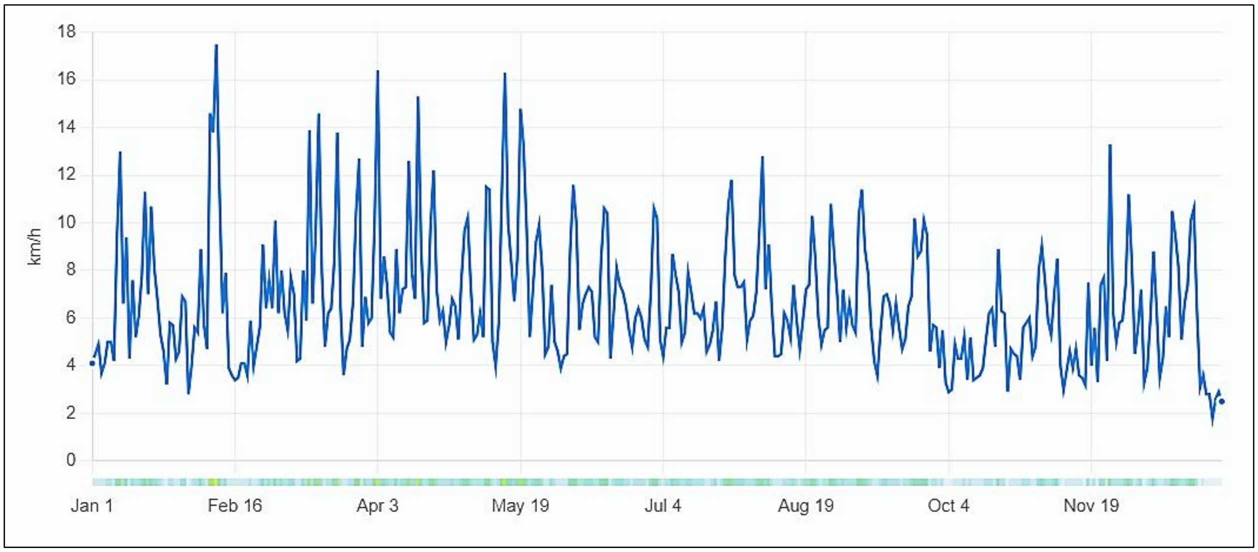


Figure 5. Daily Wind Speed in Kairouan (2023). Data source: World Meteorological Organization (WMO).

Atmospheric pressure in Kairouan throughout 2023 (Figure 6) fluctuated between approximately 1000 and 1035 hPa, with higher values typically observed during the cooler months (January–March and November–December). Sudden drops and spikes indicate the passage of weather systems such as high- and low-pressure fronts. The relatively stable pressure patterns observed during summer suggest calm, settled weather, consistent with the Mediterranean climate of the region. These pressure trends are critical for understanding weather dynamics and forecasting in semi-arid environments.

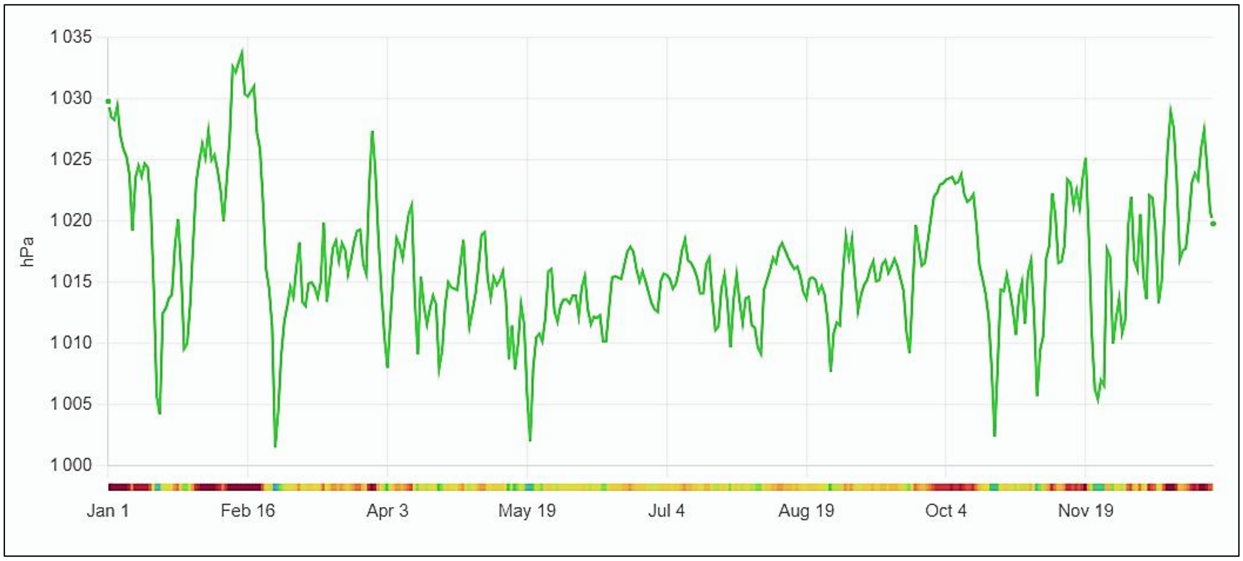


Figure 6. Daily Air Pressure Variations in Kairouan (2023). (Source: World Meteorological Organization (WMO))

3.2. Materials and methods

3.2.1. Materials

3.2.1.1. Sentinel 1 Imagery

SENTINEL-1 is an imaging radar mission providing continuous all weather, day-and night imagery at C-band (Table 1). The Sentinel-1 mission offers a range of distinct imaging modes, each with its own resolution, ranging from as low as 5 meters, and coverage extending up to 400 kilometers. It encompasses double polarization capability, rapid revisit times, and expedited product delivery. Additionally, precise spacecraft position and attitude measurements accompany every observation (Fletcher & Karen, 2012). Sentinel-1B satellite imagery was used to characterize soil moisture in this study. The Sentinel-1B data used in the study has the following specifications:

Table 1. Specifications of the Sentinel-1B data used in the study (Roland Meynart et al., 2013)

Specifications	Sentinel-1B
Polarization	VV-VH
Imaging frequency	C-Band (5.4 GHz)
Resolution mode	5 meters
Acquisition times	January 2018–June 2020

3.2.1.2. Sentinel 2 Imagery

SENTINEL-2 is a European multi-spectral imaging mission with a wide and high-resolution swath. The twin satellites, which are in the same orbit but phased at 180°, are designed to have a high revisit frequency of 5 days at the Equator. SENTINEL 2 is equipped with an optical payload that samples 13 spectral bands: four at 10 m, six at 20 m, and three at 60 m spatial resolution. The orbital swath is 290 kilometers wide (Franks & Rengarajan, 2023). During this research, the Sentinel 2 image will be used to calculate the values of the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI).

3.2.1.3. Digital elevation data

Digital Elevation Models (DEMs) are fundamental datasets in geographic information systems (GIS) used to represent terrain elevation and generate relief maps. Since 2019, the Copernicus Programme has provided a global, high-resolution, and consistent DEM freely available to the scientific community for diverse geospatial applications. DEMs are generally classified into two main categories: Digital Surface Models (DSM) and Digital Terrain Models (DTM). The DSM captures the elevation of the Earth's surface, encompassing both natural elements (e.g., vegetation, landforms) and anthropogenic structures (e.g., buildings, infrastructure), while the DTM represents the bare ground elevation without above-ground objects (F. Chen et al., 2021).

The Copernicus Digital Elevation Model (DEM) is delivered as a high-resolution DSM derived primarily from TanDEM-X mission data, offering global coverage with exceptional vertical consistency. The GLO-30 product, distributed by the European Space Agency (ESA), provides a spatial resolution of 30 m, making it particularly valuable for a wide range of applications such as hydrological and geomorphological modeling, land cover mapping, flood risk assessment, and terrain characterization (Figure 7).

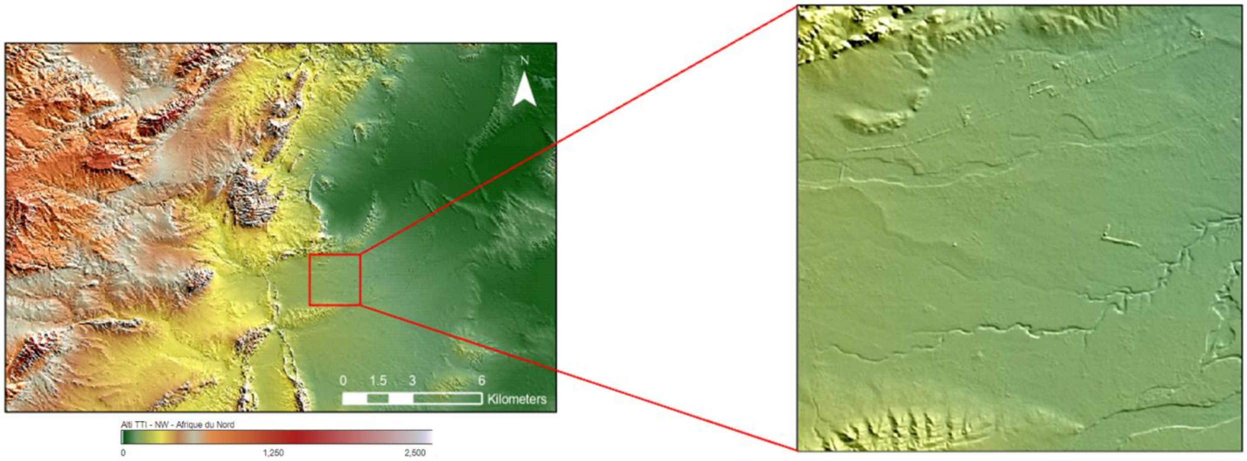


Figure 7. The DEM of the study area using Copernicus DEM GLO-30 m

The elevation showed that the study area is at the limit of the mountain range, it is mainly consolidated materials, colluvial and alluvial materials coming from the mountain. This is what fills up our study area.

The geography of this study site is primarily defined by geology, with the plain dominating and the presence of mountains such as Jebel Trozza from the north-south direction. Clayey and clayey-sandy textures may be observed in the agricultural areas. As a result, topography is the main factor affecting soil formation in the study area.

3.2.2. Methods

The diagram (Figure 8) illustrates the used workflow for characterizing soil moisture through the integration of multi-source remote sensing and environmental data. It begins with a preliminary area analysis, followed by the collection of three main data types: Sentinel-1 radar imagery, Sentinel-2 optical imagery, and climate data (e.g., ERA5, GPM). Following, covariates were extracted from these sources, the backscatter coefficients from Sentinel-1, the vegetation and water indices (NDVI, NDWI) from Sentinel-2, and environmental parameters from climatic datasets.

These covariates are then integrated and subjected to data analysis, including statistical methods such as linear correlation, to establish relationships between the derived indices and observed soil moisture. The final output of the research is the spatial and temporal characterization of soil moisture conditions across the study area. This integrated approach leverages the complementary strengths of radar, optical, and climate data for a comprehensive understanding of soil hydrological dynamics.

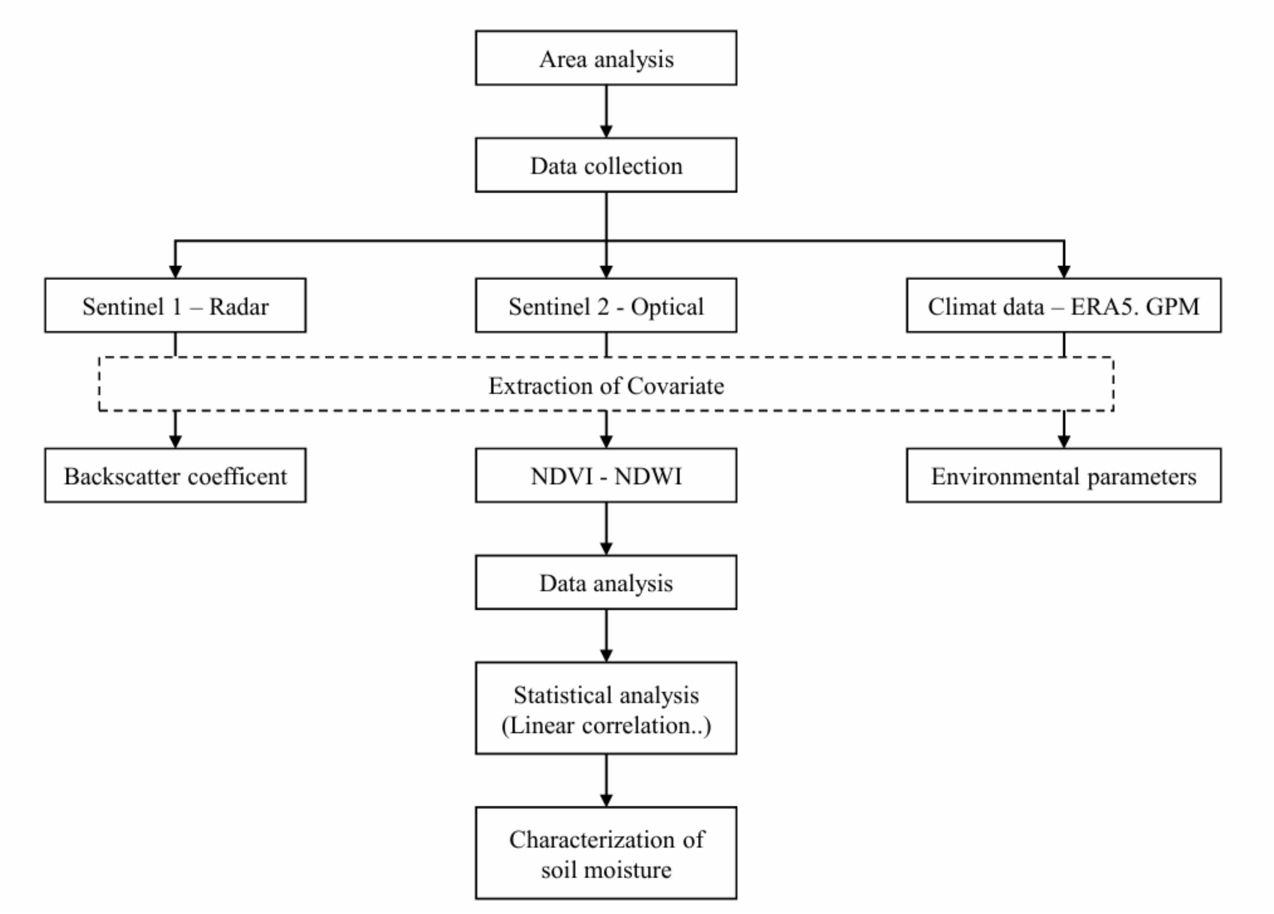


Figure 8. Methodological diagram of the soil moisture characterization

3.2.2.1. Data collection

The environmental covariates, Sentinel-1 (Backscatter coefficient), Sentinel-2 (NDVI and NDWI) and Digital terrain data were chosen as target explanatory variables to integrate. NDI (Normalized Difference Index) is calculated as the difference between reflectance values in two bands divided by the sum of those values (Dobos & Luca Montanarella, 2010)

3.2.2.2. Extraction of Covariates

A combination of literature and statistical processing was used to select certain factors. According to the literature, covariates had to follow three criteria: first, they had to represent soil-forming elements; second, they had to have a direct link with SM; and third, they had to be readily available (Dobos et al., 2000). On the basis of these specifications, three kinds of data were chosen as environmental covariates: Sentinel-1 C-band, terrain data derived from DEM, and Sentinel-2 data to explain the biomass/vegetation influence on the SM (Kibirige & Dobos, 2020).

3.4.3. Statistical Methods

In this chapter, the statistical method used to describe and analyze soil moisture is the linear correlation. In a data-poor environment, linear regression produces more realistic spatial patterns over the landscape (Q. Zhang et al., 2021).

3.2.2.4. Characterization of Soil Moisture

Many researchers have developed models to characterize soil moisture, and these models showed that soil moisture and backscattering coefficient have a fundamental relationship, e.g. (Q. Zhang et al., 2021). The backscatter coefficient and Sentinel-2 data are considered main parameters that can help to characterize the soil moisture.

3.3. Results and Discussion

3.3.1. Radar remote sensing

An optical image and a Sentinel-1 radar image shown in Figure 9 to visually analyze the area. The radar image (polarizations) is colored with an RGB composite: red = VH, green = VV, and blue = NDI, where NDI refers to the Normalized Difference Index of the VH and VV polarizations. The dominance of bluish tones, greenish tones, and the presence of reddish tones in the southern part of the study area can be immediately distinguished.

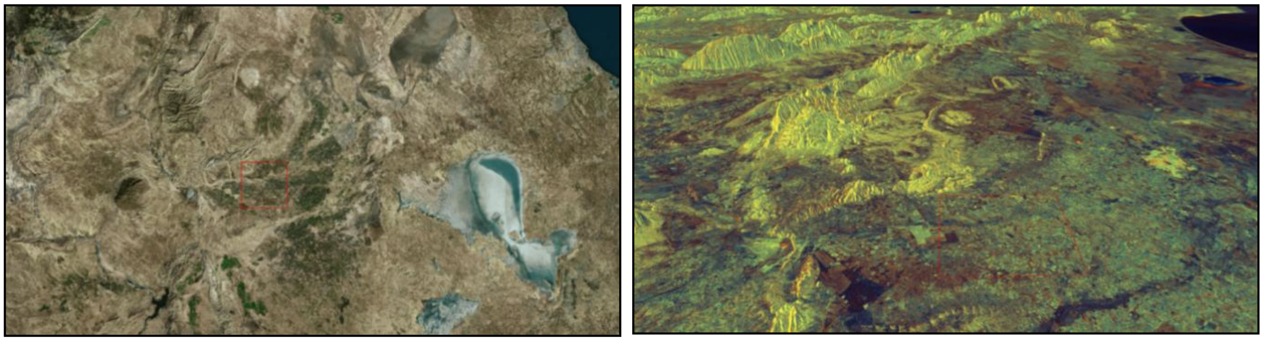
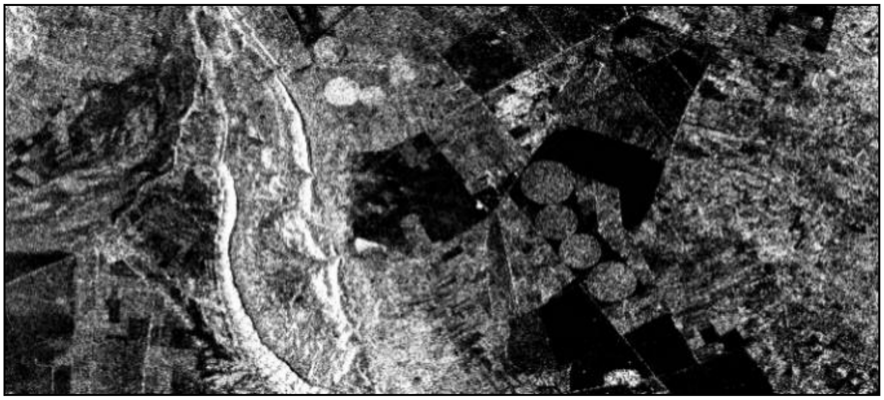


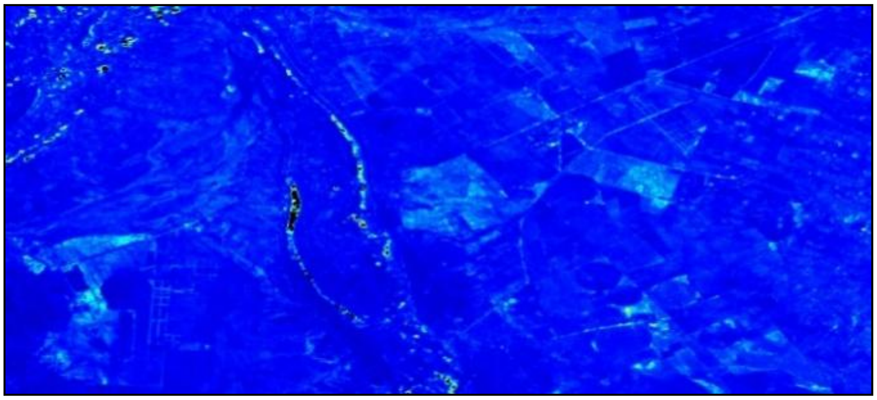
Figure 9. Optical image from the Bing Map (Left) and Sentinel-1B imagery (Right) of the study area

First, a visualization exercise was performed using the RGB composite. A more detailed analysis of the different polarizations (Figure 10) was then carried out as follows:



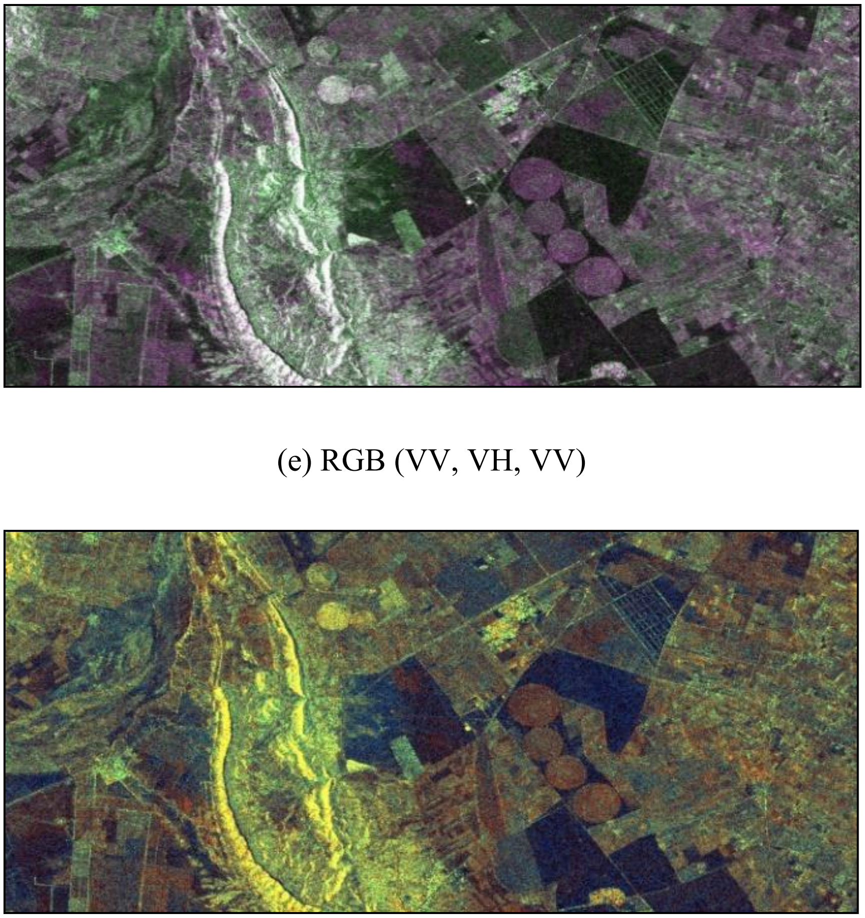
(a) Polarization VV

(b) Polarization VH



(c) Elevation

(d) Optical Image



(e) RGB (VV, VH, VV)

(f) RGB (VH, VV, NDI)

Figure 10. (a) Polarization VV; (b) Polarization VH; (c) Elevation; (d) Optical Image; (e) RGB (VV, VH, VV); (f) RGB (VH, VV, NDI)

Because SAR signals are prone to multiple scattering at various layers of vegetation, as well as at the subsurface or terrain level, interpreting vegetated parts of a SAR image demands a high level of understanding. The intensity of such scattering is heterogeneous in nature, varying from one SAR image to another, due to the changing nature of plant structures and closures. The VH polarization produces comparable results to VV polarization, with the exception of greater separation of water masses and more precise information about the water region in the generated image. The VV polarization is more informative for land use and for vegetal cover analysis which is related to the soil moisture. The wide histogram of the VV and the VH polarization image helps for a better discrimination of thematic classes including possible differentiation of identifying the open water, Irrigated area, soil roughness and the urban area. Although the RGB exercise of like-polarized and cross-polarized data can be considerably more informative. Figure 10e and 10f present a false-color composite of the Merguellil watersheds downstream. These images were acquired by the Sentinel-1 radar satellite. To aid visual interpretation, the multiple channels of polarimetric data can be used to present the data in a colored image, in which certain image features are recognizable. As a simple example, a color image can be made using a VV = Red, VH = Green and VV = Blue channel assignment (Figure 10e). This tends to look realistic, as soil water content reflections have a higher VH component than VV, and vegetation has a higher average than VH backscatter. The Sentinel-1 SAR image in VV and HV polarization modes (Figure 10f) was used, with the VH, VV, and NDI layers presented in RGB mode. Changes in the intensity of each color are related to surface conditions and covers; such as variations in surface roughness, biomasses, plant density, soil wetness, and soil types. Higher radar reflectance represented in VH/VV (yellow) is found in semi-consolidated and consolidated rocky material; lower backscatter in NDI polarization (dark blue) and VV/NDI (reddish) appears where erosional deposits and material related to fluvial erosion are abundant. Brighter colors are dominant where distinct bedrock layering can be observed. The units in bright yellow are typically caused by strong backscatter and VH polarization modes with VV represented as blue approaches zero. The reddish areas in the image are the ploughed terrain. The yellow-greenish areas are cultivated areas with high soil moisture. The green areas are due to the relatively high intensity of the VV channel, which is strongly correlated with the amount of biomass. The C-band VV and VH channels show the biomass variations over the entire region.

Three major units, dark violet, bright violet, and green, were distinguished in the Sentinel-1 SAR image in VV and HV polarization modes (Figure 10e), represented as VH/VV/VV in RGB. Higher VV/VV (violet) radar backscatter was observed over plowed terrain with high surface roughness, whereas lower backscatter in VV polarization (dark violet) was associated with areas dominated by uncultivated soils. Bedrock layering in the mountainous areas was identifiable through the predominance of brighter tones. Units in bright yellow were generally linked to high backscatter in VH polarization relative to VV, while the blue channel approached zero. Such patterns may also indicate high soil moisture conditions and the presence of cultivated areas.

Nevertheless, the interpretation of SAR signals remains challenging due to their sensitivity to multiple scattering processes in vegetation, subsurface layers, and terrain features. Therefore, further acquisition of ground-truth datasets and more comprehensive analysis are required in future research to validate these findings and strengthen the reliability of SAR-based soil moisture and land cover assessments.

3.3.2. Optical Remote Sensing

3.3.2.1. Characterization of the vegetation cover using the vegetation index (NDVI)

Figure 11 presents two normalized difference vegetation index (NDVI) images derived for the eastern part of the study area.

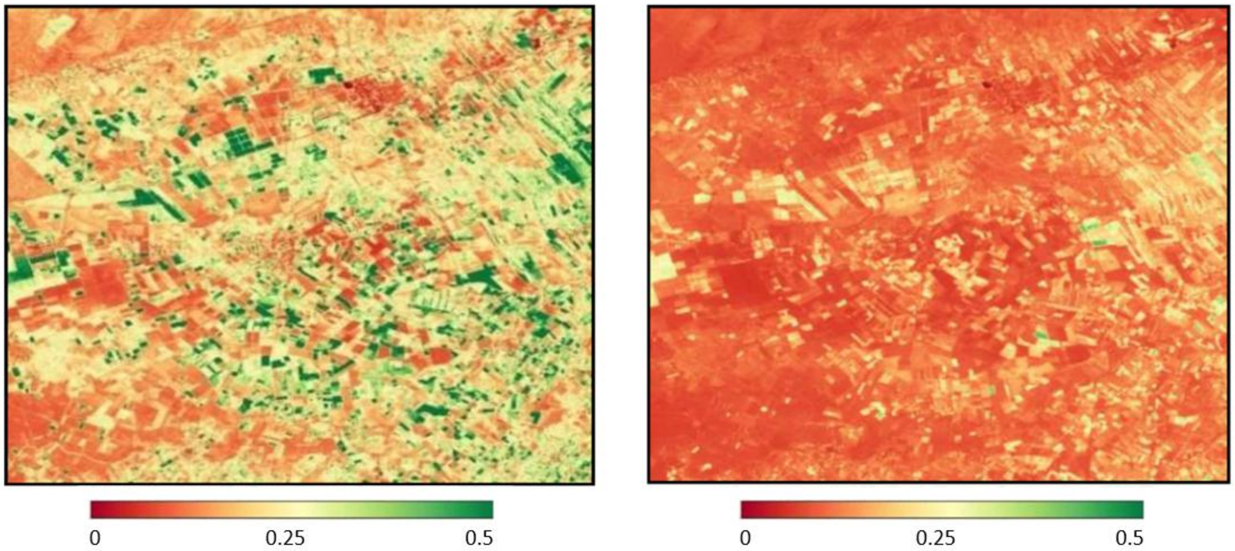
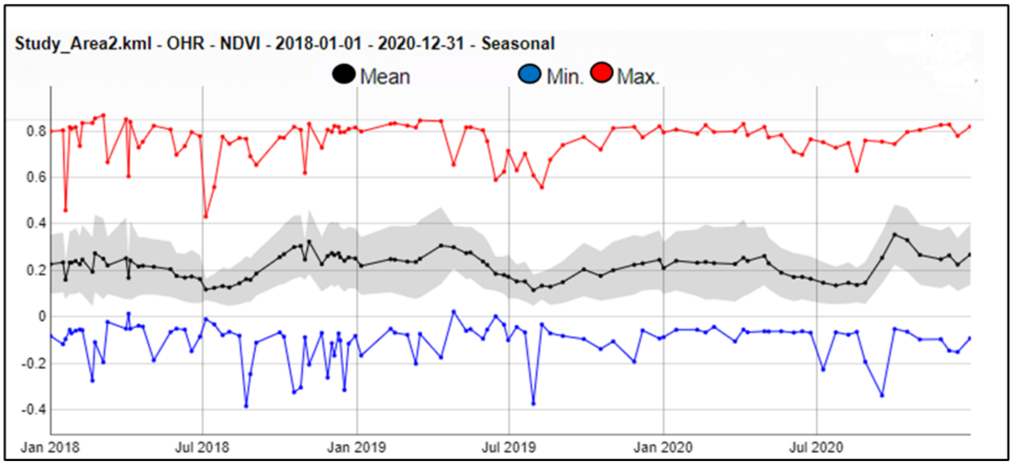


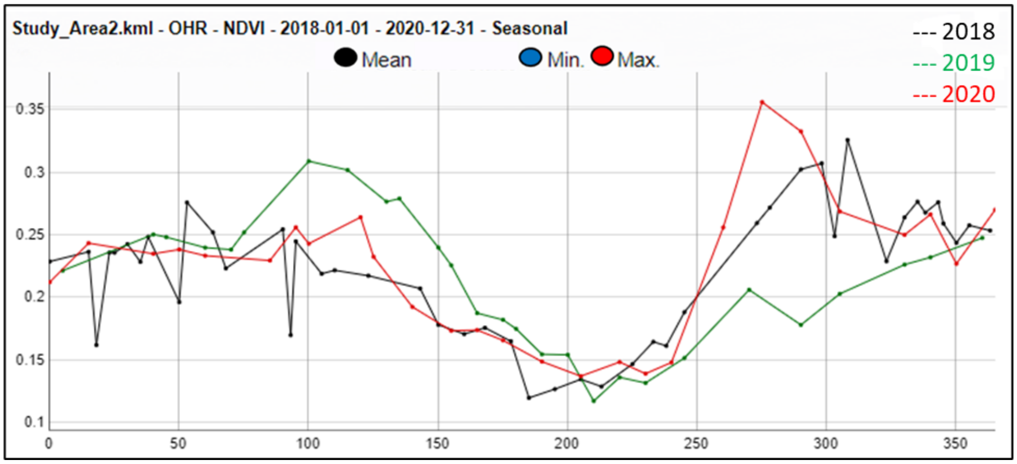
Figure 11. Normalized difference vegetation index (NDVI) of the study area during the wet season (January) and dry season (July)

The image from January 2018 depicts wet conditions, while the one from July 2019 presented dry conditions. A difference in vegetation cover between the two seasons was observed; the first image displayed relatively dense vegetation cover, whereas the second image revealed sparse vegetation cover. Figure 12a shows the variation of the normalized difference vegetation index (NDVI) of the study area. The graphs were generated using the Misbar Platform. The maximum, minimum, and mean NDVI values from January 2018 to December 2020 were analyzed. A large difference between the maximum and minimum values was observed, explained by the contrast between cultivated and non-cultivated areas, as well as by differences in plant life cycles. The variation of NDVI between 2018 and 2020 is presented in Figure 12b, where the three curves exhibit a similar seasonal trend, attributable to plant phenological cycles and cultivation practices adopted by farmers in the region.

$$\text{NDVI of the AOI} = \frac{\sum \text{Pixel Value}}{N} \qquad (\mathbf{1})$$



(a)



(b)

Figure 12. (a) The maximum and the minimum of the NDVI value of the study area, (b) variation of the NDVI value as a function of time

3.3.2.2. Characterization of soil moisture using the water index (NDWI)

Figure 13 shows two images of the Normalized Difference Water Index (NDWI) of the study region. The image for June 2018 shows the dry time, but the image of October 2018 marks the beginning of the wet season. The first image depicts several moist fields, likely corresponding to irrigated areas, while the second image reflects land surface conditions following rainfall. Based on the observed soil moisture patterns, it can be inferred that precipitation occurred less than 48 hours prior to the image acquisition date.

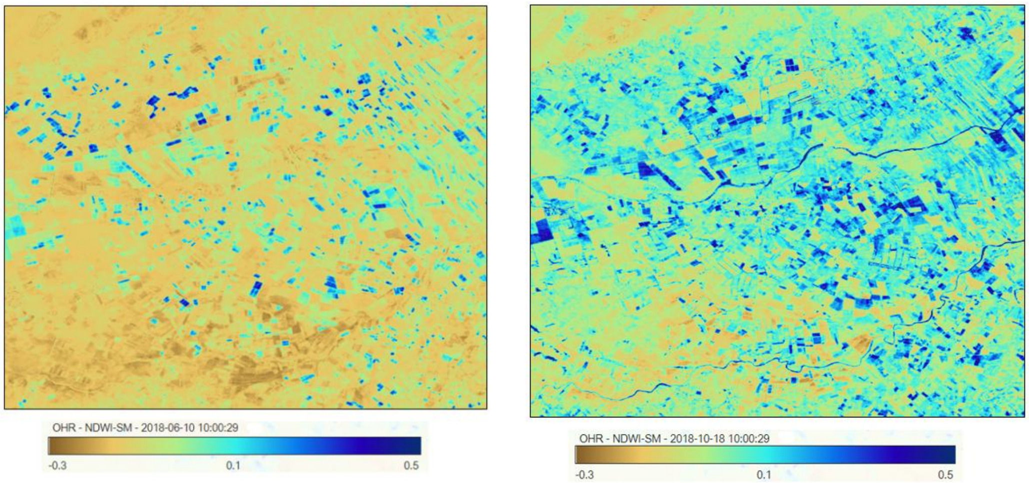
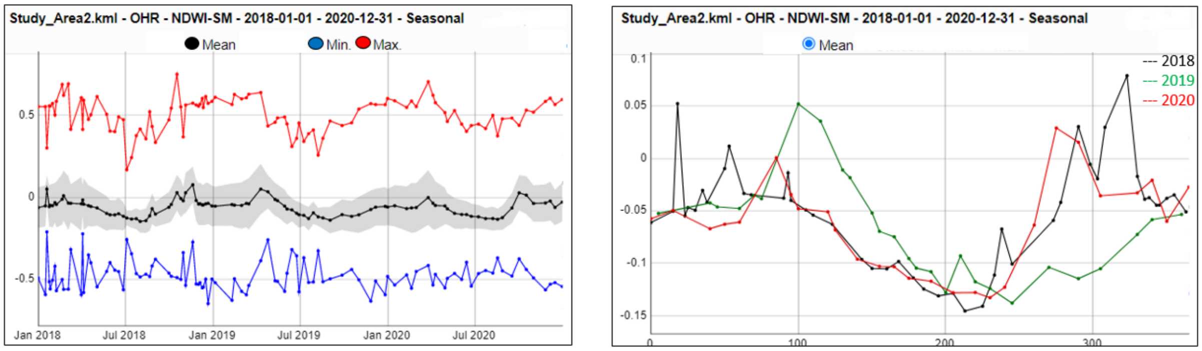

OHR - NDWI-SM - 2018-06-10 10:00:29
-0.3
0.1
0.5
OHR - NDWI-SM - 2018-10-18 10:00:29
-0.3
0.1
0.5

Figure 13. Normalized difference water index (NDWI) of the study area

The temporal variation of the Normalized Difference Water Index (NDWI) within the study area from January 2018 to December 2020 is presented in Figure 14a, illustrating the maximum, minimum, and mean values over the period. A substantial gap between the maximum and minimum values is evident, reflecting considerable heterogeneity in land use, vegetation cover, and soil properties. This suggests that soil moisture dynamics are more strongly influenced by land use patterns and soil variability than by climatic factors alone. Figure 14b depicts the interannual changes in NDWI between 2018 and 2020, where the three curves exhibit a consistent seasonal pattern, largely driven by vegetation phenology and prevailing meteorological conditions in the region. The NDVI and NDWI results showed that the study area is classified into several basic categories, which represent variations in land use /land cover and different crops. In the study area, conventional irrigation is dominant. The results showed that vegetable crops such as tomatoes, peppers and plots of olive trees are the dominant crops in the study area.



(a) (b)

Figure 14. (a) The maximum and the minimum of the NDWI value of the study area, (b) variation of the NDWI value as a function of time

Pilot Area Sub-pilot areas

To gain deeper insights into the soil moisture regime and the influence of environmental parameters, the analysis was narrowed to a specific parcel within the study area. Within this parcel, two sub-pilot zones were selected, both subjected to identical environmental and management conditions. Despite these similarities, visual inspection reveals a clear contrast in vegetation cover: Pilot 1 exhibits less developed and sparser vegetation compared to Pilot 2. To quantify these differences, a three-year analysis (2018–2020) of the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) was carried out for both sub-pilots (Figure 15).

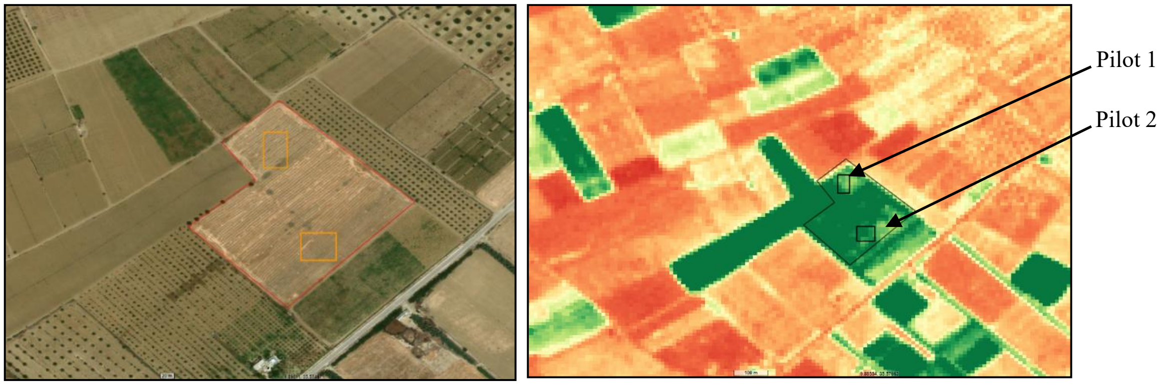
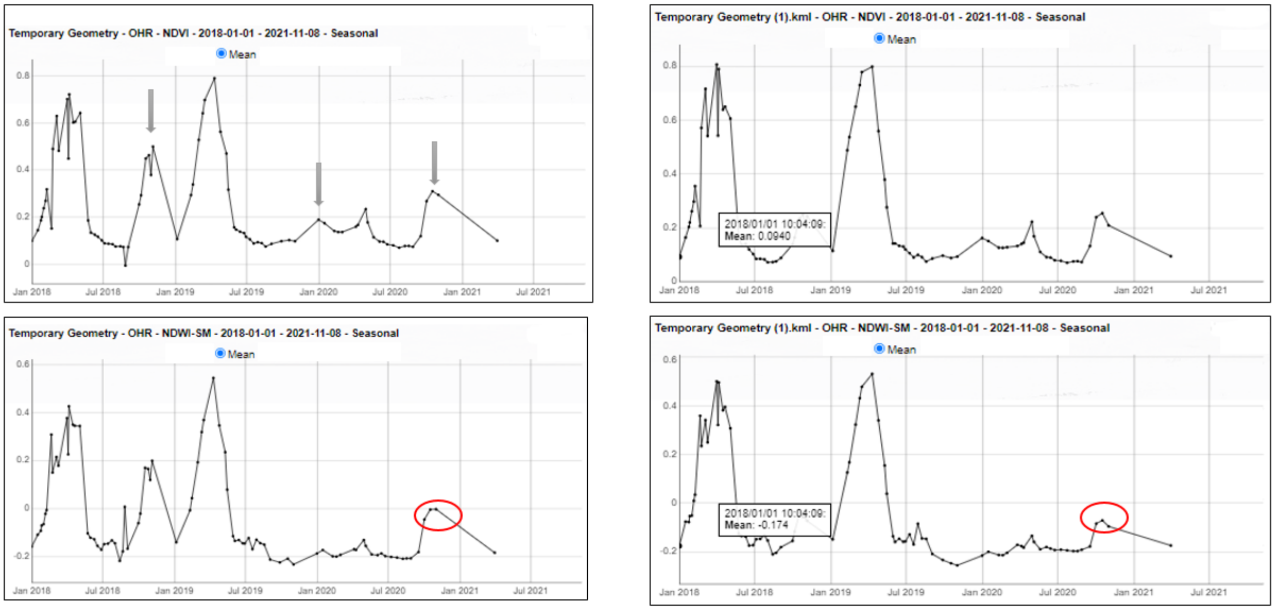


Figure 15. True-Color Imagery (Left) and NDVI (Right) for Pilot Sites 1 and 2

During the rainy season, the NDVI of Pilot 1 is lower than that of Pilot 2 (Figure 16), while it is higher during the dry seasons.



Pilot 1

Pilot 2

Figure 16. Comparison between the NDVI (top) and the NDWI (bottom) of sub-pilot areas

The NDWI values for Pilot 1 consistently exceed those of Pilot 2 in both dry and rainy seasons. This suggests that the soils in the two sub-pilots differ in their physical properties, with Pilot 1 exhibiting a higher water retention capacity than Pilot 2. However, despite this greater capacity for moisture retention, the NDVI values in Pilot 1 remain lower than those in Pilot 2. This apparent contradiction may be explained by the fact that the soil in Pilot 1 approaches water saturation more frequently, creating conditions that are less favorable for optimal vegetation growth.

3.3.2.3. Correlation between the NDVI and the NDWI

A statistical analysis was made between the NDVI and NDWI values of the study area between 2018 and 2020. Figure 17 shows the results of the linear correlation between the two indices NDVI and NDWI. A strong positive correlation is observed between the two parameters, with R = 0.79 and R² = 0.64.

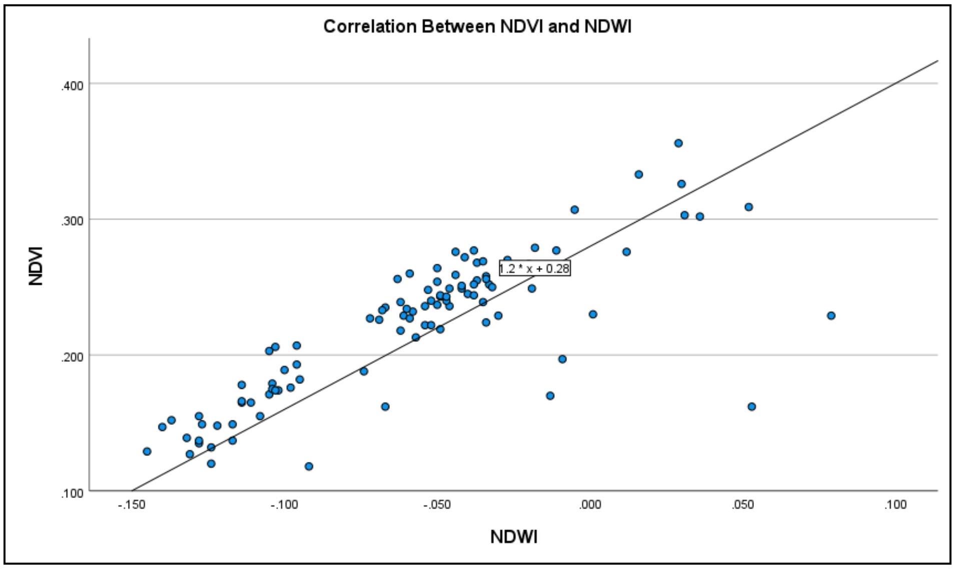
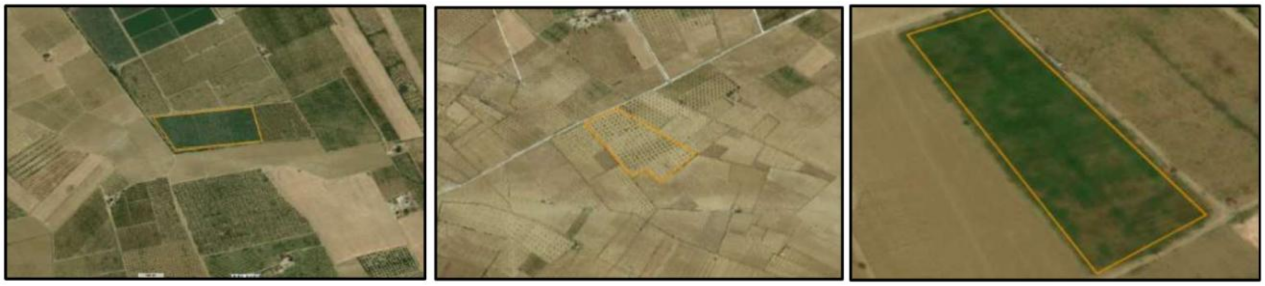


Figure 17. Correlation between NDVI and NDWI

This result is explained by the strong correlation between the water and the biomass in nature and also technically the two parameters are based on the near-infrared as one of the major points.

Some data points deviate noticeably from the correlation line, prompting further investigation to identify the underlying causes of these anomalies. As shown below, a measure of the correlation coefficient was performed between the means of NDWI-NDVI, the maximums of NDWI-NDVI and the minimums of NDWI-NDVI.

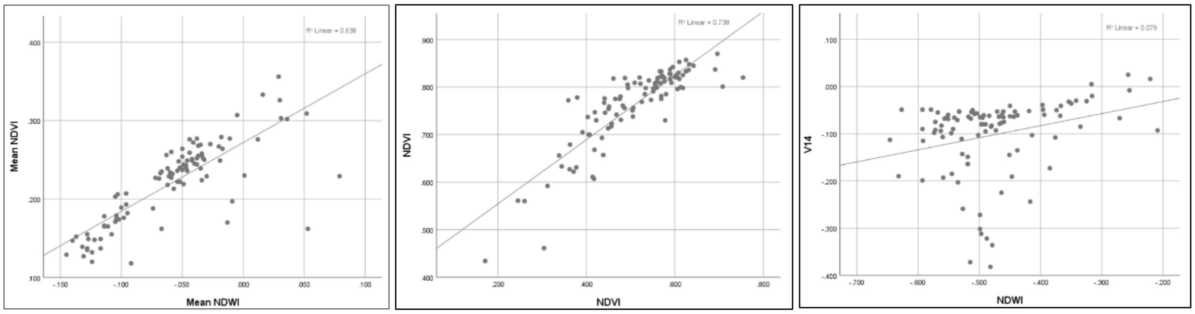


(a) Pilot 3 (b) Pilot 4 (c) Pilot 5

Figure 18. Delimitation of Pilot 3, Pilot 4 and Pilot 2

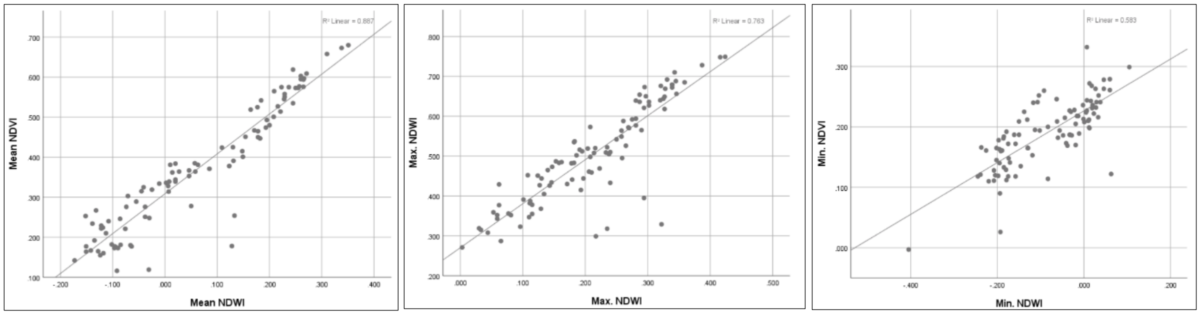
The results reveal that maximums and means have a higher correlation, whereas minimums have a lower correlation. Furthermore, a small pilot area (Figure 18a) is used for validation; its correlation coefficients were calculated as shown in the pilot curves below (Figure 19), and this pilot reveals strong correlations between maximums, means, and minimums. The weak correlation between the minimums in the study area (Figure 19) can be explained by the presence of free water or by the presence of certain urban areas. In addition, the out-of-range points could be explained by a technical error while the output results are symbolic of the same date, this error could be considered as noise.

To validate this assumption, the analysis was extended to two additional small pilot areas (Pilots 4 and 5) characterized by similar land use. Pilot 4 is predominantly covered by olive trees, whereas Pilot 5 is dedicated to horticultural crops. The results revealed a strong correlation between the mean NDWI–NDVI and the maximum NDWI–NDVI values. In contrast, the minimum NDWI–NDVI exhibited a lower correlation coefficient than both the mean and the maximum, and its behavior appeared less predictable. This weaker and less consistent relationship can be attributed to the presence of diverse land cover types, some of which may retain moisture without supporting vegetation, rendering the minimum values less representative. Furthermore, Pilots 4 and 5 (Figures 21 and 22) displayed similar outlier points to those observed in the first and second analyses. Consequently, these outliers can reasonably be considered as noise in the dataset.



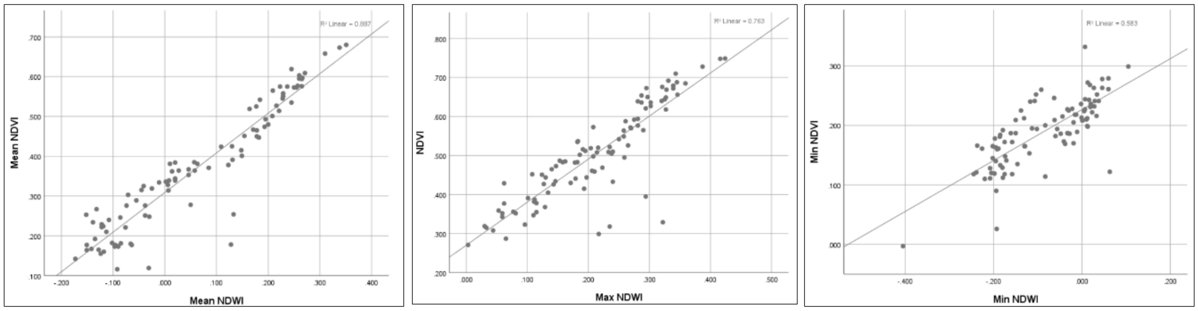
R = 0.79 R² = 0.64 R = 0.859 R² = 0.638 R = 0.28 R² = 0.079

Figure 19. Correlation between the NDWI and NDVI of the study area



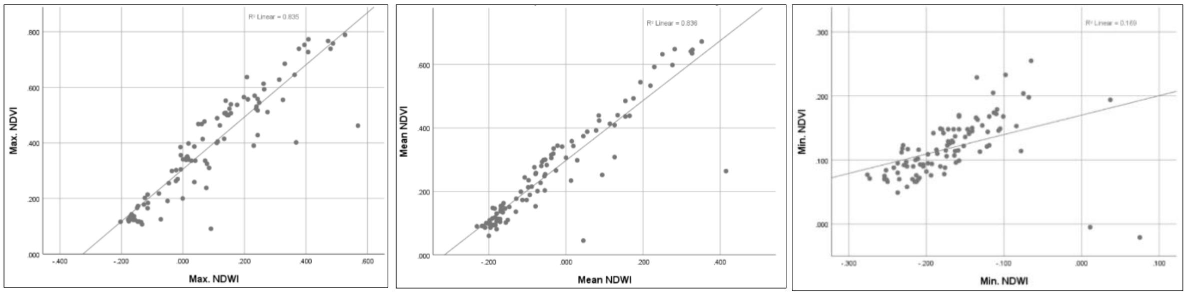
R = 0.87 R² = 0.89 R = 0.94 R² = 0.76 R = 0.764 R² = 0.58

Figure 20. Correlation between the NDWI and NDVI of Pilot 3



R = 0.94 R² = 0.89 R = 0.87 R² = 0.76 R = 0.76 R² = 0.58

Figure 21. Correlation between the NDWI and NDVI of Pilot 4



R = 0.91 R² = 0.84 R = 0.91 R² = 0.84 R = 0.41 R² = 0.17

Figure 22. Correlation between the NDWI and NDVI of Pilot 5

3.3.3. Conclusion

In this study, the correlation method was used to characterize the soil moisture regime using Sentinel-1B C-band SAR satellite data, the Normalized difference vegetation index (NDVI), and the Normalized difference water index (NDWI). Five pilots in the study area were chosen and analyzed to identify the relation between the optical and radar satellite data and the soil moisture. In order to extract the highest correlation and most informative data sources,

To validate this assumption, the analysis was extended to two additional small pilot areas (Pilots 4 and 5) characterized by similar land use. Pilot 4 is predominantly occupied by olive trees, whereas Pilot 5 is covered by horticultural crops. The correlation between different backscatters (V, H), optical data, DEM, and environmental variables was evaluated. The Normalized difference vegetation index (NDVI), the Normalized difference water index (NDWI), the Radar data (VV and VH polarization) and the soil types showed a strong correlation with the wetness and the vegetation conditions. The applied basic cognitive and statistical tools combined with local knowledge has demonstrated that these datasets have comparative potential in explaining the soil water regime and its relationship to the vegetation condition. The method is a promising and useful tool that may be used to characterize the soil moisture regime in the region. The environmental covariates should be statistically tested through the correlation methods to determine which covariates had the highest correlation with soil moisture.

Based on the conclusions above the following theses were formulated:

Thesis 1

The synergistic integration of Sentinel-1 C-band SAR backscatter (VV/VH), Sentinel-2-derived vegetation and water indices (NDVI, NDWI), and digital elevation data enables robust spatiotemporal characterization of soil moisture in the semi-arid Merguellil Basin. By capturing complementary signals related to vegetation structure, surface roughness, water content and water retention variability, this approach delineates soil wetness patterns across physiographic zones and land use types with minimal ground-based input. The method is particularly effective for characterizing topsoil moisture (0–5 cm), where SAR backscatter is most responsive, and its performance remains sensitive to vegetation density. However, limitations may arise under dense canopy cover or highly rough surfaces, where multiple scattering and signal saturation can reduce accuracy.

Thesis 2

Strong statistical correlations between radar backscatter coefficients, spectral vegetation and water indices, and topographic parameters demonstrate the capacity of multi-source remote sensing to infer soil moisture regimes under data-scarce conditions. The observed divergence in NDVI–NDWI dynamics across pilot sites further reveals the influence of soil texture, land use, and water retention variability, supporting the use of integrated optical-radar analysis for hydrological monitoring in heterogeneous agro-ecosystems. Nevertheless, performance declines under dense vegetation or highly rough surfaces, where signal penetration and retrieval accuracy are reduced.

4. Non-Invasive Soil Texture Prediction Using Random Forest and Multi-Source Environmental Data

4.1. Introduction

Traditionally, soil texture is determined through laboratory-based methods such as particle size analysis and sedimentation tests, which are labor-intensive, time-consuming, and costly (Gee & Bauder, 1986). Moreover, these methods are not always feasible for large-scale or real-time monitoring, especially in data-scarce or resource-limited regions.

Recent advances in in-situ soil sensing and remote sensing technologies have enabled the development of alternative approaches for non-invasive and scalable estimation of soil properties) (Adamchuk et al., 2004; Minasny & McBratney, 2016). Soil moisture dynamics are closely linked to soil texture, as finer-textured soils (e.g., clayey) typically exhibit greater water-holding capacity compared to coarser-textured soils (e.g., sandy) (Saxton & Rawls, 2006). In addition, vegetation indices such as the Normalized Difference Vegetation Index (NDVI), derived from satellite platforms like Sentinel-2, can serve as indirect indicators of soil conditions by capturing vegetation responses influenced by moisture and texture (Jiang et al., 2006).

Machine learning (ML) algorithms offer powerful tools for modeling complex and nonlinear relationships among environmental variables and soil properties. Among these, Random Forest Regression (RFR) has shown particular promise due to its robustness to noise, ability to model interactions, and interpretability via feature importance analysis (Breiman, 2001a) It has been successfully applied in soil science for mapping soil organic carbon, bulk density, and texture (Heung et al., 2014; Wadoux et al., 2020).

This study aims to develop a Random Forest-based predictive model for soil texture classification and fraction estimation by integrating in-situ soil moisture measurements from Sentek EnviroSCAN sensors with Sentinel-2 NDVI and atmospheric data. The objectives are threefold: (1) to demonstrate the feasibility of predicting soil particle size that defines the soil texture class without laboratory measurements, (2) to identify the most influential environmental predictors of textural variability, and (3) to assess the model's potential to support automated calibration of soil moisture sensors. The proposed approach contributes to advancing digital soil mapping and enhancing decision-making in precision agriculture and climate-resilient land management.

4.2. Study area

Hungary is located in Central Europe, within the Carpathian Basin, between 45°48′ and 48°35′ N latitude and 16°05′ and 22°58′ E longitude. The country’s topography varies from lowlands to mountainous regions, with its highest peak reaching 1014 m, while the lowest point is at 78 m above sea level. The entire territory falls within the Danube River basin, which significantly influences its hydrological and geomorphological characteristics (Kocsis et al., 2021).

The landscape of Hungary is primarily characterized by low elevation and minimal vertical dissection (Figure 23). Approximately 82.4% of the country lies below 200 m, predominantly forming the Great Hungarian Plain (Alföld). In contrast, only 0.5% of the land area exceeds 500 m in elevation, with medium-height mountains (200–500 m) covering 2.1% of the terrain. Additionally, hills and foothills constitute 15.5% of the country's relief, contributing to the diverse topographical and ecological conditions across the region (Gábris et al., 2018).

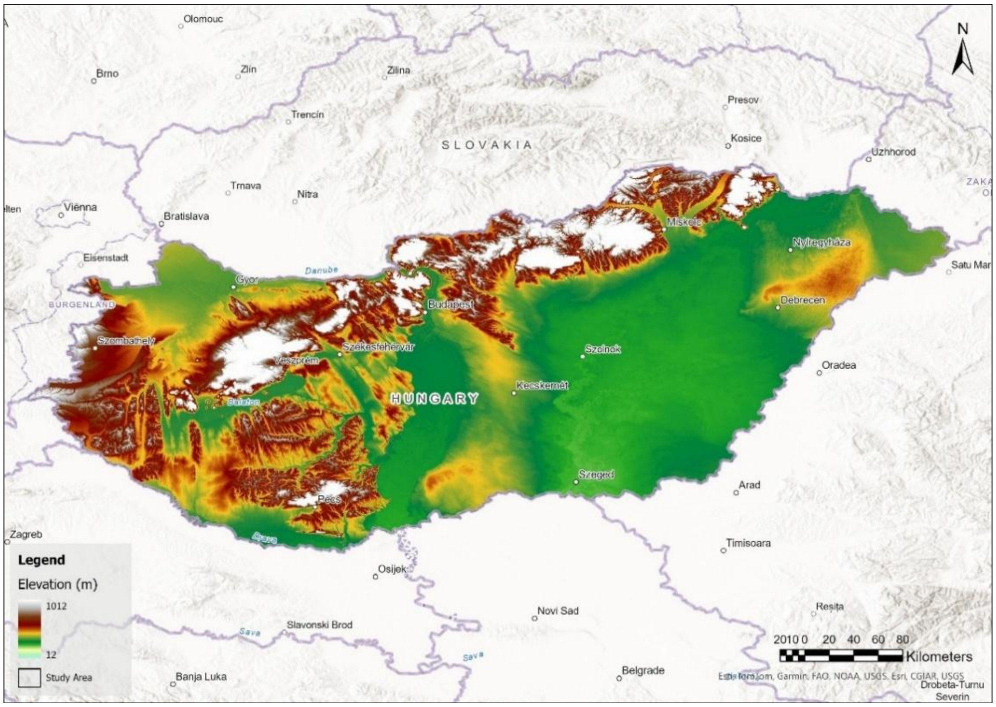


Figure 23. Location and (DEM-based) topography of the study area (Hungary)

Hungary lies within the northern temperate climatic zone, yet its weather patterns are shaped by the interaction of three major climate influences: oceanic, continental, and Mediterranean systems. The country experiences four distinct seasons, marked by significant temporal variability in temperature and precipitation. Typically, summer is the warmest season and receives the most precipitation, whereas winter is the coldest and driest. However, precipitation patterns are highly variable, both in space and time. Notably, 2010 was the wettest recorded year, featuring an average of 9 days with intense rainfall events exceeding 20 mm (Spinoni et al., 2015).

4.3. Materials and Methods

4.3.1. Materials

4.3.1.1. Soil Texture

Soil particle size distribution was analyzed from samples collected in the upper 0–30 cm soil layer at 25 field sites across Hungary, selected to represent a broad spectrum of geographic and geological conditions. The geographical distribution of these sampling locations, along with the positions of the in-situ sensors, is presented in Figure 24, highlighting the spatial variability of soil properties and monitoring points across the study region.

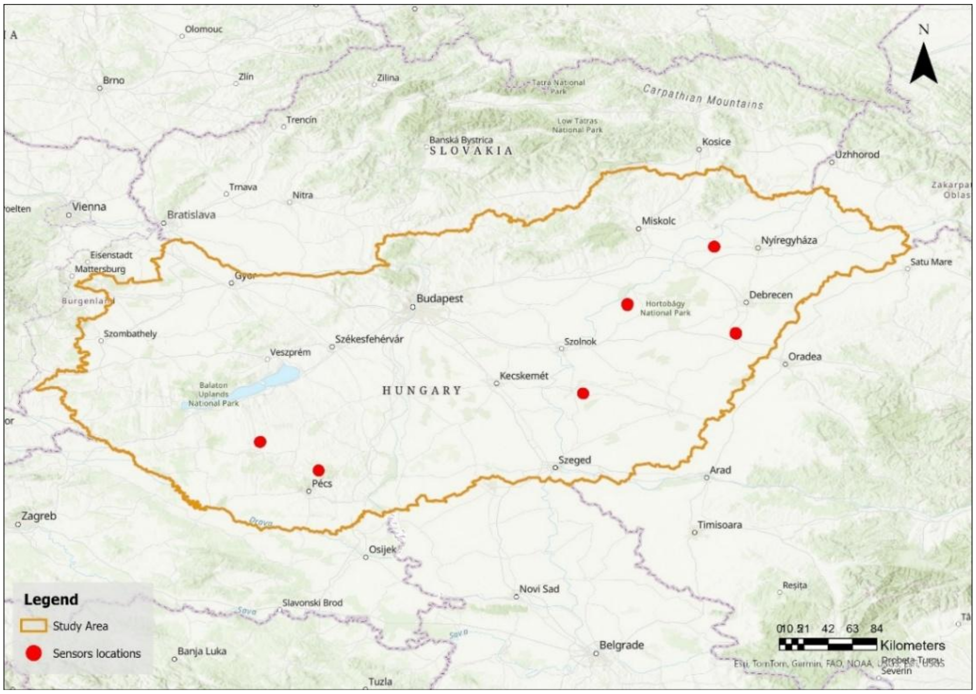


Figure 24. Sensor’s locations in the study areas

Soil samples were analyzed using standard granulometric procedures to determine the relative proportions of sand, silt, and clay fractions. These values were subsequently used to classify the soils according to the United States Department of Agriculture (USDA) soil texture classification system (Figure 25), which is widely used for its precision in categorizing soil types based on particle size distribution (Nrcs, 2017).

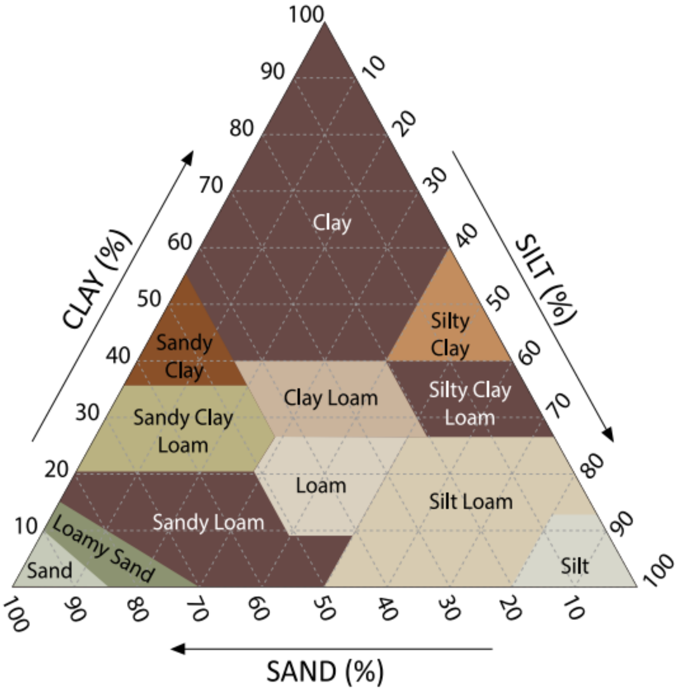


Figure 25. USDA Soil Texture Triangle used to classify soil based on proportions of sand, silt, and clay. Source: Soil Science Society of America.

The soil particle size and the soil texture classes of our dataset is provided in Table 2, showing how soil texture of this study is classified within the USDA framework.

Table 2. Soil particle size and soil texture (USDA System)

Soil Site Reference Name	Sand	Silt	Clay	Soil Texture Class
Tiszavasvari_01	37.6	38.8	23.6	Loam
Tiszavasvari_02	32.7	43.2	24.2	Loam
Tiszavasvari_03	34.4	41.6	24.0	Loam
Tiszavasvari_04	64.8	23.9	11.3	Sandy Loam
Tiszavasvari_05	25.9	46.9	27.3	Clay Loam

Tiszavasvari_17	56.3	26.2	17.5	Sandy Loam
Somodor_4	59.4	21.7	19.1	Sandy Loam
Somodor_13	52.2	26.3	21.5	Sandy Clay Loam
Somodor_21	69.0	18.6	12.4	Sandy Loam
Urbán_4	12.4	27.3	60.3	Clay
Urbán_17	51.3	22.1	26.6	Sandy Clay Loam
Tépe_06	30.3	40.4	29.3	Clay Loam
Tépe_08	38.0	39.9	22.2	Loam
Tépe_09	45.6	27.3	27.2	Sandy Clay Loam
Tépe_12	47.7	20.9	31.5	Sandy Clay Loam
Tépe_13	39.2	32.1	28.8	Clay Loam
Magyaregregy_10	60.9	19.8	19.4	Sandy Loam
Magyaregregy_11	48.6	27.7	23.7	Sandy Clay Loam
Magyaregregy_14	47.1	34.9	18.0	Loam
Kunszentmárton_7	21.0	44.5	34.5	Clay Loam
Kunszentmárton_18	22.0	30.2	47.7	Clay
Kunszentmárton_19	25.3	32.6	42.1	Clay
Matyo_12	32.5	31.4	36.1	Clay Loam
Matyo_17	26.6	27.5	45.9	Clay
Matyo_21	27.5	31.9	40.6	Clay

Figure 26 shows the USDA soil texture triangle with the distribution of 25 analyzed samples. The samples span a wide range of textural classes, including sandy loam, loam, clay loam, sandy clay loam, and clay. Most samples cluster within the loam to clay loam classes, while a few occupy sandy loam and sandy clay loam categories. This distribution reflects the heterogeneity of soil particle size composition across the study sites.

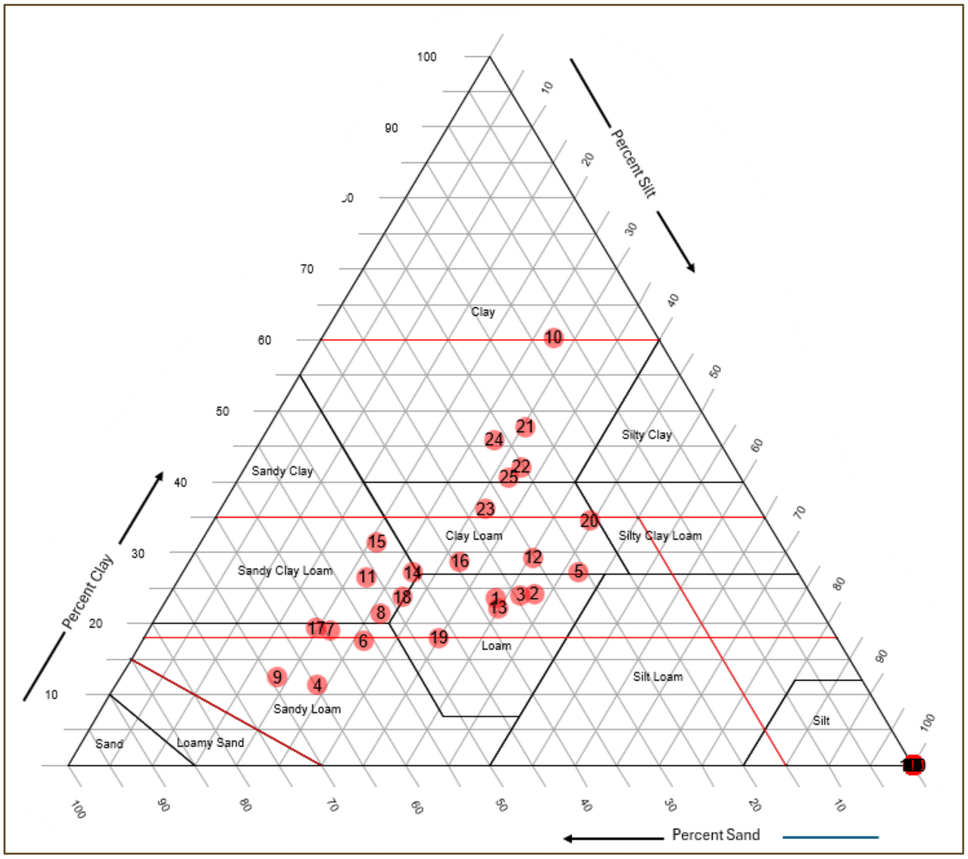


Figure 26. Soil Textiure Distribution of Soil Samples Plotted on the USDA Soil Texture Triangle

4.3.1.2. EnviroScan Sensor

The Sentek EnviroScan sensor consists of multiple capacitance-based sensors installed at 10 cm intervals (Figure 27) along an extruded plastic framework (Paltineanu & Starr, 1997); Hajdu et al., 2019). The probe had sensors at six depths, namely 10, 20, 30, 40, 60 and 100 cm. Each sensor comprises two brass rings forming a capacitor, connected to an LC oscillator, where frequency variations correspond to changes in soil capacitance (Hajdu et al., 2019). The sensor generates an oscillating capacitance field, extending beyond the PVC access pipe into the surrounding soil, allowing measurement of soil moisture content based on frequency shifts. A data logger records output counts, which are scaled between air (dry) and water (saturated) reference readings to determine soil moisture (Provenzano et al., 2020) (Z. Gao et al., 2018).

Capacitance probes are valued for their robustness, accuracy, and fast response times (Dane & Topp, 2002). However, their performance is highly dependent on good contact between the access tube and the surrounding soil, as poor contact reduces sensitivity (Kelleners et al., 2004); (de Rosny et al., 2001); (Scobie, 2006)

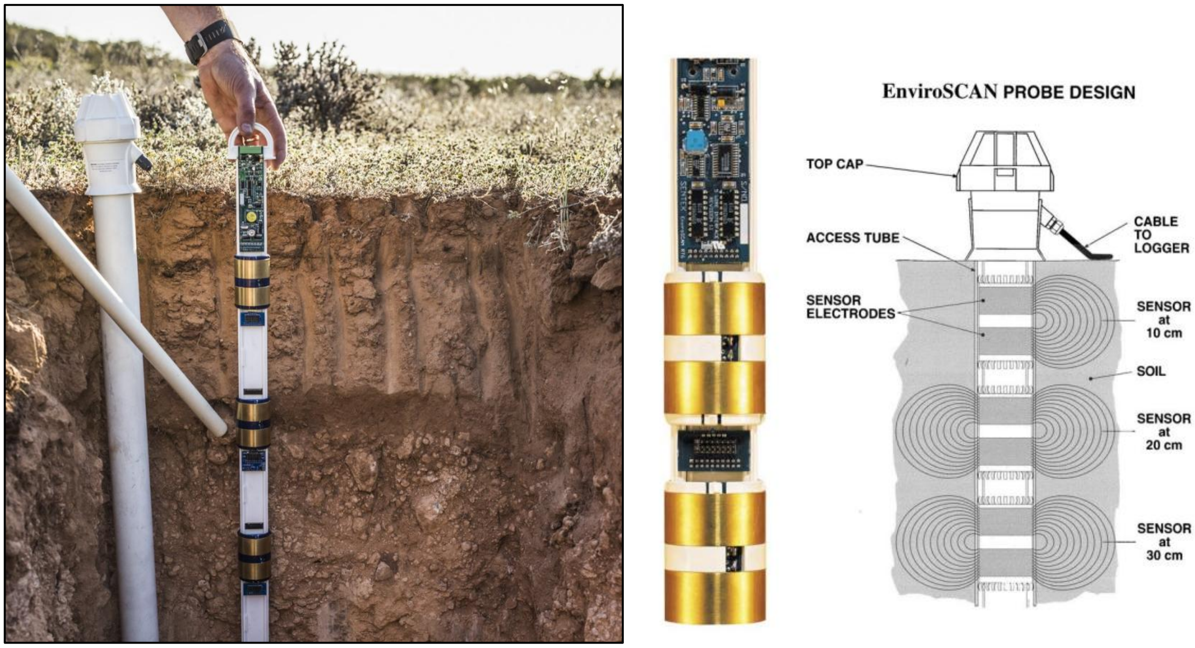


Figure 27. Sentek EnviroSCAN® Sensor for Profiling Water Content Along an Access Tube (Sentek Pty Ltd., 2011)

A total of twenty-five distinct field sites across Hungary, representing different soil textures, were selected for this study. At each site, one Sentek EnviroScan sensor access pipe was installed.

To ensure minimal soil disturbance and optimal sensor-soil contact, the installation process followed a rigorous methodology. A dedicated installation kit was employed to insert the PVC access tubes, which were stabilized using a tripod with vertical leveling capability anchored securely to the ground. A soil auger was systematically used to remove soil from within the tube, facilitating deeper penetration and ensuring a precise cut. The PVC tube was then progressively inserted into the prepared borehole until a depth of 120 cm was reached.

Following installation, the interior of the PVC tube was meticulously cleaned, and a compression rubber plug was inserted at the bottom to act as a barrier against water and vapor infiltration. To maintain structural integrity and prevent contamination, the male-threaded section at the top, designed to secure the probe’s removable screw cap, was sealed with silicone glue. These measures were critical to ensuring optimal soil-tube contact, eliminating air gaps, and preventing preferential water flow, which could otherwise compromise measurement accuracy (Starr & Paltineanu, 1998).

4.3.1.3. Remote sensing data

While vegetation can influence soil moisture dynamics, its effects are highly dependent on site-specific conditions and soil depth. According to Yang et al., 2012, slope position and aspect primarily affect soil moisture content (SMC) in shallow layers, whereas slope gradient exerts a significant control on both shallow and deep layers. The planted vegetation exhibits a negative relationship with deep soil moisture, largely due to transpiration and root water uptake, which drive spatial variability in deep SMC. Given this interaction, vegetation dynamics play a crucial role in soil moisture variability, making the assessment of vegetation conditions essential for understanding soil water distribution. On the other hand, better water supply results in better plant condition, reflecting the differences of the water supplying capacity of the soils.

The Normalized Difference Vegetation Index (NDVI) serves as a key indicator for analyzing vegetation conditions and their impact on soil moisture (Myneni et al., 1995).

The Normalized Difference Vegetation Index (NDVI) is widely recognized as a reliable indicator of vegetation health and has also been shown to exhibit a relationship with soil moisture variability under different environmental conditions (Gu et al., 2008). It is computed using the formula:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{2}$$

where:

- NIR represents the reflectance in the near-infrared band (Sentinel-2 Band 8).
- R represents the reflectance in the red band (Sentinel-2 Band 4).

3.3.1.4. Data Description

Figure 28 presents the set of input variables used in the machine learning framework together with the target feature, soil texture.

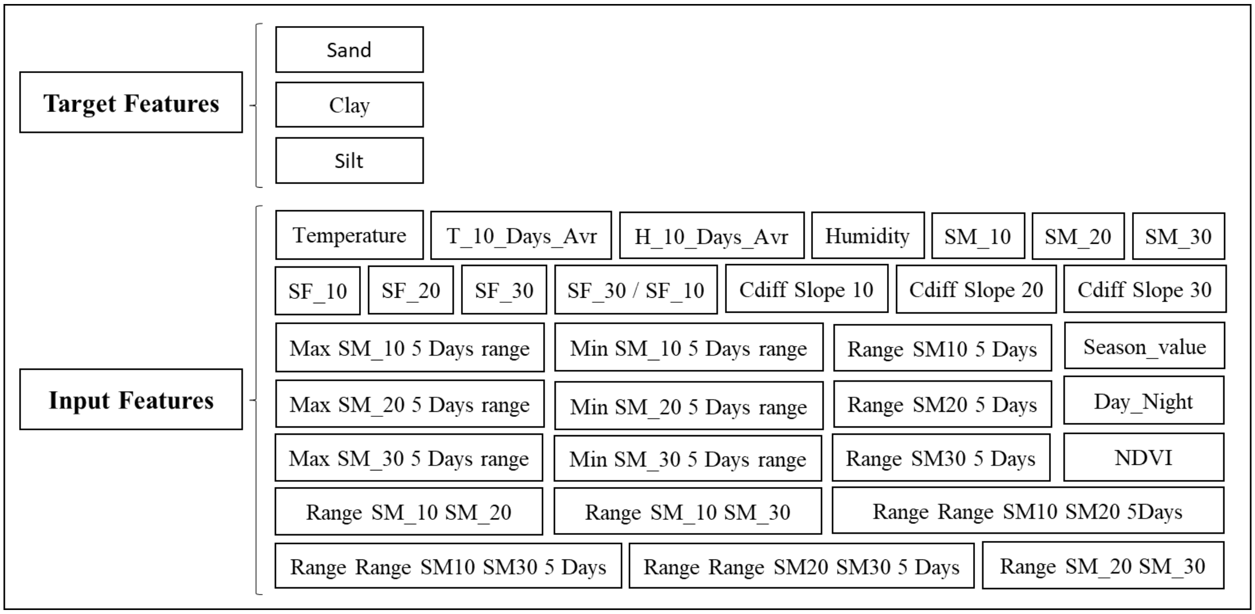


Figure 28. Input Features and Target Variable for Soil Texture Prediction

4.3.1.4.1. Target variables

The target variables in this study are the soil particle size classes, as defined by the United States Department of Agriculture (USDA). It is essential to distinguish between soil particle size and soil texture. While soil texture refers to the relative proportion of sand, silt, and clay particles and determines many physical behaviors of the soil (e.g., water retention, aeration, tillage suitability), particle size classes refer strictly to the classification of mineral particles based on their diameters, independent of their relative abundance or combined behavior (Nrcs, 2017).

According to USDA classification (Nrcs, 2017) mineral particles are grouped into three major particle size classes:

Sand: Particles with diameters between 0.05 mm and 2.00 mm. Sand particles are large and result in soils with high permeability, rapid drainage, and low water- and nutrient-holding capacities. Sandy soils are generally well-aerated but are more prone to erosion and leaching (Weil & Brady, 2017).

Silt: Particles ranging from 0.002 mm to 0.05 mm. Silt imparts a smooth, floury texture to soil and contributes to improved water-holding capacity and fertility. Silty soils retain moisture better than sandy soils and provide moderate permeability, although they are susceptible to surface crusting and compaction (Nrcs, 2017).

Clay: Particles less than 0.002 mm in diameter. Clay has the highest surface area and strong electrochemical activity. Soils dominated by clay are highly retentive of water and nutrients but have low infiltration rates and are often subject to swelling, shrinking, and structural limitations (Weil & Brady, 2017; Nrcs, 2017).

These USDA-defined particle size classes serve as the foundation for determining soil texture classes using the USDA Textural Triangle, which integrates the percentages of sand, silt, and clay to classify soils into texture classes such as loam, clay loam, or sandy loam (Nrcs, 2017).

4.3.1.4.2. Input features

Temperature: The Sentek EnviroSCAN probe is an advanced multi-depth soil moisture monitoring system that integrates temperature measurements as a key component of its data collection capabilities. By continuously measuring soil temperature and moisture at multiple depths, the probe provides comprehensive insights into thermal dynamics, soil-water interactions, and plant root-zone conditions (Sentek, 2023). These parameters are essential for understanding soil hydrological behavior, optimizing irrigation strategies, and assessing plant stress levels under varying environmental conditions.

The temperature measurement in the Sentek EnviroSCAN probe is typically performed using a thermistor or resistance temperature detector (RTD), both of which offer high-precision detection of soil temperature fluctuations. The sensor is positioned within the probe's sensor array at 10 cm depths, ensuring that temperature variations are captured concurrently with soil moisture levels. This integrated approach enables a more accurate assessment of soil-water balance, evapotranspiration rates, and thermal conductivity, all of which significantly influence water retention capacity, microbial activity, and root development (Paltineanu & Starr, 1997).

The combined measurement of temperature and soil moisture is particularly valuable in precision agriculture, climate modeling, and environmental monitoring. By correlating temperature fluctuations with soil moisture dynamics, the Sentek EnviroSCAN system facilitates the prediction of drying trends, frost risks, and soil permeability changes, contributing to data-driven decision-making in sustainable land and water management.

T_10_Days_Avr: T_10_Days_Avr represents a 10-day moving window average of temperature, where the mean temperature is computed continuously over a sliding 10-day period. This approach smooths short-term fluctuations and highlights progressive trends in temperature variation.

Humidity: The capacitance-based sensors within the Sentek EnviroSCAN probe detect changes in soil dielectric properties, which are strongly influenced by soil humidity and moisture levels. Higher soil moisture content correlates with higher soil humidity, reducing evaporation rates and influencing plant water uptake efficiency. Conversely, drier soils exhibit lower humidity levels, affecting microbial activity and increasing the likelihood of water stress conditions for vegetation (Paltineanu & Starr, 1997).

By continuously monitoring soil moisture variations, the Sentek EnviroSCAN system provides an indirect assessment of soil humidity, supporting irrigation optimization, drought monitoring, and soil health analysis. This capability makes it a valuable tool for precision agriculture, climate modeling, and environmental research, where understanding soil moisture-humidity interactions is critical for sustainable land and water management.

H_10_Days_Avr: H_10_Days_Avr represents the 10-day moving window average humidity, calculated by averaging humidity values over a continuously updating 10-day period. This moving average method smooths short-term fluctuations while capturing long-term trends in humidity variation.

SF_10, SF_20, SF_30: The Sentek EnviroSCAN probe operates on a capacitance-based sensing principle, where the scaled frequency (SF) is a key parameter used to determine soil moisture content. The probe emits an electromagnetic signal that interacts with the surrounding soil, and the resulting resonant frequency is influenced by the dielectric properties of the soil-water matrix (Sentek, 2023). This frequency is then scaled and calibrated to provide an accurate estimation of volumetric water content (VWC) across different soil types and conditions.

The scaled frequency (SF) is derived from the raw frequency output of the sensor and is normalized to minimize variations caused by sensor drift, environmental conditions, and soil texture differences (Paltineanu & Starr, 1997).

By continuously monitoring SF values, the Sentek EnviroSCAN system enables precise tracking of soil moisture fluctuations, which is crucial for irrigation management, drought assessment, and soil hydrology studies. The integration of scaled frequency data with other environmental parameters further enhances soil-water modeling, making the probe a valuable tool in agricultural and environmental research.

SF_10 is the scaled frequency at 10 cm depth, SF_20 is the scaled frequency at 20 cm depth and SF_30 is the scaled frequency at 30 cm depth,

SM_10, SM_20, SM_30: The Sentek EnviroSCAN probe determines soil moisture content using scaled frequency (SF), which is derived from the sensor's raw frequency response influenced by the dielectric properties of the soil-water matrix (Sentek, 2023). This relationship allows the probe to provide real-time, depth-specific soil moisture readings, which are essential for irrigation management, drought monitoring, and soil hydrology assessments (Paltineanu & Starr, 1997).

The SM_10 is the soil moisture at 10 cm depth, The SM_20 is the soil moisture at 20 cm depth and the SM_30 is the soil moisture at 30 cm depth.

SF_30 / SF_10: The ratio of SF30 / SF10 represents the relative difference in scaled frequency (SF) values between the 30 cm and 10 cm soil depths, providing insights into the vertical trend of soil moisture distribution. Since SF is directly related to soil moisture content, rather than inversely, a higher SF_30 / SF_10 ratio suggests that the deeper layer (30 cm) has relatively higher SF values than the upper layer (10 cm).

Max SM_10 5 Days range: The Max SM_10 5 Days Range represents the maximum recorded soil moisture (SM) value at 10 cm depth over a 5-day moving window. This metric is continuously updated as a moving value, capturing short-term fluctuations in soil moisture due to precipitation events, irrigation cycles, and evapotranspiration processes.

Min SM_10 5 Days range: The Min SM_10 5 Days Range represents the minimum recorded soil moisture (SM) value at 10 cm depth over a 5-day moving window. This metric is dynamically updated, reflecting short-term variations in surface soil moisture due to factors such as evapotranspiration, drainage, and water uptake by plant roots.

Range SM_10 5 Days: The Range SM_10 5 Days is defined as the difference between the Max SM_10 5 Days Range and the Min SM_10 5 Days Range over a 5-day moving window. This metric quantifies short-term fluctuations in soil moisture at the 10 cm depth, providing insights into moisture dynamics, infiltration efficiency, and evaporation rates.

Max SM_20 5 Days range: The Max SM_20 5 Days Range represents the highest recorded soil moisture (SM) value at 20 cm depth over a 5-day moving window. This metric captures short-term variations in subsurface moisture, reflecting water infiltration, retention capacity, and plant root-zone availability.

Min SM_20 5 Days range: The Min SM_20 5 Days Range represents the lowest recorded soil moisture (SM) value at 20 cm depth over a 5-day moving window. This metric reflects short-term trends in subsurface soil drying, providing insights into water depletion, root water uptake, and percolation losses.

Range SM_20 5 Days: The Range SM_20 5 Days is defined as the difference between the Max SM_20 5 Days Range and the Min SM_20 5 Days Range over a 5-day moving window. This metric quantifies short-term soil moisture fluctuations at the 20 cm depth, offering insights into subsurface water dynamics, infiltration rates, and plant root-zone moisture variability.

Max SM_30 5 Days range: The Max SM_30 5 Days Range represents the highest recorded soil moisture (SM) value at 30 cm depth over a 5-day moving window. This metric provides insights into subsurface moisture retention, deep infiltration dynamics, and water availability for deeper-rooted vegetation.

Min SM_30 5 Days range: The Min SM_30 5 Days Range represents the lowest recorded soil moisture (SM) value at 30 cm depth over a 5-day moving window. This metric is crucial for assessing deep soil drying trends, water availability for deep-rooted plants, and subsurface moisture retention over time.

Range SM_30 5 Days: The Range SM_30 5 Days is defined as the difference between the Max SM_30 5 Days Range and the Min SM_30 5 Days Range over a 5-day moving window. This metric quantifies short-term fluctuations in deep soil moisture, providing insights into water infiltration, retention capacity, and plant root-zone stability at 30 cm depth.

Range SM_10 SM_20: The Range SM_10 SM_20 is calculated as the difference between the soil moisture content at 10 cm (SM_10) and 20 cm (SM_20). This metric provides insights into the vertical gradient of soil moisture between shallow and intermediate soil layers, helping assess water infiltration, retention, and plant root-zone moisture availability.

Range SM_10 SM_30: The Range SM_10 SM_30 is defined as the difference between the soil moisture content at 10 cm (SM_10) and 30 cm (SM_30). This metric provides insights into vertical soil moisture distribution, infiltration efficiency, and subsurface water retention over time.

Range SM_20 SM_30: The Range SM_20 SM_30 is calculated as the difference between the soil moisture content at 20 cm (SM_20) and 30 cm (SM_30). This metric provides insights into moisture distribution and water movement between intermediate and deeper soil layers, aiding in the assessment of subsurface infiltration, retention, and plant water uptake efficiency.

Range Range SM_10 SM_20 5 days: The Range Range SM_10 SM_20 5 days is calculated as the difference between the Range SM_10 5 Days and the Range SM_20 5 Days, providing insights into short-term moisture variability between shallow (10 cm) and intermediate (20 cm) soil layers over a 5-day moving window.

A higher Range Range SM_10 SM_20 5 Days suggests that moisture fluctuations at 10 cm depth are more pronounced compared to 20 cm depth, often due to surface evaporation, rainfall events, or irrigation cycles affecting the upper soil layer more rapidly. Conversely, a lower value indicates similar moisture variability at both depths, suggesting balanced infiltration, consistent soil moisture retention, or similar drying rates across these layers.

Range Range SM_10 SM_30 5 days: The Range Range SM_10 SM_30 5 days is calculated as the difference between the Range SM_10 5 Days and the Range SM_30 5 Days, providing insights into short-term moisture variability between the shallow (10 cm) and deeper (30 cm) soil layers over a 5-day moving window.

A higher Range Range SM_10 SM_30 5 Days indicates that moisture fluctuations at 10 cm depth are more pronounced compared to 30 cm depth, often due to surface evaporation, precipitation events, or irrigation cycles that impact the upper soil layer more rapidly. Conversely, a lower value suggests more uniform moisture fluctuations between the two depths, indicating consistent infiltration, stable deep moisture retention, or minimal difference in drying rates.

Range Range SM_20 SM_30 5 days: The Range Range SM_20 SM_30 5 days is calculated as the difference between the Range SM_20 5 Days and the Range SM_30 5 Days, providing insights into short-term moisture variability between intermediate (20 cm) and deeper (30 cm) soil layers over a 5-day moving window.

A higher Range Range SM_20 SM_30 5 Days suggests that moisture fluctuations at 20 cm depth are greater than at 30 cm, indicating differences in water infiltration rates, root-zone water uptake, or varying soil retention capacities between these layers. Conversely, a lower value implies similar moisture variability across both depths, suggesting consistent water movement, uniform soil structure, or stable subsurface moisture conditions.

CDiff Slope 10: The CDiff Slope 10 represents the rate of change of the soil moisture at 10 cm depth, calculated using the central difference method. This metric quantifies how the slope varies between consecutive depth measurements, providing insights into gradients in soil properties, moisture dynamics, or other related environmental factors at this depth.

Mathematically, CDiff Slope 10 is computed as:

$$CDiff\ Slope\ 10 = \frac{SL_{10_{i+1}} - SL_{10_{i-1}}}{2} \qquad \mathbf{(3)}$$

where:

- SL_10i+1 is the next SL_10 value,
- SL10i−1 is the previous SL_10 value,
- The denominator 2 ensures a centered approximation of the slope.

A higher CDiff Slope 10 indicates steeper changes in SL_10 values, suggesting rapid variations in soil conditions at 10 cm depth, while a lower slope suggests a more stable and uniform SL_10 distribution.

CDiff Slope 20: The CDiff Slope 20 represents the rate of change of soil moisture at 20 cm depth, calculated using the central difference method. This metric quantifies how the slope varies relative to its neighboring values, providing insights into moisture distribution, soil compaction effects, and subsurface hydrological gradients at this depth.

Mathematically, CDiff Slope 20 is computed as:

$$CDiff\ Slope\ 20 = \frac{SL_{20_{i+1}} - SL_{20_{i-1}}}{2} \tag{4}$$

where:

- SL_20i+1 is the next SL_20 value,
- SL_20i−1 is the previous SL_20 value,
- The denominator 2 ensures a centered approximation of the slope.

A higher CDiff Slope 20 suggests rapid variations in SL20, indicating changes in subsurface water movement, soil structural shifts, or variations in infiltration patterns. Conversely, a lower slope suggests a more uniform SL20 profile, indicative of stable moisture retention and minimal vertical fluctuations.

CDiff Slope 30: The CDiff Slope 30 represents the rate of change of the soil moisture at 30 cm depth, calculated using the central difference method. This metric quantifies the variation in slope values relative to its neighboring points, providing insights into deep soil moisture movement, infiltration dynamics, and subsurface hydrological behavior.

Mathematically, CDiff Slope 30 is expressed as:

$$CDiff\ Slope\ 30 = \frac{SL_{30_{i+1}} - SL_{30_{i-1}}}{2} \tag{5}$$

where:

- SL_30i+1 is the next SL_30 value,
- SL_30i−1 is the previous SL_30 value,
- The denominator 2 ensures a centered approximation of the slope.

A higher CDiff Slope 30 indicates significant variations in soil moisture or structural changes at 30 cm depth, which may suggest deep percolation, subsurface compaction effects, or differential water retention. Conversely, a lower slope suggests more uniform conditions, indicating consistent deep moisture retention and minimal subsurface fluctuations.

Season_value: The Season_Value is a normalized time variable that represents the progression of a specific seasonal interval, ranging from January 30 at 23:00 to July 15 at 23:00, covering approximately 165 days or 3960 hours (165 days × 24 hours). This variable provides a continuous, scaled representation of time, where:

- January 30, 23:00 is assigned a value of 0 (seasonal start).
- July 15, 23:00 is assigned a value of 1 (seasonal end).
- Every hour within this interval is assigned a proportional value between 0 and 1, ensuring a smooth, normalized transition across the time range.

This seasonal normalization ensures that even if there are missing hourly records, each hour retains an interpolated value within the defined time scale. Season_Value is useful for capturing seasonal variations in environmental or meteorological studies, allowing for time-dependent trend analysis in soil moisture dynamics, temperature fluctuations, or climate-driven agricultural assessments.

Day_Night: The Day_Night variable, also referred to as Night_Value, is a normalized, scaled variable representing a fixed daily time interval between 23:00 (11 PM) and 04:00 (4 AM). This period spans 5 hours, and the normalization is applied consistently for every day in the dataset, ensuring a standardized representation of night-time conditions.

Mathematically, the normalization follows:

- 23:00 (11 PM) is assigned a value of 0.
- 04:00 (4 AM) is assigned a value of 1.
- Each hour within this range is assigned an incremental value based on its position within the 5-hour interval.

This approach allows Day_Night to be used as a continuous feature in modeling diurnal variations in environmental conditions, particularly in soil moisture fluctuations, temperature dynamics, evapotranspiration, and energy balance assessments.

NDVI: The Normalized Difference Vegetation Index (NDVI) is a widely used remote sensing metric for assessing vegetation health, biomass, and photosynthetic activity, which is closely linked to soil moisture availability and variability. Sentinel-2 Level-2A is a surface reflectance products, with 10 meters spatial resolution, it is a high-resolution satellite operated by the European Space Agency (ESA), provides multispectral imagery. The NDVI was pre-processed and calculated on the cloud platform using Sentinel Hub services, which provide on-demand access to Sentinel-2 Level-2A data with standardized atmospheric correction and quality filtering. Since vegetation vigor is directly influenced by soil moisture conditions, NDVI serves as a valuable proxy for monitoring soil moisture dynamics. Higher NDVI values indicate dense, healthy vegetation, which suggests adequate soil moisture availability, while lower NDVI values may indicate vegetation stress due to insufficient moisture. In this study, NDVI is utilized to assess soil moisture variability, helping to capture the spatial and temporal patterns of soil water availability.

4.3.1.5. Variable importance

The feature importance of predictor variables was evaluated using the Random Forest Regressor (RFR) algorithm. The importance of each feature was determined based on the mean decrease in impurity (MDI), which is computed as the total reduction in variance (or impurity) attributed to each feature across all decision trees within the ensemble model. Specifically, the scikit-learn implementation of Random Forest calculates feature importance by assessing how much each predictor variable contributes to improving the accuracy of the model's predictions.

For each soil property (sand, silt, and clay content), an independent Random Forest Regressor was trained using 200 estimators (decision trees) with a maximum depth of 10. After training, the feature importance scores were extracted using the .feature_importances_ attribute, which quantifies the relative contribution of each input variable to the predictive performance of the model. The features with higher scores had a greater influence on the model's predictions, whereas features with lower importance scores had a minimal impact.

4.3.1.6. Model Performance Evaluation

In machine learning, model evaluation is a crucial step to ensure the predictive reliability and generalizability of the model. In this study, the evaluation process was conducted at two levels. First, during the training phase, different hyperparameter configurations were tested using cross-validation, where the model was repeatedly trained on subsets of the data and validated on the remaining portions to optimize its predictive capability. Second, once the final model was trained using the best hyperparameter settings, it was assessed on an independent test set to evaluate its real-world performance. This two-tiered approach ensured both effective model training and a robust evaluation of its predictive accuracy (Hastie et al., 2009).

Since the Random Forest Regressor (RFR) is used for continuous variable prediction rather than classification, regression-specific evaluation metrics were employed to assess model performance. The key performance indicators used in this study include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2 Score). These metrics provide a quantitative measure of how well the predicted soil properties (Sand, Silt, and Clay fractions) align with actual observed values.

- Mean Squared Error (MSE): Measures the average squared difference between actual and predicted values. A lower MSE indicates better model performance.

$$\boldsymbol{MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{obsi} - y_{prei}\right)^2} \tag{6}$$

- Root Mean Squared Error (RMSE): The square root of MSE, providing a metric in the same unit as the target variable, making interpretation more intuitive.

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pre}\right)^2}} \tag{7}$$

- Mean Absolute Error (MAE): Computes the absolute average of prediction errors, representing the model’s average deviation from the true values.

$$\boldsymbol{MAE = (\sum\_(i = 1)\hat{}n}\text{▒}\boldsymbol{|y\_obsi -}\text{┤ ├}\boldsymbol{y\_prei}\text{┤}\boldsymbol{|\,)/n} \qquad \textbf{(8)}$$

- Coefficient of Determination (R^2 Score): Measures the proportion of variance in the dependent variable explained by the model. An R^2 value close to 1 indicates strong predictive performance, while lower values suggest greater unexplained variability.

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pre}\right)^2}{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{-obs}\right)^2}} \qquad \textbf{(9)}$$

By leveraging these evaluation metrics, this study provides a comprehensive assessment of the Random Forest model’s predictive capabilities, ensuring its reliability in soil particle size classification and environmental modeling. The inclusion of multiple regression performance indicators allows for a balanced evaluation of accuracy, consistency, and error distribution, ultimately strengthening the model’s applicability in geospatial and soil science research.

4.3.2. Methods

The flowchart (Figure 29) illustrates the methodological framework adopted in this study, organized into three main phases: data preparation, model construction, and data prediction. In the data preparation phase, field-based measurements, satellite observations, and ancillary environmental datasets were collected, pre-processed, and harmonized into a consistent spatial and temporal framework. Relevant features were extracted and compiled to create the input dataset. The model construction phase involved partitioning the dataset into training and testing subsets, followed by the implementation of three machine learning algorithms—Random Forest Regressor (RF), Extreme Gradient Boosting (XGBoost), and Gradient Boosting Regressor (GBR). Hyperparameter tuning and cross-validation were applied to optimize model performance and minimize overfitting. In the data prediction phase, the trained models were applied to generate spatially continuous predictions of soil properties, which were then validated against independent in situ measurements.

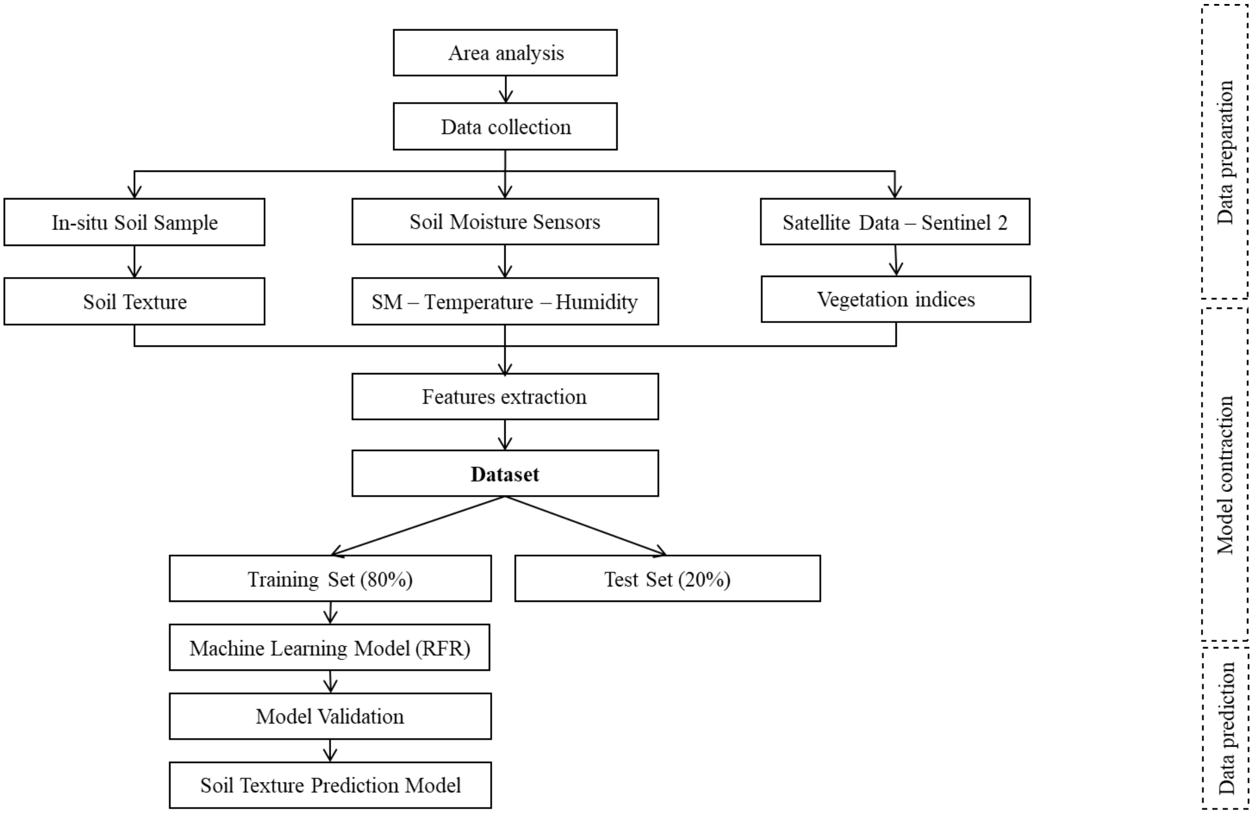


Figure 29. Soil texture prediction workflow

4.3.2.1. Retrieval of SF from the Measured Soil Moisture

In soil moisture monitoring using Sentek's EnviroSCAN, the Scaled Frequency (SF) serves as a normalized indicator of soil moisture content derived from sensor response characteristics. However, in cases where only the Volumetric Soil Water Content (SM) is available, it is necessary to retrieve the corresponding SF values through an inverse application of the default calibration equation.

The Sentek EnviroSCAN soil moisture sensors, manufactured by Sentek, use a standard default calibration equation to derive the Scaled Frequency (SF) from the Volumetric Soil Water Content (SM). The default calibration equation provided by Sentek for sands, loams, and clay loams follows the form:

$$SF = A * SM^{B} + C \tag{10}$$

where:

- SF = Scaled Frequency
- SM = Volumetric Soil Water Content (mm)
- A = 0.19570, B = 0.40400, C = 0.02852 (default calibration coefficients)

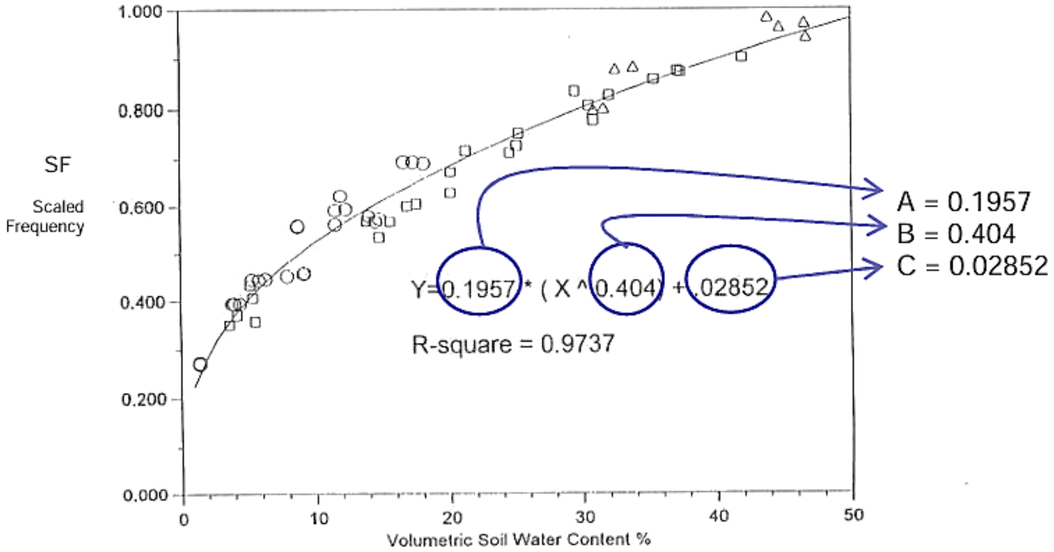


Figure 30. Empirical Relationship Between Scaled Frequency and Volumetric Soil Water Content (Sentek Pty Ltd., 2011)

Given that SF is typically obtained from sensor field measurements, the inverse approach applies when SF values must be retrieved solely from available SM data.

4.3.2.2. Remote sensing image selection and processing

4.3.2.2.1. Sentinel 2 Data: Spatial and Temporal Characteristics

Sentinel-2 satellite imagery was selected for this study due to its high spatial, spectral, and temporal resolution, which makes it a powerful tool for analyzing soil properties, vegetation dynamics, and environmental conditions. The Sentinel-2 mission, managed by the European Space Agency (ESA) under the Copernicus Program, provides multispectral imagery with optimized spectral bands for applications in vegetation monitoring, soil moisture estimation, and land cover classification.

One of the key advantages of Sentinel-2 is its multi-scale spatial resolution, which enhances the ability to analyze land surface features at different scales. The 10-meter resolution of the Visible (Red, Green, Blue) and Near-Infrared (NIR) bands allows for detailed vegetation and soil moisture analysis.

The temporal resolution of Sentinel-2 is another critical factor in its selection. The 5-day revisit time provided by the dual satellite system (Sentinel-2A and Sentinel-2B) ensures frequent image acquisitions, enabling the monitoring of seasonal changes in soil moisture, vegetation health (NDVI), and land surface dynamics. This high revisit frequency is essential for tracking soil-vegetation-atmosphere interactions and detecting short-term variations in hydrological and environmental processes.

4.3.2.2.2. Preprocessing of Sentinel-2 Data

To ensure the accuracy and reliability of the remote sensing analysis, Sentinel-2 images undergo multiple preprocessing steps to correct for atmospheric, radiometric, and geometric distortions. These corrections improve the comparability of images over time, reducing errors caused by sensor inconsistencies or atmospheric interference (SUHET, 2021).

Atmospheric Correction

Sentinel-2 Level-2A products are preferred for analysis as they provide surface reflectance values that have already undergone atmospheric correction using the Sen2Cor processor. This correction removes the effects of aerosol scattering, water vapor absorption, and atmospheric distortions, allowing for more accurate spectral analysis of soil and vegetation properties.

Geometric and Radiometric Corrections

Sentinel-2 data is georeferenced to the Universal Transverse Mercator (UTM) coordinate system, ensuring precise spatial alignment between different image acquisitions. Radiometric corrections are also applied to normalize reflectance values, compensating for sensor calibration differences and variations in illumination conditions. These preprocessing steps ensure that the extracted spectral data is consistent and suitable for temporal comparisons.

Cloud Masking

To improve the quality of the dataset, cloud-contaminated pixels are removed using the Scene Classification Map (SCL band) provided in Sentinel-2 products. This band helps distinguish cloudy, shadowed, and snow-covered areas, ensuring that only valid surface reflectance values are used in further analysis. Additional cloud removal techniques, such as the Fmask algorithm, can be employed to refine the cloud-masking process and minimize the impact of cloud-related noise in the dataset.

Band Selection and Index Computation

Specific spectral bands from Sentinel-2 are selected for further analysis, particularly for soil moisture estimation and vegetation assessment. The Near-Infrared (NIR) and Red bands are utilized to compute the Normalized Difference Vegetation Index (NDVI), which serves as an indicator of vegetation health, biomass productivity, and soil moisture availability. Higher NDVI values indicate denser vegetation cover and better soil moisture conditions, while lower values suggest drier soil and stressed vegetation.

4.3.2.2.3. Predictive Modeling

The predictive modeling process using the Random Forest Regressor (RFR) for soil composition estimation follows a structured approach, summarized as follows:

Data Acquisition and Preprocessing: The dataset contains a range of environmental and geospatial predictor variables along with soil texture components (sand, silt, and clay fractions). Initial preprocessing steps include handling missing values, standardizing variables, and ensuring consistency across all observations.

Data Splitting and Model Training: The preprocessed dataset is partitioned into training and test subsets to assess model generalizability. The Random Forest algorithm is implemented, where multiple decision trees are trained using randomly selected samples from the training set. The predictions from these trees are aggregated, reducing variance and improving predictive accuracy.

Model Evaluation and Optimization: The trained model is evaluated using Mean Squared Error (MSE), R^2 score, and feature importance analysis to assess prediction reliability. Hyperparameter tuning, including adjustments to tree depth, the number of estimators, and feature selection, is performed to optimize model performance.

The model was developed using Python 3.11.5, employing the scikit-learn library for training and evaluation. The dataset integrates environmental and soil moisture parameters, selected based on their relevance to soil texture classification. Feature importance analysis was conducted to identify the most influential predictors. The model selection module was used for data partitioning, cross-validation, and hyperparameter tuning, while the metrics module computed accuracy scores for performance evaluation. The ensemble module facilitated the implementation of the Random Forest Regressor, ensuring robust prediction capabilities.

4.3.2.2.4. RFR model application

The Random Forest Regressor (RFR) is widely recognized as one of the most effective ensembles learning methods for regression tasks. It constructs an ensemble of decision trees using bootstrap samples of the training data and aggregates their outputs to generate final predictions. This approach mitigates overfitting and enhances predictive accuracy by leveraging the diversity of individual trees (Breiman, 1996; Bakshi, 2020; Zhang et al., 2022).

Numerous studies have highlighted the RFR's robustness to noise and its strong generalization capability, which makes it particularly suitable for complex and heterogeneous datasets (Zhang et al., 2014). By averaging the outputs of multiple decision trees, RFR significantly reduces prediction variance, leading to improved model stability and reliability.

The general architecture of the RFR algorithm, illustrated in Figure 31, involves training a large number of decision trees and computing the average of their individual predictions.

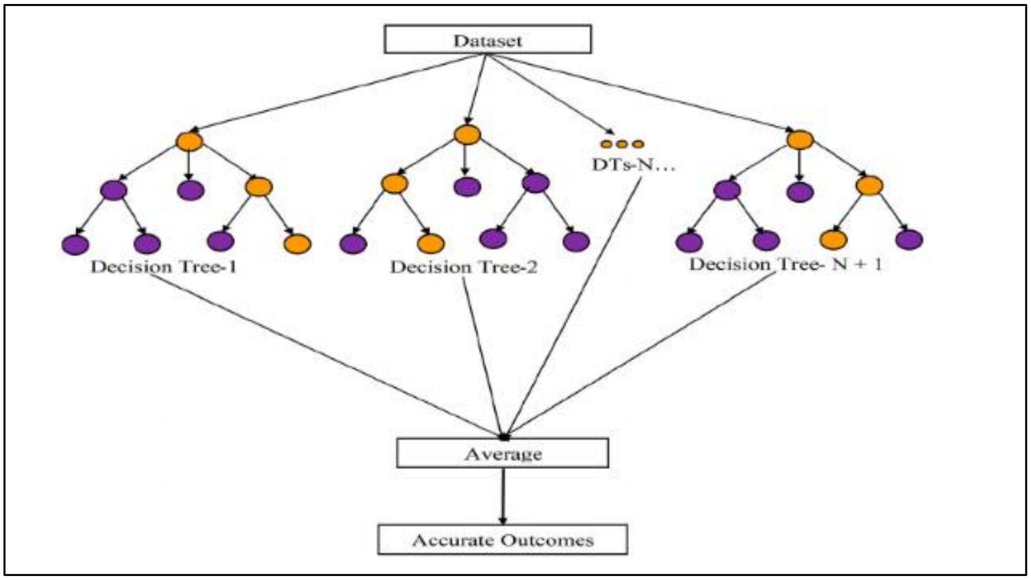


Figure 31. General diagram of a random forest regressor (Bakshi, 2020)

4.3.2.2.5. Preliminary Analysis of the Features and Targets

The dataset consists of multiple input features related to soil and environmental parameters, with the target variables being Sand, Silt, and Clay fractions. The preliminary analysis ensures that the dataset is structured, balanced, and representative of the variability in soil composition.

Model Training

The Random Forest model was employed as the predictive framework due to its robustness in handling nonlinear relationships and high-dimensional data. The training process follows these steps:

- Construction of Multiple Decision Trees: The model creates multiple decision trees, each trained on randomly selected subsets of the dataset.
- Learning Feature-Target Relationships: Each tree learns patterns between the input features and the target variables (Sand, Silt, Clay), capturing complex dependencies within the soil properties.
- Ensemble Prediction: The predictions from all trees are aggregated (averaged) to generate the final predicted values, reducing overfitting and improving generalization.

Dataset Splitting

To make sure the model was tested well, the dataset was split into training (80%) and testing (20%) sets. This follows a widely adopted convention in machine learning and data science. This splitting is common because it gives the model an adequate enough training set to learn patterns effectively, while retaining a statistically meaningful hold-out test set for fair performance evaluation. The ratio often depends on the size of the dataset, its variability, and how complex the learning task is. The 80/20 split is an optimal balance between training efficiency and validation robustness for datasets that are medium to large size (Joseph, 2022).

Hyperparameter Optimization

To enhance the predictive accuracy and generalizability of the Random Forest Regressor (RFR) model, a systematic hyperparameter optimization process was conducted using Randomized Search Cross-Validation (RandomizedSearchCV). This approach enables efficient exploration of the hyperparameter space by randomly sampling from a predefined range of values, thereby identifying optimal model configurations without the computational burden of exhaustive grid search. The optimization targeted key hyperparameters, including:

- Number of estimators (nestimatorsn_{estimators}nestimators): Specifies the number of trees in the ensemble, influencing the model's variance and stability.
- Maximum tree depth (maxdepthmax_{depth}maxdepth): Controls the depth of individual decision trees, preventing excessive complexity and mitigating overfitting.
- Minimum samples per split (minsamples_splitmin_{samples_split}minsamples_split): Defines the minimum number of samples required to split a node, affecting model flexibility.
- Minimum samples per leaf (minsamples_leafmin_{samples_leaf}minsamples_leaf): Determines the minimum number of observations required at a leaf node, balancing model complexity and generalization.
- Feature selection strategy (maxfeaturesmax_{features}maxfeatures): Regulates the number of predictor variables considered at each split, promoting diversity among decision trees.

A 5-fold cross-validation procedure was employed to ensure the robustness of the optimization process, systematically partitioning the dataset to evaluate model performance across multiple subsets. The optimal hyperparameters identified for each target variable were as follows:

- Sand: nestimators=100, minsamples_split=5, minsamples_leaf=1, maxfeatures=′sqrt′, maxdepth=Nonen
- Silt: nestimators=300, minsamples_split=2, minsamples_leaf=1, maxfeatures=′sqrt′,maxdepth=20n
- Clay: nestimators=300, minsamples_split=2, minsamples_leaf=1, maxfeatures=′sqrt′,maxdepth=20n

A systematic approach was employed to optimize the Random Forest Regressor (RFR) algorithm by selecting and fine-tuning key hyperparameters. Four critical hyperparameters were considered for optimization: the number of estimators (nestimatorsn_{estimators}nestimators), maximum tree depth (maxdepthmax_{depth}maxdepth), minimum samples per leaf (minsamples_leafmin_{samples_leaf}minsamples_leaf), and minimum samples required for a split (minsamples_splitmin_{samples_split}minsamples_split).

To explore an optimal configuration, a range of values was tested for each hyperparameter: five values for nestimatorsn_{estimators}nestimators, four values for maxdepthmax_{depth}maxdepth, three values for minsamples_leafmin_{samples_leaf}minsamples_leaf, and three values for minsamples_splitmin_{samples_split}minsamples_split, leading to a total of 180 unique hyperparameter combinations. The candidate values were selected based on default settings and prior studies, ensuring an efficient balance between model complexity and computational cost. The tested hyperparameter ranges are summarized in Table 3.

A Randomized Search Cross-Validation (RandomizedSearchCV) approach was implemented to efficiently explore the hyperparameter space while avoiding exhaustive grid search computations. Five-fold cross-validation was used to validate model performance across different subsets of the dataset, ensuring that the final hyperparameter selection was robust and generalizable. The optimized hyperparameters were then applied to train the final Random Forest models for soil particle size prediction (Sand, Silt, and Clay fractions), enhancing predictive accuracy and model stability.

Table 3. Hyperparameters of the grid search RF algorithm.

Hyperparameters	**Candidate values**
N estimators	100, 200, 300, 400, 500
MAX depth	5, 10, 20, None
MIN samples_leaf	1, 2, 4
MIN samples_split	2, 5, 10

By systematically tuning these hyperparameters, the Random Forest model usually shows significant improvements in predictive accuracy, reducing mean squared error and increasing R2 scores across all target variables.

Prediction & Model Output

Once the best-performing Random Forest model was selected through hyperparameter optimization, it was deployed to predict soil composition on the test dataset. The prediction process followed these steps:

- Input Feature Processing: For each test sample, the selected input features (such as soil moisture, temperature, and other environmental variables) were passed through the trained model.
- Independent Tree Predictions: Each individual decision tree within the Random Forest generated a separate prediction based on the relationships it learned during training.
- Ensemble Averaging: The final predicted soil composition (Sand, Silt, and Clay fractions) was computed as the average output across all decision trees, ensuring stability and accuracy in the results.

Model Selection and Cross-Validation

The best-performing Random Forest model was identified based on cross-validation scores, ensuring that the chosen configuration generalizes well across different subsets of the data. By optimizing hyperparameters, the model is expected to achieve a balance between bias and variance, improving accuracy in predicting the target variables while mitigating overfitting risks.

4.3.3. Results & Discussion

4.3.3.1. Variable Influence

The Pearson correlation coefficient was employed to evaluate the linear relationships among the predictor variables, with values ranging between -1 and 1. The correlation matrix, visualized in Figure 32, highlights the degree of association between different environmental and geospatial factors. The heatmap reveals strong positive correlations among temperature-related variables, such as Temperature and T_10_Days_Aver, with a correlation coefficient close to 0.96, indicating a strong linear dependence. Similarly, soil moisture measurements at different depths (SM_10, SM_20, and SM_30) exhibit high correlations, suggesting consistent moisture distribution patterns across layers.

Conversely, certain variables demonstrate negative correlations, such as soil moisture variability and sand content, reinforcing the expected inverse relationship between sand fraction and water retention capacity. Additionally, NDVI (vegetation index) exhibits a moderate positive correlation (0.4) with soil moisture vertical variation, indicating a potential relationship between vegetation cover and soil water fluctuation. The heatmap also highlights the interdependencies between multiple predictor variables, emphasizing the need for multi-variable analysis to accurately capture soil property variations. The observed correlation structure confirms that soil composition and environmental conditions are influenced by complex interactions among multiple factors, validating the selection of predictor variables for robust modeling.

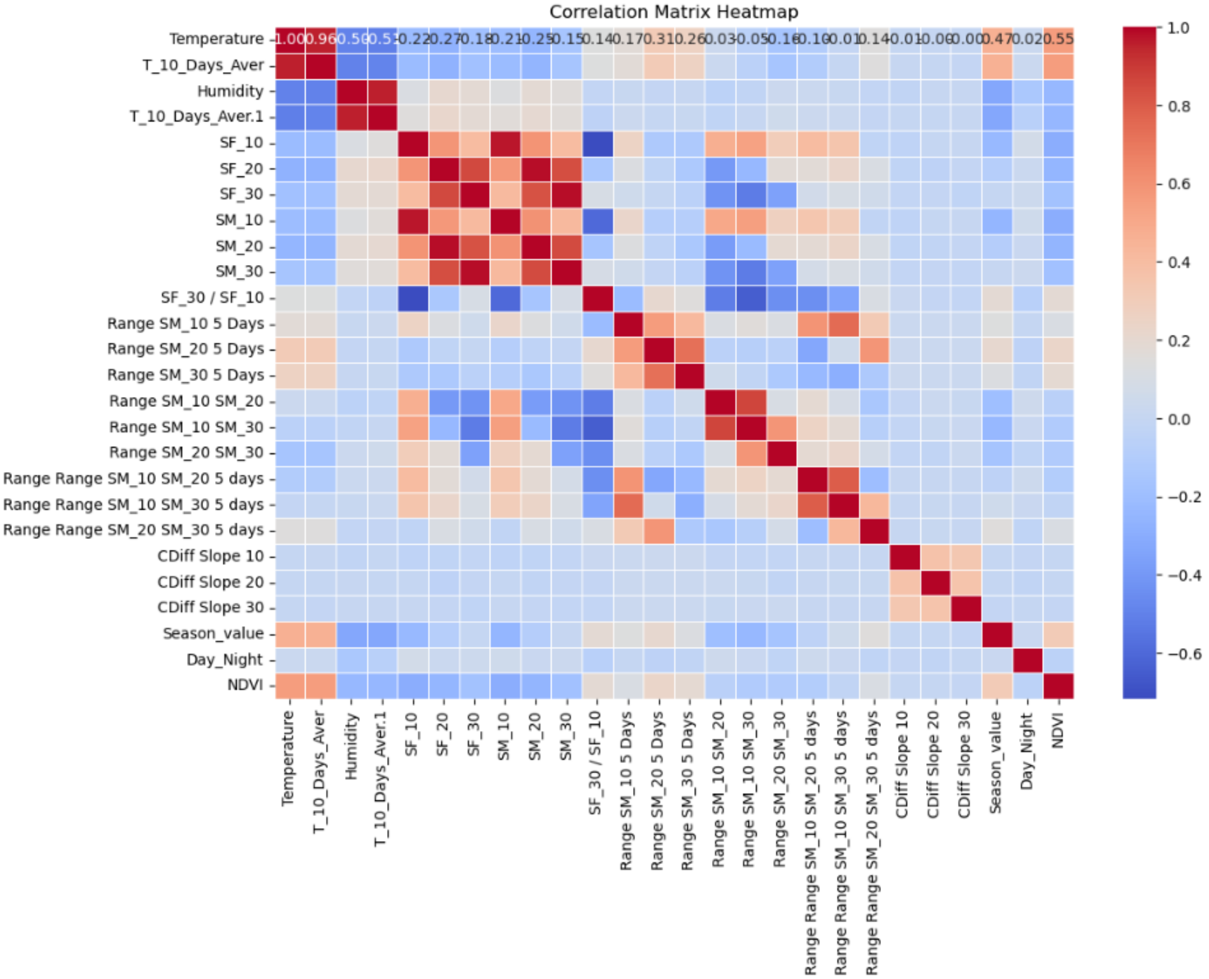


Figure 32. Correlation coefficient heatmap of the variables

4.3.3.2. Soil particle size class (Sand, Clay, Silt) Prediction Using Random Forest Regressor: Model Performance and Feature Importance Analysis

4.3.3.2.1. Model Performance Analysis

The predictive accuracy of the Random Forest model was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2 Score) across the three target variables: Sand, Silt, and Clay soil fractions. The results are summarized in table 4 below:

Table 4. Model Performance Metrics for Soil particle size class Prediction

Target Variable	MSE	RMSE	R² Score
Sand	14.97	3.87	0.92
Silt	4.85	2.2	0.93
Clay	11.8	3.44	0.85

The Random Forest model demonstrated a high predictive performance (Table 4), with R² values exceeding 0.85 for all soil types, indicating its strong explanatory power in modeling soil composition. The model effectively captured complex relationships between input features and target variables, demonstrating its reliability in soil particle size class prediction.

Among the three soil types, silt prediction achieved the highest accuracy (R² = 0.93) and the lowest error values (MSE = 4.85, RMSE = 2.20). These results suggest that silt is more predictable compared to sand or clay classes, likely due to its relatively uniform moisture distribution and intermediate particle size, which contribute to more consistent model behavior.

Sand particle size class prediction exhibited moderate accuracy (R² = 0.92), with a slightly higher error margin (MSE = 14.97, RMSE = 3.87). The higher variability in sand soil class predictions may be attributed to its coarse texture and rapid moisture drainage, making it more sensitive to environmental fluctuations such as temperature changes, evapotranspiration, and infiltration dynamics.

In contrast, clay prediction showed the lowest accuracy (R² = 0.85) and a high error (MSE = 11.80, RMSE = 3.44), indicating greater modeling challenges for clay content estimation. The higher degree of uncertainty in clay predictions may stem from its complex moisture retention properties, low permeability, and high susceptibility to compaction and shrink-swell behavior. These factors introduce nonlinear relationships that are more difficult for the model to capture with precision.

Overall, the findings indicate that silt is the most predictable, while clay content presents the greatest challenges for accurate modeling. The observed performance discrepancies highlight the importance of soil texture-specific modeling approaches and the need for further refinement in feature selection and data representation to improve predictions for clay-rich soils.

4.3.3.2.2. Model Learning Curve Analysis

The learning curve (Figure 33) illustrates the performance of the Random Forest model in predicting soil fractions, specifically clay, sand, and silt contents, based on environmental and sensor-derived input features. The green line represents the model's performance (R^2 score) on the training data, while the red line shows the cross-validation (CV) performance on unseen data. The shaded red area denotes the standard deviation of CV scores across folds, providing insight into model stability.

Initially, the training score is high (~0.91), suggesting that the model quickly fits the training data even with small sample sizes. As the number of training examples increases, the training R^2 slightly improves and plateaus around 0.96, indicating a strong fit to the known data and low bias. However, this high performance on training data may also reflect some overfitting, particularly at lower sample sizes.

The cross-validation curve starts low (~0.25) but increases steadily as more training data are added, reaching ~0.76 at 400 examples. This trend indicates that the model benefits significantly from additional data, improving its generalization capability across all three soil texture components. The gap between the training and CV curves is most prominent at lower sample sizes, which is typical of overfitting scenarios where the model learns training patterns well but fails to generalize. As the training size grows, the gap narrows, implying that model variance is reduced and that the predictions for sand, silt, and clay content become more stable.

The red shaded area around the CV curve is wide at small sample sizes, indicating high variability across validation folds, likely due to inconsistent patterns in small subsets. This variability decreases as the dataset grows, suggesting improved reliability and robustness in the model's predictions.

Overall, this learning curve confirms that Random Forest is an effective approach for predicting soil fractions. The model achieves strong predictive performance with increasing data availability, though some overfitting is present at lower sample sizes.

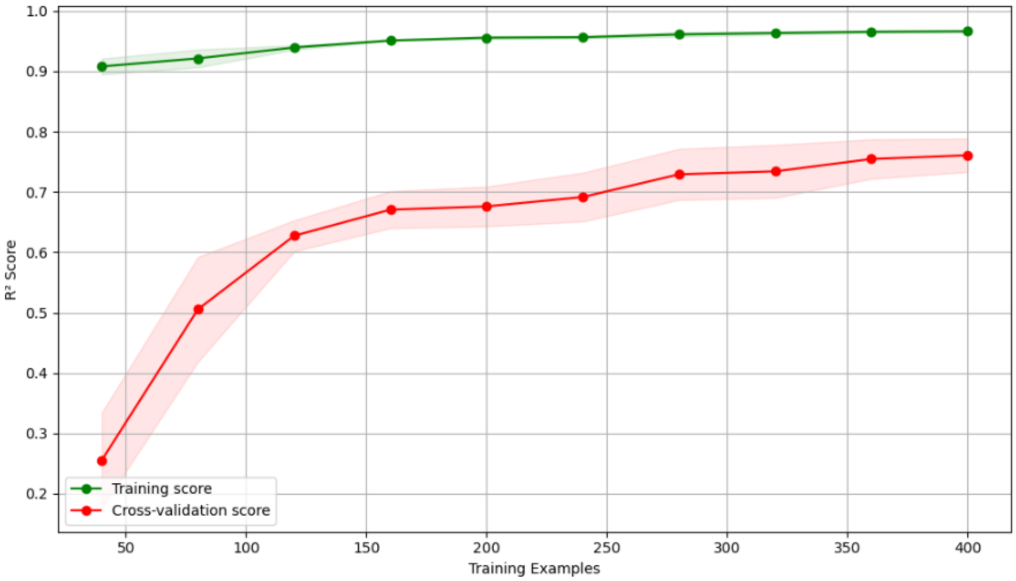
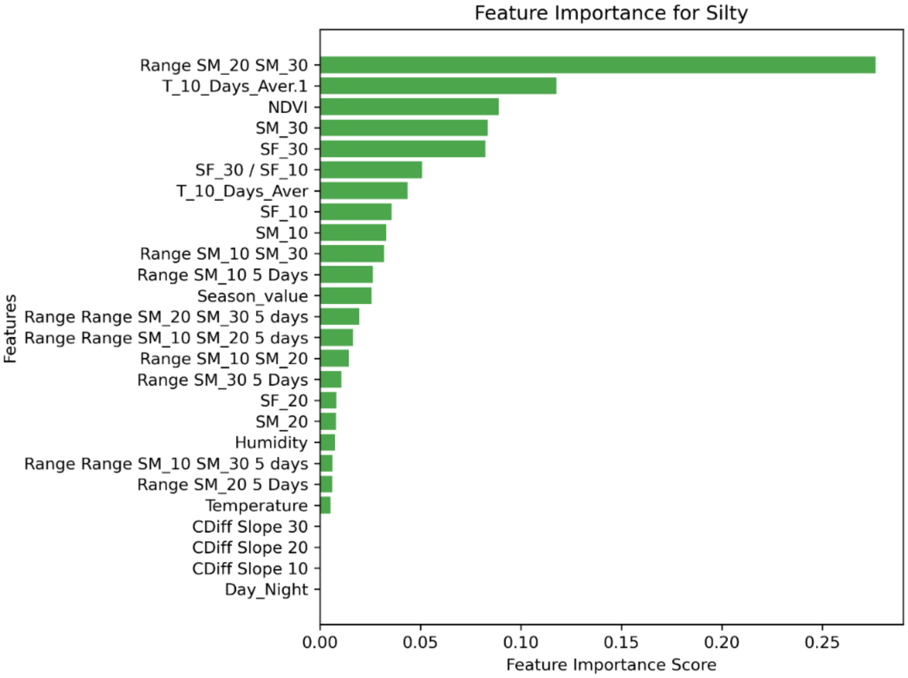


Figure 33. Learning Curve – Random Forest

4.3.3.2.3. Feature Importance Relates to Soil particle size class (Sand, Silt, Clay)



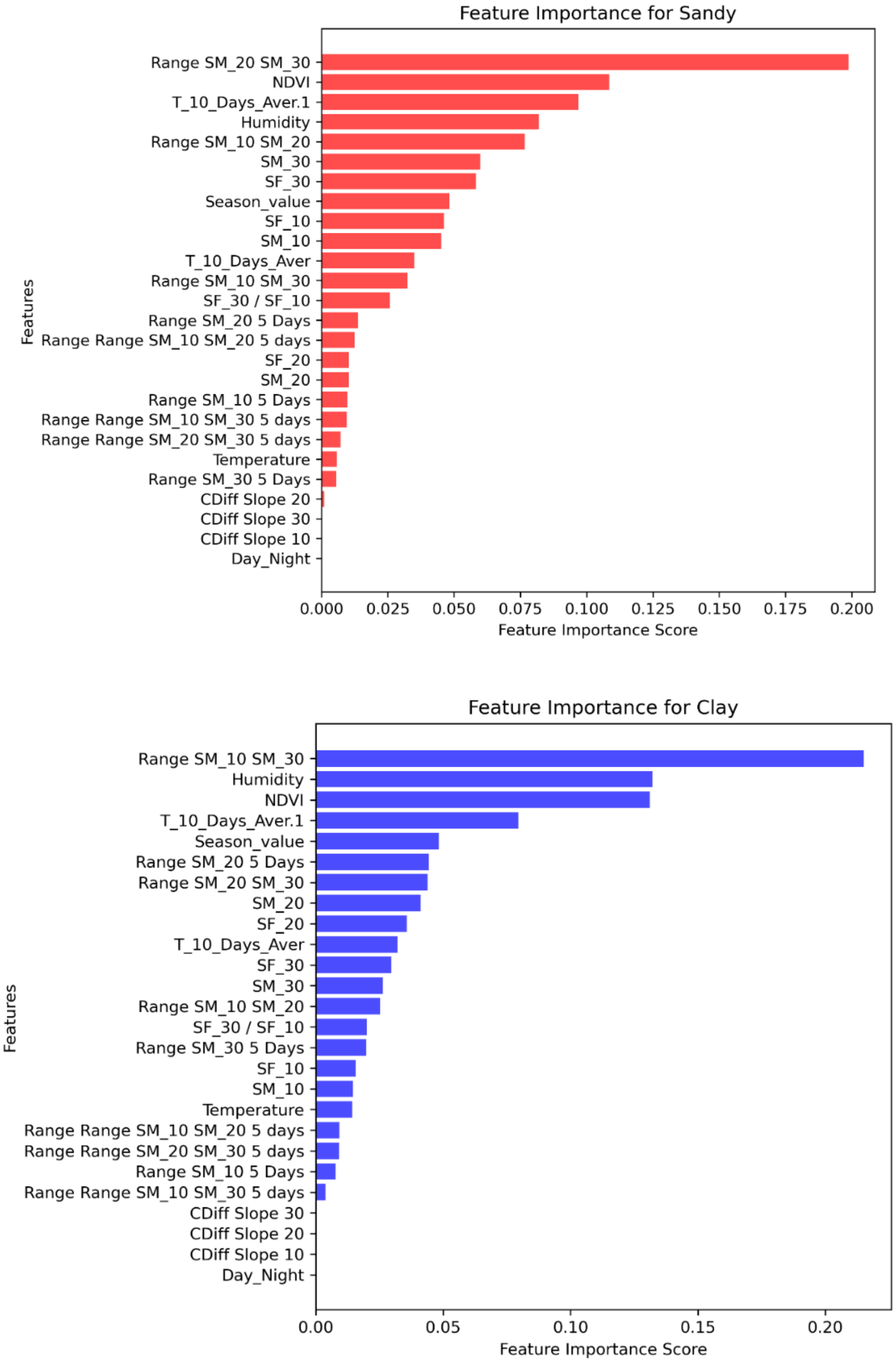


Figure 34. Feature Importance Ranking for Silt, Clay and Sand classes

The feature importance analysis (Figure 34) highlights the key environmental variables influencing the Random Forest model's ability to predict soil particle size classes, namely sand, silt, and clay.

The importance scores quantify each variable's relative contribution to the model's predictive performance.

For sand fraction prediction, the most influential variable is Range SM_20_SM_30 (importance score: 0.1989), underscoring the role of soil moisture variability between 20 cm and 30 cm depth in characterizing sandy soils. Additional important predictors include NDVI (0.1086), reflecting vegetation health, and T_10_Days_Aver.1 (0.0969), representing the average temperature over the preceding 10 days.

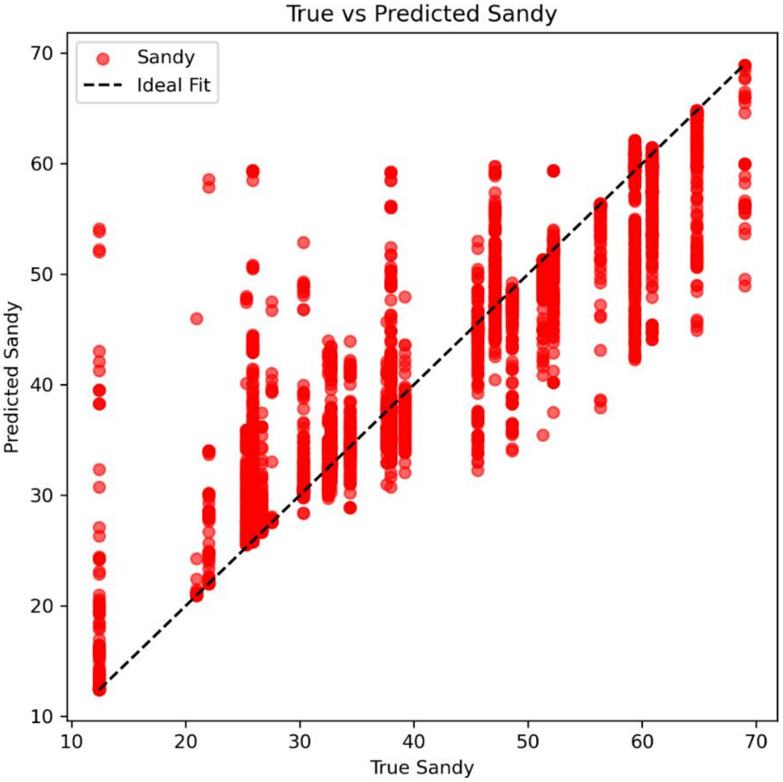
For silt fraction, Range SM_20_SM_30 again emerges as the strongest predictor (0.2765), indicating that subsurface moisture fluctuations are critical in identifying silty textures. T_10_Days_Aver.1 (0.1178) and SF_30 (0.0859), a surface factor at 30 cm depth, also contribute significantly.

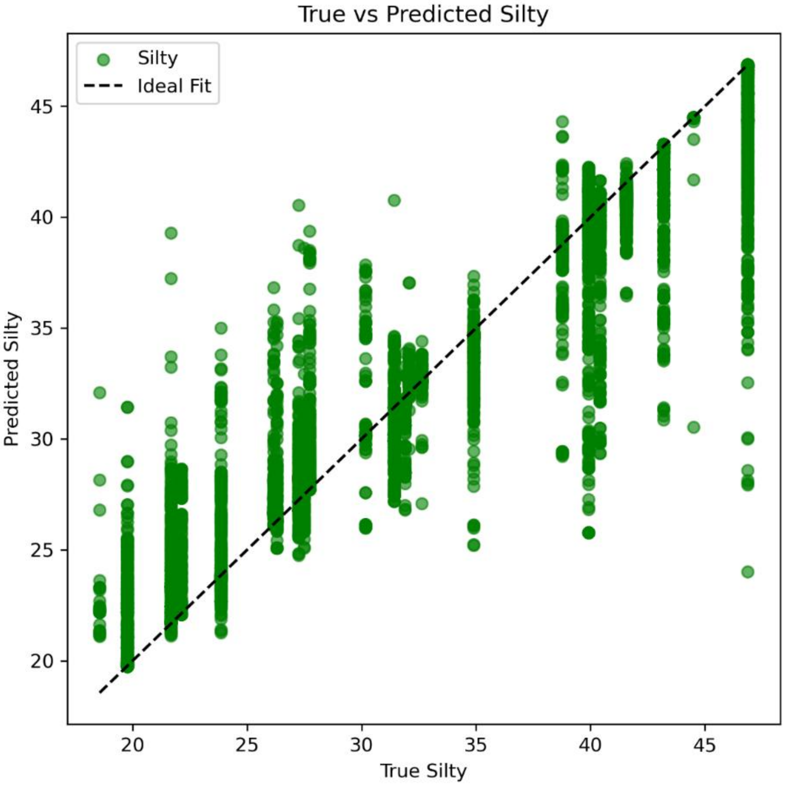
In the case of clay prediction, the leading predictor is Range SM_10_SM_30 (0.2152), which captures the vertical gradient in moisture from shallow to deeper soil layers, an important factor in clay-rich profiles. Humidity (0.1322) and NDVI (0.1312) further enhance the model's accuracy for this class.

These findings reinforce the rationale for modeling soil texture classes, as texture governs the soil's water retention and supply capacity, which in turn interacts with temperature to influence plant physiological response. Temperature modulates evapotranspiration rates and thus the demand for water, while soil texture determines how effectively moisture is retained or transmitted to the root zone. The use of moisture range between depths (e.g., 10–30 cm or 20–30 cm) as a predictor is especially insightful. These values indicate the vertical moisture gradient, which reflects both water transport capacity and differential wetting/drying dynamics. For instance, sandy soils exhibit rapid moisture changes due to high permeability, whereas clay soils display slower moisture dynamics and smaller fluctuations.

This pattern aligns well with physical expectations: coarse-textured soils (sand) offer limited water buffering, while fine-textured soils (clay) can better sustain plant water needs under fluctuating temperature regimes. The strong predictive role of temperature, while not directly altering soil moisture, serves as a proxy for the atmospheric demand placed on the soil–plant system, creating a three-dimensional interaction among soil, plant, and atmosphere.

4.3.3.2.4. Evaluation of Model Predictions: True vs. Predicted Soil Composition





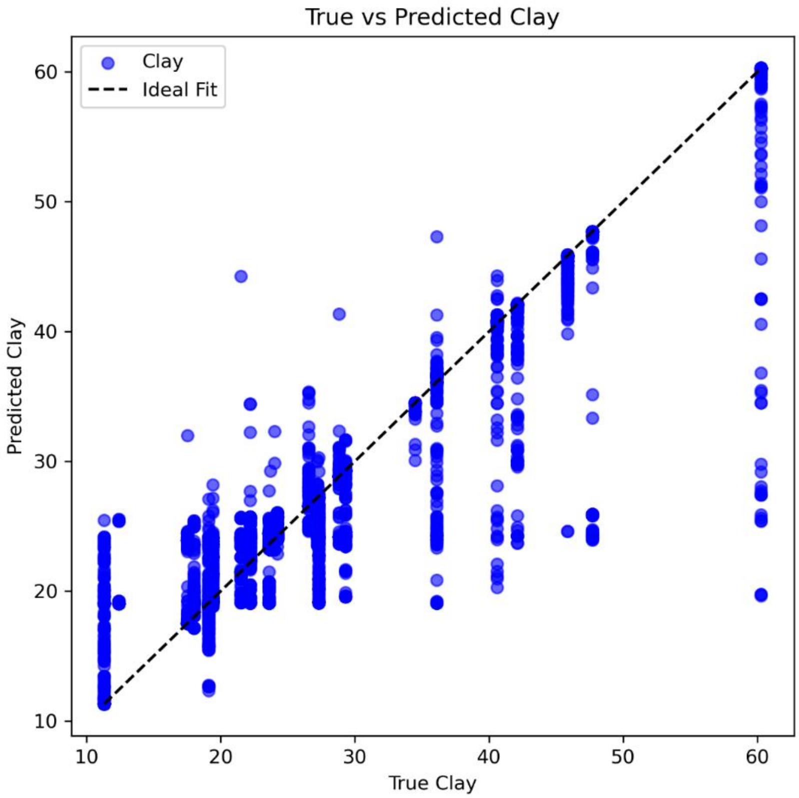


Figure 35. True vs Predicted Values of Sand, Silt, and Clay class

The scatter plots (Figure 35) illustrate the relationship between the true and predicted values for sand, silt, and clay soil compositions, providing an assessment of the model's predictive accuracy and consistency. The dashed black line represents the ideal fit line (1:1 ratio), where a perfect prediction would align all data points along with this diagonal.

The sand predictions show a strong correlation with the true values, with data points generally aligning along the ideal fit line. However, there is noticeable scatter at lower sand values, suggesting a degree of variability in the model's ability to predict lower sand content accurately. Despite this, the model maintains a consistent trend with limited extreme deviations, indicating moderate prediction reliability.

Among all soil particle size classes, the silt model exhibits the highest predictive accuracy, as evidenced by the tight clustering of points around the ideal fit line. The spread of predicted values is relatively uniform, with fewer outliers compared to sand and clay classes predictions. This result aligns with the high R² score (0.93) obtained during model evaluation, indicating that silt is the most predictable using the selected features.

The clay predictions demonstrate the highest level of dispersion, particularly for higher clay content values. While the model captures general trends, there is a wider deviation from the ideal fit line, suggesting greater difficulty in accurately predicting clay composition. This aligns with the lower R² score (0.85), indicating that clay variability presents greater modeling challenges, likely due to complex moisture retention properties and heterogeneous distribution in the dataset.

The model performs best in predicting silt, followed by sand, with clay class predictions exhibiting the highest uncertainty. The general alignment of points with the ideal fit line, however, confirms that the Random Forest model effectively captures soil composition trends, making it a reliable tool for soil texture prediction.

4.3.3.2.5. Hyperparameter Optimization Results

The optimized models were subsequently evaluated using Mean Squared Error (MSE) and the coefficient of determination (R2), key statistical metrics for assessing regression model performance. The results demonstrated high predictive accuracy, with R² values close to 1, indicating a strong correlation between predicted and observed values:

Table 5. Model Performance Metrics

Soil particle size class	**MSE**	**RMSE**	**R² Score**
Sand	5.13	2.27	0.97
Silt	1.01	1	0.99
Clay	4.88	2.21	0.94

These findings underscore the effectiveness of hyperparameter tuning in optimizing model performance, minimizing prediction error, and ensuring robust generalization to data. The results confirm that Random Forest, when systematically optimized, provides a highly reliable and scalable approach for soil composition prediction. Furthermore, the implementation of cross-validation in hyperparameter selection reinforces the stability and reproducibility of the model, ensuring that the identified parameter configurations are not biased toward a specific data subset.

By leveraging machine learning-based hyperparameter optimization, this study establishes a robust predictive framework, demonstrating that an ensemble learning approach, when properly tuned, significantly enhances soil property estimation. The findings highlight the potential of Random Forest regression in geospatial and environmental modeling, facilitating data-driven decision-making in soil texture classification and land resource management.

4.3.4. Conclusions

This study presents a robust machine learning framework for predicting soil texture composition, by prediction the sand, silt, and clay fractions, by integrating in-situ soil moisture data with Sentinel-2 derived vegetation indices and environmental variables. Among the tested models, Random Forest Regression (RFR) demonstrated a good performance, achieving coefficient of determination (R^2) values exceeding 0.90 for all soil particle size classes and peaking at 0.99 for silt class, underscoring the model's predictive accuracy and generalization capability.

The feature importance analysis revealed that short-term soil moisture dynamics at multiple depths, in combination with Range of the SM, NDVI and humidity, are key determinants of soil texture variability. These results validate the hypothesis that soil water availability and vegetative response are reliable proxies for underlying textural properties, offering a novel perspective on indirect soil characterization.

The scientific contribution of this work lies in its demonstration of a non-invasive, data-driven methodology for high-resolution soil texture prediction, with significant implications for precision agriculture, hydrological modeling, and sustainable soil management. By leveraging widely accessible remote sensing data and in-situ sensors, this approach enables scalable and cost-effective monitoring of spatial soil variability, particularly valuable in regions where traditional soil surveys are limited or impractical.

Future work will aim to enhance model transferability across diverse agroecological zones and to integrate this predictive framework into decision support systems for climate-smart land and water resource management.

Based on the conclusions above the following theses were formulated:

Thesis 3

The integration of multi-depth in-situ soil moisture dynamics from capacitance-based Sentek EnviroSCAN sensors with Sentinel-2-derived vegetation indices and environmental parameters enables the indirect, non-invasive prediction of soil texture composition. By leveraging Random Forest regression, this framework captures the complex interactions between soil moisture behavior, vegetative response, and particle size distribution, providing a scalable alternative to traditional laboratory-based soil texture analysis. Furthermore, the potential for site-specific model transferability remains an important consideration, as variations in parent material, climatic conditions, and land use may influence model performance. Addressing these factors in future research will enhance the robustness and applicability of this methodology for digital soil characterization across diverse environments.

Thesis 4

Vertical gradients and short-term variability in soil moisture, quantified through sensor-derived scaled frequency metrics and multi-depth moisture fluctuation indices, exhibit strong predictive relationships with USDA-defined soil particle size classes. The soil moisture fluctuation indices were derived from Sentek EnviroSCAN probe data, including maximum, minimum, and range values over 5-day moving windows at 10, 20, and 30 cm depths, as well as inter-depth gradients (e.g., Range SM_10–30). When combined with NDVI and environmental features within a machine learning framework, these spatiotemporal dynamics allow for accurate, high-resolution mapping of soil texture, demonstrating the viability of sensor- and satellite-based approaches for digital soil characterization in heterogeneous environments.

5. Clay Ratio prediction using Machine learning, in situ and remote sensing data

5.1. Introduction

Soil texture, particularly the clay fraction, plays a pivotal role in controlling key soil functions such as water retention, nutrient dynamics, cation exchange capacity, and root development (Daniel Hillel, 2004b). Accurate estimation of clay content is essential for informed decision-making in agriculture, hydrology, and land use planning. However, traditional methods for assessing clay ratio, such as hydrometer or pipette analyses, are often limited in spatial coverage and labor-intensive, restricting their utility for regional-scale applications.

To address these limitations, recent studies have increasingly explored the integration of in-situ sensor data, remote sensing indicators, and machine learning algorithms. Remote sensing offers spatially continuous and temporally frequent observations, such as vegetation indices (e.g., NDVI) and surface temperature, which are indirectly influenced by soil properties (Peng et al., 2017). When combined with in situ soil moisture measurements and environmental variables, these data streams enable the development of predictive models capable of estimating soil texture components with high spatial resolution.

Machine learning models, particularly ensemble methods like Random Forest, have demonstrated strong capabilities in predicting soil properties by capturing complex, non-linear interactions among environmental variables. Recent studies, such as Wang & Gao, 2023, have shown that combining Radar and optical remote sensing data with ensemble learning techniques can significantly enhance soil moisture retrieval under vegetated conditions. Building on this foundation, this section presents a Random Forest-based framework for predicting the clay ratio using a dataset that integrates remote sensing indicators, meteorological variables, and in-situ soil moisture measurements at multiple depths. Model performance is evaluated using standard regression metrics and feature importance analysis, illustrating the potential of this approach for digital soil mapping and decision support in precision agriculture.

5.2. Materials and Methods

For this chapter, the study area, datasets, and preprocessing workflow are identical to those described in the previous chapter. The key distinction lies in the prediction target: while the earlier chapter focused on soil texture classes, the present analysis specifically aims to predict the Clay Ratio. Furthermore, this chapter evaluates and compares the performance of three machine learning algorithms, Random Forest Regressor (RF), Extreme Gradient Boosting (XGBoost), and Gradient Boosting Regressor (GBR).

The workflow illustrated in Figure 36 outlines the sequential steps of the study, which are structured into three main phases: data preparation, model construction, and data prediction. In the data preparation phase, field measurements, remote sensing products, and ancillary datasets are integrated, processed, and harmonized. The model construction phase involves splitting the dataset into training and testing subsets, followed by the implementation of machine learning algorithms. Finally, in the data prediction phase, the trained models are applied to estimate the target variable, Clay Ratio, with subsequent accuracy assessment and validation against independent observations.

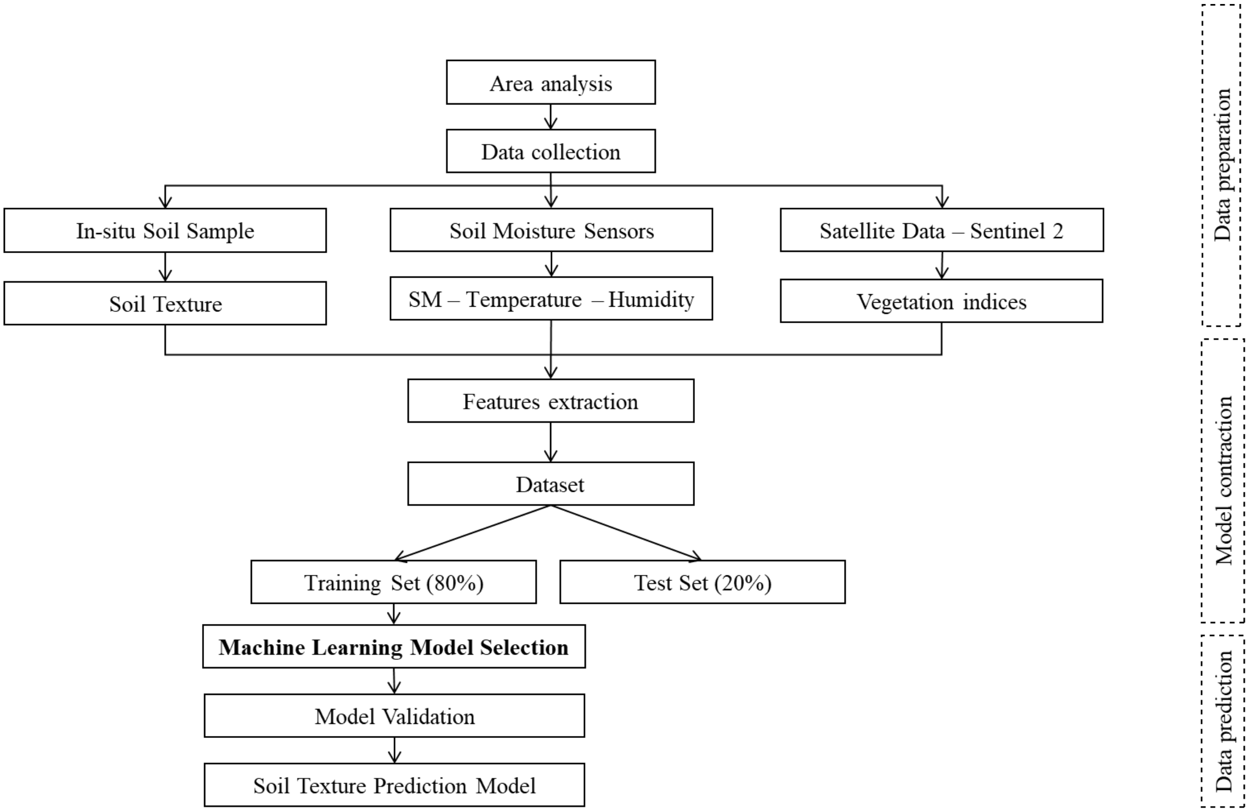


Figure 36. Clay Ration prediction workflow

The Clay Ratio was computed as the proportion of clay content to the combined content of sand and silt, following the equation:

$$\text{Clay Ratio} = \frac{\text{Clay (\%)}}{\text{Sand (\%)} + \text{Silt (\%)}} \qquad \mathbf{(11)}$$

This derived variable serves as a normalized indicator to assess the relative abundance of fine particles within the soil matrix. All predictive features, comprising in situ soil moisture measurements at multiple depths, meteorological data, and remote sensing indices, were retained to ensure methodological consistency. The Random Forest regression model was again employed due to its robustness in handling non-linear relationships and heterogeneous data sources. This consistent approach facilitates comparative analysis across different soil texture attributes within the same modeling framework.

Random Forest Regressor (RFR)

Random Forest Regressor (RFR) was employed to predict the clay ratio. RFR is an ensemble learning algorithm that operates by constructing a multitude of decision trees during training and outputs the mean prediction of the individual trees. This method reduces overfitting and increases prediction robustness by combining multiple learners trained on bootstrapped datasets and feature subsets (Breiman, 2001b).

The RFR model was implemented using the Random Forest Regressor module from the sklearn.ensemble library. A grid search was conducted to tune hyperparameters, and the final model was trained using n_estimators=200 and max_depth=10, which offered a balanced trade-off between bias and variance. The input features included a set of 26 predictors derived from in situ soil moisture, weather data, and remote sensing indices such as NDVI.

The dataset was split into training (80%) and test (20%) sets using random sampling. Model performance was assessed using standard evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2).

Gradient Boosting Regressor (GBR)

Gradient Boosting Regressor (GBR) was employed as a machine learning model to predict the clay ratio in the soil profile using the selected feature set. GBR is an ensemble learning technique that builds a predictive model by sequentially adding decision trees, where each new tree attempts to correct the residual errors made by the previous ones. It combines weak learners into a strong learner by minimizing a specified loss function using gradient descent (Friedman, 2001).

In this study, the model was implemented using the GradientBoostingRegressor module from the sklearn.ensemble library. The selected hyperparameters included n_estimators=200, max_depth=6, and a learning_rate=0.1, based on empirical tuning. The model was trained on 80% of the dataset and validated on the remaining 20% using random sampling.

Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Feature importance scores were also derived to assess the relative contribution of each predictor variable.

Extreme Gradient Boosting (XGBoost)

XGBoost Regressor (XGB) was also employed to predict the clay ratio, providing a high-performance alternative to traditional boosting methods. XGBoost is an optimized, distributed gradient boosting library designed to maximize computational efficiency while offering flexibility for complex datasets. It incorporates advanced regularization techniques to mitigate overfitting and can handle missing data internally, making it suitable for environmental modeling applications (T. Chen & Guestrin, 2016).

The model was implemented using the XGBRegressor from the xgboost Python library with key parameters set as n_estimators=200, max_depth=6, and learning_rate=0.1. Similar to GBR, the training dataset represented 80% of the total samples, and the remaining 20% was used for model evaluation.

Data Description

Figure 37 presents the set of input variables used in the machine learning framework together with the target feature, Clay Ratio.

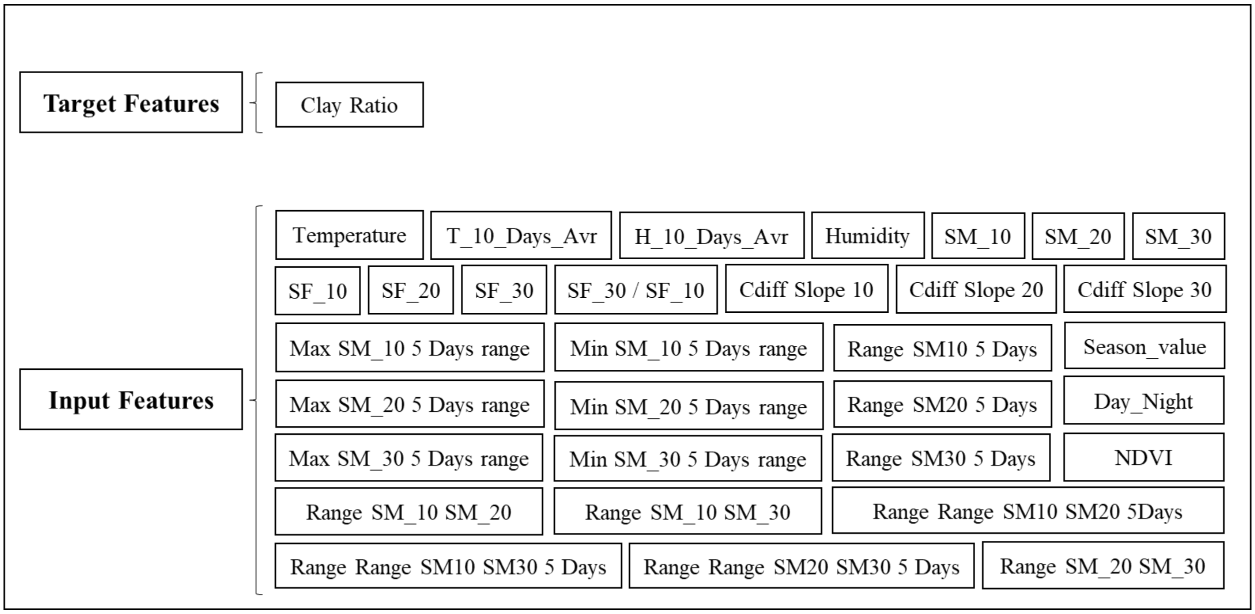


Figure 37. Input Features and Target Variable for the Clay Ratio Prediction

5.3. Results and Discussion

5.3.1. Model Performance Analysis

The comparative evaluation of the three ensemble models for Clay Ratio prediction, Random Forest, Gradient Boosting Regressor (GBR), and XGBoost Regressor (XGB), is presented in the table 6 below, based on three standard regression performance metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2 Score).

Table 6. Comparison of Model Performance Metrics for Clay Ratio Prediction

Target Variable	MSE (Mean Squared Error)	RMSE (Root Mean Squared Error)	R^2 Score
Random Forest	0.003	0.054	0.936
Gradient Boosting	0.0015	0.038	0.9747
XGBoost	**0.0015**	**0.038**	**0.9748**

The Random Forest model achieved an R^2 of 0.936, indicating that it explains 93.6% of the variance in clay ratio. However, compared to the other two boosting-based models, it yielded a higher error (MSE = 0.003, RMSE = 0.054), which suggests lower precision in capturing more complex relationships in the data.

In contrast, Gradient Boosting and XGBoost outperformed Random Forest on all metrics, both achieving an R^2 of approximately 0.975, which signifies that they explain over 97% of the variance in the clay ratio. The RMSE of 0.038 in both models reflects a substantial improvement in predictive accuracy over the Random Forest.

The slightly better R^2 score for XGBoost (0.9748) compared to Gradient Boosting (0.9747) highlights the marginal edge of XGBoost, which can be attributed to its regularization techniques and scalability. XGBoost also tends to be more robust to overfitting, particularly in high-dimensional datasets, due to its built-in shrinkage and column subsampling strategies.

These results demonstrate that boosting models, particularly XGBoost, are more effective than bagging methods like Random Forest in predicting soil clay content from integrated datasets. The enhanced performance is likely due to the ability of boosting algorithms to sequentially correct prediction errors and focus on difficult-to-predict observations.

5.3.2. Learning Curve Analysis of the Predictive Models

To assess the generalization performance and robustness of the models trained for Clay Ratio prediction, learning curves were constructed for the three ensemble regressors: Random Forest (RF), Gradient Boosting Regressor (GBR) and XGBoost Regressor (XGB). These curves plot the training and cross-validation R^2 scores as a function of increasing training set size, providing insights into potential underfitting, overfitting, and data sufficiency.

As shown in Figures 37, 38, and 39, all three models exhibit stable and consistently high training R^2 scores (0.91–0.93), indicating a strong fitting capacity. However, noticeable disparities are observed in the cross-validation scores, particularly at smaller training set sizes, where all models show significant performance variance, reflecting high sensitivity to data partitioning.

Among the models, XGBoost demonstrated the best generalization capability, achieving a smoother increase in cross-validation R^2 scores with a final value approaching 0.97. The reduced gap between training and validation scores at larger sample sizes indicates superior bias-variance tradeoff. Gradient Boosting followed closely, showing similar convergence behavior but with slightly lower validation stability, suggesting marginally higher variance.

Random Forest, while achieving high training scores, lagged behind in validation performance, especially with smaller training sets, showing greater divergence between training and cross-validation scores. This suggests a higher tendency toward overfitting in data-scarce regimes.

The shaded confidence intervals for all models decrease as training size increases, confirming that more data improves the reliability and stability of predictions. Despite fluctuations at lower data volumes, all models ultimately converge toward strong R² values with sufficient data, highlighting the viability of ensemble learning approaches for clay fraction estimation.

In conclusion, the learning curve analysis confirms that while all three ensemble regressors are effective, XGBoost offers the most balanced performance, combining high accuracy with generalization stability, making it particularly well-suited for the task of clay ratio prediction using integrated environmental features.

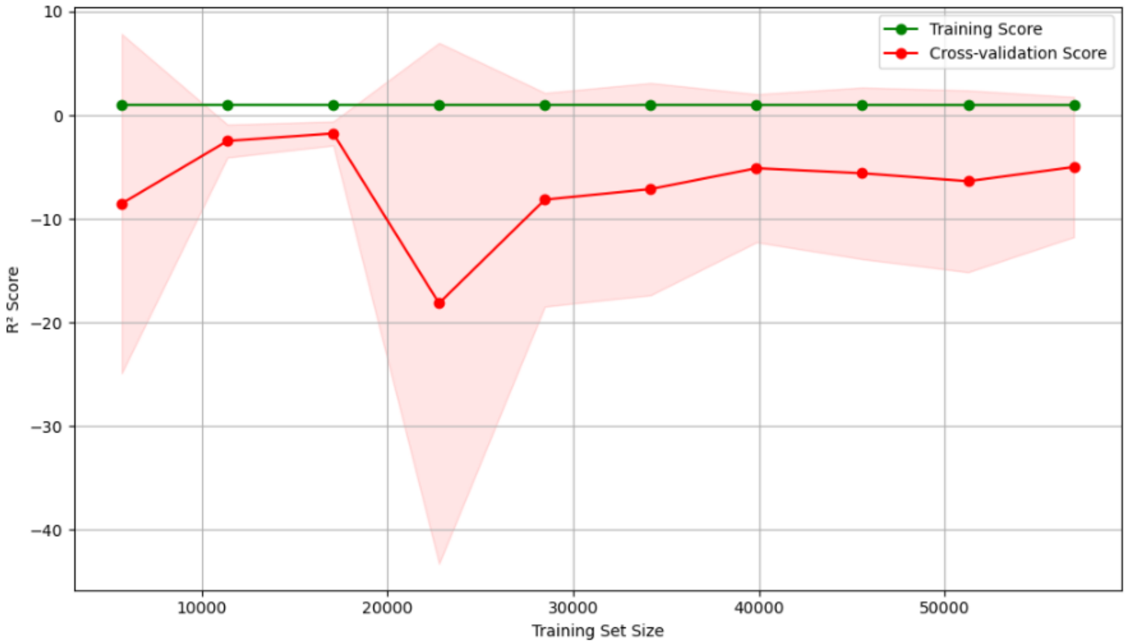


Figure 38. Learning Curve – Clay Ratio Prediction using Gradient Boosting Regressor

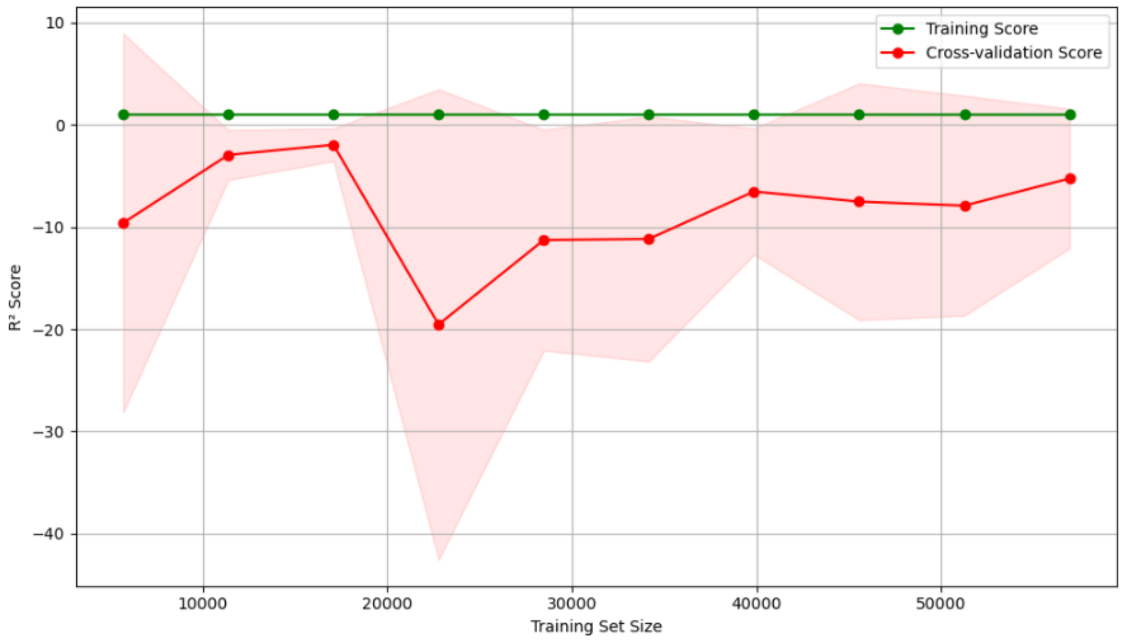


Figure 39. Learning Curve – Clay Ratio Prediction using XGBoost Regressor

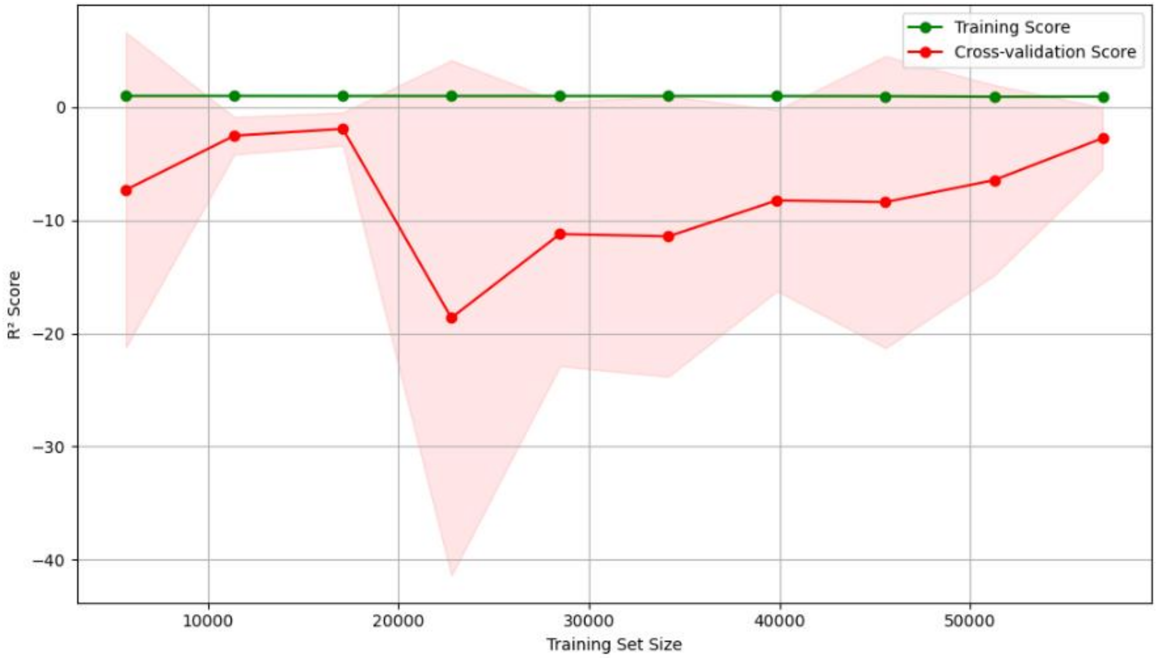


Figure 40. Learning Curve – Clay Ratio Prediction using Random Forest Regressor

5.3.3. Feature Importance Relates to Clay Ratio Prediction

5.3.3.1. Feature Importance for Predicting Clay Ratio of Random Forest Regressor

The Random Forest model highlights Range SM_10 SM_30 as the most influential feature (Figure 40), indicating that vertical moisture gradients between shallow and deeper soil layers are critical for predicting clay content. Vegetation index (NDVI), Humidity, and 10-day averaged temperature metrics (T_10_Days_Aver, T_10_Days_Aver.1) also show high importance, reflecting their role in controlling soil moisture dynamics and vegetation stress responses associated with clay-rich soils. Features related to frequency-scaled moisture (e.g., SF_30 / SF_10) and seasonal effects (Season_value) follow closely.

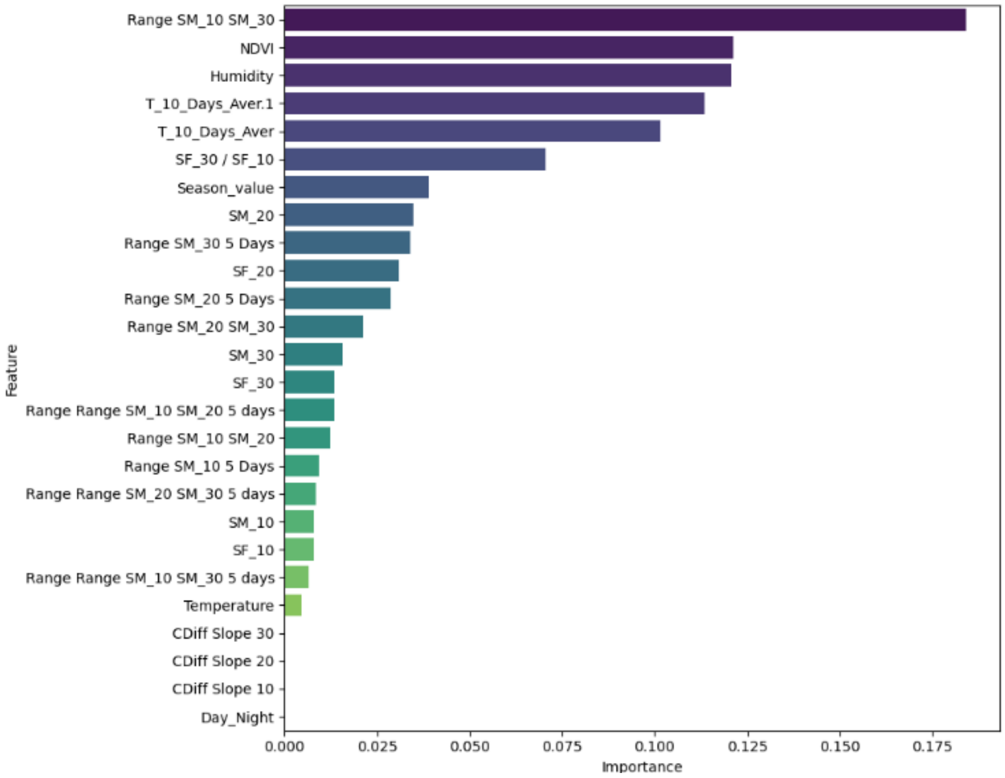


Figure 41. Feature Importance for Predicting Clay Ratio - Random Forest Regressor

5.3.3.2. Feature Importance for Predicting Clay Ratio of Gradient Boosting Regressor

Gradient Boosting confirms the dominance of Range SM_10 SM_30 as the top feature, followed by NDVI, Humidity, and T_10_Days_Aver.1, consistent with Random Forest. However, this model attributes relatively higher importance to frequency-scaled features such as SF_30 / SF_10 and SF_20, as well as moisture ranges at various depths. The distribution of importance is more gradual, showing the model's sensitivity to a broader set of hydrometeorological variables. Low influence remains with slope change metrics (CDiff Slope 10/20/30) and Day_Night.

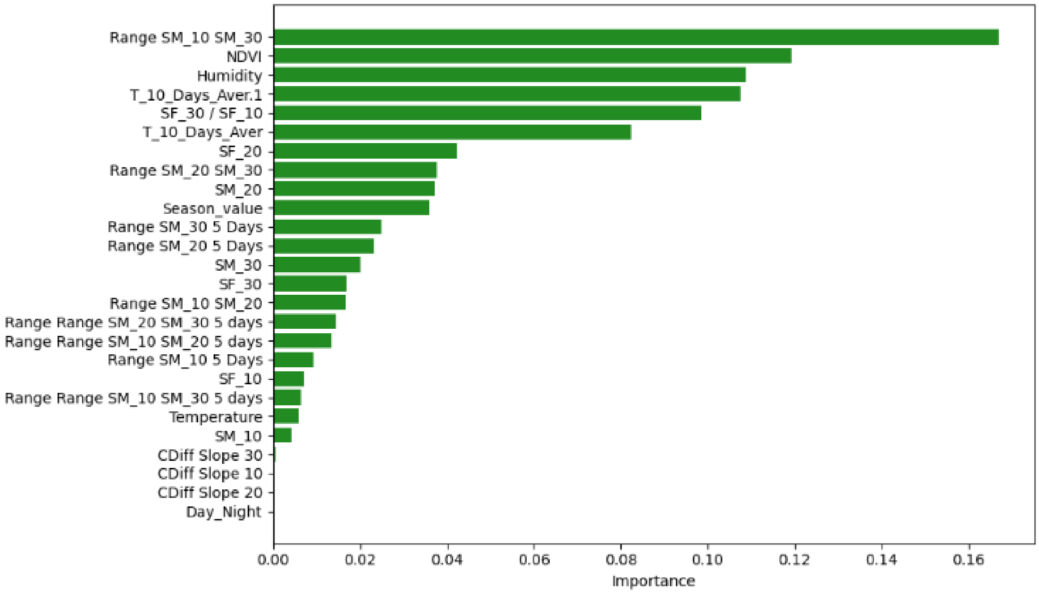


Figure 42. Feature Importance for Predicting Clay Ratio - Gradient Boosting Regressor

5.3.3.3.Feature Importance for Predicting Clay Ratio of XGBoost Regressor

XGBoost places the strongest emphasis on Range SM_10 SM_30, assigning it the highest individual score among all models, followed by SF_30 / SF_10. This underlines the significance of vertical moisture distribution and signal frequency as key indicators of clay content. While Humidity, T_10_Days_Aver.1, and NDVI remain important, the sharp drop in importance among mid-ranked features suggests that the model relies more heavily on a few dominant variables.

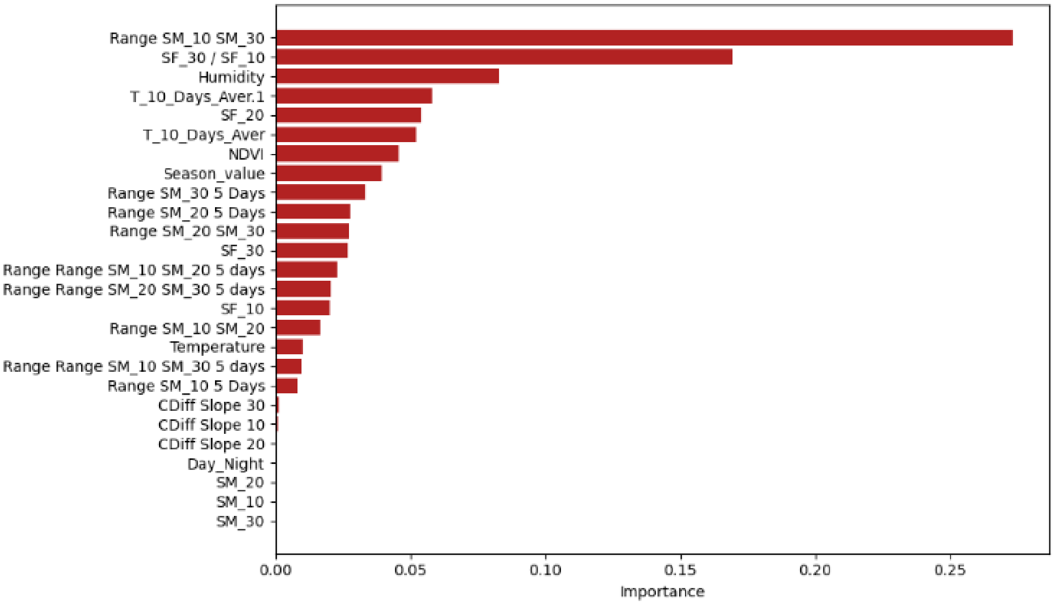


Figure 43. Feature Importance for Predicting Clay Ratio - XGBoost Regressor

5.3.4. Evaluation of Model Predictions: True vs. Predicted Soil Composition

5.3.4.1. Random Forest Regressor

The scatter plot for the Random Forest Regressor shows a moderate alignment between predicted and actual clay ratio values. Although most predictions follow the general trend of the ideal 1:1 line, the distribution exhibits noticeable dispersion, particularly at higher clay ratio levels. This spread suggests the model has difficulty generalizing for extreme values and may suffer from moderate overfitting. The reported R² of 0.936 and RMSE of 0.054 support this observation, indicating acceptable but limited predictive accuracy compared to more advanced methods.

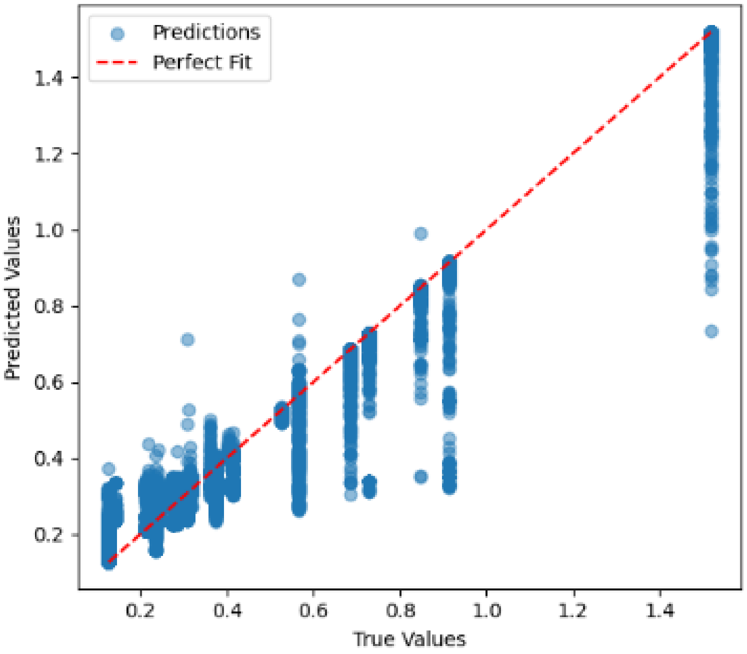


Figure 44. Random Forest Regressor: Predicted vs. True Clay Ratio Values

5.3.4.2. XGBoost Regressor

The XGBoost Regressor demonstrates a high degree of predictive accuracy. The predicted values closely follow the diagonal reference line, with minimal deviation throughout the entire range of clay ratios. The model achieves an R² of 0.993 and an RMSE of 0.019, reflecting its ability to capture complex, non-linear relationships between soil features and clay content. This performance highlights XGBoost's strength in modeling intricate patterns with high fidelity and low error rates.

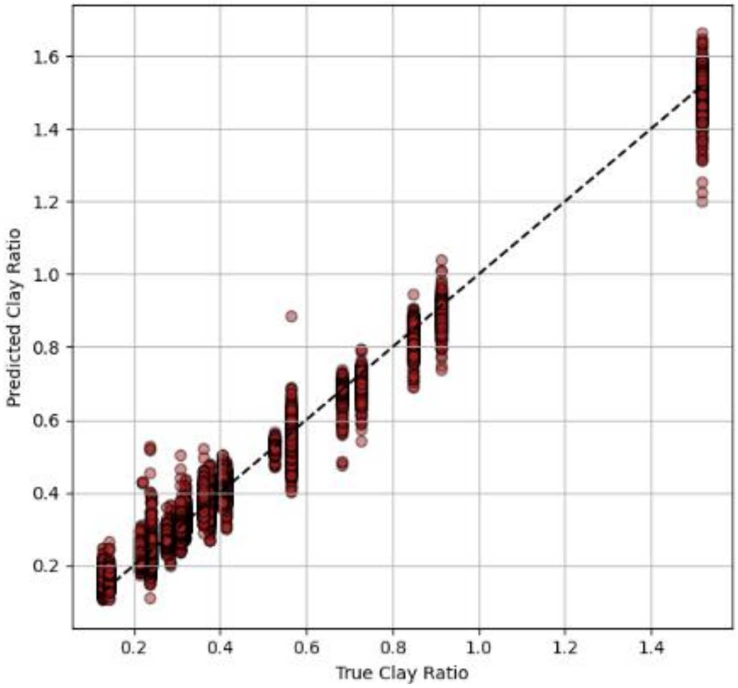


Figure 45. XGBoost Regressor: Predicted vs. True Clay Ratio Values

5.3.4.3. Gradient Boosting Regressor

The Gradient Boosting Regressor yields highly accurate predictions with a tightly clustered point distribution along the perfect fit line. It also attains an R² of 0.993, with a slightly lower RMSE of 0.018. The results suggest that Gradient Boosting offers both low bias and low variance, making it especially suitable for clay ratio prediction in datasets characterized by subtle variations and interactions among input features.

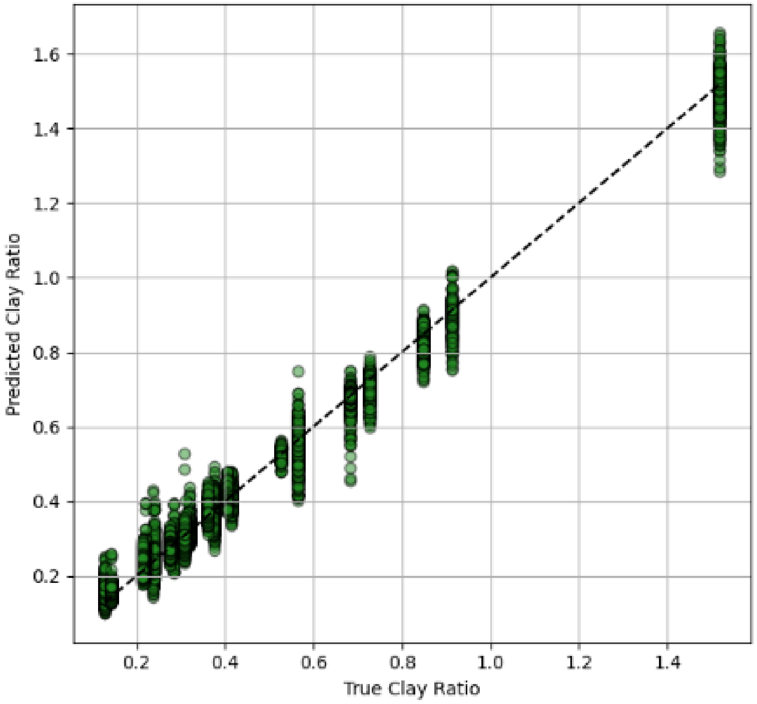


Figure 46. Gradient Boosting Regressor: Predicted vs. True Clay Ratio Values

Boosting-based models (XGBoost and Gradient Boosting) outperform Random Forest in terms of both precision and reliability. Their superior alignment with actual values and significantly lower RMSE values underscore their suitability for high-resolution clay content mapping, soil classification, and precision agricultural applications. The consistency in performance between XGBoost and Gradient Boosting also indicates their robustness in handling spatial and temporal heterogeneity in environmental data.

Model selection

Based on the comprehensive evaluation of model performance using scatter plots, error metrics, and learning curves, the Extreme Gradient Boosting (XGBoost) model emerged as the most accurate and robust algorithm for predicting the Clay Ratio among the tested machine learning methods.

The performance metrics clearly demonstrate that both the Gradient Boosting Regressor (GBR) and XGBoost models significantly outperformed the Random Forest (RF) model. The R^2 scores for GBR and XGBoost reached 0.9747 and 0.9748 respectively, compared to 0.936 for the Random Forest. Similarly, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values were substantially lower for XGBoost and GBR (MSE = 0.0015; RMSE = 0.038) compared to the Random Forest (MSE = 0.003; RMSE = 0.054). These results indicate that ensemble gradient-based methods are better suited for capturing the complex, non-linear relationships between environmental predictors and soil texture attributes.

The scatter plots of predicted versus observed Clay Ratio values reinforce these findings. XGBoost predictions aligned more closely along the 1:1 line with minimal dispersion, indicating high precision and low bias across the prediction range. GBR also exhibited strong agreement, though with slightly greater variance than XGBoost. In contrast, the Random Forest model showed more deviation from the 1:1 line, particularly at extremes, suggesting a tendency to underpredict or overpredict under certain conditions.

Learning curve analysis further confirmed the superior generalization ability of the XGBoost model. Both XGBoost and GBR demonstrated rapid convergence of training and validation scores with minimal overfitting. However, the learning curve for Random Forest revealed a wider gap between training and validation performance, suggesting some degree of overfitting and limited scalability when trained on smaller subsets.

In summary, XGBoost is identified as the best-performing model for Clay Ratio prediction in this study. Its higher predictive accuracy, strong generalization, and interpretability make it the most suitable algorithm for soil texture modeling using remote sensing, meteorological, and in-situ data in heterogeneous agro-environmental contexts.

Interpretation

The feature importance analysis of the XGBoost model for predicting the clay fraction reveals a coherent pattern grounded in the physical and functional properties of the soil–plant–atmosphere continuum. The most influential predictor, Range SM_10 SM_30 (importance = 0.2731), captures the vertical gradient of soil moisture between shallow and deeper layers. This range reflects the differential capacity of soils to store and transmit water, attributes strongly governed by soil texture. In clay-rich soils, water moves more slowly and tends to accumulate in deeper horizons due to the small pore size and high retention capacity. In contrast, sandy soils permit rapid infiltration and exhibit sharper moisture gradients, making this variable an effective indicator of underlying texture composition.

The significant contribution of the SF_30 / SF_10 ratio (0.1694) further emphasizes the relevance of vertical profiling in capturing subsurface textural transitions. Such indices likely encapsulate stratification effects and variations in soil compaction, which modulate water movement and retention across horizons, key determinants of clay content.

Climatic and biophysical variables such as humidity (0.0830), temperature (T_10_Days_Aver.1 and T_10_Days_Aver), and NDVI (0.0456) play important secondary roles. These features influence or reflect evapotranspiration demand and vegetation status, which are intimately linked to soil hydraulic behavior. Temperature, in particular, governs evapotranspiration and hence the rate at which soil moisture is depleted, while vegetation indices offer indirect cues about soil fertility and water availability, both affected by texture.

The consistent appearance of range-based moisture indicators, whether between depths (e.g., Range SM_20 SM_30) or over time (e.g., Range SM_30 5 Days), highlights the dynamic behavior of moisture redistribution. These variables capture the temporal and spatial heterogeneity of water movement, which is texture-dependent. Fine-textured soils like clay exhibit slower wetting and drying cycles, leading to more stable moisture profiles over time, while sandy soils show sharper fluctuations. These dynamics are effectively leveraged by the model to infer texture.

The minimal importance of absolute soil moisture values (e.g., SM_10, SM_20, SM_30 all at 0) underscores the added value of using relative and dynamic indicators instead. Static measurements fail to capture the complexity of water retention behavior across depth and time, especially when used in isolation. Likewise, topographic proxies such as CDiff Slope and diurnal patterns like Day_Night contribute little, indicating that under controlled or homogeneous terrain conditions, their utility in predicting textural variability is limited.

Overall, the XGBoost model's structure reveals an intricate interplay among moisture dynamics, climatic demand, and vegetative response. This reinforces the rationale for using integrated multi-source environmental data when modeling soil texture, where both vertical water movement and seasonal plant–soil–climate interactions become critical predictors.

5.4. Conclusion

In conclusion, the prediction of the clay fraction using machine learning models demonstrates the strong potential of data-driven approaches for soil texture characterization. Among the tested algorithms, XGBoost achieved the highest predictive performance, with an R^2 score of 0.9748 and the lowest error metrics, confirming its robustness in capturing complex, non-linear relationships between environmental predictors and clay content. The analysis of feature importance revealed that dynamic indicators of soil moisture variability, particularly the vertical moisture gradient between 10 cm and 30 cm depth, were the most informative variables. These findings emphasize the relevance of moisture-related features over static measurements, reflecting the intrinsic link between water retention behavior and soil texture. Additionally, variables representing climatic demand and vegetative status, such as humidity, temperature, and NDVI, further enhanced model accuracy by capturing the soil–plant–atmosphere interactions. Overall, this section illustrates the effectiveness of combining multi-source environmental data with ensemble learning techniques to improve the spatial and temporal assessment of soil physical properties, supporting precision agriculture and hydrological modeling efforts.

Based on the conclusions above the following thesis was formulated:

Thesis 5

The vertical gradient of soil moisture, quantified as the differential between shallow and subsurface layers (e.g., Range SM_10–SM_30), encapsulates the moisture behavior governed by fine particle distribution and thus emerges as a key predictor of clay ratio. When integrated with spectral vegetation proxies (NDVI) and environmental variables, this dynamic moisture signal enables physically informed, non-invasive estimation of clay-enriched soils through ensemble learning frameworks. However, in coarse-textured soils, where infiltration is rapid and vertical moisture gradients are less distinct, the predictive strength of this indicator may be reduced, highlighting the need for context-specific calibration and validation.

6. SUMMARY

This research develops and validates an integrated, machine learning–driven framework for soil moisture characterization and soil texture prediction using multi-source environmental data. By combining radar and optical remote sensing data, in-situ soil moisture profiles from Sentek EnviroSCAN sensors, and meteorological variables, the research addresses critical gaps in the spatial monitoring of soil hydrophysical properties, particularly in data-scarce and semi-arid environments.

The core innovation of this work lies in its non-invasive and data-driven approach to soil texture prediction, achieved entirely without the need for laboratory particle size analysis. This marks a significant advancement in digital soil mapping (DSM), enabling accurate estimation of sand, silt, clay fractions and Clay Ration based solely on environmental proxies derived from remote sensing and in-situ soil moisture dynamics.

For soil moisture characterization, the study leverages Sentinel-1 C-band polarizations (VV, VH) alongside vegetation (NDVI) and water indices (NDWI) from Sentinel-2, coupled with elevation and climatic data. In the Tunisian test site, temporal and spatial patterns of radar backscatter and vegetation response are found to correlate strongly with soil moisture variability, enabling effective mapping of wetness regimes across land uses and seasons.

For soil texture prediction, the thesis applies Random Forest, Gradient Boosting, and XGBoost regressors trained on a curated set of in-situ and remote sensing features, achieving high predictive accuracy for all soil fractions ($R^2 > 0.90$), with the XGBoost model reaching $R^2 = 0.9748$ for the clay ratio. Feature importance analysis reveals that dynamic and relative indicators, such as vertical soil moisture gradients (e.g., Range SM_10–30), NDVI, and short-term temperature averages, are the most powerful predictors of soil texture. These variables outperform static moisture values by capturing the nuanced interplay of infiltration, retention, and vegetation–climate–soil interactions.

The scientific contribution of this thesis is twofold:

1. It establishes a new paradigm for soil texture estimation that eliminates the need for physical sampling and lab testing, reducing cost, time, and logistical constraints.

2. It enhances the scalability of digital soil monitoring through a fusion of earth observation, sensor data, and ensemble machine learning, offering a transferable framework for agricultural, hydrological, and environmental applications.

Policy and Practical Implications

The framework has strong relevance for drought early warning systems, agricultural advisory services, and irrigation optimization. At the same time, the use of remotely derived soil texture prediction for sensor calibration without laboratory analysis opens new opportunities for low-cost, large-scale deployment of soil monitoring networks. At the policy level, the framework provides actionable information for climate adaptation, land-use planning, and sustainable groundwater management, particularly in semi-arid and data-limited regions. It also aligns with international development goals, including SDG 2 (Zero Hunger), SDG 6 (Clean Water and Sanitation), and SDG 13 (Climate Action), by promoting efficient and sustainable land and water resource management.

Limitations and Future Research

Limitations include the regional scope of the case studies (Tunisia and Hungary) and the use of C-band SAR, which limits subsurface sensitivity under dense vegetation. Future research should expand validation across diverse climates and soil types, incorporate multi-frequency SAR (L- and P-band) and thermal data, and test operational integration with real-time IoT sensor networks. Field trials should further evaluate the effectiveness of texture-driven sensor calibration under varying soil and climate conditions.

References

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, *44*(1), 71–91. https://doi.org/https://doi.org/10.1016/j.compag.2004.03.002

Ammar, S. Ben, Zouari, K., Leduc, C., & M'Barek, J. (2006). Caractérisation isotopique de la relation barrage-nappe dans le bassin du Merguellil (Plaine de Kairouan, Tunisie centrale). *Hydrological Sciences Journal*, *51*(2), 272–284. https://doi.org/10.1623/hysj.51.2.272

Anderson, M. C., Norman, J. M., Mecikalski, J. R., Otkin, J. A., & Kustas, W. P. (2007). A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research Atmospheres*, *112*(10). https://doi.org/10.1029/2006JD007506

Baghdadi, N., Zribi, M., Loumagne, C., Ansart, P., & Anguela, T. P. (2008). Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields. *Remote Sensing of Environment*, *112*(12), 4370–4379. https://doi.org/10.1016/j.rse.2008.08.004

Bakshi, C. (2020). .Random Forest Regression - Level Up Coding. . *Medium.* .

Bédir, M., Soltani, A., Belhaj Mohamed, A., Arbi, A., & Saidi, M. (2020). *Cretaceous petroleum system modeling of Kairouan Basin in eastern Tunisia*. https://doi.org/10.1007/s12517-020-05550-0/Published

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bogena, H. R., Huisman, J. A., Oberdörster, C., & Vereecken, H. (2007). Evaluation of a low-cost soil water content sensor for wireless network applications. *Journal of Hydrology*, *344*(1), 32–42. https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.06.032

Breiman, L. (1996). Bagging predictors. . *Machine Learning, 24(2), 123–140*.

Breiman, L. (2001a). *Random Forests* (Vol. 45).

Breiman, L. (2001b). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brocca, L., Ciabatta, L., Massari, C., Camici, S., & Tarpanelli, A. (2017). Soil moisture for hydrological applications: Open questions and new opportunities. *Water (Switzerland)*, *9*(2). https://doi.org/10.3390/w9020140

Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2009). Soil Moisture Temporal Stability over Experimental Areas in Central Italy. *Geoderma*, *148*, 364–374. https://doi.org/10.1016/j.geoderma.2008.11.004

Bronick, C. J., & Lal, R. (2005). Soil structure and management: a review. *Geoderma*, *124*(1), 3–22. https://doi.org/https://doi.org/10.1016/j.geoderma.2004.03.005

Carlson, T. N., Gillies, R. R., & Schmugge, T. J. (1995). An interpretation of methodologies for indirect measurement of soil water content. *Agricultural and Forest Meteorology*, *77*(3), 191–205. https://doi.org/https://doi.org/10.1016/0168-1923(95)02261-U

Cassel, D. K., & Nielsen, D. R. (1986). Field Capacity and Available Water Capacity. In *Methods of Soil Analysis* (pp. 901–926). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.2136/sssabookser5.1.2ed.c36

Cato the Elder, & V. M. T. (1934). On agriculture (W. D. Hooper & H. B. Ash, Trans.). *Loeb Classical Library 283. Harvard University Press.*

Chen, F., Van de Voorde, T., Roberts, D., Zhao, H., & Chen, J. (2021). Detection of ground materials using normalized difference indices with a threshold: Risk and ways to improve. *Remote Sensing*, *13*(3), 1–10. https://doi.org/10.3390/rs13030450

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chulli, B. , F. G. , J. N. , M. (2011). Impact of changing climate in the Kairouan Hydrological basin (central Tunisia). *Journal of Environmental Science and Engineering, 5, Pp. 682688.*

Chung, J., Lee, Y., Kim, J., Jung, C., & Kim, S. (2022). Soil Moisture Content Estimation Based on Sentinel-1 SAR Imagery Using an Artificial Neural Network and Hydrological Components. *Remote Sensing*, *14*(3), 465. https://doi.org/10.3390/rs14030465

Columella, L. J. M. (1941). On Agriculture. *Loeb Classical Library 361. Harvard University Press., Volume I.*

Dane, J. H., & Topp, G. C. (2002). Methods of Soil Analysis. Part 4: Physical Methods. . *Soil Science Society of America, Inc., Madison, USA.*

Daniel Hillel. (1980). Applications of Soil Physics. *Academic Press.*

Daniel Hillel. (2004a). Hillel, D. (2004) Introduction to Environmental Soil Physics. *Academic Press, an Imprint of Elsevier Science.* https://books.google.tn/books?id=qK7zzLIUEloC&lpg=PP1&ots=-KfrF1HuvF&lr&pg=PR6#v=onepage&q&f=false

Daniel Hillel. (2004b). Introduction to Environmental Soil Physics. Elsevier Academic Press, Amsterdam, 2004. *Elsevier Academic Press*.

Daniel J. Hillel. (1991). Out of the Earth: Civilization and the Life of the Soil. 1991. The Free Press Division of Macmillan, Inc., New York, New York. x + 321 pp. $22.50, cloth. *American Journal of Alternative Agriculture*, *6*(3), 114–114. https://doi.org/DOI: 10.1017/S0889189300004021

de Rosny, G., Chanzy, A., Pardé, M., Gaudu, J.-C., Frangi, J.-P., & Laurent, J.-P. (2001). Numerical Modeling of a Capacitance Probe Response. *Soil Science Society of America Journal*, *65*(1), 13–18. https://doi.org/https://doi.org/10.2136/sssaj2001.65113x

Deák, T., Dobai, A., Károly, K., Molnár, F., & Dobos, E. (2024). Spatial extension of soil water regime variables derived from soil moisture values using geomorphological variables in Hungary. *Hungarian Geographical Bulletin*, *73*(4), 337–353. https://doi.org/10.15201/hungeobull.73.4.1

Dean, T. J., Bell, J. P., & Baty, A. J. B. (1987). Soil moisture measurement by an improved capacitance technique, Part I. Sensor design and performance. *Journal of Hydrology*, *93*(1), 67–78. https://doi.org/https://doi.org/10.1016/0022-1694(87)90194-6

Dobos, E., & Luca Montanarella. (2010). *A quantitative procedure for building physiographic units supporting a global SOTER database Luca Montanarella European Commission*. https://www.researchgate.net/publication/288382467

Dobos, E., Micheli, E., Baumgardner, M. F., Biehl, L., & Helt, T. (2000). Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*, *97*(3), 367–391. https://doi.org/https://doi.org/10.1016/S0016-7061(00)00046-X

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

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., … Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/https://doi.org/10.1016/j.rse.2017.07.001

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

Entekhabi, D., Rodriguez-Iturbe, I., & Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology*, *184*(1), 3–17. https://doi.org/https://doi.org/10.1016/0022-1694(95)02965-6

Evett, S., & Evett, S. R. (2003). *Soil Water Measurement by Neutron Thermalization. Soil Water Measurement by Neutron Thermalization*. https://doi.org/10.1081/E-EWS

Evett, S. R., & Steiner, J. L. (1995). Precision of Neutron Scattering and Capacitance Type Soil Water Content Gauges from Field Calibration. *Soil Science Society of America Journal*, *59*(4), 961–968. https://doi.org/https://doi.org/10.2136/sssaj1995.03615995005900040001x

Famiglietti, J. S., & Wood, E. F. (1994). Multiscale modeling of spatially variable water and energy balance processes. *Water Resources Research*, *30*(11), 3061–3078. https://doi.org/https://doi.org/10.1029/94WR01498

Fletcher, & Karen. (2012). *SENTINEL 1: ESAs Radar Observatory Mission for GMES Operational Services. European Space Agency.* . Https://S Entinel.Esa.Int/Web/Sentinel/User-Guides/Sentinel-1-Sar/Overview.

Franks, S., & Rengarajan, R. (2023). Evaluation of Copernicus DEM and Comparison to the DEM Used for Landsat Collection-2 Processing. *Remote Sensing*, *15*(10). https://doi.org/10.3390/rs15102509

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gábris, G., Pécsi, M., Schweitzer, F., & Telbisz, T. (2018). Relief. National Atlas of Hungary. . *Natural Environment, 42-57.*

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, Z., Zhu, Y., Liu, C., Qian, H., Cao, W., & Ni, J. (2018). Design and test of a soil profile moisture sensor based on sensitive soil layers. *Sensors (Switzerland)*, *18*(5). https://doi.org/10.3390/s18051648

Gardner, W. H. (1986). Water Content. In *Methods of Soil Analysis* (pp. 493–544). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.2136/sssabookser5.1.2ed.c21

Gee, G. W., & Bauder, J. W. (1986). Particle-size Analysis. In *Methods of Soil Analysis* (pp. 383–411). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.2136/sssabookser5.1.2ed.c15

Grayson, R. B., & Western, A. W. (1998). Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *Journal of Hydrology*, *207*(1), 68–82. https://doi.org/https://doi.org/10.1016/S0022-1694(98)00096-1

Greacen, E. L. (1981). *Soil water assessment by the neutron method*. CSIRO.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. https://doi.org/10.5194/essd-11-717-2019

Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, *152*(3), 195–207. https://doi.org/https://doi.org/10.1016/j.geoderma.2009.06.003

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., & Verdin, J. P. (2008). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. *Geophysical Research Letters*, *35*(22). https://doi.org/10.1029/2008GL035772

Hajdu, I., Yule, I., Bretherton, M., Singh, R., & Hedley, C. (2019). Field performance assessment and calibration of multi-depth AquaCheck capacitance-based soil moisture probes under permanent pasture for hill country soils. *Agricultural Water Management*, *217*, 332–345. https://doi.org/https://doi.org/10.1016/j.agwat.2019.03.002

Hamdi, M., Zagrarni, M. F., Laouini, G., Tarhouni, J., Laaouini, G., & Müller, H. W. (2017). Assessment of groundwater flow dynamic using GIS tools and 3D geological modeling: Case of Sisseb El Alem-Nadhour Saouaf basin, Northeastern Tunisia. In *International Journal of Innovation and Applied Studies* (Vol. 19, Issue 1). http://www.ijias.issr-journals.org/

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hermassi, T., Cherif, M. A., & Habaieb, E. H. (2014). Etude du transport solide au niveau du bassin versant de Merguellil, Tunisie centrale: Cas des bassins versants d'Ettiour et de Rajela. *Houille Blanche*, *2014-August*(4), 88–96. https://doi.org/10.1051/lhb/2014043

Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma*, *214–215*, 141–154. https://doi.org/https://doi.org/10.1016/j.geoderma.2013.09.016

Idso, S. B., Jackson, R. D., Reginato, R. J., Kimball, B. A., & Nakayama, F. S. (1975). The Dependence of Bare Soil Albedo on Soil Water Content. *Journal of Applied Meteorology and Climatology*, *14*(1), 109–113. https://doi.org/10.1175/1520-0450(1975)014<0109:TDOBSA>2.0.CO;2

Jiang, Z., Huete, A. R., Chen, J., Chen, Y., Li, J., Yan, G., & Zhang, X. (2006). Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment*, *101*(3), 366–378. https://doi.org/https://doi.org/10.1016/j.rse.2006.01.003

Joseph, V. R. (2022). *Optimal Ratio for Data Splitting*. https://doi.org/10.1002/sam.11583

Kelleners, T. J., Soppe, R. W. O., Ayars, J. E., & Skaggs, T. H. (2004). Calibration of Capacitance Probe Sensors in a Saline Silty Clay Soil. *Soil Science Society of America Journal*, *68*(3), 770–778. https://doi.org/https://doi.org/10.2136/sssaj2004.7700

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., & Mecklenburg, S. (2010). The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. *Proceedings of the IEEE*, *98*(5), 666–687. https://doi.org/10.1109/JPROC.2010.2043032

Kibaroglu, A. , B. S. , & A. D. (2016). Water Resources Management in Semi-Arid Regions: Challenges and Innovations. *Springer*.

Kibirige, D., & Dobos, E. (2020). Soil moisture estimation using citizen observatory data, microwave satellite imagery, and environmental covariates. *Water (Switzerland)*, *12*(8). https://doi.org/10.3390/W12082160

Kibirige, D., & Dobos, E. (2021). Off-Site Calibration Approach of EnviroScan Capacitance Probe to Assist Operational Field Applications. *Water*, *13*(6). https://doi.org/10.3390/w13060837

Kocsis, K., Gercsák, G., Kovács, Z., Nemerkényi, Zsombor., Kincses, A., Tóth, G., Agárdi, Norbert., Koczó, Fanni., Mezei, G., & McIntosh, R. William. (2021). *National atlas of Hungary : society*. Research Centre for Astronomy and Earth Sciences, Geographical Institute.

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/https://doi.org/10.1016/j.jhydrol.2012.10.044

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., … Yamada, T. (2004). Regions of Strong Coupling Between Soil Moisture and Precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Le Goulven, P., Leduc, C., Salah Bachta, M., & Poussin, J.-C. (2009). *7 Sharing Scarce Resources in a Mediterranean River Basin: Wadi Merguellil in Central Tunisia*.

Luxmoore, R. J. (1981). *Micro-, meso-, and macroporosity of soil. 45:3*. https://doi.org/10.2136/sssaj1981.03615995004500030051x

Minasny, B., & McBratney, Alex. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311. https://doi.org/https://doi.org/10.1016/j.geoderma.2015.07.017

Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, *16*(1), 1–9. https://doi.org/10.2136/vzj2016.10.0105

Mougou, R., Mansour, M., Iglesias, A., Chebbi, R. Z., & Battaglini, A. (2011). Climate change and agricultural vulnerability: A case study of rain-fed wheat in Kairouan, Central Tunisia. *Regional Environmental Change*, *11*(SUPPL. 1), 137–142. https://doi.org/10.1007/s10113-010-0179-4

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). Interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 481–486. https://doi.org/10.1109/36.377948

Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, *184*(1), 101–129. https://doi.org/https://doi.org/10.1016/0022-1694(95)02970-2

Nrcs, U. (2017). *Soil Survey Manual Soil Science Division Staff Agriculture Handbook No. 18*.

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. A., Draper, C. S., Hagimoto, Y., Kerr, Y. H., Larson, K. M., Njoku, E. G., Small, E. E., & Zreda, M. (2013). State of the Art in Large-Scale Soil Moisture Monitoring. *Soil Science Society of America Journal*, *77*(6), 1888–1919. https://doi.org/https://doi.org/10.2136/sssaj2013.03.0093

Paltineanu, I. C., & Starr, J. L. (1997). Real-time Soil Water Dynamics Using Multisensor Capacitance Probes: Laboratory Calibration. *Soil Science Society of America Journal*, *61*(6), 1576–1585. https://doi.org/https://doi.org/10.2136/sssaj1997.03615995006100060006x

Peng, J., Loew, A., Merlin, O., & Verhoest, N. E. C. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, *55*(2), 341–366. https://doi.org/10.1002/2016RG000543

Pradipta, A., Soupios, P., Kourgialas, N., Doula, M., Dokou, Z., Makkawi, M., Alfarhan, M., Tawabini, B., Kirmizakis, P., & Yassin, M. (2022). Remote Sensing, Geophysics, and Modeling to Support Precision Agriculture—Part 1: Soil Applications. *Water (Switzerland)*, *14*(7). https://doi.org/10.3390/w14071158

Provenzano, G., Rallo, G., de Almeida, C. D. G. C., & de Almeida, B. G. (2020). Development and validation of a new calibration model for diviner 2000® probe based on soil physical attributes. *Water (Switzerland)*, *12*(12), 1–16. https://doi.org/10.3390/w12123414

Reynolds, W. D., Drury, C. F., Yang, X. M., Fox, C. A., Tan, C. S., & Zhang, T. Q. (2007). Land management effects on the near-surface physical quality of a clay loam soil. *Soil and Tillage Research*, *96*(1–2), 316–330. https://doi.org/10.1016/j.still.2007.07.003

Richards, L. A. (1931). CAPILLARY CONDUCTION OF LIQUIDS THROUGH POROUS MEDIUMS. *Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

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

Robock, A., Vinnikov, K., Srinivasan, G., Entin, J., Hollinger, S., Speranskaya, N., Liu, S., & Namkhai, A. (2000). The Global Soil Moisture Data Bank. *Bulletin of the American Meteorological Society*, *81*. https://doi.org/10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2

Roland Meynart, Steven P. Neeck, & Haruhisa Shimoda. (2013). Sensors, Systems, and Next-Generation Satellites XVII. . *Proceedings of SPIE, Volume 8889, Article 88890K.*

Sarra BEL HAJ SALEM. (2013). Utilisation des traceurs environnementaux pour l'etude des modes de recherche des eaux du bassin de Zeroud (Plaine de Kairouan). *Doctoral Dissertation, University of Sfax, Sfax, Tunisia*, *pp. 51, 2, p. 272284*.

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/https://doi.org/10.2136/sssaj2005.0117

Scobie, M. (2006). *Sensitivity of Capacitance Probes to Soil Cracks Courses ENG4111 and ENG4112 Research Project*.

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), 125–161. https://doi.org/https://doi.org/10.1016/j.earscirev.2010.02.004

Sentek Pty Ltd. (2011). *Calibration manual for Sentek soil moisture sensors (Version 2.0).* . Sentek Technologies. Https://Www.Sentek.Com.Au.

Six, J., Bossuyt, H., Degryze, S., & Denef, K. (2004). A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research*, *79*(1), 7–31. https://doi.org/https://doi.org/10.1016/j.still.2004.03.008

Spinoni, J., Szalai, S., Szentimrey, T., Lakatos, M., Bihari, Z., Nagy, A., Németh, Á., Kovács, T., Mihic, D., Dacic, M., Petrovic, P., Kržič, A., Hiebl, J., Auer, I., Milkovic, J., Štepánek, P., Zahradnícek, P., Kilar, P., Limanowka, D., … Vogt, J. (2015). Climate of the Carpathian

Region in the period 1961–2010: climatologies and trends of 10 variables. *International Journal of Climatology*, *35*(7), 1322–1341. https://doi.org/https://doi.org/10.1002/joc.4059

Starr, J. L., & Paltineanu, I. C. (1998). Soil Water Dynamics Using Multisensor Capacitance Probes in Nontraffic Interrows of Corn. *Soil Science Society of America Journal*, *62*(1), 114–122. https://doi.org/https://doi.org/10.2136/sssaj1998.03615995006200010015x

SUHET. (2021). *Sentinel-2 User Handbook*. European Space Agency (ESA). ESA Standard Document GMES-S2OP-EOPG-TN-13-0001. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook

Topp, G. C., Davis, J. L., & Annan, A. P. (1980). Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research*, *16*(3), 574–582. https://doi.org/https://doi.org/10.1029/WR016i003p00574

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/https://doi.org/10.1016/0034-4257(79)90013-0

Valentin, C., & Bresson, L.-M. (1992). Morphology, genesis and classification of surface crusts in loamy and sandy soils. *Geoderma*, *55*(3), 225–245. https://doi.org/https://doi.org/10.1016/0016-7061(92)90085-L

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, *46*(4). https://doi.org/10.1029/2008WR006829

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M. H., Amelung, W., Aitkenhead, M., Allison, S. D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., … Young, I. M. (2016). Modeling Soil Processes: Review, Key Challenges, and New Perspectives. *Vadose Zone Journal*, *15*(5), vzj2015.09.0131. https://doi.org/https://doi.org/10.2136/vzj2015.09.0131

Verhoef, A., De Bruin, H. A. R., & Van Den Hurk, B. J. J. M. (1997). *Some Practical Notes on the Parameter kB 1 for Sparse Vegetation*.

Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103359

Wang, L., & Gao, Y. (2023a). Soil Moisture Retrieval from Sentinel-1 and Sentinel-2 Data Using Ensemble Learning over Vegetated Fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 1802–1814. https://doi.org/10.1109/JSTARS.2023.3242264

Wang, L., & Gao, Y. (2023b). Soil Moisture Retrieval From Sentinel-1 and Sentinel-2 Data Using Ensemble Learning Over Vegetated Fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *PP*, 1–15. https://doi.org/10.1109/JSTARS.2023.3242264

Weil, R., & Brady, N. (2017). *The Nature and Properties of Soils. 15th edition*.

Western, A. W., Grayson, R. B., & Blöschl, G. (2002). Scaling of Soil Moisture: A Hydrologic Perspective. *Annual Review of Earth and Planetary Sciences*, *30*(Volume 30, 2002), 149–180. https://doi.org/https://doi.org/10.1146/annurev.earth.30.091201.140434

White, K. D. (1970). Roman farming. *Ithaca, N.Y., Cornell University Press*.

Yang, L., Wei, W., Chen, L., Jia, F., & Mo, B. (2012). Spatial variations of shallow and deep soil moisture in the semi-arid Loess Plateau, China. *Hydrology and Earth System Sciences*, *16*(9), 3199–3217. https://doi.org/10.5194/hess-16-3199-2012

Zhang, D., Tang, R., Zhao, W., Tang, B., Wu, H., Shao, K., & Li, Z. L. (2014). Surface soil water content estimation from thermal remote sensing based on the temporal variation of land surface temperature. *Remote Sensing*, *6*(4), 3170–3187. https://doi.org/10.3390/rs6043170

Zhang, H., Wang, S., Liu, K., Li, X., Li, Z., Zhang, X., & Liu, B. (2022). Downscaling of AMSR-E Soil Moisture over North China Using Random Forest Regression. *ISPRS International Journal of Geo-Information*, *11*(2). https://doi.org/10.3390/ijgi11020101

Zhang, Q., Liu, X., Liu, M., Zou, X., Zhu, L., & Ruan, X. (2021). Comparative analysis of edge information and polarization on sar-to-optical translation based on conditional generative adversarial networks. *Remote Sensing*, *13*(1), 1–20. https://doi.org/10.3390/rs13010128

Zribi, M., Albergel, C., & Baghdadi, N. (2020). Editorial for the special issue "Soil moisture retrieval using radar remote sensing sensors." In *Remote Sensing* (Vol. 12, Issue 7). MDPI AG. https://doi.org/10.3390/rs12071100

Zribi, M., & Dechambre, M. (2003). A new empirical model to retrieve soil moisture and roughness from C-band radar data. *Remote Sensing of Environment*, *84*(1), 42–52. https://doi.org/https://doi.org/10.1016/S0034-4257(02)00069-X