Scientific Research Publishing

Advances in Remote Sensing, 2023, 12, 99-122
https://www.scirp.org/journal/ars
ISSN Online: 2169-2688
ISSN Print: 2169-267X

Enhancing Surface Soil Moisture Estimation through Integration of Artificial Neural Networks Machine Learning and Fusion of Meteorological, Sentinel-1A and Sentinel-2A Satellite Data

Jephter Ondieki[1*], Giovanni Laneve[2], Maria Marsella[1], Collins Mito[3]

[1]DIMA-Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Roma, Italy
[2]SIA-School of Aerospace Engineering, Sapienza University of Rome, Roma, Italy
[3]Department of Physics, University of Nairobi, Nairobi, Kenya
Email: *jephterongige.ondieki@uniroma1.it

How to cite this paper: Ondieki, J., Laneve, G., Marsella, M. and Mito, C. (2023) Enhancing Surface Soil Moisture Estimation through Integration of Artificial Neural Networks Machine Learning and Fusion of Meteorological, Sentinel-1A and Sentinel-2A Satellite Data. *Advances in Remote Sensing*, **12**, 99-122.
https://doi.org/10.4236/ars.2023.124006

Received: October 10, 2023
Accepted: December 1, 2023
Published: December 4, 2023

Abstract

For many environmental and agricultural applications, an accurate estimation of surface soil moisture is essential. This study sought to determine whether combining Sentinel-1A, Sentinel-2A, and meteorological data with artificial neural networks (ANN) could improve soil moisture estimation in various land cover types. To train and evaluate the model's performance, we used field data (provided by La Tuscia University) on the study area collected during time periods between October 2022, and December 2022. Surface soil moisture was measured at 29 locations. The performance of the model was trained, validated, and tested using input features in a 60:10:30 ratio, using the feed-forward ANN model. It was found that the ANN model exhibited high precision in predicting soil moisture. The model achieved a coefficient of determination (R^2) of 0.71 and correlation coefficient (R) of 0.84. Furthermore, the incorporation of Random Forest (RF) algorithms for soil moisture prediction resulted in an improved R^2 of 0.89. The unique combination of active microwave, meteorological data and multispectral data provides an opportunity to exploit the complementary nature of the datasets. Through preprocessing, fusion, and ANN modeling, this research contributes to advancing soil moisture estimation techniques and providing valuable insights for water resource management and agricultural planning in the study area.

Keywords

Soil Moisture Estimation Techniques Fusion, Active Microwave,

Multispectral Data, Agricultural Planning

1. Introduction

Understanding hydrological processes, agricultural productivity, and water resource management all depend on having an accurate estimation of surface soil moisture [1]. At both a regional and global level, soil moisture has been retrieved using a variety of remote sensing techniques. The sensitivity of optical remote sensing to atmospheric effects and vegetation cover prevents it from accurately estimating soil moisture, despite the fact that it still offers useful information about land surface properties [2]. In order to get around the shortcomings of optical sensors for estimating soil moisture, the integration of active microwave data, such as Sentinel-1, has emerged as a promising strategy [3].

To increase the accuracy of soil moisture estimation, the fusion of remote sensing data has attracted a lot of attention recently [4]. Active microwave and multispectral data fusion has demonstrated great promise for more accurately capturing the spatiotemporal dynamics of soil moisture. Sentinel-1 is the best satellite for soil moisture monitoring because it can observe in any weather and at any time of day or night thanks to its Synthetic Aperture Radar (SAR) capabilities. Sentinel-2A, on the other hand, provides invaluable insights into the characteristics of the land surface and the dynamics of the vegetation thanks to its high spatial resolution multispectral data [5].

By examining the backscatter measurements at various polarizations, radar is a useful tool for determining soil moisture (SM). Due to their superior penetration abilities, previous research has shown that microwave responses at low frequencies (P to L-band) are particularly sensitive to SM levels on both bare and vegetated surfaces [6]. The Japanese Advanced Land Observation Satellite-2 (ALOS-2) and the Argentine satellite SAOCOM (Satélite Argentino de Observación Con Microondas) are just two examples of remote sensing satellites that have synthetic-aperture radar (SAR) systems which operate at low frequencies.

Fortunately, SAR systems that operate at higher frequencies, such as the C and X bands, exemplified by satellites like Sentinel-1, Radar Imaging Satellite 1 (RISAT-1), RADARSAT-1 & 2, and Constellation of Small Satellites for the Mediterranean Basin Observation-SkyMed (COSMO-SkyMed), and TerraSAR-X, have proven to be effective in providing accurate results for soil moisture retrieval. Numerous research studies have effectively harnessed these high-frequency SAR systems for tasks such as soil moisture mapping and retrieval [7] [8] [9] and crop monitoring [8] [10].

Numerous theoretical and empirical backscattering models have been proposed for the estimation of soil moisture based on SAR images. These models, including those by [11] [12] [13] [14], rely on quad-polarized microwave SAR images (HV, VV, HH, and VH) and sensor characteristics like wavelength and incidence angle. By inverting these models, one can derive parameters for soil

permittivity and soil roughness. Topp's model [15] is an example where soil moisture can be estimated using soil permittivity, as demonstrated in various studies [9] [16]-[24]

However, a common challenge is the limited availability of quad-polarized SAR images. To address this limitation, some researchers have developed techniques to eliminate one of the unknown parameters, typically surface roughness, by incorporating in-situ measurements [19] [20]. This modification has allowed the adaptation of existing backscattering models, both empirical and theoretical, for use with dual-polarized SAR images, thus overcoming the constraint of quad-polarized data. Once the empirical model is established, soil moisture can be calculated by solving a single equation with one unknown parameter. However, the accuracy of the derived soil moisture depends on the precision of surface roughness measurement [25].

In recent times, deep learning and machine learning models have emerged as highly effective tools for predicting surface soil moisture over extensive spatial and temporal scales [7] [26] [27]. In contrast to physical models, these machine and deep learning models rely on data-driven approaches. They utilize a range of relevant input features to map their output. Examples of such input features include brightness temperature, SAR backscatter, sensor characteristics, geographic information, and meteorological variables, as outlined by [28]. During the training phase, these machine learning models learn the relationships and patterns within the input data to make predictions about soil moisture dynamics. Subsequently, they can be applied to unobserved data to assess their predictive performance. Among the favored machine learning models for soil moisture estimation is Artificial Neural Networks (ANN), as demonstrated in studies by [29] [30] [31].

[32] Conducted soil moisture estimation with a daily spatial resolution of 12 km using data from the Advanced Very High-Resolution Radiometer (AVHRR) and Tropical Rainfall Measuring Mission (TRMM). They employed Support Vector Regression (SVR) models at six different study sites, incorporating three input features: backscatter values, incidence angle from TRMM, and normalized difference vegetation index from AVHRR. Their findings indicated that the root mean square error (RMSE) was less than 2%, and the correlation coefficients for all sites ranged from 0.34 to 0.77. They also compared the SVR model's performance with that of Artificial Neural Networks (ANN) and Multivariate Linear Regression (MLR) models, concluding that SVR outperformed both ANN and MLR in terms of performance.

[33] proposed a method based on ANN with a 10 km spatial resolution for estimating daily soil moisture. They utilized data from the International Soil Moisture Network (ISMN), along with backscatter, local incidence angle, azimuth angle, Latitude, and Longitude to train the ANN model. Their results showed that ANN performed well on test datasets, achieving a correlation coefficient (R) of 0.82 and an RMSE of 0.04 m^3/m^3.

The studies mentioned above highlight a common challenge in soil moisture products, which is finding a balance between spatial and temporal resolutions. Some products offer higher spatial resolution but lower temporal coverage, while others prioritize temporal frequency over spatial detail. This trade-off depends on the choice of input features and the specific soil moisture products used to train machine learning models.

To address this issue and enhance both spatial and temporal resolutions in soil moisture retrieval, this study leverages the capabilities of dual-polarized Sentinel-1 and multi-spectral Sentinel-2 data. Additionally, it incorporates SRTM elevation data and meteorological information. The goal is to create high-resolution and frequent estimations of surface soil moisture.

To achieve this, a novel methodology is introduced, which combines microwave data from Sentinel-1, optical data from Sentinel-2, meteorological data, and topographic data from SRTM-DEM. This approach relies on a fully connected feed-forward Artificial Neural Network (ANN) as its core component. Initially, thirteen derived features are selected as input variables for training the ANN architecture.

2. Materials and Methods

2.1. Description of the Study Area

The research is centered on the Maccarese region in Italy, which is known for its varied land cover types, encompassing agricultural fields, natural vegetation, and urban areas (**Figure 1**). The accurate estimation of soil moisture in this particular area poses notable challenges. This is due to the fact that each of these land cover types can have distinct soil properties and moisture dynamics, making it more challenging to develop a single, accurate estimation model that works across all these environments. Also the soil properties vary significantly within the region, leading to spatial heterogeneity in soil moisture content. Factors such as soil type, texture, and compaction levels can affect soil moisture retention and distribution.

2.2. Datasets

We utilized publicly available satellite data, including Sentinel-1A (SAR), Sentinel-2A (optical) images and meteorological data. The Sentinel images were downloaded from the official website of the Corpenicus data hub (https://scihub.copernicus.eu/). A comprehensive description of the Sentinel images is provided in **Table 1**.

2.3. Field Measurement

We utilized soil moisture data collected by 29 soil measurement stations of the La Tuscia University during the field campaigns from October 2022 to December 2022. A universal random grid sampling approach was used in the field to measure soil moisture. The study area was divided into small square grids measuring

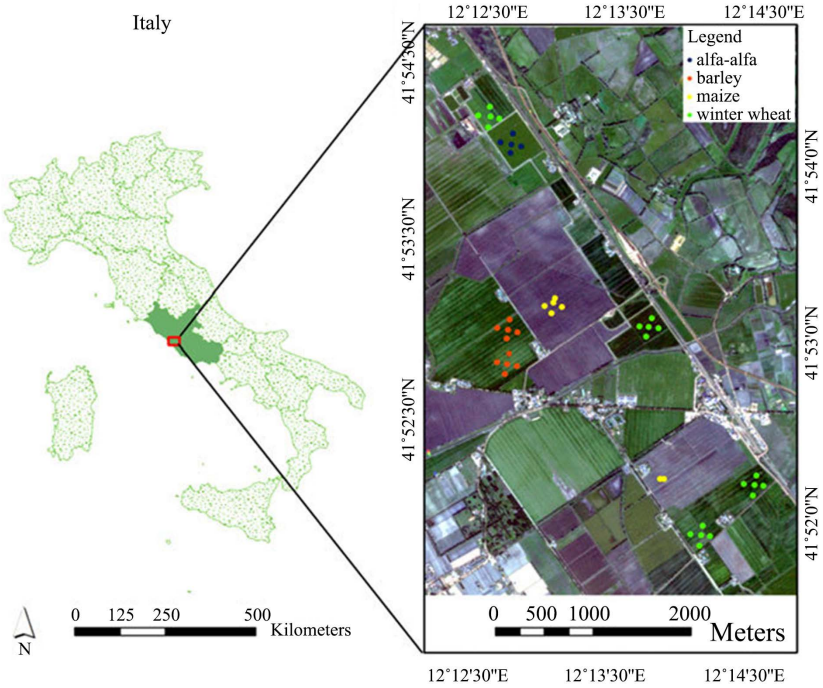


Figure 1. Location of the field test sites at Maccarese Farm, Rome, Italy.

Table 1. Detailed descriptions of the Sentinel-1A/2A images.

Date (dd/mm/yyyy) Sentinel-1A	Polarization	Pixel size(m)	Pass
06/10/2022	Dual (VH, VV)	10	Descending
07/10/202	Dual (VH, VV)	10	Ascending
18/10/2022	Dual (VH, VV)	10	Descending
19/10/2022	Dual (VH, VV)	10	Descending
30/10/2022	Dual (VH, VV)	10	Ascending
11/11/2022	Dual (VH, VV)	10	Descending
12/11/2022	Dual (VH, VV)	10	Ascending
23/11/2022	Dual (VH, VV)	10	Descending
24/11/2022	Dual (VH, VV)	10	Ascending
05/12/2022	Dual (VH, VV)	10	Descending
06/12/2022	Dual (VH, VV)	10	Ascending
17/12/2022	Dual (VH, VV)	10	Descending
18/12/2022	Dual (VH, VV)	10	Ascending
29/12/2022	Dual (VH, VV)	10	Descending
30/12/2022	Dual (VH, VV)	10	Ascending
Date (dd/mm/yyyy) Sentinel-2A	**Tale**	**Wavelength (nm)**	**Spatial Resolution (m)**
15/10/2022	76	646 - 685, 774 - 907	10
11/12/2022	76	646 - 685, 774 - 907	10

40 m × 40 m, and measurements were taken on a random grid. Each measurement was at least 40 meters apart. The sensor depth was 5 cm below the surface. The average of these measurements was then computed to obtain a representative soil moisture value for each grid. This method enabled a direct comparison of soil moisture at the point level to the corresponding satellite pixel.

2.4. Meteorological Data

The meteorological data which included precipitation, wind speed, cloud cover, solar illumination, soil temperature was obtained from the https://www.visualcrossing.com/weather-data website for the study.

2.5. Methodology

In order to extract input features for the machine learning model, we first processed the satellite images. Then, for training, validation, and testing, we engineered features and created a feed-forward multi-layer ANN model. Finally, we compared the output of the ANN model to RF benchmark algorithms and evaluated ANN performance in terms of error analysis and spatial distribution analysis. The subsections that follow go over the specifics of feature extraction and model setup.

2.5.1. Feature Extraction

1) Image processing

Sentinel-1A images were subjected to preprocessing using the Sentinel Application Platform (SNAP v8.0), which is a freely available open-source tool designed for Earth Observation data manipulation. The preprocessing of raw Sentinel-1A GRD images encompassed radiometric calibration, the removal of speckle noise, and terrain correction (Figure 2).

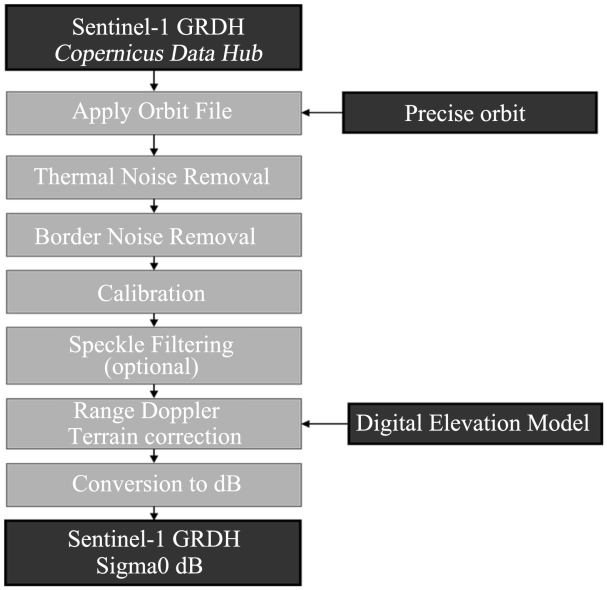


Figure 2. Sentinel-1 GRD preprocessing workflow.

In addition, Sentinel-2A images underwent processing to compute the normalized vegetation index (NDVI). This was achieved by calculating the ratio of the difference between near-infrared and red bands to their sum. The resulting NDVI image exhibited a spatial resolution of 10 × 10 meters. Pixel values in the NDVI image ranged from −1 to +1, with higher NDVI values serving as an indicator of robust and healthy vegetation [34].

2) Feature selection and scaling

The effectiveness of any machine learning model relies heavily on the quality of its input data. In our research, we carried out three distinct feature engineering procedures, namely feature extraction, generation, and scaling. Initially, we extracted 13 features from a combination of Sentinel-1A and Sentinel-2A images, as well as meteorological and DEM (Digital Elevation Model) data. These features encompassed backscatter values (σ_0) from Sentinel-1A images in both VV and VH polarizations. Radar backscatters (VH and VV) are particularly sensitive to soil moisture due to the presence of a dielectric gradient [35]. We also calculated NDVI values from Sentinel-2A imagery, and elevation data were normalized relative to mean sea level at each pixel using the DEM. Additionally, we obtained geographical coordinates (Latitude and Longitude) for every pixel in the input images.

Furthermore, we collected meteorological data from various monitoring stations, including details on soil temperature, precipitation, wind speed, cloud cover, and solar radiation. It's important to note that the well-established relationship between soil moisture and surface elevation was employed extensively in observing soil moisture patterns and developing machine learning models [36].

Geolocation variables (*i.e.*, Latitude and Longitude) were integrated into spatial machine learning applications to account for spatial dependencies in the data [31]. Additionally, we generated two synthetic features (VH/VV and VH-VV) by combining VH and VV through linear data fusion, augmenting the existing feature set (VH and VV). These synthetic features are more attuned to the geometric and dielectric properties of soil [37].

Then, the image pixels of the input features were rescaled to a standard grid size (10 × 10 m) using the nearest neighbor resampling method. Finally, we scaled each of the 13 features using the standard z-score method.

2.5.2. Model Setup

1) Feed-forward ANN

Three hidden layers were implemented in this study using TensorFlow's Keras API. The first layer, which serves as the input layer, was made up of 64 neurons, each with a Rectified Linear Unit (ReLU) activation function to add nonlinearity to the network. This layer is set up to accept input. Following that, two hidden layers were added, each with 64 neurons and ReLU activation functions. Notably, these hidden layers used L2 regularization with a regularization coefficient of 0.01, which is intended to reduce overfitting by penalizing large weight mag-

nitudes. The final layer, which consists of a single neuron, generated raw numerical predictions. This model was trained using the Adam optimizer, which has a learning rate of 0.001 and minimizes the mean squared error loss, which is a common choice for regression problems.

Furthermore, early stopping was implemented as a training callback, with a patience of 10 epochs and the option to restore the best weights, facilitating model convergence and preventing overfitting during the training process. This architecture is a versatile framework for a variety of regression tasks, providing depth and regularization techniques to optimize predictive performance while reducing the risk of overfitting.

A pixel-based approach was used in this study, with each pixel in the study area treated as an independent sample for soil moisture estimation. The extracted features are fed into the artificial neural network (ANN) model as input variables.

2.5.3. Model Evaluation

The performance of our trained model was assessed using the scaled testing dataset, with two key evaluation metrics, the coefficient of determination (R^2) and mean squared error (MSE). The MSE measures the average squared difference between the observed and predicted soil moisture values, providing insight into the model's precision. On the other hand, R-squared signifies the proportion of variability in the observed soil moisture values that can be accounted for by the model's predictions, indicating its explanatory power.

We also computed the correlation coefficient between the observed soil moisture values and the model's predictions for the training, validation, and testing datasets. This helped us understand the strength and direction of the relationship between the predicted and actual values.

To gauge the uncertainties in the model predictions, we calculated them by multiplying the square root of the MSE with the standard deviation of the respective datasets. This approach provides a measure of the dispersion or spread of the predicted values, indicating the level of confidence we can have in the model's predictions.

3. Results and Discussions

This study centered on the Maccarese region in Italy, we developed an Artificial Neural Network (ANN) model to estimate surface soil moisture. The results of the work were not only shown but also meticulously discussed, demonstrating the remarkable performance of the ANN model in accurately determining soil moisture levels in this diverse and complex region in this section.

3.1. The Temporal Changes in the Average Measured Soil Moisture and the Polarized Backscattering Coefficients of Sentinel-1 in Both VV and VH Polarizations

Figure 3 illustrates the temporal fluctuations in spatially averaged Soil Moisture

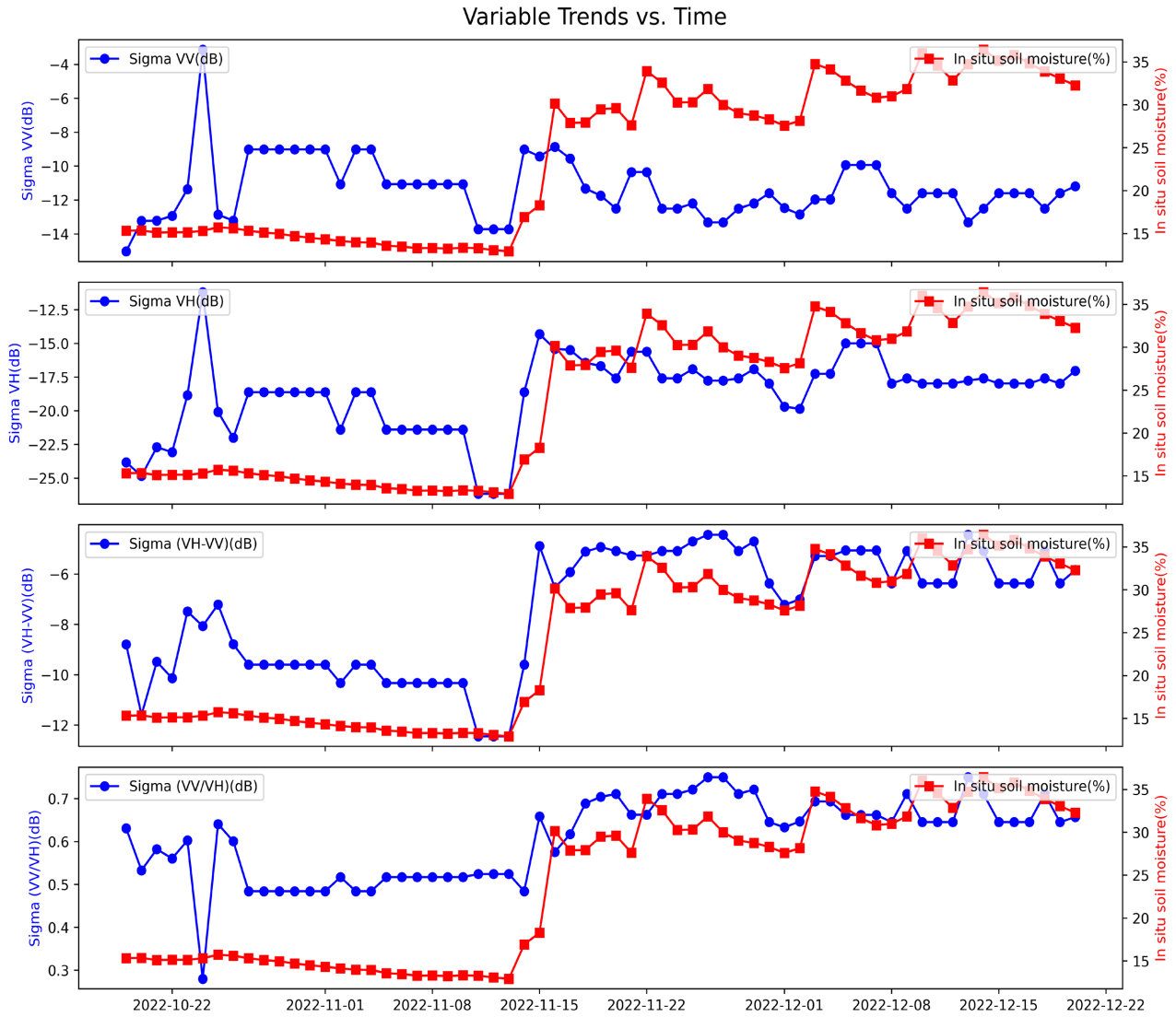


Figure 3. Presents the temporal changes in spatially averaged in-situ soil moisture (SM) measurements, alongside the corresponding spatially averaged Sentinel-1 backscattering coefficients (VV, VH, VV/VH, and VH-VV) for the Maccarese site.

(SM) values obtained from 63 selected pixels within the study area. These measurements correspond to the date of the Sentinel-1A overpass in the study region. This figure offers a visual representation of how soil moisture content changes over time in the selected area.

Additionally, the figure displays several Sentinel-1A backscattering coefficients, including VV, VH, VV/VH, and VH-VV polarized backscattering coefficients. These coefficients quantify the strength of the radar signal reflected back to the satellite from the Earth's surface. The spatial average presented in the figure reflects the average response of the radar signal across the study area. This provides valuable insights into the temporal variations in the intensity of radar signal reflections in relation to in-situ soil moisture levels.

The relationships and correlations between soil moisture and radar backscattering coefficients was assessed by examining the temporal variations of these

parameters. This data was useful for understanding the dynamics and interactions between soil moisture content and radar signals measured by Sentinel-1A.

A simple regression analysis of backscattering coefficients and soil moisture shows that Sigma VH, VV/VH, and VH-VV have a positive relationship (**Figure 4**). A positive correlation between the VH (vertical transmit and horizontal receive) radar backscattering coefficient and soil moisture suggests that as soil moisture increases, so does the VH coefficient. This means that wetter soil tends to produce stronger VH radar signals. A positive relationship between VH-VV (the difference between VH and VV) and soil moisture suggests that the significance of the difference between the VH and VV coefficients increases with soil moisture levels. This implies that the backscatter difference between these two polarizations is affected by soil moisture, with wetter soil having a more noticeable impact on VH-VV. Given that VH/VV (the ratio of VH to VV) and soil moisture have a positive correlation, the ratio tends to rise as soil moisture rises.

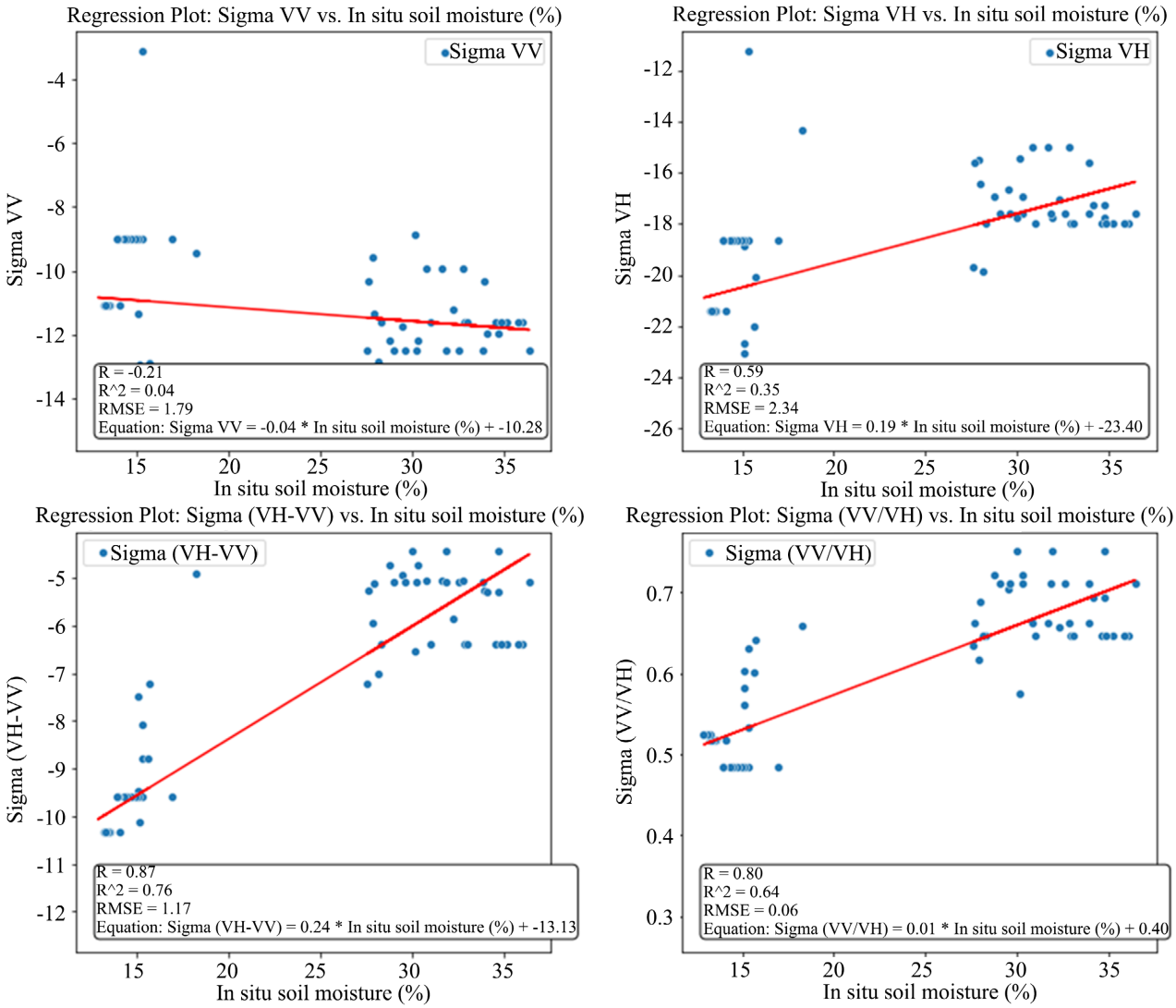


Figure 4. Scatter plot between observed SM and backscatter coefficient.

Since changes in surface roughness and dielectric properties can both be impacted by soil moisture, this ratio may be especially sensitive to those changes. In other words, as the backscattering coefficients increase, so does the soil moisture content. In fact, [35] claimed that soil moisture causes an increase in the superficial radar response because wet soil has a higher reflection coefficient than dry soil, which increases the intensity of the backscattering coefficient.

The VV (vertical transmit and vertical receive) radar backscattering coefficient has a slightly negative correlation with soil moisture, indicating that as soil moisture levels rise, the VV coefficient tends to decrease, albeit slightly. This implies that wetter soil produces weaker VV radar signals. Inverse relationships between SAR signal and soil moisture were also discovered in the literature, [38]. [39] initial hypothesis that it might be caused by variations in soil roughness contribution on the total backscatter is still being investigated as one of the causes of this behavior. The explanation given more recently was that it was a volumetric contribution [40] caused by subsurface backscattering [41].

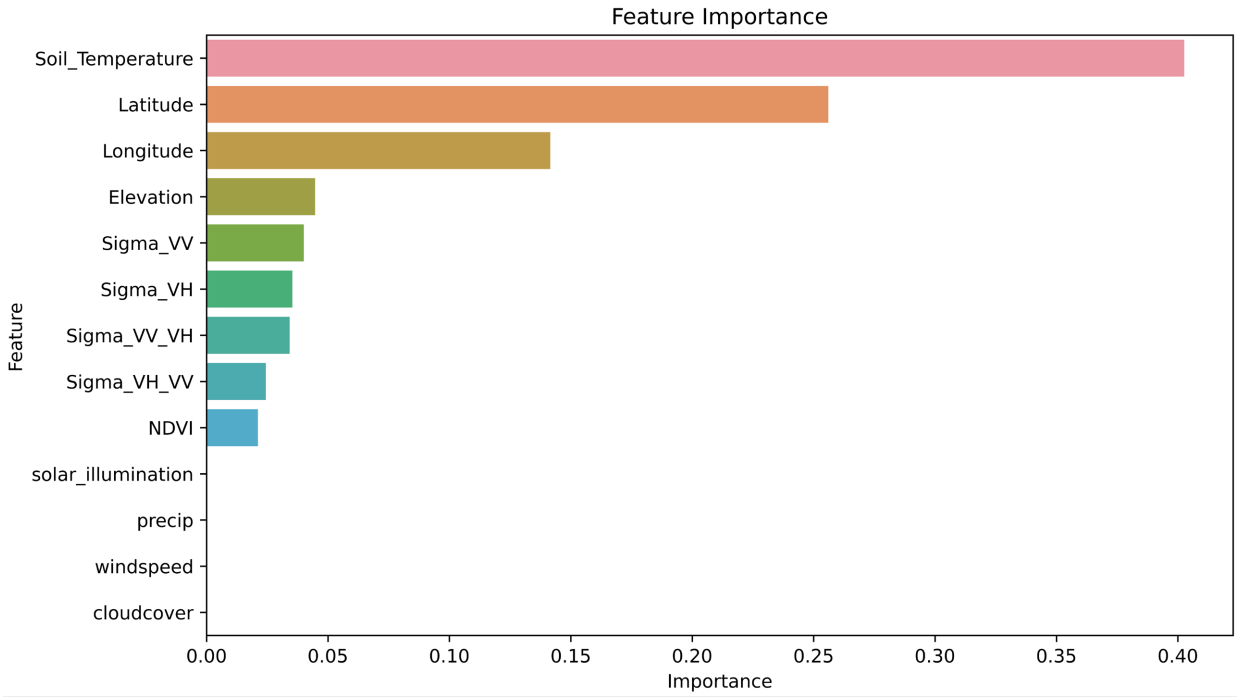
These correlations shed important light on how soil moisture affects how radar signals interact with it and how this interaction changes over time. The orientation of surface features, the makeup of the soil, and the amount of vegetation cover can all affect how each coefficient behaves specifically.

We show the relative importance score for each feature in **Figure 5(a)**. In our analysis, an ensemble of regression trees was built using the Least Squares Gradient Boosting (LSBoost) algorithm to assess the relative weights of the different features. Each tree was trained to predict the target variable using the input features. The following are the principal elements of our plan:

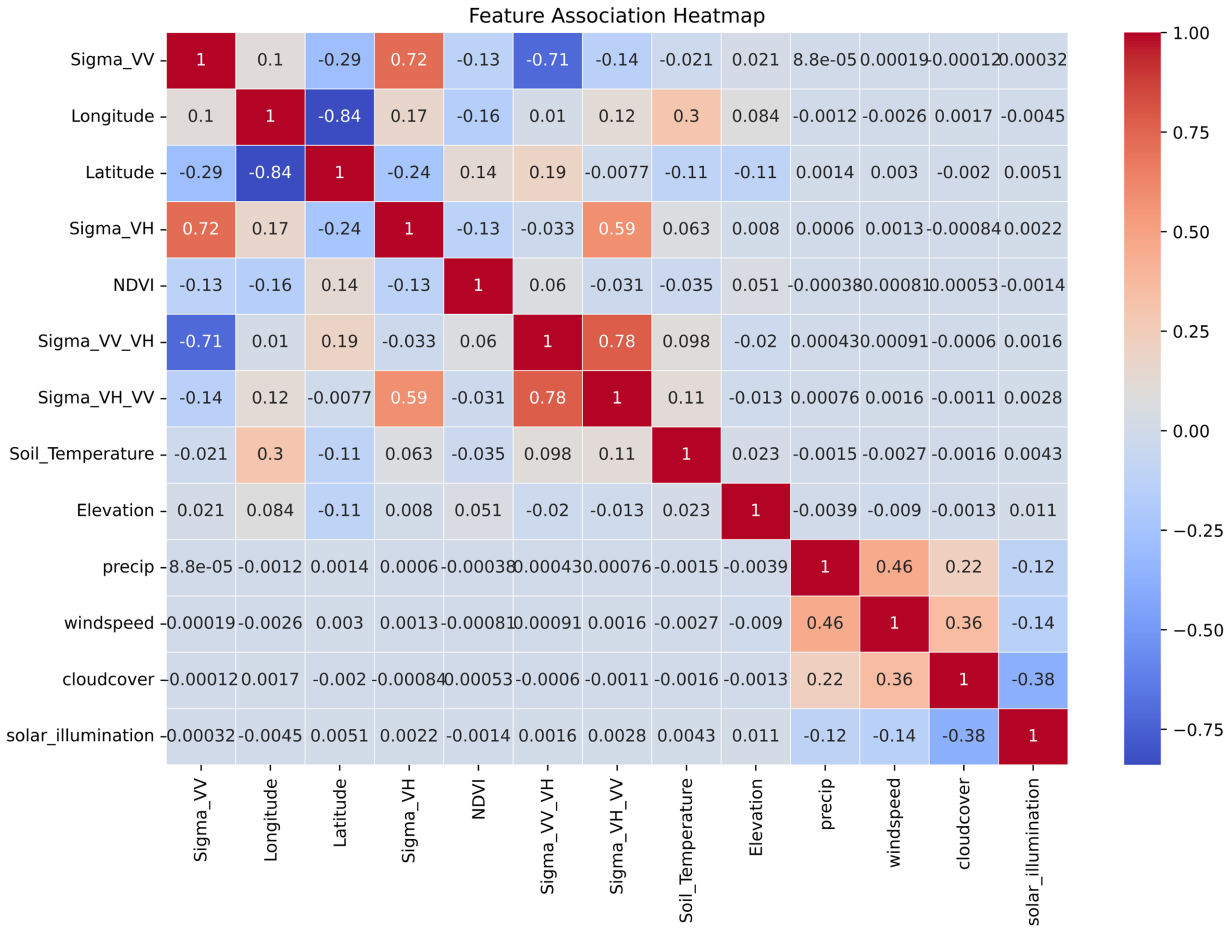
Ensemble of Regression Trees: To combine predictions from various trees and increase the reliability and precision of our predictive model, an ensemble of regression trees was used. The boosting technique was used, with each tree in the ensemble focusing on instances that had previously been incorrectly classified. Feature importance scores were calculated based on how frequently a feature was used for node splitting across all trees and how much it contributed to lowering prediction error.

More predictive power (*i.e.*, a more relevant feature) is indicated by a high value for the feature's importance. We found that the feature importance score for soil temperature, Longitude, Elevation, Sigma (σ_0) VV, and VH is very high. According to earlier research [35], the backscatter features (VV and VH) have a significant contribution. High Longitude contribution suggests that the geolocation feature may be under control. It's interesting to note that the importance scores of the synthetic features (VV/VH and VH-VV) produced through a linear data fusion are almost identical and relatively higher than those of other input features like, NDVI, solar illumination, wind speed, and cloud cover.

We generated a 13 × 13 matrix called a feature association matrix (**Figure 5(b)**). The strength of the association or similarity between any two feature pairs is represented by each cell in the matrix. Typically, the values are determined by applying a particular metric or methodology (such as correlation coefficients,


Feature Importance
Soil_Temperature
Latitude
Longitude
Elevation
Sigma_VV
Sigma_VH
Sigma_VV_VH
Sigma_VH_VV
NDVI
solar_illumination
precip
windspeed
cloudcover
Feature
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Importance

(a)


Feature Association Heatmap
Sigma_VV 1 0.1 -0.29 0.72 -0.13 -0.71 -0.14 -0.021 0.021 8.8e-05 0.00019 -0.00012 0.00032
Longitude 0.1 1 -0.84 0.17 -0.16 0.01 0.12 0.3 0.084 -0.0012 -0.0026 0.0017 -0.0045
Latitude -0.29 -0.84 1 -0.24 0.14 0.19 -0.0077 -0.11 -0.11 0.0014 0.003 -0.002 0.0051
Sigma_VH 0.72 0.17 -0.24 1 -0.13 -0.033 0.59 0.063 0.008 0.0006 0.0013 -0.00084 0.0022
NDVI -0.13 -0.16 0.14 -0.13 1 0.06 -0.031 -0.035 0.051 -0.00038 0.00081 0.00053 -0.0014
Sigma_VV_VH -0.71 0.01 0.19 -0.033 0.06 1 0.78 0.098 -0.02 0.00043 0.00091 -0.0006 0.0016
Sigma_VH_VV -0.14 0.12 -0.0077 0.59 -0.031 0.78 1 0.11 -0.013 0.00076 0.0016 -0.0011 0.0028
Soil_Temperature -0.021 0.3 -0.11 0.063 -0.035 0.098 0.11 1 0.023 -0.0015 -0.0027 -0.0016 0.0043
Elevation 0.021 0.084 -0.11 0.008 0.051 -0.02 -0.013 0.023 1 -0.0039 -0.009 -0.0013 0.011
precip -8.8e-05 -0.0012 0.0014 0.0006 -0.00038 0.00043 0.00076 -0.0015 -0.0039 1 0.46 0.22 -0.12
windspeed -0.00019 -0.0026 0.003 0.0013 -0.00081 0.00091 0.0016 -0.0027 -0.009 0.46 1 0.36 -0.14
cloudcover -0.00012 0.0017 -0.002 -0.00084 0.00053 -0.0006 -0.0011 -0.0016 -0.0013 0.22 0.36 1 -0.38
solar_illumination -0.00032 -0.0045 0.0051 0.0022 -0.0014 0.0016 0.0028 0.0043 0.011 -0.12 -0.14 -0.38 1
Sigma_VV
Longitude
Latitude
Sigma_VH
NDVI
Sigma_VV_VH
Sigma_VH_VV
Soil_Temperature
Elevation
precip
windspeed
cloudcover
solar_illumination
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

(b)

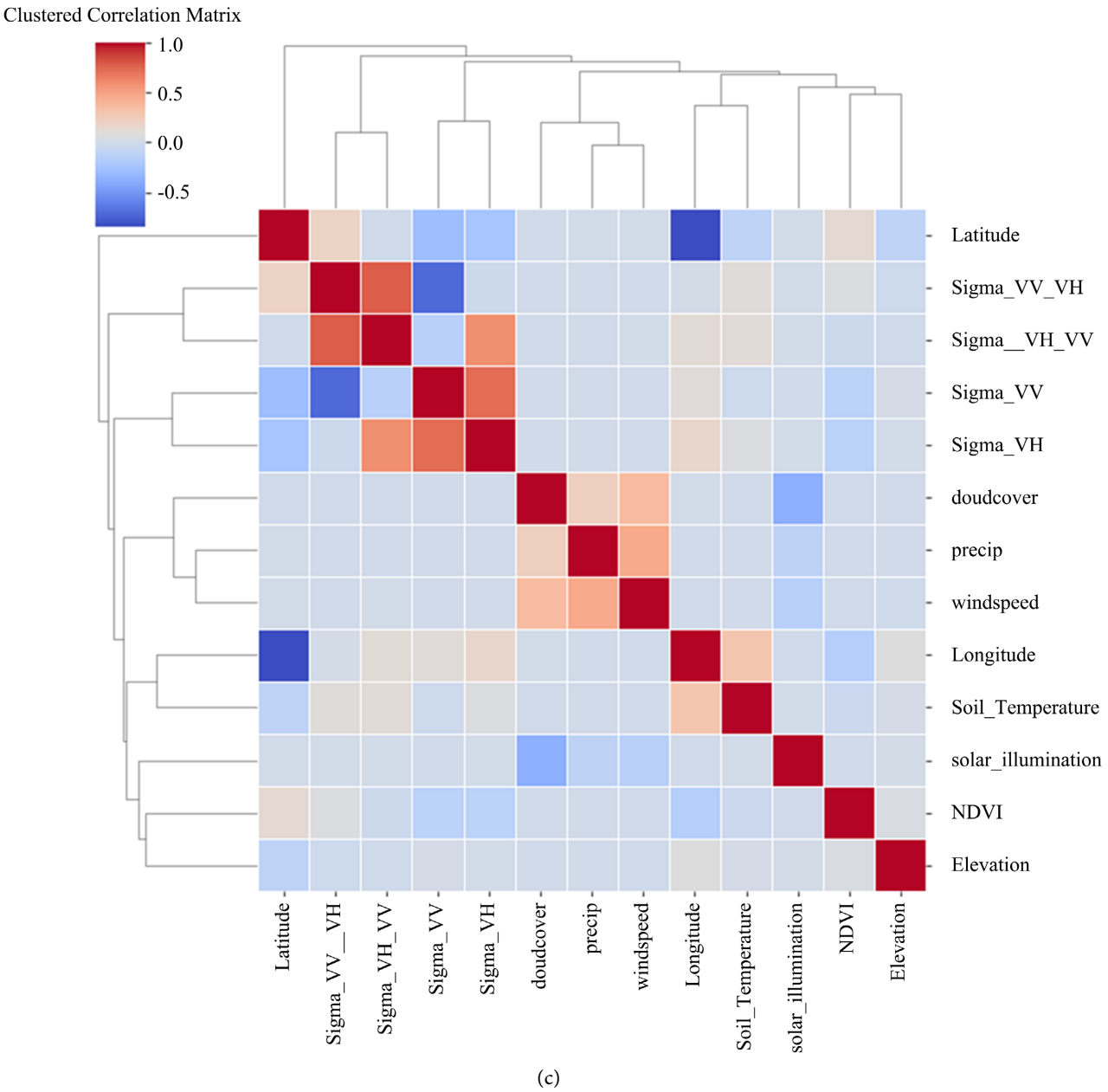


Figure 5. (a): Feature importance; (b): Feature association heat map; (c): Feature correlation matrix.

mutual information, etc.). This matrix was developed to evaluate the degree of correlation or similarity between the input features. Correlation describes how much two features vary together. When two features have a high correlation, it indicates that they tend to move in the same direction while a low correlation suggests that they move independently. As previously stated, highly correlated features can have a negative impact on the machine learning model. Here's how it's done:

Correlated features can make the model unstable by providing redundant information. Small changes in one feature can cause significant fluctuations in the model's predictions when features are highly correlated. Highly correlated fea-

tures can make the model more sensitive to data uncertainty or noise. This sensitivity can result in overfitting, in which the model fits the data noise rather than the underlying patterns.

A high correlation between the features is indicated by a high value in the feature association matrix (Figure 5(c)). We found no significant correlation between our features. This shows that the model was correctly trained using the input features without any instability or sensitivity.

3.2. Performance Evaluation of the Model

We conducted a thorough evaluation of our trained feed-forward ANN model's performance using different datasets to gauge the effectiveness of the training process. Initially, the model showed promising results on the training data, achieving an R-value of 0.85 and an R-squared value (R^2) of 0.72, as depicted in Figure 6.

However, it's essential to avoid bias by solely relying on training data. Therefore, we further assessed the model's performance using previously unseen data, including validation and testing datasets. During the parameter tuning phase,

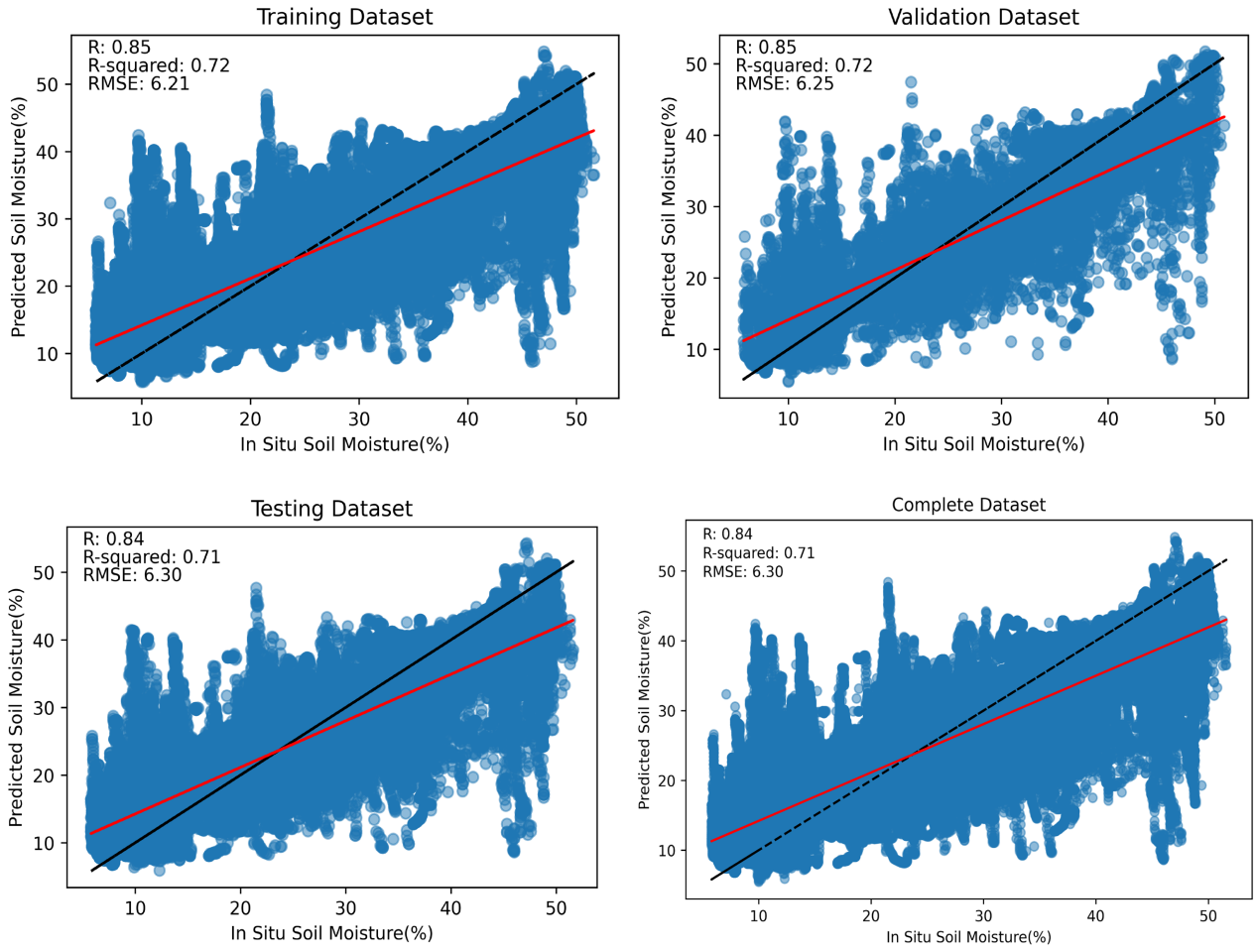


Figure 6. Model performance on training, validation, testing, and complete datasets. The dashed line in figure represents $y = x$ line.

the model's performance was evaluated using the validation data, resulting in an R-value of 0.85 and an R^2 of 0.72, demonstrating a strong agreement between in-situ and predicted soil moisture.

Subsequently, the testing data was introduced to our model, and its performance was evaluated. The measured soil moisture in the field closely matched the model's predictions, yielding an R-value of 0.84 and an R^2 of 0.71.

In summary, when considering all datasets (training, validation, and testing data), our trained model exhibited an overall accuracy with an R-value of 0.84 and an R^2 of 0.71, indicating its effectiveness in estimating soil moisture across different scenarios.

3.3. Error Analysis

Figure 7 shows the error in the training, validation, and testing process. The number of times the same error occurred is indicated by the height of the stacked bars. To represent errors ranging from −30% (leftmost bin) to +30% (rightmost bin), bin sizes of 20 were chosen. The vertical red line signifies the zero error. Underestimation is represented by the area to the left of this line, and overestimation is represented by the area to the right. We modeled the error histogram with a Gaussian distribution. The histogram would have a mean of zero and be normally distributed in a perfect world. The distribution of our histogram is almost normal, and the peak at the zero error line indicates a good fit.

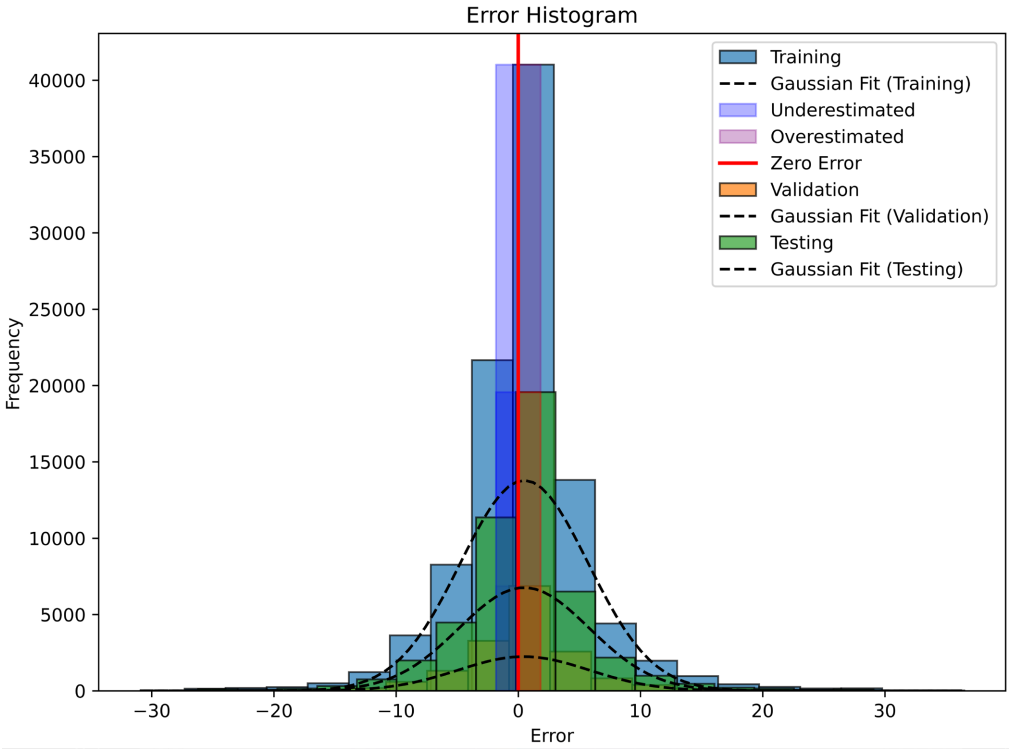


Figure 7. Shows the error histogram for the training, validation, and testing phases (with 20-bin spacing). Overestimated and underestimated regions are shown to the left and right of the zero error (red) line, respectively.

Figure 8 depicts the comparison between observed and predicted soil moisture values, along with a 95% confidence interval (C.I.). The figure shows that the predicted soil moisture values closely match the observed values, indicating a strong agreement between the model's predictions and the actual measurements. The inclusion of a 95% confidence interval provides a clear indication of the level of confidence in the model's predictions, underscoring the model's accuracy in estimating soil moisture.

3.4. Comparison with the RF Algorithms Approach

To ensure a comprehensive evaluation, we conducted a comparison between the performance of our fully connected feed-forward ANN (Figure 6) and the RF (Random Forest) (Figure 9) algorithms for predicting soil moisture using the same dataset. Our findings indicate that the proposed ANN architecture did not perform as well as the RF algorithm.

The RF algorithm achieved a higher R-squared value (R^2) of 0.88, indicating a better fit to the data, and a lower mean squared error (MSE) of 18.80, which signifies reduced prediction errors. Additionally, the mean absolute error (MAE) of 2.22 suggests that the RF algorithm had a smaller average absolute difference between predicted and observed soil moisture values.

These results are visualized in Figure 9, further highlighting the superior performance of the RF algorithm in comparison to our ANN model for soil moisture prediction

3.5. Sensitivity Analysis

To assess the consistency of our data-driven feed-forward ANN model, we conducted a sensitivity analysis. This analysis was crucial in understanding how our model responds to uncertainties in input features. We achieved this by introducing a 5% uncertainty in all the input features simultaneously while keeping other features constant. We then examined how uncertainties in individual features impacted the overall uncertainty in the response variable, which is soil moisture.

Our observations revealed that the uncertainty in the model-derived soil moisture fell within the range of 0% to +1% when input features had an uncertainty of approximately 5% (as shown in Figure 10). Notably, uncertainties in features such as VV, VH, wind speed, solar illumination, soil temperature, precipitation, and elevation did not have a significant impact on the model's performance. This suggests that when these features are subject to a 5% uncertainty, the model's predictions for soil moisture remain relatively unaffected, resulting in a smaller uncertainty in soil moisture.

On the other hand, features like NDVI, VH-VV, and longitude exhibited greater sensitivity to the presence of uncertainty. When these features experienced a 5% uncertainty, the model's predictions for soil moisture were more affected, leading to a larger uncertainty in the soil moisture estimates.

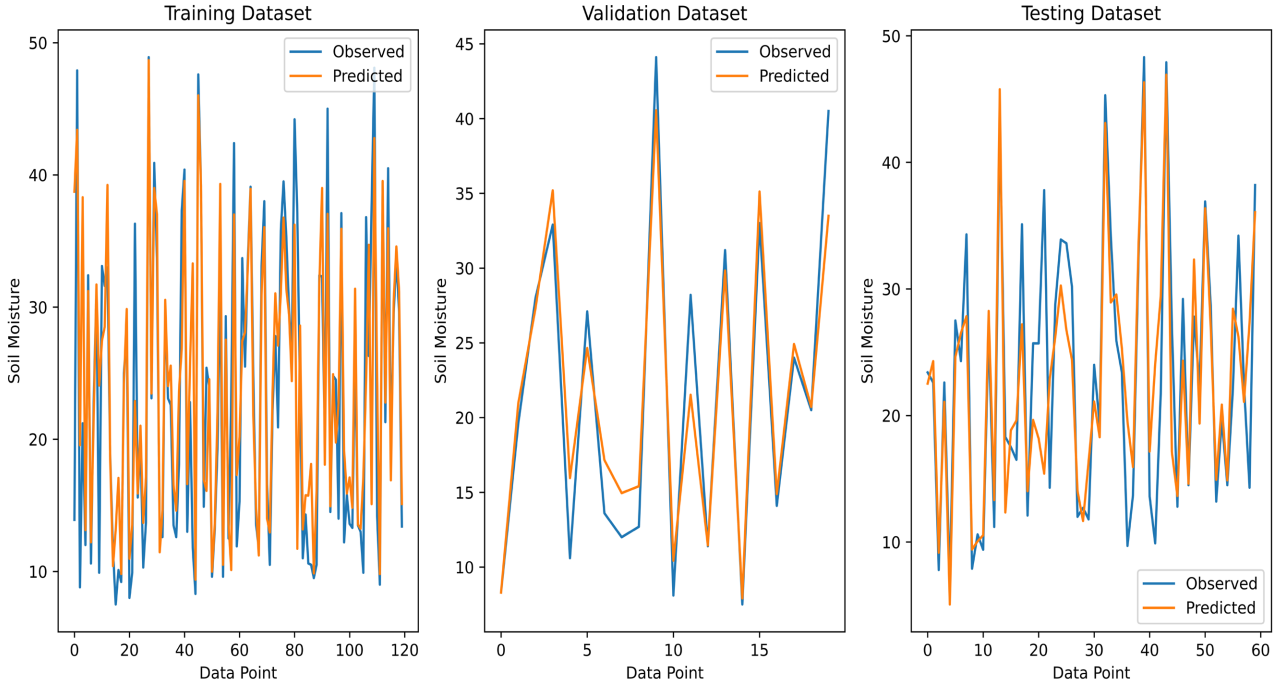


Figure 8. The line plot of observed and predicted soil moisture plotted for training, validation, and testing.

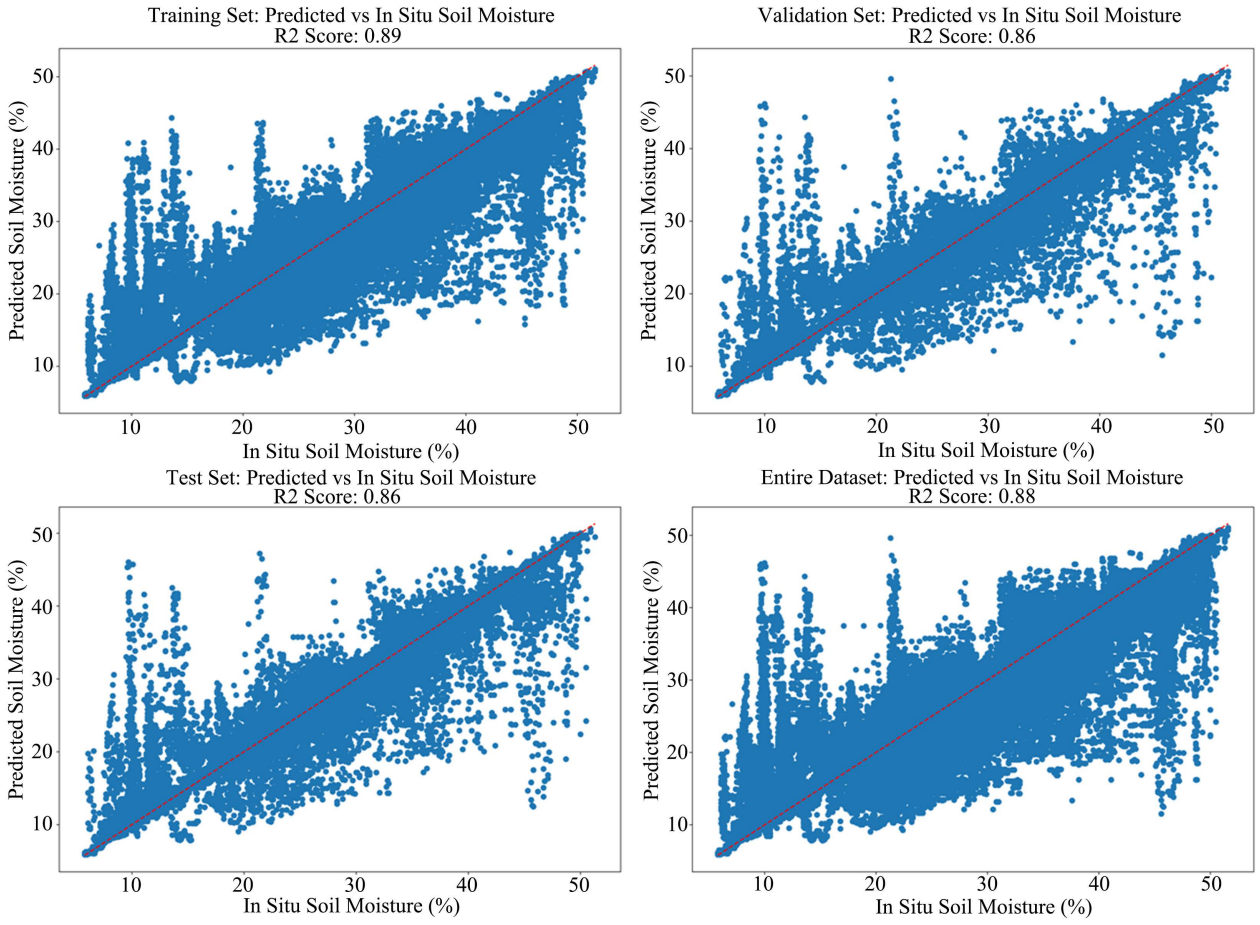


Figure 9. RF model performance on training, validation, testing, and complete datasets.

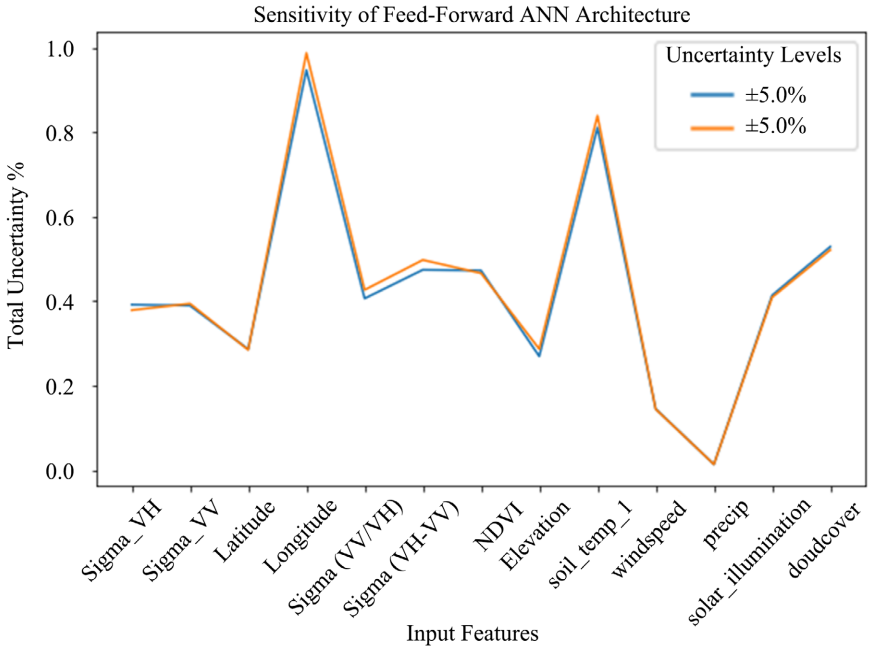


Figure 10. Sensitivity of feed-forward ANN by considering uncertainties (±5%) in input features.

3.6. Soil Moisture Map in Maccarese

Figure 11 show the surface soil moisture map generated by the ANN for the time period 19 October 2022. Soil moisture appears to be relatively high on the Maccarese at the right lower margin.

3.7. Discussions

The current study demonstrates the efficacy of estimating soil moisture in the study area using a combination of publicly available multi-sensor data, including Sentinel SAR, optical sensors, and meteorological sources. These findings support the viability of such an approach and shed light on the factors that influence soil moisture levels in this particular geographic area.

The use of a feed-forward artificial neural network (ANN) algorithm in conjunction with these various sensors produced remarkable results. The developed model's accuracy and reliability are reflected in its high coefficient of determination (R^2 greater than 0.72) and low root mean square error (RMSE) of 6.43%. These findings are consistent with previous studies [42] [43] that have been conducted in temperate regions such as Poland, Germany, and Italy. The consistency of these findings with prior research highlights the methodology's robustness and adaptability to different geographic contexts.

To estimate surface soil moisture, we used a fully connected feed-forward ANN algorithm and data fusion techniques. One of the most important contributions of this research is the identification of factors that have a significant impact on soil moisture levels in the study area. As shown in **Figure 5(a)**, soil temperature, geographic coordinates (latitude and longitude), VH (Vertical-Horizontal)

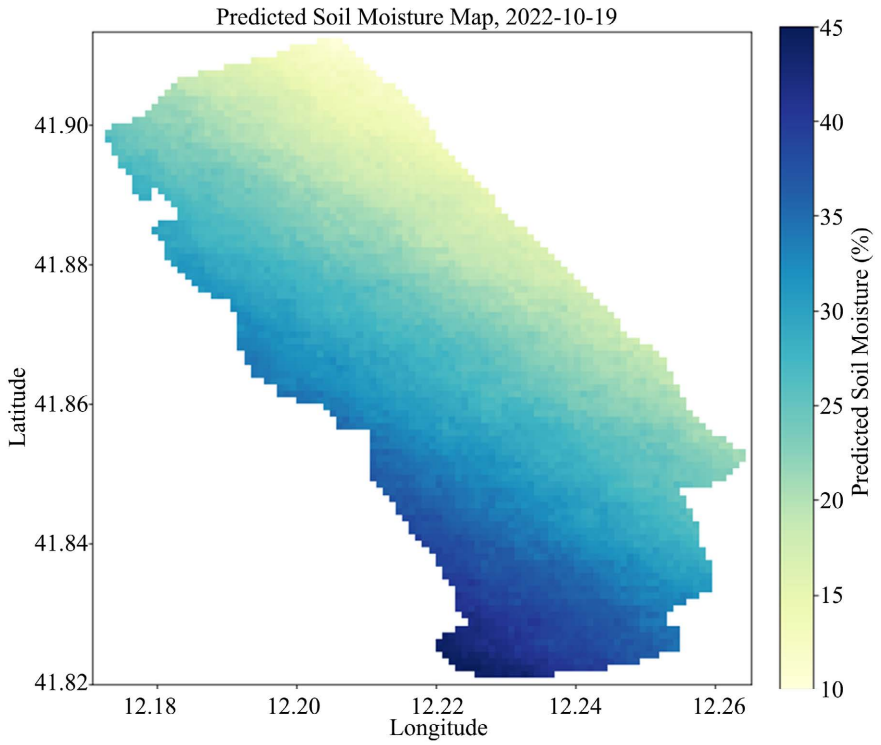


Figure 11. Predicted surface soil moisture map using ANN.

polarization, and VV (Vertical-Vertical) polarization were identified as critical variables influencing soil moisture. These factors are important in in soil moisture dynamics [44] [45] [46] [47].

Notably, this study emphasizes the importance of taking these variables into account when developing soil moisture prediction models, as they are strongly correlated with variations in soil moisture. This knowledge is useful for local agricultural practices, hydrological studies, and environmental management because it allows for more accurate forecasting and informed decision-making.

Building upon the foundations laid by previous studies such as [48] [49] [50], which highlighted the importance of accurate soil moisture estimation in environmental and agricultural contexts, this research expands our understanding of soil moisture dynamics. The comparison with Random Forest (Rf) algorithms, which achieved an R^2 of 0.89, underscores the competitive performance of the ANN model. This achievement suggests that the integration of active microwave, meteorological data, and multispectral data through preprocessing and fusion techniques can be a powerful approach for soil moisture estimation, echoing the findings of [51] who also advocated for multi-sensor data fusion techniques.

4. Conclusions and Recommendations

The culmination of this research has yielded significant results with profound implications for our comprehension of soil moisture dynamics and its applications in the study area. The Artificial Neural Network (ANN) model developed

in this study exhibited an impressive performance, achieving a coefficient of determination (R^2) of 0.71 and a correlation coefficient (R) of 0.84. These results not only underscore the model's capability to accurately estimate surface soil moisture but also emphasize its suitability for the diverse land cover types, including agricultural fields, natural vegetation, and urban areas.

Future studies should focus on the temporal analysis of soil moisture dynamics using L-band SAR data such as SAOCOM, providing insights into seasonal variations and their impacts. Additionally, research could explore L-band SAR's application in agriculture for crop health assessment and irrigation management, especially in the context of Maccarese's diverse land cover types. Understanding how impervious surfaces in urban areas affect soil moisture estimation using L-band SAR is crucial for urban planning and water management. Calibration and validation with ground-based measurements, integration with climate models, and comparative analyses with other radar frequencies should be pursued to ensure data accuracy and utility. The potential benefits of data fusion with various sensors and advanced machine learning algorithms should also be investigated, along with sensitivity analyses considering factors like vegetation, soil types, and topography, ultimately advancing the use of L-band SAR for soil moisture estimation in the region.

Acknowledgements

The author wishes to thank the La Tuscia University Italy for providing us with the *in situ* soil moisture data which aided in the completion of this work.

Data Availability

The data shall be available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Colliander, A., *et al.* (2017) Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI during SMAPVEX15. *IEEE Geoscience and Remote Sensing Letters*, **14**, 2107-2111. https://doi.org/10.1109/LGRS.2017.2753203

[2] Al-Bakri, J., Suleiman, A., Abdulla, F. and Ayad, J. (2011) Potential Impact of Climate Change on Rainfed Agriculture of a Semi-Arid Basin in Jordan. *Physics and Chemistry of the Earth, Parts A/B/C*, **36**, 125-134. https://doi.org/10.1016/j.pce.2010.06.001

[3] Wagner, W., *et al.* (2013) The ASCAT Soil Moisture Product: A Review of Its Specifications, Validation Results, and Emerging Applications. *Meteorologische Zeitschrift*, **22**, 5-33. https://doi.org/10.1127/0941-2948/2013/0399

[4] Huang, C., Chen, W., Li, Y., Shen, H. and Li, X. (2016) Assimilating Multi-Source Data into Land Surface Model to Simultaneously Improve Estimations of Soil Mois-

ture, Soil Temperature, and Surface Turbulent Fluxes in Irrigated Fields. *Agricultural and Forest Meteorology*, **230**-**231**, 142-156. https://doi.org/10.1016/j.agrformet.2016.03.013

[5] Liu, Y., Yang, Y., Jing, W. and Yue, X. (2018) Comparison of Different Machine Learning Approaches for Monthly Satellite-Based Soil Moisture Downscaling over Northeast China. *Remote Sensing*, **10**, Article No. 1. https://doi.org/10.3390/rs10010031

[6] Kang, C.S., *et al.* (2021) Global Soil Moisture Retrievals from the Chinese FY-3D Microwave Radiation Imager. *IEEE Transactions on Geoscience and Remote Sensing*, **59**, 4018-4032. https://doi.org/10.1109/TGRS.2020.3019408

[7] Chaudhary, S.K., *et al.* (2022) Machine Learning Algorithms for Soil Moisture Estimation Using Sentinel-1: Model Development and Implementation. *Advances in Space Research*, **69**, 1799-1812. https://doi.org/10.1016/j.asr.2021.08.022

[8] Kumar, S.V., Dirmeyer, P.A., Peters-Lidard, C.D., Bindlish, R. and Bolten, J. (2018) Information Theoretic Evaluation of Satellite Soil Moisture Retrievals. *Remote Sensing of Environment*, **204**, 392-400. https://doi.org/10.1016/j.rse.2017.10.016

[9] Panciera, R., *et al.* (2008) The NAFE'05/CoSMOS Data Set: Toward SMOS Soil Moisture Retrieval, Downscaling, and Assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, 736-745. https://doi.org/10.1109/TGRS.2007.915403

[10] Gaur, N. and Mohanty, B.P. (2013) Evolution of Physical Controls for Soil Moisture in Humid and Subhumid Watersheds. *Water Resources Research*, **49**, 1244-1258. https://doi.org/10.1002/wrcr.20069

[11] Attema, E.P.W. and Ulaby, F.T. (1978) Vegetation Modeled as a Water Cloud. *Radio Science*, **13**, 357-364. https://doi.org/10.1029/RS013i002p00357

[12] Dubois, P.C., van Zyl, J. and Engman, T. (1995) Measuring Soil Moisture with Imaging Radars. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 915-926. https://doi.org/10.1109/36.406677

[13] Fung, A.K., Li, Z. and Chen, K.S. (1992) Backscattering from a Randomly Rough Dielectric Surface. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 356-369. https://doi.org/10.1109/36.134085

[14] Oh, Y., Sarabandi, K. and Ulaby, F.T. (2002) Semi-Empirical Model of the Ensemble-Averaged Differential Mueller Matrix for Microwave Backscattering from Bare Soil Surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 1348-1355. https://doi.org/10.1109/TGRS.2002.800232

[15] Topp, G.C., Davis, J.L. and Annan, A.P. (1980) Electromagnetic Determination of Soil Water Content: Measurements in Coaxial Transmission Lines. *Water Resources Research*, **16**, 574-582. https://doi.org/10.1029/WR016i003p00574

[16] Baghdadi, N. and Zribi, M. (2006) Evaluation of Radar Backscatter Models IEM, OH and Dubois Using Experimental Observations. *International Journal of Remote Sensing*, **27**, 3831-3852. https://doi.org/10.1080/01431160600658123

[17] Bindlish, R. and Barros, A.P. (2000) Multifrequency Soil Moisture Inversion from SAR Measurements with the Use of IEM. *Remote Sensing of Environment*, **71**, 67-88. https://doi.org/10.1016/S0034-4257(99)00065-6

[18] Choker, M., *et al.* (2017) Evaluation of the Oh, Dubois and IEM Backscatter Models Using a Large Dataset of SAR Data and Experimental Soil Measurements. *Water*, **9**, Article No. 1. https://doi.org/10.3390/w9010038

[19] Dave, R., Kumar, G., Pandey, D.Kr., Khan, A. and Bhattacharya, B. (2021) Evaluation of Modified Dubois Model for Estimating Surface Soil Moisture Using Dual

Polarization RISAT-1 C-Band SAR Data. *Geocarto International*, **36**, 1459-1469. https://doi.org/10.1080/10106049.2019.1655801

[20] Kweon, S.-K. and Oh, Y. (2014) Estimation of Soil Moisture and Surface Roughness from Single-Polarized Radar Data for Bare Soil Surface and Comparison with Dual- and Quad-Polarization Cases. *IEEE Transactions on Geoscience and Remote Sensing*, **52**, 4056-4064. https://doi.org/10.1109/TGRS.2013.2279183

[21] Mirsoleimani, H.R., Sahebi, M.R., Baghdadi, N. and El Hajj, M. (2019) Bare Soil Surface Moisture Retrieval from Sentinel-1 SAR Data Based on the Calibrated IEM and Dubois Models Using Neural Networks. *Sensors*, **19**, Article No. 14. https://doi.org/10.3390/s19143209

[22] Walker, J.P., Troch, P.A., Mancini, M., Willgoose, G.R. and Kalma, J.D. (1997) Profile Soil Moisture Estimation Using the Modified IEM. 1997 *IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing—A Scientific Vision for Sustainable Development*, Vol. 3, 1263-1265.

[23] Weiß, T., Ramsauer, T., Löw, A. and Marzahn, P. (2020) Evaluation of Different Radiative Transfer Models for Microwave Backscatter Estimation of Wheat Fields. *Remote Sensing*, **12**, Article No. 18. https://doi.org/10.3390/rs12183037

[24] Zribi, M. and Dechambre, M. (2003) A New Empirical Model to Retrieve Soil Moisture and Roughness from C-Band Radar Data. *Remote Sensing of Environment*, **84**, 42-52. https://doi.org/10.1016/S0034-4257(02)00069-X

[25] Singh, A., Gaurav, K., Rai, A.K. and Beg, Z. (2021) Machine Learning to Estimate Surface Roughness from Satellite Images. *Remote Sensing*, **13**, Article No. 19. https://doi.org/10.3390/rs13193794

[26] Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L. and Xue, X. (2019) Research on Soil Moisture Prediction Model Based on Deep Learning. *PLOS ONE*, **14**, e0214508. https://doi.org/10.1371/journal.pone.0214508

[27] Suk Lee, C., Sohn, E., Park, J.D. and Jang, J.-D. (2019) Estimation of Soil Moisture Using Deep Learning Based on Satellite Data: A Case Study of South Korea. *GIScience & Remote Sensing*, **56**, 43-67. https://doi.org/10.1080/15481603.2018.1489943

[28] Yuan, Q., Xu, H., Li, T., Shen, H. and Zhang, L. (2020) Estimating Surface Soil Moisture from Satellite Observations Using a Generalized Regression Neural Network Trained on Sparse Ground-Based Measurements in the Continental U.S. *Journal of Hydrology*, **580**, Article ID: 124351. https://doi.org/10.1016/j.jhydrol.2019.124351

[29] Del Frate, F., Ferrazzoli, P. and Schiavon, G. (2003) Retrieving Soil Moisture and Agricultural Variables by Microwave Radiometry Using Neural Networks. *Remote Sensing of Environment*, **84**, 174-183. https://doi.org/10.1016/S0034-4257(02)00105-0

[30] Elshorbagy, A. and Parasuraman, K. (2008) On the Relevance of Using Artificial Neural Networks for Estimating Soil Moisture Content. *Journal of Hydrology*, **362**, 1-18. https://doi.org/10.1016/j.jhydrol.2008.08.012

[31] Santi, E., *et al.* (2020) Soil Moisture and Forest Biomass Retrieval on a Global Scale by Using CyGNSS Data and Artificial Neural Networks. 2020 *IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, 26 September-2 October 2020, 5905-5908. https://doi.org/10.1109/IGARSS39084.2020.9323896

[32] Ahmad, S., Kalra, A. and Stephen, H. (2010) Estimating Soil Moisture Using Remote Sensing Data: A Machine Learning Approach. *Advances in Water Resources*, **33**, 69-80. https://doi.org/10.1016/j.advwatres.2009.10.008

[33] Santi, E., Paloscia, S., Pettinato, S. and Fontanelli, G. (2016) Application of Artificial

Neural Networks for the Soil Moisture Retrieval from Active and Passive Microwave Spaceborne Sensors. *International Journal of Applied Earth Observation and Geoinformation*, **48**, 61-73. https://doi.org/10.1016/j.jag.2015.08.002

[34] Jensen, J.R. (2005) Introductory Image Processing: A Remote Sensing Perspective. Pearson Prentice-Hall, Upper Saddle River.

[35] Ulaby, F.T., Moore, R.K. and Fung, A.K. (1981) Microwave Remote Sensing: Active and Passive. Volume 1. Microwave Remote Sensing Fundamentals and Radiometry. https://ntrs.nasa.gov/citations/19820039342

[36] Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A. and Biswas, A. (2020) Comparison of Spectral and Spatial-Based Approaches for Mapping the Local Variation of Soil Moisture in a Semi-Arid Mountainous Area. *Science of the Total Environment*, **724**, Article ID: 138319. https://doi.org/10.1016/j.scitotenv.2020.138319

[37] Greifeneder, F., *et al.* (2018) The Added Value of the VH/VV Polarization-Ratio for Global Soil Moisture Estimations from Scatterometer Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **11**, 3668-3679. https://doi.org/10.1109/JSTARS.2018.2865185

[38] Ullmann, T., Jagdhuber, T., Hoffmeister, D., May, S.M., Baumhauer, R. and Bubenzer, O. (2023) Exploring Sentinel-1 Backscatter Time Series over the Atacama Desert (Chile) for Seasonal Dynamics of Surface Soil Moisture. *Remote Sensing of Environment*, **285**, Article ID: 113413. https://doi.org/10.1016/j.rse.2022.113413

[39] Baghdadi, N., Zribi, M., Loumagne, C., Ansart, P. and 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**, 4370-4379. https://doi.org/10.1016/j.rse.2008.08.004

[40] Liu, P.-W., Judge, J., DeRoo, R.D., England, A.W., Bongiovanni, T. and Luke, A. (2016) Dominant Backscattering Mechanisms at L-Band during Dynamic Soil Moisture Conditions for Sandy Soils. *Remote Sensing of Environment*, **178**, 104-112. https://doi.org/10.1016/j.rse.2016.02.062

[41] Wagner, W., *et al.* (2022) Widespread Occurrence of Anomalous C-Band Backscatter Signals in Arid Environments Caused by Subsurface Scattering. *Remote Sensing of Environment*, **276**, Article ID: 113025. https://doi.org/10.1016/j.rse.2022.113025

[42] Holtgrave, A.-K., Förster, M., Greifeneder, F., Notarnicola, C. and Kleinschmit, B. (2018) Estimation of Soil Moisture in Vegetation-Covered Floodplains with Sentinel-1 SAR Data Using Support Vector Regression. *PFG*, **86**, 85-101. https://doi.org/10.1007/s41064-018-0045-4

[43] Sadeghi, M., Babaeian, E., Tuller, M. and Jones, S.B. (2017) The Optical Trapezoid Model: A Novel Approach to Remote Sensing of Soil Moisture Applied to Sentinel-2 and Landsat-8 Observations. *Remote Sensing of Environment*, **198**, 52-68. https://doi.org/10.1016/j.rse.2017.05.041

[44] Saran, S., Sterk, G., Nair, R. and Chatterjee, R.S. (2014) Estimation of Near Surface Soil Moisture in a Sloping Terrain of a Himalayan Watershed Using ENVISAT ASAR Multi-Incidence Angle Alternate Polarisation Data. *Hydrological Processes*, **28**, 895-904. https://doi.org/10.1002/hyp.9632

[45] Ghasemloo, N., Matkan, A.A., Alimohammadi, A., Aghighi, H. and Mirbagheri, B. (2022) Estimating the Agricultural Farm Soil Moisture Using Spectral Indices of Landsat 8, and Sentinel-1, and Artificial Neural Networks. *Journal of Geovisualization and Spatial Analysis*, **6**, Article No. 19. https://doi.org/10.1007/s41651-022-00110-4

[46] Lin, R., *et al.* (2022) Improved Surface Soil Moisture Estimation Model in

Semi-Arid Regions Using the Vegetation Red-Edge Band Sensitive to Plant Growth. *Atmosphere*, **13**, Article No. 6. https://doi.org/10.3390/atmos13060930

[47] Shakya, A.K., Ramola, A. and Vidyarthi, A. (2023) Integrated Modelling of Soil Moisture by Evaluating Backscattering Models Dubois, Oh and IoT Sensor Development for Field Moisture Estimation. *Modeling Earth Systems and Environment*, **9**, 3381-3402. https://doi.org/10.1007/s40808-023-01693-7

[48] Stevanato, L., *et al.* (2019) A Novel Cosmic-Ray Neutron Sensor for Soil Moisture Estimation over Large Areas. *Agriculture*, **9**, Article No. 202. https://doi.org/10.3390/agriculture9090202

[49] Petropoulos, G.P., Ireland, G. and Barrett, B. (2015) Surface Soil Moisture Retrievals from Remote Sensing: Current Status, Products & Future Trends. *Physics and Chemistry of the Earth, Parts A/B/C*, **83**-**84**, 36-56. https://doi.org/10.1016/j.pce.2015.02.009

[50] West, H., Quinn, N. and Horswell, M. (2019) Remote Sensing for Drought Monitoring & Impact Assessment: Progress, Past Challenges and Future Opportunities. *Remote Sensing of Environment*, **232**, Article ID: 111291. https://doi.org/10.1016/j.rse.2019.111291

[51] Wang, B., Li, S., Mu, J., Hao, X., Zhu, W. and Hu, J. (2022) Research Advancements in Key Technologies for Space-Based Situational Awareness. *Space: Science & Technology*, **2022**, Article ID: 9802793. https://doi.org/10.34133/2022/9802793