




Article

Application of Multi-Source Remote Sensing Data and Machine Learning for Surface Soil Moisture Mapping in Temperate Forests of Central Japan

Kyaw Win ¹, Tamotsu Sato ^{1,2} and Satoshi Tsuyuki ^{1,*}

¹ Department of Global Agricultural Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan; kyawwin.mte@gmail.com or kyaw-win073@g.ecc.u-tokyo.ac.jp (K.W.); satoo@affrc.go.jp (T.S.)

² Forestry and Forest Products Research Institute, Tsukuba 305-8687, Japan

* Correspondence: tsuyuki@fr.a.u-tokyo.ac.jp

Abstract: Surface soil moisture (SSM) is a key parameter for land surface hydrological processes. In recent years, satellite remote sensing images have been widely used for SSM estimation, and many methods based on satellite-derived spectral indices have also been used to estimate the SSM content in various climatic conditions and geographic locations. However, achieving an accurate estimation of SSM content at a high spatial resolution remains a challenge. Therefore, improving the precision of SSM estimation through the synergies of multi-source remote sensing data has become imperative, particularly for informing forest management practices. In this study, the integration of multi-source remote sensing data with random forest and support vector machine models was conducted using Google Earth Engine in order to estimate the SSM content and develop SSM maps for temperate forests in central Japan. The synergy of Sentinel-2 and terrain factors, such as elevation, slope, aspect, slope steepness, and valley depth, with the random forest model provided the most suitable approach for SSM estimation, yielding the highest accuracy values (overall accuracy for testing = 91.80%, Kappa = 87.18%, $r = 0.98$) for the temperate forests of central Japan. This finding provides more valuable information for SSM mapping, which shows promise for precision forestry applications.

Keywords: machine learning; remote sensing; surface soil moisture; temperate forests



Citation: Win, K.; Sato, T.; Tsuyuki, S. Application of Multi-Source Remote Sensing Data and Machine Learning for Surface Soil Moisture Mapping in Temperate Forests of Central Japan. *Information* **2024**, *15*, 485. <https://doi.org/10.3390/info15080485>

Academic Editor: Marco Leo

Received: 1 July 2024

Revised: 6 August 2024

Accepted: 13 August 2024

Published: 15 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Surface soil moisture (SSM) serves as a key component of terrestrial ecosystems [1,2] and plays a critical role in regulating various hydrological, ecological, and climatic processes [3–7]. SSM typically refers to the water content of the topsoil layer (~5–15 cm) [8]. Understanding spatial and temporal variations in SSM content is essential for modulating land–atmosphere interactions [9,10], agricultural production [11,12], evaluating forest ecosystem health, assessing forest productivity, forecasting evapotranspiration [13,14], monitoring drought [15], managing water resources [16–18], and mitigating the impacts of climate change [3]. In temperate forests, where intricate interactions take place between SSM, vegetation dynamics, and topographic features, accurate monitoring of SSM is particularly important. Therefore, in the forestry field, SSM estimation with high spatial resolution is urgently required for practical applications, as soil texture and structure, vegetation coverage, and topography lead to high spatial variability in soil moisture [19,20].

Remote sensing techniques have recently emerged as powerful tools for accurate monitoring and assessment of SSM content over large spatial scales with high resolution [21]. Traditional field-based measurements (in other words, ground-truth data collection) provide valuable insights into small-scale estimation for SSM but are limited in terms of spatial coverage and temporal frequency of SSM estimation at large scales [22,23]; moreover, this measurement is labor-intensive [3,24,25], time-consuming [26], and accurate only at the

point of measurement [27]. In contrast, the application of remote sensing techniques makes it possible to observe SSM at large scales with sufficient temporal continuity, even in difficult and inaccessible locations [28–30]. Through integrating remote sensing data and field-based measurements (ground-truth data), researchers can obtain valuable insights into the large-scale spatial distribution of SSM content estimation [31].

In recent years, the integration of machine learning methods, such as neural networks, random forest (RF), support vector machine (SVM), and decision trees with remote sensing data and ground-truth data, has also been increasingly applied for SSM content estimation in various ecosystems, including in agriculture [16], irrigation management [32], landslide monitoring [33], and flood risk assessment [34]. Machine learning refers to the automatic or semi-automatic investigation and analysis of large datasets to identify meaningful correlations, patterns, and rules among data [35]. Machine learning approaches are now commonly used, having been successfully applied to predicting SSM using remotely sensed data [16,20]. The development of RF and SVM models has also led to renewed attempts to estimate SSM from remote sensing data [36,37].

Most studies have reported the application of Sentinel-1 for soil moisture estimates. For instance, Ettalbi et al. [38] evaluated soil moisture retrieval in bare agricultural areas using Sentinel-1 images and neural networks. Zhao et al. [20] addressed the scaling and mapping of soil moisture by using remote-sensing-based auxiliary variables and an RF model. Exploring a different approach, Nativel et al. [39] proposed a methodology that combines their improved change detection index with an artificial neural network (ANN) trained on Sentinel-1 and Sentinel-2 data, as well as data from the International Soil Moisture Network, to enhance the accuracy of soil moisture estimation at a 1 km scale. Chung et al. [40] trained an ANN model on Sentinel-1 SAR imagery to estimate soil moisture content. Hamze et al. [22] focused on integrating L-band-derived soil roughness into C-band synthetic aperture radar (SAR) data to improve soil moisture estimation. Each method has its own advantages and disadvantages, and it is necessary to meet the needs of practical applications [23].

To the best of our knowledge, despite the advancements in remote-sensing-based SSM estimation across several studies [41,42], limited studies have conducted SSM estimation in the forestry field. To address this challenge, this study adopts synergistic approaches to recommend the most suitable approach for SSM estimation through integrating ground-truth data, multi-source remote sensing data, and two machine learning models: random forest (RF) and support vector machine (SVM). Attarzadeh et al. [26] also suggested that the integration of data acquired with optical sensors and SAR data may provide useful information for reducing ambiguity due to the presence of vegetation. Therefore, this study explored five synergies through integrating Sentinel-1 SAR GRD: C-band synthetic aperture radar ground range detected; log scaling (S-1); Harmonized Sentinel-2 MSI: Multi-Spectral Instrument; Level-2A (S-2); and terrain information (elevation, slope, aspect, slope steepness, and valley depth), along with five spectral indices for use in cloud-based platforms, such as Google Earth Engine (GEE).

This study aimed to estimate SSM using multi-source remote sensing data in temperate forests with a variety of forest types in Central Japan. The specific objectives of this study were as follows:

- To evaluate the accuracy of SSM classification by RF and SVM models;
- To compare the model performance for each synergy;
- To derive the SSM maps of the study site based on five synergies.

2. Materials and Methods

2.1. Study Area

The study area was located in the University of Tokyo Chiba Forest (UTCBF) in Chiba Prefecture, central Japan (Figure 1). The UTCBF was established in 1894, comprising different forest types (2225 hectares), including a coniferous natural forest covering 387 hectares, a broad-leaved natural forest covering 949 hectares, plantations extend-

ing over 825 hectares, a 56-hectare exhibition forest, and nurseries, seed orchards, and other areas covering 8 hectares. The study area lies within the geographic coordinates of 35°08'25" to 35°12'51" N and 140°05'33" to 140°10'10" E. Altitudes within the study area vary from 50 to 370 m above sea level. The topography of the study area is characterized by complex and varied contour patterns, with generally steep slopes. Geologically, the area is primarily composed of brown forest soil, underlain by Neogene marine deposits overlain by Quaternary non-marine deposits. The bedrock composition comprises sandstone, conglomerates, mudstone, and tuff [43]. Climatically, the study area has a mean temperature of approximately 4 °C in midwinter and 25 °C in midsummer, with an annual mean temperature of around 14 °C. The annual precipitation averages approximately 2200 mm.

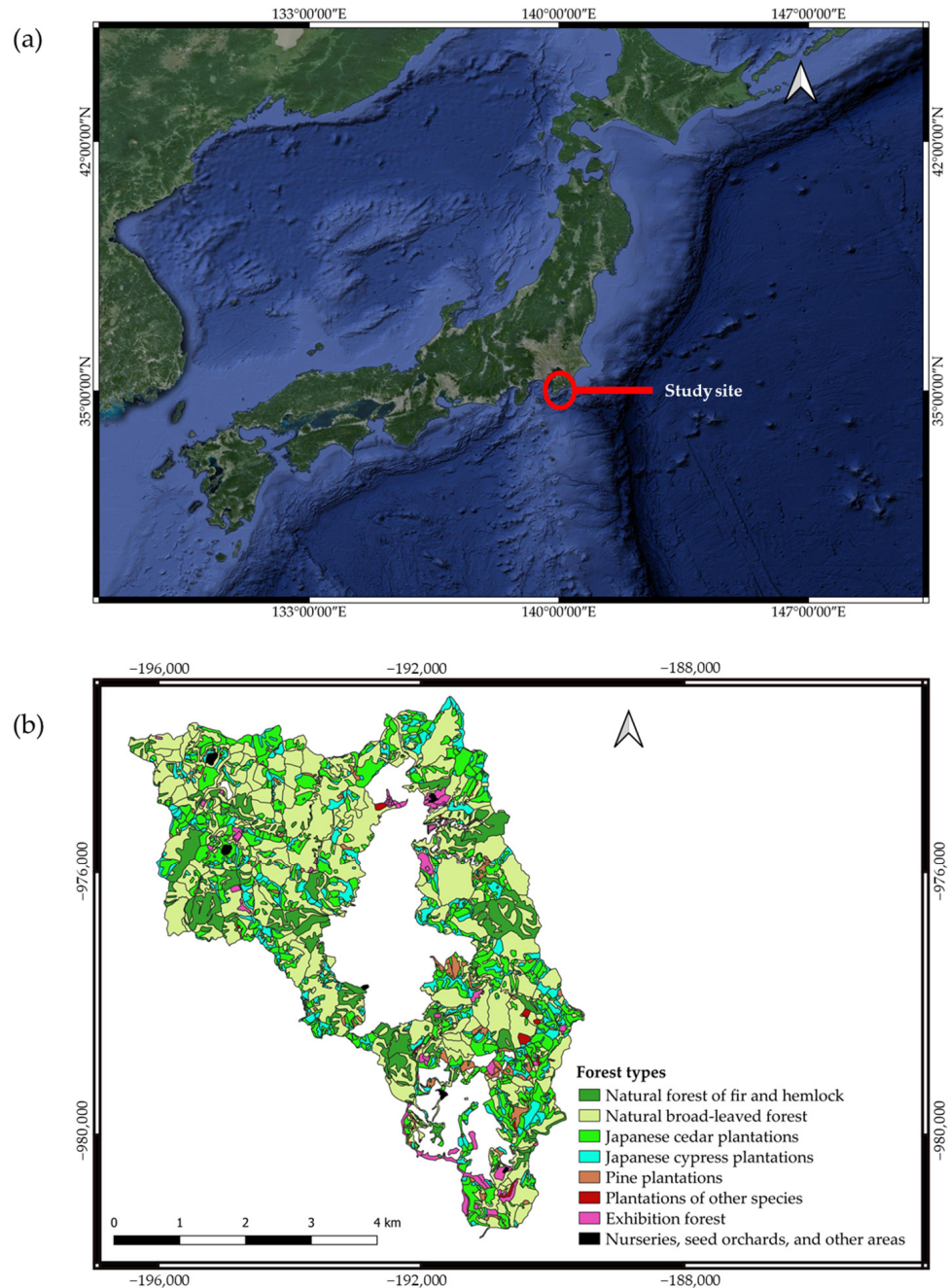


Figure 1. Cont.

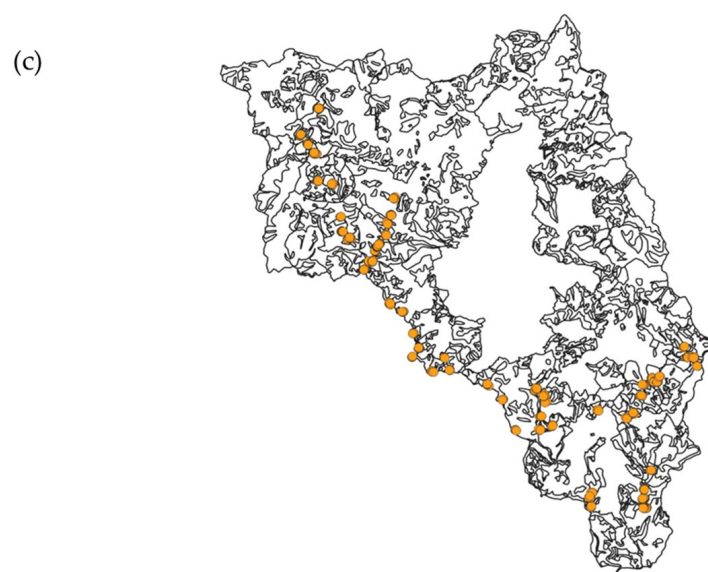


Figure 1. Overview of the study site: (a) Map of Japan. (b) Map of the University of Tokyo Chiba Forest. (c) Sample locations (the symbol of each point represents each sample).

2.2. Field Data Collection

In this study, non-destructive field measurements were conducted using a FieldScout Time-Domain Reflectometry Meter equipped with a 4.8-inch (12 cm) rod (FieldScout TDR 350; Spectrum Technologies Inc., Haltom City, TX, USA). This meter measures the percentage of volumetric water content (% VWC) and can also be entirely automated, providing accurate SSM data [44] of different forest types. The TDR measurements also achieve an accuracy level of approximately ± 2.5 vol.% [45]. The specifications of this meter are shown in Appendix A. Field measurements can be viewed with the FieldScout Mobile app, which maps out the sample points based on the recorded location coordinates.

In addition, the experimental design of SSM sampling is based on the land use or soil conditions of the study area. We randomly set up a circular sampling plot with a plot size of 0.03 ha or 0.1 ha per plot, using a stratified random sampling design. There were variations associated with the SSM levels within the plot due to elevation differences. Variations in SSM might be influenced by elevation [46]. Therefore, five to ten measurements of the SSM samples were taken in each plot. The variations in SSM measurements for each plot ranged from 0.4 to 1.8% VWC in the Japanese cypress plantations, 0.3 to 4.1% VWC in the Japanese cedar plantations, 0.2 to 10.2% VWC in the natural broad-leaved forest, 0.8 to 2.9% VWC in the natural forest of fir and hemlock, 1.7 to 8.3% VWC in the exhibition forest, 4% VWC in the plantation of other species, and 13.5% VWC in the pine plantations. The collections of SSM samples were determined by considering various protocols, including the different forest types, seasonal period of data collection, and variations in the elevation of each plot, in order to obtain the most reliable SSM data.

Adab et al. [3] collected 58 ground-truth data for six typical land-use systems, including forested areas (less than 1%), forested areas (1–5%), forested areas (5–25%), irrigated cultivated areas, rainfed cultivated areas, and rangeland areas, in order to estimate SSM using remote sensing data and machine learning. In this study, SSM ground-truth data, with a total of 375 spatially different sample points, were proportionally collected across the different forest types based on the forested areas (Table 1; Figure 1b). The large number of ground-truth data minimized the potential spatial biases in the ground-truth data collection. Moreover, the ground-truth data points were distributed across the study area (Figure 1c), ensuring coverage of the entire geographic extent. This approach was ensured to be representative of the entire study area.

Table 1. SSM ground-truth data points from field data collection.

Forest Types	Areas (ha)	Number of Ground Truth Points	
Natural forest of fir and hemlock	387	40	
Plantations	825	Pine plantations	5
		Japanese cedar plantations	95
		Japanese cypress plantations	90
		Plantations of other species	5
Natural broad-leaved forest	949	120	
Exhibition forest	56	20	
Nurseries, seed orchards, and other areas	8	0	
Total	2225	375	

Note: The information about the areas of forest types was obtained from the official website of UTCBF "https://www.uf.a.u-tokyo.ac.jp/english/our_forests/UTCBF.html" (accessed on 19 June 2024)".

We measured these ground-truth data from May to September 2023 and in March 2024. Before measuring the SSM ground-truth data, we removed the litter, rocks, and other debris from the designated area. The seasonal cycle of precipitation significantly influences soil moisture [47]. Therefore, we did not collect the SSM ground-truth data during the days on which it rained, so as to avoid large variations. Additionally, we conducted multiple measurements, at least three times, for each sample point across different seasons and times of day. Finally, we used the most stable SSM data for each sample point across different forest types to mitigate the impact of temporal variability on SSM estimation. The location of each measurement was recorded using a GPSMAP (GPSMAP 64s, Garmin Ltd., St. Olathe, KS, USA) device.

Among different forest types (Figure 1b), the values of SSM ground-truth data ranged from 0.6 to 71.9% VWC (Figure 2). We classified these values as three levels of SSM ground-truth data: (1) <20% VWC, (2) 20–40% VWC, and (3) >40% VWC. We categorized these three SSM levels based on the ranges of SSM ground-truth data for each forest type. These classifications may be helpful for practical considerations in sustainable forest management. For example, areas with SSM levels below 20% VWC may be indicative of drought stress or water scarcity. SSM levels below 20% are mostly found in natural broad-leaved forests and natural forests of fir or hemlock (Figure 2). Conversely, areas with SSM levels above 40% VWC may indicate saturated or poorly drained soils, which can affect the vegetation composition and land-use suitability. At that level, the establishment of Japanese cedar plantations that prefer SSM levels above 40% VWC might be successful (Figure 2). Through understanding the SSM level preferences of each forest type, forest managers might be able to manage the forests sustainably.

2.3. Remote Sensing Data

2.3.1. Sentinel-1

Satellite images from Sentinel-1 C-band synthetic aperture radar ground range detected (SAR GRD), log scaling (S-1), provided by the European Union/ESA/Copernicus, penetrate clouds with a spatial resolution of 10, 25, or 40 m and a temporal resolution of 6–12 days at a central frequency of 5.405 GHz. This study extracted S-1 images considering the single co-polarization, vertical transmit/vertical receive (VV) and dual-band cross-polarization, vertical transmit/horizontal receive (VH) to integrate with Sentinel-2 data. Fifteen S-1 images with 10 m resolution from April to September 2023 were extracted in this study.

2.3.2. Sentinel-2

The Harmonized Sentinel-2 MSI: Multispectral Instrument, Level-2A (S-2), provided by the European Union/ESA/Copernicus, has provided atmospherically corrected surface reflectance images with resolutions of 10, 20, and 60 m since 2017. Several studies have used

S-2 together with S-1 for land monitoring studies, including the monitoring of vegetation, soil, and water cover, as well as the observation of inland waterways and coastal areas. In this study, 40 S-2 images with 10, 20, and 60 m resolution were extracted from the GEE repository “<https://developers.google.com/earth-engine/datasets> (accessed on 18 December 2023)” for the period between April 2023 and September 2023 in order to estimate SSM. The details of the spectral bands of interest for S-2 used in this study are shown in Table 2. The blue band from S-2 is sensitive to the chlorophyll and carotenoid molecules of plants’ leaves, while the green band boosts water information [16]. The red band of the spectrum is absorbed by the vegetation [16]. Moreover, green vegetation strongly reflects NIR bands and is directly related to the moisture of the soil where it is grown [16]. In addition, the SWIR band discriminates the moisture contents of the soil and vegetation [16]. The previous results also indicate that it is possible to estimate SSM (0–7.6 cm) from visible and near-infrared reflectance [20,48].

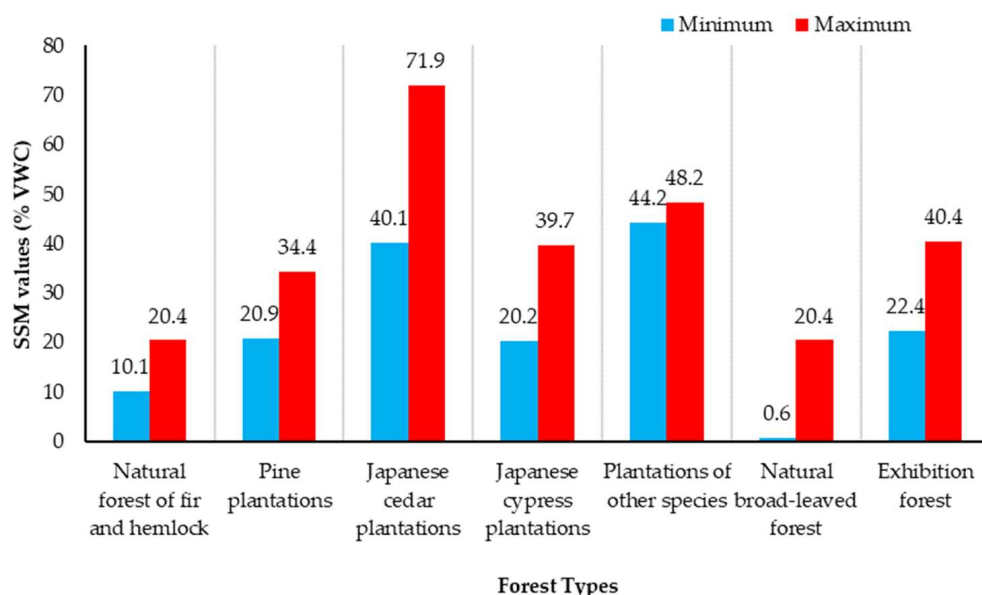


Figure 2. SSM ground-truth data values for the different forest types of the study site.

Table 2. The details of spectral bands of interest for S-2 used in this study.

Name	Pixel Size (m)	Bandwidth (nm)	Wavelength	Description
B2	10	65	496.6 nm (S2A)/492.1 nm (S2B)	Blue
B3	10	35	560 nm (S2A)/559 nm (S2B)	Green
B4	10	30	664.5 nm (S2A)/665 nm (S2B)	Red
B8	10	115	835.1 nm (S2A)/833 nm (S2B)	NIR
B11	20	90	1613.7 nm (S2A)/1610.4 nm (S2B)	SWIR1
B12	20	180	2202.4 nm (S2A)/2185.7 nm (S2B)	SWIR2
QA60	60			Cloud Mask

2.3.3. Selection of Spectral Indices

Our study selected spectral indices (SIs) derived from S-2. SIs play a crucial role in SSM classification due to their ability to capture and enhance specific characteristics of vegetation, soil, and moisture content in remote sensing data. A previous study used NDVI together with the synergetic use of S-2 and S-1 for SSM estimation [49]. Unlike that study, we evaluated the variable importance among 16 spectral indices using an RF classifier (Figure 3). Among the 16 spectral indices, the enhanced vegetation index (EVI), green normalized difference vegetation index (GNDVI), Tasseled Cap brightness (TCBrightness), Tasseled Cap wetness (TCWetness), and weighted difference vegetation index (WDVI) obviously showed the highest values (Figure 3).

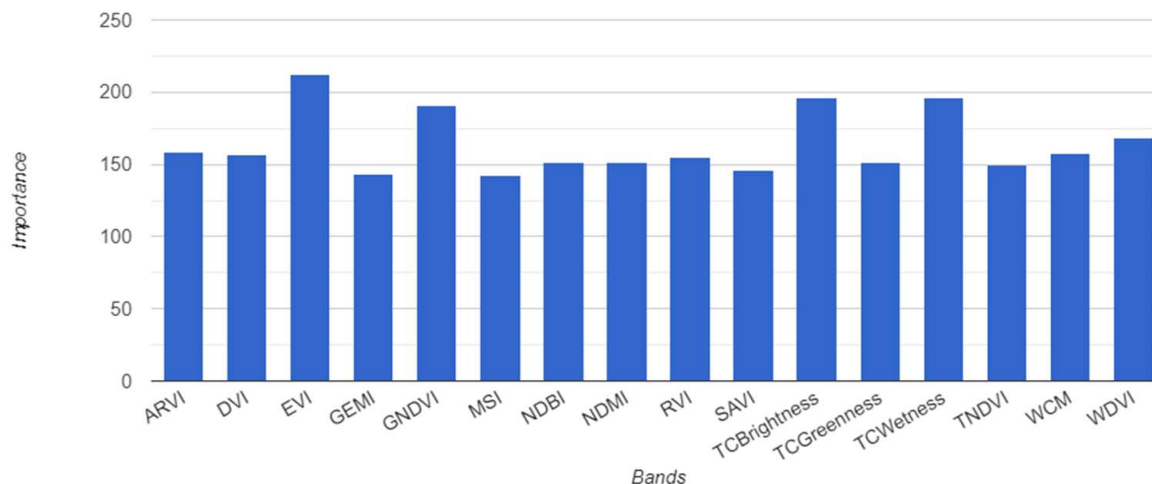


Figure 3. Result of variable importance among spectral indices.

The EVI is designed to enhance the vegetation signal through optimizing the differential response in the NIR and red bands, while the blue band is incorporated to correct atmospheric effects. The EVI corrects for some of the background noise from the soil and atmosphere, making it a more reliable index for areas with varying soil brightness or atmospheric conditions [50]. Moreover, the EVI is very effective in regions with high biomass, where the NDVI tends to saturate. Coefficients 2.5, 6.0, 7.5, and 1.0 of the EVI were used in this study. The coefficients were originally developed for the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra and Aqua satellites [51]. However, these coefficients have been widely used and validated for S-2, as well [50]. Therefore, we used the standard EVI formula (Table 3) with the mentioned coefficients for S-2. The GNDVI is more sensitive to chlorophyll content than the NDVI due to its use of the green band instead of the red band; this makes it particularly useful in monitoring crop health and vigor. As the GNDVI is highly correlated with the leaf area index (LAI) and biomass [52], it can provide indirect information about soil moisture levels, as healthier, more vigorous vegetation is typically associated with better soil moisture conditions.

TCBrightness, derived from a linear transformation of the original spectral bands, represents the overall soil and surface reflectance. It captures the variation in soil and surface reflectance [53], which can be directly related to the soil moisture content; therefore, this index is highly effective in areas with sparse vegetation where the soil surface is exposed. TCWetness is another linear transformation, emphasizing the moisture content in the soil and vegetation. Similar to TCBrightness, TCWetness is directly related to the moisture content in both soil and vegetation, making it a valuable index for estimating SSM [54]. It captures both the vegetation and soil moisture signals, providing a comprehensive measure of the overall wetness of the landscape and, therefore, is particularly useful in heterogeneous landscapes where both vegetation and soil moisture content vary [55]. The equations stated in Table 3 were used for the TCBrightness and TCWetness. IDB, the index database from the University of Bonn “<https://www.indexdatabase.de> (accessed on 29 July 2024)”, also published these equations to use for S-2 [55]. The WdVI, which is essential in estimating vegetation conditions on soils with varying reflectance properties, offers a good correction for soil background [56]. Therefore, our study finally selected these 5 spectral indices based on the highest values of variable importance (Figure 3). The other spectral indices did not obviously differ from one another in importance. The formulae for the 5 spectral indices are provided in Table 3.

Table 3. Formulae of the 5 spectral indices used in this study.

Index	Formula for Calculation	References
Enhanced Vegetation Index (EVI)	$EVI = 2.5 \times \frac{(NIR - Red)}{(NIR + 6.0 \times Red - 7.5 \times Blue + 1.0)}$	[50]
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{NIR - Green}{NIR + Green}$	[52]
Tasseled Cap Brightness (TCBrightness)	$TCBrightness = 0.3037 \times Blue + 0.2793 \times Green + 0.4743 \times Red + 0.5585 \times NIR + 0.5082 \times SWIR1 + 0.1863 \times SWIR2$	[53]
Tasseled Cap Wetness (TCWetness)	$TCWetness = 0.1509 \times Blue + 0.1973 \times Green + 0.3279 \times Red + 0.3406 \times NIR - 0.7112 \times SWIR1 - 0.4572 \times SWIR2$	[55]
Weighted Difference Vegetation Index (WDVI)	$WDVI = (NIR - 0.5 \times Red)$	[56,57]

2.3.4. Terrain Factors

Yang et al. [58] used DEM data together with soil moisture active-passive products for SSM estimation. In contrast, we used various terrain factors in this study, such as elevation, slope, slope steepness, aspect, and valley depth. To the best of our knowledge, this study is the first attempt to use various terrain factors as the input features for SSM classification of different forest types. The digital elevation model (DEM) data with 30 m resolution were obtained from the Shuttle Radar Topography Mission (SRTM) dataset provided by the US Geological Survey (USGS) [59]. The DEM data were produced during an 11-day mission in February 2000 and have been widely used in numerous studies due to their high accuracy and global coverage. Despite being collected over two decades ago, the SRTM DEM data are a more reliable source than other DEM data in hydrological application [58,60]. Therefore, the SRTM DEM data were loaded into the GEE environment, and the elevation band was then selected. Elevation directly affects climate and weather patterns, influencing soil moisture levels.

The slope and slope steepness were derived from the ee.Terrain.slope function, which represents the terrain characteristics and intensity or severity of the slope and computes the slope percentages and degrees for each pixel. Slope and slope steepness have slightly different definitions based on the specific context in which they are used. For instance, slope is the measure of the inclination or gradient of the land surface; it represents the rate of change in elevation with respect to horizontal distance, and it is typically calculated as the tangent of the angle between the land surface and the horizontal plane, often expressed as a percentage (rise over run) or in degrees. Slope steepness refers to the degree or percent to which the slope is steep and includes terrain roughness or factors affecting the perception of steepness. Slope steepness is determined by initially computing the slope. Based on the slope value, we calculated the slope steepness [61] as follows.

$$\text{Slope steepness} = 16.8 \times \sin(\theta) - 0.5, \text{ where } \theta \text{ represents slope}$$

The aspect was derived using the ee.Terrain.aspect function, which assigns a value representing the direction in degrees. This indicates the compass direction in which a slope faces; it influences sunlight exposure, which affects soil temperature and evaporation rates. The valley depth was calculated using the topographic position index (TPI), which is useful not only for understanding the accumulation and drainage of water but also for improving existing global hydrography datasets in terms of spatial coverage and representation of small streams; it compares the local mean elevation (within 250 m) to the broad mean elevation (within 2000 m). Through subtracting the broad mean from the local mean, areas that are significantly lower than their surroundings (valleys) can be identified. The greater the negative TPI value, the deeper the valley is considered to be relative to its surroundings.

2.4. Method

The processing of the images was conducted in GEE, which is a useful planetary-scale platform for analyzing and visualizing geospatial data [62]. We first applied S-2 images with 10, 20, and 60 m spatial resolution as the prime source of data for this study. Preprocessing was conducted by masking out cloudy pixels in images with more than 20% cloudy pixels, using the CFmask algorithm in GEE Code Editor “<https://code.earthengine.google.com/>” (accessed on 18 December 2023)—a cloud-based platform that provides geospatial data, tools, and computing power [63]—to detect and remove the cloud cover from the S-2 images. The algorithm used the QA60 band to identify and mask opaque and cirrus clouds, ensuring the quality of the acquired data. After filtering and masking out the cloud cover of the images based on date, location, and cloud cover percentage, we resampled all of the S-2 images to a 30 m spatial resolution using the bilinear interpolation method with the scale function in GEE. The reasons for resampling to 30 m resolution were to ensure consistency across all datasets, to capture the relevant features and patterns within the study area, and to avoid the potential for excessive data noise and variability that can accompany higher-resolution data. Moreover, GEE does have limitations in terms of handling and processing data at very high resolutions over large areas [64]. Resampling to 30 m helps mitigate these limitations, enabling more efficient processing and analysis within GEE’s environment. Then, we also resampled other remote sensing images used in this study to a 30 m resolution using the bilinear interpolation method with the scale function in GEE. After resampling all of the images, we specifically applied a temporal aggregation technique by compositing all the remote sensing images. This compositing process involves averaging the reflectance values of the selected bands and spectral indices over the period of interest. By doing so, we mitigated the short-term fluctuations and derived more stable and representative SSM estimates.

2.4.1. Synergies of Multi-Source Remote Sensing Image Combinations

Phan et al. [65] and Pratico et al. [66] found that the selection of image combinations significantly affects the classification accuracy. Other studies also proved that combining SAR and optical data allows for enhanced classification outcomes [66–71]. Moreover, the addition of spectral indices to optical data has been shown to improve classification [72,73]. These studies are especially conducted for land-use/land-cover classification and forest vegetation mapping. To the best of our knowledge, no studies have been conducted to assess the contribution of multi-source remote sensing image combinations for SSM estimation. Therefore, five synergies of integrating multi-source remote sensing images were explored to evaluate the contribution of the different band reflectance values. The datasets used for the combination of images are shown in Table 4. The purpose of these combinations was to identify the most effective image combination to achieve higher accuracy. Based on this higher accuracy, we aimed to apply the strengths of each data source to improve the reliability of SSM mapping in temperate forest ecosystems.

Table 4. Image combination protocols for five synergies.

Synergies	Image Combinations
Synergy 1	S-2 + S-1 + SI + Terrain factors
Synergy 2	S-2 + SI + Terrain factors
Synergy 3	S-2 + SI
Synergy 4	S-2 + Terrain factors
Synergy 5	S-1 + SI

2.4.2. Random Forest (RF)

Among the various machine learning methods, the RF model has demonstrated good performance, especially when compared to traditional statistical models [4]. Adab et al. [3] also showed that the RF model can outperform others for estimating SSM, and they improved the model performance with an increasing number of observed datasets. The RF

model [74] is one of the machine learning methods based on the bagging integrated learning theory [75] and the random subspace method [76]. The advantage of the RF model is that it can be robust to noise in predictors [20]. Moreover, the RF model has several major advantages over other statistical models, such as the ability to find high-dimensional nonlinear relationships using categorical and continuous predictors, and it needs no preselection of variables and only a few user-defined parameters [77,78]. In addition, the RF model has high prediction accuracy and tolerance for anomalies and noise in the data, making it relatively resistant to overfitting problems in practical applications [3,20].

The RF model is typically based on a dataset of decision trees that is determined by more than one variable. Numerous trees are composed by the algorithm, and the final predictions involve the average of the results from all the developed trees in the forest [79]. It is necessary to define three user-defined parameters a priori before running the model, i.e., the number of trees in the forest, the number of attributes for consideration at each split, and growth control, which represents the depth of individual trees and split subsets. To date, the RF model has been widely used in various applications, such as data mining [80], bioinformatics research [81], and information classification [79]; its application accuracy can be enhanced further by integrating multi-source information [20]. Therefore, five synergies of multi-source remote sensing data and ground-truth data in an RF model were applied in this study to provide a valuable tool for mapping high-resolution SSM across large areas. In this study, an RF model with an ensemble of decision trees ($n = 400$) was implemented to predict the SSM levels and develop SSM maps for each synergy. We used this RF model in GEE for classification, training, and validation purposes.

2.4.3. Support Vector Machine (SVM)

SVM refers to a class of supervised learning algorithms derived from statistical learning theory, first introduced by Vapnik [82]. SVM has been successfully applied for both classification and regression purposes [83,84]; it uses a kernel function for transforming the input data and then applying linear regression to the transformed data [85]. SVM involves two steps: (1) selecting an appropriate kernel type and setting the kernel parameter (kernel width G) and (2) specifying the penalty parameter C [86]. Therefore, in this study, a set of varying hyperparameters, with a regression cost (C) of 10, sigmoid kernel type, and radial basis function, G 0.5, was applied for the classification and prediction of SSM levels of each synergy. This model was also implemented in GEE for classification, training, and validation purposes.

2.4.4. Model Training and Testing

We divided the datasets into various training and testing sets (i.e., 50% and 50%, 60% and 40%, 70% and 30%, 80% and 20%, 90% and 10%) to ensure a comprehensive evaluation of the models' performance. Among these, 70% and 30% provided the best results for both the RF and SVM models. Therefore, the training dataset, comprising 70% of the SSM ground-truth data, was used to train both models. The remaining 30% of the SSM ground-truth data served as the testing dataset for evaluating the accuracy of each synergy. The training and testing datasets were obtained by randomly assigning a value to each data point and then filtering based on these values. Classification accuracy metrics, such as overall accuracy (OA) for training, OA for testing, and Kappa coefficient, were computed to assess the accuracy of each synergy in each model; OA represents the percentage of pixels assigned with the correct label. The Kappa coefficient shows the overall agreement between the classification and the ground-truth data; its value ranges from 0 to 1, where values close to 0 indicate no agreement between the classified results and the ground-truth data, while values close to 1 indicate agreement between the classified results and the ground-truth data.

2.4.5. Statistical Analyses

Regression metrics, such as the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE) were also used to compare the models' performance. R^2 is a statistical measure representing the proportion of variance in the dependent variable that can be explained by the independent variables; it provides an indication of how well the observed outcomes are replicated by the model based on the proportion of total variation in the outcomes explained by the model. A higher R^2 value indicates a better fit of the model to the data, signifying that a larger portion of the variance in the observed data is captured by the model. Specifically, R^2 values range from 0 to 1, where an R^2 of 1 indicates that the regression predictions perfectly fit the observed data. The RMSE measures the variation between the observed values and the values predicted by a model. A lower RMSE indicates good agreement between the estimation data and the observation data [3]. The MAE is the simplest measure of estimate accuracy, which measures the average magnitude of the errors in prediction data without considering their direction [3]; it is the absolute value of the difference between the observed and predicted values [3]. These metrics indicate the accuracy and reliability of the predicted SSM values in comparison to the SSM ground-truth data (i.e., observed values). The correlation analysis was then conducted to identify and ensure the correlation between the observed SSM ground-truth data and the predicted SSM from five synergies of each model. All of the statistical analyses were performed using R 4.3.2.

2.5. Surface Soil Moisture Classification Workflow

At first, we extracted the intended remote sensing data, as mentioned in Section 2.3. During the processing stage, all of the images were composited, and the study area was extracted from each image. All of the data were resampled to a 30 m resolution. Then, five synergies were created and processed for SSM mapping using two machine learning models, namely, RF and SVM. The flowchart of this study is shown in Figure 4. This flowchart consists of three main steps: (1) data preparation, including variable selection; (2) data splitting in machine learning; and (3) machine learning algorithms.

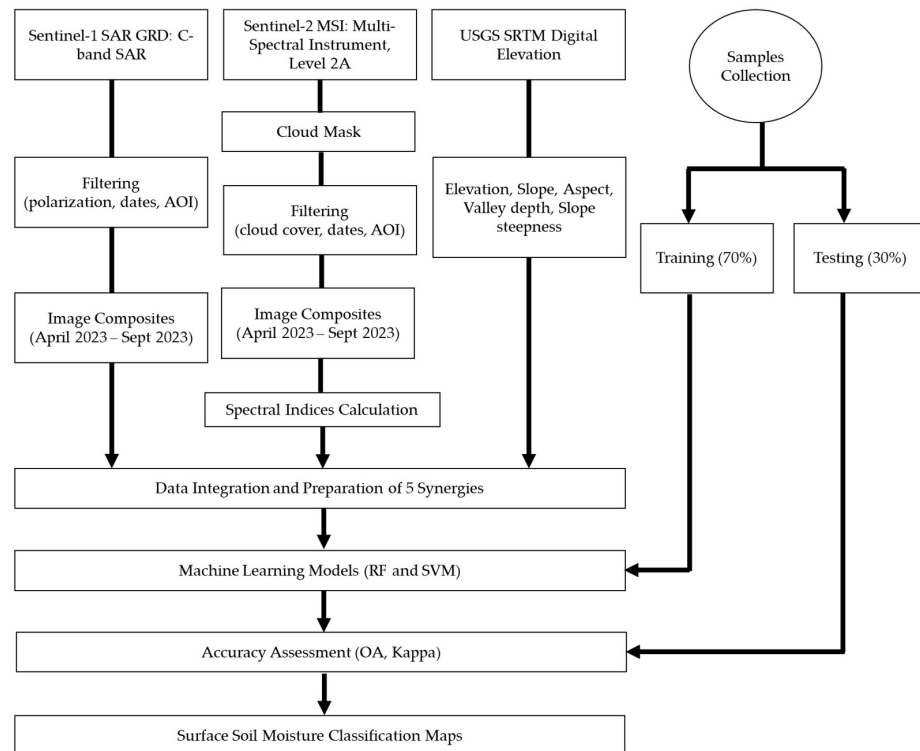


Figure 4. Flowchart of the data processing and analysis steps.

3. Results

3.1. Accuracy Evaluation of the Surface Soil Moisture Classification by RF Model

The RF model was used to evaluate the accuracy of SSM classification using multi-source remote sensing data with five synergies. The accuracy results of the RF model for these five synergies are shown in Figure 5a. Synergy 1 showed an OA of 99.24% for training and 90.18% for testing, while the Kappa coefficient was found to be 84.65%. Synergy 2 resulted in an OA of 99.25% for training and 88.89% for testing; the Kappa coefficient was slightly reduced to 83.18% compared with synergy 1. Synergy 3 also performed well, achieving an OA of 99.24% and 87.61% for training and testing, respectively, with a Kappa coefficient of 80.8%. Synergy 4 showed the highest OA for training (99.21%) and for testing (91.80%); moreover, synergy 4 showed the highest Kappa coefficient at 87.18%, indicating the strongest agreement between the observed SSM ground-truth data and predicted values. Synergy 5 achieved an OA of 99.24% for training and 83.19% for testing; the Kappa coefficient for synergy 5 was found to be 74.23%, similar to that of synergy 1.

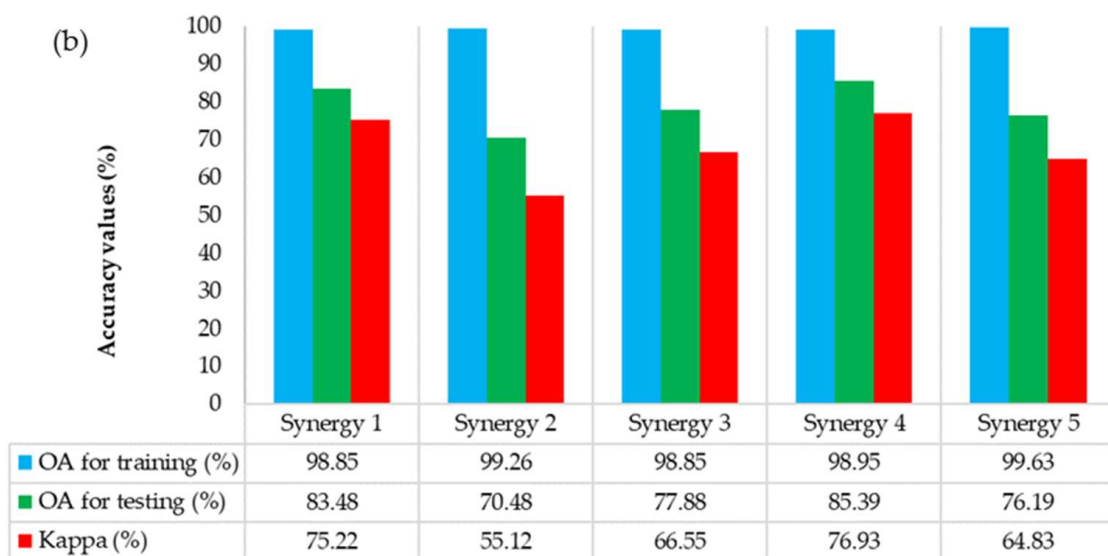
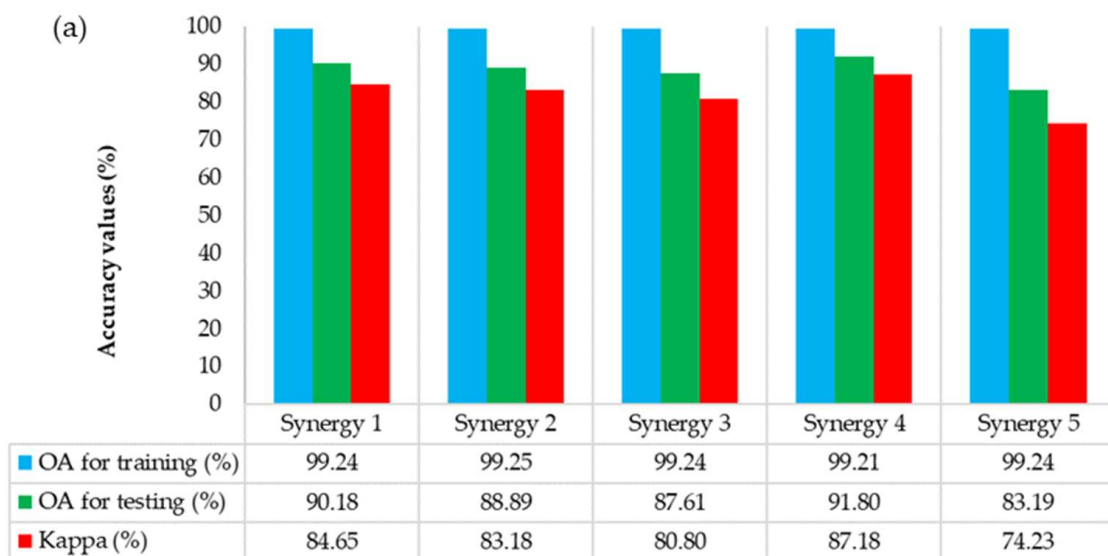


Figure 5. Accuracy results of five synergies used by (a) RF model and (b) SVM model.

In this model, the OA for the training dataset ranged from 99.21% to 99.25%, while the OA for the testing dataset ranged from 83.19% to 91.80% (Figure 5a), indicating that this model performed exceptionally well during the training and testing phases. These high levels of accuracy highlight that the RF model was able to effectively learn and capture the complex relationships between the input features and SSM levels. The Kappa coefficient ranged from 74.23% to 87.18% (Figure 5a), indicating strong agreement between the observed SSM ground-truth data and the predicted SSM values. Even the lowest value of the Kappa coefficient, observed for synergy 5, was 74.23%, suggesting strong agreement. Therefore, the Kappa coefficient values across all the synergies indicate a good to very good level of classification accuracy, with all the synergies performing well.

The high performance of synergy 4 (OA for training = 99.21%, OA for testing = 91.80%, Kappa = 87.18%) stood out as the most accurate configuration among the five synergies. This strong agreement between the observed SSM ground-truth data and predicted values highlights the robustness of synergy 4 in accurately classifying SSM levels, highlighting its potential for practical application in SSM mapping. Synergy 2, while showing the highest OA for training (99.25%), had a lower OA for testing (88.89%) and Kappa coefficient (83.18%) than synergy 4. This suggests that although the model learned the training data better in synergy 2 than in synergy 4, its generalization to new data was slightly less effective compared to synergy 4. The differences in OA for testing and Kappa between synergy 4 and the other synergies highlight the impact of data synergy on model performance. Despite the high OA for training across all synergies, the testing results highlight the importance of selecting the best combination of multi-source remote sensing data for accurate SSM classification. Therefore, the goodness of fit improved most significantly in synergy 4, followed by synergy 1, synergy 2, synergy 3, and synergy 5 (Figure 5a) based on the results of OA for testing and Kappa coefficient.

3.2. Accuracy Evaluation of the Surface Soil Moisture Classification by SVM Model

The SVM model was also applied to predict the accuracy of SSM classification using multi-source remote sensing data with the same five synergies. The accuracy results of the SVM model for those five synergies are shown in Figure 5b. Synergy 1 achieved an OA of 98.85% for training and 83.48% for testing, with a Kappa coefficient of 75.22%. Synergy 2, on the other hand, resulted in an OA of 99.26% for training and 70.48% for testing, with a notably lower Kappa coefficient of 55.12%. Synergy 3 showed an OA of 98.85% for training and 77.88% for testing, with a Kappa coefficient of 66.55%. Synergy 4 performed similarly, with an OA of 98.95% for training and 85.39% for testing, and a Kappa coefficient of 76.93%. Finally, synergy 5 achieved an OA of 99.63% for training and 76.19% for testing, with a Kappa coefficient of 64.83%, which was slightly higher than that of synergy 2.

The application of the SVM model for evaluating the accuracy of SSM classification using five different synergies (Figure 5b) also provided valuable insights into the effectiveness of various data combinations. The results indicate significant variability in model performance across the different synergies, highlighting the importance of selecting optimal data inputs. Similar to the RF model, synergy 4 also proved to be the most effective data combination for the SVM model, achieving the highest OA for testing (85.39%) and a Kappa coefficient of 76.93%, suggesting that the data sources and features included in synergy 4 provided the most informative inputs for the SVM model, leading to a high level of agreement between the observed SSM ground-truth data and the predicted values. Therefore, the strong performance of synergy 4 can be attributed to the relevant spatial variations in the SSM levels.

In contrast, synergy 2, showing the lowest performance among the five synergies, with an OA of 70.48% for testing and a Kappa coefficient of 55.12%, indicated only moderate agreement between the observed SSM ground-truth data and the predicted values. The lower performance of synergy 2 suggests that the data sources included in this synergy may not have provided sufficient or relevant information for accurately predicting SSM levels in the SVM model, but they still provided useful insights into the SSM levels, indicating

that certain features or data sources within this synergy may contribute positively to the model's performance. Therefore, the SVM model's performance in classifying SSM levels in the temperate forests of the study site was highly dependent on the chosen data synergy. This implies that the results highlight the critical role of the combinations of multi-source remote sensing data selection in SVM's application for SSM estimation, highlighting the need for careful consideration of the combinations of remote sensing data used.

3.3. Comparison of Model Performance for Each Synergy

The comparative analysis of the RF and SVM models for SSM classification across the five synergies of multi-source remote sensing data revealed large differences in model performance. The evaluation metrics R^2 , RMSE, and MAE provided comprehensive insights into the predictive capabilities and accuracy of each model (Table 5).

Table 5. Assessment of best model performance for each synergy among RF and SVM.

Synergies	Random Forest (RF)			Support Vector Machine (SVM)		
	R^2	RMSE	MAE	R^2	RMSE	MAE
Synergy 1	0.910	0.039	0.035	0.879	0.101	0.083
Synergy 2	0.905	0.064	0.056	0.884	0.020	0.019
Synergy 3	0.914	0.035	0.030	0.766	0.175	0.142
Synergy 4	0.964	0.025	0.023	0.874	0.014	0.010
Synergy 5	0.896	0.051	0.045	0.799	0.152	0.124

Synergy 1 demonstrated a strong performance with both models, but the RF model outperformed the SVM model. The RF model achieved an R^2 of 0.910, an RMSE of 0.039, and an MAE of 0.035, indicating a robust fit and low prediction error. In comparison, the SVM model achieved an R^2 of 0.879, an RMSE of 0.101, and an MAE of 0.083. While the SVM model also showed good performance, the higher RMSE and MAE values suggest that it was less accurate and consistent than the RF model for this synergy.

For Synergy 2, the RF model again showed good performance, with an R^2 of 0.905, an RMSE of 0.064, and an MAE of 0.056. The SVM model, while achieving a comparable R^2 of 0.884, showed a lower RMSE of 0.020 and MAE of 0.019. This suggests that while the SVM model was highly precise for this synergy, it may have been overfitting the data, capturing noise rather than underlying patterns.

Synergy 3 revealed a significant difference between the models. The RF model achieved an R^2 of 0.914, an RMSE of 0.035, and an MAE of 0.030, indicating high accuracy and reliability. In contrast, the SVM model's performance was considerably lower, with an R^2 of 0.766, an RMSE of 0.175, and an MAE of 0.142. This suggests that the RF model was much more effective in capturing the complex relationships within the data for this synergy.

Synergy 4 was the most suitable synergy, particularly for the RF model, which achieved the highest R^2 of 0.964, the lowest RMSE of 0.025, and an MAE of 0.023. This indicates an excellent fit and minimal prediction error, making it the most accurate configuration for SSM classification. The SVM model, while achieving good results (R^2 of 0.874, RMSE of 0.014, and MAE of 0.010), was still less effective than the RF model. The slight difference between the RMSE and MAE values of the SVM model suggests it was precise but potentially less robust.

In synergy 5, the RF model's performance (R^2 of 0.896, RMSE of 0.051, MAE of 0.045) was higher than that of the SVM model (R^2 of 0.799, RMSE of 0.152, MAE of 0.124). The higher RMSE and MAE values for the SVM model indicate less precision and higher prediction error compared to the RF model.

Among the five synergies and two models, synergy 4 (integrating S-2 data with terrain factors) with the RF model was the most effective approach, achieving the highest R^2 and the lowest RMSE and MAE. Therefore, this synergy, which contributes to the

complementary nature of spectral and topographic data, provides a detailed representation of SSM classification.

3.4. Spatial Distribution of Surface Soil Moisture by RF and SVM

The spatial distributions of SSM maps at 30 m resolution in the temperate forests of the UTCBF were analyzed based on the five synergistic approaches of integrating multi-source remote sensing data and the two machine learning models. The SSM maps derived from the five synergies with the two models revealed distinct spatial patterns of SSM content across the study area (Figure 6). In this study, the spatial distribution of SSM content in temperate forests was ranked into three levels: low (<20% VWC), moderate (20–40% VWC), and high (>40% VWC).

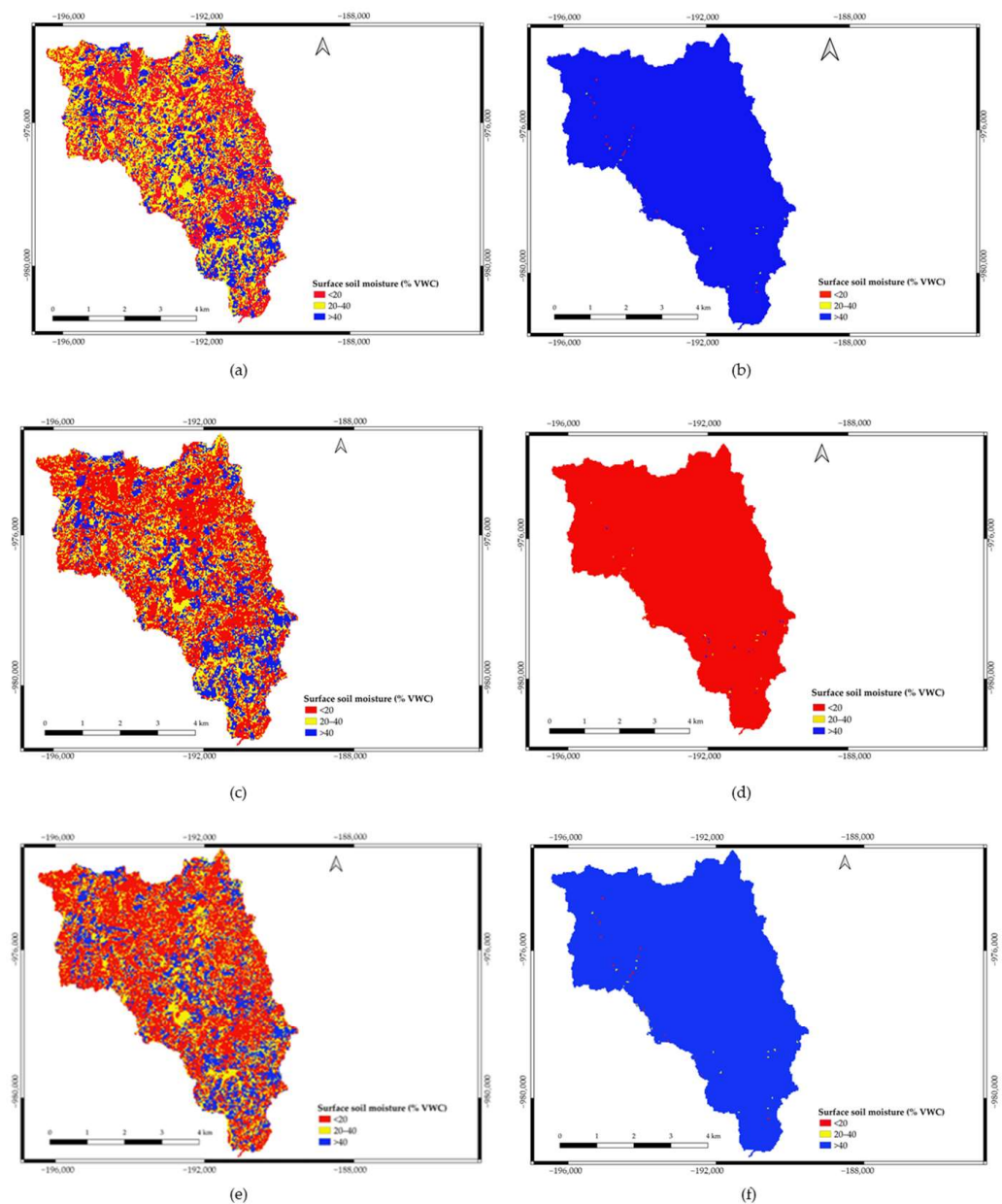


Figure 6. Cont.

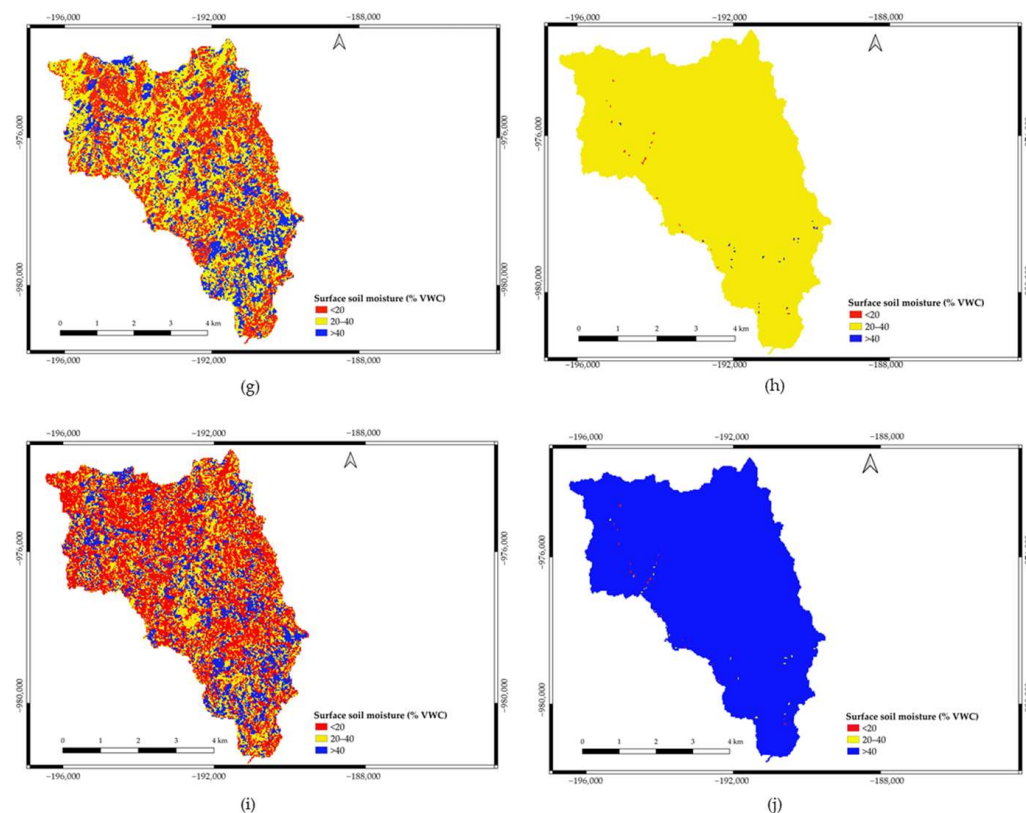


Figure 6. Spatial distribution of SSM in temperate forests of UTCBF: (a) synergy 1 by RF, (b) synergy 1 by SVM, (c) synergy 2 by RF, (d) synergy 2 by SVM, (e) synergy 3 by RF, (f) synergy 3 by SVM, (g) synergy 4 by RF, (h) synergy 4 by SVM, (i) synergy 5 by RF, and (j) synergy 5 by SVM.

In synergy 1, low and moderate SSM levels were mostly distributed in the RF model (Figure 6a), while high SSM levels were mostly distributed across all of the different forest types in the SVM model (Figure 6b). On the other hand, in synergy 2, low SSM levels were more distributed across the different forest types in the RF model (Figure 6c). In contrast to synergy 1 with the SVM model, the spatial distribution of low SSM levels was very high in synergy 2 with the SVM model (Figure 6d). Additionally, the spatial distribution of low SSM levels was more distributed across the different forest types in synergy 3 with the RF model (Figure 6e) than in synergy 2 with the RF model. In contrast, similar to synergy 1 with the SVM model, high SSM levels were observed in synergy 3 with the SVM model (Figure 6f).

In synergy 4, the spatial distributions of low and moderate SSM levels were distributed equally to the high SSM levels in the RF model (Figure 6g). In synergy 4 with the SVM model, the spatial distribution of moderate SSM levels was distributed in almost all the forest types (Figure 6h). Similar to synergy 2, low SSM levels were more distributed in synergy 5 with the RF model (Figure 6i). Similar to synergy 1 and synergy 3, the high SSM was distributed across almost all the forest types in synergy 5 with the SVM model (Figure 6j).

The spatial differences in predictions between the RF and SVM models can be attributed to differences in their learning and generalization processes. The SVM model seeks to find the optimal hyperplane that best separates the classes, which can sometimes result in overfitting to certain patterns in the data. The SVM model, while sometimes achieving high training accuracy (Figure 5b), showed signs of overfitting, where it captured noise rather than underlying patterns. This overfitting can lead to less reliable predictions in new, unseen data.

Based on the results of the SVM model (Figure 6b,d,f,h,j), the predictions tended to be either high or moderate for almost the entire area. There are many possible reasons for this. SVMs are particularly sensitive to the choice of kernel. If the chosen parameters do not capture the complexity of the data well, the model might produce less nuanced predictions. Moreover, the SVM model may not differentiate between small changes in SSM effectively if the input features used in the synergies are not highly discriminative for SSM levels. In addition, the SVM might be biased towards the more frequent classes if the training data have an imbalance in the number of ground-truth data across the SSM levels.

Another possible reason is that the SVM model might struggle to distinguish between SSM levels accurately if the boundaries for SSM levels are poorly defined due to overlapping feature values. If the SSM levels in the ground-truth data are skewed towards certain classes (e.g., more instances of high SSM levels), the SVM model might achieve high accuracy by predominantly predicting those classes. This would result in high OA and Kappa values even if the spatial distribution differs from that of the RF model.

The RF model, on the other hand, relies on an ensemble of decision trees that tend to capture more complex relationships and interactions between features. The RF model, with its ensemble approach, is typically more robust and better at generalizing from training data to testing data; its ability to handle large datasets with high dimensionality and its built-in feature importance measure make it a strong candidate for complex tasks, such as SSM classification. This allows the RF model to produce more detailed and accurate spatial distributions of SSM levels.

3.5. Assessing the Correlation between Ground-Truth Data and Predicted Values

The correlation between the observed SSM ground-truth data and the predicted SSM values derived from five synergies using RF and SVM models was evaluated to ensure the effectiveness of integrating multi-source remote sensing data and machine learning algorithms (Table 6).

Table 6. Correlation between observed SSM and predicted SSM from five synergies by RF and SVM models.

	Synergy 1	Synergy 2	Synergy 3	Synergy 4	Synergy 5
Predicted SSM by RF Model					
Observed SSM	0.95	0.95	0.96	0.98	0.95
Predicted SSM by SVM Model					
Observed SSM	0.94	0.94	0.88	0.94	0.89

The correlation coefficients between the observed SSM ground-truth data and the predicted SSM values from five synergies using the RF model ranged from 0.95 to 0.98, highlighting the superior performance of the RF model. The highest correlation was observed with synergy 4 ($r = 0.98$), followed by synergy 3 ($r = 0.96$).

The correlation coefficients between the observed SSM ground-truth data and the predicted SSM values from five synergies using the SVM model ranged from 0.88 to 0.94, indicating a high degree of predictive accuracy for each synergy. The highest correlation was observed with synergies 1, 2, and 4 ($r = 0.94$).

Consistent with the findings reported in Section 3.3, the correlation analysis also highlighted that synergy 4 (integrating S-2 and terrain factors) with the RF model consistently demonstrated a higher correlation between the observed SSM ground-truth data and the predicted values than the SVM model across all the synergies.

4. Discussion

4.1. Significance of Multi-Source Remote Sensing Data Synergies

The integration of various remote sensing data sources has become a crucial method for enhancing the accuracy of SSM mapping. Previous studies have also demonstrated the effectiveness of integrating remote sensing data to improve SSM estimation. For instance, Mohseni et al. [4] found that the integration of SMAP with MODIS products significantly improved the accuracy of SSM predictions. Yang et al. [23] also found that SMAP with MODIS products, ERA5-Land dataset, SRTM DEM, and soil texture using a light gradient-boosting machine provided and captured the temporal variation in soil moisture. In addition, Abowarda et al. [32] suggested that the integration of SMAP with the China Meteorological Administration Land Data Assimilation System showed the best performance in their study site (Haihe Basin, located in the north of China). This highlights that SMAP combined with other remote sensing data provided superior mapping performance, allowing for detailed soil analysis, which is critical for accurate SSM mapping.

In contrast, other studies have highlighted the integration of S-2 with other remote sensing data in SSM estimation. For example, Gao et al. [49] explored the integration of S-2 and S-1 to map SSM at 100 m resolution. Attarzadeh et al. [26] found that the integration of S-2 with S-1 was crucial in understanding the spatial distribution of soil moisture, providing a better approach in terms of estimation accuracy ($R^2 = 0.89$). Furthermore, Foucras et al. [87] also showed that combining S-2, S-1, and MODIS provided reliable estimates of SSM across a wide range of climatically different sites. Comparatively, Nativel et al. [39] also showed that a hybrid methodology using S-2, S-1, classical change detection SSM index, radar incidence angle, and NDVI provided an improvement in soil moisture estimations. Our study also found that the integration of S-2 with other remote sensing data provided the best performance in SSM estimation. However, among the multi-source remote sensing data, the integration of S-2 and terrain factors provided the most comprehensive dataset in this study, enhancing the model's ability to predict SSM levels accurately.

4.2. Significance of Terrain Factors in Synergy Selection for Surface Soil Moisture Estimation

In this study, we demonstrated that terrain factors are the most important factors affecting SSM estimation. Previous studies have also demonstrated the effect of terrain factors on SSM estimation. For instance, Charpentier and Groffman [88] examined the influence of topography and the magnitude of moisture content on soil moisture variability within fine-scale pixels during the First International Satellite Land Surface Climatology Project field experiment; they discovered that soil moisture variability increased with increasing topographic heterogeneity. Similarly, Mohanty et al. [89] found that slope had a significant effect on the soil moisture distribution in a gently sloping agricultural field within the Little Washita agricultural watershed. Jacobs et al. [90] also highlighted the significance of slope positions in identifying the time-stable points for mean soil moisture in four agricultural fields in Iowa. Several studies have also emphasized that relative slope position is important in determining soil moisture variations, indicating that a simple averaging of soil moisture values over the slope may lead to inaccuracies over different timescales. In the Gourma research site of West Africa, de Rosnay et al. [91] observed an inverse correlation between soil moisture and hillslope position, with lower soil moisture levels at higher elevations. In Australia, Western et al. [92] demonstrated systematic spatial variations in soil moisture, particularly in saturated areas associated with topographic convergence.

Li et al. [46] also highlighted the role of elevation in influencing soil moisture, finding that higher elevations had distinct soil moisture profiles due to climatic variations. On the other hand, Liu et al. [93] explored the influence of slope steepness on soil water content, showing that shady slopes can retain a water supply sufficient for sustaining forests, and that the existence of forests on shady slopes further reduces evaporation. Qiu et al. [94] discussed the strong and positive correlation of aspect and soil moisture, assuming that south-facing slopes experience higher evaporation rates, leading to lower soil moisture

compared to north-facing slopes. In addition, Liang et al. [95] addressed the effect of valley depth on soil moisture, finding that deeper valleys tend to retain more water, leading to higher soil moisture levels. Therefore, variations in terrain factors (e.g., elevation, slope, aspect, slope steepness, and valley depth) affect the distribution of SSM.

4.3. Application of Machine Learning and Deep Learning in Surface Soil Moisture Estimation

The use of machine learning and deep learning models has been well documented in remote sensing applications due to their robustness and ability to handle complex interactions between variables. Pasolli et al. [96] showed that support vector regression performs exceptionally well in predicting SSM when trained with multi-source data (i.e., different active and/or passive microwave measurements acquired using various sensor frequencies, polarizations, and acquisition geometries). Ahmad et al. [36] found that an SVM model integrating backscatter and incidence angle from the Tropical Rainfall Measuring Mission (TRMM), along with NDVI from the Advanced Very-High-Resolution Radiometer (AVHRR), achieved higher accuracy than ANN and MLR models, with an r of 0.57, RMSE of 2.01, and MAE of 1.97. However, this performance is notably lower in accuracy compared to our findings. Synergy 4 showed the most suitable approach with an $r = 0.94$, RMSE = 0.014, and MAE = 0.010 in the SVM model of our findings. The observed differences in model performance can be attributed to various factors, including the input data quality, the study area characteristics, and the specific pre-processing and modeling techniques applied.

Our study used multi-source remote sensing data, including S-2 and terrain factors for synergy 4 in the SVM model. The integration of high-resolution remote sensing data, which provides fine spatial resolution, likely contributed to the higher accuracy observed in our findings. However, this finding showed signs of overfitting, which is a common issue when dealing with complex datasets and high-dimensional features. In contrast, Ahmad et al. [36] reported lower accuracy metrics by using TRMM and AVHRR data, indicating a potentially more generalized model, though at the cost of precision. The signs of overfitting in our findings and lower accuracy performance in the other study indicate the limitation of the practical applicability of the SVM model.

In contrast to the SVM model, deep learning methods, particularly ANN models, have achieved higher accuracy in some studies. Hassan-Esfahani et al. [37] found that an ANN model in conjunction with field measurements and spectral images achieved better performance for SSM estimation, with an R^2 of 0.77, RMSE of 2.0, and MAE of 1.8; they suggested that their approach was particularly well suited for capturing the nonlinear relationships in SSM prediction. Chung et al. [40] also used an ANN with Sentinel-1, topography (elevation and slope), soil (percentage of sand and clay), and hydrological components, such as an antecedent precipitation index and dry days, achieving excellent performance, with an accuracy of $r = 0.78$ and RMSE = 5.60. In addition, EI Hajj et al. [97] used a neural network technique to integrate S-1 and S-2 data, achieving significant improvements in SSM prediction, with an accuracy of approximately 5 vol.%; they highlighted the potential for deep learning models to advance SSM estimation by leveraging diverse and high-dimensional data sources.

In this study, the RF model has proven to be a powerful tool for estimating SSM, as evidenced by its superior performance ($R^2 = 0.964$, RMSE = 0.025, and MAE = 0.023). Our findings also align with those of other studies. For instance, Acharya et al. [16] found that SSM prediction using RF with OPTRAM moisture values, rainfall, the standardized precipitation index, and clay percentage showed high goodness of fit ($R^2 = 0.69$). Adab et al. [3] also found that the RF model provided the highest Nash–Sutcliffe efficiency value (0.73), with an RMSE of 4.60%, an index of agreement of 0.91, and an MBE of 2.16 for SSM estimation covered by the different land-use types. Moreover, Zhao et al. [20] also demonstrated an RF model using soil evaporative efficiency, resulting in a strong predictive performance for SSM, with accuracy of $0.035 \text{ cm}^3/\text{cm}^3$, highlighting the model's capability to handle the data. Although the RF model in other studies consistently outperformed, the accuracy

metrics were notably different from those in our study. The differences in accuracy metrics highlight the robustness and precision of the model in estimating SSM. The integration of terrain factors and higher-resolution remote sensing data in our study might have led to the improved performance of the RF model compared to other studies.

4.4. Recommendation for Suitable Approach to Surface Soil Moisture Classification

Based on a comprehensive evaluation of model performance across the five synergies (Tables 5 and 6; Figure 5a,b), the integration of S-2 and terrain factors with the RF model consistently demonstrated superior performance in terms of accuracy, robustness, and reliability across all the synergies. Moreover, this approach provided the highest values of R^2 and r , along with the lowest RMSE and MAE values, indicating an excellent fit and minimal prediction error. Therefore, we recommended this approach as the most suitable approach for SSM classification to achieve the most accurate and reliable SSM levels. This recommendation could not only provide a practical application for SSM classification but also provide reliable information for water resource management and environmental monitoring, emphasizing the importance of optimal data synergy selection to achieve accurate and reliable predictions.

4.5. Implications for Sustainable Forest Management

The results of this study have practical implications for sustainable forest management in the temperate forests of the study site. Through accurately mapping SSM levels, forest managers can make informed decisions regarding water resource management, forest health monitoring, and planning of reforestation activities. For instance, areas with low SSM (<20% VWC) could be prioritized for irrigation or drought-resistant species, while areas with high SSM (>40% VWC) might be suitable for species that can survive in moist conditions, such as Japanese cedar plantations, as determined by referencing the most suitable SSM map of this study.

4.6. Limitations and Future Research

Despite the promising results, there are limitations to this study. This study was conducted over a limited time period, and long-term studies will be necessary to validate the findings across different seasons and years.

Future research should explore the integration of other remote sensing data sources, such as LiDAR, hyperspectral imagery, MODIS, or SMAP, in order to further compare SSM mapping accuracy. Additionally, investigating the impact of different machine learning algorithms, such as deep learning models, could provide insights into improving SSM estimation.

5. Conclusions

This study demonstrates the effectiveness of integrating multi-source remote sensing data and two machine learning models (RF and SVM) for accurate SSM mapping in temperate forests of Central Japan. The application of the RF model to predict SSM levels using five different synergies of multi-source remote sensing data reveals notable insights in terms of accuracy and reliability compared with the application of the SVM model. Moreover, among the five synergies with the RF model, the synergy of Sentinel-2 and terrain factors, including elevation, slope, aspect, slope steepness, and valley depth, emerged as the most suitable approach, providing valuable insight into the spatial distribution of SSM content across different forest types based on the values of classification accuracy and evaluation metrics. This finding contributes to advancing our understanding of SSM dynamics in temperate forests and has practical implications for managing land and water resources as part of forest management practice.

Author Contributions: Conceptualization, K.W.; methodology, K.W.; software, K.W.; validation, K.W.; formal analysis, K.W.; investigation, K.W.; data curation, K.W.; writing—original draft preparation, K.W.; writing—review and editing, T.S. and S.T.; supervision, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: No informed consent was required for this study, as it did not involve human participants, human data, or human tissue.

Data Availability Statement: The data presented in this study are available on request from the first author for academic purposes without commercial purposes.

Acknowledgments: We would like to thank the three anonymous reviewers for their insightful suggestions. The first author, K.W., sincerely thanks Mary McClure (Laboratory of Global Forest Environmental Studies, Department of Global Agricultural Sciences, The University of Tokyo) and Zwe Maung Maung for sharing some GEE codes. We also would like to thank Keisuke TOYAMA (Department of Forest Science, Faculty of Agriculture, Iwate University) and Yoshida Akiko (The University of Tokyo Chiba Forests) for their support during the field data collection.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Specifications of FieldScout Time Domain Reflectometry (TDR 350) Meter

Measurement unit	Percent volumetric water content (VWC)		
Accuracy and range	0.1% increment \pm 3.0%, 0% to saturation		
GNSS	Supported systems: Galileo, GLONASS, GPS, QZSS, EGNOS, MSAS, SBAS, and WAAS enabled		
Log capacity	50,000 measurements		
Available rod dimensions	Turf	1.5'' (3.8 cm)	
	Short	3.0'' (7.6 cm)	
	Medium	4.8'' (12.2 cm)	
	Long	8.0'' (20.32 cm)	

References

- Peng, J.; Albergel, C.; Balenzano, A.; Brocca, L.; Cartus, O.; Cosh, M.H.; Crow, W.T.; Dabrowska-Zielinska, K.; Dadson, S.; Davidson, M.W.J.; et al. A roadmap for high-resolution satellite soil moisture applications—Confronting product characteristics with user requirements. *Remote Sens. Environ.* **2021**, *252*, 112162. [[CrossRef](#)]
- Sabaghy, S.; Walker, J.P.; Renzullo, L.J.; Jackson, T.J. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sens. Environ.* **2018**, *209*, 551–580. [[CrossRef](#)]
- Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghalhari, G.A.F. Machine learning to estimate surface soil moisture from remote sensing data. *Water* **2020**, *12*, 3223. [[CrossRef](#)]
- Mohseni, F.; Ahrari, A.; Hurnert, J.H.; Montzka, C. The synergies of SMAP enhanced and MODIS products in a random forest regression for estimating 1 km soil moisture over Africa using Google Earth Engine. *Big Earth Data* **2023**, *8*, 33–57. [[CrossRef](#)]
- Khazaei, M.; Hamzeh, S.; Samani, N.N.; Muhuri, A.; Goita, K.; Weng, Q. A web-based system for satellite-based high-resolution global soil moisture maps. *Comput. Geosci.* **2023**, *170*, 105250. [[CrossRef](#)]
- Pioui, C.; Gay, P.E.; Benahi, A.S.; Babah Ebbe, M.A.O.; Chihrane, J.; Ghaout, S.; Cisse, S.; Diakite, F.; Lazar, M.; Cressman, K.; et al. Soil moisture from remote sensing to Forecast Desert Locust Presence. *J. Appl. Ecol.* **2019**, *56*, 966–975. [[CrossRef](#)]
- 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.; et al. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716. [[CrossRef](#)]
- Wang, L.; Qu, J.J. Satellite remote sensing applications for surface soil moisture monitoring: A review. *Front. Earth Sci. China* **2009**, *3*, 237–247. [[CrossRef](#)]
- Koster, R.D.; Suarez, M.J.; Higgins, R.W.; Van den Dool, H.M. Observational evidence that soil moisture variations affect precipitation. *Geophys. Res. Lett.* **2003**, *30*, 1241. [[CrossRef](#)]
- Patel, P.; Ankur, K.; Jamshidi, S.; Tiwari, A.; Nadimpalli, R.; Busireddy, N.K.R.; Safae, S.; Osuri, K.K.; Karmakar, S.; Ghosh, S.; et al. Impact of urban representation on simulation of Hurricane rainfall. *Geophys. Res. Lett.* **2023**, *50*, e2023GL104078. [[CrossRef](#)]
- Gu, X.; Jamshidi, S.; Sun, H.; Niyogi, D. Identifying multivariate controls of soil moisture variations using multiple wavelet coherence in the U.S. Midwest. *J. Hydrol.* **2021**, *602*, 126755. [[CrossRef](#)]
- Kashyap, B.; Kumar, R. Sensing methodologies in agriculture for soil moisture and nutrient monitoring. *IEEE Access* **2021**, *9*, 14095–14121. [[CrossRef](#)]

13. Ait Hssaine, B.; Merlin, O.; Ezzahar, J.; Ojha, N.; Er-Raki, S.; Khabba, S. An evapotranspiration model self-calibrated from remotely sensed surface soil moisture, land surface temperature and vegetation cover fraction: Application to disaggregated SMOS and MODIS data. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1781–1803. [[CrossRef](#)]
14. Alemayehu, T.; van Griensven, A.; Senay, G.B.; Bauwens, W. Evapotranspiration mapping in a heterogeneous landscape using remote sensing and global weather datasets: Application to the Mara Basin, East Africa. *Remote Sens.* **2017**, *9*, 390. [[CrossRef](#)]
15. Zhang, J.; Zhou, Z.; Yao, F.; Yang, L.; Hao, C. Validating the modified perpendicular drought index in the North China region using in situ soil moisture measurement. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 542–546. [[CrossRef](#)]
16. Acharya, U.; Daigh, A.L.M.; Oduor, P.G. Soil moisture mapping with moisture-related indices, OPTRAM, and an integrated random forest-OPTRAM algorithm from Landsat 8 images. *Remote Sens.* **2022**, *14*, 3801. [[CrossRef](#)]
17. Lennard, A.T.; Macdonald, N.; Clark, S.; Hooke, J.M. The application of a drought reconstruction in water resource management. *Hydrol. Res.* **2016**, *47*, 646–659. [[CrossRef](#)]
18. Smith, K.A.; Barker, L.J.; Tanguy, M.; Parry, S.; Harrigan, S.; Legg, T.P.; Prudhomme, C.; Hannaford, J. A multi-objective ensemble approach to hydrological modelling in the UK: An application to historic drought reconstruction. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 3247–3268. [[CrossRef](#)]
19. Piles, M.; Camps, A.; Vall-Llossera, M.; Corbella, I.; Panciera, R.; Rudiger, C.; Kerr, Y.H.; Walker, J. Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3156–3166. [[CrossRef](#)]
20. Zhao, Z.; Jin, R.; Kang, J.; Ma, C.; Wang, W. Using of remote sensing-based auxiliary variables for soil moisture scaling and mapping. *Remote Sens.* **2022**, *14*, 3373. [[CrossRef](#)]
21. Du, J.; Kimball, J.S.; Bindlish, R.; Walker, J.P.; Watts, J.D. Local scale (3-m) soil moisture mapping using SMAP and Planet SuperDove. *Remote Sens.* **2022**, *14*, 3812. [[CrossRef](#)]
22. Hamze, M.; Baghdadi, N.; El Hajj, M.M.; Zribi, M.; Bazzi, H.; Cheviron, B.; Faour, G. Integration of L-Band derived soil roughness into a bare soil moisture retrieval approach from C-Band SAR data. *Remote Sens.* **2021**, *13*, 2102. [[CrossRef](#)]
23. Yang, Z.; He, Q.; Miao, S.; Wei, F.; Yu, M. Surface soil moisture retrieval of China using multi-source data and ensemble learning. *Remote Sens.* **2023**, *15*, 2786. [[CrossRef](#)]
24. Crow, W.T.; Berg, A.A.; Cosh, M.H.; Loew, A.; Mohanty, B.P.; Panciera, R.; De Rosnay, P.; Ryu, D.; Walker, J.P. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* **2012**, *50*, RG2002. [[CrossRef](#)]
25. Ma, Y.; Hou, P.; Zhang, L.; Cao, G.; Sun, L.; Pang, S.; Bai, J. High-resolution quantitative retrieval of soil moisture based on multisource data fusion with random forests: A case study in the Zoige region of the Tibetan Plateau. *Remote Sens.* **2023**, *15*, 1531. [[CrossRef](#)]
26. Attarzadeh, R.; Amini, J.; Notarnicola, C.; Greifeneder, F. Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale. *Remote Sens.* **2018**, *10*, 1285. [[CrossRef](#)]
27. Schmutge, T.J.; Jackson, T.J.; McKim, H.L. Survey of methods for soil moisture determination. *Water Resour. Res.* **1980**, *16*, 961–979. [[CrossRef](#)]
28. Brocca, L.; Zhao, W.; Lu, H. High-resolution observations from space to address new applications in hydrology. *Innovation* **2023**, *4*, 100437. [[CrossRef](#)]
29. Mohanty, B.P.; Cosh, M.H.; Lakshmi, V.; Montzka, C. Soil moisture remote sensing: State-of-the-science. *Vadose Zone J.* **2017**, *16*, 1–9. [[CrossRef](#)]
30. Cashion, J.; Lakshmi, V.; Bosch, D.; Jackson, T.J. Microwave remote sensing of soil moisture: Evaluation of the TRMM microwave imager (TMI) satellite for the little river Watershed Tifton, Georgia. *J. Hydrol.* **2005**, *307*, 242–253. [[CrossRef](#)]
31. Das, K.; Paul, P.K. Present status of soil moisture estimation by microwave remote sensing. *Cogent Geosci.* **2015**, *1*, 1084669. [[CrossRef](#)]
32. Abowarda, A.S.; Bai, L.; Zhang, C.; Long, D.; Li, X.; Huang, Q.; Sun, Z. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sens. Environ.* **2021**, *255*, 112301. [[CrossRef](#)]
33. Liao, T.H.; Kim, S.B.; Handwerker, A.; Fielding, E.; Cosh, M.; Schulz, W. High-resolution soil-moisture maps over landslide regions in Northern California Grassland derived from SAR backscattering coefficients. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4547–4560. [[CrossRef](#)]
34. Du, J.; Kimball, J.S.; Sheffield, J.; Pan, M.; Fisher, C.K.; Beck, H.E.; Wood, E.F. Satellite flood inundation assessment and forecast using SMAP and Landsat. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6707–6715. [[CrossRef](#)] [[PubMed](#)]
35. Yeh, I.C.; Lien, C.H. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Syst. Appl.* **2009**, *36*, 2473–2480. [[CrossRef](#)]
36. Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* **2010**, *33*, 69–80. [[CrossRef](#)]
37. Hassan-Esfahani, L.; Torres-Rua, A.; Jensen, A.; McKee, M. Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sens.* **2015**, *7*, 2627–2646. [[CrossRef](#)]
38. Ettalbi, M.; Baghdadi, N.; Garambois, P.A.; Bazzi, H.; Ferreira, E.; Zribi, M. Soil moisture retrieval in bare agricultural areas using Sentinel-1 images. *Remote Sens.* **2023**, *15*, 3502. [[CrossRef](#)]

39. Nativel, S.; Ayari, E.; Rodriguez-Fernandez, N.; Baghdadi, N.; Madelon, R.; Albergel, C.; Zribi, M. Hybrid methodology using Sentinel-1/Sentinel-2 for soil moisture estimation. *Remote Sens.* **2022**, *14*, 2434. [[CrossRef](#)]
40. Chung, J.; Lee, Y.; Kim, J.; Jung, C.; Kim, S. Soil moisture content estimation based on Sentinel-1 SAR imagery using an artificial neural network and hydrological components. *Remote Sens.* **2022**, *14*, 465. [[CrossRef](#)]
41. Zhao, L.; Yang, Z.L. Multi-sensor land data assimilation: Toward a robust global soil moisture and snow estimation. *Remote Sens. Environ.* **2018**, *216*, 13–27. [[CrossRef](#)]
42. Yao, P.; Lu, H.; Shi, J.; Zhao, T.; Yang, K.; Cosh, M.H.; Gianotti, D.J.S.; Entekhabi, D. A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019). *Sci. Data* **2021**, *8*, 143. [[CrossRef](#)] [[PubMed](#)]
43. The University of Tokyo Forests Education and Research Plan (2021–2030) of the University of Tokyo Forests: Part 3 Regional Forest Plans: The University of Tokyo Chiba Forest (the 14th Period). *Misc. Inf. Univ. Tokyo For.* **2022**, *64*, 53–102.
44. Benninga, H.J.F.; Carranza, C.D.U.; Peziz, M.; Van Santen, P.; Van Der Ploeg, M.J.; Augustijn, D.C.M.; Van Der Velde, R. The Raam regional soil moisture monitoring network in the Netherlands. *Earth Syst. Sci. Data* **2018**, *10*, 61–79. [[CrossRef](#)]
45. Walker, J.P.; Willgoose, G.R.; Kalma, J.D. In situ measurement of soil moisture: A comparison of techniques. *J. Hydrol.* **2004**, *293*, 85–99. [[CrossRef](#)]
46. Li, L.; Wu, D.; Wang, T.; Wang, Y. Effect of topography on spatiotemporal patterns of soil moisture in a mountainous region of Northwest China. *Geoderma Reg.* **2022**, *28*, e00456. [[CrossRef](#)]
47. Nakaegawa, T.; Tokuhira, T.; Itoh, A.; Hosaka, M. Evaluation of seasonal cycles of hydrological processes in Japan Meteorological Agency land data analysis. *Pap. Meteorol. Geophys.* **2007**, *58*, 73–83. [[CrossRef](#)]
48. Kaleita, A.L.; Tian, L.F.; Hirschi, M.C. Relationship between soil moisture content and soil surface reflectance. *Trans. ASAE* **2005**, *48*, 1979–1986. [[CrossRef](#)]
49. Gao, Q.; Zribi, M.; Escorihuela, M.J.; Baghdadi, N. Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. *Sensors* **2017**, *17*, 1966. [[CrossRef](#)]
50. Ju, Z.; Leong Tan, M.; Samat, N.; Kiat Chang, C. Comparison of Landsat 8, Sentinel-2 and spectral indices combinations for Google Earth Engine-based land use mapping in the Johor River Basin, Malaysia. *Malays. J. Soc. Space* **2021**, *17*, 30–46. [[CrossRef](#)]
51. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [[CrossRef](#)]
52. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Weisz, R. Aerial color infrared photography for determining late-season nitrogen requirements in Corn. *Agron. J.* **2005**, *97*, 1443–1451. [[CrossRef](#)]
53. Valdivieso-Ros, C.; Alonso-Sarria, F.; Gomariz-Castillo, F. Effect of different atmospheric correction algorithms on Sentinel-2 imagery classification accuracy in a semiarid Mediterranean area. *Remote Sens.* **2021**, *13*, 1770. [[CrossRef](#)]
54. Crist, E.P.; Cicone, R.C. A physically-based transformation of thematic mapper data—the TM Tasseled Cap. *IEEE Trans. Geosci. Remote Sens.* **1984**, *GE-22*, 256–263. [[CrossRef](#)]
55. Lastovicka, J.; Svec, P.; Paluba, D.; Kobliuk, N.; Svoboda, J.; Hladky, R.; Stych, P. Sentinel-2 data in an evaluation of the impact of the disturbances on forest vegetation. *Remote Sens.* **2020**, *12*, 1914. [[CrossRef](#)]
56. Clevers, J.G.P.W. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sens. Environ.* **1989**, *29*, 25–37. [[CrossRef](#)]
57. Saad El Imanni, H.; El Harti, A.; El Iysaouy, L. Wheat yield estimation using remote sensing indices derived from Sentinel-2 time series and Google Earth Engine in a highly fragmented and heterogeneous agricultural region. *Agronomy* **2022**, *12*, 2853. [[CrossRef](#)]
58. Yang, L.; Meng, X.; Zhang, X. SRTM DEM and its application advances. *Int. J. Remote Sens.* **2011**, *32*, 3875–3896. [[CrossRef](#)]
59. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [[CrossRef](#)]
60. Forkuor, G.; Maathuis, B. Comparison of SRTM and ASTER derived digital elevation models over two regions in Ghana—Implications for hydrological and environmental modeling. In *Studies on Environmental and Applied Geomorphology*; InTech: London, UK, 2012; pp. 219–240.
61. Bircher, P.; Liniger, H.P.; Prasuhn, V. Comparing different multiple flow algorithms to calculate RUSLE factors of slope length (L) and slope steepness (S) in Switzerland. *Geomorphology* **2019**, *346*, 106850. [[CrossRef](#)]
62. Prodromou, M.; Theocharidis, C.; Gitas, I.Z.; Eliades, F.; Themistocleous, K.; Papasavvas, K.; Dimitrakopoulos, C.; Danezis, C.; Hadjimitsis, D. Forest habitat mapping in Natura2000 regions in Cyprus using Sentinel-1, Sentinel-2 and topographical features. *Remote Sens.* **2024**, *16*, 1373. [[CrossRef](#)]
63. Bartold, M.; Kluczek, M. A machine learning approach for mapping chlorophyll fluorescence at Inland Wetlands. *Remote Sens.* **2023**, *15*, 2392. [[CrossRef](#)]
64. Zhao, Q.; Yu, L.; Li, X.; Peng, D.; Zhang, Y.; Gong, P. Progress and trends in the application of Google Earth and Google Earth Engine. *Remote Sens.* **2021**, *13*, 3778. [[CrossRef](#)]
65. Noi Phan, T.; Kuch, V.; Lehnert, L.W. Land cover classification using Google Earth Engine and random forest classifier—the role of image composition. *Remote Sens.* **2020**, *12*, 2411. [[CrossRef](#)]
66. Praticò, S.; Solano, F.; Di Fazio, S.; Modica, G. Machine learning classification of Mediterranean forest habitats in Google Earth Engine based on seasonal Sentinel-2 time-series and input image composition optimisation. *Remote Sens.* **2021**, *13*, 586. [[CrossRef](#)]

67. Ghorbanian, A.; Zaghian, S.; Asiyabi, R.M.; Amani, M.; Mohammadzadeh, A.; Jamali, S. Mangrove ecosystem mapping using Sentinel-1 and Sentinel-2 satellite images and random forest algorithm in Google Earth Engine. *Remote Sens.* **2021**, *13*, 2565. [\[CrossRef\]](#)
68. Kpienbaareh, D.; Sun, X.; Wang, J.; Luginaah, I.; Kerr, R.B.; Lupafya, E.; Dakishoni, L. Crop type and land cover mapping in Northern Malawi using the integration of Sentinel-1, Sentinel-2, and PlanetScope satellite data. *Remote Sens.* **2021**, *13*, 700. [\[CrossRef\]](#)
69. Xia, J.; Yokoya, N.; Pham, T.D. Probabilistic mangrove species mapping with multiple-source remote-sensing datasets using label distribution learning in Xuan Thuy National Park, Vietnam. *Remote Sens.* **2020**, *12*, 3834. [\[CrossRef\]](#)
70. Mahdavi, S.; Salehi, B.; Amani, M.; Granger, J.; Brisco, B.; Huang, W. A dynamic classification scheme for mapping spectrally similar classes: Application to wetland classification. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *83*, 101914. [\[CrossRef\]](#)
71. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [\[CrossRef\]](#)
72. Nasiri, V.; Beloiu, M.; Asghar Darvishsefat, A.; Griess, V.C.; Maftei, C.; Waser, L.T. Mapping tree species composition in a Caspian temperate mixed forest based on spectral-temporal metrics and machine learning. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103154. [\[CrossRef\]](#)
73. De Luca, G.; Silva, J.M.N.; Di Fazio, S.; Modica, G. Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region. *Eur. J. Remote Sens.* **2022**, *55*, 52–70. [\[CrossRef\]](#)
74. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer: New York, NY, USA, 2012; pp. 157–175. ISBN 978-1-4419-9325-0.
75. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
76. Ho, K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [\[CrossRef\]](#)
77. Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using random forests analysis. *Geoderma* **2008**, *146*, 102–113. [\[CrossRef\]](#)
78. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *2/3*, 18–22.
79. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [\[CrossRef\]](#)
80. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, Central Iran. *Geoderma* **2019**, *338*, 445–452. [\[CrossRef\]](#)
81. Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [\[CrossRef\]](#)
82. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Jordan, M., Lawless, J.F., Lauritzen, S.L., Nair, V., Eds.; Springer: New York, NY, USA, 1998; Volume 3, ISBN 978-0-387-98780-4.
83. Huang, Y.; Lan, Y.; Thomson, S.J.; Fang, A.; Hoffmann, W.C.; Lacey, R.E. Development of soft computing and applications in agricultural and biological engineering. *Comput. Electron. Agric.* **2010**, *71*, 107–127. [\[CrossRef\]](#)
84. Tabari, H.; Kisi, O.; Ezani, A.; Hosseinzadeh Talaei, P. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J. Hydrol.* **2012**, *444–445*, 78–89. [\[CrossRef\]](#)
85. Mohandes, M.A.; Halawani, T.O.; Rehman, S.; Hussain, A.A. Support vector machines for wind speed prediction. *Renew. Energy* **2004**, *29*, 939–947. [\[CrossRef\]](#)
86. Sun, Z.; Guo, H.; Li, X.; Lu, L.; Du, X. Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine. *J. Appl. Remote Sens.* **2011**, *5*, 053501. [\[CrossRef\]](#)
87. Foucras, M.; Zribi, M.; Albergel, C.; Baghdadi, N.; Calvet, J.C.; Pellarin, T. Estimating 500-m resolution soil moisture using Sentinel-1 and optical data synergy. *Water* **2020**, *12*, 866. [\[CrossRef\]](#)
88. Charpentier, M.A.; Groffman, P.M. Soil moisture variability within remote sensing pixels. *J. Geophys. Res.* **1992**, *97*, 987–1005. [\[CrossRef\]](#)
89. Mohanty, B.P.; Skaggs, T.H.; Famiglietti, J.S. Analysis and mapping of field-scale soil moisture variability using high-resolution, ground-based data during the Southern Great Plains 1997 (SGP97) hydrology experiment. *Water Resour. Res.* **2000**, *36*, 1023–1031. [\[CrossRef\]](#)
90. Jacobs, J.M.; Mohanty, B.P.; Hsu, E.C.; Miller, D. SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sens. Environ.* **2004**, *92*, 436–446. [\[CrossRef\]](#)
91. de Rosnay, P.; Gruhier, C.; Timouk, F.; Baup, F.; Mougou, E.; Hiernaux, P.; Kergoat, L.; LeDantec, V. Multi-scale soil moisture measurements at the Gourma meso-scale site in Mali. *J. Hydrol.* **2009**, *375*, 241–252. [\[CrossRef\]](#)
92. Western, A.W.; Grayson, R.B.; Blöschl, G.; Willgoose, G.R.; McMahon, T.A. Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resour. Res.* **1999**, *35*, 797–810. [\[CrossRef\]](#)
93. Liu, H.; He, S.; Anenkhonov, O.; Hu, G.; Sandanov, D.; Badmaeva, N. Topography-controlled soil water content and the coexistence of forest and steppe in Northern China. *Phys. Geogr.* **2012**, *33*, 561–573. [\[CrossRef\]](#)
94. Qiu, Y.; Fu, B.; Wang, J.; Chen, L. Soil moisture variation in relation to topography and land use in a hillslope catchment of the Loess Plateau, China. *J. Hydrol.* **2001**, *240*, 243–263. [\[CrossRef\]](#)

95. Liang, W.L.; Li, S.L.; Hung, F.X. Analysis of the contributions of topographic, soil, and vegetation features on the spatial distributions of surface soil moisture in a steep natural forested headwater catchment. *Hydrol. Process.* **2017**, *31*, 3796–3809. [[CrossRef](#)]
96. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084. [[CrossRef](#)]
97. El Hajj, M.; Baghdadi, N.; Zribi, M.; Bazzi, H. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sens.* **2017**, *9*, 1292. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.