Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research xxx (xxxx) xxx

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Soil moisture estimation using combined SAR and optical imagery: Application of seasonal machine learning algorithms

Mohammad Amin Shahriari [a], Hossein Aghighi [a,*], Mohsen Azadbakht [a], Davoud Ashourloo [a], Ali Akbar Matkan [a], Foad Brakhasi [b], Jeffrey P. Walker [b]

[a] Center for Remote Sensing and GIS Research The Faculty of Earth Sciences Shahid Beheshti University Tehran Iran
[b] Department of Civil Engineering Monash University Clayton Australia

Received 15 August 2023; received in revised form 5 December 2024; accepted 27 January 2025

Abstract

Soil moisture plays a crucial role in various fields of geoscience, including agriculture, hydrology, meteorology, and climatology. This study has proposed a seasonal approach to estimate surface soil moisture at 10 m resolution using a combination of Synthetic Aperture Radar (SAR) and optical imagery, coupled with two machine learning algorithms: Support Vector Regression (SVR) and Random Forest (RF). The research utilized simultaneous observations from the Sentinel-1 SAR and Sentinel-2 optical satellites, along with near-surface soil moisture measurements (0–5 cm) from 17 stations in the Yanco study area of Australia, over a four-year period (March 2016 to August 2020). Initially, 270 features were extracted from the satellite data, which were then refined to 31 features per season based on their correlation with near-surface soil moisture. The Sequential Forward Selection (SFS) model, combined with 5-fold cross-validation and the RF algorithm using default hyperparameters, was used to rank the most important features over 100 iterations. The results revealed that important features varied across seasons, with the highest accuracy achieved using only 5 features, resulting in an average lowest Root Mean Square Error (RMSE) of 0.06 ($\pm$0.009) m^3/m^3 and a correlation of 0.65 ($\pm$0.09). The RF method consistently outperformed SVR in soil moisture estimation, with an average RMSE improvement of 0.006 m^3/m^3. Both algorithms achieved their highest accuracy in autumn (RMSE of 0.05 m^3/m^3), with winter proving the most challenging (RMSE of 0.09 m^3/m^3 for SVR and 0.08 m^3/m^3 for RF). These findings highlight the importance of selecting appropriate features for each season to enhance soil moisture estimation using combined SAR and optical remote sensing data with machine learning techniques. This seasonal approach offers promising enhancements in soil moisture monitoring, with a 0.002 m^3/m^3 reduction in RMSE and a 0.03 increase in correlation, potentially advancing our understanding and management of this critical environmental variable across various applications. However, the potential uncertainty introduced by the time lag between Sentinel-1 and Sentinel-2 acquisitions, due to dynamic changes in vegetation water content, should be acknowledged as a limitation.

Keywords: Soil Moisture; Random Forest; Synthetic Aperture Radar; Sentinel-1; Feature Selection; Seasonal Estimation

1. Introduction

Soil moisture is a critical variable in environmental science, including agriculture (Adamchuk et al., 2004, Bousbih et al., 2018), drought monitoring (AghaKouchak et al., 2015, Mishra et al., 2017), water resources manage-

* Corresponding author.
E-mail address: h_aghighi@sbu.ac.ir (H. Aghighi).

https://doi.org/10.1016/j.asr.2025.01.064

Please cite this article as: M. A. Shahriari, H. Aghighi, M. Azadbakht et al., Soil moisture estimation using combined SAR and optical imagery: Application of seasonal machine learning algorithms, Advances in Space Research, https://doi.org/10.1016/j.asr.2025.01.064

ment (Dobriyal et al., 2012), and weather forecasting (Scipal et al., 2008). There are various methods for estimating soil moisture, including ground-based, modeling, and remote sensing technology (Verstraeten et al., 2008). Although ground-based methods can provide highly accurate soil moisture estimates with high temporal resolution, and are considered a reliable benchmark for validating other techniques, their low spatial distribution restricts their applicability in many contexts. These methods are also costly and time-consuming when implemented over large spatial scales (Zhang et al., 2018). Hydrological and land surface models are commonly used to estimate soil moisture from weather variables such as precipitation, evapotranspiration, and infiltration (Lin et al., 2024, Mondal and Mishra, 2024). However, these models rely on numerous assumptions and often produce uncertain results. Additionally, the input data for these models are not always valid or readily available.

Remote sensing technology is widely used for estimating soil moisture at larger scales, with various parts of the electromagnetic spectrum, including optical (Sadeghi et al., 2017), thermal (PRICE, 1980), and microwave (Ezzahar et al., 2019, Sabaghy et al., 2018) region having been used to estimate soil moisture at the global scale. However, microwave remote sensing has proven to be the most promising approach for soil moisture estimation due to its all-weather capability, its ability to operate both day and night, and its high sensitivity to soil moisture variation within the soil medium (Ulaby et al., 1996, Filgueiras et al., 2019) given the large dielectric constant difference between water and dry soil. Compared to passive microwave, Synthetic Aperture Radar (SAR) can provide soil moisture information at higher spatial resolution (Barrett et al., 2009, Shi et al., 1997, Loew and Mauser, 2006), making it more suitable for localized applications such as farmland activities. However, the estimation of soil moisture from SAR observations is influenced by a wide range of ground parameters, including surface roughness and Vegetation Water Content (VWC), typically resulting in lower accuracy than from passive microwave (Paloscia et al., 2013). Importantly, some studies have successfully combined SAR and optical data for soil moisture estimation, with an improved accuracy in the retrieval process (Prakash et al., 2011, Bao et al., 2018).

Several methods have been proposed to estimate soil moisture from combined SAR and optical remote sensing observations, including physically-based inversions, simple linear regressions, and advanced machine learning. Inversion methods, which use Radiative Transfer Models (RTMs), offer the advantage of providing accurate soil moisture estimates across large scales due to their physically based approach. However, they can be complex and computationally intensive, requiring accurate ancillary data and calibration (Barrett et al., 2009). Simple regression models assume a linear relationship, which may not capture the complex interactions in remote sensing data, leading to potential inaccuracies. They often fail to account for multiple factors like surface roughness and vegetation, and can be sensitive to outliers (Paloscia et al., 2013). Additionally, these models may not generalize well across different environments and rely heavily on the quality and quantity of input data. Machine learning algorithms excel in handling large and complex datasets, enabling them to uncover intricate patterns and relationships that simple models might miss (Shami et al., 2024). They are highly adaptable and can therefore improve accuracy as more data becomes available (Abdalla et al., 2024).

With the abundance of data sources today, such as intensive soil moisture measurements from the Yanco study site in Australia, and free Sentinel-1 and Sentinel-2 images, machine learning algorithms can potentially leverage these rich datasets to enhance predictive performance and robustness, making them powerful tools for advanced soil moisture estimation. Accordingly, several studies have employed machine learning algorithms for estimating surface soil moisture, including Support Vector Regression (SVR) (Ahmad et al., 2010), Artificial Neural Network (ANN) (Paloscia et al., 2013), Random Forest (RF) (Hajdu et al., 2018), and eXtreme Gradient Boosting (XGBoost) (Ezzahar et al., 2023). While some of these studies utilized optical and thermal observations (Adab et al., 2020), others have specifically employed active microwave observations, including active-only (Ezzahar et al., 2019, Bhogapurapu et al., 2022, Chaudhary et al., 2022, Hou et al., 2024), or a combination of active and optical observations (Attarzadeh et al., 2018, Wang and Gao, 2023). Some studies have also compared machine learning algorithms with physical models, finding that machine learning algorithms are more suitable for soil moisture estimation due to the difficulty of obtaining ancillary variables, such as surface roughness and vegetation water content, at large scale (Ezzahar et al., 2019). Furthermore, comparisons among various machine learning algorithms have shown that RF and SVR are typically the most effective.

Although these studies demonstrated that combining radar and optical data is more effective for estimating soil moisture than using radar data alone, they have often utilized a limited number of radar and optical features. Moreover, recent research by Li et al. (2024) emphasized the need to use a broad range of optical indices for estimating soil moisture. Importantly, none of these studies have investigated the impact of using different sets of features during different seasons. Accordingly, this study has addressed these limitations by generating a variety of SAR and optical features from Sentinel-1 and Sentinel-2 satellites, respectively, including a proposed new index, (VV-VH)/VV, and evaluating different subsets of features in each season. Consequently, a seasonal approach was employed for estimating surface soil moisture in New South Wales, Australia, using RF and SVR algorithms.

2. Experimental data

This study utilized SAR measurements from the Sentinel-1 satellite and optical measurements from the Sentinel-2 satellite, along with surface soil moisture data collected from ground stations. The data were collected at the Yanco study site in New South Wales, Australia, covering the period from March 2016 to August 2020 (Fig. 1). The Yanco study site is located within the Murrumbidgee River catchment, with its landscape predominantly characterized by pasture, crops, and woody vegetation (Smith et al., 2012). The region's topography is relatively flat, with elevations ranging from 117 to 150 m, and its soil composition includes clay, red-brown earth, transitional red-brown earth, sand over clay, and entirely sandy land (Yee et al., 2016). The site experiences varying average daily temperatures throughout the year, ranging from 13 °C in July to 32 °C in January, with an average annual precipitation of 418 mm, most of which occurs during the late autumn and winter months (Yee et al., 2016). A total of 17 soil moisture stations (as shown in Fig. 1) were considered, which measured surface soil moisture at a depth of 0–5 cm in 20-minute intervals. The 20-minute time lag for Sentinel-1 refers to the maximum difference between satellite acquisition and in-situ data collection, ensuring near-simultaneous validation. A total of 95 Sentinel-1 SAR images were used in this study, acquired in Interferometric Wide Swath (IW) mode with VV and VH polarization and incident angles ranging from 35° to 38° (https://www.asf.alaska.edu/). These images, processed to 10 × 10 m Ground Range Detected (GRD) products, were derived from raw Single-Look Complex (SLC) data (5 × 20 m resolution) through multi-looking, geocoding, and subswath stitching (Small and Schubert, 2019). Additionally, 95 Sentinel-2 Level-1C images, covering 10 spectral bands in VISible (VIS), Near-InfraRed (NIR), and ShortWave InfraRed (SWIR) with spatial resolutions of 10, 20, and 60 m, were acquired. Sentinel-2 images were acquired with a time lag of less than 5 days relative to Sentinel-1, balancing data availability with the irregular imaging schedules over the Yanco region in the Southern Hemisphere. This time lag introduces uncertainties due to the dynamic nature of VWC, which can vary substantially over short timescales influenced by diurnal changes, rainfall, irrigation, and temperature. While a time lag of less than five days between Sentinel-1 and Sentinel-2 acquisitions aims to reduce mismatches, it does not fully account for the potential variability in VWC, which may affect the accuracy of the model.

Over the four-year period, a total of 1303 quality samples were collected (Table 1). The dataset was divided into different seasons: spring (September, October, and

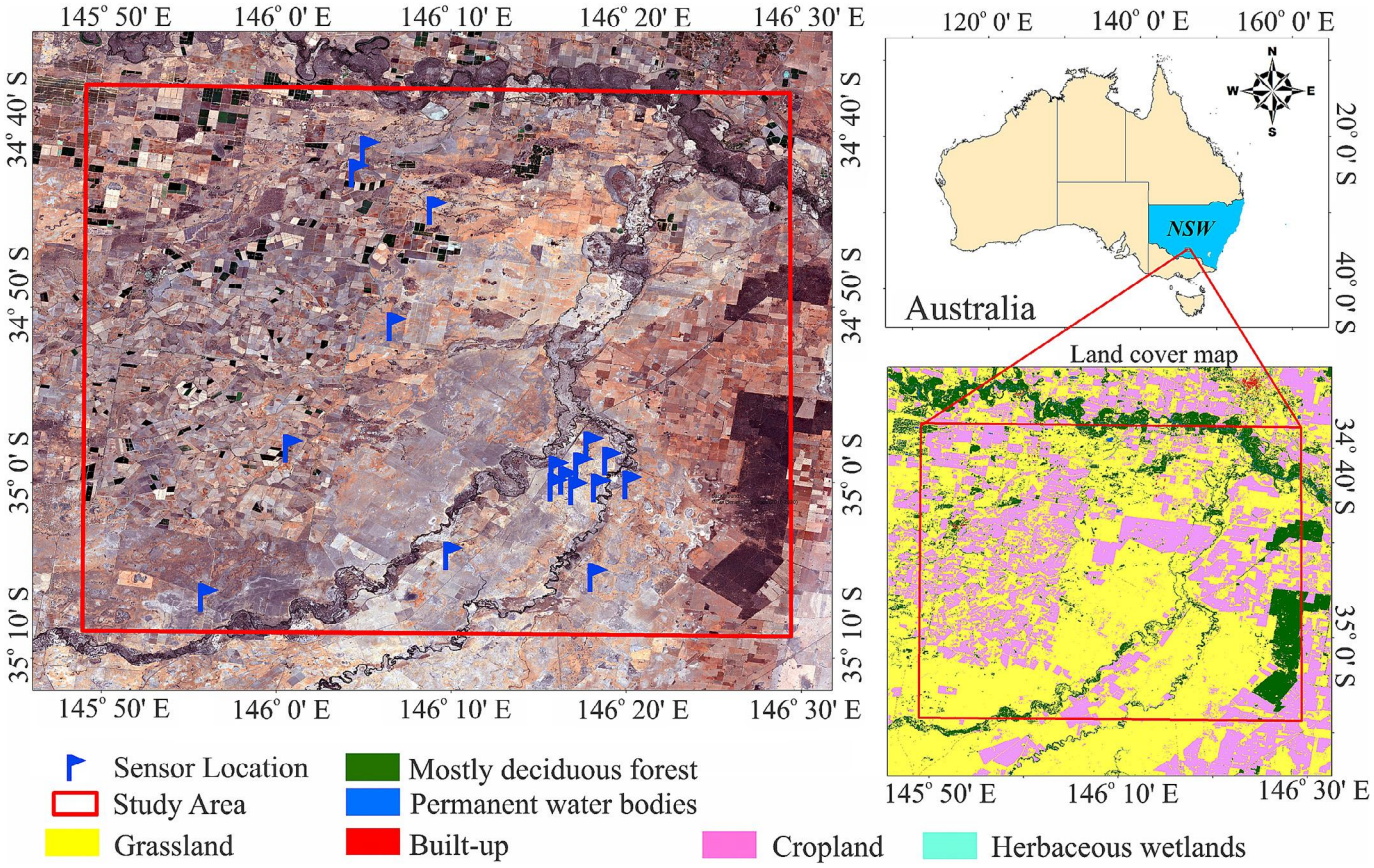

Fig. 1. Location of the study area and the soil moisture stations. The background is a mosaic of Sentinel-2 satellite images from January 2020.

Table 1
Mean values of soil moisture, NDVI, and number of collected samples for each season.

Dataset	Mean soil moisture (m^3/m^3)	Mean NDVI	Number of samples
Spring	0.13	0.21	263
Summer	0.1	0.15	343
Autumn	0.12	0.17	416
Winter	0.18	0.37	281
Entire period	0.13	0.22	1303

November), summer (December, January, and February), autumn (March, April, and May), winter (June, July, and August), and the entire four-year dataset. Seasonal variations in mean soil moisture and NDVI are shown in Table 1. Winter had the highest mean soil moisture (0.18 m^3/m^3) and NDVI (0.37), indicating higher vegetation accompanying the higher moisture content during this season. In contrast, summer exhibited the lowest mean soil moisture (0.10 m^3/m^3) and NDVI (0.15), reflecting the dry conditions typical of this season. Autumn had the largest number of samples (416) due to higher number of in suit soil moisture, while the other seasons had an average of ~290 samples each. This dataset was utilized to evaluate the performance of two commonly used machine learning algorithms, to estimate surface soil moisture from combined SAR and optical imagery.

3. Methodology

3.1. Machine learning algorithms

The SVR and RF machine learning algorithms were employed to estimate surface soil moisture using combined Sentinel-1 SAR and Sentinel-2 optical imagery (Fig. 2). The SVR algorithm is an extension of the Support Vector Machine (SVM), developed by Vapnik in the 1990s for regression problems (Vapnik and Vapnik, 1998, Vapnik, 1995). It is a supervised learning method designed to find an optimal hyperplane in a high-dimensional feature space. The SVR uses kernel functions to map data into a higher-dimensional space, enabling nonlinear modeling of complex relationships (Cristianini and Shawe-Taylor, 2000). The RF algorithm, developed by (Breiman, 2001), is an ensemble learning algorithm that combines decision trees and bagging for classification and regression tasks. It builds multiple decision trees using random subsets of features and data, enhancing model robustness and accuracy. Each tree partitions the feature space with hierarchical rules to classify data or make predictions. Bagging generates diverse training data by sampling with replacement, which reduces correlation among trees and improves stability and accuracy (Carranza et al., 2021). For both SVR and RF, a GridSearchCV method with 5-fold cross-validation was employed to determine the optimal hyperparameters using five datasets. In the study by Pasolli et al. (2011a) the SVR parameter ranges for C, epsilon, and gamma were set to [0.001:1: 1000], [0.0001:1:1000], and [0.0001:0.01:10], respectively, and this research followed the same approach. Similarly, for the RF model, this study adopted the parameter ranges for the number of trees and maximum depth from Carranza et al. (2021) which were set within [100:1:1000] and [2:1:10], respectively. The Radial Basis Function (RBF) kernel was chosen for SVR due to its superior performance in fields like hydrology and soil moisture estimation (Scholkopf et al., 1997, Asefa et al., 2006, Gill et al., 2006), owing to its ability to minimize expected error limits by automatically adjusting thresholds and weights. These intervals were used to fine-tune the parameters of the SVR and RF algorithms in this study. The list of tuned hyperparameters for RF and SVR is shown in Table 2.

3.2. Feature engineering

The satellite images including SAR and optical were processed and various features generated. Pre-processing steps for Sentinel-1 SAR images included orbital correction, Range-Doppler terrain correction, Sigma Nought conversion, and speckle noise reduction. The Lee sigma filter (LEE, 1983) with a 5×5 window was used for speckle noise reduction, with its effectiveness in soil moisture estimation demonstrated by Liu et al. (2020). For Sentinel-2 optical images, atmospheric correction (Yin et al., 2019) was performed. This involved converting Top of Atmosphere (TOA) reflectance to Bottom of Atmosphere (BOA) reflectance using the Sensor Invariant Atmospheric Correction (SIAC) function. Additionally, cloudy pixels were detected and removed using the QA60 band (Sun et al., 2019, Aybar et al., 2022). The final output images from Sentinel-1 and Sentinel-2 had a pixel size of 10×10 m. The Sentinel Application Platform (SNAP) software version 8.0.0 and Google Earth Engine were utilized to process the Sentinel-1 and Sentinel-2 images respectively. The general procedure followed in this study is illustrated in Fig. 2, depicting the flowchart of the overall process.

Feature engineering, including feature generation and selection, is one of the most important steps in the application of machine learning algorithms to remote sensing. For the purpose of soil moisture estimation, eight features were generated from Sentinel-1 SAR observations utilizing VV and VH polarizations. Previous research has highlighted various features derived from dual-polarization radar data as effective for accounting for vegetation and soil properties (Gorrab et al., 2021, Ouaadi et al., 2020, Ayari et al., 2023). Additionally, the local incidence angle was considered as one feature. Furthermore, 249 features (see list given at https://www.indexdatabase.de/db/is.php?sensor_id=96) were generated using ten bands of Sentinel-2 optical images, including blue, green, red, Vegetation Red Edge (VRE) 5, VRE 6, VRE 7, Near-InfraRed (NIR), Narrow NIR, Short-Wave Infra-Red (SWIR) 11, and SWIR 12.

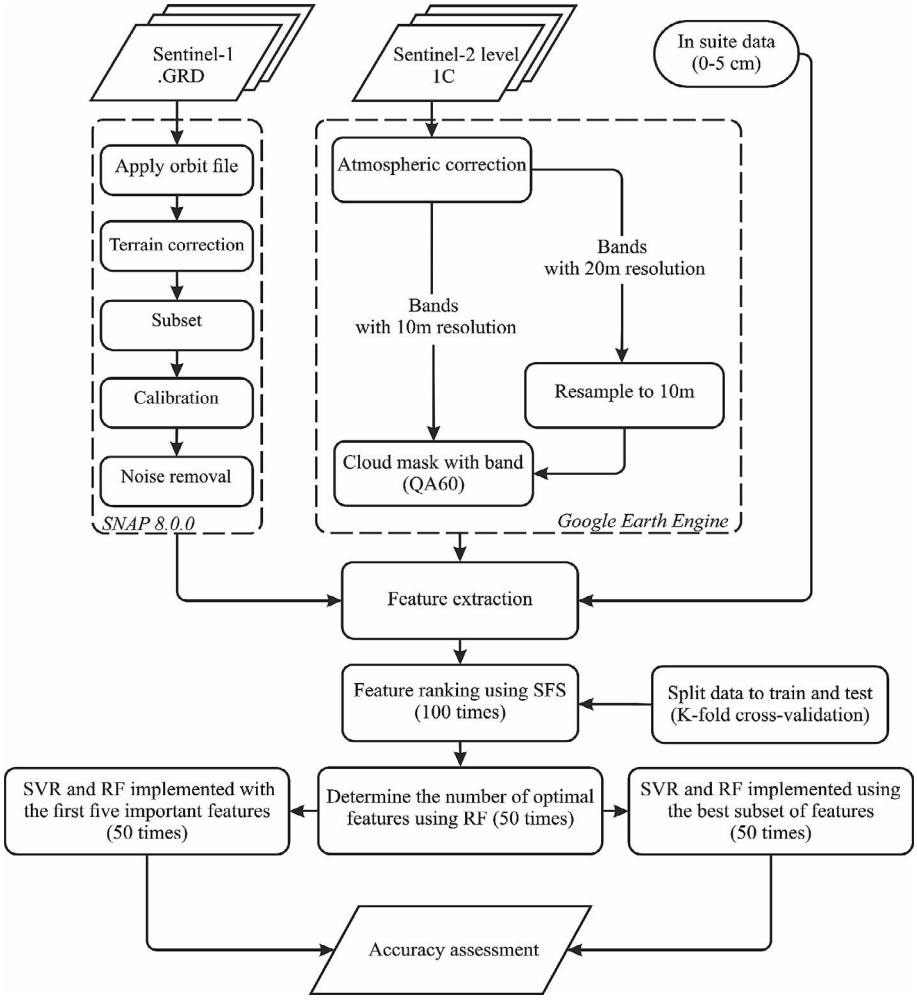

Fig. 2. Workflow for soil moisture estimation using machine learning algorithms with combined Sentinel-1 SAR and Sentinel-2 optical data.

Table 2
Tuned hyperparameters for RF and SVR algorithms.

	RF			SVR			
Dataset	Number of trees	Max depth	Max features	C	epsilon	Gamma	Kernel
Spring	160	10	Sqrt	901	6.5	0.001	RBF
Summer	600	10	Sqrt	503	3.6	0.1	RBF
Autumn	250	10	Sqrt	750	4.4	0.01	RBF
Winter	150	10	Sqrt	26	4.8	1	RBF
Entire period	150	10	Sqrt	610	5.9	0.2	RBF

A correlation analysis was performed between each of the 249 features and surface soil moisture, both seasonally and annually. The top ten features for each season were retained based on these analyses. This process ensured that the selected features had a strong correlation with soil moisture data, enhancing the model's accuracy. In total, 31 features were considered: 11 fixed features from Sentinel-1 SAR data, 10 fixed bands from Sentinel-2 optical observations, and the top ten indices calculated from Sentinel-2 optical data, which vary with each season. The list of these features is given in Tables 3 and 4 for Sentinel-1 and Sentinel-2, respectively.

Employing a large number of features in machine learning algorithms reduces performance, as the model may learn noise instead of meaningful patterns, and the presence of many unnecessary and irrelevant features in high-dimensional data further diminishes the model's ability to generalize to new data (Zebari et al., 2020). Feature selection aims to identify the most relevant subset of features from the entire set of available features, enhancing computational efficiency, especially for large datasets (May et al., 2010). The Sequential Forward Selection (SFS) (Whitney, 1971) was employed to rank and select the most important subset of the features for each dataset, with the goal being

Table 3
Features extracted from Sentinel-1 SAR data. Wi: Winter; Sp: Spring; Su: Summer; Au: Autumn; LIA: Local Incident Angle; NRPB: Normalized Ratio Procedure between Bands; PR: Polarization Ratio; RVI: RADAR Vegetation Index; VDDPI: Vertical Dual De-Polarization Index.

ID	Feature	Type	Usage	Reference
1	VV	Measured	All	Ulaby et al., 1978
2	VH	Measured	All	Ulaby et al., 1996
3	LIA	Measured	All	Pasolli et al., 2015
4	X	Calculated	All	Ulaby et al., 1996
5	VH + VV	Calculated	All	Hoskera et al., 2020
6	NRPB	Calculated	All	Filgueiras et al., 2019
7	PR1	Calculated	All	Mandal et al., 2020
8	PR2	Calculated	All	Della Vecchia et al., 2008
9	RVI	Calculated	All	Nasirzadehdizaji et al., 2019
10	VVDDPI	Calculated	All	Periasamy, 2018
11	VDDPI 2	Calculated	All	–

to find the optimal subset of features that maximizes the accuracy for performance of the model (Pudil et al., 1994). The SFS algorithm starts by individually evaluating each feature to identify the one that yielded the greatest improvement in model performance, which was then added to the feature subset. In each subsequent iteration, the algorithm evaluated the remaining unselected features and added the feature that most enhanced the model's performance when combined with the already selected features. This process continued until adding further features no longer resulted in a substantial improvement in model accuracy or performance (Aha and Bankert, 1994, Aha and Bankert, 1995).

The SFS algorithm, using the RF algorithm as a regressor with default hyperparameters, was executed 100 times for each seasonal dataset: spring, summer, autumn, winter,

Table 4
Features selected from Sentinel-2 optical data. ARI: Anthocyanin Reflectance Index; ARVI2: Atmospherically Resistant Vegetation Index 2; MTVI1: Modified Triangular Vegetation Index 1; GNDVIhyper: Normalized Difference 780/550 Green NDVI hyper; NDMI (Normalized Difference 820/1600 Normalized Difference Moisture Index); SIWSI: Normalized Difference 860/1640; NDVI: Normalized Difference Vegetation Index; NDVIre: Normalized Difference Vegetation Index red-edge; BRI: Browning Reflectance Index; Chlgreen: Chlorophyll Green; CIgreen: Chlorophyll Index Green; D678/500 Difference 678/500; GVMI: Global Vegetation Moisture Index; GLI: Green leaf index; GNDVI: Green Normalized Difference Vegetation Index; GBNDVI: Green-Blue NDVI; GRNDVI: Green-Red NDVI; MCARI: Modified Chlorophyll Absorption in Reflectance Index;MSR670: Modified Simple Ratio 670,800; SRNIR/MIR: Simple Ratio NIR/MIR; SLAVI: Specific Leaf Area Vegetation Index.

ID	Feature	Type	Usage	Reference
12	Sentinel-2 Bands	Measured	All	–
13	Alteration	Calculated	Sp	Volesky et al., 2003
14	ARI	Calculated	Sp, Au	Gitelson et al., 2003
15	ARVI2	Calculated	WY, Su	Kaufman and Tanre, 1992
16	MTVI1	Calculated	Su	Haboudane et al., 2004
17	GNDVIhyper	Calculated	Su	–
18	NDMI	Calculated	Sp	Xiao et al., 2019
19	SIWSI	Calculated	Sp	Fensholt and Sandholt, 2003
20	NDVI	Calculated	All	Rouse Jr et al., 1974
21	NDVIre	Calculated	Wi, Su	Forkuor et al., 2020
22	NormNIR	Calculated	Au	–
23	BRI	Calculated	WY, Su	Chivkunova et al., 2001
24	Chlgreen	Calculated	WY	–
25	CIgreen	Calculated	WY, Au	Hunt Jr et al., 2011
26	CRI550	Calculated	Au	Gitelson et al., 2001
27	Datt4	Calculated	Sp	–
28	Datt6	Calculated	WY, Sp, Au	–
29	D678/500	Calculated	Wi	–
30	GVMI	Calculated	WY, Sp	Glenn et al., 2010
31	GLI	Calculated	Wi	Hunt Jr et al., 2011
32	GNDVI	Calculated	Au	Ahamed et al., 2011
33	GBNDVI	Calculated	Au	Wang et al., 2007
34	GRNDVI	Calculated	Su	Main et al., 2011
35	LogR	Calculated	WY, Su, Au	–
36	MCARI	Calculated	Wi	–
37	MSR670	Calculated	Au	Pu et al., 2008
38	NormR	Calculated	Wi	–
39	SRNIR/MIR	Calculated	Wi	Malthus et al., 1993
40	SLAVI	Calculated	Sp	Lymburner et al., 2000
41	SQRT(IR/R)	Calculated	Wi	–
42	STVI1	Calculated	WY, Wi	Thenkabail et al., 1994
43	STVI2	Calculated	Sp	Thenkabail et al., 1994
44	STVI3	Calculated	WY, Su	Thenkabail et al., 1994

and the entire period. Each execution utilized 5-fold cross-validation. In each iteration, the algorithm produced a combination of features with varying numbers of features. The frequency of each feature was calculated and ranked from most important to least important. An iterative approach was employed to determine the optimal number of features for each dataset (spring, summer, autumn, winter, and the entire period), with the RF algorithm using the ranked features from the previous stage. The process involved gradually adding features to the RF model and evaluating their impact on the accuracy of soil moisture estimation. Starting from the first ranked feature with the highest frequency for each dataset, the RF algorithm with tuned hyperparameters was executed 50 times using 5-fold cross-validation. The data was shuffled and divided into training and testing sets for each iteration. This process was repeated 50 times for the first ranked feature and the average of the 50 RMSE and R values calculated. The next step included selecting the first and second ranked features, and the process repeated as before until the last feature was included. The optimal number of features for each dataset was determined by identifying the point at which adding an additional feature did not increase the R value or decrease the RMSE.

3.3. Soil moisture estimation

The RF and SVR algorithms were employed to estimate soil moisture using two approaches: (i) selecting the number of features determined by the RF algorithm, and (ii) fixing the top 5 important features identified by the SFS algorithm. In the feature ranking process using the SFS algorithm, it was found that, on average, the first 5 features had the highest frequency of importance. The objective of this comparison was to evaluate whether determining the number of features using the RF algorithm, which is computationally expensive, is necessary. Similar to the feature selection, in estimating soil moisture using each of the approaches, the data was divided into training and testing using 5-Fold cross validation. To ensure robustness and reliability, each of the employed machine learning algorithms were executed 50 times by shuffling the data in each iteration. This process was executed for each dataset, including spring, summer, autumn, winter, and the entire dataset. In this study, RMSE, Bias, and R were used as mathematical criteria to evaluate the results of the RF and SVR algorithms. Additionally, the statistical criterion Analysis of Variance (ANOVA) was employed with a confidence level of $\alpha = 5$ % (Devore et al., 2021). The performance of the SVR and RF methods was then compared for each season and for the entire dataset.

The null hypothesis, H0, posits that the performance of SVR and RF algorithm is similar across different seasons or the entire dataset, while the alternative hypothesis, H1, suggests that a significant difference in performance exists. The hypotheses are represented as follows:

$$\begin{cases} H0: \ \mu\ SVR = \mu\ RF for\ \ each\ \ dataset \\ H1: \ \ at\ \ least\ \ one\ \ is\ \ different \end{cases}$$

where μ denotes the mean.

4. Results and discussion

4.1. Feature selection results

From the 31 features employed for each season and the entire dataset (Tables 3 and 4), the results of the feature ranking by the SFS algorithm showed that only 30, 27, 29, 22, and 26 features for spring, summer, autumn, winter, and the entire dataset, appeared in the feature combinations (Fig. 3), respectively. The other features were not important and therefore did not appear in the final results. Although the LIA is not directly tied to the physical properties of soil moisture, it has a significant impact on the backscatter signals in SAR images, which in turn influences the accuracy of soil moisture estimation. LIA primarily affects how SAR signals interact with the surface, rather than contributing directly to the estimation of soil moisture. However, due to the spatial distribution of soil moisture stations in the study area, variations in near and far incidence angles become particularly important for interpreting backscatter signals, underscoring LIA's indirect but crucial role in soil moisture estimation across all datasets. The VRE5 band emerged as the next most important feature, especially during the spring season, underscoring its relevance in soil moisture estimation. This contrasts with previous studies, which have primarily focused on using the red band and NIR for soil moisture estimation (Amani et al., 2016, Zhan et al., 2007). The importance of VRE5 has been largely unexplored due to its absence in most sensors. The SWIR2 and VV bands also demonstrated high importance, with frequencies of 83 % and 80 %, respectively, due to their moisture sensitivity and backscatter effects. The importance of SWIR (Babaeian et al., 2018, Sadeghi et al., 2015) and VV (El Hajj et al., 2019, Dubois et al., 1995) in soil moisture estimation has been emphasized in previous studies. Additionally, the innovative (VV-VH)/VV index proved effective for soil moisture estimation in the overall dataset.

In spring, NDVI was the next most frequently selected feature, appearing in all of the iterations. The VRE5 band was also prominent, emphasizing its importance during spring when vegetation growth is substantial. Other important indices in spring included the ARI and VH + VV index, highlighting the combined effectiveness of Sentinel-1 and Sentinel-2 data in capturing soil and vegetation interactions. In the summer season, characterized by low vegetation (NDVI = 0.15) and low soil moisture (SM = 0.1 m^3/m^3) as shown in Table 1, the blue band emerged as a crucial feature, appearing in 81 % of iterations. Its sensitivity to atmospheric conditions makes it particularly effective during the drier summer months, despite being less useful in

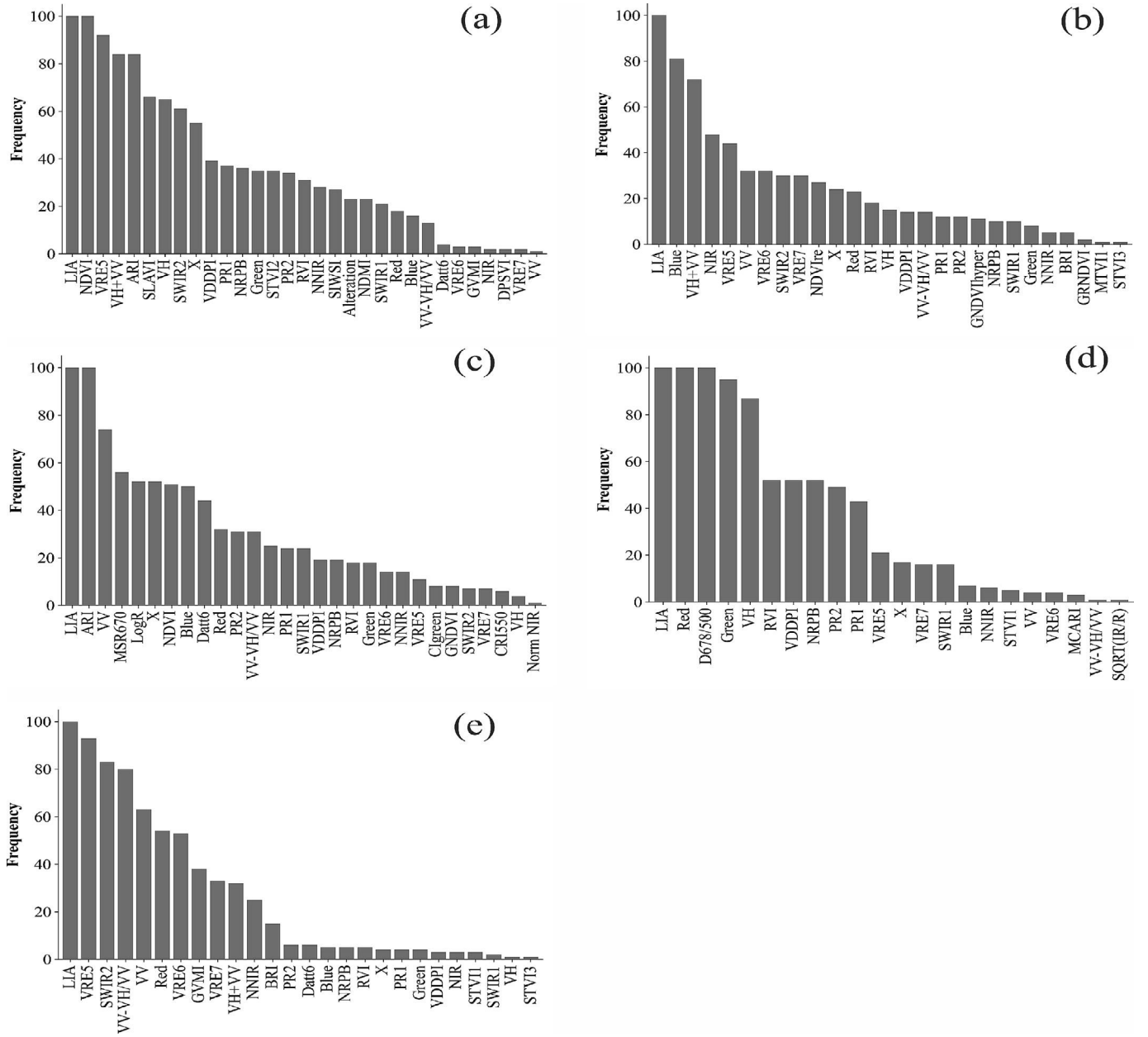

Fig. 3. Frequency of the best performance features provided by the SFS algorithm in: (a) spring, (b) summer, (c) autumn, (d) winter, and (e) entire dataset.

other seasons due to its susceptibility to atmospheric water vapor. The reduced precipitation, evaporation, transpiration, and soil moisture in summer enhance the blue band effectiveness. The VH + VV index, along with the NIR and VRE5 bands, also proved important for accurate soil moisture estimation during this season. These findings underscore the significance of integrating multi-sensor data to improve soil moisture estimation in summer, leveraging both optical and radar indices to address the unique challenges posed by summer's dry conditions.

In autumn, LIA and ARI were consistently identified as the most important features. These, along with VV backscatter, were crucial because the reduced vegetation cover during autumn allows for more direct interactions with the soil surface. Autumn and summer have similar soil moisture and vegetation conditions (Table 1), which accounts for the frequent selection of ARI and the MSR670 index from Sentinel-2 data. With low vegetation levels in autumn, VV backscatter was more important than VH backscatter, as it primarily reflects signals from the ground (McNairn and Brisco, 2004). The X index, calculated as VH-VV, was also effective in capturing soil moisture dynamics, further highlighting its usefulness in assessing ground conditions. The top features identified for the winter season were LIA, Red, and D678/500 indices, each appearing in all of the iterations. The D678/500 index, derived from subtracting the red band from the blue band, is particularly sensitive to vegetation

chlorophyll, which makes it important in winter when vegetation is prevalent. The winter season recorded the highest NDVI (0.37) and soil moisture content (0.18 m^3/m^3), indicating higher vegetation and moisture presence compared to other seasons. The red and blue bands are strongly influenced by plant chlorophyll and are absorbed by vegetation, while the green band, frequently observed in the winter dataset, was mainly reflected during the winter. Studies have suggested that vegetation indices, which reflect properties such as vegetation growth, biomass, or stress levels, can provide indirect estimates of soil moisture content, particularly in scenarios where the soil surface is not visible (Petropoulos et al., 2015, Wang et al., 2010, Xiao et al., 1994). These indices are influenced by the relationship between vegetation and available soil moisture, though it is important to note that this relationship is complex and context-dependent. Surface soil moisture can fluctuate rapidly due to environmental interactions, while vegetation response, linked to root zone moisture, may remain more stable over time. In winter with the presence of vegetation (NDVI = 0.37), VH backscatter was dominant due to its sensitivity to volumetric scattering mechanisms, which are influenced by the structural characteristics of the vegetation and the scene. Previous research, such as Vreugdenhil et al. (2018) has emphasized that VH backscatter often dominates in vegetated environments, though it primarily reflects volumetric properties rather than direct soil moisture information. In contrast, VV backscatter, which interacts more with the soil surface, especially in areas with sparse vegetation or bare soil. Studies using combined SAR and optical data for soil moisture estimation (Pasolli et al., 2011b, Ahmad et al., 2010, El Hajj et al., 2019) have also emphasized the crucial importance of NDVI, though relying on a single feature across different environmental conditions can lead to inaccuracies due to variations in water content, vegetation, and soil parameters. The NDVI, while important, has limitations such as saturation and susceptibility to atmospheric and soil conditions (Weiss et al., 2020), explaining its lower frequency in most seasons except spring. While some features, like LIA, consistently ranked highest across all datasets, the effectiveness of SAR and optical features varied by season; SAR was more effective in winter and autumn, whereas optical features were more beneficial in spring and summer. Overall, combining both SAR and optical data is essential for achieving the most accurate soil moisture estimation throughout the year.

The number of important features using the RF algorithm is shown in Fig. 4, which illustrates the average performance of the algorithm across 50 iterations for each dataset. On average, using the top five ranked features in each dataset provided the highest accuracy for soil moisture estimation, while adding the sixth feature did not substantially improve the accuracy. Therefore, the two approaches were compared for training the RF and SVR algorithms based on: (i) the recommended number of features with the lowest RMSE (Fig. 5a and b), and (ii) the five top-ranked features with the highest frequency (Fig. 5c and d). Both approaches achieved similar accuracy, with an average difference of 0.0008 m^3/m^3 across all of the datasets, indicating that additional features do not substantially enhance the accuracy of soil moisture estimation, and thus supporting the use of the second approach.

4.2. Soil moisture estimation results

The RF and SVR algorithms were executed 50 times each, utilizing the five most important features from each dataset (refer to Fig. 5c and d). An ANOVA test was conducted to evaluate the differences in performance between the RF and SVR algorithms across various datasets. This statistical test was applied to the RMSEs obtained from 50 iterations of the SVR and RF models. The null hypothesis posited that the means of the RMSEs for the two models were equal, with a p-value of less than 0.05 considered indicative of a statistically significant difference, thus leading to the rejection of the null hypothesis. The results indicated that only the RF model, when applied to the entire dataset and spring dataset, had p-values greater than 0.05 ($p = 0.12$), suggesting no significant differences in performance for these datasets. However, for all other datasets, the p-values were less than 0.05, indicating significant differences and leading to the rejection of the null hypothesis. The RF algorithm was found to consistently outperform the SVR algorithm across all datasets. Specifically, the RF algorithm achieved reductions in RMSE (and increases in R) by 0.005 m^3/m^3 (0.1) in spring, 0.004 m^3/m^3 (0.12) in summer, 0.006 m^3/m^3 (0.08) in autumn, 0.01 m^3/m^3 (0.14) in winter, and 0.01 m^3/m^3 (0.12) across the entire dataset. Therefore, in the following analysis, the focus is placed on the result of the RF algorithm.

The average RMSE (R) in Fig. 5d (Fig. 6b) for the autumn dataset was found to be lower than that of the other datasets, with values of 0.05 m^3/m^3 (0.78) when the RF algorithm was used. Summer was the second, with mean RMSE and R values of 0.06 m^3/m^3 and 0.50, respectively. The accuracy of estimated soil moisture for the spring and winter datasets was found to be lower compared to the summer and autumn datasets, with RMSE of 0.06 m^3/m^3 and R of 0.60 for spring, and RMSE of 0.08 m^3/m^3 and R of 0.70 for winter when the RF algorithm was employed. There are two possible reasons for having higher RMSE for winter and spring as compared with summer and autumn. The first reason is that summer and autumn exhibited the lowest mean NDVI and soil moisture values compared to spring and winter datasets. SAR signals can penetrate the canopy and come from deeper layers of the soil, which represent the average 0–5 cm soil moisture recorded by the station. However, with a denser canopy in winter and spring, the radar backscatter has reduced interaction with the soil surface, resulting in radar and optical data that primarily reflect the canopy characteristics. The second reason is the number of representative

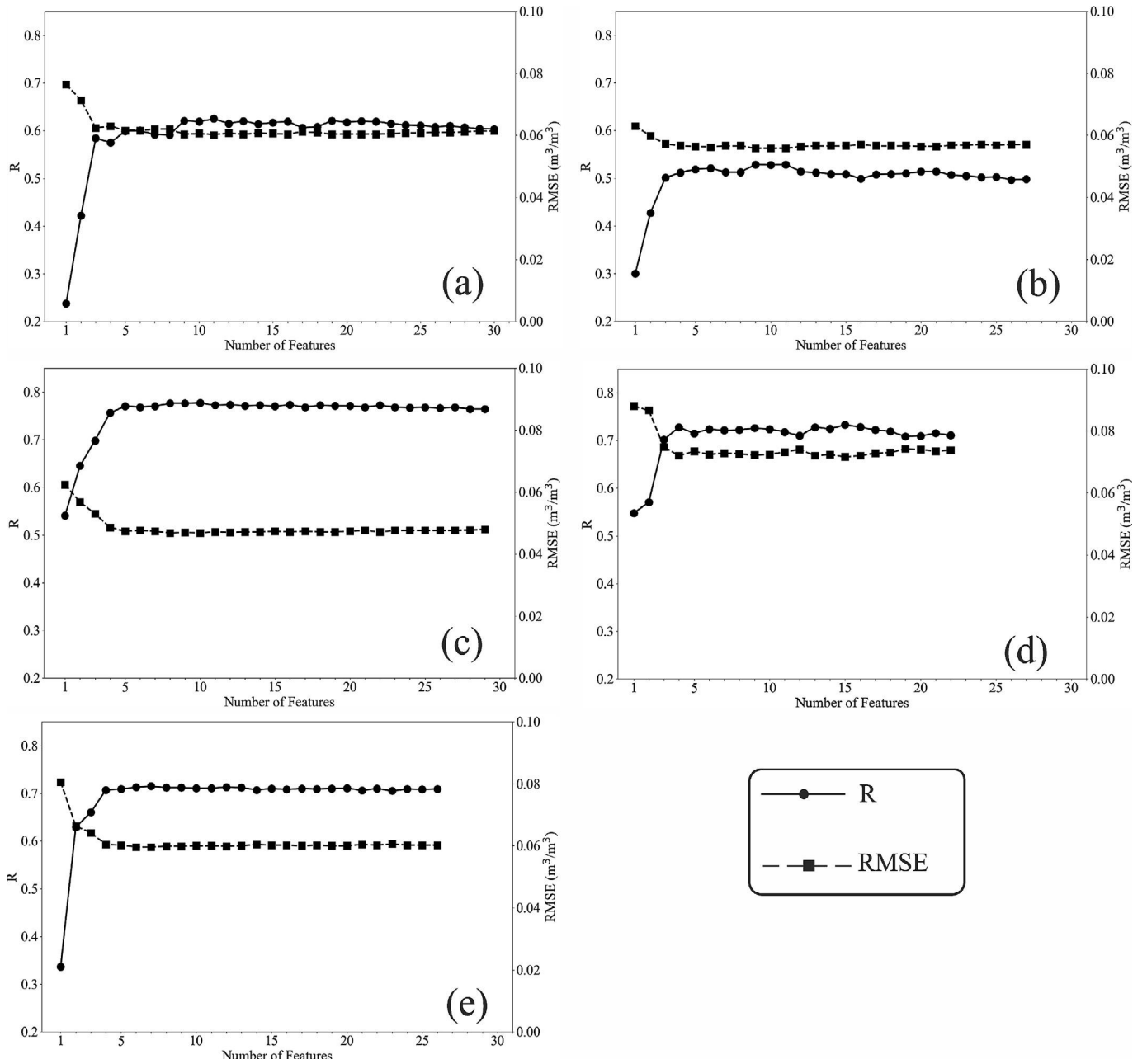

Fig. 4. Number of optimal features obtained from the RF algorithm based on RMSE (m^3/m^3) and R for: (a) spring, (b) summer, (c) autumn, (d) winter, and (e) entire period (the lowest RMSE value indicates the best accuracy in the number of features).

training samples (Fig. 6b), which was higher during summer and autumn as compared with winter and spring datasets. It was also evident that the standard deviation of the accuracy of soil moisture estimation during winter was higher (0.02) than that for the other datasets, being followed by spring (0.02), autumn (0.01) showing the greater robustness of the RF algorithm during summer and autumn as compared with winter and spring.

Fig. 7 presents the best result from 50 runs of the RF algorithm, showing a positive correlation between observed and estimated soil moisture. The results show that as soil moisture increased, accuracy decreased. This decline can be attributed to two factors: (i) reduced sensitivity of backscatter to changes in soil moisture at higher levels, leading to saturation (the phenomenon of SAR data becoming saturated for soil moisture above 0.25 m^3/m^3 was confirmed by Zribi et al. (2013); and (ii) the insufficiency of training data for soil moisture levels above 0.25 m^3/m^3, as only 7 % of the dataset falls within this range. The model performance across the entire datasets showed a good fit between the observed and estimated soil moisture values with the best estimate in autumn and the worst estimate in winter.

The models used in this study were calibrated and validated based on the specific environmental conditions of the study area, which included variations in vegetation cover,

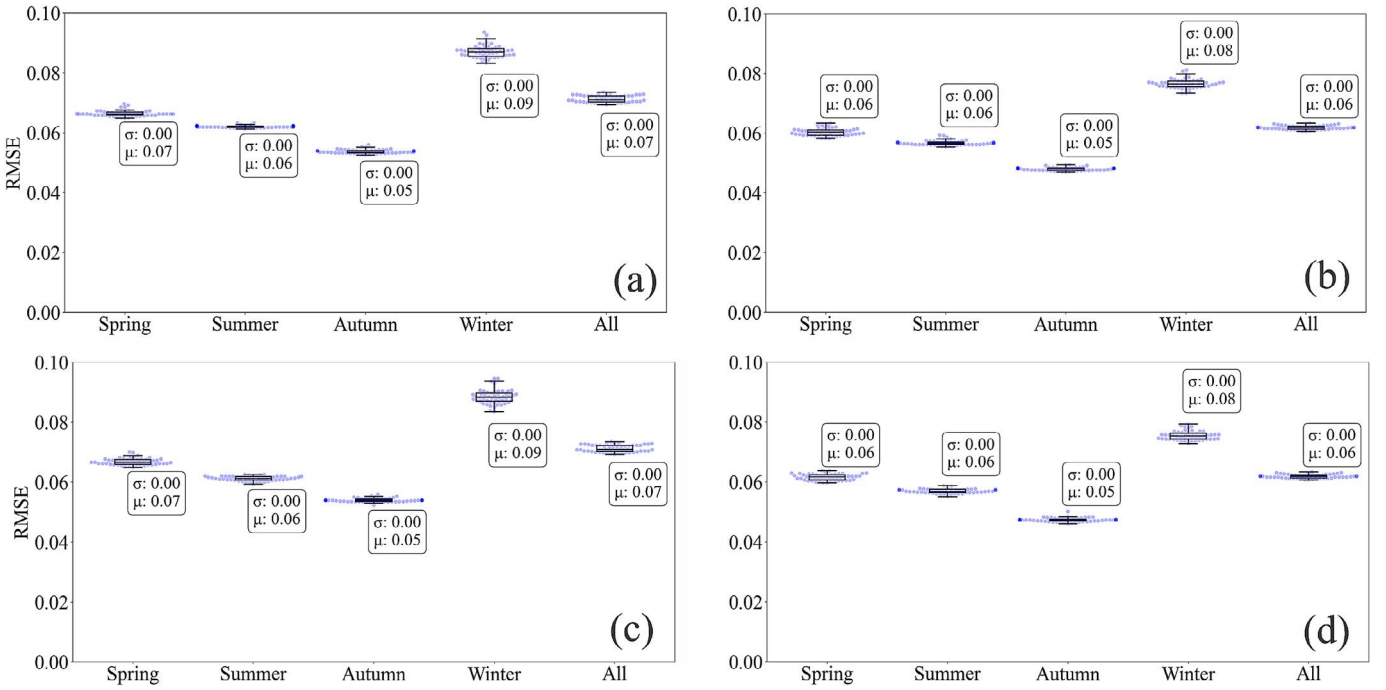

Fig. 5. Average RMSE between observed and estimated soil moisture for all the features suggested by the RF algorithm, using (a) SVR and (b) RF, and for the top five features suggested by the RF algorithm using (c) SVR and (d) RF. Note that μ is mean, while σ indicates the standard deviation. All refers to the entire period.

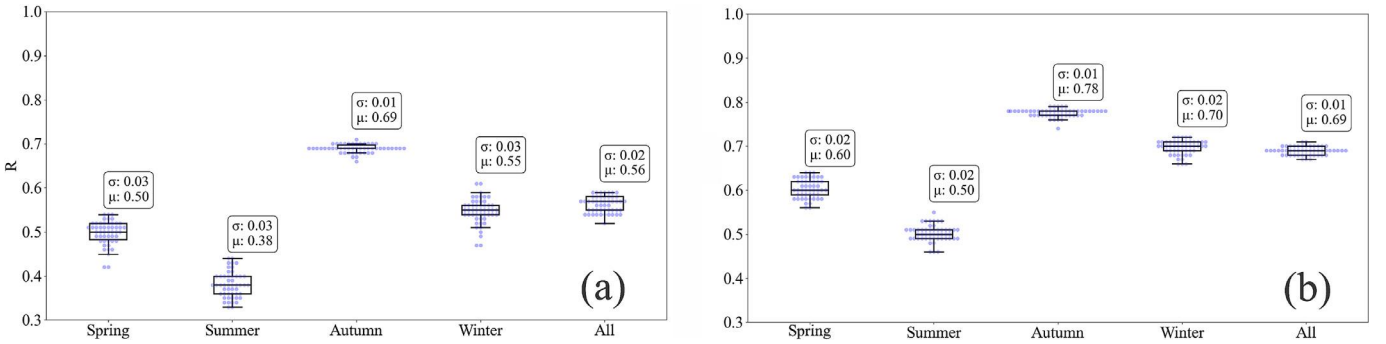

Fig. 6. Average correlation between observed and estimated soil moisture for the top five features suggested by the RF algorithm using (a) SVR and (b) RF. Note that μ is mean, while σ indicates the standard deviation. All refers to the entire period.

soil properties, and atmospheric conditions. These factors significantly influence soil moisture retrieval, as radar backscatter reflects a combination of soil and vegetation contributions. For instance, in densely vegetated areas, the backscattered signal is dominated by volumetric scattering, making soil moisture estimation more challenging, while sparser vegetation allows for more direct interaction with the soil surface, improving estimation accuracy. Additionally, the temporal misalignment between Sentinel-1 and Sentinel-2 acquisitions introduces additional uncertainty in soil moisture estimation due to changes in VWC. VWC is highly dynamic and can fluctuate on diurnal and sub-daily timescales, driven by environmental factors such as rainfall, irrigation, and temperature changes. Although the less-than-five-day time lag helps reduce mismatches, it does not eliminate the discrepancies between the observed radar and optical signals. These variations in VWC may particularly affect the interpretation of results during periods of rapid vegetation growth or stress, potentially influencing the accuracy of soil moisture estimation. Future studies should explore methods to minimize these uncertainties, such as using observations acquired closer in time or integrating auxiliary data that capture sub-daily VWC variations. It is also important to acknowledge a key limitation of machine learning models: their performance heavily depends on the quality and quantity of the training data. While high-quality, localized datasets improve accuracy and reliability, the effectiveness of these models can vary significantly based on the specific characteristics of the data they are trained on. Additionally, improper selection or adjustment of model parameters can hinder the ability to achieve optimal accuracy, making careful calibra-

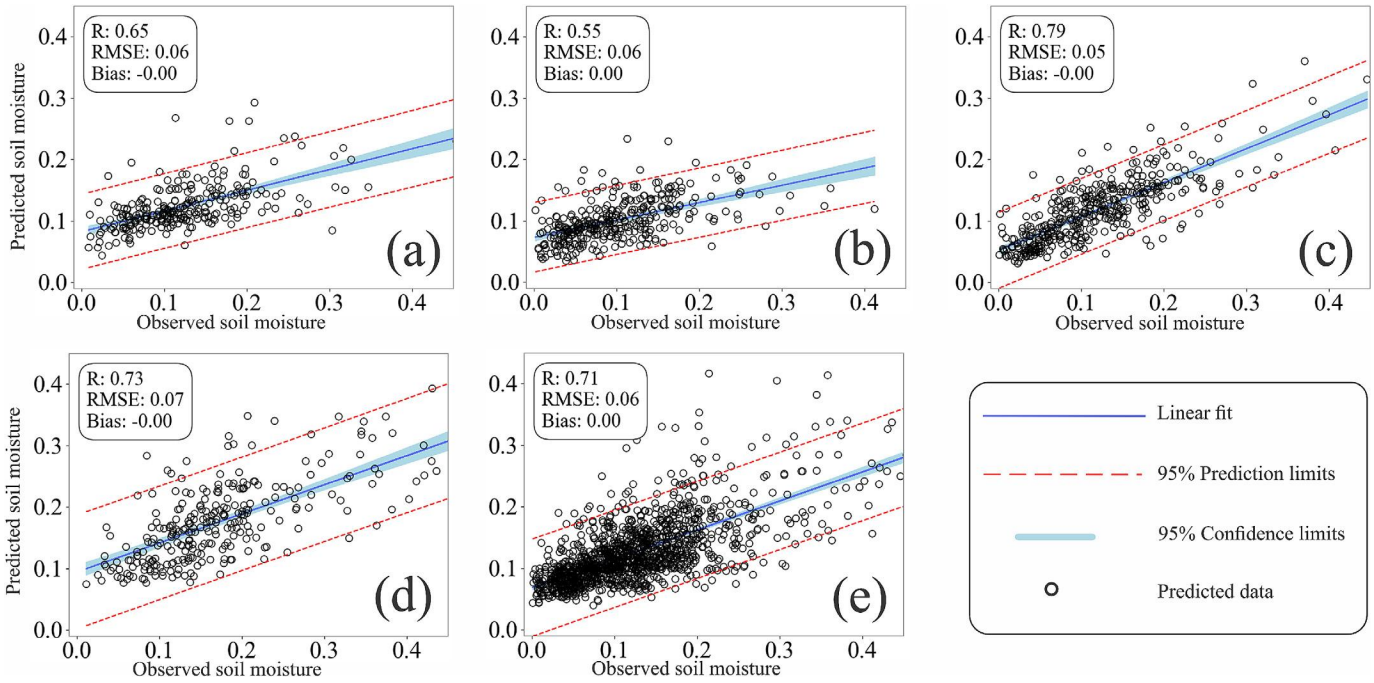

Fig. 7. Estimated seasonal soil moisture using the RF model for: (a) spring, (b) summer, (c) autumn, (d) winter, and (e) the entire period.

tion and fine-tuning essential. Although the employed methods demonstrated strong performance within the study area, its applicability to other regions with different vegetation types, climates, and soil properties requires further investigation. Future research should focus on testing these approaches in diverse regions, incorporating broader datasets, and developing transferable models. Such efforts would enhance the generalizability of soil moisture estimation methods across varying environmental conditions and mitigate the limitations associated with localized model training.

5. Conclusion

Two machine learning algorithms, Random Forest (RF) and Support Vector Regression (SVR), were employed to estimate near-surface soil moisture at 10 m resolution in the Yanco study area of Australia from combined Sentinel-1 SAR and Sentinel-2 optical imagery collected from March 2016 to August 2020. A total of 270 features were generated: 249 features from the 10 bands of Sentinel-2 optical images and 8 features from the two polarizations of Sentinel-1 SAR data. A seasonal feature selection approach was considered based on the correlation of the features with soil moisture, and then employing the SFS model with the RF algorithm (with default hyperparameters) to select the most important features for each season. The results showed that with only 5 features, the average lowest RMSE (0.06 m^3/m^3) and highest correlation (0.65) for soil moisture estimation could be achieved. However, the importance of seasonal feature selection was highlighted, showing that the optimal features may vary across different seasons, with the Local Incidence Angle (LIA) being the only feature consistently important across all seasons. Moreover, vegetation played a critical role in feature selection: in seasons with lower vegetation, such as autumn and summer, VV backscatter was more influential, while in denser vegetation seasons like winter and spring, VH backscatter was more prominent. Additionally, indices derived from Sentinel-1 images were particularly important in winter, likely due to the impact of plant water content on radar backscatter.

Results of soil moisture estimation using the RF and SVR algorithms with tuned hyperparameters and the optimal number of features showed that the RF model consistently outperformed the SVR model. This was particularly notable in autumn, where it achieved the highest accuracy with an RMSE of 0.05 m^3/m^3, while the lowest accuracy was observed in winter, with an RMSE of 0.07 m^3/m^3. Additionally, the findings underscored the challenge of accurately estimating soil moisture at higher moisture levels, where SAR backscatter has a reduced sensitivity, and further exacerbated by the scarcity of training data in these conditions. Despite these limitations, the study provided valuable insights into the potential of machine learning algorithms combined with multi-sensor data for improving soil moisture estimation, particularly when tailored to seasonal variations. It is important to note that these conclusions are based on data from the Yanco study area, and further validation in diverse environmental conditions is needed to confirm the robustness of these methods. Future research could focus on expanding the training dataset and exploring additional features or alternative machine learning approaches to further enhance the accuracy of soil moisture estimation across different climatic regimes. Furthermore, the temporal misalignment

between Sentinel-1 and Sentinel-2 acquisitions, due to the less-than-five-day time lag, poses a limitation, as dynamic changes in vegetation water content during this period may affect soil moisture estimation accuracy. Future work should aim to mitigate this uncertainty, potentially through the use of multi-sensor data with higher temporal resolution or the integration of models capable of accounting for diurnal VWC variations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their sincere gratitude to OzNet for providing the in-situ data and to the European Space Agency (ESA) for the freely available Sentinel-1 and Sentinel-2 data, which were crucial for this study. The authors also thank the Reviewers for their valuable comments, which substantially improved the quality of this paper.

References

Abdalla, A., Shami, S., Shahriari, M.A., Azar, M.K., 2024. Estimation of land displacement in East Baton Rouge Parish, Louisiana, using InSAR: Comparisons with GNSS and machine learning models. *Egypt. J. Remote Sens. Space Sci.* 27, 204–215.

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., Ghalhari, G.A.F., 2020. Machine learning to estimate surface soil moisture from remote sensing data. *Water* 12, 3223.

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. *Comput. Electron. Agric.* 44, 71–91.

Aghakouchak, A., Farahmand, A., Melton, F.S., Teixeira, J., Anderson, M.C., Wardlow, B.D., Hain, C.R., 2015. Remote sensing of drought: Progress, challenges and opportunities. *Rev. Geophys.* 53, 452–480.

Aha, D.W., Bankert, R.L., 1994. Feature selection for case-based classification of cloud types: An empirical comparison. In: Proceedings of the AAAI-94 Workshop on Case-Based Reasoning, p. 112.

Aha, D.W., Bankert, R.L., 1995. A comparative evaluation of sequential feature selection algorithms. In: Pre-proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics. PMLR, pp. 1–7.

Ahamed, T., Tian, L., Zhang, Y., Ting, K., 2011. A review of remote sensing methods for biomass feedstock production. *Biomass Bioenergy* 35, 2455–2469.

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* 33, 69–80.

Amani, M., Parsian, S., Mirmazloumi, S.M., Aieneh, O., 2016. Two new soil moisture indices based on the NIR-red triangle space of Landsat-8 data. *Int. J. Appl. Earth Obs. Geoinf.* 50, 176–186.

Asefa, T., Kemblowski, M., McKee, M., Khalil, A., 2006. Multi-time scale stream flow predictions: The support vector machines approach. *J. Hydrol.* 318, 7–16.

Attarzadeh, R., Amini, J., Notarnicola, C., Greifeneder, F., 2018. Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale. *Remote Sens. (Basel)* 10, 1285.

Ayari, E., Zribi, M., Lili-Chabaane, Z., Kassouk, Z., Jarlan, L., Rodriguez-Fernandez, N., Baghdadi, N., 2023. Sensitivity of surface soil moisture retrieval to satellite-derived vegetation descriptors over wheat fields in the Kairouan plain. *Eur. J. Remote Sens.* 56 2260555.

Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., 2022. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Sci. Data* 9, 782.

Babaeian, E., Sadeghi, M., Franz, T.E., Jones, S., Tuller, M., 2018. Mapping soil moisture with the OPtical TRApezoid Model (OPTRAM) based on long-term MODIS observations. *Remote Sens. Environ.* 211, 425–440.

Bao, Y., Lin, L., Wu, S., Deng, K.A.K., Petropoulos, G.P., 2018. Surface soil moisture retrievals over partially vegetated areas from the synergy of Sentinel-1 and Landsat 8 data using a modified water-cloud model. *Int. J. Appl. Earth Obs. Geoinf.* 72, 76–85.

Barrett, B.W., Dwyer, E., Whelan, P., 2009. Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sens. (Basel)* 1, 210–242.

Bhogapurapu, N., Dey, S., Homayouni, S., Bhattacharya, A., Rao, Y., 2022. Field-scale soil moisture estimation using sentinel-1 GRD SAR data. *Adv. Space Res.* 70, 3845–3858.

Bousbih, S., Zribi, M., el Hajj, M., Baghdadi, N., Lili-Chabaane, Z., Gao, Q., Fanise, P., 2018. Soil moisture and irrigation mapping in A semi-arid region, based on the synergetic use of Sentinel-1 and Sentinel-2 data. *Remote Sens. (Basel)* 10, 1953.

Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.

Carranza, C., Nolet, C., Pezij, M., van der Ploeg, M., 2021. Root zone soil moisture estimation with Random Forest. *J. Hydrol.* 593 125840.

Chaudhary, S.K., Srivastava, P.K., Gupta, D.K., Kumar, P., Prasad, R., Pandey, D.K., Das, A.K., Gupta, M., 2022. Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation. *Adv. Space Res.* 69, 1799–1812.

Chivkunova, O.B., Solovchenko, A.E., Sokolova, S., Merzlyak, M.N., Reshetnikova, I., Gitelson, A.A., 2001. Reflectance spectral features and detection of superficial scald–induced browning in storing apple fruit.

Cristianini, N., Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press.

Della Vecchia, A., Ferrazzoli, P., Guerriero, L., Ninivaggi, L., Strozzi, T., Wegmuller, U., 2008. Observing and modeling multifrequency scattering of maize during the whole growth cycle. *IEEE Trans. Geosci. Remote Sens.* 46, 3709–3718.

Devore, J.L., Berk, K.N., Carlton, M.A., 2021. Regression and correlation. Modern Mathematical Statistics with Applications. Springer.

Dobriyal, P., Qureshi, A., Badola, R., Hussain, S.A., 2012. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydrol.* 458, 110–117.

Dubois, P.C., Zyl, J.V., Engman, T., 1995. Measuring soil moisture with imaging radars. *IEEE Trans. Geosci. Remote Sens.* 33.

el Hajj, M., Baghdadi, N., Zribi, M., 2019. Comparative analysis of the accuracy of surface soil moisture estimation from the C-and L-bands. *Int. J. Appl. Earth Obs. Geoinf.* 82 101888.

Ezzahar, J., Ouaadi, N., Zribi, M., Elfarkh, J., Aouade, G., Khabba, S., Er-Raki, S., Chehbouni, A., Jarlan, L., 2019. Evaluation of backscattering models and support vector machine for the retrieval of bare soil moisture from Sentinel-1 data. *Remote Sens. (Basel)* 12, 72.

Ezzahar, J., Chehbouni, A., Ouaadi, N., Madiafi, M., Khabba, S., Er-Raki, S., Laamrani, A., Chakir, A., Chabaane, Z.L., Zribi, M., 2023. Sentinel-1 backscatter and interferometric coherence for soil moisture retrieval in winter wheat fields within a semi-arid south-Mediterranean climate: Machine learning versus semi-empirical models. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*.

Fensholt, R., Sandholt, I., 2003. Derivation of a shortwave infrared water stress index from MODIS near-and shortwave infrared data in a semiarid environment. *Remote Sens. Environ.* 87, 111–121.

Filgueiras, R., Mantovani, E.C., Dias, S.H.B., Fernandes Filho, E.I., da Cunha, F.F., Neale, C.M.U., 2019. New approach to determining the surface temperature without thermal band of satellites. *Eur. J. Agron.* 106, 12–22.

Forkuor, G., Zoungrana, J.-B.-B., Dimobe, K., Ouattara, B., Vadrevu, K. P., Tondoh, J.E., 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets-A case study. *Remote Sens. Environ.* 236 111496.

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. Soil moisture prediction using support vector machines 1. *JAWRA J. Am. Water Resour. Assoc.* 42, 1033–1046.

Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160, 271–282.

Gitelson, A.A., Merzlyak, M., Zur, Y., Stark, R., Gritz, U., 2001. Non-destructive and remote sensing techniques for estimation of vegetation status.

Glenn, E.P., Nagler, P.L., Huete, A.R., 2010. Vegetation index methods for estimating evapotranspiration by remote sensing. *Surv. Geophys.* 31, 531–555.

Gorrab, A., Ameline, M., Albergel, C., Baup, F., 2021. Use of sentinel-1 multi-configuration and multi-temporal series for monitoring parameters of winter wheat. *Remote Sens. (Basel)* 13, 553.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I. B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90, 337–352.

Hajdu, I., Yule, I., Dehghan-Shear, M.H., 2018. Modelling of near-surface soil moisture using machine learning and multi-temporal sentinel 1 images in New Zealand. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1422–1425.

Hoskera, A.K., Nico, G., Irshad Ahmed, M., Whitbread, A., 2020. Accuracies of soil moisture estimations using a semi-empirical model over bare soil agricultural croplands from sentinel-1 SAR data. *Remote Sens. (Basel)* 12, 1664.

Hou, C., Tan, M.L., Li, L., Zhang, F., 2024. Comparison of three machine learning algorithms for retrieving soil moisture information from Sentinel-1A SAR data in northwest Shandong plain, China. *Adv. Space Res.* 74, 75–88.

Hunt Jr, E.R., Daughtry, C., Eitel, J.U., Long, D.S., 2011. Remote sensing leaf chlorophyll content using a visible band index. *Agron. J.* 103, 1090–1099.

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* 30, 261–270.

Lee, J., 1983. Digital image smoothing and the sigma filter. Comput. Vision Graph. Image Process.

Li, Z., Yuan, Q., Su, X., 2024. High-spatial-resolution surface soil moisture retrieval using the Deep Forest model in the cloud environment over the Tibetan Plateau. *Geo-spatial Inf. Sci.*, 1–20.

Lin, Z., Wang, Q., Xu, Y., Luo, S., Zhou, C., Yu, Z., Xu, C.-Y., 2024. Soil moisture dynamics and associated rainfall-runoff processes under different land uses and land covers in a humid mountainous watershed. *J. Hydrol.* 636 131249.

Liu, Y., Qian, J., Yue, H., 2020. Combined Sentinel-1A with Sentinel-2A to estimate soil moisture in farmland. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 1292–1310.

Loew, A., Mauser, W., 2006. A semiempirical surface backscattering model for bare soil surfaces based on a generalized power law spectrum approach. *IEEE Trans. Geosci. Remote Sens.* 44, 1022–1035.

Lymburner, L., Beggs, P.J., Jacobson, C.R., 2000. Estimation of canopy-average surface-specific leaf area using Landsat TM data. *Photogramm. Eng. Remote Sens.* 66, 183–192.

Main, R., Cho, M.A., Mathieu, R., O'Kennedy, M.M., Ramoelo, A., Koch, S., 2011. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS J. Photogramm. Remote Sens.* 66, 751–761.

Malthus, T.J., Andrieu, B., Danson, F.M., Jaggard, K.W., Steven, M.D., 1993. Candidate high spectral resolution infrared indices for crop cover. *Remote Sens. Environ.* 46, 204–212.

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H., Rao, Y.S., 2020. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sens. Environ.* 247 111954.

May, R.J., Maier, H.R., Dandy, G.C., 2010. Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Netw.* 23, 283–294.

McNairn, H., Brisco, B., 2004. The application of C-band polarimetric SAR for agriculture: A review. *Can. J. Remote. Sens.* 30, 525–542.

Mishra, A., Vu, T., Veettil, A.V., Entekhabi, D., 2017. Drought monitoring with soil moisture active passive (SMAP) measurements. *J. Hydrol.* 552, 620–632.

Mondal, S., Mishra, A., 2024. Quantifying the precipitation, evapotranspiration, and soil moisture network's interaction over global land surface hydrological cycle. *Water Resour. Res.* 60 e2023WR034861.

Nasirzadehdizaji, R., Balik Sanli, F., Abdikan, S., Cakir, Z., Sekertekin, A., Ustuner, M., 2019. Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. *Appl. Sci.* 9, 655.

Ouaadi, N., Jarlan, L., Ezzahar, J., Zribi, M., Khabba, S., Bouras, E., Bousbih, S., Frison, P.-L., 2020. Monitoring of wheat crops using the backscattering coefficient and the interferometric coherence derived from Sentinel-1 in semi-arid areas. *Remote Sens. Environ.* 251 112050.

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., Reppucci, A., 2013. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.* 134, 234–248.

Pasolli, L., Notarnicola, C., Bruzzone, L., 2011a. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* 8, 1080–1084.

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Chiesa, S.D., Niedrist, G., Tappeiner, U., Zebisch, M., 2011b. Polarimetric RADARSAT-2 imagery for soil moisture retrieval in alpine areas. *Can. J. Remote. Sens.* 37, 535–547.

Pasolli, L., Notarnicola, C., Bertoldi, G., Bruzzone, L., Remelgado, R., Greifeneder, F., Niedrist, G., Della Chiesa, S., Tappeiner, U., Zebisch, M., 2015. Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8, 262–283.

Periasamy, S., 2018. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. *Remote Sens. Environ.* 217, 537–549.

Petropoulos, G.P., Ireland, G., Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Phys. Chem. Earth, Parts a/b/c* 83, 36–56.

Prakash, R., Singh, D., Pathak, N.P., 2011. A fusion approach to retrieve soil moisture with SAR and optical data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5, 196–206.

Price, J.C., 1980. The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation. *Water Resour. Res.* 16, 787–795.

Pu, R., Gong, P., Yu, Q., 2008. Comparative analysis of EO-1 ALI and Hyperion, and Landsat ETM+ data for mapping forest crown closure and leaf area index. *Sensors* 8, 3744–3766.

Pudil, P., Novovičová, J., Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recogn. Lett.* 15, 1119–1125.

Rouse Jr., J.W., Haas, R. H., Deering, D., Schell, J., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Sabaghy, S., Walker, J.P., Renzullo, L.J., Jackson, T.J., 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sens. Environ.* 209, 551–580.

Sadeghi, M., Jones, S.B., Philpot, W.D., 2015. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. *Remote Sens. Environ.* 164, 66–76.

Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B., 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sens. Environ.* 198, 52–68.

Scholkopf, B., Sung, K.-K., Burges, C.J., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V., 1997. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process.* 45, 2758–2765.

Scipal, K., Drusch, M., Wagner, W., 2008. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water Resour.* 31, 1101–1112.

Shami, S., Shahriari, M.A., Nilfouroushan, F., Forghani, N., Salimi, M., Reshadi, M.A.M., 2024. Surface displacement measurement and modeling of the Shah-Gheyb salt dome in southern Iran using InSAR and machine learning techniques. *Int. J. Appl. Earth Obs. Geoinf.* 132 104016.

Shi, J., Wang, J., Hsu, A.Y., O'Neill, P.E., Engman, E.T., 1997. Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. *IEEE Trans. Geosci. Remote Sens.* 35, 1254–1266.

Small, D., Schubert, A., 2019. Guide to sentinel-1 geocoding. Remote Sensing Lab. Univ. Zurich (RSL), Zürich, Switzerland, Tech. Rep. UZHS1-GC-AD.

Smith, A.B., Walker, J.P., Western, A.W., Young, R., Ellett, K., Pipunic, R., Grayson, R., Siriwardena, L., Chiew, F.H., Richter, H., 2012. The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.* 48.

Sun, Z., Xu, R., Du, W., Wang, L., Lu, D., 2019. High-resolution urban land mapping in China from sentinel 1A/2 imagery based on Google Earth Engine. *Remote Sens. (Basel)* 11, 752.

Thenkabail, P.S., Ward, A.D., Lyon, J., Merry, C.J., 1994. Thematic Mapper vegetation indices for determining soybean and corn growth parameters.

Ulaby, F.T., Batlivala, P.P., Dobson, M.C., 1978. Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-bare soil. *IEEE Trans. Geosci. Electron.* 16, 286–295.

Ulaby, F.T., Dubois, P.C., van Zyl, J., 1996. Radar mapping of surface soil moisture. *J. Hydrol.* 184, 57–84.

Vapnik, V.N., Vapnik, V., 1998. Statistical learning theory.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer.

Verstraeten, W.W., Veroustraete, F., Feyen, J., 2008. Assessment of evapotranspiration and soil moisture content across different scales of observation. *Sensors* 8, 70–117.

Volesky, J.C., Stern, R.J., Johnson, P.R., 2003. Geological control of massive sulfide mineralization in the Neoproterozoic Wadi Bidah shear zone, southwestern Saudi Arabia, inferences from orbital remote sensing and field studies. *Precambr. Res.* 123, 235–247.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., Strauss, P., 2018. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study. *Remote Sens. (Basel)* 10, 1396.

Wang, L., Gao, Y., 2023. Soil moisture retrieval from sentinel-1 and sentinel-2 data using ensemble learning over vegetated fields. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 16, 1802–1814.

Wang, F.-M., Huang, J.-F., Tang, Y.-L., Wang, X.-Z., 2007. New vegetation index and its application in estimating leaf area index of rice. *Rice Sci.* 14, 195–203.

Wang, H., Li, X., Long, H., Xu, X., Bao, Y., 2010. Monitoring the effects of land use and cover type changes on soil moisture using remote-sensing data: A case study in China's Yongding River basin. *Catena* 82, 135–145.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236 111402.

Whitney, A.W., 1971. A direct method of nonparametric measurement selection. *IEEE Trans. Comput.* 100, 1100–1103.

Xiao, Q., Chen, W., Sheng, Y., Li, J., 1994. A study on soil moisture monitoring using NOAA satellite. *Quart. J. Appl. Meteorol.* 5, 312–318.

Xiao, C., Li, P., Feng, Z., 2019. Monitoring annual dynamics of mature rubber plantations in Xishuangbanna during 1987-2018 using Landsat time series data: A multiple normalization approach. *Int. J. Appl. Earth Obs. Geoinf.* 77, 30–41.

Yee, M.S., Walker, J.P., Monerris, A., Rüdiger, C., Jackson, T.J., 2016. On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *J. Hydrol.* 537, 367–381.

Yin, F., Lewis, P.E., Gomez-Dans, J.L., Wu, Q., 2019. A sensor-invariant atmospheric correction method: Application to Sentinel-2/MSI and Landsat 8/OLI.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J., 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends* 1, 56–70.

Zhan, Z., Qin, Q., Ghulan, A., Wang, D., 2007. NIR-red spectral space based new method for soil moisture monitoring. *Sci. China Ser. D Earth Sci.* 50, 283–289.

Zhang, Y., Gong, J., Sun, K., Yin, J., Chen, X., 2018. Estimation of soil moisture index using multi-temporal Sentinel-1 images over Poyang lake ungauged zone. *Remote Sens. (Basel)* 10, 12.

Zribi, M., Gorrab, A., Baghdadi, N., Lili-Chabaane, Z., Mougenot, B., 2013. Influence of radar frequency on the relationship between bare surface soil moisture vertical profile and radar backscatter. *IEEE Geosci. Remote Sens. Lett.* 11, 848–852.