Soil Moisture Retrieval From Sentinel-1 and Sentinel-2 Data Using Ensemble Learning Over Vegetated Fields

Liguo Wang and Ya Gao

***Abstract*—Soil moisture (SM) is valuable basic data in climate, hydrological models, and agricultural applications. The rapid development of remote sensing technology can be used to monitor changes in SM at multiple spatial and temporal scales. In this article, we unfolded an SM retrieval method using ensemble learning combined with the Water Cloud Model (WCM) by Sentinel-1 and Sentinel-2 with multisource datasets. First, using the WCM, the influence of vegetation cover on the backscattering coefficient was removed, where we use three vegetation index (enhanced vegetation index (EVI), normalized difference vegetation index, and normalized difference water index) for analysis and comparison. Then, combined with other multisource datasets, an SM retrieval model was established based on the ensemble learning algorithm. Here, we choose two familiar ensemble learning algorithms for analysis and comparison, using Pearson correlation significance analysis, which are the random forest (RF) and the adaptive boosting (AdaBoost). The results revealed that the RF model performed is slightly superior to the AdaBoost model. The optimal performance mean absolute error, root-mean-square error (RMSE), and the unbiased RMSE of RF model are 2.289 vol%, 2.934 vol%, 2.934 vol%, respectively, which are slightly better than the AdaBoost model. EVI is suitable for WCM model to remove vegetation scattering effect. It shows that it is attainable to utilize the ensemble learning method to inversion of SM using radar data. The proposed framework maximizes the potential of WCM, RF model, and multisource datasets in deriving spatiotemporally continuous SM estimates, which should be valuable for SM inversion development.**

***Index Terms*—Adaptive boosting (AdaBoost), ensemble learning, random forest (RF), Sentinel-1/2, soil moisture (SM), Water Cloud Model (WCM).**

I. Introduction

SOIL moisture (SM) is an earthshaking basis for a variety of hydrological, biological, and biogeochemical processes, which supervises the energy, carbon, and water alternation between global terrestrial ecosystem and the atmosphere via transpiration [1], [2], [3]. SM information plays an important role in irrigation regulation, drought monitoring, yield prediction, promoting water-saving agriculture and ensuring food security [4], [5], [6], [7].

Recent years have witnessed a spurt of progress in remote sensing technology; new sensors have improved their performance in precision and spatial resolution. It provides richer, flexible alternative methods to obtain SM over a large area and worldwide, especially using optical/thermal infrared and microwave sensors [8], [9], [10], [11], [12]. Compared with optical remote sensing, microwave remote sensing is highly valued for its all-day, all-weather observation capability and high sensitivity to SM [13], [14], [15].

Up to now, multitudinous scholars have presupposed distinctive algorithms and methods for SM inversion on surface of exposed land, including physical analog (integral equation model [16], the advanced integral equation model (AIEM) [17]), empirical models (Oh Model [18], Dubois Model [19]), and semiempirical models (Shi Model [20]). If the algorithm of retrieving bare surface SM is applied to vegetation-covered area, it will be underestimated or overestimated. Therefore, many scholars have studied semiempirical models in vegetation-covered areas. Ordinary vegetation scattering models incorporate the Water Cloud Model (WCM [21]) and the Michigan Microwave Canopy Scattering Model [22].

However, retrieving soil parameters from radar signals is a matter of uncertainty, as multiple combinations of SM, soil roughness, and vegetation characteristics may result in the same electromagnetic response. In the last two decades, artificial neural networks (ANNs) have been extensively applied to retrieval of SM. They are able to retrieve tanglesome, dynamic, and nonlinearity patterns on the basis of the datasets [23], [24], [25]. Therefore, a variety of machine learning methods with empirical models or semiempirical models have diffuse application in radar SM inversion. Mohammad et al. used a neural network (NN) method for SM inversion with Sentinel-1 radar data. They tested disparate inversion SAR architectures.

1) The input of radar signal only in VV polarization.
2) Radar signal input only VH polarization mode.
3) Both VV and VH polarization methods are used as radar signal input.

The results show that the accuracy of soil water estimation is heightened by using prior information to analyze soil [15].

Manuscript received 2 October 2022; revised 22 January 2023; accepted 31 January 2023. Date of publication 6 February 2023; date of current version 15 February 2023. This work was supported by the National Natural Science Foundation of China under Grant 62071084 and Grant 62001434. *(Corresponding author: Ya Gao.)*

Liguo Wang is with the Harbin Engineering University, Harbin 150000, China, and also with the College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China (e-mail: wangliguo@hrbeu.edu.cn).

Ya Gao is with the Harbin Engineering University, Harbin 150001, China (e-mail: gaoya0001@hrbeu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2023.3242264, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3242264

Mirsoleimani et al. used NNs under the calibrated integral equation model and the modified Dubois model to estimate SM. The outcomes indicate the productivity of the VV polarization data for retrieving soil surface moisture [26]. An inversion approach was developed to invert the SAR data and estimate the SM using NNs of the *C*-bands (Sentinel-1) and L-bands (PALSAR). The experiment result expressed that the L-band provided marginally more minor exact SM estimates than the *C*-band [27]. Gao et al. performed SM retrieval using ALOS-2 data based on an optimized back propagation (GA-BP) NN technique with WCM. This way displayed higher sensitivity in the L-band to SM even under vegetation-covered area [28]. Liu et al. based on four algorithms, generalized regression neural network (GRNN), support vector regression (SVR), random forest regression (RFR), and deep neural network (DNN) algorithms to retrieve SM, merging with Sentinel-1A with Sentinel-2A images. It can be seen that the regression algorithm has higher accuracy in estimating SM, and the accuracy of DNN in estimating SM exceeded in GRNN and RFR, and is superior to SVR [29]. ALOS-2 and Sentinel-1 data were used. An ANN SMC inverse algorithm combining a WCM, an AIEM, and an Oh model database was also used. That confirms that Sentinel-1 and ALOS-2 SM inversions have higher accuracy and correspond to lower vegetation areas (crops, grasses, and shrubs) [30].

This article demonstrates the ability of a retrieval approach for SM to perform robust and accurate. The suggested method is in the light of the retrieval of the WCM using the ensemble learning over vegetated fields. First, we established three vegetation indexes using Sentinel-2 as a vegetation parameter in WCM. Second, through the first step, we can get different combinations of soil backscattering. We combine DEM and measured data to establish different datasets. Third, in accordance with two ensemble learning algorithms [random forest (RF) and adaptive boosting (AdaBoost)], the training data are used to train the algorithm, respectively. Finally, to evaluate the applicability of RF and AdaBoost, the predicted and measured SM values were used to identify the trained RF and AdaBoost. In this dissertation, the remaining parts show the materials and methods. Section III presents the results and Section IV describes the discussion. Finally, the main outcomes are shown in Section V.

II. Materials and Methods

A. Study Area

The research area represents the vegetated fields of the Luan River in northern Hebei, which is the second biggest river in the North China. It is the only river that flows into the sea from the Bohai Sea. The Luan River Basin is a vital water source in the Beijing–Tianjin–Hebei region [31].

The study area of this article selects the advanced position of the Luan River watershed (Shandian River and Xiaoluan River watershed), which covers an area of 41–43 °N and 115.5–117.5 °E. Research area land cover is dominated by grassland, farmland, forest, a few shrublands, and exposed soil ground. Fig. 1 shows the locality geographical map of the research area.

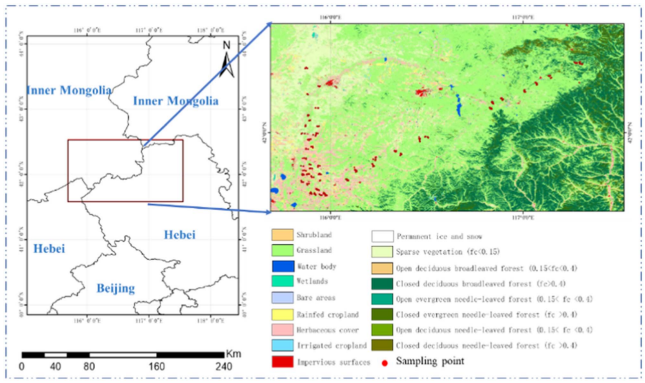


Fig. 1. Location of the Luan River Basin study site, geographical location of the study area, vegetation coverage and sampling point distribution (from GLC_FCS30).

B. Datasets

1) Sentinel-1: Sentinel-1 involves a couple of polar-orbiting satellites, Sentinel-1A (S1A) and Sentinel-1B (S1B) at *C*-band. The sensors carried by the two satellites are SAR, which are active microwave imaging radar satellites. The sensor is equipped with *C*-band (https://search.asf.alaska.edu/) [32], [33], [34]. The Sentinel-1 IW Ground Range Detected product level-1 data under space resolution of 10 m $*$ 10 m was selected for this study. Three SAR images acquired by Sentinel-1 were calculated hereinafter (20180912, 20180916, 20180919). The data are processed by SNAP software, followed by thermal noise removal, radiometric calibration, multilook filtering, coherent speckle filtering, terrain correction, and finally converted to backscatter coefficient in dB units. The formula for decibelization is as follows:

$$dB = 10 * \log_{10}(P/P_0). \tag{1}$$

Among them, P and P_0 represent the target amount and the reference amount, respectively. For the backscattering coefficient σ_0, decibelization actually performs the following logarithmic transformation:

$$\sigma_0(dB) = 10 * \log_{10}\sigma_0. \tag{2}$$

2) Sentinel-2: Sentinel-2 has a couple of satellites (Sentinel-2A, Sentinel-2B) whose assignment is to support vegetation, land cover, and environmental monitoring. They bestow all of the surface of earth land, large islands, inland and coastal waters on the earth's surface every five days (https://scihub.copernicus.eu/) [35], [36], [37]. Sentinel-2 data with the same or similar time as the radar data were selected for stitching and cropping processing. In this study, six optical images obtained by Sentinel-2A were used to approximate the dates of the S1 images. The preprocessing process includes atmospheric correction, radiometric calibration, resampling with the same resolution as Sentinel-1 to 10 m, and then calculating normalized difference water index (NDWI), normalized difference vegetation index (NDVI), and enhanced vegetation index (EVI) [38], [39], [40], [41].

3) Synchronous Observation Datasets: This dataset is developed by the National Tibetan Plateau Data Center (TPDC) (http://data.tpdc.ac.cn). The datasets are divided into two parts, one is synchronous observation dataset of soil temperature and SM and the other is synchronous observation dataset of surface roughness of Luan River in 2018 [42], [43]. In this study, 0–5 cm SM (volume water content, %) and 0–5 cm soil temperature (°C) were selected. This dataset is a field sampled dataset by the team of Zhao et al. The dataset contains surface and soil temperature and SM data measured simultaneously on the ground during the 2018 Soil Moisture Remote Sensing Experiment aerial flight test in the Luan River Basin. It is used to verify the "true value" of the remote sensing inversion. The ground-synchronous samples were collected in the upper reaches of the Luan River (Shandian River upper course and Xiaoluan River Basin) in September 2018, using a portable SM meter, an external probe type temperature logger, and the ring knife method to obtain the data [31]. The study area is divided into two directions. The topography in the north-south direction is complex, and there are many types of typical features, such as grassland, agricultural land, wasteland, bare land, and woodland. The surface undulations of different land types in the same large sample are more varied; in the northeast-southwest direction, the land types are simple and mostly grassland. They were measured by laboratory calculations. The measured SM was obtained in the range of 1.9 vol.%–71.17 vol.%. The range of root mean square height was 0.46–4.44. The range of correlation length was 6.305–31.965 cm.

The interaction between microwave and surface is not only related to the characteristics of microwave and soil dielectric properties but also closely related to the microscopic and geometric characteristics of surface. In microwave remote sensing research, the surface roughness used to describe the geometric characteristics of the surface is generally represented by root mean square height and correlation length. There is an important relationship between soil structure and microwave backscattering. Zribi et al. [64] analyzed using a fractional Brownian model. The surface roughness dataset is derived from the ground synchronous observation in the SM remote sensing experiment in the Luan River Basin. Surface roughness is expressed as RMS height and correlation length, where RMS height is a metric of roughness in the vertical direction, and autocorrelation length is a metrical of roughness in the horizontal direction. This dataset is obtained through the steps of soil surface height digitization, slope correction, period correction, and roughness calculation [31].

4) Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER-GDEM): ASTER-GDEM is an easy-to-use, high-precision DEM at 30 m spatial resolution that covers almost all places on earth [44], [45], [46]. By processing, we fetched the aspect, elevation, and slope of SM samples therefrom the ASTER-GDEM as the input datasets (http://reverb.echo.nasa.gov/reverb/).

5) GLC_FCS30: The land cover type data are selected from global 30-m land-cover dynamic monitoring products with fine classification system from 1985 to 2020 (GLC_FCS30-1985_2020) [47], [48], [49]. These data are according to all Landsat satellite data (Landsat TM, ETM+, and OLI) since 1984 to 2020, and united to the change detection results to realize the dynamic update of land cover by region and period. As a result, the GLC_FCS30 products from 1985 until 2020 were produced. The data contains a total of 29 land cover types (https://data.casearth.cn/).

TABLE I
MULTISOURCE DATASETS USED IN THIS STUDY

Dataset	Details	Spatial resolution	Temporal resolution
Sentinel-1	Soil backscattering (VV, VH)	10m	12d
Sentinel-2	EVI	10/20m	5d
	NDVI		
	NDWI		
Topography	Elevation	30m	-
	Slope		
	Aspect		
In-situ soil moisture	0-5cm SM	Point scale	-
In-situ surface roughness	RMS height	Point scale	-
	Correlation length		

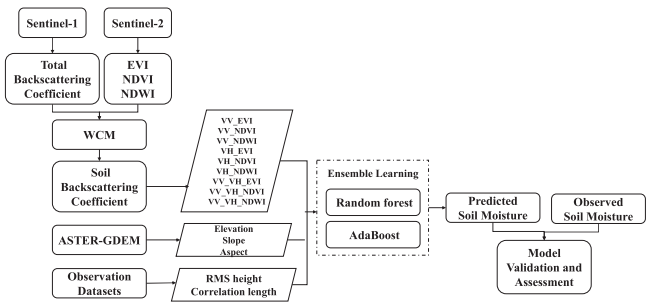


Fig. 2. Flowchart of the retrieval framework.

C. Methods

1) Overall Retrieval Framework: SAR is the main means of monitoring SM in active microwave remote sensing. The radar backscattering coefficient is affected by SM, surface roughness parameters, soil texture parameters, and the configuration parameters of the satellite sensor system. We collected Sentinel-1, Sentinel-2, DEM, soil roughness data, and SM data, which provide a wealth of soil information. Combining all these datasets, we utilize two integrated learning models to retrieve SM. Table I summarizes the specific dataset information, and Fig. 2 gives the flowchart of the proposed retrieval framework. First, we chose to study the area covered by vegetation, because the presence of vegetation not only produces direct backscattering but also attenuates backscattering from the surface, resulting in the observed backscattering that includes both vegetation, the surface,

and the interplay among vegetation with the surface amount of. So we have to remove the effect of vegetation scattering first. The WCM model considers both the scattering contributions of vegetation and soil, and has been extensively used in SM inversion under different vegetation cover. With the WCM, we get the backscatter of bare soil (VV, VH).

In this study, we choose NDWI, NDVI, and EVI as the description of vegetation based on formula (4). There are nine SAR compound modes for estimating the SM (VV_NDWI, VV_NDVI, VV_EVI, VH_NDWI, VH_NDVI, VH_EVI, VV_VH_NDWI, VV_VH_NDVI, VV_VH_EVI).

Then, we will extract the required variable information from the multisource dataset, which consisted of soil backscatter from Sentinel-1/2, measured SM and soil roughness from TPDC and elevation, slope, and aspect from ASTER GDEM. Nonlinear interrelation among multiple input variate with target variate (SM) was approximated using two integrated learning models, RF and AdaBoost. Insomuch the partition of the training and test part may bias the properties of the integrated learning, we used tenfold cross-validation means to value the arithmetic and optimize the hyperparameters of various methods. Eventually, different evaluation indicators are selected to the accuracy of assessment and applicability of the SM inversion model.

2) Water Cloud Model: In 1978, Attema and Ulaby took crops as the study goal and proposed a semiempirical vegetation backscattering method in accordance with the first-order solution of the radiative transfer equation, namely WCM [21]. In the WCM, the vegetation canopy is assumed to be an isotropic scatterer, so the total backscattering of the surface covered by vegetation is just represented as two parts, one is the volume scattering term directly reflected by the vegetation canopy and the other is backscattering term that reaches the ground after being attenuated twice by the vegetation layer. The model is more simple and practical in describing the radar scattering mechanism of low vegetation cover ground surface, so it is widely used as a tool for inversion SM in crop coverage areas. The common form of the WCM model is denoted as

$$\sigma^0_{\text{can}} = \sigma^0_{\text{veg}} + \tau^2 \sigma^0_{\text{soil}} \tag{3}$$

$$\sigma^0_{\text{veg}} = AV \cos\theta \left(1 - \tau^2\right) \tag{4}$$

$$\tau^2 = \exp\left(-2BV_2/\cos\theta\right) \tag{5}$$

where σ^0_{can} is the radar total backscattering coefficient received, σ^0_{veg} is the signal directly reflected by the vegetation, σ^0_{soil} is the scattered signal of the soil, τ^2 is the attenuation coefficient of the signal attenuated twice by the vegetation, θ is the signal incident angle, V is vegetation related parameters, and NDWI, NDVI, and EVI are described as vegetation in this study. A and B are empirical factors for this model, related to the type of vegetation and radar parameters. In this study, the values of A and B refer to the results of Bindlish and Barros [63], that is, $A = 0.0012$ and $B = 0.091$.

3) Random Forest: RF is earliest discovered by Breiman, which represents a very representative bagging (bootstrap aggregating) ensemble algorithm [50]. RF is a randomly constructed forest that incorporates quite a little unrelated decision trees. Integrated learning works out individual prediction problems by building combinations of multiple models [51], [52]. RFs have the advantage of aggregating different outputs of a single decision tree, reducing the variance that can lead to decision tree errors. According to the majority voting algorithm, we can find the average output given from many individual trees, smoothing the variance. This way the model may produce results that are close to the true values [53]. Scikit-learn package is a powerful freeware machine learning library for the Python programming language [54]. We use sklearn package to develop the RF model. N_estimators is an important parameter of RF. The effect of this parameter on the precisely of the RF model is monotonic. The bigger the n_estimators, the better the model effect tends to be. The quantity of trees and the number of features are the two main parameters in the RF, and other parameters use the default values. We set the number of trees to 10, 20, 50, 100, 200, 500, and all the features are generally used in this study.

4) AdaBoost: Boosting is a method in ensemble learning. It is a gradual process of optimizing the overall learner, and each individual learner is making up for the deficiencies of the overall learner, thus achieving overall optimization. AdaBoost is a classic algorithm of Boosting [55], [56], [57]. AdaBoost package was accustomed to train the AdaBoost pattern. Adaboost framework parameters included base_estimator, n_estimators, learnig_rate, loss and base learner parameters. The method of selecting the optimal parameters we have chosen for cross-validation is in this article. Adaboost has excellent resistance to overfitting, although increasing the number of training epochs does not increase the generalization error.

Base_estimator specifies the learning content. Generally, this parameter is the default and we do not need to change it. N_estimators is the maximum number of iterations of the weak learner, where we set the number of layers of the tree in the weak model to 20, 30, and 50, and the number of weak models to 100, 300, and 500. The optimal value is obtained by learning. Learning_rate is called a regularization parameter, and adjusting this parameter reasonably can alleviate the overfitting problem. We set this parameter to 0.8. The meaning of the parameter loss is to define the loss function. The optional parameters are linear, square, and exponential, which correspond to the linear loss function, the squared loss function, and the exponential loss function. Here we choose linear. The base learner parameters also refer to the decision tree parameters, where we set the number of layers of the tree in the weak model to 20, 30, and 50, and the number of weak models to 100, 300, and 500. The Adaboost interface in sklearn is in sklearn.ensemble. For details, please read the official documentation of sklearn.

5) Model Validation and Assessment: Model validation has been generally applied in SM inversion evaluation: R, mean absolute error (MAE), Bias, root-mean-square error (RMSE), and unbiased RMSE (ubRMSE) in this study. The model validation is computed as follows:

$$R = \frac{E\left[\left(sm_{est} - E\left[sm_{est}\right]\right)\left(sm_{true} - E\left[sm_{true}\right]\right)\right]}{\sigma_{est}\sigma_{true}} \tag{6}$$

$$\text{Bias} = E\left[sm_{est}\right] - E\left[sm_{true}\right] \tag{7}$$

$$\text{MAE} = E\left[\left(sm_{est} - sm_{true)}\right)\right] \tag{8}$$

$$\text{RMSE} = \sqrt{E\left[\left(sm_{est} - sm_{true)}\right)^2\right]} \tag{9}$$

$$\text{ubRMSE} = \sqrt{E\left\{\left[\left(sm_{est} - E\left[sm_{est}\right]\right) - \left(sm_{true} - E\left[sm_{true}\right]\right)\right]^2\right\}} \tag{10}$$

where $E[.]$ is the mean value, sm_{true} displays the measured SM, and sm_{est} indicates the united SM inversion of algorithms. σ_{ture} and σ_{est} are the standard deviation of the measured SM and inversion SM, severally.

There are three conditions for the Pearson correlation significance test.

1) The predicted SM and measured SM data come in pairs from a normally distributed population.
2) The gap between the predicted SM and measured SM data should not be too wide.
3) Each group of samples was sampled independently.

When the above conditions are met, the Pearson correlation coefficient significance analysis can be carried out. To test whether the data conform to a normal distribution, we chose a P-P plot. P-P diagram is a scatterplot painted from the cumulative probability of a variable corresponding to the cumulative probability of a specified theoretical distribution. It is used to visually detect whether the sample data conforms to a certain probability distribution. If the data being tested conforms to the specified distribution, the points representing the sample data should lie substantially on the diagonal representing the theoretical distribution. For Pearson correlation significance analysis, we chose *T*-test.

When we do the Pearson correlation analysis, we also do the Pearson correlation significance test (*t*-test). When the tested samples basically meet the conditions of the Pearson correlation test, we can perform the significance test with the Pearson correlation coefficient. There are many mathematicians here who have proved that the Pearson correlation coefficient can construct a statistic *t*, the structure of which is given as follows:

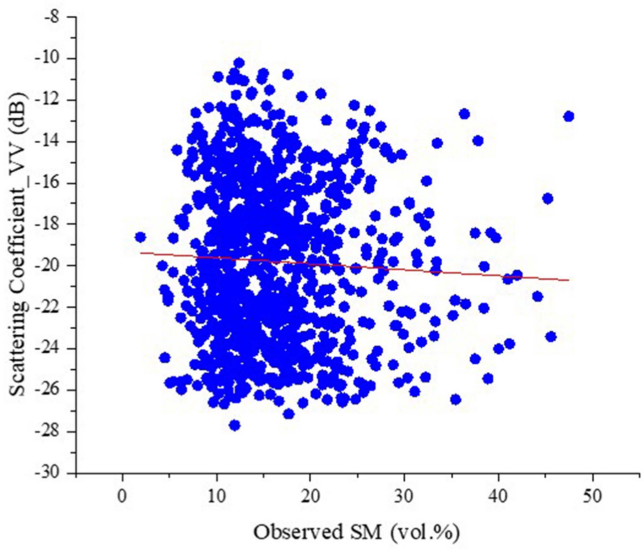
$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} \tag{11}$$

where *n* is the quantity of samples and *r* is the Pearson correlation coefficient. This statistic is attested to be in line with the *t* distribution with *t*–2 degrees of freedom.

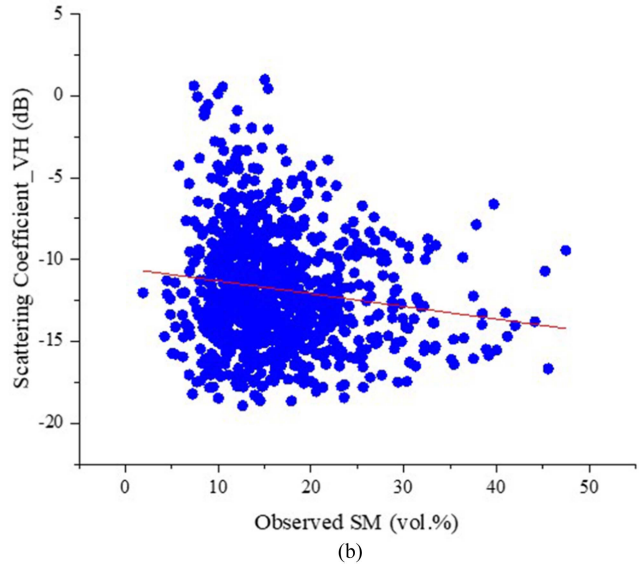
III. Results

A. Radar Data Analysis

Analyzing the sensitivity of the backscattering coefficient with respect to the SM helps us to choose the most sensitive method for the SM. This part of the work will analyze the role of SAR information to observe SM of Sentinel-1 (VV, VH). The results were shown in Fig. 3 and Table II. In the study area, grassland, farmland, forest, a few shrublands, and bare ground are mainly land cover. We obtain the total backscattering coefficient from Sentinel-1. Fig. 2 shows the connection of the SAR backscattering coefficient and Observed SM. The SAR signal of VV polarization [Fig. 3(a)] and VH polarization [Fig. 3(b)] both have low sensitivity to Observed SM. The consequence of the backscatter coefficient and SM fitting is represented in



(a)



(b)

Fig. 3. SAR backscattering coefficient versus Observed SM. (a) VV. (b) VH.

TABLE II
Result of Backscatter Coefficient and SM Fitting

Relationship	Linear fit	Logarithm fit
VV-SM	y = -0.0195x - 19.481	y = -0.234ln(x) - 19.162
	R=0.024	R=0.034
VH-SM	y = -0.0777x - 10.537	y = -1.214ln(x) - 8.5099
	R=0.136	R=0.149

Table II. We both calculate two curve fittings, one is a linear fit and the other is a logarithm fit. We can find that the *R* values are 0.024 and 0.136 of VV and VH polarization, respectively (see Table II), and the *R* are 0.034 and 0.149. It means that σ^0_{can} both of VV and VH are low sensitivity with SM.

Compared with the results of the backscatter coefficient and SM fitting, we can find the logarithm fit shows higher coefficients than the linear fit. This finding may be due to the effect of surface roughness or vegetation, or it may be due to the effect of SAR sensors. A very complex relationship between them. The machine learning method can effectively simulate any nonlinear function, so the integrated learning method is introduced in this study to try to conduct SM inversion. In addition, we also found that the study area is covered by vegetation, and the total backscattering coefficients obtained from Sentinel-1 include vegetation scattering and soil scattering. Due to the effect of vegetation scattering, the relationship between radar signal and SM is less sensitive, resulting in the underestimation or overestimation of SM and surface roughness. It is challenging to describe the effect of plant, surface roughness, and intricacy system between adjust SAR information and SM. The complex relationship among the input with output data can be described by an ensemble learning approach, and it can replace traditional numerical modeling techniques [58], [59]. Therefore, before inversion of surface SM, vegetation scattering should be removed first to improve the sensitivity of soil scattering and SM. Therefore, the WCM was used in this article to remove the effect of plant and obtain the backscattering coefficient of bare soil. Then, surface soil water retrieval was carried out by the ensemble learning method.

B. Results of RF

Given that the best results were obtained using both the VV and VH together, the specific details are described in Section III-A1. At the same time, EVI was applied in WCM as a vegetation description, and the result was due to NDVI and NDWI. The results are discussed in Section III-A2.

1) Estimating SM by RF: Fig. 4 shows the performance of the RF that takes σ^0_{soil}, DEM, and soil roughness datasets as input parameters. Among them, by using WCM model, NDWI, NDVI, and EVI as vegetation description, nine different σ^0_{soil} are obtained, which are VV_NDWI, VV_NDVI, VV_EVI, VH_NDWI, VH_NDVI, VH_EVI, VV_VH_NDWI, VV_VH_NDVI, VV_VH_EVI, respectively. For the purpose of evaluating the relationship between the observed SM and the predicted SM, we selected four evaluation indicators, namely MAE, Bias, RMSE, and ubRMSE.

In all cases shown in Fig. 4, there is a nonbiased estimation of SM. When the σ^0_{soil} is described as VV, the RF built with VV_NDWI shows the MAE of 2.43 vol.%, the RMSE of 3.32 vol.%, and ubRMSE of 3.32 vol.% [see Fig. 4(a1)]. For a reference VV_NDVI VV_EVI, the accuracy of sm estimates (MAE, RMSE, and ubRMSE) is improved [see Fig. 4(a2) and (a3)]. The use of VV_NDVI yields a smaller MAE "estimated SM—reference SM" (2.19 vol.%), an RMSE (3.02

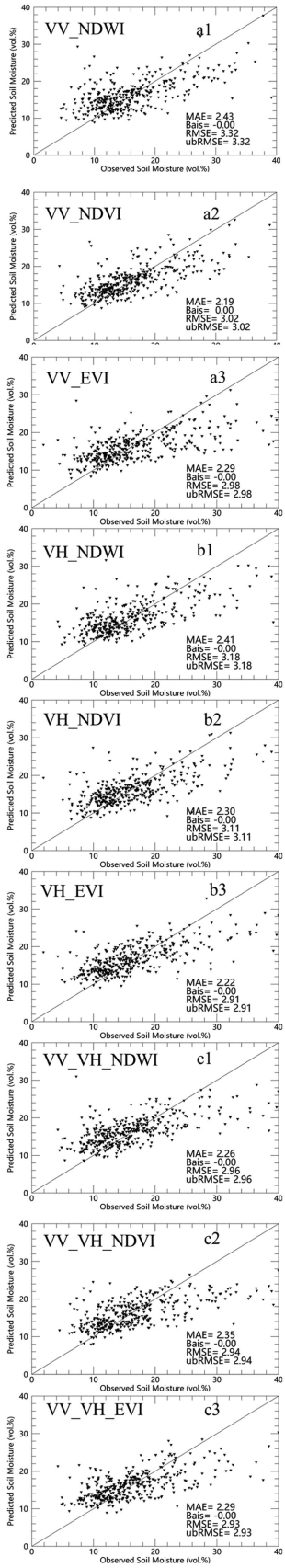


Fig. 4. Predicted SM and observed SM dispersion point diagram of RF: VV(a), VH(b), VV_VH(c); 1(NDWI), 2(NDVI), 3(EVI).

vol.%), and ubRMSE (3.02 vol.%) in comparison to VV_NDWI [see Fig. 4(a2)]. Similarly, the use of VV_EVI improves the MAE by 0.14 vol.% (MAE = 2.29 vol.%), the RMSE by 0.34 vol.% (RMSE = 2.98 vol.%), and the ubRMSE by 0.34 vol.% (ubRMSE = 2.98 vol.%) [see Fig. 4(a3)].

When VH is used as the input parameter, the results obtained are shown in Fig. 4(b). The results of VH_NDWI represent that MAE is 2.41 vol.%, RMSE is 3.18 vol.%, and ubRMSE is 3.18 vol.% [see Fig. 4(b1)]. Likewise, both VH_NDVI and VH_EVI have increased results [see Fig. 4(b2) and (b3)]. MAE reduced 0.11 vol.% (MAE = 2.30 vol.%), RMSE descended 0.77 vol.% (RMSE = 3.11 vol.%), and ubRMSE shortened 0.77 vol.% (ubRMSE = 2.98 vol.%) with the input of VH_NDVI [see Fig. 4(b2)]. In the case of VH_EVI [see Fig. 4(b3)], the MAE varies from 2.41 vol.% to 2.22 vol.%, the RMSE changes 3.18 vol.% into 2.91 vol.%, and the ubRMSE fluctuates from 3.18 vol.% to 2.91 vol.%.

The first two use a single polarization as the input, and Fig. 4(c) shows that the two polarizations of VV and VH are jointly used as the input. The result of VV_VH_NDWI is that MAE is 2.26 vol.%, RMSE is 2.96 vol.%, and ubRMSE is 2.96 vol.% [see Fig. 4(c1)]. VV_VH_NDVI displays the 2.35 vol.% of MAE, 2.94 vol.% of RMSE, and 2.94 vol.% of ubRMSE [Fig. 4(c2)]. VV_VH_EVI depictes the best results (MAE = 2.29 vol.%, RMSE = 2.93 vol.%, and ubRMSE = 2.93 vol.%) [Fig. 4(c3)].

From what has been discussed above, we can get the σ^0_{soil} obtained by using EVI as the vegetation parameter, and then as the RF input parameter, the inversion result is better than NDVI and NDWI. At the same time, we found that the results obtained by using both VV and VH polarizations as the input parameters of the RF are superior than those obtained by using only VV or VH.

2) Relative Error Between Predicted and Measured Value of RF: Fig. 5 shows an error plot of predicted and observed values from soil inversion using RF. In RF training, we divided the dataset into a training set and a test set with a ratio of 0.6 to 0.4. Due to the large amount of data, we randomly selected 100 points for plotting analysis when making the error analysis chart. Fig. 5(a)–(c) shows the common situation of input polarization mode VV, VH, and both VV and VH, respectively. The error range is mostly between −5 and 5, and very few have lower or higher error values. When the input is VV polarization, the ratio of the error rates among the inversion results with the observed values of σ^0_{soil} computed by different vegetation parameters are 78% [VV_NDWI, Fig. 5(a1)], 78% [VV_NDVI, Fig. 5(a2)], and 80% [VV_EVI, Fig. 5(a3)] between –5 and 5, respectively. As the VH polarization of the RF, we can gain that the probability between –5 and 5 is 74% [VH_NDWI, Fig. 5(b1)], 74% [VH_NDVI, Fig. 5(b2)], and 84% [VH_EVI, Fig. 5(b3)]. When both VV and VH are input, the probability of VV_VH_NDWI is 74% [Fig. 5(c1)], VV_VH_NDVI is 83% [Fig. 5(c2)], and VV_VH_EVI is 71% [Fig. 5(c3)] from –5 to 5. Through Section III-A1, we find that the final inversion results are best when vegetation is described as EVI, and both VV, VH as the input of the RF.

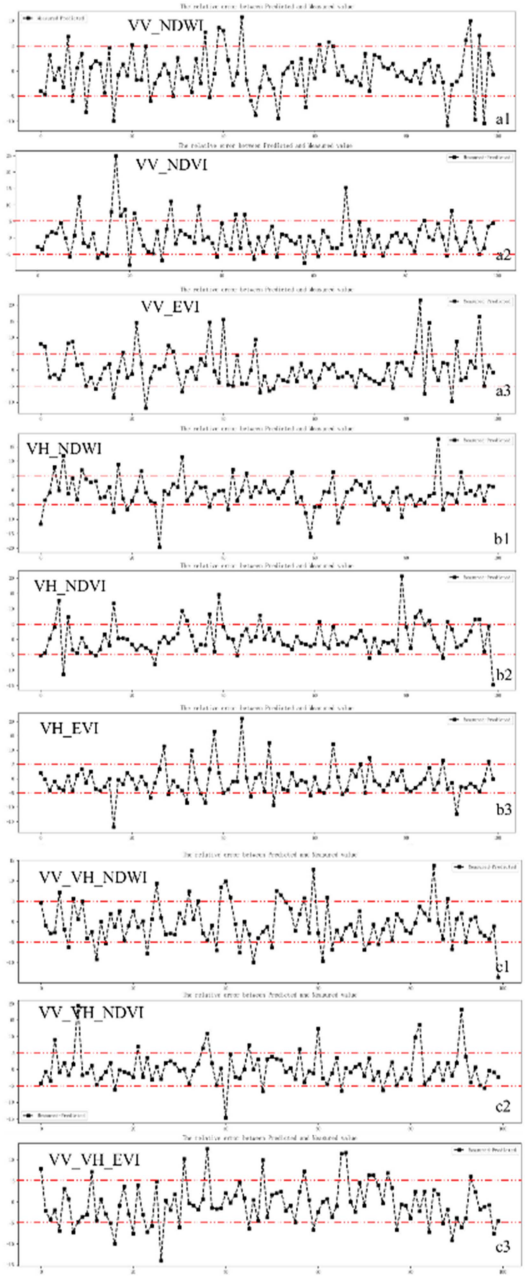


Fig. 5. Relative error between predicted and measured value of RF: VV(a), VH(b), VV_VH(c); 1(NDWI), 2(NDVI), 3(EVI).

C. Results of AdaBoost

1) Estimating SM by AdaBoost: Fig. 6 shows the performance of the AdaBoost for estimating SM. For comparative analysis with the RF algorithm, we set the same input parameters. For comparative analysis with RF algorithm, we set the same input parameters. Also, each set of parameters is studied the same number of times. All the results represent nonbiased estimation of SM in Fig. 6. When the polarization mode is VV, the AdaBoost built with VV_NDWI shows MAE of 2.78 vol.%, RMSE of 3.74 vol.%, and ubRMSE of 3.74 vol.% [Fig. 6(a1)]. The results of VV_NDVI are MAE of 2.69 vol.%,

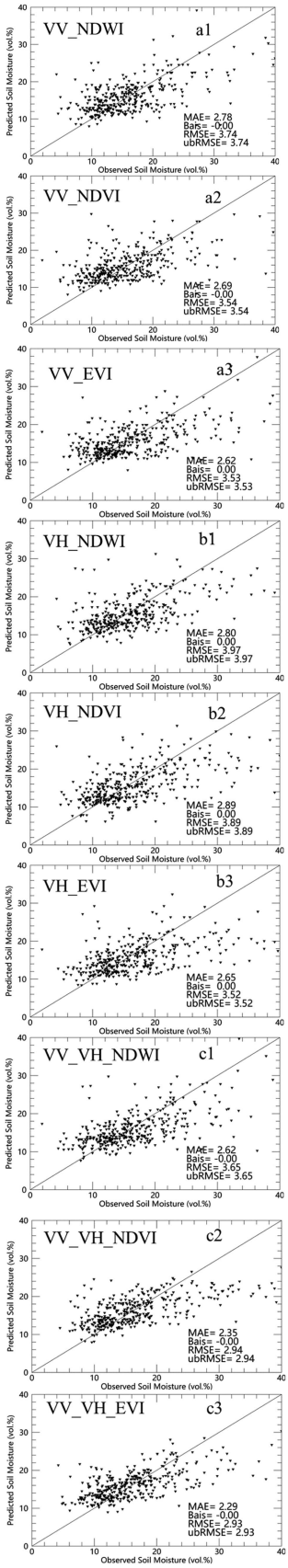


Fig. 6. Predicted SM and observed SM dispersion point diagram of AdaBoost: VV(a), VH(b), VV_VH(c); 1(NDWI), 2(NDVI), 3(EVI).

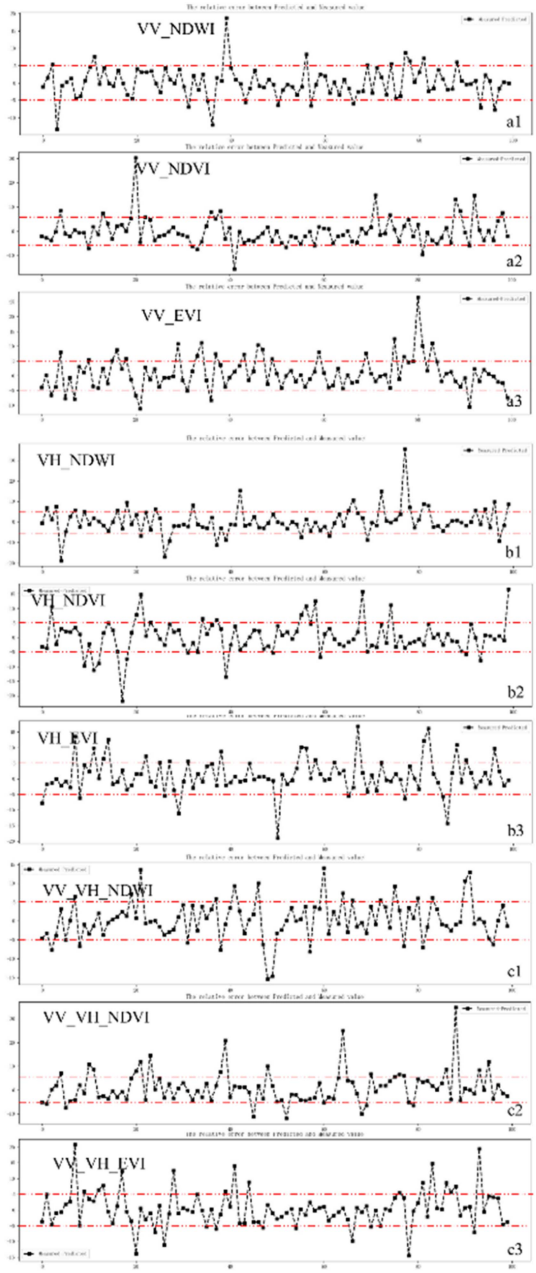


Fig. 7. Relative error between predicted and measured value of AdaBoost; VV(a), VH(b), VV_VH(c); 1(NDWI), 2(NDVI), 3(EVI).

RMSE of 3.54 vol.%, and ubRMSE of 3.54 vol.% [Fig. 6(a2)]. The VV_EVI indicates MAE of 2.62vol.%, RMSE of 3.53 vol.%, and ubRMSE of 3.53 vol.% [Fig. 6(a3)].

With VH as the input, we get the result shown in Fig. 7(b). MAE is 2.80 vol.%, RMSE is 3.97 vol.%, and ubRMSE is 3.97 vol.% with the input as VH_NDWI [Fig. 6(b1)]. For a reference VH_NDVI, the MAE on the mv estimates is 2.89 vol.%, the RMSE is 3.89 vol.%, and the ubRMSE is 3.89 vol.% [Fig. 6(b2)]. Fig. 6(b3) depicts MAE of 2.65 vol. %, RMSE of 3.52 vol.%, and ubRMSE of 3.52 vol.% with the input of VH_EVI.

Likewise, both the polarization (VV and VH) modes together serve as the input of AdaBoost. VV_VH_NDWI reflects MAE

of 2.62 vol.%, RMSE of 3.65 vol.%, and ubRMSE of 3.65 vol.% [Fig. 6(c1)]. The results of VV_VH_NDVI improve the MAE on SM estimates to 2.56 vol.%, the RMSE to 3.33 vol.%, and the ubRMSE to 3.33 vol. % [Fig. 6(c2)]. Similarly, the use of VV_VH_EVI, the results (MAE = 2.27 vol.%, RMSE = 3.14 vol.%, and ubRMSE = 3.14 vol. %) are higher with the VV_VH_NDWI and VV_VH_NDVI [Fig. 6(c3)].

Through the learning of all combinations of AdaBoost, the obtained inversion results show that, like RF, σ^0_{soil} obtained by calculating EVI as a vegetation description, performs the best inversion. At the same time, when VV and VH are both the input of AdaBoost, the inversion result has the highest accuracy.

We can see from the results (see Figs. 4 and 6) that the inverse SM is underestimated in relatively wet areas. This is not a limitation of the model, it is related to the radar signal. In wet areas, the radar signal is less sensitive to SM, so the predicted values are lower than the measured SM.

2) Relative Error Between Predicted and Measured Value of AdaBoost: The relative error between predicted and measured value of AdaBoost is described in Fig. 7. The distribution of error rates mostly ranges from –5 to 5. Similarly, due to the large number of samples, 100 samples were randomly selected for image display analysis. Looking at Fig. 7(a), we can get that, in the error rate range –5 to 5, VV_NDWI has 74% [Fig. 7(a1)], VV_NDVI has 84% [Fig. 7(a2)], and VV_EVI has 74% [Fig. 7(a3)]. As the input of VH, between the rate of –5 and 5, we can get 74% [VH_NDWI, Fig. 7(b1)], 80% [VH_NDVI, Fig. 7(b2)], and 78% [VH_EVI, Fig. 7(b3)]. VV_VH_NDWI is 76% [Fig. 7(c1)], VV_VH_NDVI is 78% [Fig. 7(c2)], and VV_VH_EVI is 77% [Fig. 7(c3)].

D. Pearson Correlation Coefficient and T-Test

In Sections III-A and III-B, we obtained the comparison results of SM inversion and measured SM using RF and AdaBoost methods. In this chapter, we carry out the correlation analysis of the Pearson correlation coefficient, which further proves the accuracy of our inversion results.

1) P-P Plot of RF: Fig. 8 shows a P-P plot of measured SM and predicted SM with RF. As can be seen from the figure, the points representing the sample data should essentially be on the diagonal line representing the theoretical distribution. The sample data should be distributed on and near the diagonal of the theoretical distribution. At this point, we can prove that the measured SM [Fig. 8(a)] and the predicted SM [Fig. 8(b)] conform to the normal distribution. The SM inversion results obtained by RF method can be correlated with the measured SM.

First, the description of the simple indicators of the predicted SM and the measured SM is described in Table III. The average and standard deviation of the two groups of data are calculated using SPSS, and the number of data in each group is counted. So we can see that the total number of samples is 370. The mean value and standard deviation of observed SM are 16.2175 and 6.55297, respectively. The mean value is 16.2253 and the standard deviation is 3.77122 of the predicted SM.

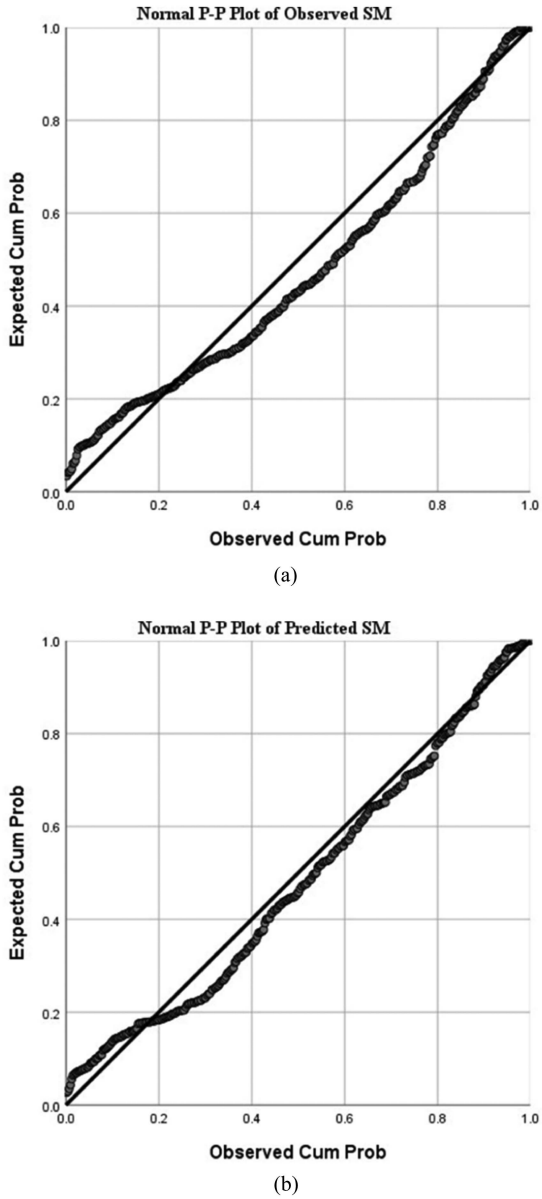


Fig. 8. Normally distributed P-P diagram of RF.

TABLE III
DESCRIPTIVE STATISTICS OF RF

	Mean	Std. Deviation	N
Observed SM	16.2175	6.55297	370
Predicted SM	16.2253	3.77122	370

Then, Table IV shows the consequences of the Pearson correlation significance analysis of RF. In this case, the correlation coefficient between observed SM and predicted SM was 0.627, and the Sig value (significance test result) was 0.000 ($P < 0.01$), showing that there was a vitally important moderate positive correlation between observed SM content and predicted SM content, which was consistent with our hypothesis.

TABLE IV
PEARSON CORRELATION SIGNIFICANCE ANALYSIS OF RF

		Observed SM	Predicted SM
Observed SM	Pearson Correlation	1	.627**
	Sig. (2-tailed)		0.000
	N	370	370
Predicted SM	Pearson Correlation	.627**	1
	Sig. (2-tailed)	0.000	
	N	370	370

**. Correlation is significant at the 0.01 level (2-tailed).

TABLE V
DESCRIPTIVE STATISTICS OF ADABOOST

	Mean	Std. Deviation	N
Observed SM	15.7766	6.12124	370
Predicted SM	15.8333	3.88558	370

TABLE VI
PEARSON CORRELATION SIGNIFICANCE ANALYSIS OF ADABOOST

		Observed SM	Predicted SM
Observed SM	Pearson Correlation	1	.588**

2) P-P Plot of AdaBoost: There are P-P plots of AdaBoost in Fig. 9. We can clearly observe that both measured data [Fig. 9(a)] and predicted data [Fig. 9(b)] are distributed on or very close to the theoretical line. The results of AdaBoost also conform to a normal distribution.

There are 370 samples of observed SM and predicted SM separately. Table V shows descriptive statistics of AdaBoost. The mean is 15.7766 and 15.8333 of observed SM and predicted SM, and the standard deviation is 6.12124 and 3.88558.

There is the result of the Pearson correlation significance analysis of AdaBoost in Table VI. The correlation coefficient between observed SM and predicted SM was 0.588, and the Sig value (significance test result) was 0.000 ($P < 0.01$), indicating an essential moderate positive correlation among observed SM with predicted SM. Through the Pearson correlation significance analysis, we can clearly prove that the inversion SM results obtained using AdaBoost have a strong correlation with the measured SM. It further shows that the use of AdaBoost can be very good for SM prediction.

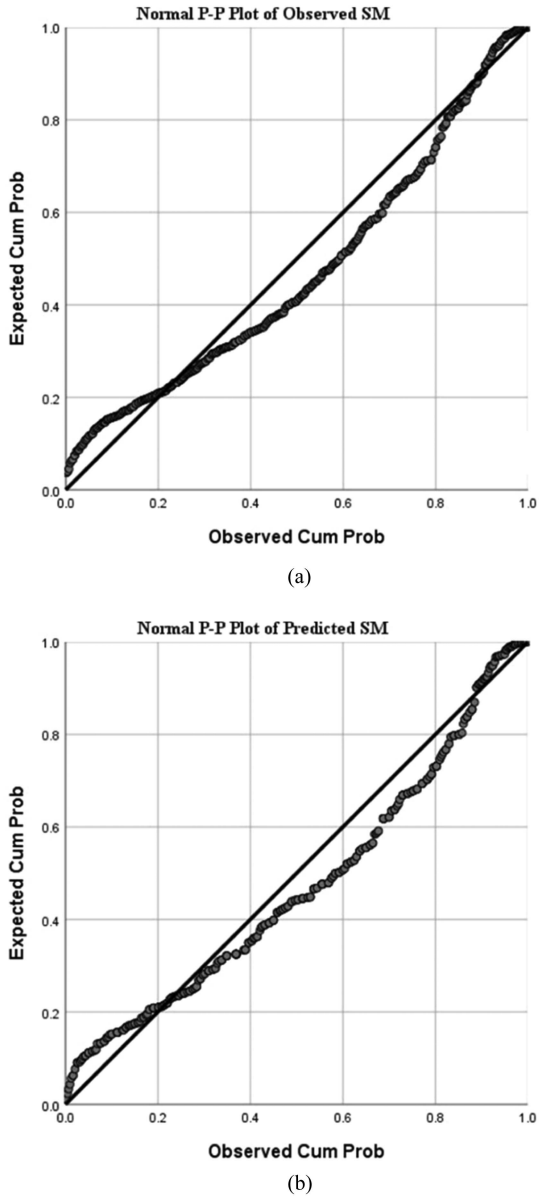


Fig. 9. Normally distributed P-P diagram of AdaBoost.

IV. DISCUSSION

In this study, we presupposed an algorithm in accordance with two ensemble learning models (RF and AdaBoost) to retrieve SM, incorporating radar data and ground datasets. We assume that ensemble learning models can be used to learn complex nonlinear relationships between SM extracted from multiple data sources and various environmental variables. There are

TABLE VII
MEASURED SM AND PREDICTED SM RESULTS WITHOUT DEM OF RF

	Sig. (2-tailed)		0.000
	N	370	370
Predicted SM	Pearson	.588**	1
	Correlation		
	Sig. (2-tailed)	0.000	
	N	370	370

**. Correlation is significant at the 0.01 level (2-tailed).

TABLE VIII
MEASURED SM AND PREDICTED SM RESULTS WITHOUT DEM OF ADABOOST

	MAE	Bais	RMSE	ubRMSE
VV_NDWI_RF	3.03	0.00	4.21	4.21
VV_NDVI_RF	2.92	0.00	3.87	3.87
VV_EVI_RF	2.93	-0.00	3.93	3.93
VH_NDWI_RF	2.48	-0.00	3.29	3.29
VH_NDVI_RF	3.02	0.00	3.79	3.79
VH_EVI_RF	2.81	0.00	3.82	3.82
VV_VH_NDWI_RF	2.94	-0.00	3.82	3.82
VV_VH_NDVI_RF	2.85	0.00	3.76	3.76
VV_VH_EVI_RF	2.44	-0.00	3.19	3.19

TABLE IX
MEASURED SM AND PREDICTED SM RESULTS WITHOUT SOIL ROUGHNESS OF RF

	MAE	Bais	RMSE	ubRMSE
VV_NDWI_AdaBoost	4.49	-0.00	5.99	5.99
VV_NDVI_AdaBoost	4.41	0.00	6.29	6.29
VV_EVI_AdaBoost	4.56	-0.00	6.40	6.40
VH_NDWI_AdaBoost	4.14	0.00	6.00	6.00
VH_NDVI_AdaBoost	3.96	0.00	5.67	5.67
VH_EVI_AdaBoost	4.21	-0.00	5.96	5.96
VV_VH_NDWI_AdaBoost	3.24	-0.00	4.34	4.34
VV_VH_NDVI_AdaBoost	2.98	0.00	4.06	4.06
VV_VH_EVI_AdaBoost	3.11	-0.00	4.17	4.17

TABLE X
MEASURED SM AND PREDICTED SM RESULTS WITHOUT SOIL ROUGHNESS OF ADABOOST

	MAE	Bais	RMSE	ubRMSE
VV_NDWI_AdaBoost	2.65	-0.00	4.08	4.08
VV_NDVI_AdaBoost	2.70	0.00	3.57	3.57
VV_EVI_AdaBoost	2.67	0.00	3.63	3.63
VH_NDWI_AdaBoost	3.31	0.00	5.07	5.07
VH_NDVI_AdaBoost	2.69	-0.00	3.91	3.91
VH_EVI_AdaBoost	3.30	-0.00	4.50	4.50
VV_VH_NDWI_AdaBoost	2.94	0.00	3.79	3.79
VV_VH_NDVI_AdaBoost	2.74	0.00	4.11	4.11
VV_VH_EVI_AdaBoost	2.82	-0.00	3.75	3.75

three inversion parts, as input, the radar signal provides only VV polarization data, the radar signal provides only VH polarization data, and both VV and VH were developed. And three vegetation parameters were used in the WCM (EVI, NDVI, and NDWI).

The developed RF and AdaBoost were trained with multisource datasets and validated using real and observed datasets. Training is performed with multisource datasets, whereas the real and observed databases are used for validation. The consequence displays that the RF algorithm can raise the accuracy of SM estimation. On the side, the results show that higher accuracy is obtained when VV and VH are used together than when VH and VV are used alone. At the same time, it was found that, when the vegetation parameter input by WCM is EVI, the calculated soil backscatter is obtained. When VV_VH_EVI and other datasets are used as input of the ensemble learning model, the best inversion results are obtained. Overall bias is nearly zero, and the correlation between predicted SM and measured SM was significant. In contrast, the best results of RF improved the accuracy of the predicted SM product (MAE = 2.289 vol.%, RMSE = 2.934 vol.%, and ubRMSE = 2.934 vol.%).

There are several differences between our retrieval algorithm and other machine learning SM retrieval studies. While several studies use radar backscatter of different polarizations as input to machine learning models, the results show that VV alone and VV and VH together have the same results and are better than VH alone [15], [60], [61]. While our method shows that using VV and VH together outperforms using VV alone. This is due to the difference in dataset selection and data resolution. The radar data, optical data, and DEM data we selected are all data with relatively high resolution, and some measured data are used as verification. DEM data were used in this study. We removed the DEM data from the dataset and reused the method of this article to invert the SM in the absence of DEM data. From the obtained results (see Tables VII and VIII), it can be seen that the inversion results obtained in the absence of DEM data are lower than those with DEM data. It indicates that DEM increases the accuracy of SM inversion and topography affects soil sensitivity.

To verify that soil roughness is important to improve the accuracy of soil inversion. We readded another set of experiments to observe the inversion accuracy when no soil roughness data were available. We obtained the following results (see Tables IX and X). We can find that the inversion results obtained are worse when there is no soil roughness data as input. It means that the microwave and the surface interact with each other and the soil roughness plays an important role.

There are some limitations in this study. First, in the establishment of datasets, there are some deficiencies in data quantity, limitations of measured data, and the absence of site data. In Figs. 5 and 6, we can see that although the analysis accuracy of the predicted SM results and measured SM results is relatively high, there is a certain underestimation when the SM exceeds about 25 vol.%. This may be due to the relatively small amount of data in this range when SM exceeds 25 vol.%, resulting in insufficient training and underestimation of prediction. In future research, we should increase the amount of data and look for data that is closely related to SM. Second, we chose the Luanhe Basin as the research area for inversion analysis, and the research area is relatively small. In the following study, we will expand the study area and increase the amount of data, measured data, and

auxiliary data. At the same time, we will increase the radar data of different time resolutions to further improve the accuracy of prediction.

V. Conclusion

In this article, we used Sentinel-1, Sentinel-2, DEM, and measured data as datasets for SM inversion analysis by ensemble learning algorithm (RF and AdaBoost). Since the study area is covered by vegetation, vegetation scattering affects soil scattering, so we use the WCM model to decrease the influence of vegetation scattering. Simultaneously, we selected three vegetation indices as vegetation parameters in the WCM model, namely EVI, NDVI, and NDWI. The main discoveries of this article are as follows.

1) Both generally integrated learning algorithms are trained on the ground of features extracted from multisource datasets. The optimal performance MAE, RMSE, and ubRMSE of the RF model are 2.289 vol%, 2.934 vol%, and 2.934 vol%, respectively, which are slightly better than the AdaBoost model.
2) Different conditions were developed (single VV, single VH, and both VV and VH) as the input of ensemble learning algorithm. The results show that inversion results using both VV and VH as inputs due to VV alone and VH alone.
3) When WCM model was used to remove vegetation influence, three vegetation indices (EVI, NDVI, NDWI) were selected, and the results showed that EVI as vegetation parameter had the best overall effect in the inversion results.

In conclusion, our results show that the ensemble learning method is effective for SM inversion. Future studies may consider integrating finer radar satellite data to expand the study area and obtain high-precision SM products.

Acknowledgment

Leading Talents Project of the State Ethnic Affairs Commission. The dataset is provided by National Tibetan Plateau Data Center (http://data.tpdc.ac.cn).

References

[1] J. Peng, A. Loew, O. Merlin, and N. E. C. Verhoest, "A review of spatial downscaling of satellite remotely sensed soil moisture," *Rev. Geophys.*, vol. 55, pp. 341–366, 2017.
[2] S. I. Seneviratne et al., "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Sci. Rev.*, vol. 99, pp. 125–161, 2010.
[3] H. Vereecken et al., "Soil hydrology: Recent methodological advances, challenges, and perspectives," *Water Resour. Res.*, vol. 51, pp. 2616–2633, 2015.
[4] M. E. Holzman, F. Carmona, R. Rivas, and R. Niclòs, "Early assessment of crop yield from remotely sensed water stress and solar radiation data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 297–308, 2018.
[5] A. Ghulam, Q. Qin, T. Teyip, and Z. Li, "Modified perpendicular drought index (MPDI): A real-time drought monitoring method," *ISPRS J. Photogrammetry Remote Sens.*, vol. 62, pp. 150–164, 2007.
[6] H. Lievens et al., "SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia," *Remote Sens. Environ.*, vol. 168, pp. 146–162, 2015.
[7] A. Berg and J. Sheffield, "Climate change and drought: The soil moisture perspective," *Curr. Climate Change Rep.*, vol. 4, pp. 180–191, 2018.
[8] M. Sadeghi, E. Babaeian, M. Tuller, and S. B. Jones, "The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations," *Remote Sens. Environ.*, vol. 198, pp. 52–68, 2017.
[9] W. Zhao, N. Sanchez, and A. Li, "Triangle space-based surface soil moisture estimation by the synergistic use of in situ measurements and optical/thermal infrared remote sensing: An alternative to conventional validations," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4546–4558, Aug. 2018.
[10] H. Cui et al., "Evaluation and analysis of AMSR-2, SMOS, and SMAP soil moisture products in the Genhe area of China," *J. Geophys. Res., Atmos.*, vol. 122, pp. 8650–8666, Aug. 2017.
[11] W. W. Verstraeten, F. Veroustraete, C. J. van der Sande, I. Grootaers, and J. Feyen, "Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests," *Remote Sens. Environ.*, vol. 101, pp. 299–314, 2006.
[12] P. Leng, X. Song, S. Duan, and Z. Li, "A practical algorithm for estimating surface soil moisture using combined optical and thermal infrared data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 338–348, 2016.
[13] D. Zheng et al., "Assessment of the SMAP soil emission model and soil moisture retrieval algorithms for a Tibetan Desert ecosystem," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3786–3799, Jul. 2018.
[14] N. Baghdadi, M. Zribi, C. Loumagne, P. Ansart, and T. Anguela, "Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields," *Remote Sens. Environ.*, vol. 112, pp. 4370–4379, Dec. 2008.
[15] M. El Hajj, N. Baghdadi, M. Zribi, and H. Bazzi, "Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas," *Remote Sens.*, vol. 9, Dec. 2017, Art. no. 1292.
[16] A. K. Fung, Z. Li, and K. S. Chen, "Backscattering from a randomly rough dielectric surface," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 356–369, Mar. 1992.
[17] K. S. Chen, T.-D. Wu, L. Tsang, Q. Li, J. Shi, and A. K. Fung, "Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 90–101, Jan. 2003.
[18] Y. Oh, K. Sarabandi, and F. T. Ulaby, "An empirical model and an inversion technique for radar scattering from bare soil surfaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 370–381, Mar. 1992.
[19] P. C. Dubois, J. van Zyl, and T. Engman, "Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 915–926, Jul. 1995.
[20] J. Shi, J. Wang, A. Y. Hsu, P. E. O'Neill, and E. T. Engman, "Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1254–1266, Sep. 1997.
[21] E. P. W. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Sci.*, vol. 13, no. 2, pp. 357–364, Mar./Apr. 1978.
[22] F. T. Ulaby, K. McDonald, K. Sarabandi, and M. C. Dobson, "Michigan microwave canopy scattering models (MIMICS)," in *Proc. Int. Geosci. Remote Sens. Symp., 'Remote Sens., Moving Toward 21st Century'*, 1988, p. 1009.
[23] E. Santi, M. Dabboor, S. Pettinato, and S. Paloscia, "Combining machine learning and compact polarimetry for estimating soil moisture from C-band SAR data," *Remote Sens.*, vol. 11, Oct. 2019, Art. no. 2451.
[24] Z. Xiao et al., "Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 209–223, Jan. 2014.
[25] K. Jia et al., "Global land surface fractional vegetation cover estimation using general regression neural networks from MODIS surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4787–4796, Sep. 2015.
[26] H. R. Mirsoleimani, M. R. Sahebi, N. Baghdadi, and M. El Hajj, "Bare soil surface moisture retrieval from Sentinel-1 SAR data based on the calibrated IEM and dubois models using neural networks," *Sensors*, vol. 19, Jul. 2019, Art. no. 3209.
[27] M. El Hajj, N. Baghdadi, and M. Zribi, "Comparative analysis of the accuracy of surface soil moisture estimation from the C- and L-bands," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 82, 2019, Art. no. 101888.

[28] Y. Gao, M. Gao, L. Wang, and O. Rozenstein, "Soil moisture retrieval over a vegetation-covered area using ALOS-2 L-band synthetic aperture radar data," *Remote Sens.*, vol. 13, 2021, Art. no. 3894.
[29] Y. Liu, J. Qian, and H. Yue, "Combined Sentinel-1A with Sentinel-2A to estimate soil moisture in farmland," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1292–1310, Jan. 2021.
[30] H. Cui et al., "The potential of ALOS-2 and Sentinel-1 radar data for soil moisture retrieval with high spatial resolution over agroforestry areas, China," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
[31] T. Zhao et al., "Soil moisture experiment in the Luan River supporting new satellite mission opportunities," *Remote Sens. Environ.*, vol. 240, 2020, Art. no. 111680.
[32] N. Baghdadi et al., "Potential of Sentinel-1 images for estimating the soil roughness over bare agricultural soils," *Water (Basel)*, vol. 10, 2018, Art. no. 131.
[33] N. Baghdadi, H. Bazzi, M. El Hajj, and M. Zribi, "Detection of frozen soil using Sentinel-1 SAR data," *Remote Sens.*, vol. 10, 2018, Art. no. 1182.
[34] S. Bousbih et al., "Soil moisture and irrigation mapping in a semi-arid region, based on the synergetic use of Sentinel-1 and Sentinel-2 data," *Remote Sens.*, vol. 10, 2018, Art. no. 1953.
[35] S. Bousbih et al., "Sentinel-1 and Sentinel-2 data for soil moisture and irrigation mapping over semi-arid region," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 7022–7025.
[36] S. Bousbih et al., "Sentinel-1 and Sentinel-2 data for the characterisation of the states of continental surface over a semi-arid region en Tunisia," in *Proc. Mediterranean Middle-East Geosci. Remote Sens. Symp.*, 2020, pp. 285–288.
[37] L. Lozac'h, H. Bazzi, N. Baghdadi, M. E. Hajj, M. Zribi, and R. Cresson, "Sentinel-1/Sentinel-2-derived soil moisture product at plot scale (S^2 MP)," in *Proc. Mediterranean Middle-East Geosci. Remote Sens. Symp.*, 2020, pp. 168–171.
[38] N. Djamai and R. Fernandes, "Comparison of SNAP-derived Sentinel-2A L2A product to ESA product over Europe," *Remote Sens.*, vol. 10, 2018, Art. no. 926.
[39] M. Arekhi, C. Goksel, F. Balik Sanli, and G. Senel, "Comparative evaluation of the spectral and spatial consistency of Sentinel-2 and Landsat-8 OLI data for Igneada Longos forest," *ISPRS Int. J. Geo-Inf.*, vol. 8, 2019, Art. no. 56.
[40] S. Bhatnagar et al., "Mapping vegetation communities inside wetlands using Sentinel-2 imagery in Ireland," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 88, 2020, Art. no. 102083.
[41] Z. Ju, M. Leong Tan, N. Samat, and C. K. Chang, "Comparison of Landsat 8, Sentinel-2 and spectral indices combinations for Google Earth engine-based land use mapping in the Johor River Basin, Malaysia," *Malaysian J. Soc. Space*, vol. 17, pp. 30–46, 2021.
[42] Z. Tianjie et al., "Synchronous observation data set of soil temperature and soil moisture in the upstream of Luan River (2018)," National Tibetan Plateau/Third Pole Environment Data Center, 2021.
[43] G. Peng, "Synchronous observation data set of soil surface roughness in the upstream of Luan River (2018)," National Tibetan Plateau/Third Pole Environment Data Center, 2021.
[44] M. F. Gómez, J. D. Lencinas, A. Siebert, and G. M. Díaz, "Accuracy assessment of ASTER and SRTM DEMs: A case study in Andean Patagonia," *GISci. Remote Sens.*, vol. 49, pp. 71–91, 2012.
[45] S. Zhao et al., "Accuracy assessment of the ASTER GDEM and SRTM3 DEM: An example in the Loess Plateau and North China Plain of China," *Int. J. Remote Sens.*, vol. 32, pp. 8081–8093, 2011.
[46] T. Tachikawa, M. Hato, M. Kaku, and A. Iwasaki, "Characteristics of ASTER GDEM version 2," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 3657–3660.
[47] X. Zhang et al., "Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth engine platform," *Earth Syst. Sci. Data*, vol. 12, pp. 1625–1648, 2020.
[48] L. Liu, X. Zhang, Y. Gao, X. Chen, X. Shuai, and J. Mi, "Finer-resolution mapping of global land cover: Recent developments, consistency analysis, and prospects," *J. Remote Sens.*, vol. 2021, 2021, Art. no. 5289697.
[49] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery," *Earth Syst. Sci. Data*, vol. 13, pp. 2753–2776, 2021.
[50] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[51] H. Ishwaran and J. D. Malley, "Synthetic learning machines," *BioData Mining*, vol. 7, 2014, Art. no. 28.
[52] E. Scornet, "Random forests and kernel methods," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1485–1500, Mar. 2016.
[53] A. Aulia, D. Jeong, I. M. Saaid, D. Kania, M. T. Shuker, and N. A. El-Khatib, "A random forests-based sensitivity analysis framework for assisted history matching," *J. Petroleum Sci. Eng.*, vol. 181, 2019, Art. no. 106237.
[54] A. Swami and R. Jain, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, 2013.
[55] P. Dou and Y. Chen, "Remote sensing imagery classification using Adaboost with a weight vector (WV Adaboost)," *Remote Sens. Lett.*, vol. 8, pp. 733–742, 2017.
[56] M. Nakamura, H. Nomiya, and K. Uehara, "Improvement of boosting algorithm by modifying the weighting rule," *Ann. Math. Artif. Intell.*, vol. 41, pp. 95–109, 2004.
[57] S. Merler, B. Caprile, and C. Furlanello, "Parallelizing Adaboost by weights dynamics," *Comput. Statist. Data Anal.*, vol. 51, pp. 2487–2498, 2007.
[58] S. Paloscia, S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci, "Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation," *Remote Sens. Environ.*, vol. 134, pp. 234–248, 2013.
[59] D. Alexakis, F. Mexis, A. Vozinaki, I. Daliakopoulos, and I. Tsanis, "Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A hydrological approach," *Sensors*, vol. 17, 2017, Art. no. 1455.
[60] M. E. Hajj, N. Baghdadi, M. Zribi, and H. Bazzi, "Coupling Sentinel-1 and Sentinel-2 images for operational soil moisture mapping," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5537–5540.
[61] N. Baghdadi, M. E. Hajj, and M. Zribi, "An operational high resolution soil moisture retrieval algorithm using Sentinel-1 images," in *Proc. Photon. Electromagn. Res. Symp. Spring*, 2019, pp. 4086–4092.
[62] W. G. Sombroek, "Towards a global soil resources inventory at scale 1:1 million," Int. Soil Ref. Inf. Centre, Wageningen, The Netherlands, Discuss. Paper 84/4, 1984.
[63] R. Bindlish and A. P. Barros, "Parameterization of vegetation backscatter in radar-based, soil moisture estimation," *Remote Sens. Environ.*, vol. 76, pp. 130–137, 2001.
[64] M. Zribi, V. Ciarletti, O. Taconet, J. Paille, and P. Boissard, "Characterisation of the soil structure and microwave backscattering based on numerical three-dimensional surface representation: Analysis with a fractional Brownian model," *Remote Sens. Environ.*, vol. 72, pp. 159–169, 2000.

Liguo Wang received the M.A. degree in mathematics education in 2002 and the Ph.D. degree in signal and information processing in 2005 from Harbin Institute of Technology, Harbin, China.

From 2006 to 2008, he was a Postdoctoral Researcher with the College of Information and Communications Engineering, Harbin Engineering University, Harbin, China. He is currently a Professor with the College of Information and Communications Engineering, Dalian Minzu University, Dalian, China, and a Doctor Supervisor with the College of Information and Communications Engineering, Harbin Engineering University. His research interests include remote sensing image processing and machine learning.

Ya Gao is currently working toward the Ph.D. degree in information and communication engineering with Harbin Engineering University, Harbin, China.

Her research interests include agricultural drought monitoring and early warning, remote sensing image processing, radar soil moisture inversion, machine learning, and neural networks.