remote sensing

MDPI

Article

Soil Moisture Inversion Based on Data Augmentation Method Using Multi-Source Remote Sensing Data

Yinglin Wang [1,2,3], Jianhui Zhao [1,2,3], Zhengwei Guo [1,2,3,*], Huijin Yang [1,2,3] and Ning Li [1,2,3]

[1] College of Computer and Information Engineering, Henan University, Kaifeng 475004, China
[2] Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475004, China
[3] Henan Engineering Research Center of Spatial Information Processing, Kaifeng 475004, China
* Correspondence: gzw@henu.edu.cn

Abstract: Soil moisture is an important land environment characteristic that connects agriculture, ecology, and hydrology. Surface soil moisture (SSM) prediction can be used to plan irrigation, monitor water quality, manage water resources, and estimate agricultural production. Multi-source remote sensing is a crucial tool for assessing SSM in agricultural areas. The field-measured SSM sample data are required in model building and accuracy assessment of SSM inversion using remote sensing data. When the SSM samples are insufficient, the SSM inversion accuracy is severely affected. An SSM inversion method suitable for a small sample size was proposed. The alpha approximation method was employed to expand the measured SSM samples to offer more training data for SSM inversion models. Then, feature parameters were extracted from Sentinel-1 microwave and Sentinel-2 optical remote sensing data, and optimized using three methods, which were Pearson correlation analysis, random forest (RF), and principal component analysis. Then, three common machine learning models suitable for small sample training, which were RF, support vector regression, and genetic algorithm-back propagation neural network, were built to retrieve SSM. Comparison experiments were carried out between various feature optimization methods and machine learning models. The experimental results showed that after sample augmentation, SSM inversion accuracy was enhanced, and the combination of utilizing RF for feature screening and RF for SSM inversion had a higher accuracy, with a coefficient of determination of 0.7256, a root mean square error of 0.0539 cm^3/cm^3, and a mean absolute error of 0.0422 cm^3/cm^3, respectively. The proposed method was finally used to invert the regional SSM of the study area. The inversion results indicated that the proposed method had good performance in regional applications with a small sample size.

Keywords: surface soil moisture; synthetic aperture radar; data augmentation; feature optimization; machine learning

check for updates

Citation: Wang, Y.; Zhao, J.; Guo, Z.; Yang, H.; Li, N. Soil Moisture Inversion Based on Data Augmentation Method Using Multi-Source Remote Sensing Data. *Remote Sens.* **2023**, *15*, 1899. https://doi.org/10.3390/rs15071899

Academic Editor: Jeroen Meersmans

Received: 21 February 2023
Revised: 29 March 2023
Accepted: 30 March 2023
Published: 31 March 2023

1. Introduction

Surface soil moisture (SSM) is a critical land environment variable that connects agriculture, ecology, and hydrology, as well as a key parameter in hydrology, meteorology, and agricultural applications. SSM monitoring can be used to plan irrigation, monitor water quality, manage water resources, and estimate crop yield [1,2]. Understanding the spatial and temporal distribution and dynamic changes of SSM can help guide agricultural management.

Traditional SSM monitoring often employs the gravimetric method or the probe method. Although the precision is reasonably good and the operation is simple, it necessitates a significant amount of personnel and material resources, and is easily influenced by the surrounding environment and human variables. Furthermore, because the number of sample locations is restricted, it is hard to obtain a substantial amount of SSM information in a short period of time [3].

Remote sensing technology offers a potent approach to detecting SSM on a broad scale and with great spatial-temporal resolution. Synthetic aperture radar (SAR) is a promising method for assessing SSM with high spatial-temporal resolution [4,5]. In contrast to optical remote sensing, SAR does not require sunshine, and microwave signals may penetrate the surface soil to estimate and monitor SSM in real time [6]. SAR data demonstrate the vast potential and promising practice of mapping global SSM at medium and high spatial resolution [7]. SAR is sensitive to the dielectric and geometric properties of the target [8,9]. Fung et al. [10] built an integral equation model to estimate soil moisture. Empirical models such as the Oh model established by Oh et al. [11] and the Dubois model established by Dubois et al. [12] can estimate soil moisture within their effective range. Bao et al. [13] modified the water cloud model (WCM) based on optical indicators, and introduced the vegetation index to reduce the impact of vegetation cover on SSM. Compared with empirical, semi-empirical, and theoretical models, machine learning can avoid complex physical relationships and solve nonlinear problems, and is widely used in SSM inversion. Gao et al. [14] used Sentinel-1 and Sentinel-2 data to determine SSM using the change detection method. Guo et al. [15] used Sentinel-1 and Sentinel-2 data to determine SSM using support vector regression (SVR) and generalized regression neural network (GRNN) methods. Datta et al. [16] compared the applicability of different machine learning and linear regression models in SSM inversion using Sentinel-1 and Sentinel-2 data.

Because artificial neural network (ANN) has high nonlinear fitting abilities and can learn autonomously, it is increasingly being employed to solve the problem of SSM inversion. Arnicola et al. [17] discovered that by increasing the number of input ANN characteristics, the SSM inversion accuracy may be gradually increased. Pasolli et al. [18,19] applied the SVR model to retrieve SSM using microwave remote sensing data. Using different input parameters can also increase the accuracy of SSM inversion. Said et al. [20] estimated SSM using an ANN with several input parameters. Multiple regression is inferior to ANN inversion. In addition to traditional machine learning methods, many deep learning methods have also been employed in SSM monitoring in recent years. Cai et al. [21] developed an SSM prediction model using a deep learning regression network (DNNR) with big data fitting capabilities. To obtain reliable results, the deep learning method requires a large number of training samples.

In the case of small samples, it is critical to select the suitable machine learning model and then refine the model parameters. When there are too many input factors, screening some distinctive parameters can significantly enhance the accuracy of soil moisture inversion. Lin et al. [22] inverted the SSM of winter wheat fields using RADARSAT-2 data and polarization decomposition method to enhance the number of input factors, and used several feature selection and machine learning methods to improve the model performance and estimate SSM effectively and accurately. Zhang et al. [23] extracted several features from passive microwave remote sensing data, optical remote sensing data, land surface model (LSM) and other auxiliary data, assessed the value of different features to SSM retrieval, and then proposed an SSM retrieval method based on random forest (RF) model.

In practical applications, most machine learning techniques require amounts of sample data to assure adequate training. When there are few training samples, the model trained with tiny samples is prone to over-fitting of small samples and under-fitting target tasks. Therefore, when the number of samples is insufficient, increasing the sample size is a crucial way to raise inversion accuracy. Based on multi-time camera-borne SAR and ground measurement data and the change detection theory, Balenzano et al. [24] investigated the link between SSM changes and SAR signal changes of two crops in different wave bands, polarizations, and incident angles, and provided the quantitative equation that connects them, i.e., the alpha approximation method. He et al. [25] expanded the alpha approximation approach by using a time series of L-band SAR data and simultaneous ground observations from SMAPEx-3 to retrieve SSM. Xu et al. [26] used the alpha approximation method to augment the measured data for training the SVR model and further improved the SSM inversion accuracy. However, the input parameters and machine learning models used in

these studies were specified in advance, lacking more optimizations of input parameters and inversion models to improve the SSM inversion accuracy further.

There are various constraints in SSM inversion for a small size of sample data. To improve the accuracy of SSM inversion for small samples, an SSM inversion method combining sample augmentation, feature optimization, and machine learning models was investigated in this paper. Firstly, assuming that the surface roughness and vegetation conditions remain unchanged in the short term, the field-measured SSM data were augmented by using the alpha approximation method to provide more training data for the machine learning models. Secondly, feature parameters were extracted from Sentinel-1 and Sentinel-2 remote sensing data, and optimized by using Pearson correlation analysis, RF, and principal component analysis (PCA) methods. Then, three common machine learning models suitable for small sample training, which were genetic algorithm-back propagation neural network (GA-BP), SVR, and RF, were built to retrieve SSM and evaluate the accuracy. Finally, after comparing various combinations of feature optimization methods and machine learning models, the optimal inversion model was chosen to retrieve the regional SSM of the study area.

2. Materials and Methods

2.1. Study Area and Sampling Procedures

The study area was the eastern part of the Danjiangkou Ecological Service Area, which spanned Henan and Hubei provinces, China. The Danjiangkou Reservoir is the water source of the Middle Route Project of South-to-North Water Transfer. The Danjiangkou Ecological Service Area is a national first-class water source protection zone that was declared as one of China's ecological function protection zones in 2015. Its landscape is sloping from northwest to southeast, with low mountains in the northwest, hills in the center, and hills and alluvial plains in the southeast. The soil types in the study area are mainly yellow-brown soil and brown soil [27]. The study area has a monsoon environment ranging from the north subtropical zone to the warm temperate zone, with a mild climate, and four distinct seasons. In recent years, the annual precipitation here is about 800 mm to 1300 mm. It is a transitional zone between north and south, with a wide range of vegetation types and an abundance of plant resources. The study area is mostly made up of agricultural land, building land, and a body of water, as shown in Figure 1.

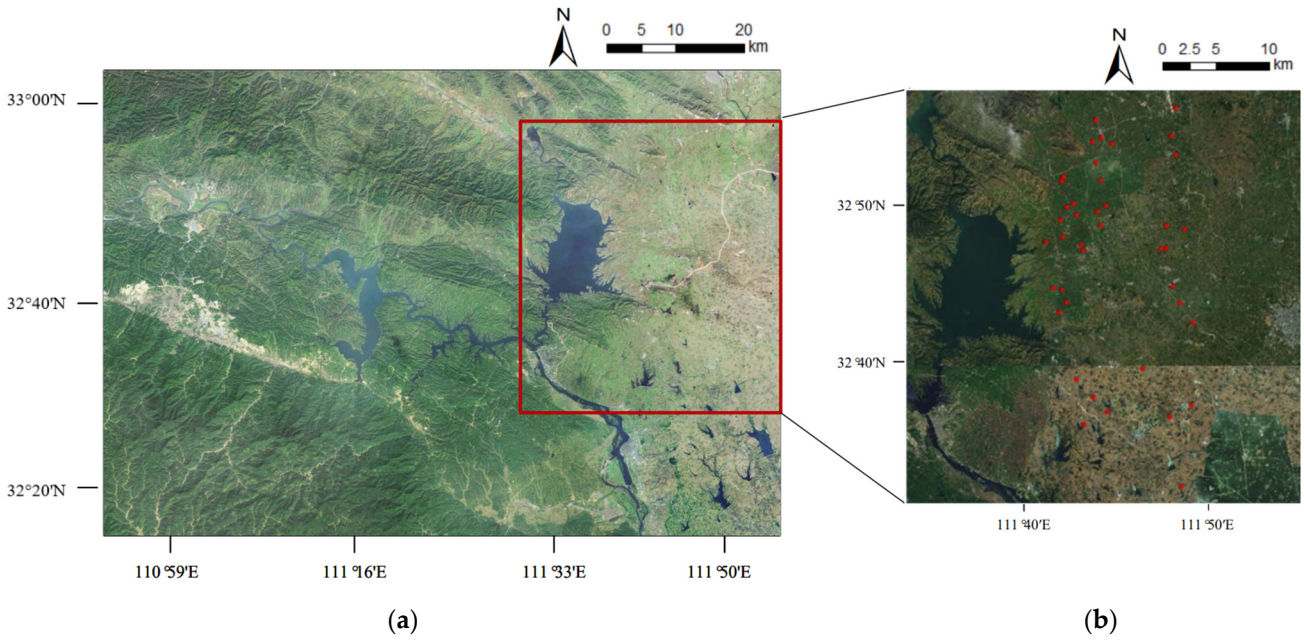


Figure 1. Location of the study area and the sampling points: (**a**) the Danjiangkou ecological service area; (**b**) the study area and sampling points.

Sentinel-1 SAR remote sensing images used in this study were acquired on 3 dates, which were 11 September, 23 September, and 5 October 2021. The obtained SAR images were preprocessed using the Sentinel application platform (SNAP) software from European Space Agency (ESA), including radiometric calibration, multi-viewing, Refined Lee filtering, and terrain correction. Simultaneously, the PolSARpro software was used to decompose Sentinel-1 SAR data to extract polarization information.

Sentinel-2 optical remote sensing images used in this study were acquired on 3 dates quasi-synchronous with Sentinel-1 data, which were 12 September, 22 September, and 2 October 2021. All Sentinel-2 data were L2A products with 12 bands. Details and acquisition dates for Sentinel-1 and Sentinel-2 image data utilized in this study are shown in Table 1.

Table 1. Remote sensing data information.

Data Source	Acquisition Date	Product Type	Polarization Mode
Sentinel-1 (SAR Data)	11 September 2021 23 September 2021 5 October 2021	1W SLC + GRD	Dual-Polarization
Sentinel-2 (Optical Data)	12 September 2021 22 September 2021 2 October 2021	L2A	

A field survey was carried out on 23 September 2021. A total of 41 sample points were set-up in the study area, as shown in Figure 1. Data gathered in the field included SSM value and coordinates of each sampling point. A portable TDR350 SSM meter was used to measure field SSM value. At each sampling point, the volumetric soil moisture content of the farmland surface layer was measured 5 times at 5 different places in a cross shape, and the average value of these 5 SSM values was used as the final measured SSM value at this sampling point. An outdoor portable UG905 locator with a positioning accuracy of 1 to 3 m was used to determine the latitude and longitude of each sampling point. The WGS84 coordinate system was used to record the coordinate of each sampling point.

2.2. Methods

The technical roadmap of the proposed method is shown in Figure 2.

The first step was data augmentation. The alpha approximation method was used to increase the sample size.

The second step was feature extraction. To obtain the necessary characteristic parameters, Sentinel-1 SAR data were preprocessed and H/A/αpolarization decomposition was carried out. The band data were extracted from the Sentinel-2 optical data, and the corresponding vegetation indices were calculated as the characteristic parameters.

The third step was feature optimization. The extracted feature parameters were optimized using 3 methods separately, including Pearson correlation analysis, RF, and PCA. The most advantageous feature subset was chosen based on the correlation between the characteristic parameters and the field-measured SSM values.

The fourth step was model building. To guarantee the training and inversion correctness of the models, GA-BP, SVR, and RF models were built and tweaked individually.

The fifth step was accuracy assessment. The inversion accuracy of 9 combinations of feature optimization methods and machine learning models was evaluated, and the optimal combination was chosen to retrieve the regional SSM of the study area.

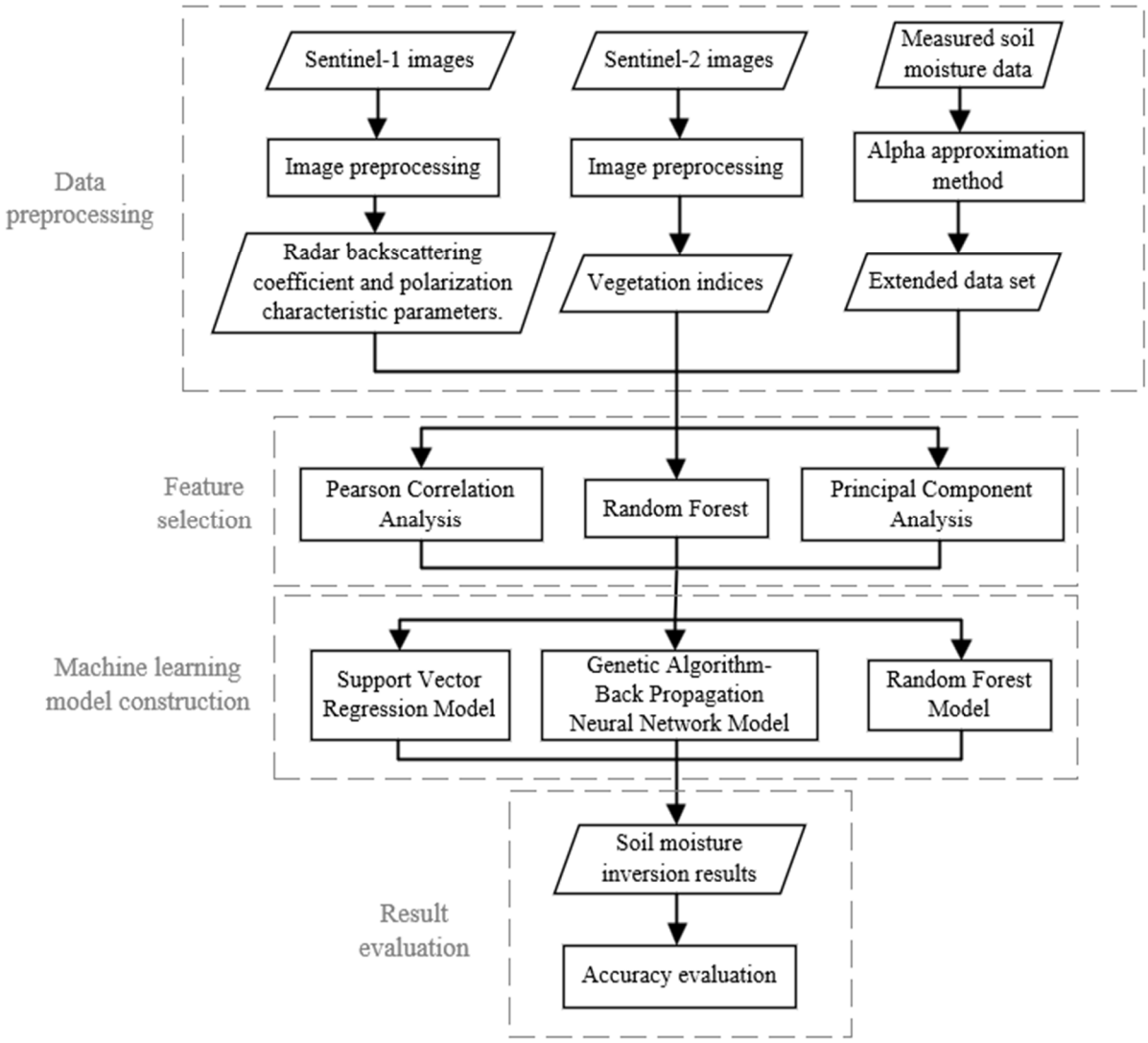


Figure 2. Technology roadmap of the proposed method.

2.2.1. Data Augmentation

For the problem of SSM inversion accuracy affected by the small size of the field measured SSM sample data, the alpha approximation method was adopted in this study to expand the sample size.

The alpha approximation method was proposed by Balenzano et al. [23]. Assuming that vegetation conditions and surface roughness remain constant throughout time, the change in backscattering is only affected by changes in soil moisture [25]. The quantitative link between backscattering coefficients and SSM is defined as Equations (1)–(3).

$$\frac{\sigma_0^2}{\sigma_0^1} \approx \left| \frac{\alpha_{pp}^2(\varepsilon_s, \theta)}{\alpha_{pp}^1(\varepsilon_s, \theta)} \right|^2 \tag{1}$$

$$\alpha_{HH}(\varepsilon_s, \theta) = \left| \frac{\varepsilon_s - 1}{\left(cos\theta + \sqrt{\varepsilon_s - sin^2\theta}\right)^2} \right| \tag{2}$$

$$\alpha_{VV}(\varepsilon_s, \theta) = \left| \frac{(\varepsilon_s - 1)\left[sin^2\theta - \varepsilon_s\left(1 + sin^2\theta\right)\right]}{\left(cos\theta + \sqrt{\varepsilon_s - sin^2\theta}\right)^2} \right| \tag{3}$$

where σ_0^i is the backscattering coefficient at time i, θ is the incidence angle, ε_s is the soil dielectric constant, PP is the polarization (HH or VV), and α_{PP} is a function of the soil dielectric constant and the incident angle.

Equation (1) can be written as Equation (4).

$$\left|\alpha_{PP}^1(\varepsilon_s,\theta)\right| - \sqrt{\frac{\sigma_0^1}{\sigma_0^2}}\left|\alpha_{pp}^2(\varepsilon_s,\theta)\right| = 0 \tag{4}$$

When N successive SAR image scenes are employed, the $N-1$ equations are summarized as Equation (5).

$$\begin{bmatrix} 1 & -\sqrt{\frac{\sigma_0^1}{\sigma_0^2}} & 0 & & \cdots & 0 & 0 \\ 0 & 1 & & -\sqrt{\frac{\sigma_0^2}{\sigma_0^3}} & & 0 & 0 \\ & \vdots & & & \ddots & & \vdots \\ & 0 & 0 & 0 & \cdots & 1 & -\sqrt{\frac{\sigma_0^{N-1}}{\sigma_0^N}} \end{bmatrix}_{(N-1)\times N} \begin{bmatrix} \alpha_{PP}^1(\varepsilon_s,\theta) \\ \alpha_{PP}^2(\varepsilon_s,\theta) \\ \alpha_{PP}^3(\varepsilon_s,\theta) \\ \cdots \\ \alpha_{PP}^N(\varepsilon_s,\theta) \end{bmatrix}_{N\times 1} \approx \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}_{(N-1)\times 1} \tag{5}$$

Equation (5) can be expressed as Equation (6) when three SAR image scenes are available.

$$\begin{cases} \left|\alpha_{PP}^1(\varepsilon_s,\theta)\right| - \sqrt{\frac{\sigma_0^1}{\sigma_0^2}}\left|\alpha_{pp}^2(\varepsilon_s,\theta)\right| \approx 0 \\ \left|\alpha_{PP}^2(\varepsilon_s,\theta)\right| - \sqrt{\frac{\sigma_0^2}{\sigma_0^3}}\left|\alpha_{pp}^3(\varepsilon_s,\theta)\right| \approx 0 \end{cases} \tag{6}$$

where σ_0^1, σ_0^2 and σ_0^3 can be acquired from the 3 Sentinel-1imgaes that are currently accessible. This indicates that there are 3 unknown parameters (specifically $\left|\alpha_{PP}^1(\varepsilon_s,\theta)\right|$, $\left|\alpha_{PP}^2(\varepsilon_s,\theta)\right|$, $\left|\alpha_{PP}^3(\varepsilon_s,\theta)\right|$) that need to be determined. After obtaining $\left|\alpha_{PP}^1(\varepsilon_s,\theta)\right|$ as a prior information through ground estimates, $\left|\alpha_{PP}^2(\varepsilon_s,\theta)\right|$ and $\left|\alpha_{PP}^3(\varepsilon_s,\theta)\right|$ can be acquired using Equation (6).

Because the premise of this study was that vegetation conditions and surface roughness remain intact in a short period of time, Sentinel-1A data with a repetition period of 12 days is suitable for the experiment. In this paper, a field survey was carried out on 23 September 2021, the same day when Sentinel-1 satellite transited over the study area. When the other two Sentinel-1 scenarios are known and taken as previous knowledge, the SSM data on 11 September 2021, and 5 October 2021, can be simply inversed using the empirical expression of Equation (6).

In the field survey, 41 measured SSM samples were obtained. In the subsequent experiment, the measured samples were randomly split into two sets, which were the training set with 26 samples and the testing set with 15 samples. Only the training set was expanded using the alpha approximation method. The testing set remained unchanged and was utilized to assess the experimental accuracy. A total of 93 sampling points were obtained after data augmentation. To get rid of the interference, any outliers that may be present in the expanded data were removed.

2.2.2. Feature Parameter Extraction

The training accuracy of a machine learning model is highly connected to the number and quality of the training data. The model will converge too slowly if there is too much training data. This can impair the model's ability to train on its own, lead to incorrect predictions, and lower the model's accuracy. The prediction accuracy of the machine learning model can be increased while reducing consumption by analyzing the feature parameter set and choosing the feature parameters with strong correlation as the input data.

- Polarization Feature Parameters

SAR works by sending microwave beams to objects and picking up echoes from those items to identify distinguishing traits. Radar information is directly impacted by both object characteristics and radar parameters, including the target object's physical characteristics and the wavelength, incidence angle, and polarization mode [28].

The incident angle (θ), VV, and VH polarization backscattering coefficients were extracted from the preprocessed Sentinel-1 SAR data and used as the defining parameters of the following experiments based on the latitude and longitude of each sampling point.

Both $cos(\theta)$ and $sin(\theta)$ are connected to soil moisture [29]. The correlation between the backscattering coefficient and $sin(\theta)$ is larger in soils with higher soil moisture levels, and the correlation between the backscattering coefficient and $cos(\theta)$ is higher in soils with lower soil moisture levels. When the incident angle is constant, the backscattering coefficient increases with the increase of volumetric soil moisture content, and the combination of different polarization backscattering coefficients of (σ^0_{VV} + σ^0_{VH}), ($\sigma^0_{VH} - \sigma^0_{VV}$), ($\sigma^0_{VV} \times \sigma^0_{VH}$) and ($\sigma^0_{VH}/\sigma^0_{VV}$) are also increased. More characteristic parameters from SAR remote sensing data can be extracted using polarization decomposition. H/A/α decomposition is used for eigenvalue decomposition of coherent matrix or covariance matrix of target features on Sentinel-1 dual polarization data, from which scattering entropy (H), inverse entropy (A), average scattering angle (α) and eigenvalues (λ_1 and λ_2) can be extracted [30].

- Vegetation Indices

Many vegetation indices can be generated from optical remote sensing data to describe surface vegetation information [28]. The backscattering coefficient of SAR is not only related to its own polarization mode, incidence angle, and SSM, but also to the vegetation coverage and roughness of the surface. It is necessary for SSM inversion to remove or weaken the impact of vegetation and surface roughness. The vegetation index is the combination of ground reflectivity in two or more wavelength bands to accentuate a certain feature or detail of plants. Varied vegetation indices have different band application ranges and fields due to sensor kinds and band combinations.

According to the multi-band data provided by the multispectral imager (MSI) carried by Sentinel-2 and the actual vegetation coverage in the study area, six vegetation indices commonly used in SSM inversion research, including normalized difference vegetation index (NDVI), normalized difference moisture index (NDWI), specific vegetation index (RVI), water stress index (MSI), water body index (WBI) and fused vegetation index (FVI) [31], were finally selected for this study. Their calculation formulas are shown in Equations (7)–(12).

$$NDVI = \frac{\rho_{842} - \rho_{665}}{\rho_{842} + \rho_{665}} \tag{7}$$

$$NDWI = \frac{\rho_{842} - \rho_{1610}}{\rho_{842} + \rho_{1610}} \tag{8}$$

$$RVI = \frac{\rho_{842}}{\rho_{665}} \tag{9}$$

$$MSI = \frac{\rho_{1610}}{\rho_{842}} \tag{10}$$

$$WBI = \frac{\rho_{865}}{\rho_{945}} \tag{11}$$

$$FVI = \frac{2\rho_{842} - \rho_{665} - \rho_{1610}}{2\rho_{842} + \rho_{665} + \rho_{1610}} \tag{12}$$

where ρ_{490}, ρ_{665}, ρ_{842}, ρ_{865}, ρ_{945} and ρ_{1610} represent the band values corresponding to 490, 665, 842, 865, 945, 1610 nm in Sentinel-2 data, respectively. The 490 nm and 665 nm bands represent the blue and red of visible light. The 842 nm and 865 nm bands represent Near

Infrared (NIR) and Narrow NIR. The 945 nm band represents water vapor. The 1610 nm band represents Short Wave Infrared (SWIR).

- Surface Roughness

The surface roughness of the soil influences the microwave backscattering coefficient. The surface roughness information changes depending on the band frequency, incident angle, and polarization mode. Removing the influence of surface roughness on SSM inversion can increase accuracy. The surface combined roughness *Zs* [32] used in this study was calculated using SAR data and represented by Equations (13)–(15).

$$Z_s = \exp\left(\frac{\sigma^0_{HV} - \sigma^0_{VV} - B_v(\theta)}{A_v(\theta)}\right) \tag{13}$$

$$A_v = -2.6408sin^3(\theta) + 5.293sin^2(\theta) - 3.838\sin(\theta) + 2.2042 \tag{14}$$

$$B_v = 4.1522sin^3(\theta) - 13.1sin^2(\theta) + 16.9472\sin(\theta) - 16.4228 \tag{15}$$

where A_v and B_v are coefficients only applicable to the combined roughness model using C-band data, and only change with incident angle.

A total of 21 feature parameters were extracted from Sentinel-1 and Sentinel-2 data, as shown in Table 2.

Table 2. Summary of parameters extracted from Sentinel-1 and Sentinel-2 data.

No.	Parameter	No.	Parameter	No.	Parameter
1	θ	8	$\sigma^0_{VV} \times \sigma^0_{VH}$	15	NDVI
2	σ^0_{VV}	9	$\sigma^0_{VH}/\sigma^0_{VV}$	16	NDWI
3	σ^0_{VH}	10	H	17	RVI
4	cos(θ)	11	A	18	MSI
5	sin(θ)	12	α	19	WBI
6	$\sigma^0_{VV} + \sigma^0_{VH}$	13	λ_1	20	FVI
7	$\sigma^0_{VH} - \sigma^0_{VV}$	14	λ_2	21	Zs

2.2.3. Feature Parameter Optimization

In this study, the extracted feature parameters were evaluated using Pearson correlation analysis, RF, and PCA methods separately to select the suitable feature parameter subset for the subsequent machine learning models.

- Pearson Correlation Analysis Method

The Pearson correlation coefficient, which has a value between −1 and 1, is the simplest approach to determine whether two variables are linearly connected. The sign indicates the positive-negative correlation. The closer its absolute value is to 1, the stronger the linear association between the two variables is. Conversely, the closer it is to 0, the weaker the linear relationship between the two variables is. It is calculated using Equation (16).

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{16}$$

where *Cov* (*X*, *Y*) represents the covariance of two variables *X* and *Y*, *Var*(*X*) is the variance of *X*, and *Var*(*Y*) is the variance of *Y*.

- Random Forest Method

The RF method can calculate the relevance of each variable during the model-building process [33]. In feature selection process with RF, the importance of each feature is first calculated and arranged in descending order. The proportion to be deleted is then established, and the matching proportion of characteristics is eliminated based on their relevance,

yielding a new feature set. The preceding procedure with a new feature set is repeated until only m features remained, among which m is a preset value. Finally, the feature set with the lowest error rate is chosen based on each feature set acquired in the preceding process and its related error rate.

- Principal Component Analysis Method

PCA is a method for finding a way to minimize the dimension of data while minimizing information loss [34]. It is an important tool in data analysis and frequently used in machine learning to minimize the dimension of high-dimensional data, since it can extract the key characteristic variables from the data. Each vector has a correlation in high-dimensional data sets, whereas it has a linear independence in low-dimensional data sets, allowing the overlapping information in high-dimensional data sets to be removed [35]. High-dimensional data are reduced to fulfill the goals of data dimensionality reduction, compression, and noise reduction. The data dimensionality is reduced, but the most relevant information is maintained, and certain unimportant aspects are deleted.

2.2.4. Construction of Machine Learning Models

Machine learning excels in nonlinear fitting. It is useful in resolving issues with excessive factors and convoluted structures in SSM inversion models. Even after sample augmentation, the number of samples is still limited due to the small number of field-measured SSM samples. Three field-measured typical machine learning models, GA-BP, SVR, and RF, that are appropriate for small sample training, were chosen for the study in order to prevent over-fitting.

- Genetic Algorithm–Back Propagation Neural Network Model

Both neural networks and evolutionary algorithms are ways for imitating biological treatment modes and obtaining practical answers to complicated issues. The BP neural network is capable of adaptive learning and powerful nonlinear simulation. However, it is prone to local minima. In addition, the network's design is not theoretically guided and is instead dependent on the designers' expertise and repeated experimentation in the sample space, which restricts the network's ability to find the overall optimal solution. GA can converge to the global optimal solution and has strong stability, but it lacks adaptive learning capabilities. As a result, combining a neural network with the genetic algorithm can enhance not only the neural network's ability to generalize mapping, but also its rate of convergence, capacity for global optimization, and learning capacity [36]. The entire prediction model is extensively upgraded in terms of accuracy and fitting capacity.

- Support Vector Regression Model

SVR is a regression analysis technique that uses the support vector machine (SVM). The majority of the sample points are situated outside the two decision boundaries thanks to the separation hyperplane that SVM discovers by maximizing the interval. In contrast to SVM, SVR also takes into account the maximum interval, but it also takes into account the points within the decision boundary to ensure that the majority of the sample points are situated within the interval. The most significant advantage of SVR is that it uses the kernel function rather than the inner product operation in high-dimensional space, transforming a high-dimensional nonlinear regression problem into a two-dimensional linear regression problem [37].

- Random Forest Model

RF can be used not only for parameter optimization, but also for parameter inversion. It is an integrated algorithm based on decision trees, with each decision tree acting as a classifier. When decision trees are being trained, randomness is incorporated, and samples and features are chosen at random. There will be n trees with n classification outcomes for each input sample. All the RF-categorized voting results are combined and the one with the most votes is chosen as the final result. In this process, integration and randomness coexist.

RF model has the advantages of increasing prediction accuracy, decreasing over-fitting, and being unaffected by missing data and multi-collinearity [38]. One advantage of RF is that it has good generalization performance due to the use of multiple regression trees, which helps to reduce model variability. It simply has two parameters, the number of trees and the number of features, therefore it doesn't require complicated parameter adjustment [17].

3. Results

In this section, the results of sample augmentation and SSM inversion were analyzed, and the spatial distributions of regional SSM of the study area were obtained.

3.1. Sample Augmentation Results

The inversion accuracy before and after sample augmentation was compared using the same input parameters by various inversion methods, to confirm the efficacy of sample augmentation and the benefits of various machine learning models for SSM inversion. The incident angle (θ), the VV and VH polarization backscattering coefficients, and NDVI were chosen as the input parameters in this experiment, while GA-BP, SVR, and RF models were employed as the SSM inversion models.

In this experiment, three precision evaluation indexes, which were determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE), were employed to assess SSM inversion accuracy. The average values acquired after repeated experiments were recorded as the experimental results to lessen the randomness of the outcomes, as indicated in Table 3.

Table 3. Comparison of SSM inversion accuracy before and after sample augmentation.

Method	Model	R^2	RSME (cm^3/cm^3)	MAE (cm^3/cm^3)
Using original measured SSM samples before augmentation	GA-BP	0.3868	0.0757	0.0719
	SVR	0.3258	0.0708	0.0566
	RF	0.4802	0.0667	0.0522
Using extended measured SSM samples after augmentation	GA-BP	0.5411	0.0606	0.0607
	SVR	0.4484	0.0644	0.0546
	RF	0.5906	0.0578	0.0488

According to the experimental results, the SSM inversion accuracy was clearly improved once the samples were enhanced for all these three machine learning models, and the RF model was the best prediction model among them.

3.2. Optimal Model Construction

3.2.1. Feature Selection

Sentinel-1 and Sentinel-2 data were preprocessed, and a total of 21 feature parameters were extracted. In order to improve the model performance, Pearson correlation analysis, RF and PCA methods were used to reduce the redundant features and improve the accuracy of model estimation.

There are differences in the selection criteria of these three methods in feature selection. Pearson correlation analysis method measures the linear correlation between the data, and the higher correlation indicates that it is more sensitive to SSM inversion. The ranking of feature importance is shown in Table 4. RF is measured according to the average contribution of each feature on each tree. Higher contribution indicates that it is more sensitive to SSM inversion, and the ranking of feature importance is shown in Table 5. PCA shows the characteristics of data in a smaller dimension by dimensionality reduction, and finally the new variable is the linear combination of the original variables. The correlation results of different features are shown in Figure 3.

Table 4. Ranking of feature importance using the Pearson correlation analysis method.

No.	Parameter	Correlation Coefficient	No.	Parameter	Correlation Coefficient
1	cos(θ)	−0.3942	12	RVI	0.1286
2	θ	−0.3837	13	$\sigma^0_{VH} - \sigma^0_{VV}$	−0.1039
3	sin(θ)	0.3832	14	H	0.1027
4	$\sigma^0_{VH}/\sigma^0_{VV}$	−0.3608	15	$\sigma^0_{VH} \times \sigma^0_{VV}$	0.0976
5	FVI	0.3180	16	$\sigma^0_{VH} + \sigma^0_{VV}$	−0.0853
6	NDVI	0.2374	17	λ1	−0.0701
7	α	0.2004	18	λ2	−0.0654
8	NDWI	0.1915	19	σ^0_{VH}	−0.0471
9	MSI	−0.1743	20	A	−0.0438
10	Zs	0.1556	21	WBI	0.0388
11	σ^0_{VV}	−0.1462			

Table 5. Ranking of feature importance using RF method.

No.	Parameter	Correlation Coefficient	No.	Parameter	Correlation Coefficient
1	cos(θ)	0.522	12	σ^0_{VH}	0.01696
2	sin(θ)	0.4129	13	FVI	−0.01457
3	α	0.3683	14	λ_1	0.1441
4	NDVI	0.3457	15	H	0.1364
5	θ	0.2896	16	$\sigma^0_{VV} \times \sigma^0_{VH}$	−0.0929
6	$\sigma^0_{VH} - \sigma^0_{VV}$	0.2601	17	λ_2	0.0896
7	MSI	−0.2418	18	$\sigma^0_{VV} + \sigma^0_{VH}$	0.0644
8	σ^0_{VV}	−0.2447	19	$\sigma^0_{VH}/\sigma^0_{VV}$	−0.0624
9	Zs	0.2035	20	NDWI	0.0541
10	RVI	0.1965	21	WBI	−0.0522
11	A	0.1703			

Different feature selection methods were applied to the subsequent three machine learning models respectively. For the verification of different input parameters, the comparison and analysis showed that the results obtained by selecting the first eight features as ideal feature subsets were generally accurate, so the following experiments all selected the first eight parameters to ensure the homogeneity of the experiments.

3.2.2. Machine Learning Model

To verify the efficiency of the suggested method, a comparative experiment was carried out utilizing measured SSM data after sample augmentation, and the application performance of various feature selection methods and machine learning models in SSM inversion was reviewed.

In this experiment, R^2, RMSE and MAE were employed to assess SSM inversion accuracy. The average values acquired after repeat experiments were recorded as the experimental results to lessen the randomness of the outcomes, as indicated in Table 6.

According to the experimental results, the combination of employing RF for feature selection and RF for SSM inversion offered the maximum inversion accuracy, with R^2, RMSE and MAE of 0.7256, 0.0539 cm3/cm^3 and 0.0422 cm^3/cm^3, respectively.

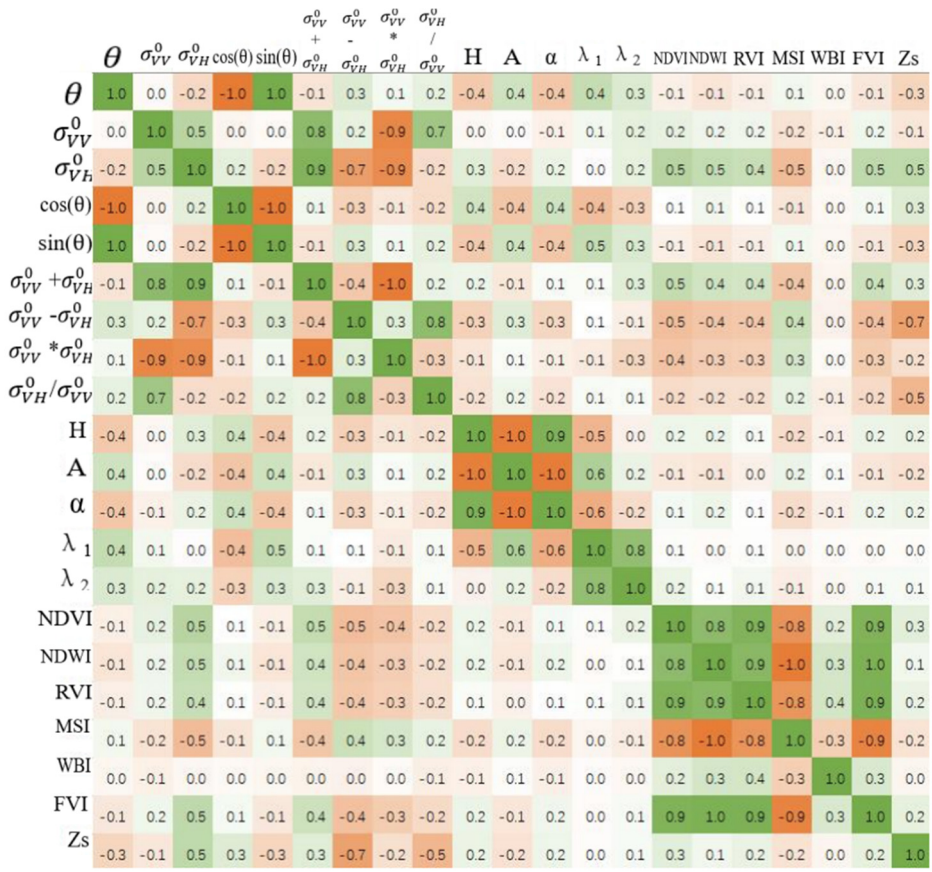


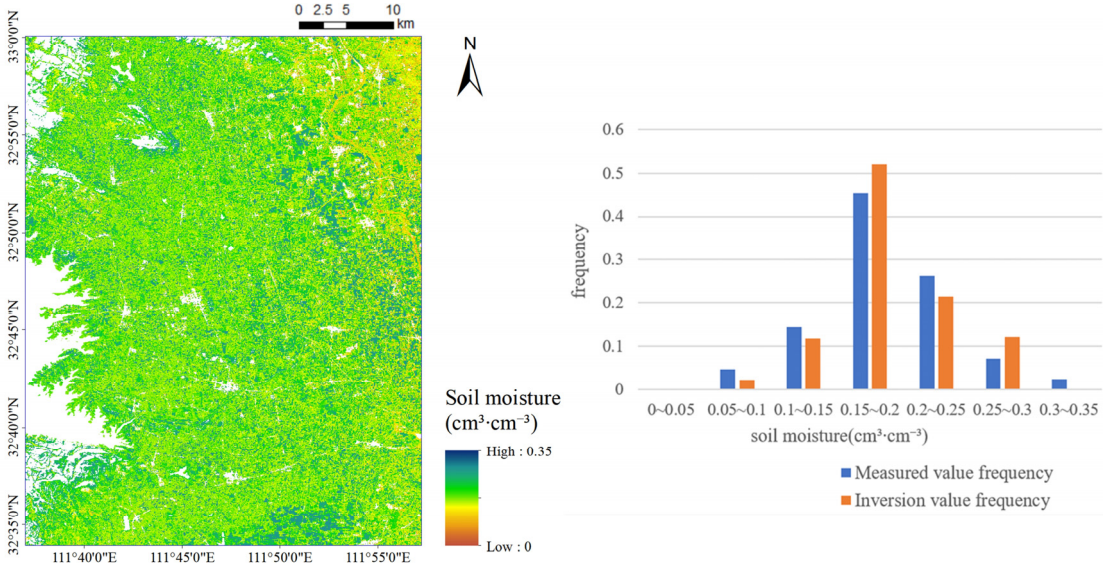
Figure 3. Correlation between features obtained by PCA method.

Table 6. Comparison of the accuracy of the inversion results under different model combinations.

Feature Selection Method	Machine Learning Method	R^2	RSME (cm^3/cm^3)	MAE (cm^3/cm^3)
Pearson correlation analysis	GA-BP	0.6656	0.0627	0.0519
Pearson correlation analysis	SVR	0.5223	0.0608	0.0466
Pearson correlation analysis	RF	0.5750	0.0567	0.0448
RF	GA-BP	0.6420	0.0776	0.0658
RF	SVR	0.5910	0.0571	0.0455
RF	RF	0.7324	0.0534	0.0413
PCA	GA-BP	0.5943	0.0770	0.0586
PCA	SVR	0.4969	0.0786	0.0622
PCA	RF	0.5824	0.0574	0.0450

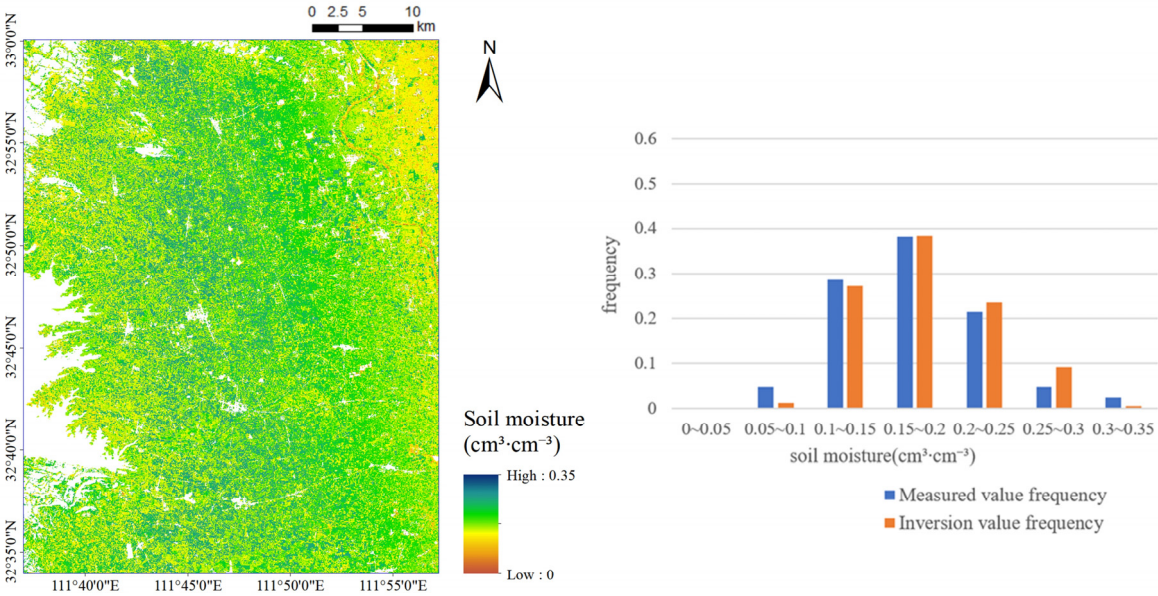
3.3. Spatial Distribution of SSM

Based on Sentinel-1 SAR data and Sentinel-2 optical data, RF model was used for feature selection and SSM inversion, and the spatial and frequency distributions of regional SSM in the study area were obtained, as shown in Figures 4–6. In the study area, the average values of measured SSM on three dates after sample augmentation were 0.1892, 0.1861 and 0.1808 cm^3/cm^3, respectively. The average values of retrieved SSM on corresponding dates were 0.1876, 0.1833 and 0.1783 cm^3/cm^3, respectively, which were consistent with the field measured SSM in general.



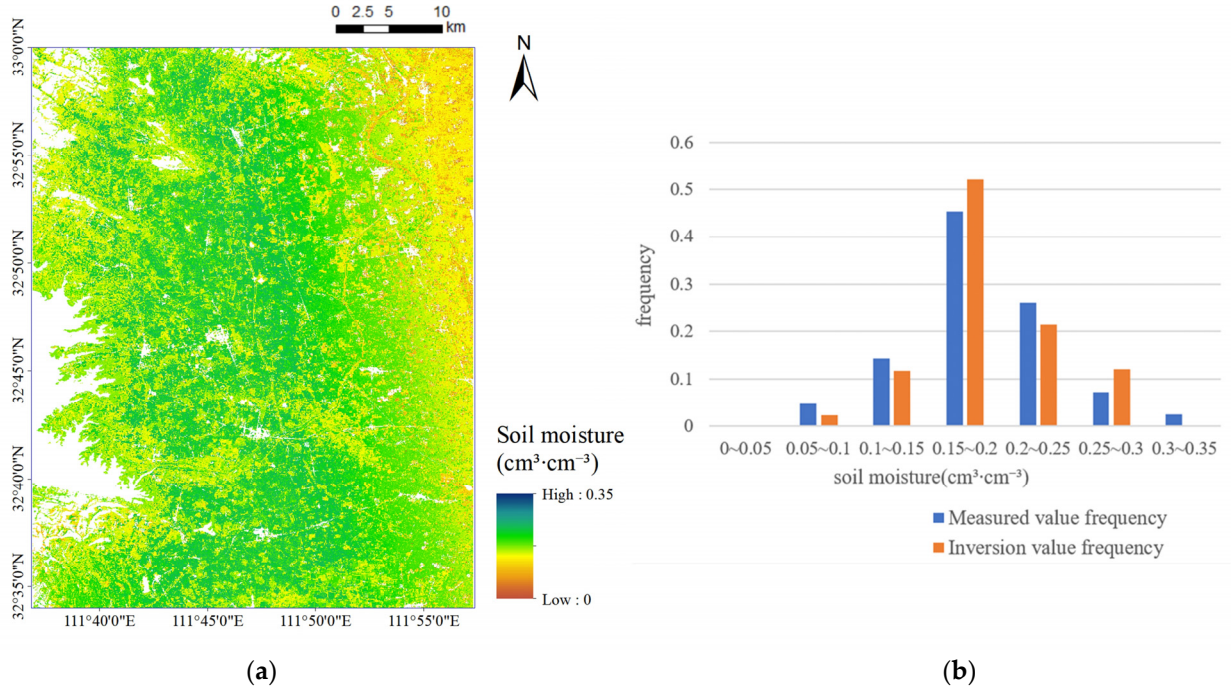
(**a**) (**b**)

Figure 4. Inversion results of regional SSM in the study area on 11 September 2021: (**a**) spatial distribution of retrieved SSM; (**b**) frequency distribution of retrieved and measured SSM.



(**a**) (**b**)

Figure 5. Inversion results of regional SSM in the study area on 23 September 2021: (**a**) spatial distribution of retrieved SSM; (**b**) frequency distribution of retrieved and measured SSM.



(**a**) (**b**)

Figure 6. Inversion results of regional SSM in the study area on 5 October 2021: (**a**) spatial distribution of retrieved SSM; (**b**) frequency distribution of retrieved and measured SSM.

4. Discussion

4.1. Data Augmentation

Field measurements are necessary for soil moisture inversion. In practice, there are two main methods of field measuring. One is the traditional manual measuring method based on manual ground sampling and measurement on the date of satellite transits [1,4,6,8,12–20,22,25,26,31,33,36,38]. The other is the automatic measuring method based on ground-based observation stations or networks in the study area [2,3,7,21,23,24,29]. Compared with the automatic measuring method, the field measured SSM data obtained through the traditional manual measuring method are often more difficult to collect, and usually in a limited number of times and in small quantities.

For those areas without any ground-based observation sites or automatic observation networks, like the study area in this paper, due to the limitations of time and space, the data obtained by manual field measurement are generally limited. The small size of field measured SSM data have a negative effect on SSM inversion accuracy, since there is insufficient data to train the inversion model and make a meaningful evaluation on the inversion results.

The experimental results shown in Tables 3 and 6 demonstrated that, the proposed inversion method based on data augmentation was effective to supply more sample data for SSM inversion and further improved the inversion accuracy, providing a feasible reference for SSM inversion studies based on small sample size of field measured data.

It is worth noting that, the alpha approximation method used in this paper for data augmentation has a certain precondition, which assumed that the vegetation conditions and surface roughness remain unchanged in the short spanned period. So, the proposed method in this paper is not suitable for SSM inversion in those areas with large changes in the vegetation conditions and surface roughness in the study period. In fact, even in a short period, this precondition is difficult to strictly meet. In this study, although the dates of 11 September 2021 and 5 October 2021 were close to the middle date of 23 September 2021, and the vegetation conditions and surface roughness kept constant on the whole according to the field survey and actual situation, small changes in some parts of the study area were ineluctable. This fact affected the application of the alpha approximation method and

further the SSM inversion accuracy of the proposed method. In the future, more reliable and effective data augmentation methods will be explored to expand the sample size, thus to further improving the SSM inversion accuracy.

4.2. Accuracy Analysis

After data augmentation, the parameters extracted from remote sensing data and machine learning models for SSM inversion were optimized to improve the SSM inversion accuracy further.

Three parameter optimization methods, which were Pearson correlation analysis, RF and PCA, were proven to be effective for parameter optimization in SSM inversion [22,34,38], and so chosen in this study to reduce the redundant features and improve the accuracy of model estimation. The experimental results shown in Tables 4 and 5 and Figure 3 indicated that, it was hard to get a uniform optimal feature subset through these three different optimization methods, due to their different optimization principles and evaluation criteria. Inspired by the research in reference [22], the extracted parameters and the used inversion models were optimized in the whole in this study by using different combinations of parameter optimization methods and machine learning models. In order to ensure the homogeneity of the experiments, the first eight features in each ranking result of the feature optimization method were uniformly selected as the ideal feature subset for the subsequent experiments. However, it was not ensured that the subsets with the first eight features for all these three methods could all reach the final maximum inversion accuracy for all these nine model combinations. Different sizes of optimal feature subsets for different feature selection methods may be more reasonable for the proposed method and will be explored in this study to further improve the SSM inversion accuracy.

Three typical machine learning models commonly used in SSM inversion, which were GA-BP [36,38], SVR [15,18,19,26,33], and RF [16,23,33,34,38], were selected in this study because of their good performance in SSM inversion based on the small size of samples [15,16,19,33,36,38]. Combined with three parameter optimization methods, the performance of nine different model combinations in SSM inversion was compared in the experiments. According to the experimental results shown in Table 6, the combination of employing RF for feature selection and RF for SSM inversion offered the maximum inversion accuracy, with higher R^2 and lower RSME and MAE than other combinations.

The performance of GA-BP and SVR models was a little worse than that of the RF model in this study, although they are generally considered to have good generalization ability when the sample size is small. One possible reason is that some parameters of GA-BP and SVR models may be not set properly and could be further optimized. Another possible reason is that there was an over-fitting issue in their training process because the sample size in this study was too small, which also affected the performance of the RF model in this study.

A total of 41 measured SSM samples were obtained in the field survey, and expanded to 93 samples after data augmentation. Compared with the original sample data and previous SSM inversion studies based on small samples [1,4,6,15,16,18,19,33,36,38], the sample size had increased. However, due to the limited initial sample size and the limitations of the alpha approximation method, the sample size was still limited, which was prone to over-fitting issues in practice. A considerable sample set is still the guarantee for sufficient model training and reasonable inversion accuracy. More field measurements are planned in this study in the future.

Even though, for those SSM inversion studies based on a small sample size of field measured data, as demonstrated in this study, the data augmentation method was still an effective way to supply more sample data and further improved the inversion accuracy.

5. Conclusions

An SSM inversion method combining sample augmentation, feature optimization, and machine learning models was investigated in this paper. First, sample augmentation

was applied to the field-measured SSM data to address the issue that inversion accuracy was impacted by the sample size. Sentinel-1 SAR data and Sentinel-2 optical data were integrated to extract and select several feature parameters. The optimal inversion model was chosen by combining various feature selection methods with machine learning models at the same time to achieve optimal inversion accuracy. The experimental results indicated that the inversion accuracy had improved after sample augmentation, and the combination of employing RF for feature selection and RF for SSM inversion offered the maximum inversion accuracy, with R^2, RMSE and MAE of 0.7256, 0.0539 cm^3/cm^3 and 0.0422 cm^3/cm^3, respectively. The proposed method was finally used to invert the regional SSM of the study area. The inversion results indicated that the proposed method had good performance in regional applications with a small sample size, and provided a feasible way for SSM inversion in those areas where the vegetation conditions and surface roughness remain unchanged within a certain time span. In the future, more effective methods of data augmentation and machine learning can be explored to further improve the SSM inversion accuracy.

Author Contributions: Methodology, Y.W., Z.G., J.Z., H.Y. and N.L.; investigation, J.Z. and Y.W.; experiment and visualization, Y.W.; validation, Z.G., J.Z., H.Y. and N.L.; writing-original draft, Y.W. and J.Z.; writing-review and editing, Y.W., Z.G., J.Z., H.Y. and N.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (42101386, 61871175), the Plan of Science and Technology of Henan Province (222102110439, 212102210093), the College Key Research Project of Henan Province (22A520021), the Plan of Science and Technology of Kaifeng City (2102005), the Key R&D Project of Science and Technology of Kaifeng City (22ZDYF006), and the Key Laboratory of Natural Resources Monitoring and Regulation in Southern Hilly Region, Ministry of Natural Resources of the People's Republic of China (NRMSSHR2022Z01).

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the European Space Agency (ESA) for providing the Sentinel-1 and Sentinel-2 data, and the anonymous reviewers and editors for their valuable comments, that are crucial in improving the quality of this paper. The authors would also like to thank all the teachers and students of the SAR information processing team of Henan University for helping this study.

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

References

1. Wang, H.; Magagi, R.; Goita, K. Potential of a two-component polarimetric decomposition at C-band for soil moisture retrieval over agricultural fields. *Remote Sens. Environ.* **2018**, *217*, 38–51. [CrossRef]
2. Gill, M.K.; Asefa, T.; Kemblowski, M.W.; McKee, M. Soil moisture prediction using support vector machines. *J. Am. Water Resour. Assoc.* **2006**, *42*, 1033–1046. [CrossRef]
3. Ge, L.; Hang, R.; Liu, Y.; Liu, Q. Comparing the Performance of Neural Network and Deep Convolutional Neural Network in Estimating Soil Moisture from Satellite Observations. *Remote Sens.* **2018**, *10*, 1327. [CrossRef]
4. Zhang, X.; Chen, B.; Fan, H.; Huang, J.; Zhao, H. The Potential Use of Multi-Band SAR Data for Soil Moisture Retrieval over Bare Agricultural Areas: Hebei, China. *Remote Sens.* **2016**, *8*, 7. [CrossRef]
5. Zhang, W.F.; Chen, E.X.; Li, Z.Y.; Yang, H.; Zhao, L. Review of applications of radar remote sensing in agriculture. *J. Radars* **2020**, *9*, 444–461.
6. Zhang, X.; Tang, X.; Gao, X. Soil Moisture Retrieval Over Early Corn Covered Area Using Radarsat-2 and TerraSAR-X Data. In Proceedings of the 2019 6th APSAR, Xiamen, China, 26–29 November 2019.
7. Pierdicca, N.; Pulvirenti, L.; Bignami, C. Soil moisture estimation over vegetated terrains using multitemporal remote sensing data. *Remote Sens. Environ.* **2010**, *114*, 440–448. [CrossRef]
8. Bhogapurapu, N.; Dey, S.; Bhattacharya, A.; Rao, Y.S. Soil moisture estimation using Simulated NISAR Dual Polarimetric GRD Product over croplands. In Proceedings of the 2021 7th APSAR, Bali, Indonesia, 1–3 November 2021.
9. Fu, Z.; Zhang, H.; Zhao, J.; Li, N.; Zheng, F. A Modified 2-D Notch Filter Based on Image Segmentation for RFI Mitigation in Synthetic Aperture Radar. *Remote Sens.* **2023**, *15*, 846. [CrossRef]
10. Fung, A.K.; Li, Z.; Chen, K.S. Backscattering from a randomly rough dielectric surface. *Remote Sens.* **1992**, *30*, 356–369. [CrossRef]
11. Oh, Y.; Sarabandi, K.; Ulaby, F.T. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 370–381. [CrossRef]

12. Dubois, P.C.; van Zyl, J.; Engman, T. Measuring soil moisture with imaging radars. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 915–926. [CrossRef]
13. Bao, Y.; Lin, L.; Wu, S.; Deng, K.A.W.; Petropoulos, G.P. 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.* **2018**, *72*, 76–85. [CrossRef]
14. Gao, Q.; Zribi, M.; Escorihuela, M.J.; Baghdadi, N. Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. *Sensors* **2017**, *17*, 1966. [CrossRef]
15. Guo, J.; Liu, J.; Ning, J.; Han, W.T. Construction and validation of farmland surface soil moisture retrieval model based on sentinel multi-source data. *Trans. CSAE* **2019**, *35*, 71–78.
16. Datta, S.; Das, P.; Dutta, D.; Giri, R.K. Estimation of Surface Moisture Content using Sentinel-1 C-band SAR Data Through Machine Learning Models. *J. Remote Sens.* **2021**, *49*, 887–896. [CrossRef]
17. Notarnicola, C.; Angiulli, M.; Posa, F. Soil moisture retrieval from remotely sensed data: Neural network approach Versus Bayesian method. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 547–557. [CrossRef]
18. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084. [CrossRef]
19. Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Chiesa, S.D.; Niedrist, G.; Bruzzone, L.; Tappeiner, U.; Zebisch, M. Soil moisture monitoring in mountain areas by using high-resolution SAR images: Results from a feasibility study. *Eur. J. Soil Sci.* **2014**, *65*, 852–864. [CrossRef]
20. Said, S.; Kothyari, U.C.; Arora, M.K. ANN-based soil moisture retrieval over bare and vegetated areas using ERS-2 SAR data. *J. Hydrol. Eng.* **2008**, *13*, 461–475. [CrossRef]
21. Cai, Y.; Zheng, W.G.; Zhang, X.; Zhangzhong, L.L.; Xue, X.Z. Research on soil moisture prediction model based on deep learning. *PLoS ONE* **2019**, *14*, e0214508. [CrossRef] [PubMed]
22. Lin, C.; Xing, M.F.; He, B.B.; Wang, J.F.; Shang, L.J.; Huang, M.D. Estimating Soil Moisture Over Winter Wheat Fields During Growing Season Using Machine-Learning Methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3706–3718.
23. Zhang, L.; Zhang, Z.; Xue, Z.; Li, H. Sensitive Feature Evaluation for Soil Moisture Retrieval Based on Multi-Source Remote Sensing Data with Few In-Situ Measurements: A Case Study of the Continental U.S. *Water* **2021**, *13*, 2003. [CrossRef]
24. Balenzano, A.; Mattia, F.; Satalino, G.; Davidson, M.W.J. Dense Temporal Series of C- and L-band SAR Data for Soil Moisture Retrieval Over Agricultural Crops. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 439–450. [CrossRef]
25. He, L.; Qin, Q.; Panciera, P.; Tanase, M.; Walker, J.P.; Hong, Y. An Extension of the Alpha Approximation Method for Soil Moisture Estimation Using Time-Series SAR Data Over Bare Soil Surfaces. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1328–1332. [CrossRef]
26. Xu, W.; Zhang, Z.; Qin, Q.; Hui, J.; Long, Z. Soil Moisture Estimation with SVR and Data Augmentation Based on Alpha Approximation Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3190–3201. [CrossRef]
27. Chen, S.S. Evaluation of Ecological Service Function of Water Retention and Soil Conservation in Water Source Area for the South-to-North Water Transfer-A Case Study in Shangluo City. Master's Thesis, Northwest University, Xi'an, China, 2016.
28. Wang, C.; Zhang, H.; Chen, X. *Quad Polarization Synthetic Aperture Radar Image Processing*; Science Press: Beijing, China, 2008.
29. Lin, L.B. Soil Moisture Retrieval under Vegetation Cover Using Multi-Source Remote Sensing Data. Master's Thesis, Nanjing University of Information Science and Technology, Nanjing, China, 2018.
30. Cloude, S.R.; Pottier, E. A review of target decomposition theorems in radar polarimetry. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 498–518. [CrossRef]
31. Zhao, J.H.; Zhang, B.; Li, N.; Guo, Z.W. Cooperative Inversion of Winter Wheat Covered Surface Soil Moisture Based on Sentinel-1/2 Remote Sensing Data. *J. Electron. Inf.* **2021**, *43*, 692–699.
32. Tong, L.; Chen, Y.; Jia, M.Q. *Mechanism of Radar Remote Sensing*; Science Press: Beijing, China, 2014.
33. Hoa, P.V.; Giang, N.V.; Binh, N.A.; Hai, L.V.H.; Pham, T.-D.; Hasanlou, M.; Tien Bui, D. Soil Salinity Mapping Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms: A Case Study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sens.* **2019**, *11*, 128. [CrossRef]
34. Alhowaide, A.; Alsmadi, I.; Tang, J. PCA, Random-Forest and Pearson Correlation for Dimensionality Reduction in IoT IDS. In Proceedings of the 2020 IEMTRONICS, Vancouver, BC, Canada, 9–12 September 2020.
35. Narasimhan, S.; Shah, S.L. Model identification and error covariance matrix estimation from noisy data using PCA. *Control Eng. Pract.* **2008**, *16*, 146–155. [CrossRef]
36. Yu, F.; Zhao, Y.S.; Li, H.T. Soil moisture retrieval based on GA-BP neural networks algorithm. *J. Infrared Millim. Waves* **2012**, *31*, 283–288. [CrossRef]
37. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sens.* **2015**, *7*, 16398–16421. [CrossRef]
38. Zhao, J.; Zhang, C.; Min, L.; Guo, Z.; Li, N. Retrieval of Farmland Surface Soil Moisture Based on Feature Optimization and Machine Learning. *Remote Sens.* **2022**, *14*, 5102. [CrossRef]