Retrieving Surface Soil Moisture over Wheat-Covered Areas Using Data from Sentinel-1 and Sentinel-2

Yan Li [1,2], **Chengcai Zhang** [1,*] **and Weidong Heng** [1]

1　School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China; yanli@gs.zzu.edu.cn (Y.L.); hengweidong@gs.zzu.edu.cn (W.H.)
2　Zhengzhou Vocational College of Industrial Safety, Zhengzhou 451192, China
*　Correspondence: zhangcc@zzu.edu.cn; Tel.: +86-135-2348-0466

Abstract: Surface soil moisture (SSM) is a major factor that affects crop growth. Combined microwave and optical data have been widely used to improve the accuracy of SSM retrievals. However, the influence of vegetation indices derived from the red-edge spectral bands of multi-spectral optical data on retrieval accuracy has not been sufficiently analyzed. In this study, we retrieved soil moisture from wheat-covered surfaces using Sentinel-1/2 data. First, a modified water cloud model (WCM) was proposed to remove the influence of vegetation from the backscattering coefficient of the radar data. The vegetation fraction (FV) was then introduced in this WCM, and the vegetation water content (VWC) was calculated using a multiple linear regression model. Subsequently, the support vector regression technique was used to retrieve the SSM. This approach was validated using in situ measurements of wheat fields in Hebi, located in northern Henan Province, China. The key findings of this study are: (1) Based on vegetation indices obtained from Sentinel-2 data, the proposed VWC estimation model effectively eliminated the influence of vegetation; (2) Compared with vertical transmit and horizontal receive (VH) polarization, vertical transmit and vertical receive (VV) polarization was better for detecting changes in SSM key phenological phases of wheat; (3) The validated model indicates that the proposed approach successfully retrieved SSM in the study area using Sentinel-1 and Sentinel-2 data.

Keywords: surface soil moisture; sentinel-1 SAR; Sentinel-2; vegetation water content; water cloud model; support vector regression

1. Introduction

Surface soil moisture (SSM) is a key variable that couples the land and the atmosphere, as well as energy and water cycles. SSM, therefore, plays an essential role in hydrology, climatology, meteorology, ecology, and agronomy [1–3]. SSM is particularly important in arid and semi-arid agricultural regions, where its spatiotemporal distribution affects crop growth and development [4,5]. Despite its importance, it is difficult to accurately retrieve SSM over large scales due to the complexity of natural surfaces [6,7].

Retrieving SSM using remote sensing technology has been investigated for more than 30 years. Among the remote sensing methods, most optical methods estimate SSM by using the spectral reflectance indices, which are easy to implement but can be easily affected by weather [8]. In thermal infrared methods, SSM is mainly estimated from thermal inertia [9]. It is important to note that the vegetation canopy can conceal soil radiation information, thereby affecting the accuracy of SSM retrievals in areas with high vegetation coverage [10]. Microwaves have stronger penetrating ability and are not influenced by clouds due to their longer wavelengths.

Previous quantitative studies showed that microwave remote sensing methods could effectively estimate SSM. Passive sensors measure the intensity of naturally emitted microwaves from the Earth's surface. However, the atmospheric effect is not negligible

frequencies above 6 GHz, which limits the advantages of microwave remote sensing measurements in all-weather conditions [11,12]. As active instruments, synthetic aperture radar (SAR) sensors supply their own source of illumination and subsequently determine the energy difference between the transmitted and received electromagnetic radiation [13,14]. They can make SSM observations over large areas in all weather conditions. Further, they can operate during the day and night. However, it is inevitable that vegetation will microwave SSM retrieval [15]. Therefore, multi-sensor synergistic methods for retrieving SSM have attracted more attention, of which SAR and optical data synergy is one of the most widely used approaches [16–18].

The Sentinel-1 mission is the European radar observatory for the Copernicus joint initiative of the European Commission (EC) and the European space agency (ESA). The Sentinel-1A and Sentinel-1B satellites include C-band imaging operating in four exclusive imaging modes and operate in a constellation, collect free high spatial (5 m × 5 m) and temporal resolution (6 d) data.

Subsequently, the Sentinel-2A and Sentinel-2B satellites were launched, carrying a multi-spectral instrument (MSI) that also collects high spatial (10 m) and temporal resolution (5 d) data. Based on the synergy between Sentinel-1 SAR and Sentinel-2 data, Attarzadeh et al. [19] estimated the SSM in vegetated areas of Kenya using the support vector regression (SVR) technique. Their results showed that the coefficient of determination (R^2) between the measured and estimated SSM values was 0.89. Bousbih et al. [20] retrieved SSM in the semi-arid region of the Kairouan plain using VV polarization backscattering coefficients recorded by Sentinel-1 and normalized difference vegetation index (NDVI) data from Sentinel-2 images. They compared the SSM values using the water cloud model (WCM) and a neural network (NN) and found that both of the measurements were accurate, with a root mean square error (RMSE) of less than 5%. Wang et al. [21] used Sentinel-1 SAR data to estimate SSM in wheat farmlands based on the WCM and the advanced integral equation model (AIEM). To evaluate the accuracy of the different optical datasets used in SSM retrieval, they compared seven vegetation indices based on Landsat-8, Sentinel-2, and GF-1 data and concluded that the Sentinel-2 data achieved higher accuracy than the data from the other sources.

Previous studies mostly focused on vegetation parameters derived from Sentinel-2 data for the purpose of estimating the impact of vegetation. Sentinel-2 data have three red-edge spectral bands centered at 705, 740, and 783 nm, which are closely related to important biochemical parameters of green plants [22,23]. However, few researchers have investigated using these red-edge spectral data to retrieve SSM.

Considering this, the main aim of this study was to evaluate the synergistic use of Sentinel-1 and Sentinel-2 data to retrieve SSM over wheat-covered areas, with a focus on developing a method for eliminating vegetation influence from the radar backscattering. To accomplish this, we developed and utilized the WCM. The vegetation spectral indices obtained from Sentinel-2 data were used to express the VWC, which is an important parameter in the WCM. Moreover, the SVR technique [24], which has good robustness for the limited availability of reference results in various application domains, was applied to retrieve SSM.

2. Study Area and Dataset

2.1. Study Area

The study area was Hebi (35°26′–36°02′ N, 113°59′–114°45′ E), which is located in northern Henan Province, China (Figure 1). The agricultural fields analyzed in this study area are mainly located on the Huang-Huai-Hai Plain, which has a homogeneous soil texture. Hebi has a sub-humid warm temperate continental monsoon climate. The average rainfall ranges from 349.2 to 970.1 mm, with peaks between May and August. Dry spells occur between November and March, so the seasonal climatic characteristics are favorable for crop farming. The cropping systems in this area usually include wheat, corn, cotton, and canola, and wheat cycles occur between late September (emergence) and the middle of

June (harvest) of the following year. A total of 28 samplings in this region were selected on slopes between 0% and 5% (Figure 1c). According to the statistic of Hebi in the 2019 National economic and social development communique, the wheat fields in the study area covered approximately 897.4 km^2 in 2019.
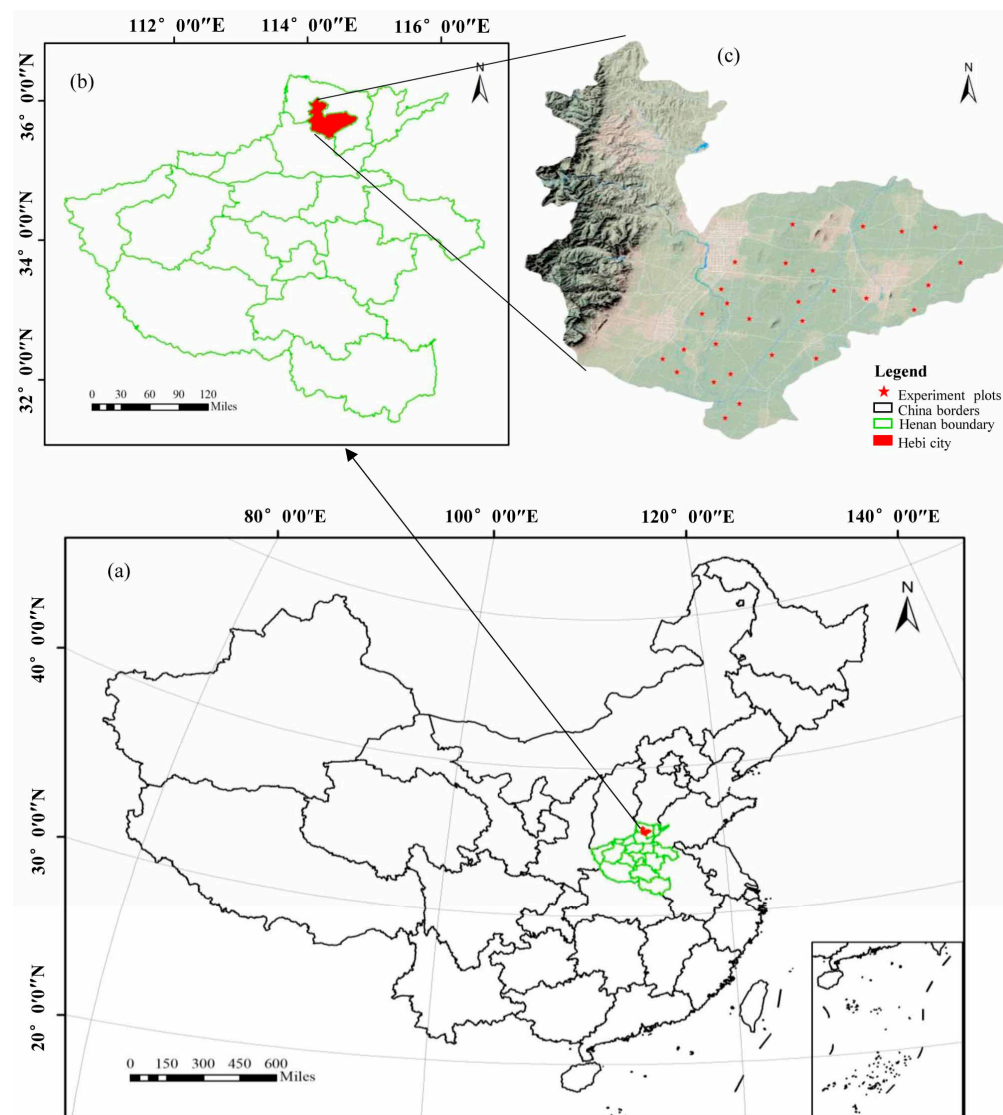

Figure 1. Overview of the study area and map showing the locations of the sampling sites.

2.2. Remote Sensing Data

2.2.1. Sentinel-1 Data

Although these two satellites pass over the study area at least 2–3 times per month and considering that the sensitivity of wheat differs at different growth stages, this study only analyzed the SSM when wheat was in key phenological stages (jointing, heading, and filling stages). Therefore, Sentinel-1 images from 9 April, 3 May, and 27 May 2019 were used. The acquired SAR images were single look complex (SLC) and level-1B data, which need to be pre-processed prior to retrieving SSM. The Sentinel-1pre-processing used for the Sentinel-1 images was based on a common software platform, namely the Sentinel application platform (SNAP). The specific processes were conducted as follows:

(1) The SAR images were corrected using radiometric calibration so that their pixel values truly represented the radar backscatter of the reflecting surface. Radiometric calibration was conducted based on the following expression:

$$\sigma_{i,j}^0 = 10log_{10}\left(\frac{|DN_{ij}|^2}{A_{ij}^2}\right) \tag{1}$$

where σ^0 is the backscattering coefficient (dB); i and j represent the i-th row and the j-th column, respectively; (DN_{ij} is the digital number of the SAR image; and A_{ij} is the calibration parameter;

(2) Image mosaics were constructed, and geometric corrections were made;

(3) The refined Lee filter method was applied to the multi-look images for speckle noise removal [25]. In this method, the weights of neighboring pixels were first estimated using kernel density, after which the value of the central pixel was calculated using linear weighting. Following this process, the speckle was effectively eliminated while preserving the edge information of the image. Figure 2a shows the pre-processing results of Sentinel-1 SAR data from the study area for 9 April 2019.
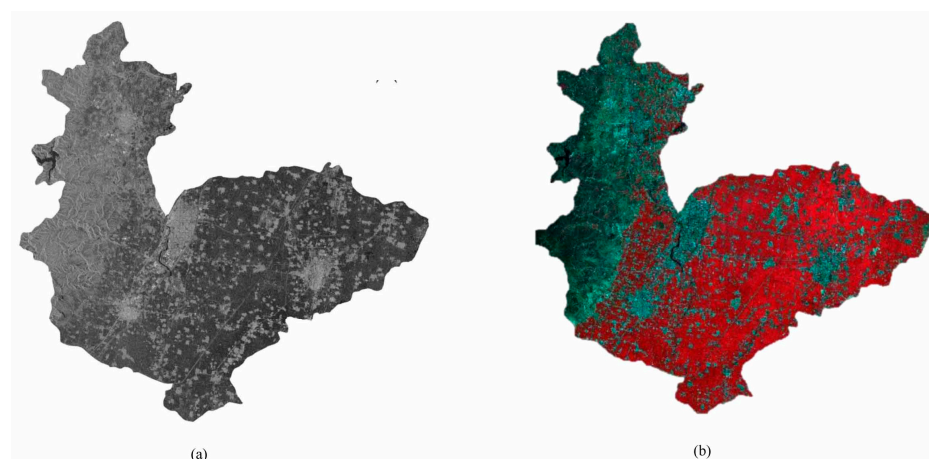

(a) (b)

Figure 2. Remote sensing images of the study area. (**a**) Sentinel-1 SAR image (VV polarization). (**b**) Sentinel-2 composite false red green blue (RGB) images (Band 8 = red, Band 4 = green, and Band 3 = blue).

2.2.2. Sentinel-2 Data

Three Sentinel-2 optical images were used over the study sites and were located as close as possible to the SAR data (depending on cloud coverage; Table 1). These acquired optical images were level-2A data, in which bottom-of-atmosphere reflectance was corrected in cartographic geometry. We resampled the data to a 10 m resolution. Figure 2b shows the Sentinel-2 optical image of the study area on 3 April 2019.

Table 1. List of satellite-based and in situ data used over the study area.

Phenology	Sentinel-1	Sentinel-2	In Situ
Jointing stage	4-9-2019	4-3-2019	4-7-2019
Heading stage	5-3-2019	5-3-2019	5-3-2019
Filling stage	5-27-2019	5-23-2019	5-27-2019

2.3. In Situ Measurements

2.3.1. SSM Measurements

Contemporaneous with the satellites acquisitions, gravimetric soil moisture (GSM) at depths of 0–10 cm were conducted during the key phenological phases of wheat (jointing,

heading, and filling stage). Data collection was divided into three periods, with 28 sampling sites selected for each period. At each sampling site, three sampling points were selected that were separated by approximately 10 m. In addition, the location of each in situ measurement was recorded using a global positioning system (GPS) device. It should be noted that, as soon as the soil samples were returned to the laboratory, they were immediately weighed with an electronic balance (accuracy of 0.1 g) and dried in an oven at 105 °C for 20–24 h, until a constant temperature was reached. GSM was obtained using the following formula:

$$\text{GSM}(\%) = \frac{W_{wet} - W_{dry}}{W_{dry}} \times 100\% \tag{2}$$

where W_{wet} and W_{dry} are the weights of soil samples collected before and after drying, respectively.

The volumetric soil moisture (VSM) was then obtained by multiplying the GSM by the soil bulk density. Therefore,

$$\text{VSM}(\%) = \text{GSM}(\%) \times \rho_b \tag{3}$$

where ρ_b is the soil bulk density.

The SSM used in this study was the VSM.

2.3.2. Vegetation Parameters

To obtain the vegetation parameters, wheat samples were collected at the same times as the soil moisture samples. The main vegetation parameters were plant height, plant density, and VWC. The height of the wheat was measured using a meter scale over a 20×20 cm area; the plant density was determined for 1 m^2 of each wheat field. The fresh weight (W_F) was measured after the wheat samples were brought back to the laboratory, and the dry weight (W_D) was obtained by drying the wheat in an oven. The specific equation for VWC is as follows:

$$\text{VWC} = (W_F - W_D) \times \rho) \tag{4}$$

where ρ is the plant density of wheat.

The details of the experimental sites are summarized in Table 1.

3. Methods

In this study, we propose a proper SSM retrieval method for wheat-covered fields based on the data described in Section 2. After a variety of tests, the decomposed scattering model and SVR were implemented for SSM retrieval in this region. A flowchart of the processing steps for SSM estimation is shown in Figure 3. The SSM retrieval method used in this study comprised three main phases. The first phase was obtaining the parameters for the model from remote sensing; i.e., the backscattering coefficient images of the study areas from Sentinel-1 SAR were obtained by pre-processing, and the backscattering coefficients for each sample were extracted according to their latitudes and longitudes. The vegetation spectral indices obtained from Sentinel-2 underwent a similar process. The second phase was the central part of the processing chain. To eliminate the influence of vegetation on radar backscattering, a modified WCM with the vegetation fraction was constructed. VWC, as one of the important parameters in the WCM, can be expressed using the multiple linear regression model combined with the vegetation indices obtained from the Sentinel-2 data. Finally, to achieve an efficient and robust SSM retrieval algorithm, a machine learning approach was used. In detail, an SVR technique was applied that allowed for non-linear relationships between a target variable and several input features. Further, the coefficient of determination (R^2) and root mean square error (RMSE) were calculated to evaluate the accuracy of the SSM estimates, and the SSM values were mapped throughout the study area. More details on each part of the soil moisture retrieval algorithm are given below.
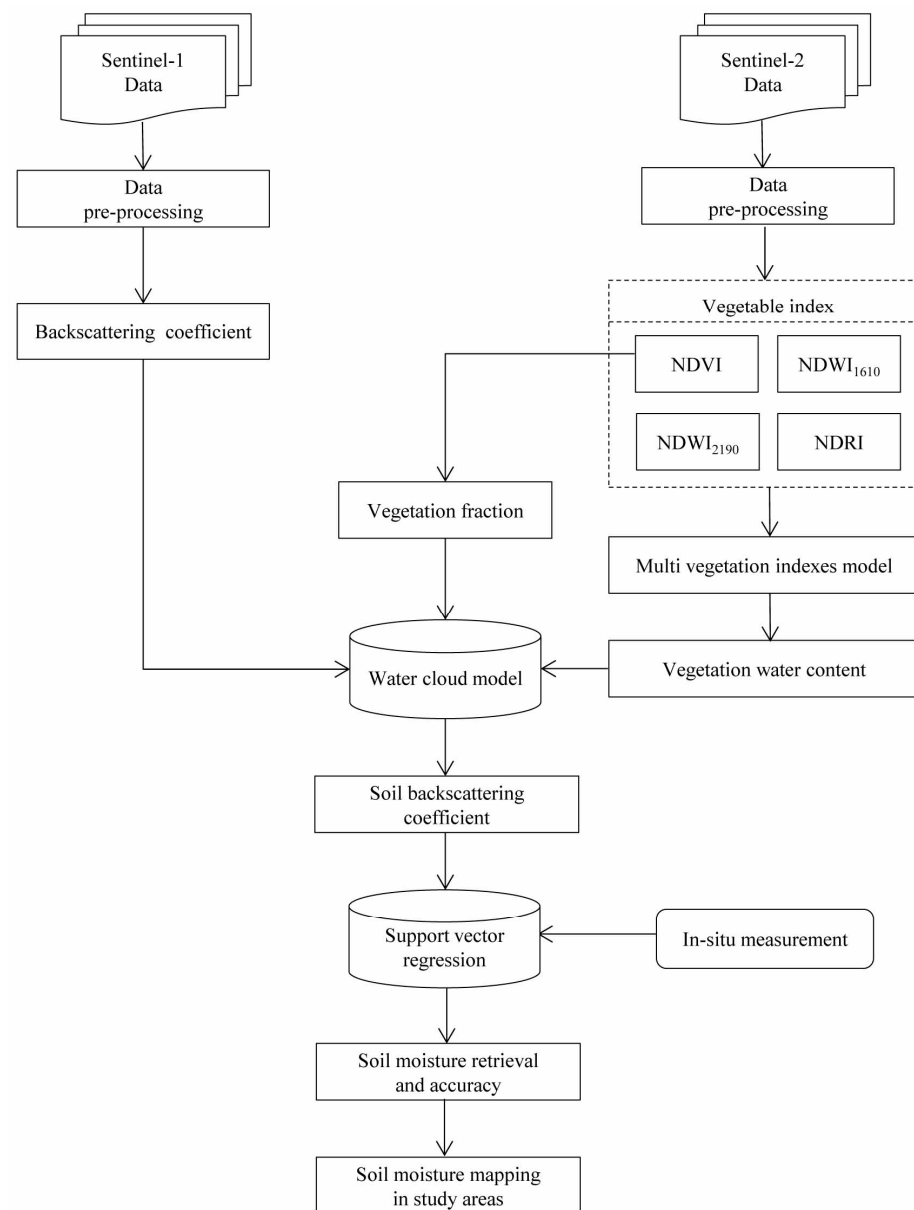
Figure 3. Flowchart of the soil moisture retrieval process based on Sentinel-1 and Sentinel-2 data.

3.1. Modified WCM

Vegetation canopies reduce the sensitivity of radar measurements to soil moisture, thereby affecting the accuracy of soil moisture estimations. Addressing this issue was the main aim of this study. The WCM is based on the radiation transport model and was proposed by Attema and Ulaby in 1978 [26]. According to the model, the total backscattering term (σ_{pp}^0) over the vegetated fields is simply divided into two parts: the backscatter contribution from the vegetation canopy (σ_{veg}^0) and the backscatter contribution from the soil surface (σ_{soil}^0). To better describe the backscattering of the soil and vegetation during different periods in wheat-covered areas, the vegetation fraction was introduced into the WVC. For a given incidence angle, the model is described as follows:

$$\sigma_{pp}^0 = f_v \left(\sigma_{veg}^0 + L^2 \sigma_{soil}^0 \right) + (1 - f_v)\sigma_{soil}^0 \tag{5}$$

$$\sigma_{veg}^0 = a V_1 \cos(\theta) \left(1 - L^2 \right) \tag{6}$$

$$L^2 = e^{-2bV_2/\cos(\theta)} \tag{7}$$

where θ is the incident angle, f_v is the vegetation fraction; σ^0_{pp} is the co-polarized total backscattering coefficient, σ^0_{veg} is the backscatter contribution from the vegetation canopy, σ^0_{soil} is the backscatter contribution from the soil surface, and L^2 is the double attenuation factor. V_1 and V_2 denote the vegetation descriptors. These can be VWC, NDVI, the leaf area index (LAI), or other vegetation descriptors [27–29]. In this study, V_1 and V_2 refer to VWC (kg/m^2). The empirical parameters a and b depend on the vegetation type and the incident angle. Bindlish and Barros [30] proposed values for a and b for different land cover types. As the crop in the study area was wheat, winter wheat was selected as the land cover type, with values of 0.0018 and 0.138 for a and b, respectively.

Moreover, f_v is an additional parameter in the vegetation scattering model used to distinguish the proportions of vegetation coverage and bare soil in pixels. f_v can be calculated using the mixed pixel decomposition model [31]:

$$f_v = (\mathrm{NDVI} - \mathrm{NDVI}_{min})/(\mathrm{NDVI}_{max} - \mathrm{NDVI}_{min}) \tag{8}$$

where NDVI_{min} denotes a bare soil pixel, which is theoretically close to zero, and NDVI_{max} denotes a pure vegetation pixel, which is theoretically close to one. In order to reduce the influence of weather conditions, a 0.5% confidence level was used to obtain the thresholds for NDVI_{min} and NDVI_{max}.

3.2. Building the VWC Model

Based on previously published studies, the vegetation indices can be used to estimate VWC [32]. The vegetation indices commonly used for VWC estimation include *NDVI* [33] and *NDWI* [34]. These two vegetation indices are be calculated as follows:

$$\mathrm{NDVI} = (R_{\mathrm{NIR}} - R_{\mathrm{Red}})/(R_{\mathrm{NIR}} + R_{\mathrm{Red}}) \tag{9}$$

$$\mathrm{NDWI} = (R_{\mathrm{NIR}} - R_{\mathrm{SWIR}})/(R_{\mathrm{NIR}} + R_{\mathrm{SWIR}}) \tag{10}$$

where R_{NIR} is the reflectivity in the near-infrared band, R_{Red} is the reflectivity in the red band, and R_{SWIR} is the reflectivity in the shortwave infrared band. Sentinel-2 data have three shortwave infrared bands: SWIR$_1$ (central wavelength = 1.374 μm), SWIR$_2$ (central wavelength = 1.610 μm), and SWIR$_3$ (central wavelength = 2.190 μm). As the spatial resolution of SWIR$_1$ is only 60 m, SWIR$_2$ and SWIR$_3$ were selected to calculate the vegetation indices (NDWI$_{1610}$ and NDWI$_{2190}$) in this study.

In recent years, studies have focused on the red-edge band between the red and near-infrared bands because of the abrupt changes in leaf reflectivity that occurs within this band. This focus has produced good applications for identifying surface types, calculating parameters, distinguishing vegetation growth states, and estimating vegetation leaf area indices. The normalized difference red-edge index (NDRI) can be calculated as follow:

$$\mathrm{NDRI} = \left(R_{Red-edge1} - R_{Red-edge2}\right)/\left(R_{Red-edge1} + R_{Red-edge2}\right) \tag{11}$$

where $R_{Red-edge}$ is the reflectivity in the red-edge infrared band. The Sentinel-2 data have three red-edge infrared bands, and the central wavelengths of red-edge bands 1 and 2 are located at the valley value (0.705 μm) and peak value (0.740 μm) of the red-edge band range, respectively. Therefore, the two red-edge infrared bands were used to calculate the NDRI in this study.

In this study, the model used to estimate VWC was divided into two steps. The first step established the exponential relationship between the VWC and the vegetation index. This exponential expression is as follows:

$$C = \alpha e^{\beta x} \tag{12}$$

where C is the VWC, α and β are the parameters to be solved, and x is the vegetation index (NDVI, NDWI$_{1610}$, NDWI$_{2190}$, or NDRI).

In the second step, the VWC obtained in the first step was used as the characteristic parameter of the multiple linear regression equation. Thus, the modified VWC can be expressed as

$$m_{veg} = \gamma_0 + \gamma_1 C_1 + \gamma_2 C_2 + \ldots + \gamma_k C_k + \varepsilon \tag{13}$$

where m_{veg} is the modified VWC, k is the number of vegetation indices, ε is the bias that obeys the normal distribution N(0,σ^2), and γ is a parameter. If $(y_1; x_{11}, x_{21}, \ldots, x_{k1}, \ldots, (y_n; x_{1n}, x_{2n}, \ldots, x_{kn})$ is a sample of capacity n, then:

$$\begin{cases} y_1 = \sum\limits_{i=0}^{k} \gamma_i C_{i1} + \varepsilon_1 \\ \quad \ldots \\ y_n = \sum\limits_{i=0}^{k} \gamma_i C_{in} + \varepsilon_n \end{cases} \tag{14}$$

The value of the parameter γ_i ($i = 0, 1, 2, \ldots, k$) can then be estimated.

3.3. SVR Estimation of SSM

SVR is a supervised regression technique that allows the modeling of multi-dimensional and non-linear relationships between target variables [35]. According to functional theory, as a kernel function satisfies Mercer's theorem, it can correspond to some type of inner product in the high-dimensional feature space. Owing to its accurate estimation, good intrinsic generalization ability, and its ability to deal with complex non-linear problems, the SVR technique can be applied for soil moisture estimation [36].

In this study, the SVR technique was chosen to retrieve SSM from the backscattering coefficients, and the entire procedure can be divided into two main phases: the training and estimation phases.

During the training phase, the field measurements, coupled with the features extracted from remote sensing data, were exploited to determine the underlying relationship between the input features and the output target value. According to our previous analysis, the features of the remote sensing data were estimated using the modified WCM, which reduced the ambiguity in the SAR signal due to the presence of vegetation. The relationships between the backscattering coefficient and the field measurements were implemented using MATLAB, and these reference samples were divided into two subsets, that is, 75% (63 samples) for training and 25% (21 samples) used for the quantitative assessment of the estimation performance.

Typically, the SVR technique has different configurations of the free model parameters, namely hyper-parameters that can control the learning process of the estimation method. These hyper-parameters are composed of the regularization parameter C, the tolerance to errors ε, and the kernel parameters. The Gaussian radial basis function (RBF) kernel function was chosen due to its limited computational overhead [37], and a grid search strategy was adopted to drive the selection of the best parameter configuration.

After the above learning phase (carried out off-line), the trained SVR regressor was ready for the estimation phase. During the estimation phase, independent test samples were used to quantitatively assess the estimation performance using common metrics such as the RMSE and R^2.

4. Results and Discussion

4.1. Estimating VWC from Sentinel-2 Data

A total of 84 in situ measurements were acquired, of which 68 random samples were used to build the VWC models, while the remaining 16 samples were used to validate the performance of the models. According to Equation (12), the fitting relationship between the measured VWC and the estimated VWC was based on a multiple linear regression

model under different combinations of vegetation indexes, as shown in Figure 4. The R^2 and RMSE between the estimated and measured VWC values based on the 16 validation data points are shown in Table 2.
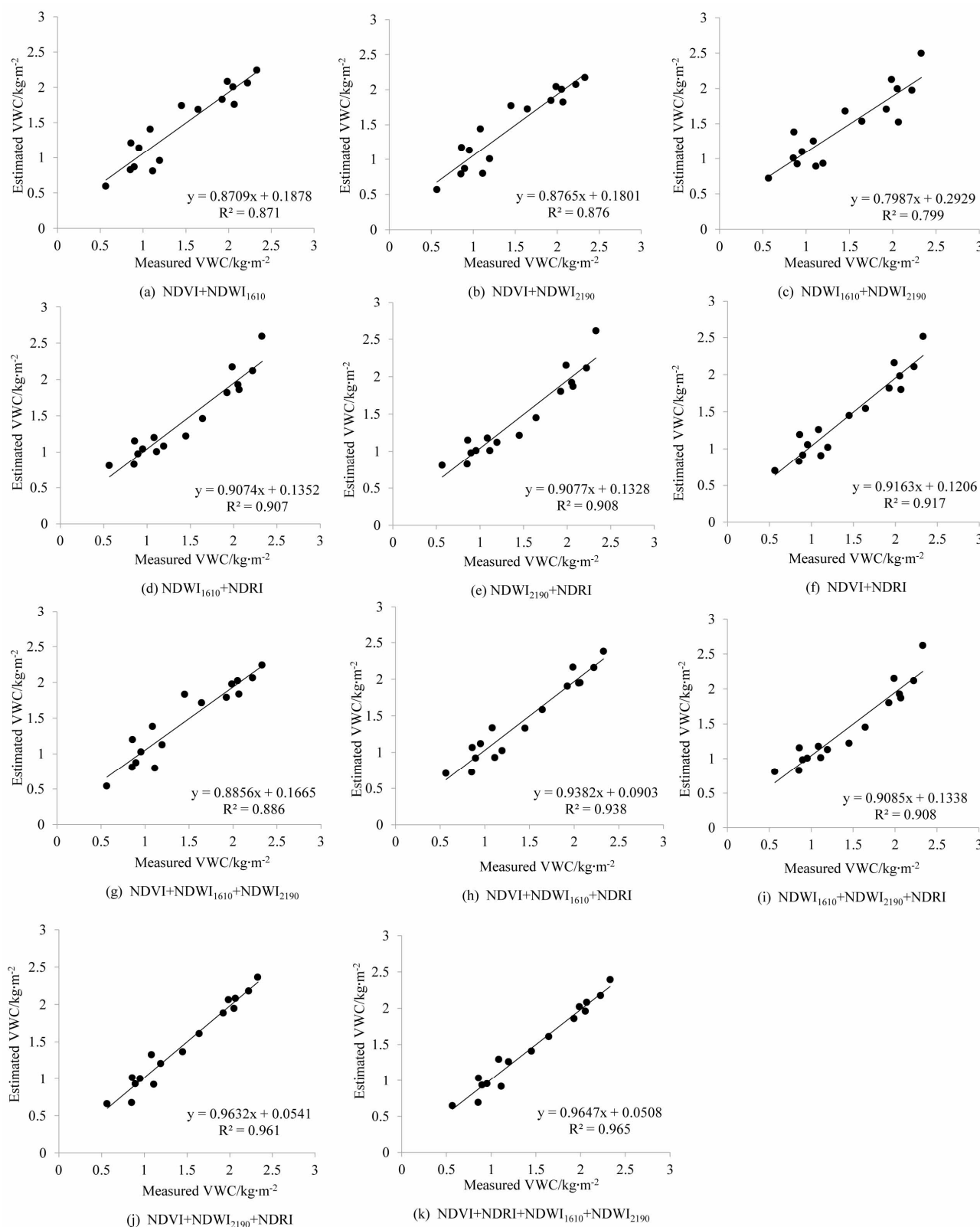

Figure 4. Scatter plots of measured and estimated VWC, using combinations of vegetation indices.

Table 2. Correlations between the modified VWC estimation models and the in situ measured VWC data.

No.	The Multiple Regression Models	R^2	RMSE/kg·m^{-2}	Note
a	$y = 0.3e^{2.538x_1} - 0.019e^{3.949x_2} - 0.058$	0.871	0.201	
b	$y = 0.384e^{2.538x_1} - 0.14e^{2.635x_3} - 0.051$	0.876	0.197	
c	$y = 0.15e^{3.949x_2} + 0.22e^{2.635x_3} + 0.04$	0.799	0.251	
d	$y = 0.376e^{3.933x_4} - 0.064e^{3.949x_2} + 0.151$	0.907	0.171	x_1 is NDVI,
e	$y = 0.391e^{3.933x_4} - 0.088e^{2.635x_3} + 0.169$	0.908	0.17	x_2 is NDWI$_{1610}$,
f	$y = 0.101e^{2.538x_1} + 0.218e^{3.933x_4} + 0.019$	0.917	0.162	x_3 is NDWI$_{2190}$, and
g	$y = 0.402e^{2.538x_1} - 0.26e^{3.949x_2} - 0.439e^{2.635x_3} - 0.034$	0.886	0.189	x_4 is NDRI
h	$y = 0.199e^{2.538x_1} - 0.234e^{3.949x_2} + 0.306e^{3.933x_4} + 0.048$	0.938	0.14	
i	$y = 0.019e^{3.949x_2} - 0.439e^{2.635x_3} + 0.39e^{3.933x_4} + 0.171$	0.908	0.17	
j	$y = 0.32e^{2.538x_1} - 0.433e^{2.635x_3} + 0.375e^{3.933x_4} + 0.088$	0.961	0.11	
k	$y = 0.261e^{2.538x_1} + 0.127e^{3.949x_2}$ $-0.604e^{2.635x_3} + 0.428e^{3.933x_4} + 0.092$	0.965	0.105	

By examining the characteristic parameters of the multiple linear regression model, which include VWC with two vegetation indices, Figure 4a–f show that NDVI + NDRI yielded the best result, with R^2 and RMSE values of 0.917 and 0.162, respectively. The R^2 gradually increased with an increasing number of characteristic parameters, whereas the RMSE gradually decreased. Regarding the characteristic parameters of multiple regression models that included VWC with three vegetation indices (Figure 4g–j), NDVI + NDWI$_{2190}$ + NDRI yielded the best result, with R^2 and RMSE values of 0.963 and 0.108, respectively. NDVI + NDWI$_{1610}$ + NDWI$_{2190}$ + NDRI yielded the best result of all combinations. The R^2 between the estimated and measured SSM was 0.965, and the correlation between them was statistically significant at the 0.01 level. Compared with the inputs of the three characteristic parameters, the R^2 values for model No. k were 0.079, 0.027, 0.057, and 0.002 higher, respectively. RMSE decreased by 0.084, 0.035, 0.065, and 0.006, respectively. Therefore, model No. k was used as the modified VWC estimation model in this study.

4.2. SSM Retrieval Results Using the Modified WCM

As stated in Section 4.1, the VWC values from the Sentinel-2 vegetation index were used in the formula of the WCM to obtain the soil backscattering coefficient. The corresponding backscatter coefficients for the 84 samples in their image according to the coordinates, their relationships before and after removing the VV vegetation influence, and the VH polarization backscattering coefficient are shown in Figure 5. The backscattering coefficients of VV polarization (Figure 5a) were higher than those of VH polarization (Figure 5b) at the same sampling points. After removing the influence of vegetation, the value of the soil backscatter coefficient was generally lower than that of the total backscatter coefficient. The variation in the VV polarization backscattering coefficient was −2.35 dB, while that of the VH polarization was −2.92 dB. The variation in the VH polarization backscattering before and after the correction was greater than that of the VV polarization. This indicates that VH polarization was more easily affected by the vegetation layer during the transmission process.

After extracting the most relevant features using SVR, training was performed, as shown in the flowchart in Figure 6. The R^2 and RMSE values are presented in Table 3. In order to evaluate the accuracy of the modified model, we compared the performances with the original WCM (where VWC is composed of NDVI) and the radar backscatter coefficient, which ignores the influence of vegetation.
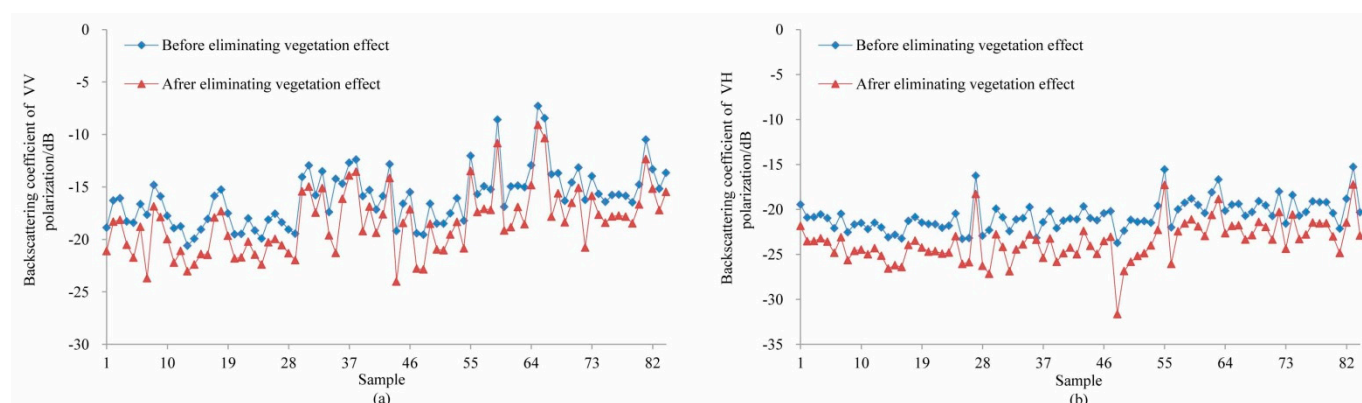
Figure 5. Graph of relationship between VV, VH polarization backscattering coefficients. (**a**) VV polarization. (**b**) VH polarization.

Table 3. Summary of results for different reference samples by SVR.

Reference Samples	VV		VH	
	R^2	RMSE (%)	R^2	RMSE (%)
radar backscattering coefficient	0.661	3.314	0.451	4.192
soil backscattering coefficient by original WCM	0.801	2.992	0.586	3.994
soil backscattering coefficient by modified WCM	0.86	2.119	0.667	3.629

As shown in Figure 6 and Table 3, for the VV polarization, the R^2 between the estimated and measured SSM was 0.86 for the modified WCM, and the RMSE of the estimated SSM was 2.119 %, while the R^2 and RMSE values of SSM were 0.801 and 2.992%, and 0.661 and 3.314% respectively, for the original WCM and radar backscatter coefficient. The correlation between the estimated and measured SSM was statistically significant at the 0.01 level.

In contrast, the scattered VH polarization points deviated more from the 1:1 line than those of the VV polarization. For the VH polarization, the R^2 between the estimated and measured SSM was 0.667 for the modified WCM, and the RMSE of the estimated SSM was 3.629%, while the R^2 and RMSE values of SSM were 0.586 and 3.994%, and 0.451 and 4.192% respectively, for the original WCM and radar backscatter coefficient. The correlation between the estimated and measured SSM was statistically significant at the 0.05 level.

From the results, this study confirms previous findings regarding the significance of using vegetation indices in the soil moisture retrieval process in vegetated areas. Further, VV polarization had good accuracy and stability for retrieving SSM in the study area, and the modified WCM yielded satisfactory results in retrieving SSM with Sentinel-1 and Sentinel-2 data.

The spatial distribution of SSM retrievals and frequency diagram of SSM in the study area are shown in Figure 7. Based on the supervised classification technology of threshold segmentation by the environment for visualizing images (ENVI) software, the non-wheat areas, such as towns, rivers, and other non-agricultural areas in the Sentinel-1 SAR image of the study area were removed.

As seen in Figure 7, the retrieved SSM in the study areas were mainly distributed in the range of 25–40%. The result of the SSM retrieval on 3 April was slightly drought, but the SSM in a small part of the central and southern were significantly higher than that in other areas (Figure 7a). There is a possibility of being irrigated of the wheat fields in the dry spring. Due to the fact that there was continuous rainfall in the south of the study area, the results of SSM retrieval were relatively moist in the southern on 3 May (Figure 7b). According to the meteorological data, there was no amount of rainfall in the study area before the satellite's transit, and this may explain why the SSM was slightly lower on 23 May than that on 3 May (Figure 7c). The SSM retrieval results were basically

consistent with the measured SSM. Therefore, the method proposed in this study had strong applicability for the study area.
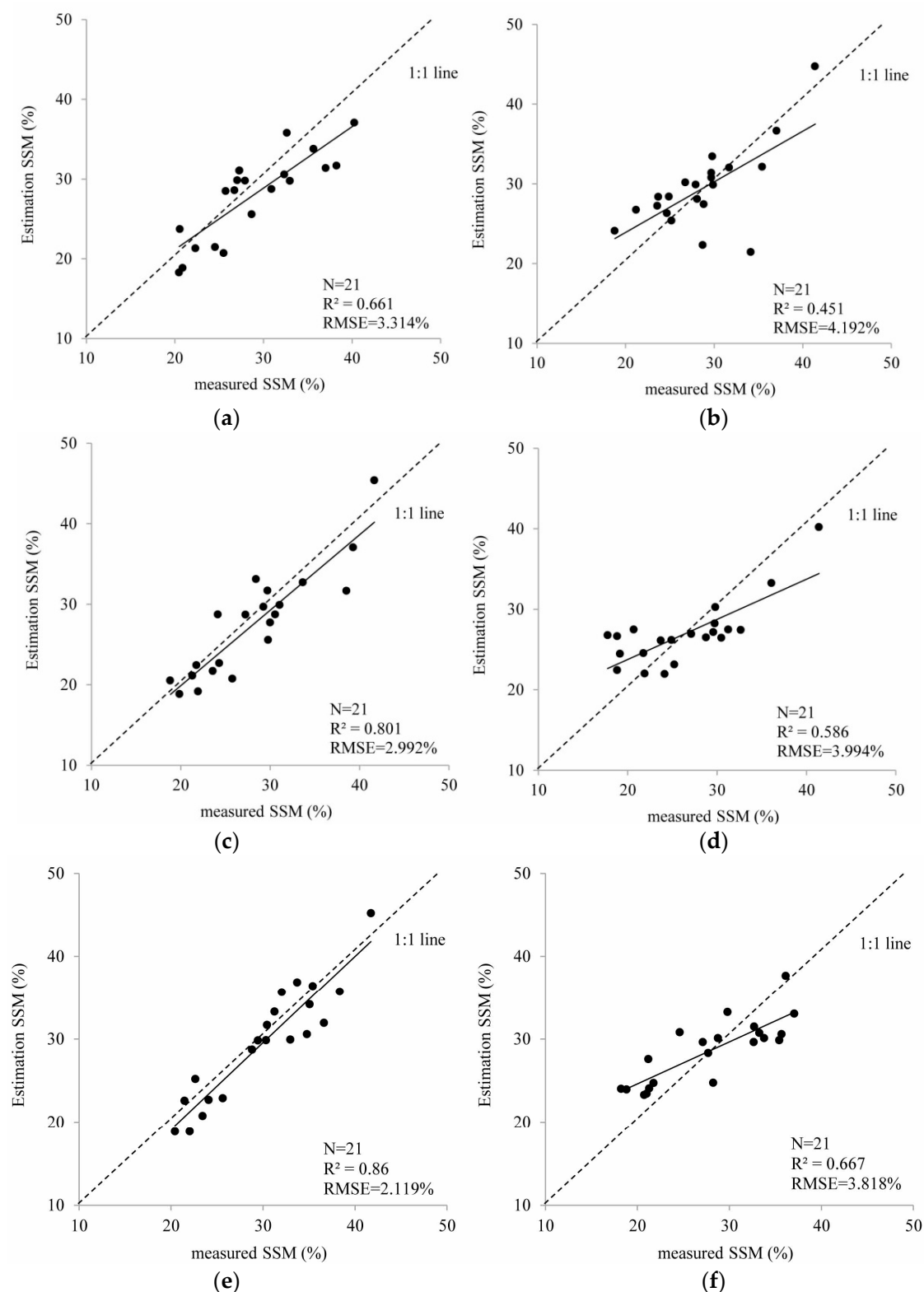

Figure 6. Scatterplot of estimated versus measured SMC (%) values. (**a**) Radar backscattering coefficient of VV polarization. (**b**) Radar backscattering coefficient of VH polarization. (**c**) Soil backscattering coefficient of VV polarization by the original WCM. (**d**) Soil backscattering coefficient of VH polarization by the original WCM. (**e**) Soil backscattering coefficient of VV polarization by the modified WCM. (**f**) Soil backscattering coefficient of VH polarization by the modified WCM.
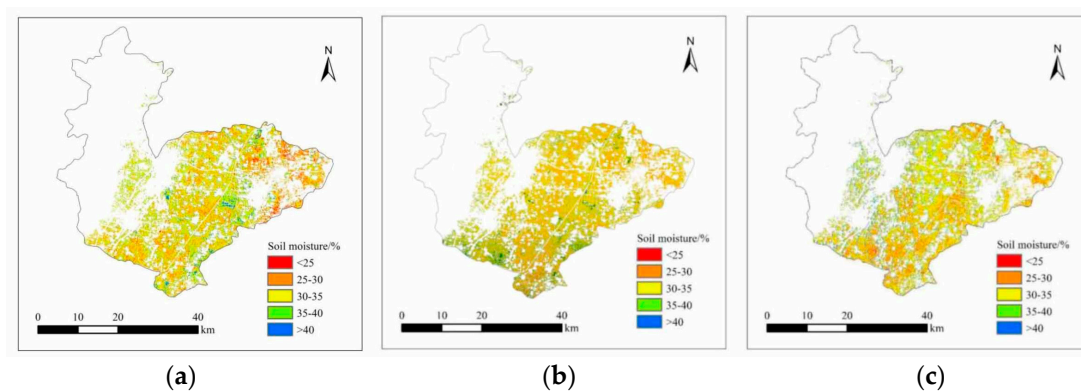
Figure 7. SSM maps of wheat-covered fields in study area. (**a**) 3 April 2019. (**b**) 3 May 2019. (**c**) 23 May 2019.

4.3. Discussion

As a classical model, the WCM has been widely used to retrieve soil moisture information from areas with vegetation cover. In recent years, some studies have used Sentinel-1 and Sentinel-2 data to estimate SSM. Guo et al. [38] estimated farmland SSM using multi-source remote sensing data from Sentinel-1 radar and Sentinel-2 optical images. They applied the Oh model [39], SVR, and a generalized regression neural network (GRNN) to retrieve SSM. They used the WCM to remove the influence of vegetation. The inputs of SVR included the dual-polarization radar backward scattering coefficient, altitude, local incident angle, and vegetation indices (NDVI, the modified soil adjusted vegetation index (MSAVI), and the difference vegetation index (DVI)). Their results indicated that combining multi-characteristic parameters based on SVR delivered the best retrieval accuracy with the R^2 of 0.903 and an RMSE of 0.014 cm^3/cm^3. Zhao et al. [40] estimated the SSM for winter wheat fields using Sentinel-1 and Sentinel-2 data. Based on near-infrared, red, and shortwave infrared bands, they proposed a new fusion vegetation index (FVI) to estimate VWC. They used the Maclaurin series to improve the WCM and considered that the single-polarization backscattering coefficients could be replaced by VV/VH. As a result of their retrieval analysis, they obtained an R^2 value of 0.7642 and an RMSE of 0.0209 cm^3/cm^3 in VV/VH; their R^2 and RMSE values were 0.6791 and 0.0249 for VV polarization, and 0.5151 and 0.0289 for VH polarization, respectively. In the present study, vegetation indices including NDVI, $NDWI_{1610}$, $NDWI_{2190}$, and NDRI were used (NDRI was composed of two red edge bands). To the best of the authors' knowledge, this study is the first to propose removing the influence of vegetation on SSM estimation by using the red side bands in Sentinel-2 data. Baghdadi et al. [7] estimated the SSM of crop fields and grasslands from Sentinel-1/2 data. They combined the WCM with the integral equation model (IEM) [41] using real data composed of a C-band radar backscattered signal, NDVI, soil moisture, and surface roughness values. Their results indicated that the soil contribution to the total radar backscatter signal was lower in VH polarization than in VV polarization. Zeng et al. [17] studied SSM under different vegetation covers based on Sentinel-1A and SVR techniques and concluded that VV polarization could achieve high retrieval accuracy. Wang et al. [42] combined full polarization Radarsat-2 SAR data and SVR techniques to estimate soil moisture in sparsely vegetated arid areas. They determined that the inversion accuracy of the co-polarization data (VV or HH polarization) was higher than that of the cross-polarization data (VH or HV polarization). Comparing the inversion results reported in this study with those of the previous studies mentioned above, it is possible to conclude that VV polarization is more sensitive to SSM than VH polarization.

One of the limitations of this study is that it only focused on wheat fields. In addition, only a small number of sampling sites were measured. Furthermore, as wheat is a drought-resistant crop, it is also planted on hills and mountains in China, while this study only examined wheat on plains. However, soil moisture retrieval methods that are based on

other kinds of crops and those that encompass many terrain conditions may have more practical significance than the method reported here.

5. Conclusions

Using Hebi, a representative wheat planting area in Henan Province, China, as the study area, we investigated the potential for synergy between C-band Sentinel-1 SAR and Sentinel-2 optical data for SSM retrieval in wheat fields. To extract the soil backscattering coefficient (σ_{soil}^0) from the Sentinel-1 SAR data, the WCM was selected to remove the influence of the vegetation layer from the radar backscattering coefficient. Then, combined with the WCM and SVR algorithms, the SSM of the wheat-covered fields were retrieved and analyzed under different polarization modes (VV, VH). The main conclusions of this study can be summarized as follows:

(1) A modified WCM was constructed using FV and VWC values calculated from Sentinel-2 data. The FV was used to distinguish the proportions of vegetation coverage and bare soil in pixels, and the VWC model, which was based on the combination of four vegetation indices (NDVI, $NDWI_{1610}$, $NDWI_{2190}$, and NDRI), was able to effectively remove the influence of the vegetation canopy on the backscattering coefficient of the Sentinel-1SAR data;

(2) Compared with Sentinel-1 VH polarization data (after removing the vegetation influence using the WCM), VV polarization data produced higher estimation accuracies regarding the SSM retrieval. This result indicates that the VV polarization contains more soil backscattering information and is more sensitive to changes in SSM than VH polarization;

(3) C-band Sentinel-1 SAR and Sentinel-2 optical data were used to study wheat-covered fields. The results indicate that the estimated SSM based on these two kinds of satellite data are applicable to agricultural environments for wheat. To further this research, we intend to examine the proposed algorithm with regard to other crops and different regions. Furthermore, to estimate SSM more accurately on a large scale, some advanced algorithms, such as deep learning, should be implemented.

Author Contributions: Y.L. conceived the idea, implemented the algorithm, interpreted results, and analyzed the data. C.Z. discussed the details of the algorithm, contributed to the manuscript's preparation, and revised the manuscript. W.H. designed experiments, performed field experiments, and processed data. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study will be made available on request from the corresponding author.

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

References

1. Miralles, D.G.; Teuling, A.J.; Heerwaarden, C.; De Arellano, J.V. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.* **2014**, *7*, 345–349. [CrossRef]
2. Vereecken, H.; Huisman, J.A.; Franssen, H.H.; Brüggemann, N.; Bogena, H.R.; Kollet, S. Soil hydrology: Recent methodological advances, challenges, and perspectives. *Water Resour. Res.* **2015**, *51*, 2616–2633. [CrossRef]

3.   Cho, E.; Choi, M. Regional scale spatio-temporal variability of soil moisture and its relationship with meteorological factors over the Korean peninsula. *J. Hydrol.* **2014**, *516*, 317–329. [CrossRef]

4.   Brocca, L.; Tullo, T.; Melone, F.; Moramarco, T.; Morbidelli, R. Catchment scale soil moisture spatial–temporal variability. *J. Hydrol.* **2012**, *422*, 63–75. [CrossRef]

5.   Jiang, J.; Hu, D.; Li, Y.; Tang, X.; Li, J. Research of soil moisture retrieval model of wheat covered surface based on mimics model. *J. Triticeae Crop* **2015**, *35*, 707–713.

6.   Dong, J.Z.; Wade, C.; Kenneth, J.T.; Michael, H.C. Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation. *Remote Sens. Environ.* **2020**, *242*, 111756. [CrossRef]

7.   Baghdadi, N.; El, H.M.; Zribi, M.; Bousbih, S. Calibration of the water cloud model at C-band for winter crop fields and grasslands. *Remote Sens.* **2018**, *9*, 969. [CrossRef]

8.   Bai, X.; Zhang, L.; He, C.; Zhu, Y. Estimating regional soil moisture distribution based on NDVI and land surface temperature time series data in the upstream of the heihe river watershed, northwest china. *Remote Sens.* **2020**, *12*, 2508. [CrossRef]

9.   Soliman, A.; Heck, R.J.; Brenning, A.; Brown, R.; Miller, S. Remote sensing of soil moisture in vineyards using airborne and ground-based thermal inertia data. *Remote Sens.* **2013**, *5*, 3729–3748. [CrossRef]

10. Wang, J.; Ding, J.L.; Chen, W.Q.; Yang, A.X. Microwave modeling of soil moisture in oasis regional scale based on sentinel-1 radar images. *J. Infrared Millimeter Waves* **2017**, *36*, 120–126.

11. Peischl, S.; Walker, J.P.; Ye, N.; Ryu, D.; Kerr, Y. Sensitivity of multi-parameter soil moisture retrievals to incidence angle configuration. *Remote Sens. Environ.* **2014**, *143*, 64–72. [CrossRef]

12. Pan, M.; Sahoo, A.K.; Wood, E.F. Improving soil moisture retrievals from a physically-based radiative transfer model. *Remote Sens. Environ.* **2014**, *140*, 130–140. [CrossRef]

13. Zhu, L.; Walker, J.P.; Shen, X. Stochastic ensemble methods for multi-SAR-mission soil moisture retrieval. *Remote Sens. Environ.* **2020**, *251*, 112099. [CrossRef]

14. Aubert, M.; Baghdadi, N.N.; Zribi, M.; Ose, K. Toward an operational bare soil moisture mapping using terrasar-x data acquired over agricultural areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 900–916. [CrossRef]

15. He, B.; Xing, M.; Bai, X. A synergistic methodology for soil moisture estimation in an alpine prairie using radar and optical satellite data. *Remote Sens.* **2014**, *6*, 10966–10985. [CrossRef]

16. Zhou, P.; Ding, J.L.; Wang, F.; Guljamal, U.; Lei, L.P. Retrieval methods of soil water content in vegetation covering areas based on multi-source. *J. Remote Sens.* **2010**, *14*, 959–973.

17. Zeng, X.; Xing, Y.; Wei, S.; Zhang, Y.; Wang, C. Soil water content retrieval based on sentinel-1a and landsat 8 image for bei'an-heihe expressway. *J. Eco-Agric.* **2017**, *25*, 118–126.

18. Temimi, M.; Leconte, R.; Chaouch, N.; Sukumal, P.; Khnbilvardi, R.; Brissette, F. A combination of remote sensing data and topographic attributes for the spatial and tempeoral monitoring of soil wetness. *J. Hydrol.* **2010**, *388*, 28–40. [CrossRef]

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

20. Bousbih, S.; Zribi, M.; El Hajj, M.; Baghdadi, N.; Lilichabaane, Z.; Gao, Q.; Fanise, P. Soil moisture and irrigation mapping in a semi-arid region, based on the synergetic use of Sentinel-1 and Sentinel-2 data. *Remote Sens.* **2018**, *10*, 1953. [CrossRef]

21. Wang, Q.; Li, J.; Jin, T.; Chang, X.; Li, D. Comparative analysis of Landsat-8, Sentinel-2, and GF-1 data for retrieving soil moisture over wheat farmlands. *Remote Sens.* **2020**, *12*, 2708–2714.

22. Kim, H.O.; Yeom, J.M. Effect of red-edge and texture features for object-based paddy rice crop classification using rapideye multi-spectral satellite image data. *Int. J. Remote Sens.* **2014**, *35*, 7046–7068. [CrossRef]

23. Fang, C.; Wang, L.; Hanqiu, X.U. A comparative study of different red edge indices for remote sensing detection of urban grassland health status. *J. Geo-Inform. Sci.* **2017**, *19*, 1382–1392.

24. 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]

25. Lee, J.S. Speckle filtering of synthetic aperture radar images: A review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [CrossRef]

26. Attema, E.P.W.; Fawwaz, T. Vegetation modeled as a water cloud. *Radio Sci.* **1978**, *13*, 357–364. [CrossRef]

27. Bao, Y.; Lin, L.; Wu, S.; Abdalla, K.; 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]

28. Yadav, V.P.; Prasad, R.; Bala, R.; Vishwakarma, A.K. Estimation of soil moisture through water cloud model using sentinel -1A SAR data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.

29. Yang, G.; Yue, J.; Li, C.; Feng, H.; Lan, Y. Estimation of soil moisture in farmland using improved water cloud model and radarsat-2 data. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 146–153.

30. Bindlish, R.; Barros, A.P. Parameterization of vegetation backscatter in radar-based, soil moisture estimation. *Remote Sens. Environ.* **2001**, *76*, 130–137. [CrossRef]

31. Gutman, G.; Ignatov, A. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int. J. Remote Sens.* **1998**, *19*, 1533–1543. [CrossRef]

32. Lin, L.; Bao, Y.; Zuo, Q.; Fang, S. Soil moisture retrieval over vegetated areas based on sentinel-1 and fy-3c data. *Remote Sens. Technol. Appl.* **2018**, *33*, 750–758.

33. Deering, D.W. *Rangeland Reflectance Characteristics Measured by Aircraft and Spacecraft Sensors*; Texas A&M University: College Station, TX, USA, 1978.

34. Gao, B. NDWI-a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]

35. Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Bruzzone, L.; Remelgado, R.; Greifeneder, F. Estimation of soil moisture in mountain areas using svr technique applied to multiscale active radar images at c-band. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *8*, 262–283. [CrossRef]

36. Bertoldi, G.; Chiesa, S.D.; Notarnicola, C.; Pasolli, L.; Niedrist, G.; Tappeiner, U. Estimation of soil moisture patterns in mountain grasslands by means of SAR radarsat2 images and hydrological modeling. *J. Hydrol.* **2014**, *516*, 245–257. [CrossRef]

37. Wang, J.; Cai, L.; Zhao, X. Multiple-instance learning via an RBF kernel-based extreme learning machine. *J. Intell. Syst.* **2016**, *26*, 185–195. [CrossRef]

38. Guo, J.; Liu, J.; Ning, J.; Han, W. Construction and validation of soil moisture retrieval model in farmland based on sentinel multi-source data. *Trans. CSAE* **2019**, *35*, 71–78.

39. Oh, Y. Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 596–601. [CrossRef]

40. 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. Inform. Technol.* **2021**, *43*, 692–699.

41. Fung, A.K.; Li, Z. Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 356–369. [CrossRef]

42. Wang, Y.; Kong, J.; Yang, L.; Li, J.F.; Zhang, W. Remote sensing inversion of soil moisture in vegetation-sparse arid areas based on SVR. *J. Geo-inform. Sci.* **2019**, *21*, 1275–1283.