Article

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

Yan Li [1,2], Chengcai Zhang [1,*], Weidong Heng [1], Feng, Yang [3] and Jisheng Liu [3]

[1] School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China; yanli@gs.zzu.edu.cn (Y.L.); zhangcc@zzu.edu.cn; hengweidong@gs.zzu.edu.cn (W.H)
[2] Zhengzhou Vocational College of Industrial Safety, Zhengzhou 451192, China; yanli@gs.zzu.edu.cn (Y.L.)
[3] Chushandian Reservoir Management Bureau, Xinyang 464000, China
* Correspondence: Correspondence: zhangcc2000@163.com (Z.Z.); Tel.: +86-135-2348-0466

Abstract: Surface soil moisture (SSM) is a significant factor affecting crop growth. This paper presents a method for retrieving SSM over wheat-covered areas using synergy dual-polarization C-band Sentinel-1 synthetic aperture radar and Sentinel-2 optical data. Firstly, 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 was then introduced in this WCM, and the vegetation water content (VWC) was calculated using 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 the wheat field in Hebi, in the north of Henan Province. The key findings of this study are as follows: (1) Based on vegetation indices obtained from Sentinel-2; the proposed VWC estimation model can effectively eliminate the influence of vegetation; (2) compared with vertical transmit and horizontal receive polarization, vertical transmit and vertical receive polarization is better for detecting changes in SSM at different growth stages of wheat; and, (3) the validation results indicated that the proposed approach, based on Sentinel-1 and Sentinel-2 data, successfully retrieved SSM in the study area.

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, and the energy and water cycles. It therefore plays an essential role in hydrology, climatology, meteorology, ecology, and agronomy [1–3]. SSM is especially important in arid or 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].

In recent decades, several quantitative studies have proven that microwave remote sensing, especially the synthetic aperture radar (SAR), can effectively estimate SSM [8–10]. As active instruments, SAR sensors provide SSM observations over large areas in all weather conditions. Furthermore, they can operate during both the day and night. SAR sensors are mainly based on the difference in the energy between the transmitted and received electromagnetic radiation [11–12]. Important SAR satellites include the European Space Agency (ESA)'s ERS-1/2, ENVISAT-1 and Sentinel-1; the Canadian Space Agency (CSA)'s Radarsat-1/2; the German Aerospace Center (DLR)'s TerraSar-X; China's GF3, and Japan's ALOS-2. Sentinel-1 is an earth observation satellite that is part of the ESA's global monitoring for environment and security. It provides more channels for observing soil moisture at a high spatial resolution (10 m), has a regular temporal coverage (5–6 days when both the Sentinel-1A and Sentinel-1B satellites are considered), and its data is freely available [13–14].

The relationship between soil moisture and the radar backscattering coefficient is well documented [15–18]. According to these studies, the backscattering of SAR is significantly affected by the natural surfaces (i.e., the dielectric constant of the soil, the roughness of the soil surface, and the presence of vegetation) [19–20]. In vegetated areas, a frequently used model to estimate the backscattering of SAR under different conditions is the semi-empirical water cloud model (WCM). Attema and Ulaby [17] assumed that the vegetation canopy could be modeled as a homogeneous horizontal cloud of identical water spheres, uniformly distributed over the surface of soil. Chai *et al.* [21] estimated SSM from C-band Radarsat-2 SAR data by using the WCM to fit the Chen model [22] and the Dubois model [23] to study the grassland vegetation area of the northeastern Tibetan Plateau. Bindlish *et al.* [24] utilized a modified WCM in conjunction with the integrated equation model (IEM); they allowed for the explicit representation of vegetation backscattering effects, without the need to specify a large number of parameters. Lin *et al.* [25] developed the WCM by using the microwave polarization difference index (MPDI) to express vegetation water content. They concluded that vertical transmit and vertical receive (VV) polarization based on Sentinel-1A data, when combined with MPDI, could achieve a high retrieval accuracy for soil moisture.

It should be noted that spectral reflectance indices computed with visible, near-infrared, and even short-infrared data are highly sensitive to vegetation information. Therefore, the synergistic application of SAR and optical data has great advantages for SSM estimation [26–29]. Zeng *et al.* [30] proposed an integrated approach to improve the retrieval accuracy of soil moisture content (SMC) using the WCM and the support vector regression (SVR) technique. They concluded that combining VV polarization with the normalized differential vegetation index (NDVI) could achieve a high retrieval accuracy. Bao *et al.* [31] presented a new methodology for retrieving soil moisture based on the synergy between Sentinel-1 SAR and Landsat 8 data. In order to remove the effect of vegetation, they obtained different vegetation spectral indices from Landsat 8 to retrieve the vegetation water content (VWC). Their results showed that the normalized difference water index (NDWI), which was built from the 1.57–1.65 μm band was the best removing the effect of vegetation effect. However, Sentinel-2 incorporates three red-edge spectral bands that centered at 705, 740, and 783 nm, which are closely related to important biochemical parameters of green plants. However, few researchers have explored the advantages of using the red-edge spectral data to retrieve SSM.

Considering the above facts, 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 the methodology for eliminating the influence of vegetation from the radar backscattering of wheat. To this end, the WCM was developed and used. The vegetation spectral indices obtained from Sentinel-2 were used to express the VWC, which is an important parameter of the WCM. Moreover, the SVR technique [32–33], which has good robustness to the limited availability of reference results in various application domains, was applied to retrieve SSM.

The remainder of this paper is organized as follows. Section 2 introduces the study area on which our analysis is focused, and details the available remote sensing and in-situ measurements. Section 3 describes the methods of retrieval for SSM, the WCM and SVR technique. Section 4 performs the experimental activities and soil moisture maps with respect to the algorithm test and discusses the advantages and limitations of the approach. Finally, Section 5 presents the conclusions of the study.

2. Study Area and Dataset

2.1. Study area

The study site chosen for this study was Hebi (113°59′E–114°45′E, 35°26′N–36°02′N), which is located in the north of Henan province, China (Fig. 1). The agricultural fields analyzed in this study area mainly lie on the Huang-Huai-Hai Plain, which has a homo-

geneous soil texture. Hebi has a sub-humid warm temperate continental monsoon climate. The average rainfall ranges between 349.2 mm and 970.1 mm, with peaks occurring between May and August. Dry spells occur between November and March, so the climate's seasonal characteristics are favorable for crop farming. The cropping systems in this area usually involve wheat, corn, cotton, and canola, and wheat cycles occur between late September (emergence) and middle June (harvest) of the following year. A total of 28 sampling sites were selected, on slopes between 0% and 5% (Fig. 1c).

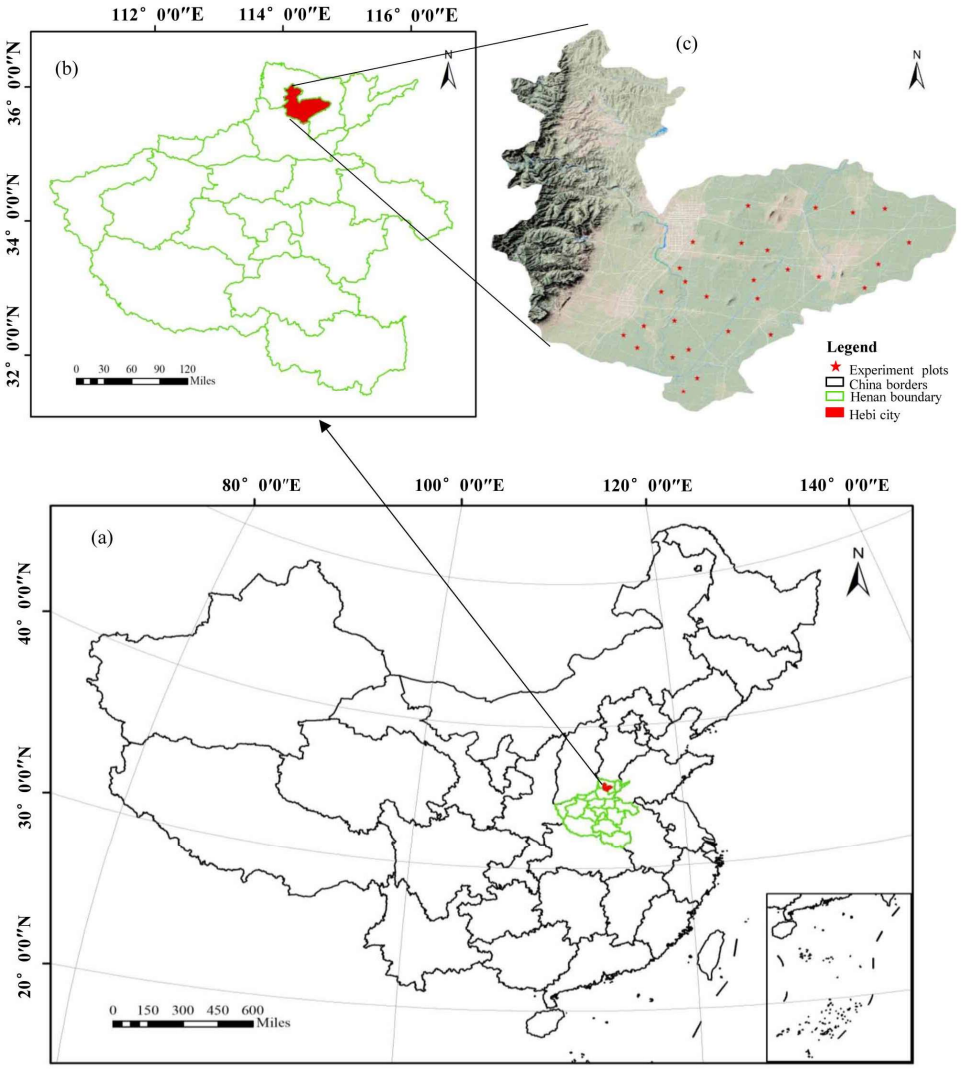


Figure 1. General overview of the study area and map showing locations of sampling sites.

2.2. Remote sensing data

2.2.1. Sentinel-1 data

The Sentinel-1A satellite was launched in April 2014, and was followed by Sentinel-1B in April 2016. The two satellites are equipped with C-band (5.405 GHz) SAR instruments, share the same orbital plane with a 180° orbital phasing difference, and provide data with a spatial resolution of 10 m and a temporal resolution of six days, in both VV and vertical transmit and horizontal receive (VH) polarizations.

In the study areas, three Sentinel-1 images from April 9, May 3, and May 27, 2019 were acquired. The acquired SAR images are single look complex (SLC) product and level-1B data, which need to be pre-processed prior to analysis. The Sentinel-1 data were pre-processed using the simplified numerical automatic programmer (SNAP) software, using the following steps:

(1) Radiometric calibration. SAR images were corrected through 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:

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

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 and geometric correction.

(3) Speckle noise removal. The refined Lee filter method was applied to the multi-look images [34]. In this method, the weights of neighboring pixels were first estimated using kernel density, and then 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. Fig. 2(a) shows the pre-processing results of Sentinel-1 SAR data from the study area for April 9, 2019.

2.2.2. Sentinel-2 data

Sentinel-2 satellites provide wide-swath, high-resolution, multi-spectral imaging, and are designed to provide full and systematic coverage of the Earth's land surfaces [35]. Moreover, Sentinel-2 offers a high revisit of five days at the equator under the same viewing conditions, high spatial resolution (the spatial resolutions of bands 2, 3, 4, and 8 can reach 10 m) and a wide multispectral field of view (from 13 spectral bands in the visible region). It is important to note that Sentinel-2 is the only multispectral satellite with three bands (bands 5, 6, and 7) in the red edge range.

In order to avoid effects of cloud contamination, three optical images were acquired that had either no or few clouds (April 3, May 3, and May 27, 2019). These acquired optical images were level-2A data, which do not require radiometric calibration or atmospheric correction. Fig. 2(b) shows the Sentinel-2 optical image of the study area on April 3, 2019.

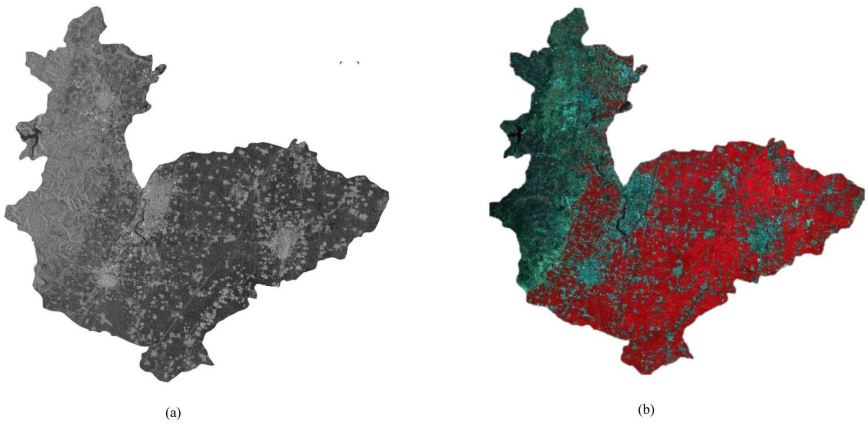


Figure 2. Remote sensing images of the study area. (a) Sentinel-1 SAR image (VV polarization); (b) Sentinel-2 composite false red green blue (RGB) images (Band 8 = red, Band 4 = green, and Band 3 = blue).

2.3. In-situ measurements

2.3.1. SSM measurements

The gravimetric soil moisture (GSM) was obtained using the cutting ring method. Almost simultaneous with the satellites overpass, soil moisture measurements over depths of 0–10 cm were conducted in different stages of wheat growth. The data collection was divided into three periods, with 28 sampling sites selected for each period. Samples were collected during the jointing, heading, and filling stages of wheat. For each sampling site, three sampling points were selected; the distance between each sampling point was 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 were then dried in an oven at 105° for 20–24 h, until a constant temperature was reached. GSM was obtained using the following formula:

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

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

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

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

where ρ_b is soil bulk density.

The SSM used in this study was the VSM.

2.3.2. Vegetation parameter

To obtain the vegetation parameter, wheat samples were collected at the same times that soil moisture measurements were taken. The main vegetation parameters were plant height, plant density, and VWC. The height of wheat was measured using a meter scale over a 20 × 20 cm area of wheat; the plant density was determined for a 1 m2 area in each wheat field. The fresh weight (WF) was measured after the wheat samples were brought back to the laboratory, and the dry weight (WD) was obtained by drying the wheat in an oven. The specific equation for VWC is as follows:

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

where ρ is plant density of wheat.

The main information of the experimental sites is summarized in Table 1.

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

Sentinel-1	Sentinel-2	In-situ	Temperature	Notes
2019-4-9	2019-4-3	2019-4-7	8 – 20 °C	There was a small amount of rain on the time of the Sentinel-1 transit.
2019-5-3	2019-5-3	2019-5-3	14 – 27 °C	/
2019-5-27	2019-5-27	2019-5-27	20 – 37 °C	There was a small amount of rain a week before the satellite transit, but high temperatures meant that evaporation was also high.

3. Methods

Here, a proper soil moisture retrieval method is proposed 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 illustrated in Fig. 3. The SSM retrieval method used here comprised three main phases. The first one was obtaining parameters of 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 of 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 built. 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 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. Furthermore, the coefficient of determination (R^2) and root mean square error (RMSE) were calculated to evaluate the accuracy of the SSM estimations, and the SSM values were mapped across the study area. More details on each part of the soil moisture retrieval algorithm are given below.

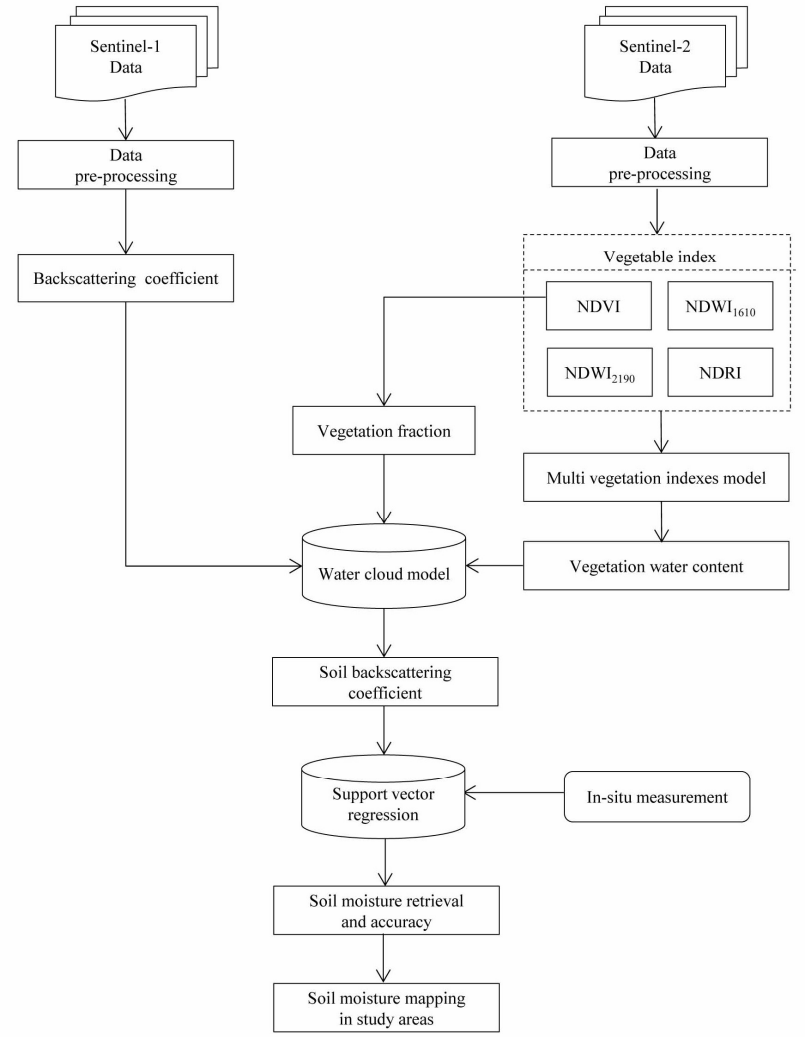


Figure 3. Flowchart of the soil moisture retrieval process based on Sentinel-1 and Sentinel-2 data.

3.1. the modified WCM

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

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

$$\sigma_{veg}^0 = aV_1\cos(\theta)(1 - L^2) \quad (6)$$

$$L^2 = \exp[-2bV_2/\cos(\theta)] \quad (7)$$

where θ is the incident angle, f_v is the vegetation fraction; σ_{pp}^0 is the co-polarized total backscattering coefficient, σ_{veg}^0 is the backscatter contribution of the vegetation canopy, σ_{soil}^0 is the backscatter contribution of the soil surface, and L^2 is the double attenuation factor. V_1 and V_2 denote the vegetation descriptors. It can be VWC, NDVI, the leaf area index (LAI), or other vegetation descriptors [31, 36–38]. In this study, V_1 and V_2 refer to VWC (kg/m^2). The empirical parameters of *a* and *b* depend on the vegetation type and the incident angle. Bindlish and Barros [24] proposed values for *a* and *b* for different land cover types. As the crop in the study area is wheat, winter wheat was selected, so values of 0.0018 and 0.138 were used for *a* and *b*, respectively.

Moreover, f_v is an additional parameter of the vegetation scattering model; it is used to distinguish the proportion of vegetation coverage and bare soil in pixels. f_v can be calculated using the mixed pixel decomposition model [39]:

$$f_v = (NDVI - NDVI_{min})/(NDVI_{max} - NDVI_{min}) \quad (8)$$

where $NDVI_{min}$ denotes the bare soil pixel, which is theoretically close to zero, and $NDVI_{max}$ denotes the 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 of $NDVI_{min}$ and $NDVI_{max}$.

3.2. Building a VWC model

Based on previous published studies, the vegetation indices can be used to estimate VWC [3, 13, 40]. The vegetation indices commonly used for VWC estimation include NDVI [41] and NDWI [42]. These two vegetation indices are be calculated as follows:

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

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

where R_{NIR} is the reflectivity at the near-infrared band, R_{Red} is the reflectivity at the red band, and R_{SWIR} is the reflectivity at 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 delivered good applications for identifying surface types, calculating parameters, distinguishing vegetation growth states, and estimating vegetation leaf area indices [43–44]. The normalized difference red-edge index (NDRI) can be calculated as follow:

$$NDRI = (R_{Red-edge1} - R_{Red-eage2})/(R_{Red-edge1} + R_{Red-eage2}) \quad (11)$$

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

Statistical analysis is often used to establish the relationship between a vegetation index and in-situ measured VWC data [29–30, 45]. Studies have used linear, one-variable quadratic, and exponential models to quantitatively describe the correlation between them [31, 35, 45].

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

$$C = \alpha e^{\beta x} \quad (12)$$

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 taken as the characteristic parameter of the multiple linear regression equation. Thus, the modified VWC can be expressed as

$$m_{veg} = \gamma_0 + \gamma_1 C_1 + \gamma_2 C_2 + \cdots + \gamma_k C_k + \varepsilon \quad (13)$$

where m_{veg} is the modified VWC, k is the number of vegetation indices, ε is the bias that obeys the normal distribution $N(0, \sigma^2)$, 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:

$$\begin{cases} y_1 = \sum_{i=0}^{k} \gamma_i C_{i1} + \varepsilon_1 \\ \cdots \\ y_n = \sum_{i=0}^{k} \gamma_i C_{in} + \varepsilon_n \end{cases} \quad (14)$$

The value of parameter γ_i (i=0,1,2,…,k) can be estimated.

3.3. SVR estimation of SSM

The retrieval process, which is an approach to derive SMC information from SAR backscattering coefficients, was carried out by applying a machine learning technique, namely SVR. In order to estimate the VSM from the dual-polarization SAR Sentinel-1 backscattering coefficients, the retrieval process was carried out using SVR. The entire procedure was divided into two main phases: training and validation. For the 28 samples that were acquired in each period, 20 random samples were used for training and the rest were used for validation. Owing to its accurate estimation, good intrinsic generalization ability and ability to deal with complex nonlinear problems, the SVR technique can be widely applied for soil moisture estimation [32–33].

SVR is a supervised regression technique that transforms the nonlinear problem of the soil backscattering coefficient, normalized red edge index, vegetation coverage, and soil moisture into the linear problem of a higher dimensionality space, as follows:

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*) K(x_i, x) + b \quad (15)$$

where n is the number of training samples, a_i^* represent the Lagrange multipliers of the optimization problem; k(. . .) is a kernel function, where x_i and x are the predictor variables; and b is the bias.

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. Furthermore, different kernel functions can be used to construct different support vector machine [35–36]. The commonly adopted kernel is the Gaussian radial basis function (RBF) kernel. The RBF kernel has good anti-interference ability, and less numerical difficulties than other kernel functions [46]. It can deal with samples when the relationship between class labels and features is nonlinear:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\delta^2}\right), \quad \delta > 0 \tag{16}$$

where δ is the reach rate, that is, the rate at which soil moisture falls to 0.

4. Results and discussion

4.1. Estimating VWC from Sentinel-2 data

A total of 84 in-situ measurement samples were acquired, of which 68 random samples were used to build establish VWC models, and the remaining 16 samples were applied to validate the performance of the models. According to equation (12), the fitting relationship between the measured VWC and the estimated VWC that based on the multiple linear regression model under different combinations of vegetation indexes is shown in Fig. 4. The R^2 and RMSE between the estimated and measured VWC based on the 16 sample data are shown in Table 2.

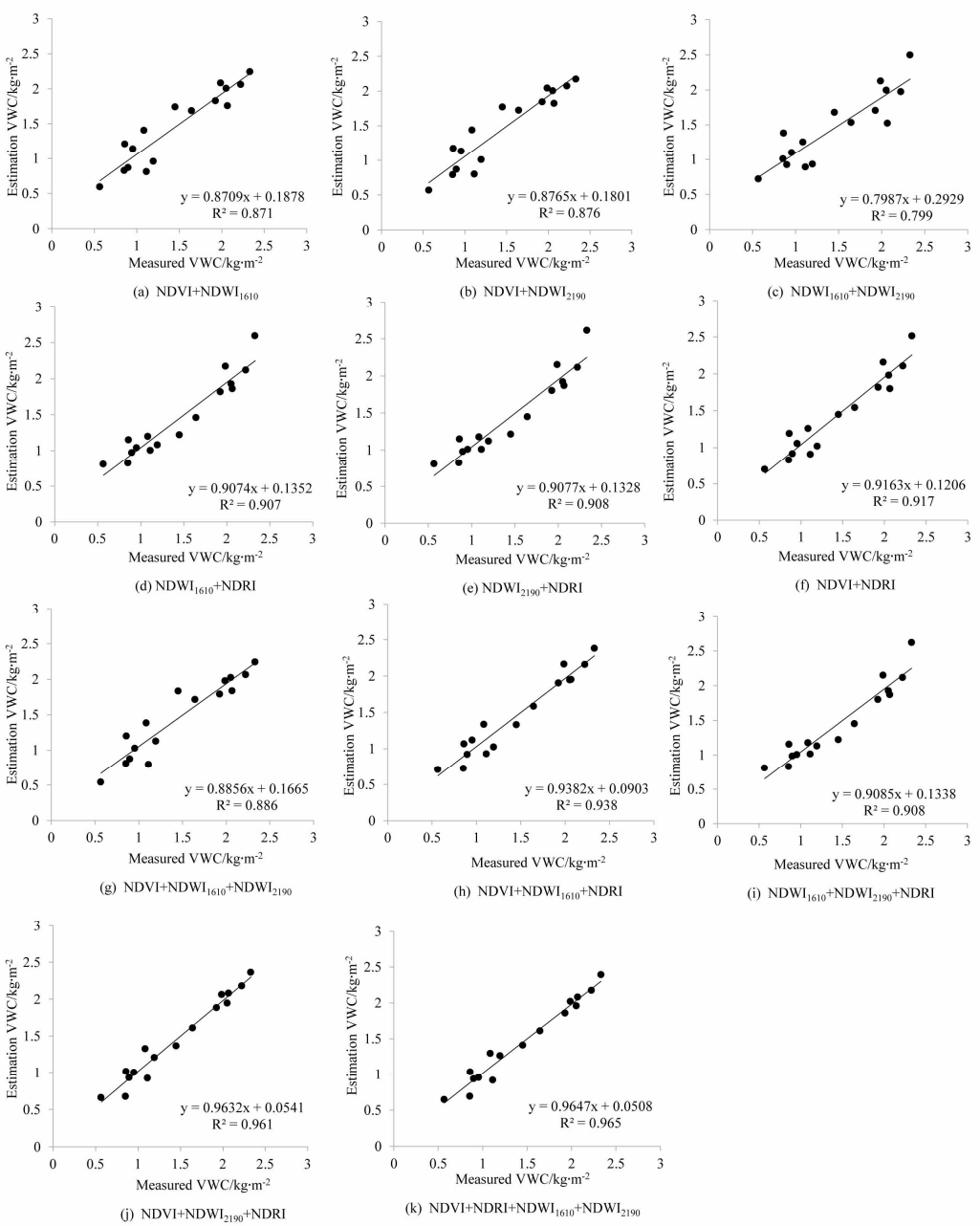


Figure 4. Scatter plot between measured and VWC estimations, with different combinations of vegetation indices.

Examining the characteristic parameters of the multiple linear regression model, that include VWC with two vegetation indices (Nos. a–f), it can be seen that NDVI + NDRI obtained the best result, with R^2 and RMSE values of 0.917 and 0.162, respectively. R^2 gradually increased with increasing numbers of characteristic parameters, whereas RMSE gradually decreased. Regarding the characteristic parameters of multiple regression models including VWC with three vegetation indices (Nos. g–j), NDVI + $NDWI_{2190}$ + NDRI obtained the best result, with R^2 and RMSE of 0.963 and 0.108 respectively. NDVI+$NDWI_{1610}$+$NDWI_{2190}$+NDRI obtained the best result over all combinations. Compared with the inputs of the three characteristic parameters, the R^2 values of model No. k were 0.079, 0.027, 0.057, and 0.002 higher, respectively; its RMSE values 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 our study.

Table 2. Correlations of 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	
e	$y = 0.391e^{3.933x_4} - 0.088e^{2.635x_3} + 0.169$	0.908	0.17	x_1 is NDVI,
f	$y = 0.101e^{2.538x_1} + 0.218e^{3.933x_4} + 0.019$	0.917	0.162	x_2 is $NDWI_{1610}$,
g	$y = 0.402e^{2.538x_1} - 0.26e^{3.949x_2} - 0.439e^{2.635x_3} - 0.034$	0.886	0.189	x_3 is $NDWI_{2190}$, and
h	$y = 0.199e^{2.538x_1} - 0.234e^{3.949x_2} + 0.306e^{3.933x_4} + 0.048$	0.938	0.14	x_4 is NDRI
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	$\mathbf{y = 0.261e^{2.538x_1} + 0.127e^{3.949x_2} - 0.604e^{2.635x_3} + 0.428e^{3.933x_4} + 0.092}$	0.965	0.105	

4.2. SSM retrieval results of the modified WCM

As stated in Section 4.1, the values of VWC from Sentinel-2 vegetation index were brought into the formula of the WCM to obtain the soil backscattering coefficient. The corresponding backscatter coefficients of the 84 samples in their image according to their coordinates, their relationships between before and after removing the vegetation influence of VV, and the VH polarization backscattering coefficient are shown in Fig. 5. The backscattering coefficients of VV polarization (Fig. 5a) were higher than those of VH polarization (Fig. 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 average variation of the VV polarization backscattering coefficient was -2.35 dB, whereas for the VH polarization it was -2.92 dB. The variation of VH polarization backscattering before and after correction was greater than that of VV polarization. This shows that VH polarization was more easily affected by the vegetation layer during the transmission process.

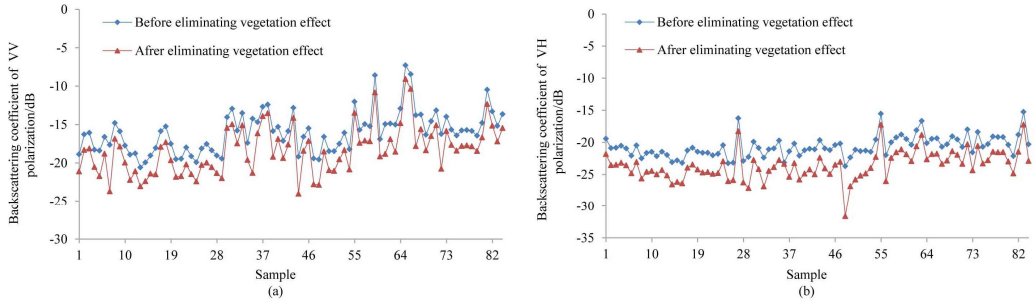


Figure 5. Relationship between VV and VH polarization backscattering coefficients: (a) VV polarization and (b) VH polarization.

The relationships between the estimated SSM values and the in-situ measurements are shown in Fig.6, and the R^2 and RMSE values are summarized in Table 3.

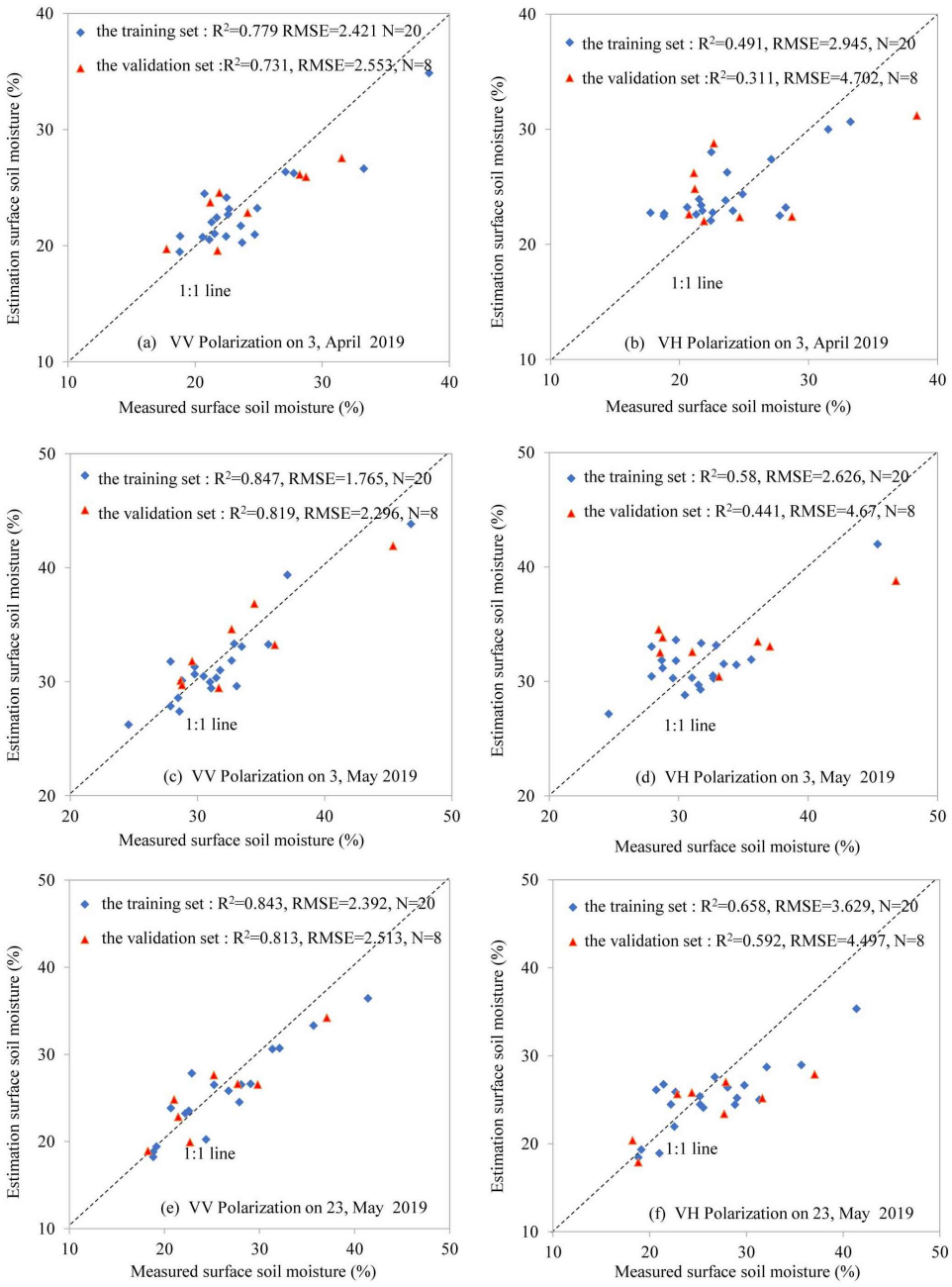


Figure 6. Comparisons between measured SM values and those retrieved using SVR for different stages of wheat: (a) VV polarization on 3 April, 2019; (b) VH polarization on 3 April, 2019; (c) VV polarization on 3 May, 2019; (d) VH polarization on 3 May, 2019; (e) VV polarization on 23 May, 2019; and (f) VH polarization on 23 May, 2019.

Table 3. Summary of training and validation results of the SVR for each dataset.

Data set	Training results				Validation results			
	VV		VH		VV		VH	
	R^2	RMSE (%)	R^2	RMSE (%)	R^2	RMSE (%)	R^2	RMSE (%)
3 April, 2019 (jointing stage)	0.779	2.421	0.491	2.945	0.731	2.553	0.311	4.702
3 May, 2019 (heading stage)	0.847	1.765	0.58	2.626	0.819	2.296	0.441	4.67
23 May, 2019 (infilling stage)	0.843	2.392	0.658	3.629	0.813	2.513	0.592	4.497

As seen in Fig. 6, the scattered points of the VV polarization were closer to the 1:1 line than that of VH polarization. However, when the SSM was relatively wet (more than 35%), the inversion error was relatively large.

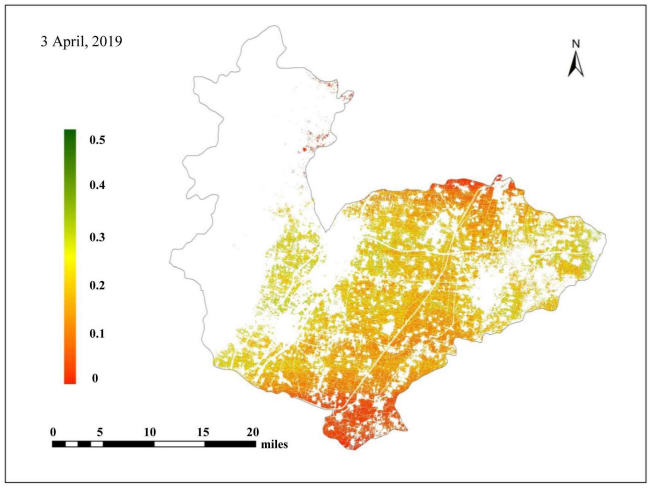
During the jointing stage of wheat (3 April, 2019), the R^2 and RMSE values of SSM based on the SVR were 0.779 and 2.421 for the training phase at VV polarization, respectively, while they were 0.731 and 2.553 respectively for the validation phase (Fig. 6a). The R^2 and RMSE values of the VH polarization were (0.491, 2.945) and (0.311, 4.702) for the training and the validation phases, respectively (Fig. 6b).

During the heading stage of wheat (3 May, 2019), the R^2 and RMSE values of the VV polarization were 0.847 and 1.765 for the training phase, and were 0.819 and 2.296 for the validation phase (Fig. 6c) respectively; the R^2 and RMSE values of the VH polarization were 0.58 and 2.626 respectively for the training phase, and were 0.441 and 4.67 respectively for the validation phase (Fig. 6d).

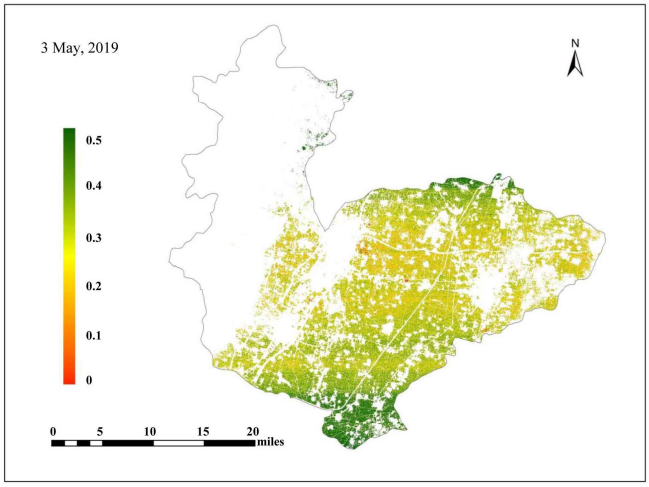
During the filling stage of wheat (23 May, 2019), the R^2 and RMSE values of SSM based on the SVR at VV polarization were 0.843 and 2.392 respectively for the training phase; they were 0.813 and 2.513 respectively for the validation phase (Fig. 6e). The R^2 and RMSE values of the VH polarization were (0.658, 3.629) and (0.592, 4.497) for the training and validation phases, respectively (Fig. 6f).

According to the training and validation results, VV polarization has good accuracy and stability for retrieval SSM in the study area, and the modified WCM presented satisfactory results in retrieval SSM with Sentinel-1 and Sentinel-2 data. Therefore, using the generated model, the SSM of all wheat fields in the study area was retrieved with three different periods.

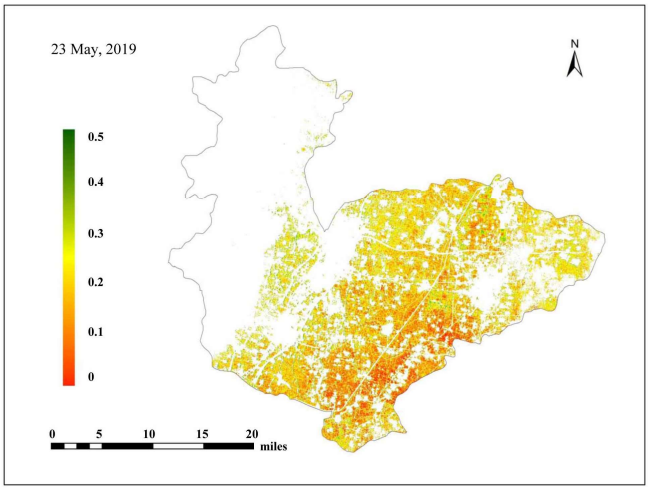
The spatial distribution of SSM retrievals and frequency diagram of SSM in the study area are shown in Fig. 7. Based on the supervised classification technology of threshold segmentation by the environment for visualizing images (ENVI) software, the non-wheat areas, such as towns, rivers and other non-agricultural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



(a)



(c)



(e)

Figure 7. Temporal SSM mapping for wheat-covered fields of study area (left), and the frequency of SSM (right).

As shown in Fig. 7, the temporal variations of retrieved SSM in the study areas were found to be in the range of 10 – 50 %. The results of SSM retrieval on 3 April were slightly

drought-related (Fig. 7(a)). The frequency of the SSM retrieval value was highest in the range of 20–30%. This was because the weather was clear and there was no rainfall during the previous period. The results of SSM retrieval on 3 May were relatively high, which is due to the fact that there was continuous rainfall, so the SSM was moist (Fig. 7(b)). The frequency of the SSM retrieval value was highest in the range of 30 – 40 %. The results of SSM retrieval on 20 May were drier than wetter than Figure 7(a), but drier than Figure 7(b), and the frequency of the SSM retrieval value was highest in the range of 20 – 40 % (Fig. 7(c)). According to the meteorological data, there was only a small amount of rainfall in the experimental area one week before the satellites transit. But for the high temperature, the values of SSM were still not high. This may explain why the SSM was slightly lower on this day than on 3 May.

The frequency distributions of the soil moisture retrieval results of these three times were basically consistent with the measured values on each day. The model therefore had strong applicability for the study area.

4.3. Discussion

As a classical model, WCM has been widely used to retrieve soil moisture information from areas with vegetation cover. In recent years, some new studies have used Sentinel-1 and Sentinel-2 data to estimate SSM [29, 36, 47]. Guo *et al.* [47] investigated the estimation of farmland SSM using multi-source remote sensing data from Sentinel-1 radar and Sentinel-2 optical satellites. They applied the Oh model [48], SVR, and the generalized regression neural network (GRNN) to retrieve SSM. They used the WCM to remove the influence of vegetation. The inputs of SVR include 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)). These results have shown that combining multi-characteristic parameters based on SVR delivered the best retrieval accuracy with the R^2 of 0.903 and RMSE of 0.014 cm^3/cm^3, respectively. Zhao *et al.* [49] estimated the SSM of winter wheat fields using Sentinel-1 and Sentinel-2 data. Based on near-infrared, red, and short-wave 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 a RMSE value of 0.0209 cm^3/cm^3 in VV/VH; their R^2 and RMSE values were (0.6791, 0.0249) for VV polarization and (0.5151, 0.0289) for VH polarization, respectively. In the current 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 band of sentinel-2. Baghdadi *et al.* [36] estimated the SSM of crop fields and grasslands from Sentinel-1/2 data. They combined the WCM combined with the IEM using real data composed of a C-band radar backscattered signal, NDVI, soil moisture, and surface roughness values. Their results showed that the soil contribution to the total radar backscattered signal was lower in VH polarization than in VV polarization. Zeng *et al.* [30] studied SSM under different vegetation covers based on Sentinel-1A and SVR techniques; they concluded that VV polarization could achieve high retrieval accuracy. Wang *et al.* [50] combined full polarization Radarsat-2 SAR data and SVR techniques to estimate soil moisture in vegetation-sparse arid areas. They determined that the inversion accuracy of the co-polarization data (VV polarization or HH polarization) was higher than that of the cross-polarization data (VH polarization or HV polarization). Comparing the inversion results reported here with those of the above studies, it is possible to consistently conclude that VV polarization is more sensitive to SSM than VH.

One of the limitations of this study is that it only focused on wheat fields. Furthermore, only a small number of sample points were measured. Furthermore, as wheat is a drought-resistant crop, it is also planted in the hills and mountains of China. This study

only examined wheat in plain areas, however. Soil moisture retrieval methods that are based on more kinds of crops, and that encompass many terrain conditions, may have more practical significance.

5. Conclusions

Taking Hebi, a representative area of wheat planting in Henan Province, as a study area, here the potential of synergy between C-band Sentinel-1 SAR and Sentinel-2 optical data for SSM retrieval was investigated regarding wheat fields. To extract the soil backscattering coefficient (σ^0_{soil}) from the Sentinel-1 SAR data, first 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 wheat-covered fields in different periods of growth 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 performed using experimental dataset of vegetation fraction and VWC values calculated from optical images. Vegetation fraction was used to distinguish the proportion 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 vegetation canopy on the backscattering coefficient of Sentinel-2 radar data.

(2) Compared with Sentinel-1 VH polarization data (after removal of vegetation influence using the WCM), VV polarization data resulted in a higher estimation accuracy regarding the SSM retrieval. This result shows that the VV polarization has more soil backscattering information and is more sensitive to the change of SSM than VH polarization.

(3) In this study, C-band Sentinel-1 SAR and Sentinel-2 optical data have been used to study the wheat-covered fields. Results obtained show that the estimated SSM based on these two kinds of satellite data are applicable to agricultural environments for wheat. To further our research, we intend to examine the proposed algorithm on other crops and different regional studies. Furthermore, to estimate SSM more accurately on a large scale, some advanced algorithms, such as deep learning, would be implemented.

Author Contributions: Y. L. conceived the idea, implemented the algorithm, interpreted results, and data analysis. C. Z. discussed the details of the algorithm, contributed to the manuscript's preparation, and revised it. W. H. designed experiments, performed field experiments and processed data, F. Y. and J. L. contributed analysis tools.

Funding: This research was supported by The National Natural Science Foundation of China (Grant 51739009), Key Projects of Water Conservancy Science and Technology by Henan Province (Grant No. GG201902), Major Projects of Science and Technology in Henan Province (NO. 181100310300), Open Research fund of Key Open Laboratory of Agro Meteorological Support and Application Technology in Henan Province (Grant No. AMF201807) and Training plan for Young Core Teachers in Colleges of Henan Province (Grant No. 2019gzgg073).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank ESA for providing free access to the Sentinel-1 SAR and Sentinel-2 satellite data used in the present study.

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

References

1. D. G. Miralles et al., "Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation," *Nat. Geosci.* **7**, 345-349 (2014).
2. H. Vereecken et al., "Soil hydrology: Recent methodological advances, challenges, and perspectives," *Water Resour. Res.* **51**, 2616–2633 (2015).
3. E. Cho et al., "Regional scale spatio-temporal variability of soil moisture and its relationship with meteorological factors over the Korean peninsula," *J. Hydrol.* **516**, 317-329 (2014).
4. X. Bai et al., "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.* **12**(15), 2508 (2020).
5. Mikunthan T et al., "Estimation of Groundwater Recharge in Limestone Aquifer using an Improved Soil Moisture Balance Method: A Case Study in Jaffna District," *Trop. Agric. Res.* **21**(1), 51–61 (2009).
6. J. Wang et al., "Microwave modeling of soil moisture in Oasis regional scale based on Sentinel-1 radar images," *J. Infrared Millimeter Waves* **36**(1), 120-126 (2017).
7. L. Brocca et al., "Catchment scale soil moisture spatial–temporal variability," *J. Hydrol.* **422**, 63–75 (2012).
8. G. Bertoldi et al., "Estimation of soil moisture patterns in mountain grasslands by means of SAR Radarsat2 images and hydrological modeling," *J. Hydrol.* **516**(4), 245-257 (2014).
9. S. Paloscia et al., "Soil moisture mapping using sentinel-1 images: algorithm and preliminary validation," *Remote Sens. Environ.* **134**, 234-248 (2013).
10. J. Dong et al., "Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation," *Remote Sens. Environ.* **242**, 111756 (2020).
11. M. Aubert et al., "Toward an operational bare soil moisture mapping using TerraSAR-X data acquired over agricultural areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(2), 900-916 (2013).
12. B. W. Barret et al., "Soil moisture retrieval from active space borne microwave observations: an evaluation of current techniques," *Remote Sens.* **1**(3), 210–242 (2009).
13. P. Zhou et al., "Retrieval methods of soil water content in vegetation covering areas based on multi-source remote sensing data," *J. Remote Sens.* **14**(5), 959-973(2010).
14. D. D. Alexakis et al., "Soil moisture content estimation based on sentinel-1 and auxiliary earth observation products. A hydrological approach," *Sensors.* **17**(6),1455–1461(2017).
15. Y. Oh et al., "An empirical model and an inversion technique for radar scattering from bare soil surfaces," *IEEE Trans. Geosci. Remote Sens.* **30**(2), 370- 381 (1992).
16. M. R. Sahebi et al., "Estimation of the moisture content of bare soil from RADARSAT-1 SAR using simple empirical models," *Int. J. Remote Sens.* **24**(12),2575-2582(2003).
17. E. P. W. Attema et al., "Vegetation modeled as a water cloud," *Radio Sci.* **13**(2), 357-364(1978).
18. F. Ulaby et al., "Michigan microwave canopy scattering model," *IEEE Trans. Geosci. Remote Sens.* **11**(7), 477-491(1990).
19. M. Zribi et al., "Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation," *Hydrol. Earth Syst. Sci.* **15**, 345–358(2011).
20. K. Kumar et al., "Study of water cloud model vegetation descriptors in estimating soil moisture in Solani catchment," *Hydrol. Process.* **29**(9), 2137-2148 (2015).
21. X. Chai et al., "Modeling and mapping soil moisture of plateau pasture using Radarsat-2 imagery," *Remote Sens.* **7**(2), 1279-1299 (2015).
22. K. S. Chen et al., "A simple-model for retrieving bare soil-moisture from radar-scattering coefficients," *Remote Sens. Environ.* 54, 121–126 (1995).
23. P. C. Dubois et al., "Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sens.* 33, 915–926 (1995).
24. R. Bindlish et al., "Parameterization of vegetation backscatter in radar-based, soil moisture estimation," *Remote Sens. Environ.* **76**(1), 130-137 (2001).
25. L. N. Lin et al., "Soil moisture retrieval over vegetated areas based on Sentinel-1 and FY-3C data," *Remote Sens. Technol. Appl.* **33**(4), 750-758 (2018).
26. J. B. Jiang et al., "Research of soil moisture retrieval model of wheat covered surface based on MIMICS model," *J. Triticeae Crops* **35**(5), 707-713 (2015).
27. B. He et al., "A synergistic methodology for soil moisture estimation in an alpine prairie using radar and optical satellite data," *Remote Sens.* **6**(11), 10966–10985 (2014).
28. L. Bruzzone et al., "Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data," *IEEE Trans. Geosci. Remote Sens.* **43**(1), 159-174 (2005).
29. A. Reza et al., "Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale," *Remote Sens.* **10**(8), 1285 (2018).
30. X. J. Zeng et al., "Soil water content retrieval based on Sentinel-1A and Landsat 8 image for Bei'an-Heihe expressway," *J. Eco-Agric.* **25**(1), 118–126 (2017).
31. Y. S. Bao et al., "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. and Geo.* **72**, 76-85 (2018).

32. L. Pasolli et al., "Bruzzone L . Estimating soil moisture with the support vector regression technique," *IEEE Geo. Remote Sens. Letters* **8**(6), 1080-1084 (2011).
33. S. Ahmad et al., "Estimating soil moisture using remote sensing data: A machine learning approach," *Advances in Water Resour.* **33**(1), 69-80 (2010).
34. J. S. Lee et al., "Speckle filtering of synthetic aperture radar images: a review," *Remote Sens. Reviews* **8**(4), 313–340 (1994).
35. M. Pesaresi et al., "Assessment of the added-value of Sentinel-2 for detecting built-up areas," *Remote Sens.* **8**(4), 299 (2016).
36. N. Baghdadi, et al. "Calibration of the water cloud model at C-Band for winter crop fields and grasslands," *Remote Sens.* **9**,(969),13(2018).
37. Yadav, V. P et al., "Estimation of soil moisture through water cloud model using sentinel-1A SAR data," *IEEE Int. Geo. Remote Sens. Symposium* (2019).
38. G. Gutman et al., "The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models," *Int. J. Remote Sens.* **19** (8), 1533–1543. (1998).
39. G. J. Yang et al., "Estimation of soil moisture in farmland using improved water cloud model and Radarsat-2 data," *Trans. Soci. Agri. Enger.* **32**(22), 146-153 (2016).
40. C. Z. Wang et al., "Soil moisture estimation in a semiarid rangeland using ERS-2 and TM imagery," *Remote Sens. Environ.* **90**(2), 178-189 (2004).
41. D. W. Deering. "Rangeland reflectance characteristics measured by aircraft and spacecraft sensors," College Station, TX: Texas A&M University, 1978.
42. B. C. Gao. NDWI—A normalized difference water index for remote sensing of vegetation quid water from space," *Remote Sens. Environ.* **58**(3), 257-266 (1996).
43. H. O. Kim et al., "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.* **35**(19), 7046-7068 (2014).
44. C. Y. Fang et al., "A comparative study of different red edge indices for remote sensing detection of urban grassland health status," *J. Geo-inform. Sci.* **19**(10), 1382-1392 (2017).
45. M. Zribi et al., "Derivation of wild vegetation cover density in semi-arid regions: ERS2/SAR evaluation," *Int. J. Remote Sens.* **24**(6), 1335-1352 (2003).
46. J.Wang, et al. "Multiple-Instance Learning via an RBF Kernel-Based Extreme Learning Machine," *J. Intell. Syst.* 26,1(2016).
47. J. Guo et al., "Construction and validation of soil moisture retrieval model in farmland based on Sentinel multi-source data," *Transactions of the CSAE.* 35(14), 71-78(2019).
48. Oh Y. "Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces," *IEEE Trans. Geo. Remote Sens.*42(3), 596-601(2004).
49. J. H. Zhao et al., "Cooperative inversion of winter wheat covered surface soil moisture based on Sentinel-1/2 remote sensing data," *J. Electron. Inform. Technol.43,* (2021).
50. Y. T. Wang., "Remote sensing inversion of soil moisture in vegetation-sparse arid areas based on SVR," *J. Geo-inform. Sci.* 21(8):1275-1283 (2019).