remote sensing

MDPI

Article

Estimation of Soil Moisture Using Multi-Source Remote Sensing and Machine Learning Algorithms in Farming Land of Northern China

Quanshan Liu [1], Zongjun Wu [1], Ningbo Cui [1,*], Xiuliang Jin [2], Shidan Zhu [1], Shouzheng Jiang [1], Lu Zhao [1] and Daozhi Gong [3]

[1] State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China
[2] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences/Key Laboratory of Crop Physiology and Ecology, Ministry of Agriculture, Beijing 100081, China
[3] Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing 100081, China
* Correspondence: cuiningbo@scu.edu.cn

Abstract: Soil moisture is a key parameter for the circulation of water and energy exchange between surface and the atmosphere, playing an important role in hydrology, agriculture, and meteorology. Traditional methods for monitoring soil moisture suffer from spatial discontinuity, time-consuming processes, and high costs. Remote sensing technology enables the non-destructive and efficient retrieval of land information, allowing rapid soil moisture monitoring to schedule crop irrigation and evaluate the irrigation efficiency. Satellite data with different resolutions provide different observation scales. Evaluating the accuracy of estimating soil moisture based on open and free satellite data, as well as exploring the comprehensiveness and adaptability of different satellites for soil moisture temporal and spatial observations, are important research contents of current soil moisture monitoring. The study utilized three types of satellite data, namely GF-1, Landsat-8, and GF-4, with respective temporal and spatial resolutions of 16 m (every 4 days), 30 m (every 16 days), and 50 m (daily). The gray relational analysis (GRA) was employed to identify vegetation indices that selected sensitivity to soil moisture at varying depths (3 cm, 10 cm, and 20 cm). Then, this study employed random forest (RF), Extra Tree (ETr), and linear regression (LR) algorithms to estimate soil moisture at different depths with optical satellite data sources. The results showed that the accuracy of soil moisture estimation was different at different growth stages. The model accuracy exhibited an upward trend during the middle and late growth stages, coinciding with higher vegetation coverage; however, it demonstrated a decline in accuracy during the early and late growth stages due to either the absence or limited presence of vegetation. Among the three satellite images, the vegetation indices derived from GF-1 exhibited were more sensitive to vegetation characteristics and demonstrated superior soil moisture estimation accuracy (with R^2 ranging 0.129–0.928, RMSE ranging 0.017–0.078), followed by Landsat-8 (with R^2 ranging 0.117–0.862, RMSE ranging 0.017–0.088). The soil moisture estimation accuracy of GF-4 was the worst (with R^2 ranging 0.070–0.921, RMSE ranging 0.020–0.140). Thus, GF-1 is suitable for vegetated areas. In addition, the ETr model outperformed the other models in both accuracy and stability (ETr model: R^2 ranging from 0.117 to 0.928, RMSE ranging from 0.021 to 0.091; RF model: R^2 ranging from 0.225 to 0.926, RMSE ranging from 0.019 to 0.085; LR model: R^2 ranging from 0.048 to 0.733, RMSE ranging from 0.030 to 0.144). Utilizing GF-1 is recommended to construct the ETr model for assessing soil moisture variations in the farming land of northern China. Therefore, in cases where there are limited ground sample data, it is advisable to utilize high-spatiotemporal-resolution remote sensing data, along with machine learning algorithms such as ETr and RF, which are suitable for small samples, for soil moisture estimation.

Keywords: soil moisture; multisource remote sensing; machine learning; farming land

Citation: Liu, Q.; Wu, Z.; Cui, N.; Jin, X.; Zhu, S.; Jiang, S.; Zhao, L.; Gong, D. Estimation of Soil Moisture Using Multi-Source Remote Sensing and Machine Learning Algorithms in Farming Land of Northern China. *Remote Sens.* **2023**, *15*, 4214. https://doi.org/10.3390/rs15174214

Academic Editor: Dusan Gleich

Received: 25 July 2023
Revised: 17 August 2023
Accepted: 23 August 2023
Published: 27 August 2023

1. Introduction

Soil moisture is an essential parameter in the fields of hydrology, meteorology, and agricultural science. It plays a crucial role in controlling surface evapotranspiration and vegetation photosynthesis, and serves as a link between the groundwater, surface water, and atmospheric water in the hydrological cycle [1,2]. It has significant implications in global water cycle, energy balance, and climate change research. Therefore, the timely and accurate monitoring of soil moisture is of great importance for agriculture, the water cycle, and related aspects. The real-time monitoring of soil moisture at large regional scales is an essential indicator in modern agriculture. Comprehensive soil moisture monitoring plays a crucial and guiding role in crop yield, drought assessment, and precise irrigation decision making [3].

The measurement of soil moisture poses challenges due to its pronounced spatiotemporal variability [4]. Conventional techniques employed for soil moisture monitoring, such as the gravimetric method, time domain reflectometry, and neutron probe, are limited to capturing singular or localized measurements. The methods, as a result, demonstrate spatial discontinuity, necessitate time-consuming procedures, entail substantial expenses, and are susceptible to the influence of external conditions on measurement accuracy [5]. Remote sensing technology encompasses real-time monitoring, extensive coverage, and cost-effectiveness, endowing it with indispensable utility in the domain of soil moisture assessment. Currently, soil moisture retrieval using remote sensing methods can be broadly classified into two categories based on the data sources employed: microwave remote sensing and optical remote sensing. Microwave remote sensing employs microwave equipment to detect and capture the electromagnetic radiation and scattering characteristics of a target object within the microwave frequency range, facilitating the identification of distant objects, and possesses the capability to penetrate through vegetation and effectively detect subsurface targets. Despite its independence from cloud cover, the limited spatial and temporal resolution of microwave remote sensing poses challenges for its widespread application in precision agriculture. Optical remote sensing exhibits high spatial and temporal resolution, and can detect the differences in absorption characteristics between soil and water in the visible and near-infrared bands, and uses this difference to monitor soil moisture [6], which presents significant advantages in the domains of agricultural irrigation, crop growth monitoring, and yield prediction. But optical remote sensing is susceptible to weather.

In a recent investigation conducted by Fan et al. [7], the scholars employed the triple collocation analysis approach, with GLDAS data as the reference, to discern the accuracy of soil moisture products derived from microwave remote sensing datasets, namely SMOS, SMAP, AMSR2, and FY-3C. Yao et al. [8] employed the Optical Gradient Model (OPTRAM) for soil moisture retrieval and generated irrigation district maps for the arid and semi-arid regions in northwest China spanning the past 30 years. Microwave remote sensing presents notable advantages, including high temporal resolution and broad swath coverage. Nevertheless, its utility is impeded by its coarse spatial resolution, exemplified by AMSR-E with a resolution of 5 km and SMOS with a resolution of 50 km [9]. Active microwave sensors, such as Synthetic Aperture Radar (SAR), exhibit spatial resolutions ranging from 10 to 100 m, but they have a relatively low spatial resolution [10]. Optical remote sensing overcomes the limitations of traditional soil moisture monitoring in terms of low efficiency and high cost. It also provides higher spatiotemporal resolution, and thus it is increasingly applied in regional-scale soil moisture monitoring [11–13]. Nie et al. [6] calculated PDI, MPDI, and VAPDI using GF-1 and Landsat-8 satellites to construct a soil moisture inversion model for the 0–30 cm depth. The results showed that NIR-R spectral indices exhibited greater sensitivity to surface soil moisture. Liu et al. [13] utilized random forest (RF) regression to optimize the sensitivity of Sentinel-2 spectral bands to soil moisture at the 0–5 cm depth and constructed SMMI and PDI in different feature spaces. The outcomes demonstrated that Sentinel-2 effectively estimated surface soil moisture at the 0–5 cm depth in complex agricultural environments.

The soil moisture content within the root zone (0–15 cm) exerts a profound influence on crop physiological states. Theoretically, it can be indirectly assessed through the analysis of crop canopy reflectance or vegetation indices [14,15]. The relationship between soil moisture content and vegetation spectra is intricate, posing challenges for classical regression methods in achieving unbiased and effective parameter estimation [16,17]. The machine learning method has significant advantages in addressing complex relationships such as nonlinearity and heteroscedasticity, and it has been widely applied in remote sensing. This method has emerged as a pivotal research focus within agricultural remote sensing, showcasing remarkable performance in modeling and inversion studies [18–21]. Cheng et al. [22] explored the performance of soil moisture inversion at different soil depths and vegetation cover levels based on Landsat-8 satellite data and the RF algorithm. The results showed that soil moisture inversion reaped optimal outcomes specifically beneath low-density grassland cover. Adab et al. [23] used Landsat-8 data to establish SVM, ANN, EN, and RF soil water inversion models on land utilization categories in semi-arid Iran, and the results indicated that the RF model had the highest accuracy in estimating soil moisture at a depth of 5 cm. Owing to its superlative predictive and classificatory prowess, the RF model has been widely adopted for unraveling non-linear associations across an array of disciplines. Machine learning has been effectively employed for soil moisture prediction based on remote sensing data. However, there are still some special cases of interest in terms of using machine learning in remotely sensed soil moisture estimation. The accuracy of soil moisture estimation by machine learning is influenced by the size of the sample and the optimization of machine learning hyperparameters. In addition, evaluating the accuracy of estimating soil moisture using open and freely accessible satellite data, as well as exploring the comprehensiveness and adaptability of different satellites for temporal and spatial observations, is an important research focus in current soil moisture monitoring. Currently, there is an increasing availability of free high temporal-spatial resolution satellite data. However, limited research has been conducted on the utilization of diverse satellite data for constructing soil moisture models, evaluating model estimation accuracy, and comprehensively assessing the spatiotemporal coverage and applicability of satellites in soil moisture assessment. The accurate and rapid estimation of regional soil moisture enables the elucidation of spatial and temporal heterogeneity, thereby enhancing our understanding of the regional water cycle process. It is also serves as a valuable reference for implementing agricultural management strategies, making informed decisions regarding agricultural irrigation, and effectively managing regional water resources.

Multispectral satellites have high spatiotemporal resolution. There have been few studies on harnessing the potential of these satellites to cultivate machine-learning-driven soil moisture inversion models. In this study, a variety of vegetation indices were calculated based on the reflectance of different remote sensing satellites (GF-1, Landsat-8 and GF-4). Then we employed Gray Correlation Analysis (GRA) to select sensitive vegetation indices for soil moisture at different depths in the study area. And, we utilized the RF, ETr, and linear regression (LR) machine learning algorithms to construct soil moisture estimation models at different depths. The objective of this study was to analyze the adaptability of freely available satellite images for soil moisture estimation, evaluate the accuracy performance of different machine learning algorithms in estimating soil moisture, investigate the temporal and spatial distribution patterns of soil moisture in Shandian River basin, and validate the potential application of domestic remote sensing satellite data in precision agriculture.

2. Materials and Methods

2.1. Study Area

The Shandian River Basin is located in northern China, originates from the northern foothills of the Bayan-Gol-Tu Mountains in Hebei Province, and is situated within the Inner Mongolia Autonomous Region (115.5°E–116.5°E, 41.5°N–42.5°N). Serving as a vital water source conservation area in the Beijing-Tianjin-Hebei region, it exhibits a pronounced

agricultural-pastoral-forestry ecotone. The study area is relatively flat with an altitude of 1300~1400 m, belonging to the temperate continental climate. The average annual precipitation in most areas is 300~500 mm, and the rainfall is mostly concentrated from July to September. Referred to as a typical seasonal frost region, this area is characterized by arid and cold winters. The dominant land cover types include croplands, grasslands, and forests, with minor extents of shrubs and bare land [24]. The study area was depicted in Figure 1.

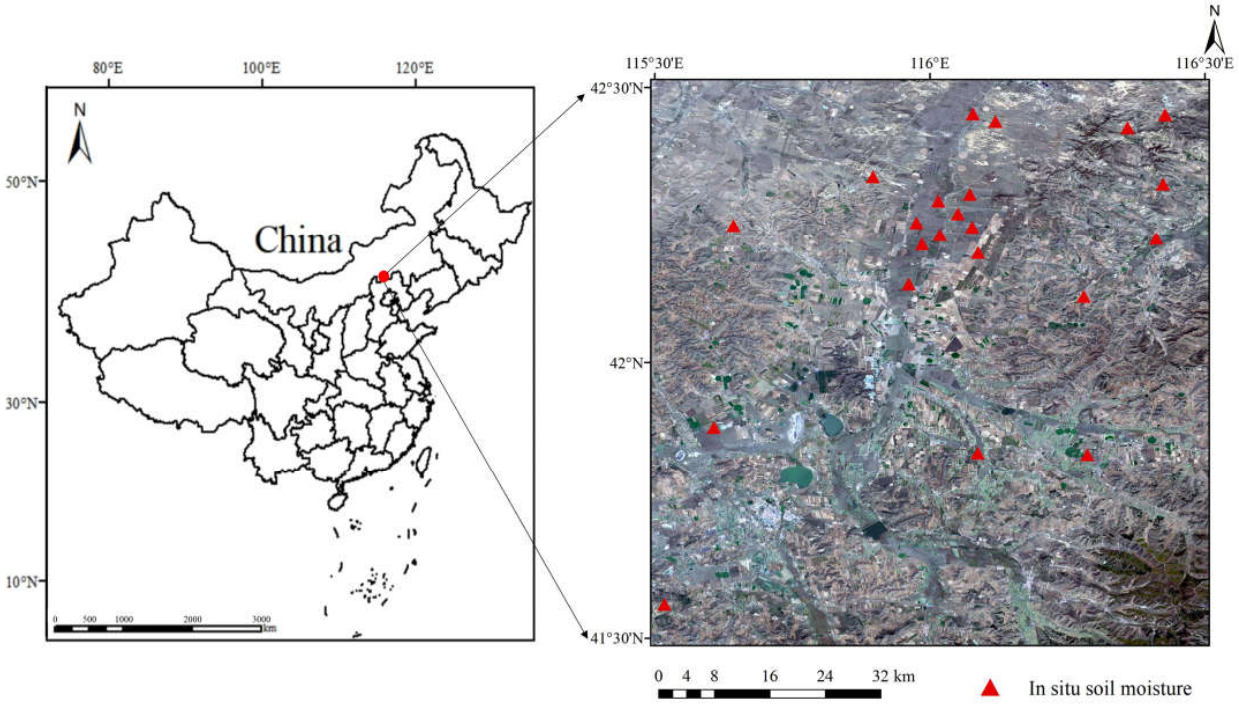

Figure 1. Shandian river basin and ground measured sites.

2.2. Data Sources and Processing

2.2.1. Ground Data

This study utilized the soil temperature and moisture wireless sensor network observation dataset of the Shandian River Basin (2019) [25]. The dataset includes in situ measurements of soil moisture, soil temperature, and precipitation from 34 stations in the Shandian River Basin. The monitoring network encompasses an extensive area of approximately 10,000 km^2 (115.5°E–116.5°E, 41.5°N–42.5°N) and was strategically deployed within the basin using an optimized deployment approach. The soil volumetric moisture content and soil temperature at five measured depths (3 cm, 5 cm, 10 cm, 20 cm, and 50 cm) were continuously monitored at each site using Decagon 5TM soil moisture sensors (Austria, Pessl Instruments, Weiz, Austria). The HOBO rain gauge was used to monitor rainfall. The Shandian River Basin, where the sensors were deployed, is characterized by a predominantly flat topography and is primarily covered by grasslands and croplands. Once the measurement data stabilized, soil samples were collected periodically for each soil layer at each station. These samples were analyzed for parameters such as gravimetric/volumetric water content, bulk density, and soil texture to calibrate the raw measurement data. In this study, the soil moisture data from the surface layer at depths of 3 cm, 10 cm, and 20 cm were selected as the ground truth measurements.

2.2.2. Satellite Data and Preprocessing

GF-1 is the first high-resolution satellite for Earth observation in China, equipped with a panchromatic/multispectral camera (PMS sensor) and a multispectral camera (WFV sensor). It offers a revisit cycle of 4 days, capturing multispectral images with 4 bands at a spatial resolution of 16 m [26]. Landsat-8, the eighth installment of the renowned

Landsat program by the United States, carries an Operational Land Imager (OLI) and a Thermal Infrared Sensor (TIRS). Its revisit cycle spans 16 days, with OLI data consisting of 8 bands at a spatial resolution of 15 m for panchromatic images and 30 m for multispectral images [27]. GF-4, a Chinese geostationary orbit remote sensing satellite, employs a push-broom imaging technique, providing the advantages of high temporal resolution and relatively higher spatial resolution. It carries a multispectral camera and a thermal infrared camera. The PMS multispectral data of GF-4 consists of 5 bands with a spatial resolution of 50 m [28]. These satellite datasets can be freely accessed and downloaded (GF-1 and GF-4 data download address: https://data.cresda.cn, accessed on 30, June 2022, Landsat-8 data download address: http://glovis.usgs.gov/, accessed on 22, July 2022). Table 1 presents the band information for each satellite sensor. To prepare the satellite imagery for analysis, several preprocessing steps are necessary, including radiometric calibration, atmospheric correction, orthorectification, and image registration, as illustrated in Figure 2.

Table 1. Satellite data used in this study.

Satellite	GF-1 WFV	Landsat-8 OLI	GF-4 PMS
Spectral range (μm)	0.450–0.520 0.520–0.590 (b2) 0.630–0.690 (b3) 0.770–0.890 (b4)	0.433–0.453 (b1) 0.450–0.515 (b2) 0.525–0.600 (b3) 0.630–0.680 (b4) 0.845–0.885 (b5) 1.560–1.660 (b6) 2.100–2.300 (b7) 0.500–0.680 (b8) 1.360–1.390 (b9)	0.450–0.900 (b1) 0.450–0.520 (b2) 0.520–0.600 (b3) 0.630–0.690 (b4) 0.760–0.900 (b5)
Spatial resolution (m)	16	30	50
Width (km)	60	185	400
Revisit time (d)	4	16	1

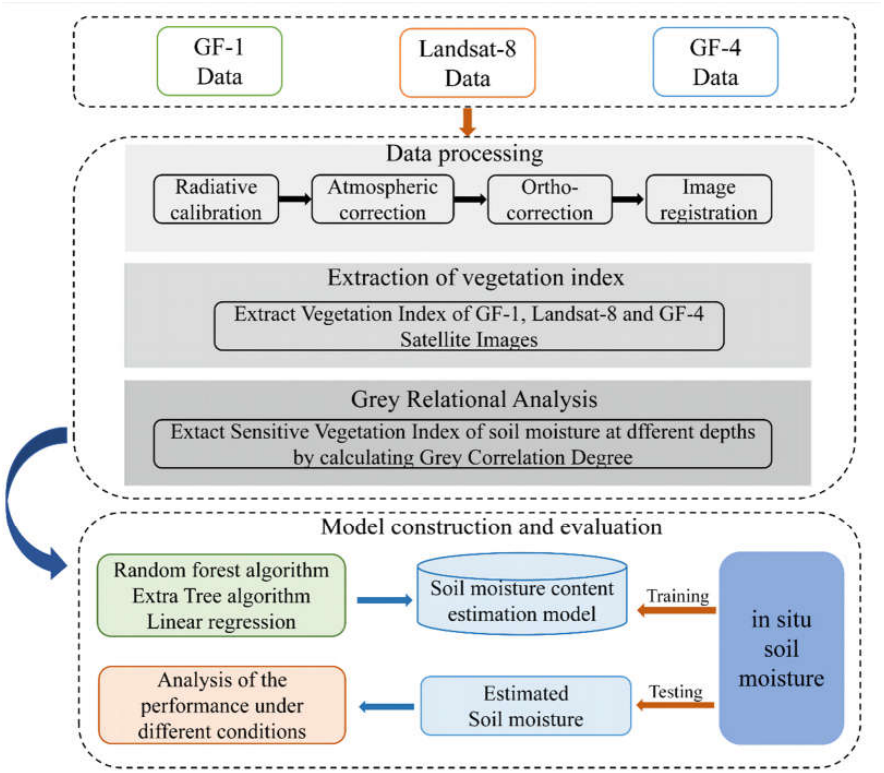

Figure 2. Flowchart of the soil moisture retrieval based on GF-1, Landsat-8, and GF-4.

This study selected cloud-free remote sensing images from GF-1, Landsat-8, and GF-4 satellites. Four remote sensing images were selected for each satellite according to the 4 growth stages of the plants and the underlying vegetation (Table 2) [29]. The study area primarily comprises agricultural land and grassland, encompassing crops such as carrots and potatoes. Due to the differences in phenology among different kinds of vegetation, the growth cycle of vegetation can be categorized into four stages: the early growth stage indicates the seedling phase, middle growth stage signifies rapid growth, late growth stage represents maturity, and end growth stage denotes completion.

Table 2. Image selection of GF-1, Landsat-8, and GF-4 satellite.

Growth Stage	Satellite	Date
Early	GF-1	14 April 2019
	Landsat-8	2 April 2019
	GF-4	14 April 2019
Middle	GF-1	14 June 2019
	Landsat-8	24 August 2019
	GF-4	24 August 2019
Late	GF-1	29 September 2019
	Landsat-8	25 September 2019
	GF-4	28 September 2019
End of growth	GF-1	31 October 2019
	Landsat-8	27 October 2019
	GF-4	30 October 2019

2.2.3. Remote Sensing Vegetation Indices

This study calculated soil moisture spectral indices at different growth stages and depths, eliminating the influence of vegetation growth, based on a total of 18 different bands from GF-1, Landsat-8, and GF-4. Representative spectral indices (20 in total) that are correlated with soil moisture were selected. The specific calculation formulas could be found in Table 3.

Table 3. Vegetation index summary.

Vegetation Index	Abbreviation	Formula	References
Comprehensive spectral response index	COSRI	$(B + G) \times (NIR - R)/(R + NIR)^2$	[30]
Differential vegetation index	DVI	$NIR - R$	[31]
Enhanced vegetation index	EVI	$2.5(NIR - R)/(1 + NIR + 6R - 7.5B)$	[32]
Green DVI	GDVI	$NIR - G$	[33]
Green leaf index	GLI	$(2G - B - R)/(2G + B + R)$	[34]
Green NDVI	GNDVI	$(NIR - G)/(NIR + G)$	[35]
Green OSAVI	GOSAVI	$(NIR - G)/(NIR + G + 0.16)$	[36]
Green RVI	GRVI	NIR/G	[37]
Green SAVI	GSAVI	$(NIR - G)/(NIR + G + 0.5)$	[38]
Infrared vegetation index	IPVI	$NIR/(NIR + R)$	[39]
Modified soil-adjusted vegetation index	MSAVI2	$NIR + 0.5 - 0.5 \times [(2NIR + 1)^2 - 8(NIR - R)]^{0.5}$	[40]
Normalized difference vegetation index	NDVI	$(NIR - R)/(NIR + R)$	[41]
Normalized NIR	NNIR	$NIR/(NIR + R + G)$	[42]
Normalized Red	NR	$R/(NIR + R + G)$	[43]
Optimized soil-adjusted vegetation index	OSAVI	$(NIR - R)/(NIR + R + 0.16)$	[36]
Red vegetation index	RI	$(R - G)/(R + G)$	[44]
Ratio vegetation index	RVI	NIR/R	[45]
Transformed vegetation index	TVI	$60(NIR - G) - 100(R - G)$	[46]
Visible atmospheric resistance index	VARI	$(G - R)/(G + R - B)$	[47]
Wide dynamic range vegetation index	WDRVI	$(0.12NIR - R)/(0.12NIR + R)$	[48]

2.3. Methodology

2.3.1. Gray Relational Analysis

Gray Relational Analysis (GRA) serves as a valuable tool for assessing the correlation and strength of association between variables. It provides an effective approach for analyzing, modeling, and predicting systems with limited information, thereby reducing the need for a large number of samples [49,50]. The basic idea is to identify the primary and secondary relationships among factors in the system by calculating the degree of association between variables, thus determining the most influential factor. To ensure precise outcomes, the data are subjected to mean normalization prior to computing the correlation coefficient, thereby mitigating errors arising from disparate dimensions [51]. The main steps involved in calculating gray correlation degree (GCD) are as follows:

Given that the reference sequence $X_0 = \{x_0(k), k = 1, 2, \ldots, n\}$, and the compared sequence is $X_i = \{x_0(k), k = 1, 2, \ldots, n\}$. The formula for calculating the GCD between X_0 and X_i is as follows:

$$GCD = \frac{1}{n}\sum_{i=1}^{n}\gamma(x_0(k),\ x_i(k)) \tag{1}$$

$$\gamma(x_0(k), x_i(k)) = \frac{\min\limits_{i}\min\limits_{k}|x_0(k) - x_i(k)| + \rho\max\limits_{i}\max\limits_{k}|x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho\max\limits_{i}\max\limits_{k}|x_0(k) - x_i(k)|} \tag{2}$$

where ρ is the resolution coefficient, taken as 0.5.

2.3.2. Machine Learning Algorithms

Machine learning has significant advantages in handling intricate relationships, including non-linearity and heteroscedasticity. Ensemble learning, by constructing and combining multiple machine learning models, aims to obtain a more comprehensive and robust supervised model with better learning accuracy [52,53]. In this study, we have employed bagging models, namely random forest (RF) and Extra Trees (ETr), in conjunction with the widely utilized linear regression (LR) algorithm, to establish the estimation model.

The RF and ETr both rely on decision tree algorithm. The RF algorithm is a non-parametric machine learning algorithm that utilizes multiple decision trees trained on samples and integrates their predictions. It achieves this by randomly sampling observations and feature variables from the modeling dataset. Each sampling generates a tree, and each tree generates rules and decision values that are specific to its own characteristics. The random forest algorithm combines the rules and decision values from all decision trees in the forest to achieve regression [16]. For this study, n_estimators and max_depth were set to 200 and 50, respectively.

(1) By employing bootstrap resampling, a set of n training samples was generated, which is equivalent in size to the original dataset. Subsequently, n regression trees were constructed $\{h(x, \theta_n), n = 1, 2, \ldots, N\}$, The formula is as follows:

$$h(x) = \frac{1}{Q}\sum_{n=1}^{N} h(x, \theta_n) \tag{3}$$

where θ_n is an independent equally distributed random variable; n represents number of regression trees; and $h(x)$ represents regression trees.

(2) The process of regression tree growth involves each split node randomly selecting a feature subset consisting of variables from the entire set of variables, and pruning is not required during the splitting process.

(3) During each bootstrap resampling, the unsampled data were utilized to estimate the internal error and determine the significance. Taking x_p ($p = 1, 2, 3, 4$) as input data, the importance score of the q-th tree is as follows:

$$I(x_p) = \sum_{n=1}^{N} I_n(x_p)/N \tag{4}$$

where x_p represents input variable; $I(\cdot)$ represents discriminant function.

The ETr algorithm, similar to random forests, demonstrates high randomness. It incorporates stochastic features and random thresholds for node partitioning in decision trees, thereby introducing greater and more diverse variations in the architecture of each tree. It has few key parameters and utilizes reasonable heuristic methods for parameter configuration. Each decision tree in ETr uses the original training set, resulting in faster training speed and stable results in the presence of outliers and noise in the training data [54]. ETr not only enhances the randomness of decision trees but also improves the accuracy of suboptimal solutions and the flexibility of solution computation [55]. The algorithm utilizes the training data sample as the input for each base classifier, and employs the Gini coefficient to select optimal features for node splitting until a decision tree is constructed. The final decision is obtained by iteratively constructing a multitude of decision trees. For this study, n_estimators and max_depth were set to 150 and 10, respectively. The formula is as follows:

$$Gini = 1 - \sum_{i=1}^{n} p_n^2 \tag{5}$$

$$y = \underset{i=1,2,\ldots,n}{\arg\max} \sum_{m=1}^{M} \lambda(y_m = i) \tag{6}$$

where p_n represents the probability that the selected sample belongs to category n. $y_m = i$ means that the output result of the fault classification of the m-th decision tree is i.

The LR algorithm is the simplest and most foundational form of supervised learning in machine learning, which can be divided into simple linear regression and multiple linear regression. By analyzing the loss function or utility function associated with the problem, the LR algorithm strives to determine the optimal model through the optimization of said function. Traditionally, the least squares method is employed to minimize the loss function [56]. Nevertheless, the least squares method encounters limitations when applied to large datasets. Its loss function is as follows. In this study, we opted for the gradient descent method to minimize the loss function within the LR algorithm, offering an effective alternative.

$$(w*, b*) = \arg\min_{(w,b)} \sum_{i=1}^{n} (wx_i + b + y_i)^2 \tag{7}$$

where w and b are the parameters of the first order equation; y_i represents true value.

2.3.3. Model Evaluation

In this study, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were employed as evaluation metrics to assess the accuracy of the model. The specific expressions are as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i)^2} \tag{9}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |Y_i - X_i| \tag{10}$$

In the equation, X_i represents the observed soil moisture values, Y_i represents the predicted soil moisture values. $\overline{X_i}$ and $\overline{Y_i}$ represents the mean of the corresponding values, and n represents the total number of data points. The higher R^2 is close to 1, the better models perform; the closer RMSE and MAE are to 0, the smaller the models' simulation error.

3. Results

3.1. Sensitive Vegetation Index Selection of Soil Moisture Based on GRA

In this study, GRA was conducted to explore the relationship between vegetation indices derived from multi-source satellite data (GF-1, Landsat-8 and GF-4) and soil moisture at varying depths of 3 cm, 10 cm, and 20 cm. The GCD values were used to assess the correlation between vegetation indices and soil moisture at different depths. The top five vegetation indices ranked by GCD were selected as input variables for modeling. The GCD between vegetation indices from different stages of GF-1, Landsat-8, and GF-4 satellites and soil moisture at different depths was shown in Figure 3.

Figure 3a illustrated a ranked GCD heatmap derived from a GF-1 satellite. Table 4 presents the vegetation indices that exhibited high sensitivity to soil moisture at different depths, as determined by the GCD analysis using GF-1 data. From Figure 3a and Table 4, it can be observed that the GCD between vegetation indices and soil moisture varied across each growth period. Overall, NR, IPVI, and GLI showed a higher correlation with soil moisture in the early stage, while EVI exhibited good correlation with soil moisture at different depths in the middle of the growth stage. On the other hand, RI, VARI, and WDRVI exhibited lower correlation rankings across all time periods.

In Figure 3b, the heatmap illustrated the ranking of the GCD between Landsat-8 vegetation indices and soil moisture across different growth stages. In the middle growth stage, there was a significant difference in the GCD ranking between vegetation indices and soil moisture at different depths. This could be attributed to the vigorous growth of crops during this period, resulting in substantial variations in soil moisture at different depths due to their heightened water uptake efficiency. In the early growth stage, WDRVI and NR derived from Landsat-8 exhibited good correlation with soil moisture at different depths. However, during the peak growth period in August, these vegetation indices showed significant differences in sensitivity to soil moisture at different depths. Towards the late stages of plant growth, GLI demonstrated a stronger association with soil moisture at different depths, while the WDRVI ranked lower in terms of correlation.

Figure 3c displays the heatmap depicting the GCD rankings between GF-4 vegetation indices and soil moisture at different depths. Table 4 presents the curated selection of soil-moisture-sensitive vegetation indices derived from GF-4 data. During the vigorous growth stage of plants, RVI derived from GF-4 exhibited good correlation with soil moisture at different depths. In the late growth stage, GLI demonstrated heightened sensitivity to soil moisture. However, EVI, RI, TVI, and VARI consistently exhibited lower GCD rankings across different growth stages and soil depths, indicative of their diminished responsiveness to soil moisture variations.

In conclusion, it could be concluded that the correlation between vegetation indices derived from the three satellite and soil moisture varied during different periods. RI and VARI consistently ranked lower in GCD during all growth stages. The spectral characteristics of plant leaves are determined by their internal cellular structure, which accounts for this phenomenon. The multiple reflections between the cell walls and cell gaps result in high reflectance in the near-infrared band. Due to the absence of the NIR band in RI and VARI, their responsiveness to the water content of vegetation is weakened. As a result, these indices demonstrated relatively low sensitivity to variations in soil moisture.

3.2. Estimation of Soil Moisture at Different Depths Using Multi-Source Remote Sensing Data and Machine Learning Methods

The results of soil moisture estimation in early, middle, late, and end growth stages are presented in Tables 5–8. The soil moisture estimation model based on GF-1 was the most accurate (with R^2 ranging 0.129–0.928, RMSE ranging 0.017–0.078), followed by Landsat-8 (with R^2 ranging 0.117–0.862, RMSE ranging 0.017–0.088), while the soil moisture estimation model based on GF-4 showed relatively lower accuracy (with R^2 ranging 0.070–0.921, RMSE ranging 0.020–0.140). For the three machine learning models, both the ETr model and the RF model displayed similar accuracy (the R^2 of ETr model ranged from 0.117 to 0.928,

RMSE ranging from 0.021 to 0.091; The R^2 of RF model ranged from 0.225 to 0.926, RMSE ranging from 0.019 to 0.085), while the LR model was the least accurate (R^2 ranging from 0.048 to 0.733, RMSE ranging from 0.030 to 0.144). The accuracy of soil moisture estimation models varied across different soil depths, with the optimum performance observed at a depth of 3 cm.

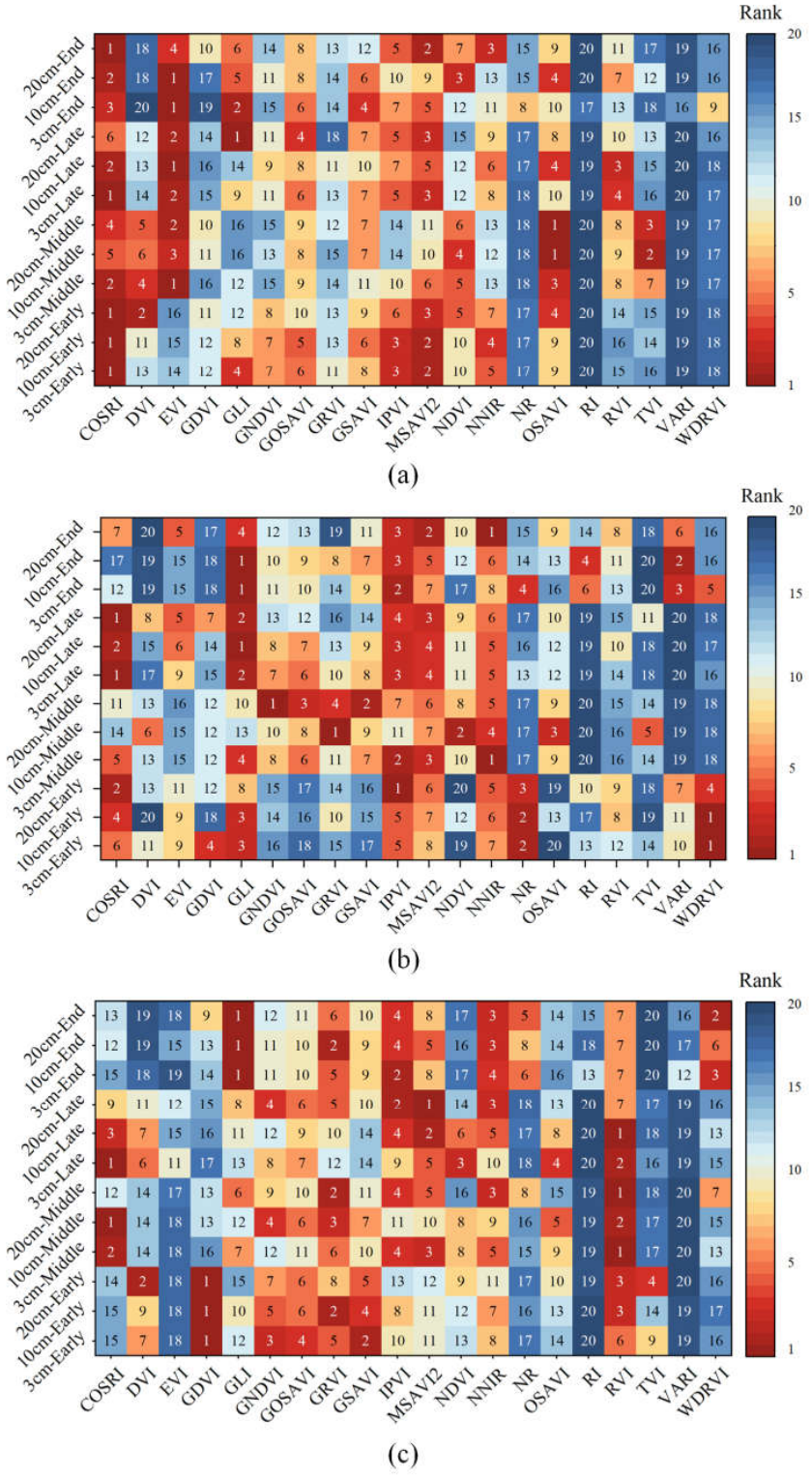

Figure 3. Heat map of gray correlation degree between soil moisture and different vegetation indices based on different satellites. (**a**–**c**) show the heat maps of GF-1, Landsat-8, and GF-4.

Table 4. Screening results of VIs variables based on GRA.

Satellite	Growth Stages	3 cm		10 cm		20 cm	
		VIs	**GCD**	**VIs**	**GCD**	**VIs**	**GCD**
GF-1	Early	NR	0.934	NR	0.938	IPVI	0.926
		WDRVI	0.932	WDRVI	0.931	NR	0.924
		GLI	0.931	IPVI	0.931	NNIR	0.923
		IPVI	0.926	GLI	0.930	GLI	0.922
		NNIR	0.924	GOSAVI	0.929	MSAVI2	0.922
	Middle	EVI	0.898	OSAVI	0.907	OSAVI	0.891
		COSRI	0.894	TVI	0.897	EVI	0.888
		OSAVI	0.882	EVI	0.896	TVI	0.884
		DVI	0.882	NDVI	0.895	COSRI	0.878
		NDVI	0.878	COSRI	0.894	DVI	0.878
	Late	COSRI	0.802	EVI	0.798	GLI	0.884
		EVI	0.802	COSRI	0.791	EVI	0.884
		MSAVI2	0.801	RVI	0.788	MSAVI2	0.884
		RVI	0.801	OSAVI	0.786	GOSAVI	0.883
		IPVI	0.800	MSAVI2	0.786	IPVI	0.882
	End	EVI	0.755	EVI	0.767	COSRI	0.707
		GLI	0.755	COSRI	0.755	MSAVI2	0.704
		COSRI	0.754	NDVI	0.751	NNIR	0.704
		GSAVI	0.752	OSAVI	0.750	EVI	0.704
		MSAVI2	0.750	GLI	0.748	IPVI	0.703
Landsat-8	Early	WDRVI	0.720	WDRVI	0.743	IPVI	0.688
		NR	0.719	NR	0.742	COSRI	0.685
		GLI	0.710	GLI	0.736	NR	0.684
		GDVI	0.704	COSRI	0.734	WDRVI	0.684
		IPVI	0.700	IPVI	0.731	NNIR	0.683
	Middle	NNIR	0.978	GRVI	0.976	GNDVI	0.970
		IPVI	0.978	NDVI	0.975	GSAVI	0.970
		MSAVI2	0.978	OSAVI	0.975	GOSAVI	0.970
		GLI	0.977	NNIR	0.975	GRVI	0.970
		COSRI	0.976	TVI	0.975	NNIR	0.970
	Late	COSRI	0.806	GLI	0.798	COSRI	0.748
		GLI	0.805	COSRI	0.796	GLI	0.744
		IPVI	0.800	IPVI	0.792	MSAVI2	0.743
		MSAVI2	0.798	MSAVI2	0.789	IPVI	0.742
		NNIR	0.797	NNIR	0.789	EVI	0.74
	End	GLI	0.767	GLI	0.770	NNIR	0.702
		IPVI	0.757	VARI	0.763	MSAVI2	0.702
		VARI	0.756	IPVI	0.763	IPVI	0.701
		NR	0.756	RI	0.762	GLI	0.701
		WDRVI	0.754	MSAVI2	0.762	EVI	0.700
GF-4	Early	NR	0.827	WDRVI	0.813	IPVI	0.790
		WDRVI	0.827	NR	0.813	NNIR	0.790
		GLI	0.825	IPVI	0.813	MSAVI2	0.790
		IPVI	0.824	GLI	0.812	GRVI	0.790
		NNIR	0.823	COSRI	0.812	RVI	0.789
	Middle	RVI	0.945	COSRI	0.939	RVI	0.924
		COSRI	0.944	RVI	0.936	GRVI	0.923
		MSAVI2	0.944	GRVI	0.936	NNIR	0.923
		IPVI	0.942	GNDVI	0.935	IPVI	0.922
		NNIR	0.942	OSAVI	0.935	MSAVI2	0.922
	Late	COSRI	0.850	RVI	0.841	MSAVI2	0.838
		RVI	0.847	MSAVI2	0.839	IPVI	0.838
		NDVI	0.847	COSRI	0.836	NNIR	0.837
		OSAVI	0.846	IPVI	0.835	GNDVI	0.837
		MSAVI2	0.845	NNIR	0.835	GRVI	0.837
	End	GLI	0.907	GLI	0.905	GLI	0.880
		IPVI	0.906	GRVI	0.905	WDRVI	0.880
		WDRVI	0.906	NNIR	0.904	NNIR	0.879
		NNIR	0.905	IPVI	0.904	IPVI	0.879
		GRVI	0.905	MSAVI2	0.904	NR	0.878

Table 5. Soil moisture estimation model in early growth stage.

Satellite	Machine Learning Algorithm	Depth (cm)	Training		Testing	
			R^2	RMSE	R^2	RMSE
GF-1	Random Forest	3	0.764	0.024	0.588	0.044
		10	0.814	0.026	0.463	0.044
		20	0.769	0.026	0.404	0.055
	Extra Tree	3	0.812	0.019	0.601	0.041
		10	0.818	0.027	0.518	0.037
		20	0.762	0.034	0.496	0.03
	Linear Regression	3	0.158	0.05	0.129	0.059
		10	0.345	0.039	0.162	0.054
		20	0.155	0.053	0.149	0.078
Landsat-8	Random Forest	3	0.759	0.024	0.308	0.059
		10	0.748	0.028	0.261	0.054
		20	0.79	0.027	0.239	0.054
	Extra Tree	3	0.753	0.029	0.296	0.046
		10	0.798	0.026	0.28	0.047
		20	0.764	0.022	0.117	0.077
	Linear Regression	3	0.292	0.043	0.306	0.055
		10	0.247	0.045	0.23	0.064
		20	0.173	0.052	0.108	0.058
GF-4	Random Forest	3	0.826	0.035	0.349	0.063
		10	0.802	0.046	0.255	0.054
		20	0.803	0.04	0.225	0.082
	Extra Tree	3	0.848	0.035	0.451	0.051
		10	0.823	0.038	0.384	0.075
		20	0.820	0.033	0.372	0.091
	Linear Regression	3	0.389	0.048	0.232	0.084
		10	0.249	0.063	0.07	0.076
		20	0.153	0.096	0.086	0.052

Table 6. Soil moisture estimation model in middle growth stage.

Satellite	Machine Learning Algorithm	Depth (cm)	Training		Testing	
			R^2	RMSE	R^2	RMSE
GF-1	Random Forest	3	0.926	0.019	0.876	0.039
		10	0.913	0.021	0.874	0.061
		20	0.852	0.03	0.814	0.056
	Extra Tree	3	0.928	0.021	0.885	0.024
		10	0.913	0.026	0.828	0.034
		20	0.879	0.028	0.803	0.065
	Linear Regression	3	0.708	0.03	0.566	0.068
		10	0.733	0.039	0.606	0.064
		20	0.612	0.057	0.612	0.048
Landsat-8	Random Forest	3	0.803	0.041	0.731	0.024
		10	0.837	0.032	0.719	0.059
		20	0.717	0.05	0.427	0.064
	Extra Tree	3	0.794	0.033	0.748	0.06
		10	0.809	0.035	0.723	0.07
		20	0.79	0.044	0.602	0.048
	Linear Regression	3	0.221	0.062	0.167	0.091
		10	0.208	0.071	0.174	0.102
		20	0.183	0.089	0.116	0.065
GF-4	Random Forest	3	0.832	0.028	0.716	0.064
		10	0.83	0.031	0.701	0.073
		20	0.745	0.047	0.629	0.061
	Extra Tree	3	0.839	0.028	0.72	0.052
		10	0.834	0.022	0.7	0.08
		20	0.753	0.046	0.682	0.055
	Linear Regression	3	0.185	0.041	0.17	0.12
		10	0.179	0.042	0.131	0.14
		20	0.128	0.089	0.129	0.077

Table 7. Soil moisture estimation model in late growth stage.

Satellite	Machine Learning Algorithm	Depth (cm)	Training		Testing	
			R^2	RMSE	R^2	RMSE
GF-1	Random Forest	3	0.838	0.038	0.776	0.027
		10	0.856	0.035	0.728	0.051
		20	0.796	0.036	0.495	0.08
	Extra Tree	3	0.854	0.037	0.8	0.031
		10	0.829	0.038	0.828	0.071
		20	0.823	0.04	0.612	0.054
	Linear Regression	3	0.277	0.066	0.275	0.083
		10	0.231	0.085	0.248	0.062
		20	0.213	0.076	0.215	0.083
Landsat-8	Random Forest	3	0.838	0.038	0.606	0.024
		10	0.811	0.032	0.776	0.062
		20	0.811	0.029	0.62	0.077
	Extra Tree	3	0.808	0.033	0.693	0.056
		10	0.862	0.031	0.698	0.082
		20	0.832	0.031	0.568	0.073
	Linear Regression	3	0.284	0.078	0.21	0.045
		10	0.334	0.078	0.216	0.058
		20	0.216	0.075	0.219	0.094
GF-4	Random Forest	3	0.845	0.037	0.693	0.028
		10	0.8	0.042	0.689	0.043
		20	0.771	0.044	0.652	0.051
	Extra Tree	3	0.797	0.035	0.778	0.06
		10	0.835	0.03	0.751	0.079
		20	0.818	0.032	0.748	0.08
	Linear Regression	3	0.115	0.073	0.152	0.082
		10	0.245	0.044	0.172	0.118
		20	0.147	0.086	0.131	0.074

Table 8. Soil moisture estimation model at end of the growth stage.

Satellite	Machine Learning Algorithm	Depth (cm)	Training		Testing	
			R^2	RMSE	R^2	RMSE
GF-1	Random Forest	3	0.849	0.025	0.623	0.05
		10	0.793	0.03	0.462	0.078
		20	0.779	0.036	0.579	0.072
	Extra Tree	3	0.836	0.035	0.736	0.046
		10	0.813	0.034	0.732	0.049
		20	0.804	0.036	0.693	0.068
	Linear Regression	3	0.214	0.059	0.255	0.063
		10	0.3	0.066	0.414	0.054
		20	0.276	0.063	0.269	0.088
Landsat-8	Random Forest	3	0.748	0.022	0.316	0.08
		10	0.735	0.031	0.312	0.079
		20	0.769	0.04	0.348	0.071
	Extra Tree	3	0.772	0.026	0.477	0.062
		10	0.766	0.028	0.479	0.075
		20	0.743	0.044	0.472	0.061
	Linear Regression	3	0.117	0.057	0.1	0.078
		10	0.178	0.06	0.151	0.082
		20	0.101	0.068	0.101	0.107
GF-4	Random Forest	3	0.734	0.034	0.233	0.063
		10	0.746	0.037	0.22	0.067
		20	0.725	0.04	0.216	0.085
	Extra Tree	3	0.799	0.033	0.285	0.045
		10	0.78	0.038	0.256	0.063
		20	0.758	0.048	0.189	0.041
	Linear Regression	3	0.188	0.058	0.138	0.078
		10	0.102	0.071	0.071	0.073
		20	0.129	0.059	0.048	0.114

In the early growth stage, the soil moisture estimation models based on Landsat-8 and GF-4 exhibited inadequate accuracy and instability at different depths. This could be attributed to the lower spatial resolution of these two satellites compared to GF-1, as well as the relatively limited vegetation coverage in the early growth stage, which resulted in

restricted spectral reflectance capabilities and the overfitting of the estimation models. In the middle growth stage, the underlying vegetation experiences vigorous growth, resulting in higher vegetation coverage and stronger spectral reflectance. Consequently, the models demonstrated better estimation accuracy and stability compared to other growth stages. In the late growth stage, the vegetation reached the later growth phase, resulting in reduced vegetation coverage compared to the peak growth phase, which led to the decreased accuracy and stability of the models. At the end growth stage, the soil moisture estimation models based on three satellites showed noticeable instability. As the depth increased, both the accuracy and stability of the models deteriorated. Due to the decrease in plant coverage at the end of the growing stage and weak spectral reflectance, there was a decline in the correlation between vegetation index and soil moisture. Additionally, the soil moisture estimation models based on Landsat-8 and GF-4 satellites exhibited significant overfitting, possibly due to their lower spatial resolution, which resulted in model instability.

In conclusion, the GF-1 satellite had better spatial resolution than Landsat-8 and GF-4; it had better accuracy and stability in soil moisture estimation models in four different stages, and it was suitable for vegetated areas. The accuracy of RF was close to that of the ETr model, but the RF model exhibited a poorer stability and was more prone to to overfitting. The models performed best in estimating soil moisture at a depth of 3 cm, while their accuracy decreased and instability increased at a depth of 20 cm. Additionally, compared to GF-1 and Landsat-8 satellite imagery, the retrieval results based on the GF-4 satellite were inferior due to its lower resolution. Moreover, as the vegetation coverage decreased, the accuracy of machine learning models in inverting soil moisture was significantly decreased.

3.3. Comprehensive Evaluation of Soil Moisture Estimation Accuracy Based on Different Remote Sensing Imagery and Machine Learning Models

Figure 4 shows the R^2 for soil moisture estimation at depths of 3 cm, 10 cm, and 20 cm using different satellites. It can be observed that at a depth of 3 cm, the modeling and validation R^2 values for GF-1, Landsat-8, and GF-4 were generally higher for all models when vegetation coverage was relatively high. The models with GF-1 had good fitting results at different depths. As the depth increased, the R^2 for both training and testing stages declined, accompanied by an increase in RMSE. Moreover, the models were prone to overfitting in the testing phase, leading to diminished stability. Therefore, machine learning models based on vegetation indices performed well in estimating surface soil moisture (3 cm and 10 cm depths) in the periods of high vegetation coverage.

Figure 4 and Tables 5–8 show that the ETr model based on GF-1, Landsat-8, and GF-4 generally outperformed the RF and LR models in terms of R^2 at different depths. Additionally, the ET model displays relatively smaller RMSE. Although the RF model demonstrated comparable R^2 and RMSE in training stage, it tended to experience heightened instances of overfitting in the testing stage, resulting in lower model stability. The LR model showed good generalization ability at different depths and periods, but its predicting accuracy was relatively low. In conclusion, the ETr model had good estimation accuracy and stability.

What's more, the models based on the three satellites had similar RMSEs in different stages. However, the R^2 values of Landsat-8 and GF-4 were lower than those of GF-1. Specifically, all models showed signs of overfitting based on Landsat-8 when vegetation coverage was low. The acquisition time of the three satellite images were close in the latter two growth stages, so we analyzed the accuracy of each model. The results showed that GF-4 had the poorest estimation accuracy because it had the lowest resolution, and the models were overfitting at each soil depth. Conversely, GF-1 exhibited the highest estimation accuracy, albeit with a relatively mild degree of overfitting. GF-1 was suitable for vegetated areas, and had high resolution, high accuracy and strong model stability in different levels of vegetation coverage, so it was suitable for retrieving soil moisture at a depth of 3 cm and 10 cm.

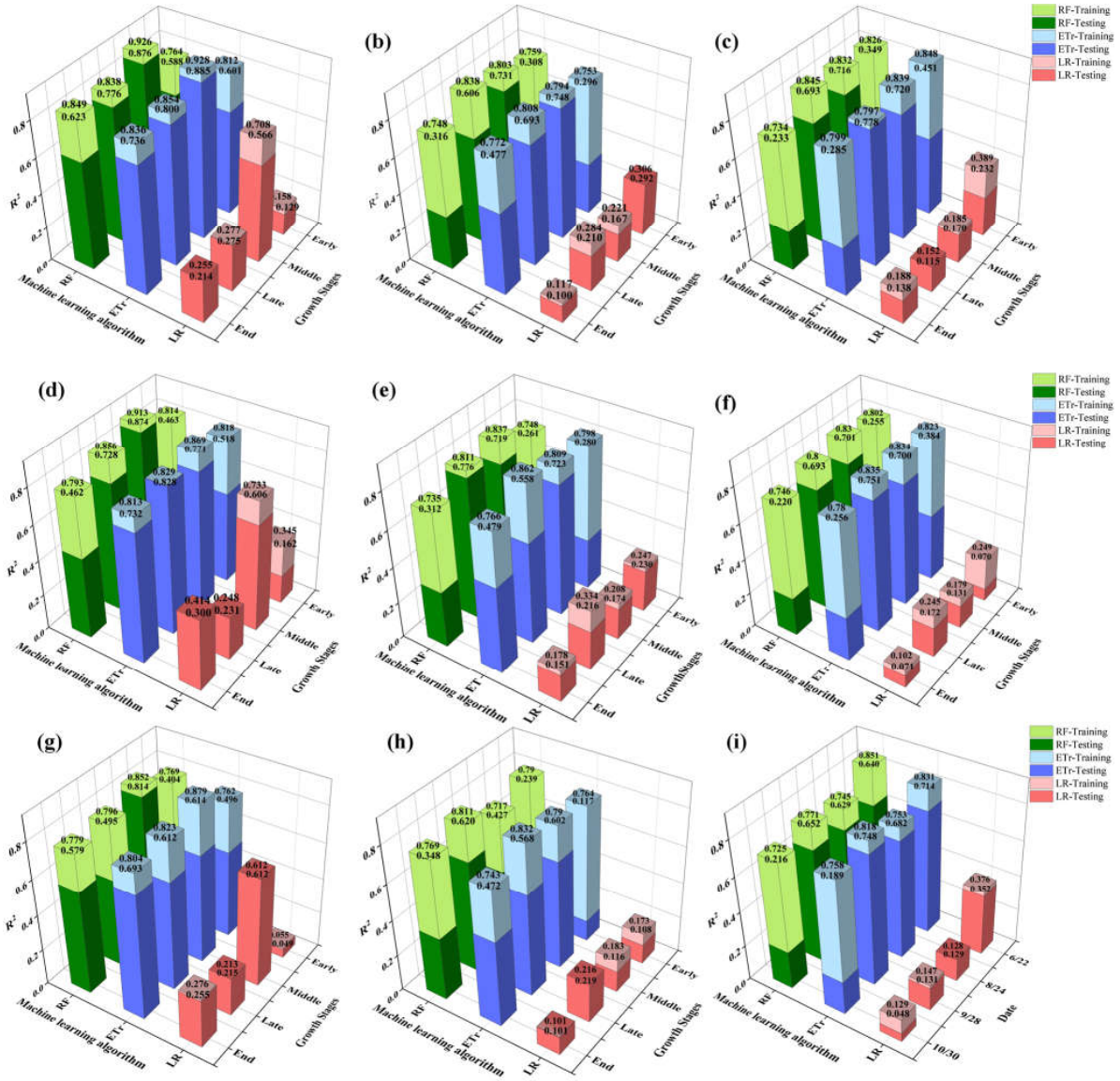

Figure 4. R^2 of 3 cm, 10 cm, and 20 cm soil moisture estimation based on different satellites. (**a**) R^2 of 3 cm depth based on GF-1, (**b**) R^2 of 3 cm depth based on Landsat-8, (**c**) R^2 of 3 cm depth based on GF-4, (**d**) R^2 of 10 cm depth based on GF-1, (**e**) R^2 of 10 cm depth based on Landsat-8, (**f**) R^2 of 10 cm depth based on GF-4, (**g**) R^2 of 20 cm depth based on GF-1, (**h**) R^2 of 20 cm depth based on Landsat-8, and (**i**) R^2 of 20 cm depth based on GF-4.

3.4. Spatial and Temporal Distribution of Soil Moisture in Shandian River Basin

This study found that the ETr model based on GF-1 exhibited good accuracy in estimating soil moisture at different depths in the farming land of northern China. Therefore, the ETr model constructed by GF-1 was employed to simulate the spatial distribution of soil moisture at various depths in the basin (Figure 5). The study area spans from the southwest to the northeast, encompassing diverse surface types including wetlands, crops, grasslands, and woodlands. As can be seen from Figure 5, regions characterized by higher soil moisture were predominantly concentrated in the central and southern parts of the study area during the early, middle, and late growth stages. The main land use types in these regions are farmland and grassland, characterized by a high vegetation cover and a well-developed water system, so these regions are the main agricultural production area of the Shandian River basin. Furthermore, crops in these growing stages would have the corresponding irrigation measures, resulting in higher soil moisture levels. In the end growth stage, the

high values of soil moisture were mainly concentrated in the northern region, while the low values were mainly distributed in the southern farmland region. This was because the land use type in the northern region is grassland and wetland, which had a good soil water retention capacity. Additionally, as crops in the southern region entered the late growth stage without corresponding irrigation measures and with a decline in vegetation coverage, soil moisture was relatively impacted by surface evapotranspiration, resulting in lower levels of soil water during this period. In summary, the study area exhibited significant spatiotemporal variation in soil moisture, which was closely associated with land use types.

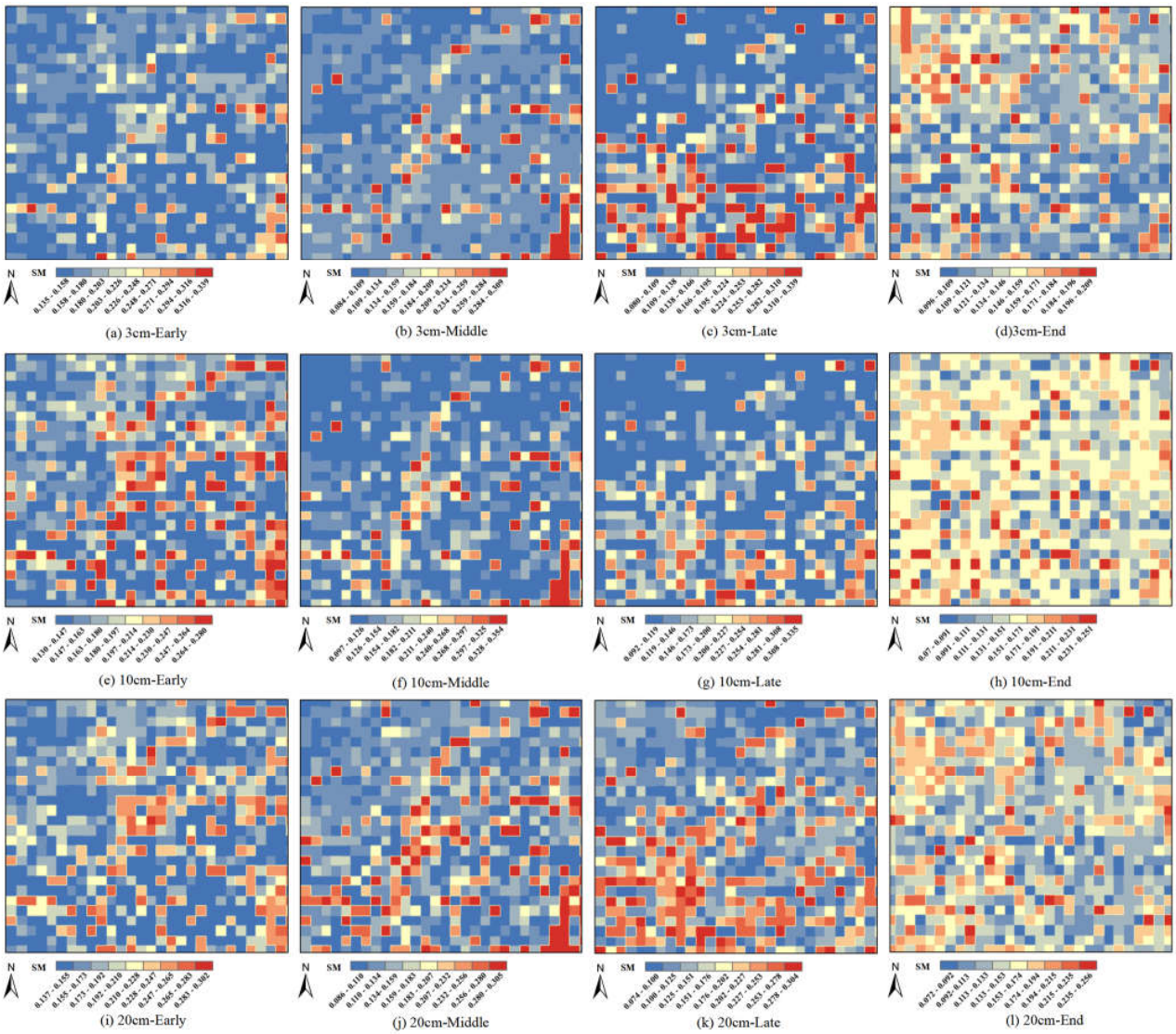

Figure 5. Spatial and temporal distribution of soil moisture at different depths in farming land of northern China based on GF-1 data and ETr model.

Moreover, the spatial distribution of soil moisture at depths of 3 cm, 10 cm, and 20 cm exhibited significant differences during the same period, while there was little difference in the spatial distribution of soil water between depths of 10 cm and 20 cm. The soil moisture at depths of 3 cm, 10 cm, and 20 cm exhibited lower values during the early and middle growth stages, reached its peak in the late growth stage, and declined to its lowest level towards the end of the growth stage. This was attributed to the higher water demand during the early and middle stages of vigorous crop growth resulting in low soil water content. In the late growth stage, water requirements decreased as crops reach maturity, and precipitation occurred prior to image acquisition, so the soil moisture was relatively

high. At the end of the growth stage, no irrigation measures were implemented in the study area, and considering the approaching winter season, arid climate with reduced rainfall, soil moisture was relatively low with insignificant spatial distribution differences compared to the other growth stages.

4. Discussion

4.1. The Sensitive Vegetation Indices of Different Screening Periods Are Obviously Different

In this study, we used R, G, B, and NIR bands of GF-1, Landsat-8, and GF-4 to calculate vegetation indices at different time periods. Then, we employed GRA to select the vegetation indices that were sensitive to soil moisture at different depths. The COSRI index derived from GF-1 exhibited robust correlation with soil moisture at different depths in different periods, showing high accuracy in the estimation of soil moisture. However, Landsat-8 and GF-4 satellites showed different sensitive vegetation indices in different periods. The estimation models of these satellites showed overfitting issues in April and October when vegetation coverage was relatively low. The selection of sensitive vegetation indices varied among different periods due to the factors influencing the reflectance of green plant leaves, which differ significantly in the visible and NIR bands. The water demand of crops varies during different growth stages, and changes in underlying surface conditions also impact soil moisture. Therefore, the trends and magnitudes of reflectance changes in different spectral bands do not align perfectly, resulting in significant differences in vegetation indices constructed based on different spectral bands in different periods [5,57]. Additionally, the vegetation coverage of the underlying surface varies within the growth stages. Vegetation began to grow in April, and gradually withered into the seasonal frozen soil period at the end of October. In two stages, vegetation coverage is relatively low, making it challenging to accurately retrieve vegetation reflectance information, particularly when the spatial resolution of the imagery is low. Consequently, the accuracy and stability of estimation models in these periods deviate significantly.

4.2. Multisource Remote Sensing Data and Machine Learning Model Had Significant Differences in Soil Water Estimation Accuracy at Different Depths

Soil moisture is a crucial parameter for reflecting regional water resource changes and agricultural soil moisture conditions. There are many factors that affect soil moisture, and many researchers have conducted relevant studies on soil moisture estimation [13,14,58]. This study found differences in the accuracy of remote sensing soil moisture estimation at different growth stages of crops. Overall, under the same coverage, the GF-1 image had the highest estimation accuracy, followed by Landsat-8, and GF-4 had the lowest estimation accuracy. This discrepancy can be attributed to the varying spatial resolutions of the satellites: GF-1 has a resolution of 16 m, Landsat-8 has a resolution of 30 m, and GF-4 has a resolution of 50 m. This indicates that resolution has a significant impact on soil moisture estimation accuracy [6,59]. Therefore, GF-1 was suitable for vegetated areas and can monitor temporal and spatial changes in soil moisture.

Machine learning algorithms have been widely studied in the field of remote sensing and have shown excellent performance in solving nonlinear relationship problems [60]. Machine learning algorithms such as RF and SVM have been extensively applied to soil moisture prediction due to their high accuracy and stability [61,62]. He et al. [63] integrated the "trapezoid" model and multiple learning techniques (RF and XGBoost) to estimate soil moisture on the Tibetan Plateau based on MODIS data. The results showed that the ensemble model outperformed the separate model. Hence, machine learning algorithms offer significant advantages for soil moisture estimation. Zhao et al. [64] selected the features extracted from Sentinel-1/2 and Radarsat-2 remote sensing data and constructed a soil moisture inversion model based on the RF, RBFNN, GRNN, SVM, GBPNN, and ELM algorithms. The experimental results showed that among the six models, the random forest model had a higher inversion accuracy, with an R^2 of 0.6395 and RMSE of 0.0264. Cheng et al. [65] evaluated the SMC-estimation accuracy provided by multimodal data

fusion and four machine learning algorithms: PLSR, KNN, RF, and BPNN, which showed that the RF algorithm provided more accurate SMC estimates than the other three algorithms. It was found that the RF algorithm demonstrates exceptional performance for soil moisture estimation because it is relatively resistant to overfitting problems [66]. The RF algorithm demonstrates robustness in handling high-dimensional data with a high degree of fault tolerance, and can also be applied to small sample data sets [67]. The neural network algorithm requires a large amount of sample data for effective learning, and its performance is contingent upon the network structure and sample complexity. In situations where the sample size is limited, overfitting may occur in the neural network algorithm [68]. Additionally, several studies have indicated that the utilization of small sample data sets may not be enough for the optimal training of the SVM algorithm, so it is difficult to estimate soil moisture well [16,68]. In this study, the ETr, RF, and LR algorithms, which are suitable for small samples, were used to construct estimation models for soil moisture at different depths. The ETr and RF models exhibited high estimation accuracy, while the LR model performed relatively poorly. But, the ETr and RF models were prone to overfitting when the vegetation coverage was low, which may be due to the poor correlation between spectral information and soil moisture in that period. Both RF and ETr models are based on decision trees, and the RF model obtains the best splitting attribute within a random subset, while the ETr model randomly selects partition points for feature values rather than optimal ones. Consequently, the generated decision trees of the ETr model are generally larger than those of RF, resulting in lower model variance. Thus, the ETr model tends to yield better results and generalization capabilities to some extent [69,70]. Considering accuracy and stability, the ETr model can be considered the optimal model for soil moisture estimation in the Shandian River Basin.

The findings of this study demonstrate that the spatial resolution of remote sensing imagery significantly influences the accuracy of soil water estimation. Moreover, various machine learning algorithms exhibit distinct levels of accuracy based on different sample sizes. In summary, in cases of limited soil water sample data, the integration of ETr or RF algorithms with high-resolution spatial and temporal imagery can accurately estimate regional-scale soil water content.

Due to limited data and other constraints, this study considered the relationship between vegetation indices and soil moisture to construct the estimation model. In subsequent research, the influence of surface temperature, meteorological factors, and other factors on soil moisture will be considered. Additionally, the retrieval of soil water by the fusion of multi-source remote sensing data is also a hot research topic at present. Negahbani et al. [71] used ESTARFM, which combines Landsat8 and MODIS data, to obtain the daily surface SM with a spatial resolution of 100 m. The outcomes of the study indicated the high ability of the proposed fusion approach for achieving accurate and consistent SM monitoring by using the specified ESTARFM model. Thus, we will estimate soil moisture by fusing multi-source satellite remote sensing data and evaluate its inversion accuracy in the follow-up study. At the same time, we will consider higher-resolution remote sensing images (such as Sentinel-2, GF-1 PMI), and use data assimilation algorithms to couple high spatial-temporal resolution remote sensing data with hydrological models which can predict the changes in soil moisture at different temporal scales and provide relevant references for the optical remote sensing estimation of soil moisture. In this study, the GF-4 satellite offers a wide swath of 400 km and a high temporal resolution (with a revisit period of 20 s), but relatively lower spatial resolution (50 m). In future research, data fusion methods will be considered to improve the spatial resolution of GF-4 and achieve the efficient and continuous monitoring of soil moisture at a large regional scale.

5. Conclusions

In this study, three satellites (GF-1, Landsat-8, and GF-4) with different resolutions were used to the vegetation indices sensitive to soil moisture by using GRA. Three different machine learning algorithms (Extra Tree, random forest, and linear regression) were

employed to build soil moisture models at different depths (3 cm, 10 cm, and 20 cm). The goal was to explore the accuracy of soil moisture estimation based on different satellite images and machine learning algorithms in order to obtain the optimal estimation model for soil moisture in the farming land of northern China. The results indicated that the selection of sensitive vegetation indices by using the GRA varied in different periods. This phenomenon arose due to the significant differences in water consumption intensity during the various growth stages of the crops, resulting in the disparate utilization of soil moisture. Furthermore, the reflectance properties of crop leaves undergo transformations because of their growth, further contributing to the dynamic correlation between vegetation indices constructed from different spectral bands and soil moisture levels. However, vegetation indices predominantly containing near-infrared bands exhibit higher sensitivity to moisture. Among three satellite datasets, the higher-resolution GF-1 imagery had the best soil estimation accuracy, followed by Landsat-8. The soil moisture estimation accuracy of GF-4 was the worst. Therefore, GF-1 is suitable for estimating soil moisture in vegetated areas.

Among the three machine learning models, both the ETr model and RF model exhibited similar accuracy, whereas the LR model demonstrated relatively inferior accuracy. Overall, the ETr model showed superior prediction accuracy and stability in estimating soil moisture. Due to the influence of surface reflectance, different satellites exhibited optimal accuracy for soil moisture estimation at a depth of 3 cm, thereby highlighting the significant potential of optical remote sensing imagery in monitoring surface soil water. In this study, the ETr model based on GF-1 had the best accuracy in soil moisture estimation at different growth stages (with an R^2 of 3 cm depth ranging from 0.601 to 0.928, with an R^2 of 10 cm depth ranging from 0.518 to 0.913, with an R^2 of 20 cm depth ranging from 0.496 to 0.879). It is recommended to utilize GF-1 WFV data to construct the ETr model for monitoring surface soil moisture (3 cm and 10 cm) in the farming land of northern China. Therefore, in cases where there are limited ground sample data, it is advisable to utilize high-spatiotemporal-resolution remote sensing data along with machine learning algorithms such as ETr and RF, which are suitable for small samples, for soil moisture estimation.

Author Contributions: Q.L.: Conceptualization, Methodology, Investigation, Writing—Original Draft, Investigation, Writing—Original Draft. Z.W.: Formal analysis, Investigation. N.C.: Conceptualization, Methodology, Writing—Reviewing & Editing, Funding Acquisition. X.J.: Writing—Reviewing & Editing. S.Z.: Writing—Reviewing & Editing. S.J.: Investigation, Writing—Reviewing & Editing. D.G.: Methodology. L.Z.: Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No.2021YFD1600803-1, 2022YFD1900805), the National Natural Science Foundation of China (51922072, 52279041), and the Science and Technology Projects of Sichuan (2022YFQ0066).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The dataset is provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn, accessed on 22 July 2022). We would like to thank the National Tibetan Plateau Data Center.

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

References

1. Pangaluru, K.; Velicogna, I.; Geruo, A.; Mohajerani, Y.; Ciracì, E.; Cpepa, S.; Basha, G.; Rao, S.V.B. Soil Moisture Variability in India: Relationship of Land Surface–Atmosphere Fields Using Maximum Covariance Analysis. *Remote Sens.* **2019**, *11*, 335. [CrossRef]
2. Zheng, X.; Feng, Z.; Li, L.; Li, B.; Jiang, T.; Li, X.; Li, X.; Chen, S. Simultaneously estimating surface soil moisture and roughness of bare soils by combining optical and radar data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102345. [CrossRef]
3. West, H.; Quinn, N.; Horswell, M.; White, P. Assessing Vegetation Response to Soil Moisture Fluctuation under Extreme Drought Using Sentinel-2. *Water.* **2018**, *10*, 838. [CrossRef]
4. Vereecken, H.; Huisman, J.A.; Pachepsky, Y.; Montzka, C.; van der Kruk, J.; Bogena, H.; Weihermüller, L.; Herbst, M.; Martinez, G.; Vanderborght, J. On the spatio-temporal dynamics of soil moisture at the field scale. *J. Hydrol.* **2014**, *516*, 76–96. [CrossRef]

5. Zhang, B.; Wu, Y.; Zhao, B.; Chanussot, J.; Hong, D.; Yao, J.; Gao, L. Progress and Challenges in Intelligent Remote Sensing Satellite Systems. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1814–1822. [CrossRef]
6. Nie, Y.; Tan, Y.; Deng, Y.; Yu, J. Suitability Evaluation of Typical Drought Index in Soil Moisture Retrieval and Monitoring Based on Optical Images. *Remote Sens.* **2020**, *12*, 2587. [CrossRef]
7. Fan, J.; Luo, M.; Han, Q.; Liu, F.; Huang, W.; Tan, S. Evaluation of SMOS, SMAP, AMSR2 and FY-3C soil moisture products over China. *PLoS ONE* **2022**, *17*, e266091. [CrossRef]
8. Yao, Z.; Cui, Y.; Geng, X.; Chen, X.; Li, S. Mapping Irrigated Area at Field Scale Based on the OPtical TRApezoid Model (OPTRAM) Using Landsat Images and Google Earth Engine. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4409011. [CrossRef]
9. Chai, X.; Zhang, T.; Shao, Y.; Gong, H.; Liu, L.; Xie, K. Modeling and Mapping Soil Moisture of Plateau Pasture Using RADARSAT-2 Imagery. *Remote Sens.* **2015**, *7*, 1279–1299. [CrossRef]
10. Al-Bakri, J.; Suleiman, A.; Berg, A. A comparison of two models to predict soil moisture from remote sensing data of RADARSAT II. *Arab. J. Geosci.* **2014**, *7*, 4851–4860. [CrossRef]
11. Jing, W.; Song, J.; Zhao, X. Evaluation of Multiple Satellite-Based Soil Moisture Products over Continental U.S. Based on In Situ Measurements. *Water Resour. Manag.* **2018**, *32*, 3233–3246. [CrossRef]
12. Menenti, M.; Li, X.; Jia, L.; Yang, K.; Pellicciotti, F.; Mancini, M.; Shi, J.; Escorihuela, M.J.; Zheng, C.; Chen, Q.; et al. Multi-Source Hydrological Data Products to Monitor High Asian River Basins and Regional Water Security. *Remote Sens.* **2021**, *13*, 5122. [CrossRef]
13. Liu, Y.; Qian, J.X.; Yue, H. Comprehensive Evaluation of Sentinel-2 Red Edge and Shortwave-Infrared Bands to Estimate Soil Moisture. *Ieee J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 7448–7465. [CrossRef]
14. Cao, M.; Chen, M.; Liu, J.; Liu, Y. Assessing the performance of satellite soil moisture on agricultural drought monitoring in the North China Plain. *Agric. Water Manag.* **2022**, *263*, 107450. [CrossRef]
15. Li, W.; Wang, Y.; Yang, J.; Deng, Y. Time-Lag Effect of Vegetation Response to Volumetric Soil Water Content: A Case Study of Guangdong Province, Southern China. *Remote Sens.* **2022**, *14*, 1301. [CrossRef]
16. Bayad, M.; Chau, H.W.; Trolove, S.; Muller, K.; Condron, L.; Moir, J.; Yi, L. Time series of remote sensing and water deficit to predict the occurrence of soil water repellency in New Zealand pastures. *Isprs-J. Photogramm. Remote Sens.* **2020**, *169*, 292–300. [CrossRef]
17. Koley, S.; Jeganathan, C. Estimation and evaluation of high spatial resolution surface soil moisture using multi-sensor multi-resolution approach. *Geoderma* **2020**, *378*, 114618. [CrossRef]
18. Carranza, C.; Nolet, C.; Pezij, M.; van der Ploeg, M. Root zone soil moisture estimation with Random Forest. *J. Hydrol.* **2021**, *593*, 125840. [CrossRef]
19. Hosoda, M.; Tokonami, S.; Suzuki, T.; Janik, M. Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil. *Radiat. Meas.* **2020**, *138*, 106402. [CrossRef]
20. Senanayake, I.P.; Yeo, I.Y.; Walker, J.P.; Willgoose, G.R. Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. *Sci. Total Environ.* **2021**, *776*, 145924. [CrossRef]
21. Zhu, Q.; Xu, X.; Sun, Z.; Liang, D.; An, X.; Chen, L.; Yang, G.; Huang, L.; Xu, S.; Yang, M. Estimation of Winter Wheat Residue Coverage Based on GF-1 Imagery and Machine Learning Algorithm. *Agronomy.* **2022**, *12*, 1051. [CrossRef]
22. Cheng, M.; Li, B.; Jiao, X.; Huang, X.; Fan, H.; Lin, R.; Liu, K. Using multimodal remote sensing data to estimate regional-scale soil moisture content: A case study of Beijing, China. *Agric. Water Manag.* **2022**, *260*, 107298. [CrossRef]
23. Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghalhari, G.A.F. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Water* **2020**, *12*, 3223. [CrossRef]
24. Zhao, T.J.; Shi, J.C.; Xu, H.X.; Sun, Y.L.; Chen, D.Q.; Cui, Q.; Jia, L.; Huang, S.; Niu, S.D.; Li, X.W.; et al. Comprehensive remote sensing experiment of water cycle and energy balance in the Shandian river basin. *Natl. Remote Sens. Bull.* **2021**, *25*, 871–887. [CrossRef]
25. Zhao, T.; Ji, D.; Jiang, L.; Cui, Q.; Chen, D.; Zheng, J.; Zhang, Z.; Hu, L.; Shi, J. *In-Situ Measurement Data Set (2019) of the Soil Moisture and Temperature Wireless Sensor Network within the Shandian River Basin*; National Tibetan Plateau Data Center: Lanzhou, China, 2021.
26. Luo, L.; Wang, X.; Lasaponara, R.; Xiang, B.; Zhen, J.; Zhu, L.; Yang, R.; Liu, D.; Liu, C. Auto-Extraction of Linear Archaeological Traces of Tuntian Irrigation Canals in Miran Site (China) from Gaofen-1 Satellite Imagery. *Remote Sens.* **2018**, *10*, 718. [CrossRef]
27. Sánchez, N.; Alonso-Arroyo, A.; Martínez-Fernández, J.; Piles, M.; González-Zamora, Á.; Camps, A.; Vall-llosera, M. On the Synergy of Airborne GNSS-R and Landsat 8 for Soil Moisture Estimation. *Remote Sens.* **2015**, *7*, 9954–9974. [CrossRef]
28. Sun, Z.; Wei, J.; Zhang, N.; He, Y.; Sun, Y.; Liu, X.; Yu, H.; Sun, L. Retrieving High-Resolution Aerosol Optical Depth from GF-4 PMS Imagery in Eastern China. *Remote Sens.* **2021**, *13*, 3752. [CrossRef]
29. Gadi, V.K.; Garg, A.; Rattan, B.; Raj, P.; Gaurav, S.; Sreedeep, S.; Sahoo, L.; Berretta, C.; Peng, L. Growth dynamics of deciduous species during their life period: A case study of urban green space in India. *Urban For. Urban Green.* **2019**, *43*, 126380. [CrossRef]
30. Fernández-Buces, N.; Siebe, C.; Cram, S.; Palacio, J.L. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico. *J. Arid Environ.* **2006**, *65*, 644–667. [CrossRef]
31. Xu, L.Y.; Xie, X.D.; Li, S. Correlation analysis of the urban heat island effect and the spatial and temporal distribution of atmospheric particulates using TM images in Beijing. *Environ. Pollut.* **2013**, *178*, 102–114. [CrossRef]

32. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
33. Wu, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sens.* **2014**, *6*, 1211–1233. [CrossRef]
34. Gobron, N.; Pinty, B.; Verstraete, M.M.; Widlowski, J.L. Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications. *Ieee Trans. Geosci. Remote Sensing.* **2000**, *38*, 2489–2505.
35. Ranjan, R.; Chopra, U.K.; Sahoo, R.N.; Singh, A.K.; Pradhan, S. Assessment of plant nitrogen stress in wheat (*Triticum aestivum* L.) through hyperspectral indices. *Int. J. Remote Sens.* **2012**, *33*, 6342–6360. [CrossRef]
36. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]
37. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
38. Ren, H.; Zhou, G. Determination of green aboveground biomass in desert steppe using litter-soil-adjusted vegetation index. *Eur. J. Remote Sens.* **2014**, *47*, 611–625. [CrossRef]
39. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [CrossRef]
40. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
41. Schnell, J.A. *Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation*; NASA/GSFCT Type Final Report; NASA: Pasadena, CA, USA, 1974.
42. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Weisz, R. Aerial Color Infrared Photography for Determining Late-Season Nitrogen Requirements in Corn. *Agron. J.* **2005**, *97*, 1443–1451. [CrossRef]
43. Noborio, K. Measurement of soil water content and electrical conductivity by time domain reflectometry: A review. *Comput. Electron. Agric.* **2001**, *31*, 213–237. [CrossRef]
44. Williamson, H.D. Reflectance from shrubs and under-shrub soil in a semiarid environment. *Remote Sens. Environ.* **1989**, *29*, 263–271. [CrossRef]
45. Mishra, S.; Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* **2012**, *117*, 394–406. [CrossRef]
46. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, *76*, 156–172. [CrossRef]
47. Schneider, P.; Roberts, D.A.; Kyriakidis, P.C. A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. *Remote Sens. Environ.* **2008**, *112*, 1151–1167. [CrossRef]
48. Gitelson, A.A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [CrossRef]
49. Tao, J.; Sun, X.H.; Cao, Y.; Ling, M.H. Evaluation of water quality and its driving forces in the Shaying River Basin with the grey relational analysis based on combination weighting. *Environ. Sci. Pollut. Res.* **2022**, *29*, 18103–18115. [CrossRef]
50. Wang, Y.; Liu, X.; Ren, G.; Yang, G.; Feng, Y. Analysis of the spatiotemporal variability of droughts and the effects of drought on potato production in northern China. *Agric. For. Meteorol.* **2019**, *264*, 334–342. [CrossRef]
51. Wang, X.; Zhang, F.; Kung, H.; Johnson, V.C. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sens. Environ.* **2018**, *218*, 104–118. [CrossRef]
52. Karmakar, C.; Datcu, M. A Framework for Interactive Visual Interpretation of Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4509105. [CrossRef]
53. Xie, J.; Qi, T.; Hu, W.; Huang, H.; Chen, B.; Zhang, J. Retrieval of Live Fuel Moisture Content Based on Multi-Source Remote Sensing Data and Ensemble Deep Learning Model. *Remote Sens.* **2022**, *14*, 4378. [CrossRef]
54. Rhee, J.; Yang, H. Drought Prediction for Areas with Sparse Monitoring Networks: A Case Study for Fiji. *Water* **2018**, *10*, 788. [CrossRef]
55. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
56. Wang, S.; Ning, Y.; Shi, H. A new uncertain linear regression model based on equation deformation. *Soft Comput.* **2021**, *25*, 12817–12824. [CrossRef]
57. Tan, C.; Zhang, Z.; Xu, C.; Ma, Y.; Yao, Z.; Wei, G.; Li, Y. Soil water content inversion model in field maize root zone based on UAV multispectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 63–74.
58. Nguyen, T.T.; Ngo, H.H.; Guo, W.; Chang, S.W.; Nguyen, D.D.; Nguyen, C.T.; Zhang, J.; Liang, S.; Bui, X.T.; Hoang, N.B. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Sci. Total Environ.* **2022**, *833*, 155066. [CrossRef]
59. Wang, Q.; Li, J.; Jin, T.; Chang, X.; Zhu, Y.; Li, Y.; Sun, J.; 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. [CrossRef]
60. Qun'Ou, J.; Lidan, X.; Siyang, S.; Meilin, W.; Huijie, X. Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms—A case study in the Miyun Reservoir, China. *Ecol. Indic.* **2021**, *124*, 107356. [CrossRef]

61. Jeon, G.; Bellandi, V.; Chehri, A. Editorial for the Special Issue Advanced Machine Learning for Time Series Remote Sensing Data Analysis. *Remote Sens.* **2020**, *12*, 2815. [CrossRef]
62. Zhou, P.; Du, L.; Liu, X.; Shen, Y.; Fan, M.; Li, X. Self-Paced Clustering Ensemble. *Ieee Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 1497–1511. [CrossRef]
63. He, L.; Cheng, Y.; Li, Y.X.; Li, F.; Fan, K.L.; Li, Y.Z. An Improved Method for Soil Moisture Monitoring With Ensemble Learning Methods Over the Tibetan Plateau. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2833–2844. [CrossRef]
64. Zhao, J.H.; Zhang, C.Y.; Min, L.; Guo, Z.W.; Li, N. Retrieval of Farmland Surface Soil Moisture Based on Feature Optimization and Machine Learning. *Remote Sens.* **2022**, *14*, 5102. [CrossRef]
65. Cheng, M.H.; Jiao, X.Y.; Liu, Y.D.; Shao, M.C.; Yu, X.; Bai, Y.; Wang, Z.X.; Wang, S.Y.; Tuohuti, N.; Liu, S.B.; et al. Estimation of soil moisture content under high maize canopy coverage from UAV multimodal data and machine learning. *Agric. Water Manag.* **2022**, *264*, 107530. [CrossRef]
66. Han, Z.Y.; Zhu, X.C.; Fang, X.Y.; Wang, Z.Y.; Wang, L.; Zhao, G.X.; Jiang, Y.M. Hyperspectral Estimation of Apple Tree Canopy LAI Based on SVM and RF Regression. *Spectrosc. Spectr. Anal.* **2016**, *36*, 800–805.
67. Breiman. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
68. Yuan, H.H.; Yang, G.J.; Li, C.C.; Wang, Y.J.; Liu, J.G.; Yu, H.Y.; Feng, H.K.; Xu, B.; Zhao, X.Q.; Yang, X.D. Retrieving Soybean Leaf Area Index from Unmanned Aerial Vehicle Hyperspectral Remote Sensing: Analysis of RF, ANN, and SVM Regression Models. *Remote Sens.* **2017**, *9*, 309. [CrossRef]
69. Alfian, G.; Syafrudin, M.; Fahrurrozi, I.; Fitriyani, N.L.; Atmaji, F.T.D.; Widodo, T.; Bahiyah, N.; Benes, F.; Rhee, J. Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. *Computers* **2022**, *11*, 136. [CrossRef]
70. Ganaie, M.A.; Tanveer, M.; Suganthan, P.N.; Snasel, V. Oblique and rotation double random forest. *Neural Netw.* **2022**, *153*, 496–517. [CrossRef]
71. Negahbani, S.; Momeni, M.; Moradizadeh, M. Improving the Spatiotemporal Resolution of Soil Moisture through a Synergistic Combination of MODIS and LANDSAT8 Data. *Water Resour. Manag.* **2022**, *36*, 1813–1832. [CrossRef]