agronomy

MDPI

Article

Estimation of Winter Wheat Residue Coverage Based on GF-1 Imagery and Machine Learning Algorithm

Qilei Zhu [1,2,3], Xingang Xu [1,2,*], Zhendong Sun [1,2], Dong Liang [3], Xiaofei An [4], Liping Chen [4], Guijun Yang [1,2], Linsheng Huang [3], Sizhe Xu [1,2] and Min Yang [1,2]

1 Key Laboratory of Quantitative Remote Sensing in Agriculture, Ministry of Agriculture and Rural Affairs, Beijing 100097, China; p19301153@stu.ahu.edu.cn (Q.Z.); 201983020071@sdust.edu.cn (Z.S.); yanggj@nercita.org.cn (G.Y.); 2222016068@ujs.edu.cn (S.X.); yangmin@stu.hubu.edu.cn (M.Y.)
2 Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
3 National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei 230601, China; dliang@ahu.edu.cn (D.L.); linsheng0808@ahu.edu.cn (L.H.)
4 Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; anxf@nercita.org.cn (X.A.); chenliping@iea.ac.cn (L.C.)
* Correspondence: xuxg@nercita.org.cn

Abstract: Crop residue is an important component of farmland ecosystems, which is of great significance for increasing soil organic carbon, mitigating wind erosion and water erosion and conserving soil and water. Crop residue coverage (CRC) is an important parameter to characterize the number and distribution of crop residues, and also a key indicator of conservation tillage. In this study, the CRC of wheat was taken as the research object. Based on the high-resolution GF-1 satellite remote sensing imagery from China, decision tree (DT), gradient boosting decision tree (GBDT), random forest (RF), least absolute shrinkage and selection operator (LASSO), extreme gradient boosting regression (XGBR) and other machine learning algorithms were used to carry out the estimation of wheat CRC by remote sensing. In addition, the comparisons with sentinel-2 imagery data were also utilized to assess the potential of GF satellite data for CRC estimates. The results show the following: (1) Among the spectral indexes using shortwave infrared characteristic bands from sentinel-2 imagery, the dead fuel index (DFI) was the best for estimating wheat CRC, with an R^2 of 0.54 and an RMSE of 10.26%. The ratio vegetation index (RVI) extracted from visible and near-infrared characteristic bands from GF-1 data performed the best, with an R^2 of 0.46 and an RMSE of 11.39%. The spectral index extracted from GF-1 and sentinel-2 images had a significant response relationship with wheat residue coverage. (2) When only the characteristic bands from the visible and near-infrared spectral ranges were applied, the effects of the spatial resolution differences of different images on wheat CRC had to be taken into account. The estimations of wheat CRC with the high-resolution GF-1 data were significantly better than those with the Sentinel-2 data, and among multiple machine learning algorithms adopted to estimate wheat CRC, LASSO had the most stable capability, with an R^2 of 0.46 and an RMSE of 11.4%. This indicates that GF-1 high-resolution satellite imagery without shortwave infrared bands has a good potential in applications of monitoring crop residue coverage for wheat, and the relevant technology and method can also provide a useful reference for CRC estimates of other crops.

Keywords: wheat residue coverage; GF-1 remote sensing satellite; Sentinel-2 remote sensing satellite; machine learning regression model

Citation: 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. https://doi.org/10.3390/agronomy12051051

Academic Editor: Gniewko Niedbała

Received: 13 March 2022
Accepted: 25 April 2022
Published: 27 April 2022

1. Introduction

As the final product of agricultural crop production, residue coverage plays a significant role in soil ecological protection and secondary crop planting. In addition, residues can reduce the evaporation of water and retain water moisture in the surface soil [1]. At

the same time, residue mulch can also slow down the weathering and erosion of farmland soil. Meanwhile, if the straw is returned to the field, the soil fertility can be increased by the mineralization of the straw after ploughing. This provides natural organic matter for crop growth, meaning straw has a good regulation effect on the crop growth process [2–4]. Studies have shown that residue plays an important role in improving the quality of agricultural products, inhibiting the growth of weeds, reducing the cost of planting and fertilization and increasing grain yield [4–7]. Crop residue coverage (CRC) is an important index reflecting the crop residue distribution in farmland. Quantitative remote sensing technology can monitor straw returning and mulching in a wide range, dynamically and accurately, which is of great significance to the decision of straw returning conservation tillage.

Conventional methods, such as the random sampling method and drawstring method, are often time-consuming and laborious in estimating crop residue coverage [8,9]. With the development of remote sensing spatial information technology, estimation of crop residue coverage based on remote sensing technology has become an important research direction in the field of conservation tillage. At present, research on remote sensing monitoring of crop residue coverage has made some progress. Using ASD hyperspectral data, Daughtry et al. analyzed the response characteristics between the residue coverage and reflectance spectrum of a variety of crops, such as wheat and corn, and pointed out that 2100 nm is a unique cellulosic lignin spectral absorption characteristic of crop residue; the cellulose absorption index (CAI) was proposed to accurately estimate the CRC of crops [10,11]. Based on the spectral features of Sentinel-2 remote sensing images, Cai et al. introduced texture features to estimate wheat residue coverage and achieved a good estimation accuracy [12]. On the basis of analyzing the spatial heterogeneity of soil types, Huang et al. constructed a remote sensing monitoring model of residue coverage based on soil type partitioning by using the NDTI and STI extracted from sentinel-2 remote sensing imagery in the shortwave infrared band and sufficiently estimated the residue coverage of maize crops [13]. Zheng et al. extracted the mini-NDTI spectral index based on Landsat data for crop straw estimation analysis [9,14]. Yue et al. proposed the broadband spectral angle index (BAI) using sentinel-2 data to estimate CRC in wheat [15–17]. Ding et al. used machine learning methods such as ANN, SVR and GPR to estimate corn residue coverage based on sentinel-2 remote sensing image features [18]. Li et al. used the satellite image data from Landsat-8 OLI simulated by ASD hyperspectral data to construct the index NDI_{OLI} to estimate the CRC of wheat [8]. Zhang et al. used IRS data from the China Environmental Satellite (HJ-1B) to construct the index NDI_{IRS} to monitor wheat residue coverage. However, the low spatial resolution of IRS image data limits their application in complex farmland environments [19]. The above studies show that the estimation of crop residue coverage based on Sentinel, Landsat and other satellite remote sensing image data mostly uses an index containing shortwave infrared characteristic bands to realize the remote sensing estimation of crop residue coverage. In recent years, with the continuous development of satellite remote sensing technology in China, a series of high-resolution satellites have been launched successively, and the earth observation image data are increasingly abundant. However, there are few studies on the monitoring of crop residue coverage with high-resolution GF image data. On the other hand, with the rapid development of machine learning application technology in recent years, research on remote sensing estimation of crop target parameters based on machine learning algorithms has gradually become a hotspot [20], but it is rarely seen in the application of remote sensing monitoring of wheat crop residue.

GF-1 is China's first high-resolution satellite, equipped with panchromatic and multispectral cameras, with a short revisit cycle and many available times and phases. It plays an increasingly important role in China's agricultural resource survey. In this paper, based on GF-1 remote sensing data, combined with various typical machine learning algorithms, the remote sensing estimation of wheat residue coverage was carried out and compared with sentinel-2 image data. This paper analyzed whether the high spatial resolution advantage of GF-1 can make up for the lack of shortwave infrared characteristic bands to monitor

residue coverage and then discusses the potential of domestic GF satellite image data in crop residue coverage remote sensing monitoring, in order to provide a new technology and method reference for remote sensing estimation of crop residue coverage.

2. Data and Methods

2.1. Overview of the Study Area

The study area was the Xiaotangshan National Precision Agriculture Research Demonstration Base in Changping District, Beijing, located between 40°00′ N and 40°21′ N, and 116°34′ E and 117°00′ E, with an average altitude of about 36 m. The study area is flat with distinct seasons. The study area has a warm temperate sub-humid continental monsoon climate, with a dry and windy spring, a hot and rainy summer and a cold and dry winter. The annual average temperature in most areas is 10–12 °C, and the annual average precipitation is 500–600 mm. The sampling time of the coverage measurement was 7 July 2020, with clear weather. The wheat harvest in the experimental area was in the middle of June. The distribution of GF-1 image data and sampling points in the study area is shown in the Figure 1.

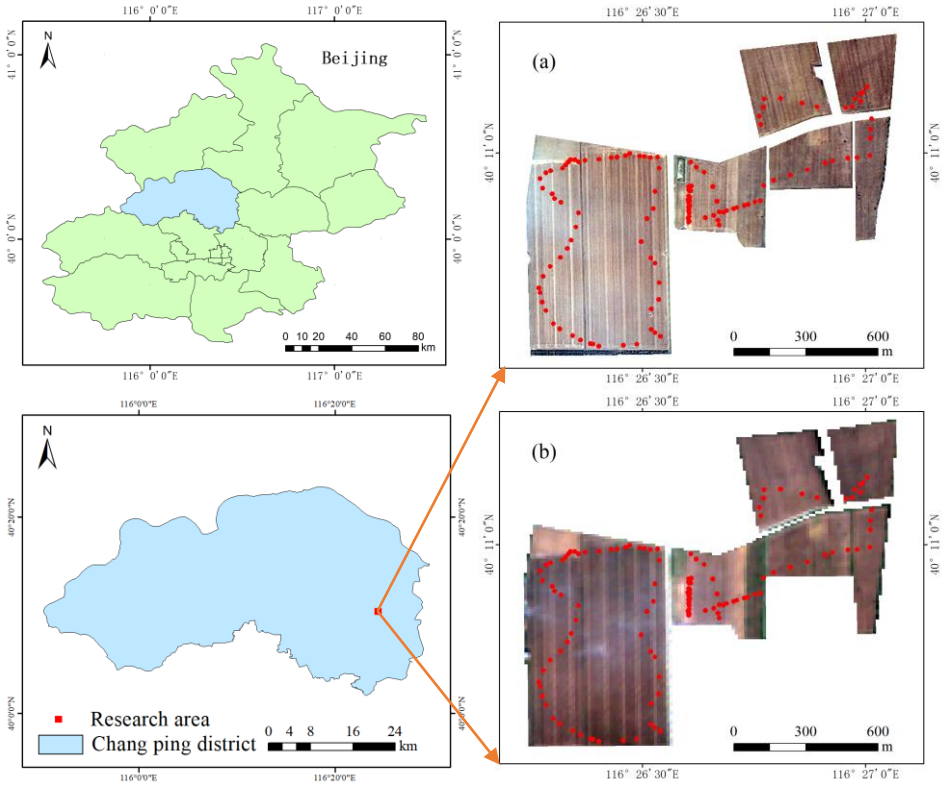

Figure 1. Location of study area and remote sensing imagery ((**a**) GF-1; (**b**) Sentinel-2).

2.2. Data Acquisition and Preprocessing

2.2.1. Acquisition of CRC of Wheat Measured in Field

Field-measured data of wheat residue coverage were sampled on 7 July 2020. Wheat residue coverage data were obtained using a RedEdge-M multispectral camera (Micasense, Seattle, WA, USA, sensor field angle 47.2°, 5-band channel, wavelength range 400–900 nm). When taking samples in the test area, the camera lens was vertically downward, the wheat residue ground samples were kept 1.5 m away from the camera and the geographic coordinates of the field samples were recorded. A total of 121 effective samples were taken. The selected plots and locations of sample points are shown in Figure 1. In order to eliminate the problem of the shooting time interval, this paper used the method of time node partitioning to process sample point data in a short time interval to reduce the influence of light intensity change. The image mosaic method is shown in Figure 2.

Multiple sample images with multispectral information are spliced into one picture with the same spectral information.

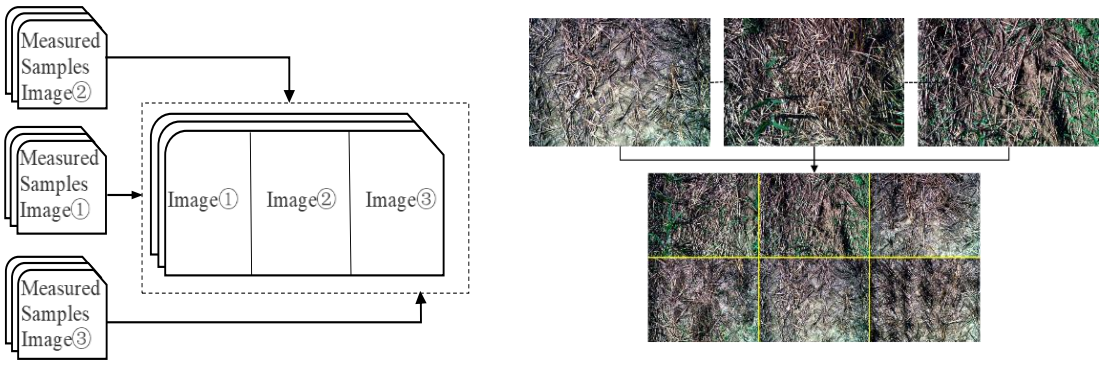

Figure 2. Partition splicing of sample points.

The obtained sample data were divided into different test areas according to the shooting time. After splicing, the band was calculated to obtain a series of spectral index features and obtain the HSV color space transformation features. Then, based on the manual visual interpretation method, different categories of regions of interest (ROIs) were constructed from the sample image data with multispectral characteristics after splicing and made into training samples. The real residual cover value of wheat residue was obtained by using supervised classification, in which the samples were divided into five types: residue, soil, green vegetation, soil shadow and residue shadow. All types of pixels were evenly distributed, and the sample size was sufficient. Supervised classification algorithms such as random forest (RF), support vector machine (SVM) and Mahalanobis distance were used in the experiment to accurately classify the ground object information of various points [21]. Random forest classification is the best supervised learning algorithm based on accuracy analysis. The random forest algorithm also has a good performance in large-scale classification. In the ground crop recognition with GF-1 remote sensing data, combined with the advantages of the high spatial resolution of the satellite images, a good classification accuracy can be obtained.

Field harvest images are shown in Figure 2, which can be divided into five categories: soil, residue, soil shadow, residue shadow and green vegetation. The ROI training samples of real features were selected through artificial visual interpretation. In order to more efficiently identify, classify and reduce part of the workload, the multispectral images of experimental sample points in a similar time area were combined and processed, the training samples were set up in batches and the classification operation was supervised. The method to calculate the coverage is to add the number of pixels of residue, residue shadow and green vegetation as the numerator a, and the denominator as the number of pixels of the whole classification $a + b$.

$$\begin{aligned} CRC &= \tfrac{a}{a+b} \\ a &= Straw + StrawShadows + Vegetables \\ b &= Soil + SoilShadows \end{aligned} \quad (1)$$

2.2.2. Remote Sensing Data Preprocessing

The remote sensing data obtained in this paper are panchromatic/multispectral images of the GF-1 PMS satellite with a spatial resolution of 2 m panchromatic and 8 m multispectral. The acquisition date of the image data was 1 July 2020. After radiometric calibration, atmospheric correction, geometric correction, image fusion and other processing, the final image resolution was 2 m for the fusion multispectral data. Sentinel-2 remote sensing data were obtained from the Google Earth Engine platform on 6 July 2020. The Level 2A product is the reflectivity product of the lower atmosphere after radiometric calibration

and atmospheric correction. The information parameters of the GF-1 and Sentinel-2 sensors are shown in Tables 1 and 2.

Table 1. Main parameters of the GF-1 PMS sensor.

Name	Spectral Range (μm)	Pixel Size/Meters	Revisit Time (Day)	Description
Pan	0.45~0.90	2	4	Panchromatic
B1	0.45~0.52	8		Blue
B2	0.52~0.59			Green
B3	0.63~0.69			Red
B4	0.77~0.89			NIR

Table 2. Main parameters of the Sentinel-2 sensor.

Name	Pixel Size/Meters	Wavelength	Description
B1	60	443.9 nm (S2A)	Aerosols
B2	10	496.6 nm (S2A)	Blue
B3	10	560 nm (S2A)	Green
B4	10	664.5 nm (S2A)	Red
B5	20	703.9 nm (S2A)	Red Edge 1
B6	20	740.2 nm (S2A)	Red Edge 2
B7	20	782.5 nm (S2A)	Red Edge 3
B8	10	835.1 nm (S2A)	NIR
B8A	20	864.8 nm (S2A)	Red Edge 4
B9	60	945 nm (S2A)	Water vapor
B11	20	1613.7 nm (S2A)	SWIR 1
B12	20	2202.4 nm (S2A)	SWIR 2

2.3. Remote Sensing Spectral Index for CRC Estimation

In the experiment, GF-1 and Sentinel-2 remote sensing data were used to construct the CRC model. The spectral indexes selected mainly include the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and soil-adjusted vegetation index (SAVI), and the index information involved in this paper is shown in Table 3.

Table 3. Spectral indexes.

Vegetation Index	Abbreviation	Formula	Reference
Dead fuel index	DFI	$100\left(1-\frac{SWIR2}{SWIR1}\right)\times\frac{R}{NIR}$	[22]
Normalized differential tillage index	NDTI	$\frac{SWIR1-SWIR2}{SWIR1+SWIR2}$	[23]
Simple tillage index	STI	$\frac{SWIR1}{SWIR2}$	[23]
Normalized difference vegetation index	NDVI	$\frac{NIR-R}{NIR+R}$	[24]
Differential vegetation index	DVI	$NIR-R$	[25]
Enhanced vegetation index	EVI	$2.5\frac{NIR-R}{1+NIR+6R-7.5B}$	[26]
Ratio vegetation index	RVI	$\frac{NIR}{R}$	[27]
Optimized soil-adjusted vegetation index	OSAVI	$(1+0.16)\frac{NIR-R}{NIR+R+0.16}$	[28]
Two-band enhanced vegetation index	EVI2	$2.5\frac{NIR-R}{NIR+2.4R+1}$	[29]
Transformed vegetation index	TVI	$60(NIR-G)-100(R-G)$	[30]
Modified soil-adjusted vegetation index	MSAVI2	$\frac{2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-R)}}{2}$	[31]
Wide dynamic range vegetation index	WDRVI	$\frac{0.12NIR-R}{0.12NIR+R}$	[32]
Soil-adjusted vegetation index	SAVI	$1.5\frac{NIR-R}{NIR+R+0.5}$	[33]
Green NDVI	GNDVI	$\frac{NIR-G}{NIR+G}$	[34]
Red vegetation index	RI	$\frac{R-G}{R+G}$	[35,36]

2.4. Technical Route

The overall research route of the experiment is shown in Figure 3. In Figure 3, the red dotted box corresponds to the sentinel-2 data preprocessing method (①②③), and the blue dotted box corresponds to the GF-1 data preprocessing method (①②③④). Spectral features of the corresponding sample locations were extracted from the processed remote sensing images. After the classification and processing of measured sample images, the corresponding coverage value was extracted. Two types of regression models, namely, a univariate regression model and a machine learning model, were constructed based on the spectral characteristics and coverage information.

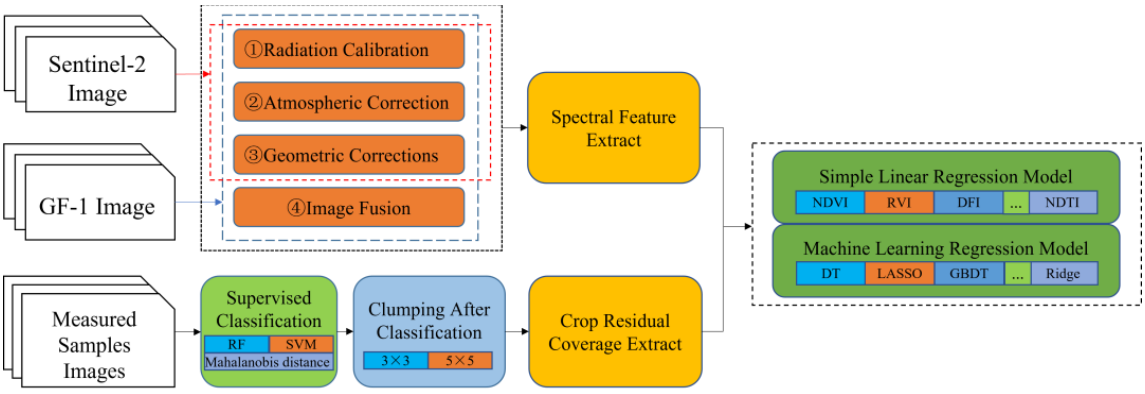

Figure 3. Overall technical route of the experiment.

2.5. Machine Learning Methods

The supervised classification methods adopted in this paper include random forest (RF), support vector machine (SVM), Mahalanobis distance classification and other algorithms, and the supervised classification can effectively reflect the ground object information of real sample points [21,37,38]. In the experiment, the random forest supervised learning method had the highest classification accuracy and the smallest error. RF performs well in many fields with relatively good operation speed and accuracy [21,37–40]. The RF classification algorithm is a type of ensemble algorithm, made up of many of the decision tree branch collection. Many of the decision tree results can work together to attribute data classification and prediction tasks, and this algorithm adopts the bootstrap sampling method for randomly generated data, multiple decision trees and the introduction of a random model, which means it is not easy to fall into overfitting [41]. The main machine learning algorithms used in the regression model constructed by characteristic variables and coverage include decision tree, LASSO regression, ridge regression and the GBDT algorithm. These machine learning methods all belong to a supervised integrated learning method.

The decision tree (DT) algorithm can easily obtain the target result by giving the formulated decision rules. Researchers can easily understand and grasp this structure, and it has a strong explanatory character [42–44].

The LASSO algorithm, first proposed by Robert in 1996, can automatically eliminate insignificant regression coefficients when building models. The complexity of the model is reduced, that is, the overall regression coefficient is reduced, and the function of automatic variable screening is realized [45,46].

Ridge regression was proposed in 1970. It was first used to deal with the case that the sample size was less than the input characteristic parameters. By introducing regular terms into the objective function, the unsolved condition was solved [47,48].

Similar to the previous supervised learning methods, GBDT can be used to solve classification and regression problems simultaneously. The GBDT algorithm can combine multiple weak classifiers to realize the process of strong classifiers, and the feedback residual after each run is matched with the result of the previous one, in order to achieve

a stronger prediction effect through combination. XGBR was further developed from the traditional GBDT [49,50].

2.6. Evaluation of Fitting Accuracy

In the coverage evaluation after classification, the regression relationship between the real ground classification results and the satellite images was established, and the spectral index with a better fitting effect was selected using correlation analysis to establish the regression model. In this paper, 70% of the modeling process was used to establish the training set, and 30% was used to establish the test set. There were 121 samples in the GF-1 dataset, including 84 training samples and 37 test samples. A total of 115 samples were collected from Sentinel-2 data, including 80 training samples and 35 test samples (there are 121 experimental sample points, among which sentinel-2 has multiple sample points with different geographical locations on the same pixel due to the pixel scale). The determination coefficient (R^2) and root mean square error (*RMSE*) were used to judge the fitting degree of the model. The closer R^2 is to 1, the smaller the *RMSE* value is, and the better the inversion effect; otherwise, the worse the model fitting effect. The determination coefficient and root mean square error can be calculated as follows, where the coefficient n is the total number of test samples, y_i is the true measured value of the ith sample, $\hat{y}_i$ is the model predicted value of the ith sample and $\overline{y}$ is the mean value. The root mean square error shows the difference between the true value and the predicted value. It makes the error more significant.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (3)$$

3. Results

3.1. Extraction of Wheat Straw Coverage and Comparison of Different Filtering Results

In the experiment, there were three types of coverage in the regression model, which were the results of the random forest classification, and morphological transformation was performed under 5 × 5 and 3 × 3 clustering processing windows after classification. Morphological transformation can alleviate the salt and pepper phenomenon and increase the coherence and smoothness between classes to some extent. The original sample image and classification filtering results are shown in Figure 4.

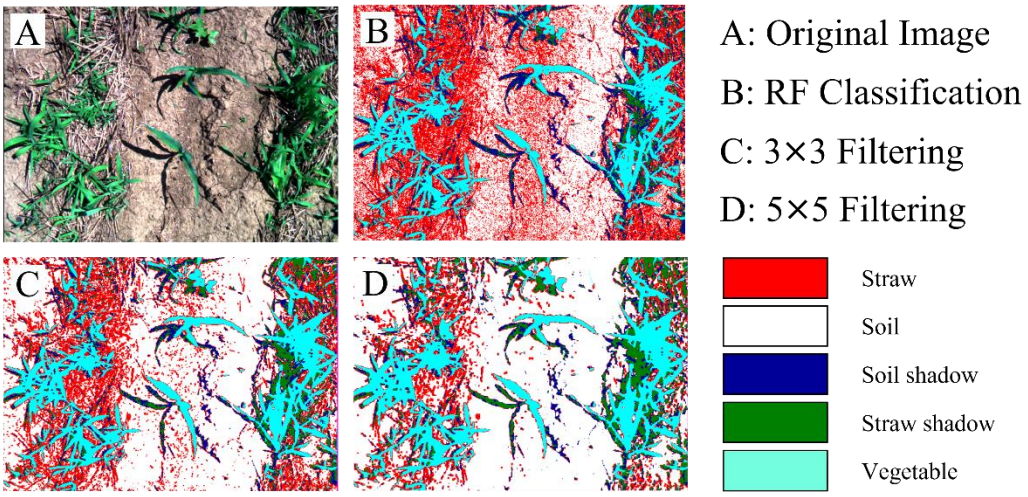

Figure 4. Comparison of sample points before and after classification. Note: In the figure, (**A**), (**B**), (**C**) and (**D**), respectively, represent the original sample image, RF classification image, 3 × 3 salt and pepper filtering processing and 5 × 5 filtering processing after RF classification.

The confusion matrix evaluation coefficient and Kappa coefficient were used to evaluate the classification accuracy. In this paper, the overall classification accuracy and Kappa coefficient of experimental ground samples from high to low were 88% and 0.85 for random forest, 82% and 0.77 for support vector machine, 75% and 0.67 for Mahalanobis distance, and 69% and 0.59, respectively.

3.2. Sensitivity Analysis of Spectral Index Based on GF Image Data to CRC in Wheat

The evaluation results of the GF-1 data spectral index and wheat CRC construction model are shown in Figure 5. It can be found from the Figure 5 that under the advantage of a higher spatial resolution, a variety of spectral indexes can construct the coverage model with a good fitting effect. Through correlation analysis, 12 spectral indexes and coverages were selected to construct the CRC estimation model. Among them, the spectral index with a higher CRC estimation accuracy in the univariate regression model was the RVI. In different morphological window processing, the 5 × 5 clustering results show the best performance in the $CRC_{5 \times 5}$-RVI model (Figure 6). Meanwhile, as can be seen from the Figure 5, the spectral indexes in the GF-1 data analysis results all had a good fitting accuracy in CRC estimation. The CRC model based on the spectral characteristics of CRC and the RVI after different window processing had good stability.

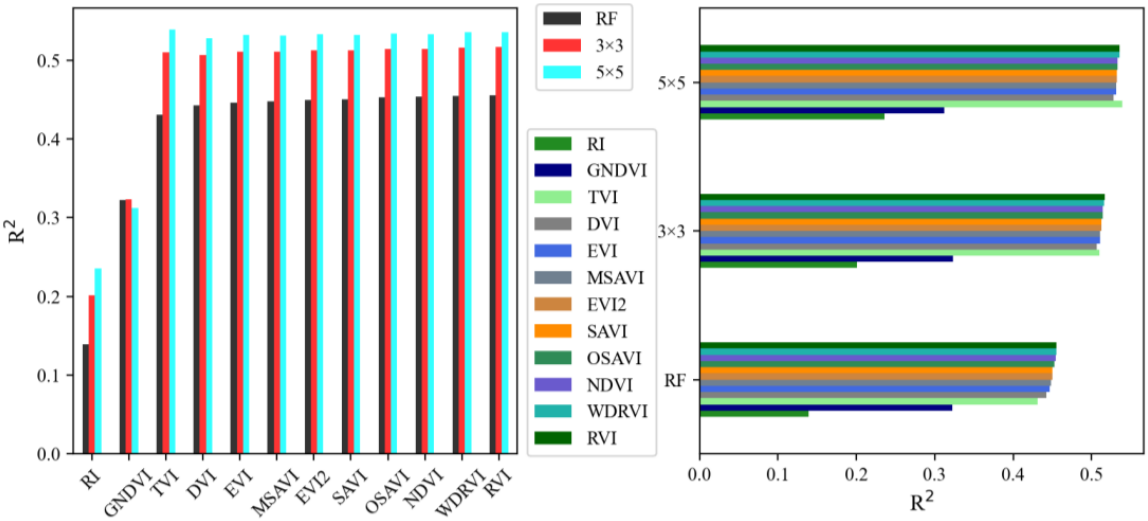

Figure 5. Model determination coefficient (R^2) of GF-1 data.

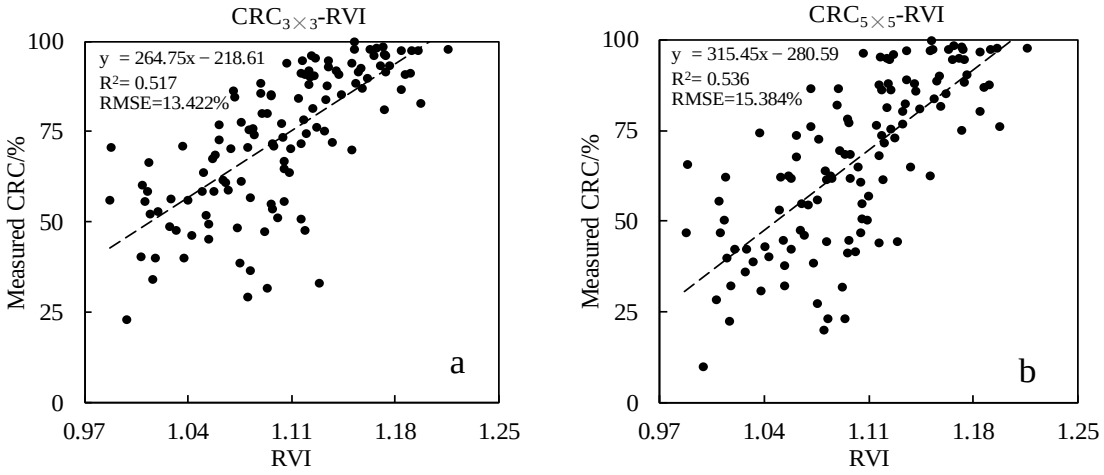

Figure 6. Remote sensing estimation of wheat coverage based on RVI after two filtering processes of GF-1 data (filter scale 3 × 3 (**a**), filter scale 5 × 5 (**b**)).

At the same time, it can be observed from the Figure 7 that the determination coefficient (R^2) of the models without morphological processing of RF classification was generally lower than that of the other two cases and was concentrated around 0.45. By using the mathematical statistical results of the boxplot, it was found that there were some abnormal samples in the results of RF classification processing, while the results of clustering processing were stable. Then, according to the results obtained from the boxplot analysis, the corresponding abnormal sample data were found and removed, and the CRC-RVI model was constructed again. By comparing the CRC models before and after the elimination of abnormal samples, the coefficient of determination (R^2) increased from 0.45 to 0.52. The results show that the accuracy of the CRC model can be improved to a certain extent after the abnormal data are removed by reasonable data analysis tools, which also reflects the better applicability of the data pretreatment method of different morphological windows after classification adopted in this paper. Figure 7 shows the statistical results after boxplot analysis of different classification processing. The left side of Figure 7 is the univariate regression model constructed by the CRC data after RF classification and the RVI spectral index, and the right side is the CRC-RVI estimation model after excluding the RF classification outliers.

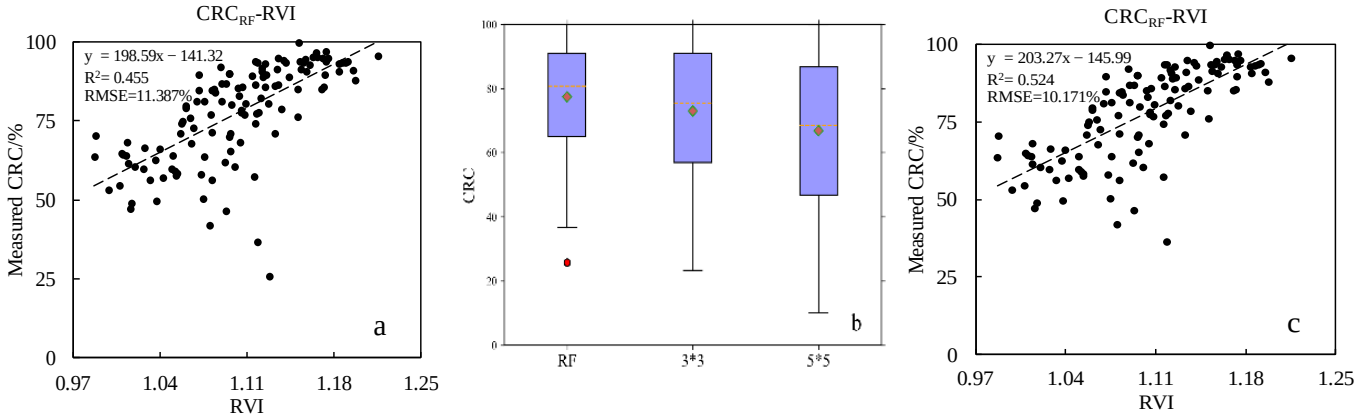

Figure 7. Comparison of GF-1 experimental sample CRC$_{RF}$-RVI before and after the removal of outliers (elimination of outliers from the boxplot (**b**), before (**a**), after (**c**)).

3.3. Sensitivity Analysis of Spectral Index to CRC in Wheat Based on Sentinel-2 Image Data

In this experiment, we combined sentinel-2 remote sensing data and extracted relevant spectral information of the experimental area. Through analysis, it was found that the NDTI, DFI and STI performed well in the univariate regression model, and the highest determination coefficient (R^2) was 0.54. However, the spectral index RVI synthesized by Sentinel-2 in the visible–near-infrared band range was not as accurate as the CRC-RVI model constructed by the GF-1 data and had almost no correlation. A machine learning model will be used for further exploration and research in the subsequent experimental analysis. Figure 8 shows the results of the CRC estimation model analysis of Sentinel data based on spectral information synthesis. Figures 9 and 10 show the univariate regression model constructed by the different spectral indexes in Figure 8, wherein the characteristic index RVI synthesized by Sentinel-2 in the visible–near-infrared spectral range does not show a good response correlation with the estimation model constructed by CRC.

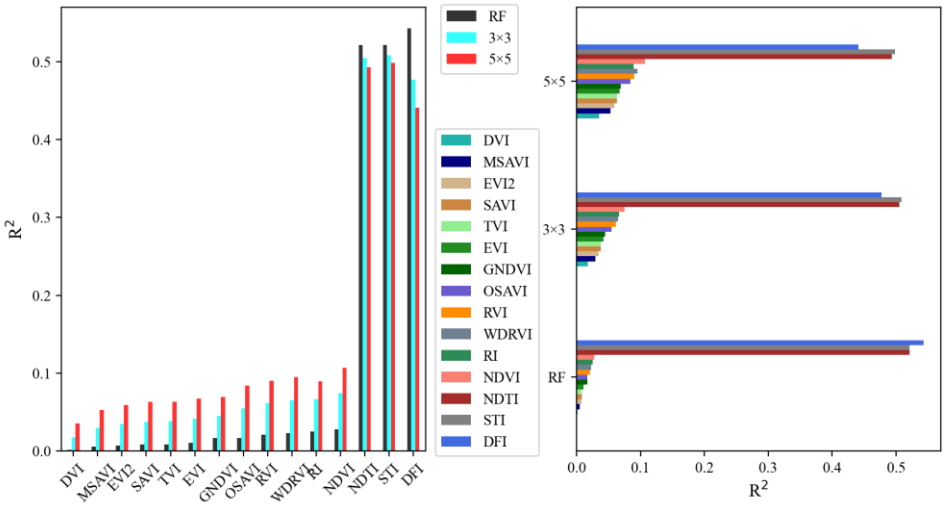

Figure 8. Estimation of wheat coverage R^2 based on Sentinel-2 image data index.

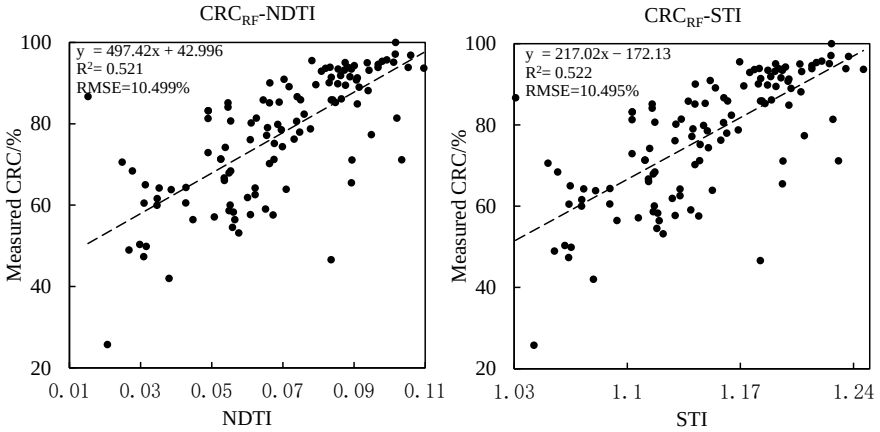

Figure 9. Response relationship between NDTI or STI and CRC_{RF} based on Sentinel-2.

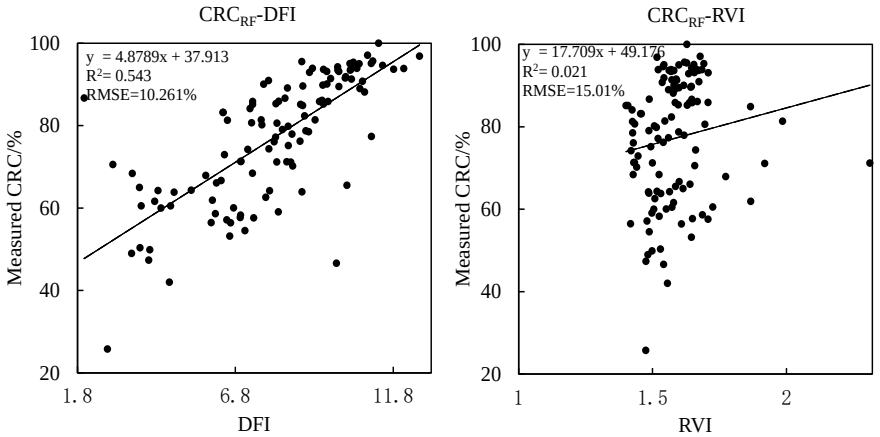

Figure 10. Response relationship between Sentinel-2 DFI or RVI and CRC_{RF} of winter wheat.

It can be seen from Figure 10 that in the sentinel-2 data with a low spatial resolution, in the same univariate model, the RVI model synthesized by Sentinel-2 data in visible light and near-infrared does not have the fitting accuracy of the GF-1 data. Sentinel-2 has a high spectral resolution dominance in the shortwave region, and the construction of various spectral indexes has a model accuracy similar to the GF-1 imagery in the CRC estimation. A spectral index synthesized by the shortwave band such as the DFI can be used as a

reference for accurate estimation of residue coverage, but there are some limitations in achieving accurate estimation of residue coverage on a more precise regional scale. From the perspective of conservation tillage, this study shows that a coverage of more than 30% is defined as conservation tillage. As the experimental base of modern precision agriculture, the coverage of most sample plots in this experiment exceeds 30%. It can be seen from the image that the coverage distribution is relatively uniform. In the experiment, the phenomenon of "supersaturation" described in previous experiments may occur in areas with high coverage [9,11,51]. The possible causes of this phenomenon are further analyzed through the display of the original images below. Compared with the GF-1 data, the relatively coarse spatial resolution of the Sentinel-2 data has some mixed pixel problems, and it is still feasible to roughly estimate CRC in farmland, but it is still insufficient to accurately estimate the crop state in small areas with a scale that is too large. Sentinel-2 data were used as reference objects for GF-1 remote sensing image data in this experiment, highlighting their advantages.

3.4. Remote Sensing Estimation of Wheat CRC Based on Machine Learning and Two Remote Sensing Images

A variety of regression models were selected in the experiment, such as DT, RF and GBDT, among which the DT regression, LASSO regression and ridge regression algorithms have a better model fitting effect. In this paper, the RVI, NDVI and WDRVI constructed by GF-1 and S2 in the visible and near-infrared bands were selected as the parameter combination of the machine learning model because these three characteristic indexes have a good correlation with CRC. In the experiment, reasonable parameter combinations were selected and provided to the basic model for selection through a grid search, and the parameters of the model were tuned by the fitting results after ten cross-fitting tests. The results of various machine learning algorithms for GF-1 and Sentinel-2 data training sets and test sets are shown in Table 4.

Table 4. Comparison of machine learning regression models for GF-1 and Sentinel-2 data.

Characteristic Parameter	Data Source	Machine Learning Method		DT		RF		LASSO		Ridge		GBDT		XGBR	
				GF	S2	GF	S2	GF	S2	GF	S2	GF	S2	GF	S2
RVI, NDVI, WDRVI	CRC_{RF}	training set	R^2	0.53	0.11	0.53	0.1	0.43	0.03	0.43	0.03	0.52	0.09	0.54	0.12
			RMSE/%	10.67	13.62	10.58	13.71	11.67	14.24	11.68	14.24	10.81	13.92	10.52	13.73
		test set	R^2	0.49	0.19	0.49	0.28	0.5	0.18	0.49	0.18	0.52	0.3	0.51	0.37
			RMSE/%	10.78	15.22	10.77	14.87	10.76	15.8	10.77	15.81	10.51	15.56	10.53	14.8
	$CRC_{3 \times 3}$	training set	R^2	0.57	0.14	0.57	0.13	0.51	0.07	0.51	0.07	0.58	0.13	0.62	0.14
			RMSE/%	12.48	17.64	12.6	17.83	13.42	18.36	13.43	18.36	12.55	18.06	11.89	18.16
		test set	R^2	0.52	0.33	0.55	0.38	0.52	0.24	0.52	0.24	0.55	0.39	0.56	0.39
			RMSE/%	13.35	17.47	12.91	17.36	13.54	18.49	13.55	18.51	12.95	18.42	12.85	17.79
	$CRC_{5 \times 5}$	training set	R^2	0.61	0.17	0.6	0.15	0.53	0.11	0.53	0.11	0.62	0.17	0.63	0.17
			RMSE/%	13.87	20.55	14.11	20.85	15.37	21.28	15.33	21.28	13.8	21.07	13.72	22.68
		test set	R^2	0.55	0.34	0.59	0.39	0.55	0.25	0.55	0.25	0.57	0.37	0.57	0.35
			RMSE/%	15.27	20.12	14.52	19.8	15.64	21.24	15.63	21.26	14.92	21.43	14.78	22.39

Note: Of the two columns of data under each machine learning method in the table, GF-1 data are on the left and Sentinel-2 data are on the right.

The CRC results after different filtering (CRC_{RF}, $CRC_{3 \times 3}$, $CRC_{5 \times 5}$) were compared and analyzed by machine learning algorithms. It can be found that compared with the sentinel-2 data results, the training set and verification set of the CRC estimation model constructed by GF-1 are more accurate and stable. The accuracy of the two models after different cluster processing was compared with that of the previous analysis only using the RVI spectral index, which was generally consistent with the results of the previous analysis. Through observation, it was found that the estimation accuracy of the RVI, NDVI and WDRVI constructed by GF-1 in visible and near-infrared light as machine learning model parameters for CRC is better than that of S2.

Figure 11 shows the CRC models constructed by the GF-1 feature index combined with machine learning in Table 4 and their R^2 and *RMSE*. The R^2 and *RMSE* of the training set and test set of the GBDT regression algorithm were 0.52, 10.81%, 0.52 and 10.51%, and those of the training set and test set of the LASSO regression were 0.43, 11.67%, 0.49 and 10.76%. In sentinel-2 (S2) data, the R^2 and *RMSE* of the training set and test set with a better performance of the decision tree regression algorithm were 0.11, 13.62%, 0.19 and 15.22%. The results show that the training set and validation set R^2 of the XGBR, GBDT and DT models has high accuracy and low errors. However, it can be seen from the data distribution that the LASSO model is more concentrated on both sides of the 1:1 line.

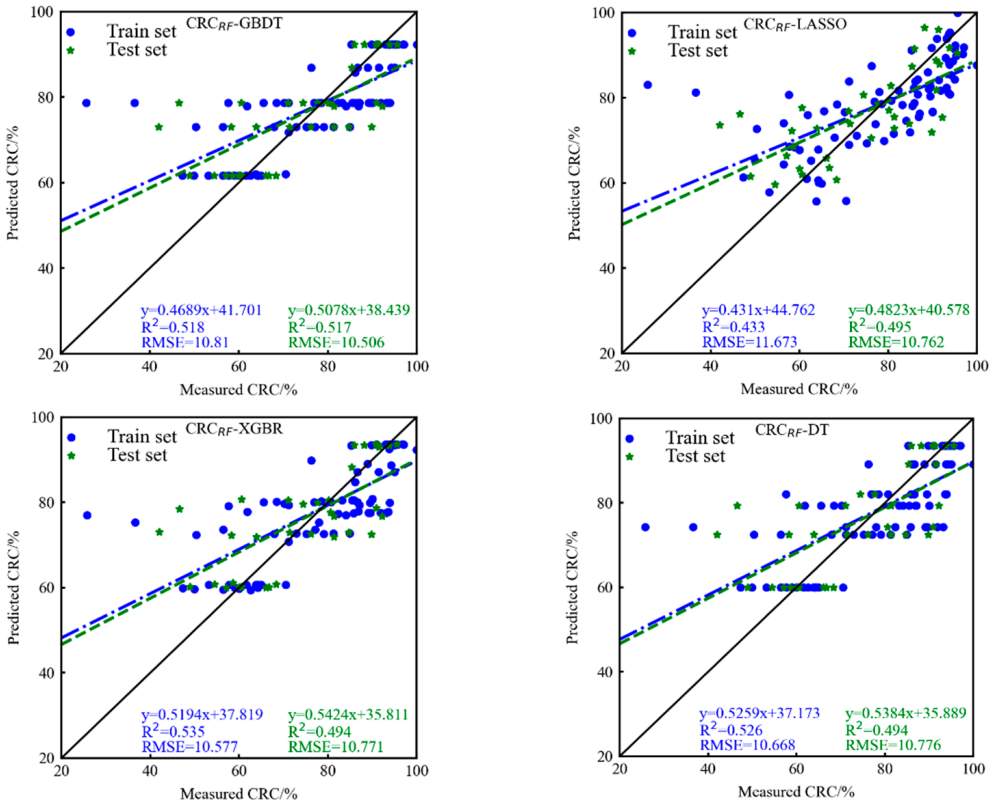

Figure 11. Regression relationship between CRC and XGBR, LASSO, GBDT and DT regression models of GF-1 data.

Based on the above analysis results, the GF-1 index combined with the LASSO model and S2 combined with the DFI model had a good correlation with CRC. In this paper, the respective advantages of the two satellite data models were used to evaluate the distribution of CRC in wheat in the whole test base. The results are shown in Figure 12. As a national agricultural demonstration base, it can be seen from the figure that the CRC distribution after wheat harvest was relatively uniform, and the overall coverage level reached the standard of returning to the field. This paper aimed to improve the accuracy of the model and reduce the error in the process of CRC inversion, but it is difficult to avoid the error in the modeling process. It can be seen from the distribution image that there are some high-coverage or low-coverage areas around the coverage beyond the estimation range, so there are individual blank values.

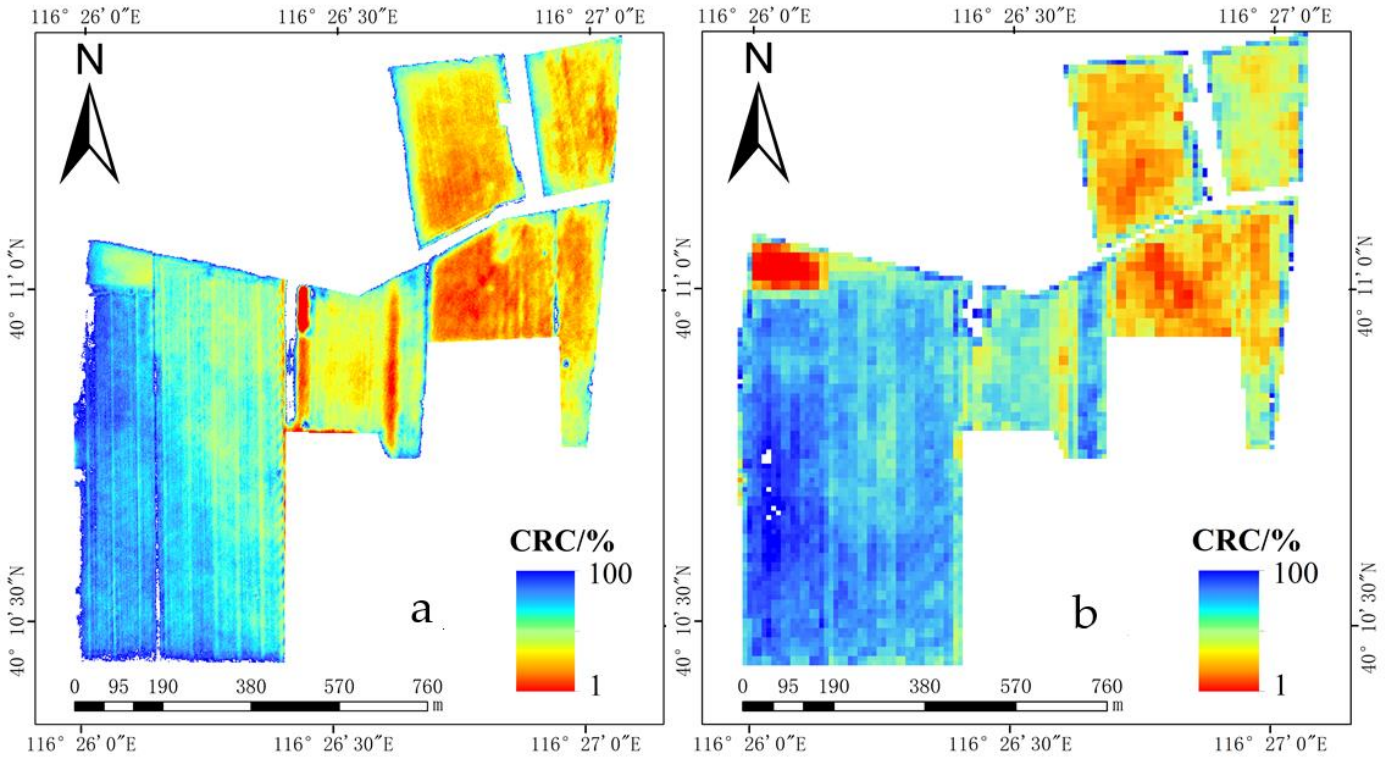

Figure 12. Estimation of wheat CRC ((**a**) based on GF-1 data of the CRC_{RF}-LASSO model; (**b**) based on S2 data of the CRC_{RF}-DFI model).

4. Discussion

According to the analysis results of GF-1 and Sentinel-2 data, sentinel-2 images can sufficiently distinguish residue and soil by the spectral difference in the wavelength range of 2100 nm, depending on their own high spectral resolution. This experiment has similar results to the research on the residue coverage rate in the existing literature. The NDTI, DFI and STI are common spectral indexes [22,52]. In the sentinel-2 data analysis in this paper, the DFI had a good estimation accuracy in the fitting of residue coverage identification. In the univariate regression model, the maximum determination coefficient (R^2) reached 0.54. However, compared with GF-1, sentinel-2 data in the same spectral index synthesized in the range of the visible light and near-infrared wavelengths could not be used to construct a regression model with a certain fitting accuracy with CRC, while the model constructed by GF-1 had a higher fitting accuracy. The spectral index constructed from GF-1 data had a certain degree of correlation with CRC. At the same time, more fine-scale CRC monitoring can be realized.

In Table 5, Sentinel-2 uses the spectral index model synthesized by shortwave advantages, and GF-1 uses the spectral index model synthesized by visible and near-infrared advantages. It can be found that both of them can obtain a great CRC estimation accuracy by virtue of their respective advantages. Among them, the CRC_{RF}-LASSO and $CRC_{3\times3}$-LASSO models' R^2 and *RMSE* are the statistical results of model accuracy for the training set and test set.

Table 5. Comparison of machine learning regression models for GF-1 and Sentinel-2 data.

Data	Model	R^2	*RMSE*/%	Reference
Sentinel-2	CRC_{RF}-DFI	0.54	10.26	[22]
	CRC_{RF}-NDTI	0.52	10.50	[23]
	CRC_{RF}-STI	0.52	10.50	[23]
GF	CRC_{RF}-RVI	0.46	11.39	In this paper
	CRC_{RF}-LASSO	0.46	11.4	In this paper
	$CRC_{3\times3}$-LASSO	0.52	13.46	In this paper

According to the analysis results of the estimation model, it can be seen that the CRC of a small number of sample points has a low value overestimation phenomenon, which is similar to the results of previous studies [13], as shown in the results shown in Figure 12 under the condition of a high spatial resolution. As shown in Figure 13, in the case of high-spatial resolution photography, the shadow produced by the straw occlusion leads to a certain misclassification of the pixel number of the soil and straw. In this experiment, due to the HSV color space transformation, a variety of characteristics of the vegetation index join, and different categories of the ROI area are constructed with a detailed observation of the different objects in the shadow. To a certain extent, the classification of real samples can be effectively separated, the experimental error can be reduced and the estimation accuracy of the model can be improved. The shadow produced by the light irradiation on the residue is projected on the soil surface and the residue surface at the same time. From the classification results, it is a problem to distinguish shadows belonging to the residue or the soil. The classification accuracy of experimental samples has a certain influence on the model. Because the shadow produced by sunlight in other periods is smaller and covers the residue surface, and the results of the shadow on the soil and residue are different, most of the shadows in the experiment can be correctly classified.

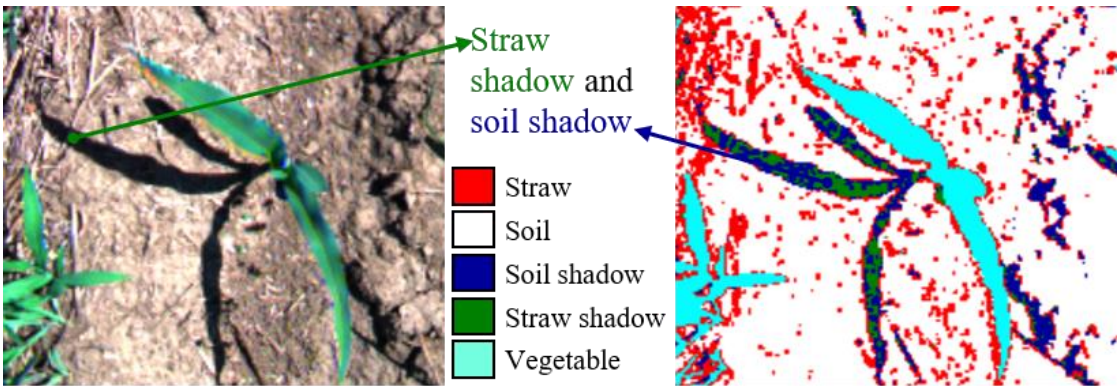

Figure 13. Display of shadow classification error.

In order to verify the above-mentioned reasons and the feasibility of the experimental results, some comparative experiments were conducted. In the experiment, filter windows of different scales were selected to cluster and merge the adjacent regions. Then, a univariate model and a machine learning estimation model were constructed and compared with the CRC results obtained from the three classification processes. From the classification results, the category continuity of random forest classification is poor, because of the existence of residual residue debris on the ground, but its statistical results are comprehensive, and the overall accuracy of the confusion matrix is good. Due to their lower spatial resolution than GF-1, Sentinel-2 data can only make a rough estimation of CRC in areas with a large crop area, but there is still a problem with the large scale in terms of accurately determining the plot coverage in small areas. The high spatial resolution of the GF-1 data has an obvious advantage for regression modeling analysis, and the sentinel-2 data regression model

has a better accuracy in the DFI univariate features under the condition of the same data sample coverage.

The model fitting accuracy of the GF-1 data is similar to that of Sentinel-2, which provides empirical support for accurate point-to-point data experimental analysis under a high spatial resolution and makes up for the experimental blank of coverage analysis in the range of visible–near-infrared bands. In future studies, the impact of shadow will be solved by feasibility analysis and combination with ground handheld multi-hyperspectral, UAV image and GF-1 satellite remote sensing data.

5. Conclusions

In this paper, a new method based on the visible-to-near-infrared spectral information from multi-source satellite images and machine learning algorithms was put forward to estimate wheat residue coverage. The performance of sentinel-2 and GF-1 remote sensing data in estimating wheat residual coverage was compared and analyzed under the advantages of the spatial resolution and spectral resolution. The results show that the spectral index extracted from GF-1 and Sentinel-2 images had a significant response relationship with wheat residue coverage. Based on the machine learning regression algorithms, GF-1 was much better than the sentinel-2 spectral index in the visible and near-infrared bands at estimating wheat residue coverage. The results show that the GF satellite images with spectral information in the visible–near-infrared region and high spatial resolution advantages have potential for the estimation of large-area crop residue coverage. Related studies can also provide certain ideas and method references for the accurate quantitative monitoring of residue coverage in a large area.

Author Contributions: Conceptualization, X.X. and Q.Z.; methodology, Q.Z.; software, Q.Z.; validation, Q.Z.; formal analysis, X.X.; investigation, Z.S., D.L., X.A., L.C., G.Y., L.H., S.X. and M.Y.; resources, Q.Z.; data curation, Q.Z.; writing—original draft preparation, Q.Z.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (Grant No. 2019YFE0125300), the Beijing Academy of Agricultural and Forestry Sciences Innovation Capacity Construction Specific Projects (Grant No. KJCX20210433) and the China Agriculture Research System of MOF and MARA (Grant No. CARS-03).

Data Availability Statement: Data are available on the China Centre for Resources Satellite Data and Application website (http://www.gscloud.cn, accessed on 1 July 2020) and the Google Earth Engine website (https://code.earthengine.google.com, accessed on 7 July 2020).

Acknowledgments: We extend our warm thanks to the technical teams at the Beijing Academy of Agricultural and Forestry Sciences who participated in the ground truth measurement campaigns and data processing.

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

References

1. Yan, Q.-Y.; Dong, F.; Lou, G.; Yang, F.; Lu, J.-X.; Li, F.; Zhang, J.-C.; Li, J.-H.; Duan, Z.-Q. Alternate row mulching optimizes soil temperature and water conditions and improves wheat yield in dryland farming. *J. Integr. Agric.* **2018**, *17*, 2558–2569. [CrossRef]
2. Zahra, K.; Hossein, R.M.; Maryam, D.; Jafari, A.; Abdolreza, K.S.; Jafar, N.M. Feasibility of satellite and drone images for monitoring soil residue cover. *J. Saudi Soc. Agric. Sci.* **2018**, *19*, 56–64.
3. Hrm, A.; Zm, B.; Srm, C.; Ar, C. Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. *Catena* **2021**, *202*, 105258.
4. Shu, W.; Xh, A.; Yang, Z.A.; Cya, C.; Ar, B. The effect of corn straw return on corn production in Northeast China: An integrated regional evaluation with meta-analysis and system dynamics—ScienceDirect. *Resour. Conserv. Recycl.* **2021**, *167*, 105402.
5. Chen, S.; Xu, C.; Yan, J.; Zhang, X.; Zhang, X.; Wang, D. The influence of the type of crop residue on soil organic carbon fractions: An 11-year field study of rice-based cropping systems in southeast China. *Agric. Ecosyst. Environ.* **2016**, *223*, 261–269. [CrossRef]
6. Büchi, L.; Wendling, M.; Amossé, C.; Jeangros, B.; Charles, R. Cover crops to secure weed control strategies in a maize crop with reduced tillage. *Field Crops Res.* **2019**, *247*, 107583. [CrossRef]

7. Xf, A.; Av, A.; Bm, B.; An, A. Winter cover crops in Dutch maize fields: Variability in quality and its drivers assessed from multi-temporal Sentinel-2 imagery—ScienceDirect. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102139.
8. Li, Z.; Wang, C.; Pan, X.; Liu, Y.; Li, Y.; Shi, R. Estimation of wheat residue cover using simulated Landsat-8 OLI datas. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 145–152. (In Chinese)
9. Zheng, B.; Ca Mpbell, J.B.; Beurs, K. Remote sensing of crop residue cover using multi-temporal Landsat imagery. *Remote Sens. Environ.* **2012**, *117*, 177–183. [CrossRef]
10. Daughtry, C.; Hunt, E.R.; Mcmurtrey, J.E. Assessing crop residue cover using shortwave infrared reflectance. *Remote Sens. Environ.* **2004**, *90*, 126–134. [CrossRef]
11. Nagler, P.L.; Daughtry, C.; Goward, S.N. Plant Litter and Soil Reflectance. *Remote Sens. Environ.* **2000**, *71*, 207–215. [CrossRef]
12. Cai, W.T.; Zhao, S.H.; Wang, Y.M.; Peng, F.C. Estimation of winter wheat residue cover using spectral and textural information from Sentinel-2 data. *J. Remote Sens.* **2020**, *24*, 1108–1119. (In Chinese)
13. Huang, J.-Y.; Liu, Z.; Wan, W.; Liu, Z.-Y.; Wang, J.-Y.; Wang, S. Remote sensing retrieval of maize residue cover on soil heterogeneous background. *J. Appl. Ecol.* **2020**, *31*, 474–482. (In Chinese)
14. Zheng, B.; Campbell, J.B.; Serbin, G.; Daughtry, C. Multitemporal remote sensing of crop residue cover and tillage practices: A validation of the minNDTI strategy in the United States. *J. Soil Water Conserv.* **2013**, *68*, 120–131. [CrossRef]
15. Yue, J.; Tian, Q.; Dong, X.; Xu, N. Using broadband crop residue angle index to estimate the fractional cover of vegetation, crop residue, and bare soil in cropland systems. *Remote Sens. Environ.* **2019**, *237*, 111538. [CrossRef]
16. Jyab, C.; Qta, B. Estimating fractional cover of crop, crop residue, and soil in cropland using broadband remote sensing data and machine learning—ScienceDirect. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102089.
17. Yue, J.; Tian, Q.; Dong, X.; Xu, K.; Zhou, C. Using Hyperspectral Crop Residue Angle Index to Estimate Maize and Winter-Wheat Residue Cover: A Laboratory Study. *Remote Sens.* **2019**, *11*, 807. [CrossRef]
18. Ding, Y.; Zhang, H.; Wang, Z.; Xie, Q.; Wang, Y.; Liu, L.; Hall, C.C. A Comparison of Estimating Crop Residue Cover from Sentinel-2 Data Using Empirical Regressions and Machine Learning Methods. *Remote Sens.* **2020**, *12*, 1470. [CrossRef]
19. Zhang, M.; Meng, J.; Li, Q.; Wu, B.; Du, X.; Zhang, F. Estimation of winter wheat residue cover with HJ-1B data. *Trans. Chin. Soc. Agric. Eng.* **2011**, *27*, 352–357. (In Chinese)
20. Zhang, C.; Marzougui, A.; Sankaran, S. High-resolution satellite imagery applications in crop phenotyping: An overview. *Comput. Electron. Agric.* **2020**, *175*, 105584. [CrossRef]
21. Feyisa, G.L.; Palao, L.K.; Nelson, A.; Gumma, M.K.; Paliwal, A.; Win, K.T.; Nge, K.H.; Johnson, D.E. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Comput. Electron. Agric.* **2020**, *175*, 105595. [CrossRef]
22. Cao, X.; Chen, J.; Matsushita, B.; Imura, H. Developing a MODIS-based index to discriminate dead fuel from photosynthetic vegetation and soil background in the Asian steppe area. *Int. J. Remote Sens.* **2010**, *31*, 1589–1604. [CrossRef]
23. Deventer, V.; Ward, A.D.; Gowda, P.H.; Lyon, J.G. Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 87–93.
24. Schell, J.A. Monitoring vegetation systems in the great plains with ERTS. *Nasa Spec. Publ.* **1973**, *351*, 309.
25. Richardson, A.J.; Wiegand, C.L. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.
26. Hui, Q.L.; Huete, A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465.
27. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [CrossRef]
28. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]
29. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [CrossRef]
30. 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]
31. Qi, J.G.; Chehbouni, A.R.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
32. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [CrossRef]
33. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
34. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
35. Escadafal, R. Remote sensing of arid soil surface color with Landsat Thematic Mapper. *Adv. Space Res.* **1989**, *9*, 159–163. [CrossRef]
36. Tian, Q.J.; Min, X.J. Advances in study on vegetation indices. *Adv. Earth Sci.* **1998**, *13*, 327–333. (In Chinese)
37. Zhang, Z.; Wei, M.; Pu, D.; He, G.; Wang, G.; Long, T. Assessment of Annual Composite Images Obtained by Google Earth Engine for Urban Areas Mapping Using Random Forest. *Remote Sens.* **2021**, *13*, 748. [CrossRef]
38. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *Isprs J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]

39. Chen, Y.; Yun, W.; Zhou, X.; Peng, J.; Li, S.; Zhou, Y. Classification and extraction of land use information in hilly area based on MESMA and RF classifier. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 136–144. (In Chinese)
40. Hou, M.J.; Yin, J.P.; Ge, J.; Li, Y.C.; Feng, Q.S.; Liang, T.G. Land Cover Remote Sensing Classification Method of Alpine Wetland Region Based on Random Forest Algorithms. *Trans. Chin. Soc. Agric. Mach.* **2020**, *7*, 220–7227. (In Chinese)
41. Zhang, L.; Gong, Z.N.; Wang, Q.W.; Jin, D.; Wang, X. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images. *J. Remote Sens.* **2019**, *23*, 313–326. (In Chinese)
42. Friedl, M.A.; Brodley, C.E. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* **1997**, *61*, 399–409. [CrossRef]
43. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [CrossRef]
44. Xu, M.; Watanachaturaporn, P.; Varshney, P.K.; Arora, M.K. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* **2005**, *97*, 322–336. [CrossRef]
45. Zou, H. The Adaptive Lasso and Its Oracle Properties. *Publ. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [CrossRef]
46. Naik, P.; Dalponte, M.; Bruzzone, L. Prediction of Forest Aboveground Biomass Using Multitemporal Multispectral Remote Sensing Data. *Remote Sens.* **2021**, *101*, 1282. [CrossRef]
47. Cao, J.; Zhang, Z.; Tao, F.; Zhang, L.; Li, Z. Identifying the Contributions of Multi-Source Data for Winter Wheat Yield Prediction in China. *Remote Sens.* **2020**, *12*, 750. [CrossRef]
48. Hoerl, A.; Kennard, R. Citation Classic-ride Regression-Biased-Estimation for nonorthogonal problems. *Curr. Contents/Eng. Technol. Appl. Sci.* **1982**, *35*, 18.
49. Yang, B.; Ma, J.; Yao, X.; Cao, W.; Zhu, Y. Estimation of Leaf Nitrogen Content in Wheat Based on Fusion of Spectral Features and Deep Features from Near Infrared Hyperspectral Imagery. *Sensors* **2021**, *21*, 613. [CrossRef]
50. Zamani Joharestani, M.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. PM2. 5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere* **2019**, *10*, 373. [CrossRef]
51. Galloza, M.S.; Crawford, M.M.; Heathman, G.C. Crop Residue Modeling and Mapping Using Landsat, ALI, Hyperion and Airborne Remote Sensing Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 446–456. [CrossRef]
52. Chai, G.; Wang, J.; Wang, G.; Han, L.; Wang, Z. Estimation of fractional cover of non-photosynthetic vegetation in typical steppe based on MODIS data. *Remote Sens. Land Resour.* **2019**, *31*, 234–241. (In Chinese)