Article

Evaluation of Crop Type Classification with Different High Resolution Satellite Data Sources

Jinlong Fan [1,*], Xiaoyu Zhang [2], Chunliang Zhao [1,3], Zhihao Qin [3], Mathilde De Vroey [4] and Pierre Defourny [4]

1   National Satellite Meteorological Center, Beijing 100081, China; zcluestc@163.com
2   Ningxia Institute of Meteorological Sciences, Yinchuan 750002, China; zhang_xynet@163.com
3   MOA Key Laboratory of Agricultural Remote Sensing, Institute of Agro-Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; qinzhihao@caas.cn
4   Earth and Life Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; mathilde.devroey@student.uclouvain.be (M.D.V.); pierre.defourny@uclouvain.be (P.D.)
*   Correspondence: fanjl@cma.gov.cn

Abstract: Crop type classification with satellite imageries is widely applied in support of crop production management and food security strategy. The abundant supply of these satellite data is accelerating and blooming the application of crop classification as satellite data at 10 m to 30 m spatial resolution have been made accessible easily, widely and free of charge, including optical sensors, the wide field of viewer (WFV) onboard the GaoFen (GF, high resolution in English) series from China, the MultiSpectral Instrument (MSI) onboard Sentinel 2 (S2) from Europe and the Operational Land Imager (OLI) onboard Landsat 8 (L8) from USA, thanks to the implementation of the open data policy. There are more options in using the satellite data as these three data sources are available. This paper explored the different capability of these three data sources for the crop type mapping in the same area and within the same growing season. The study was executed in a flat and irrigated area in Northwest China. Nine types of crop were classified using these three kinds of time series of data sources in 2017 and 2018, respectively. The same suites of the training samples and validation samples were applied for each of the data sources. Random Forest (RF) was used as the classifier for the crop type classification. The confusion error matrix with the OA, Kappa and F1-score was used to evaluate the accuracy of the classifications. The result shows that GF-1 relatively has the lowest accuracy as a consequence of the limited spectral bands, but the accuracy is at 93–94%, which is still excellent and acceptable for crop type classification. S2 achieved the highest accuracy of 96–98%, with 10 available bands for the crop type classification at either 10 m or 20 m. The accuracy of 97–98% for L8 is in the middle but the difference is small in comparison with S2. Any of these satellite data may be used for the crop type classification within the growing season, with a very good accuracy if the training datasets were well tuned.

Keywords: crop mapping; classification; GF-1; Sentinel-2; Landsat 8

1. Introduction

Global food security has been attracting great concern around the world as the world population is projected to continually grow and the water and thermal condition of farmland is dramatically influenced by this global change. Effectively and efficiently monitoring agricultural production is critical to tackle food insecurity issues and support farming management. Rapid development of Earth observation satellites provides a great opportunity to fulfill this requirement globally. In particular, free and uncharged access to satellite images is accelerating methodology development for agricultural monitoring.

GF, the acronym of GaoFen in Chinese, which means high resolution in English, is one of the key Earth observation programs in China. The China Space Administration has announced at the GEO Ministerial Summit in December 2019 in Australia that the global data of GF-1 data [1] at 16 m spatial resolution will be made freely available to the world

as a contribution from China in building the GEOSS, Global Earth Observing System of Systems. Sentinel [2] is a key component of the Copernicus program for Earth observation in Europe. Sentinel data is already freely accessed via the Copernicus Open Access Hub or other Hubs. Landsat [3] is a USA satellite mission and has a long history, dating back to the 1970s, providing 30-m resolution satellite multispectral data for the world. Landsat 8 is providing high quality satellite data free and openly since its successfully launched in 2013. S2 and L8 [4,5] are freely available via a well-developed data sharing server; however, GF is still new and an online data sharing platform is still in development. GF data are accessible online to some extent via the current web portal, while GF series satellite data are quite easily available in several operational satellite application centers in China in which there are dedicated land lines to connect to the data distributor, CRESDA (China Center for Resource Satellite Data and Application). The user may have more options for the input satellite data in agricultural monitoring as the abundant abovementioned data sources are freely available and at no charge.

As one of key elements of agricultural monitoring [6–8], crop area estimation with the satellite images has been implemented for many years. Crop type classification of using remote sensing time series is an important tool to deliver such information. A bunch of applications and the tool of cloud based and open source in crop type classification have been developed recently, of which the Sen2-Agri system is free and open source, allowing any user generating near-real time products tailored to their needs at their own premises or on cloud computing infrastructure [9]. Additionally, the Google Earth Engine provides a free cloud computing capacity with the various hosted datasets [10,11]. A large variety of crop classification methods at different scales and various levels of accuracy can be found in the literature [12–14]. From the first use of satellites for agricultural monitoring in the 1970s [15] to the latest study [16–21], crop classification strategies have evolved tremendously.

A common method of crop classification is a pixel-based approach in which an individual pixel is assigned to a specific class according to its spectral similarity to the class, as a satellite image is a pixel-based n-dimension data array. However, with the increase of spatial resolution, the pixel-based classification may bring more negative effects because the pixel size is too small to match the target class. Therefore, the object-based method [22–24] was adopted to improve the classification. A key step in the objected-based method is that it takes the spectral-, textural-, neighborhood- and object-specific shape parameters into account as segmentation; hence, it operates with a group of similar pixels instead of a single pixel. In addition to the advantages, limitations are also obvious in this method [25]. The overall effect of the method strongly relies on the segmentation scales. Inappropriate scale for the segmentation may create a negative impact on input features of classification. Spatial resolution of satellite image is another unavoidable problem. The sensors with a high revisit frequency and a large swath width usually offer a coarser spatial resolution. Using such sensors to map heterogeneous agricultural areas is problematic when the field areas are inferior to the sensor's pixel size, causing mixed pixels. In order to overcome the mixed pixel problem, the sub-pixel classification approach was developed for the coarse resolution satellite images [26–30]. The classification approach takes into account the possibility of a pixel of belonging to different classes with a method of pixel unmixing. Working with time series images instead of single one allows extracting temporal features that are a great asset for classifying different crop types [11,31,32].

Developing an algorithm for classification is also extremely important for crop type classification. Traditionally, unsupervised and supervised classifications are the two algorithms of classification. In unsupervised classification, clusters are created according to their similar spectral characteristics of the input images without the training samples. The class type is assigned later based on expert knowledge. The K-means and the ISO-DATA (Iterative Self-Organizing Data Analysis Technique Algorithm) are the two most frequently and widely used unsupervised algorithms. However, accuracy of unsuper-

vised classification is hard to be improved to a very high level in comparison with the supervised classification.

In supervised classification, training samples have to be provided in association with the input images. The final class for each pixel will be decided by such commonly used classifiers as Maximum Likelihood Classification (MLC), Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), Decision Trees (DT) and so on [32]. Many papers [9,33,34] have concluded that RF and other similar classifiers have the best performance in practice. Therefore, RF has been a very popular classification method in many applications, such as land cover and crop type identification. Fusion of optical data with SAR time series images for the crop type classification has also made steady progress in recent years [35–37], especially in reducing the cloudy effects of pixels.

It has been demonstrated the combination of S2 and L8 for classification can obviously improve the capability of change detection over the ground [14,38]. A systematic review on recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping has been presented in Chaves [39], which highlighted the possibility of using a medium-resolution (Landsat-like, 10–30 m) time series and multispectral optical data to provide a harmonization between these sensors and data cube architectures for analysis-ready data that are permeated by publicizations, open data policies and open science principles.

As a newcomer, GF-1 provides alternative high spatial resolution satellite data for classification. A combination of GF, S2 and L8 for classification has not been reported [40–43]. The objective of this study is to evaluate the influence of independent satellite data sources on the crop type classification. We conducted crop type classification with these three data sources independently. Then, we reveal the best option for crop type classification by evaluating the accuracy of the classifications. The question we intend to answer is how well the GF may achieve for crop type classification in comparison with S2 and L8. Results from the study will support the expectation of combining these three data sources for crop type classification in the near future.

2. Study Area and Data

2.1. Study Area

The study area locates at the northern part of the Yellow River Irrigation Area of Ningxia Hui Autonomous Region in Northwest China. With a total territory of 66.400 km^2, Ningxia is naturally composed of three geographical zones strongly governing economic development in the region [44]: the mountainous and loess hilly district (MLHD) in the south, the dry and desert district (DDD) in the center and the Yellow River irrigated district (YERID) in the north. In the MLHD, 70% of the cultivated land consists of slope farmlands. The DDD is a very arid area, mostly covered by grassland, which accounts for 77% of Ningxia's total. The YERID is the most important agricultural production region in Ningxia. It has only one third of the total farmland in the region, but accounts for two thirds of Ningxia's total grain production and agricultural production value. Annual rainfall is extremely low in Ningxia, but the Yellow River brings abundant water from Qinhai-Tibet Plateau for an efficient irrigation to the agricultural fields in the region. In YERID, wheat, rice and corn are by far the major crops. This study focused on the YERID because it is the most important agricultural district in the region. The study area also represents a very typical irrigated agricultural region in China. Availability of clear sky images is high, as the rainfall is extremely low. Figure 1 shows the study area with six Sentinel-2 tiles. The crop growing season in the region is from May to September; therefore, only the satellite data observed in the growing season were collected.
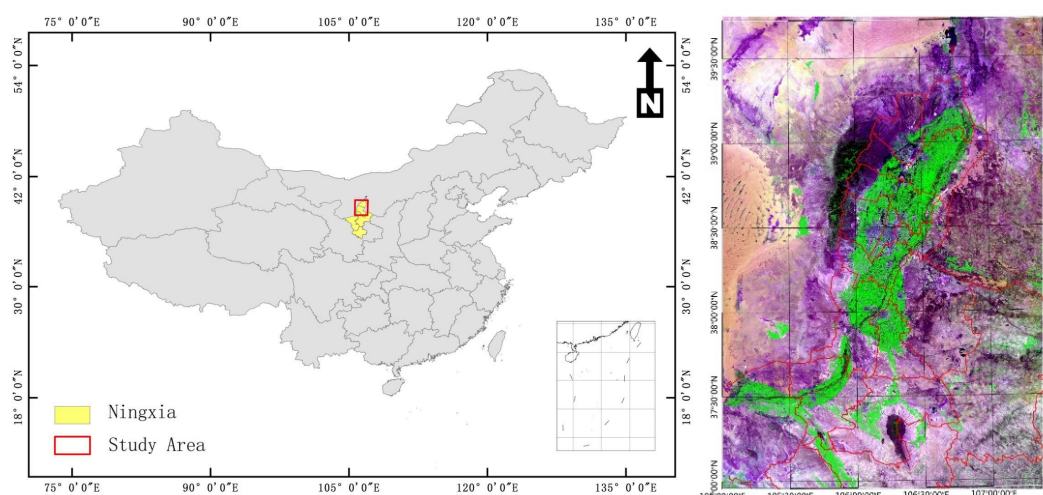
Figure 1. The S2 composite image of the study area showing the landscape and its geographical location in China.

2.2. Satellite Data and Processing

GF-1 Satellite, as the first satellite of the Chinese High Resolution Earth Observation System (GF), which includes series of high resolution optical and SAR satellites, was successfully launched on 26 April 2013 [1]. Onboard, there are four sets of multiple spectral cameras (wide field of viewer, WFV) with a mosaiced swatch of 800 km at 16-meter spatial resolution and a four-day revisit frequency [45]. WFV has only four bands as listed in Table 1. The L1B data of GF-1 WFV data covering the study area and over the growing season of 2017 and 2018 were downloaded from the CRESDA. The L1B data of GF-1 WFV have included the calibration coefficients and the RPC information which were used to implement the image georeferencing with the RPC Orthorectification approach. Then, the FLAASH approach was used to perform the atmospheric correction [46]. In order to compare with S2 and L8, the spatial resolution of GF-1 WFV in this study was set as 15 m.

Table 1. The band specification and spatial resolution of GF-1, S2 and L8.

Band Number	Central Wavelength (nm)	Bandwidth (nm)	Resolution (m)	Central Wavelength (nm)	Bandwidth (nm)	Resolution (m)	Central Wavelength (nm)	Bandwidth (nm)	Resolution (m)
	GF-1 WFV			S2A/B MSI			L8 OLI		
1				443	20	60	443	16	30
2	485	70	16	490	65	10	482	60	30
3	555	70	16	560	35	10	562	57	30
4	660	60	16	665	30	10	655	37	30
5				705	15	20			
6				740	15	20			
7				783	20	20			
8	830	120	16	842	115	10			
8a				865	20	20	865	28	30
9				945	20	60			
10				1375	30	60	1372	21	30
11				1610	90	20	1609	85	30
12				2190	180	20	2200	187	30

The Sentinel satellites are part of European Space Agency's Copernicus program. Sentinel-2A was launched in June 2015 and complemented by Sentinel-2B in March 2017. These two Sentinel missions were equipped with high resolution multispectral (MSI) sensors. Sentinel-2 provides an unprecedented 10-m spatial resolution with a five-day global revisit frequency, a 13 bands imager and a 290 km swath width. Table 1 also lists the spectral and spatial specifications of Sentinel 2A/B. Through the Sen2-Agri system [47], Sentinel-2 L1C (top of atmosphere) images were automatically downloaded for the six tiles covering the study area and over the growing season of 2017 and 2018 from the Copernicus Open Access Hub. Using Sen2-Agri's L2A processor, a Multi-sensor Atmospheric Correction

and Cloud Screening (MACCS algorithm) [48,49] was performed on top of the atmosphere images. The 60-m spectral bands were not used in this study due to the coarse resolution.

Landsat 8, launched on February 11, 2013, consists of two science instruments—the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). These two sensors provide seasonal coverage of the global landmass at a spatial resolution of 30 meters (visible, NIR, SWIR), 100 meters (thermal) and 15 meters (panchromatic). The Landsat 8 image files consist of 16-bit GeoTIFF images, packaged as Georeferenced Tagged Image File Format. Table 1 lists the band information for Landsat 8 [5]. Each scene of Landsat images may be referred to by the path and row. In this study, the scenes of 129/33, 129/34, 130/33 and 130/34 cover the study area. Thus, all the Landsat 8 data during the crop growth season from May to September in 2017 and 2018 were downloaded from the USGS EarthExplorer. The FLAASH approach was used to perform the atmospheric correction for all Landsat 8 images [50]. The thermal bands, the panchromatic band and the Cirrus band were not used in this study.

After a visual check of all images, the images with higher than 25% cloudy coverage were completely removed. The images taken on the same day were mosaiced and tailored to the study area only. The finally used images are listed in Table 2.

Table 2. The used satellites data from GF-1, S2 and L8.

	GF		L8		S2	
	2017	**2018**	**2017**	**2018**	**2017**	**2018**
May	20170526	20180502 20180514 20180530	20170517	20180504 20180511 20180520 20180527	20170517	20180502 20180515 20180517 20180522 20180530
June	20170616 20170628	20180627	20170625	20180605 20180612 20180621 20180628	20170606 20170626	20180604 20180609 20180614
July	20170701 20170706 20170710 20170731	20180705 20180722	20170704	20180714 20180730	20170709 20170711 20170729 20170731	20180704 20180711 20180719 20180721
August	20170804 20170816	20180804 20180815	20170805	20180808 20180815 20180824		
September				20170906 20170922 20170929		

2.3. Field Data and Training Samples

The reference data needed for training and validation of the crop classification were collected in the field. During the field campaign, the georeferenced pictures were taken with a GPS camera along the roads in the study area following predefined itineraries. At home, the land cover and crop type classes with the longitude and latitude coordinates were retrieved by visually screening the pictures with the tool developed for the photo data interpretation. The final output of this process is a formatted file gathering all GPS points with corresponding classes, class codes, author, roadside (left or right), collecting dates and times and the corresponding photo file names. A three-day field campaign with a random sampling approach was carried out in middle June 2017, and collected about 1500 ground truth photos with spatial reference and associated crop or other land cover classes. Three field campaigns with a random block sampling approach were carried out

again in June, July and September in 2018. All those sample points are distributed over the irrigated area, as shown on Figure 2.

The field samples include both the crop classes and non-crop classes. These samples are point-based ground truth and not ideally and evenly distributed in the study area. As a result of the land ownership policy in China, the field of each household is quite small and often a parcel of a few meter width and a few hundred meters length. The field boundary data in the study area are not available, nor are the boundaries in most arable land areas in China. Therefore, it is not feasible to use the field-based training sample, as other papers did [9,47]. However, it looks like a large field with a single crop as the fields often grow the same crop in practice in China. More well-distributed training samples were obtained by visually interpreting the satellite images. Thus, based on those ground sample points, using Google Earth and the Landsat 8 images in the growing season, referring to a 10 km grid frame, the systematic samples approach was implemented and the spatial size of 3 × 3 pixels or a little bit larger of polygons for crop types and non-crop types were drawn on the images. Samples within each 10 km grid were further identified and dramatically increased and the spatial distribution of samples was improved.

All samples were randomly separated into two parts with a ratio of 70% to 30%. The 70% samples were used for classification and 30% for validation. The distance between two samples, in fact, the centers of every two polygons, was taken into account in the separation process. If they are too close, all pixels in both polygons are classified as either training or validation datasets. The threshold for the distance in this study was set as 900 m. This avoids a strong spatial correlation. For instance, all samples taken in one field represent only one field at the level of image pixel. If some pixels of the samples are used for the classification and other for the validation, the accuracy will be very high in normal case, but this accuracy does not represent the true value. Table 3 lists the number of samples and the proportion for each crop type at the 30-m level.
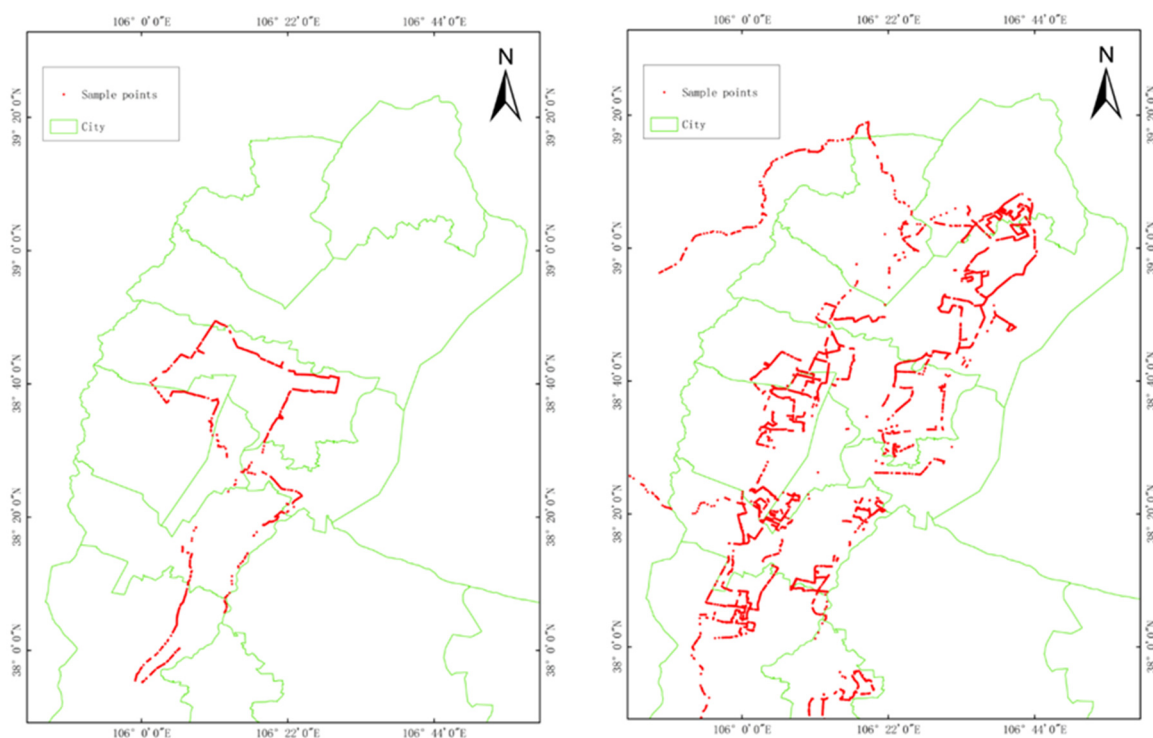

Figure 2. The field samples in the study area for 2017 (**left**) and 2018 (**right**).

Table 3. The number of training samples at 30 m and the proportion for each crop type.

Class Name	Forage Grass	Corn	Grape	Greenhouse	Sward	Medlar	Rice	Vegetable	Wheat
Pixel Counts 2017	2094	2433	1302	1642	494	281	7681	991	509
Pixel Counts 2018	1258	4631	1366	1451	103	167	4134	1727	1793
Proportion 2017	12.0	14.0	7.5	9.4	2.8	1.6	44.1	5.7	2.9
Proportion 2018	7.6	27.8	8.2	8.7	0.6	1.0	24.9	10.4	10.8

3. Methodology

3.1. Classification Algorithm

Two methods are available for crop type and land cover classification: unsupervised and supervised [51]. The K-means and ISO-data are two typical unsupervised classification algorithms which are easy to apply but the accuracy is hard to reach the potential ceiling. The supervised classification is widely used at present. Maximum Likelihood (ML), Support Vector Machine (SVM) and Neural Net (NN) are three typically and widely used supervised algorithms for crop type classification with satellite images. However, Random Forest (RF) has become the more popular classifier in recent years because it is robust and easy to apply and only a few parameters are required to be set and adjusted accordingly. It has commonly proved that the accuracy of RF is often better than that of others such as ML and SVM. Therefore, RF was selected as the classifier in this study.

RF is a supervised machine learning algorithm and a kind of ensemble of the decision tree. The decision will be made based on the bagging method. At each time, a random subset of input training samples will be used and the error of out of bag will be used to evaluate the accuracy. The final result will be decided by merging all trees together and get a more accurate and stable prediction. The detailed RF algorithm may refer to the literature [52–55]. There are only two key parameters to be considered: one is the number of features and the other is the number of trees. In general, the number of features may be set as the square root or the logarithm of the number of input bands of the image. The number of trees should be tested in advance. In practice, 100 or 200 for the number of the tree is fine. The larger both parameters are, the more time it will take to compute, but the accuracy will certainly not be the highest. After many tests, the square root of all input features and 100 was set for the number of features and the number of trees in this study, respectively.

3.2. Increasing Features

In order to achieve the highest accuracy, the classification should take all spatial, temporal and spectral information of the input images into account. For the optical image, clouds always contaminate the images. Some images have been removed due to high cloud coverage. In reality, only a few scenes of time series images may be used for the classification. Therefore, all available satellite image should be used during the crop growing season to capture the crop phenology changes as much as possible. The different crop phenology information is the key to discriminate the different crop types in the same growing season. The spatial information is decided by the spatial resolution of the image. The data source has an influence on the spatial information. More features also may increase the accuracy of the classified image [5,25]. In this study, every two spectral bands may be used to calculate an NDVI-like index (Equation (1)), and thus, all possible NDVI-like indices were calculated and added to the spectral bands as the input features for the classification:

$$F = \frac{b_{i+1} - b_i}{b_{i+1} + b_i} \tag{1}$$

where F is a new feature and b_i and b_{i+1} are spectral bands.

3.3. Validation Methods

An error confusion matrix is usually used as the quantitative evaluation method to calculate the image classification accuracy. The approach of the error confusion matrix was also applied for the validation in this study. Table 4 lists the typic error confusion matrix. The overall accuracy, OA, Kappa and F1-scores may be further calculated based on the error confusion matrix. In Table 4, i represents the ground truth and j the classified result. Nij is the number of pixels that were in class i according to the ground truth but were classified to class j in the resulting image. J is the digital code of each crop type class.

Table 4. The typical error confusion matrix.

| | | Classified | | | |
		j = 1	j = 2	...	j = J
Ground Truth	i = 1	N_{11}	N_{12}		N_{1J}
	i = 2	N_{21}	N_{22}		N_{2J}
	...				
	i = J	N_{J1}	N_{J2}		N_{JJ}

The overall accuracy (OA) can be computed as the proportion of fully correct classes in the validation samples:

$$\text{OA} = \frac{\sum_{i=1}^{J} N_{ii}}{\sum_{i=j=1}^{J} N_{ij}} \tag{2}$$

The Kappa coefficient is another index to evaluate the accuracy of the classified image:

$$\text{Kappa} = \frac{M \sum_{i=j=1}^{J} N_{ij} - \sum_{i=j=1}^{J} N_i N_j}{M^2 - \sum_{i=j=1}^{J} N_i N_j} \tag{3}$$

where M is total number of validation samples. The Kappa coefficient is good to evaluate the unbalanced validation samples, but it is difficult to understand the meaning or what Kappa represents. Kappa gives a sense to evaluate how much mistake was avoided in the classification process. Kappa is always less than or equal to 1. A value of 1 implies perfect agreement and values less than 1 imply less than perfect agreement. Normally, 0.8 and higher shows very good agreements.

In addition, the F1-Score is also used to evaluate the classified image. The F1-score (Equation (6)) is derived from the precision and the recall for each class, which are computed through Equations (4) and (5) respectively. It gives an indication of the classification performance per class:

$$\text{Precision} = \frac{N_{ii}}{\sum_{j=1}^{J} N_{ij}} \tag{4}$$

$$\text{Recall} = \frac{N_{jj}}{\sum_{i=1}^{J} N_{ij}} \tag{5}$$

$$\text{F1} - \text{score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

In this study, the validation of the classified images was carried out independently with the same suite of the validation samples. The statistic was computed by counting the number of pixels for each class.

4. Results and Analysis

4.1. Accuracy Analysis for the Different Experiments of Input Data

GF-1 WFV has only four bands, namely blue, green, red and near infrared. S2 has four bands at 10 m and six bands at 20 m. L8 has seven bands at 30 m. These bands are the basic features that may be used for the classification. It is already known that the classification accuracy may increase as more features are used as input data [5,25]. Thus, in this study, three experiments were designed to perform the crop type classification. The first one uses four GF-1 WFV available bands and four similar bands for S2 and L8. The second one uses four similar bands, with all calculated feature indices accordingly. The third one uses all available bands and all calculated feature indices accordingly for these three satellite data sources. Table 5 lists the spectral bands used and the number of increased features accordingly for GF-1, S2 and L8 in three experiments. The training datasets and the validation datasets are the same for all these experiments. The RF classifier algorithm is also the same. Because 10 S2 bands have two spatial resolutions, the datasets for S2 were prepared into two datasets, one with 10 m spatial resolution and another with 20 m spatial resolution. For the 10-m S2 datasets, the pixels in the window of 2 × 2 took the value of corresponding one pixel at 20 m. For the 20-m S2 datasets, the pixel at 20 m took the averaged value of the corresponding 2 × 2 pixels at 10 m.

Table 5. The used spectral bands and the number of increased features for GF-1, S2 and L8.

Experiment	Bands/Features	GF-1 WFV	S2A/B MSI	L8 OLI
1	Used Bands	485, 555, 660, 830	490, 560,665,865	482,562, 655, 865
	Increased Features	0	0	0
2	Used Bands	485, 555, 660, 830	490, 560,665,865	482,562, 655, 865
	Increased Features	6	6	6
3	Used Bands	485, 555, 660, 830	490, 560,665,705,740, 783, 842, 865,1610,2190	443, 482,562, 655, 865, 1609, 2200
	Increased Features	6	45	21

Figure 3 shows the accuracies of experiment 1. It is clear that the classification accuracy from the GF-1 WFV is lower than those from S2 and L8 in both years with regards to OA, Kappa and F1-score. The accuracy from the S2 data has the relatively highest score in 2017, while it is slightly lower than L8 in 2018 because of a lack of data after July in 2018. There was no difference found between the 10-m S2 and the 20-m S2 for both years.

In experiment 2, six feature indices were calculated in reference to Equation (1) and added as the input features. Figure 4 shows the accuracies of these three satellite data sources in experiment 2. It is clear that the classification accuracy from the GF-1 WFV data has the relatively lowest score for both years with regards of the OA, Kappa and F1-score. The accuracies of 10-m S2 and 20-m S20 are close in 2017, while it is slight larger in 2018. The accuracies between L8 and S2 are also close, though the one from S2 is a little bit better than L8 in 2017. Due to a lack of S2 data after July in 2018, it did not show the higher value of S2 in 2018 than L8, but both are very close. From the experiments 1 and 2, it is obvious that the accuracies arose for each input datasets as the input features increased.
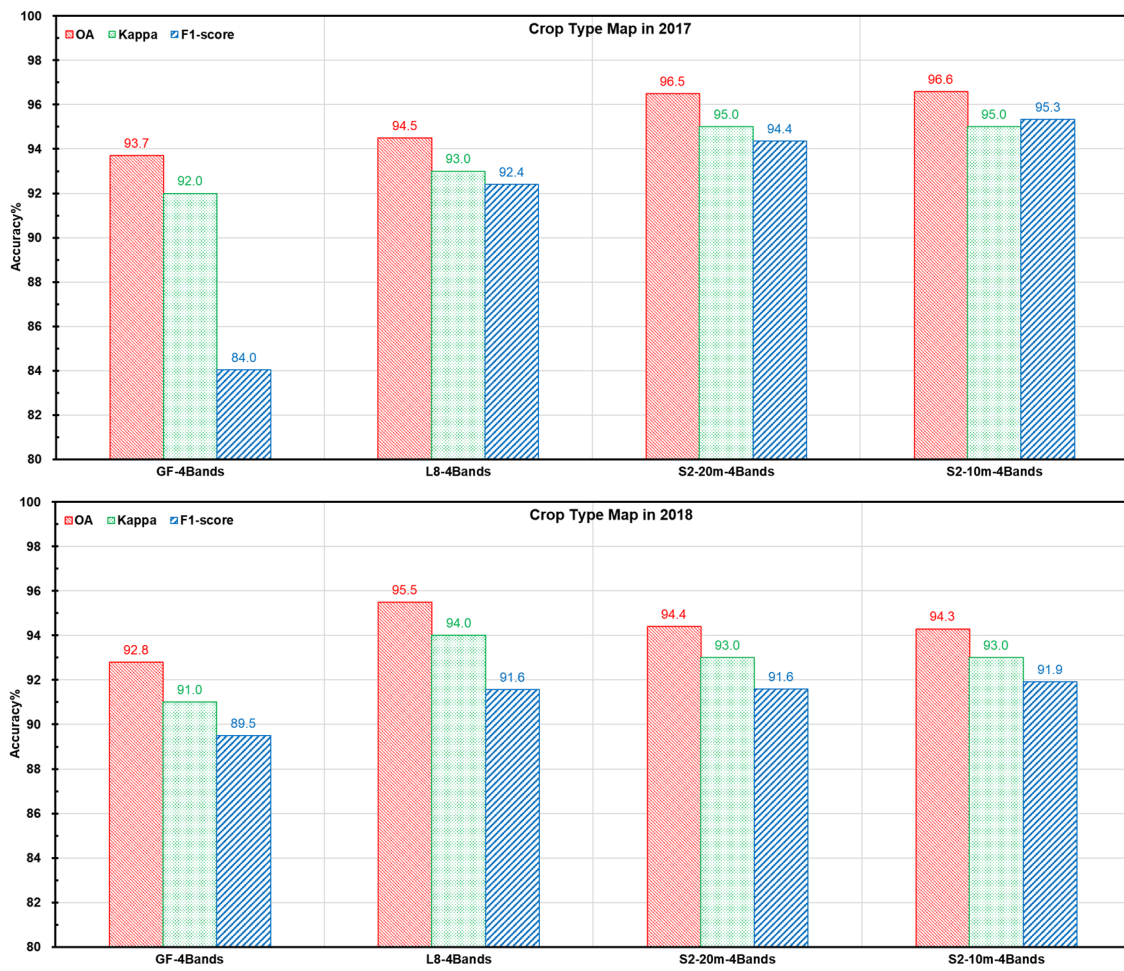
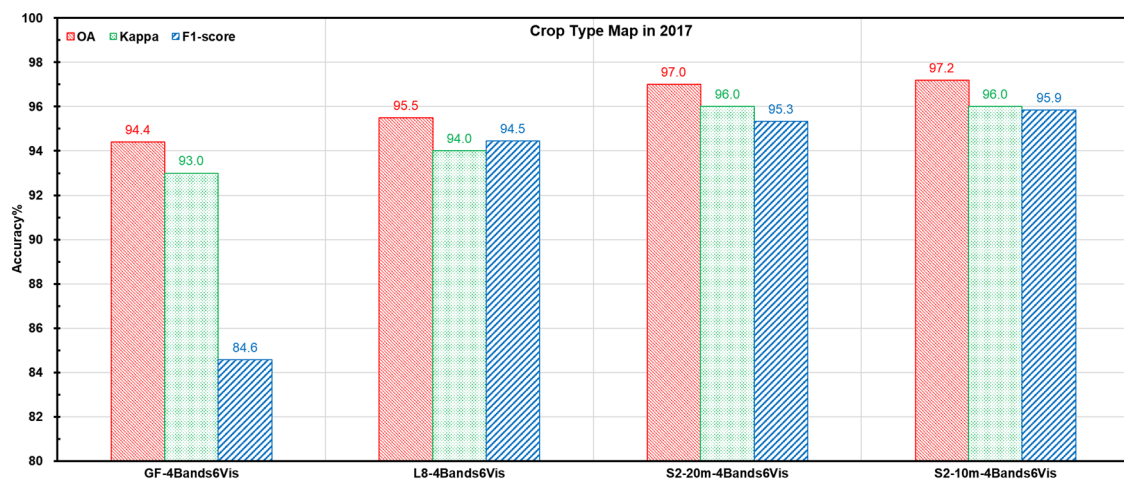**Figure 3.** The accuracies of experiment 1.
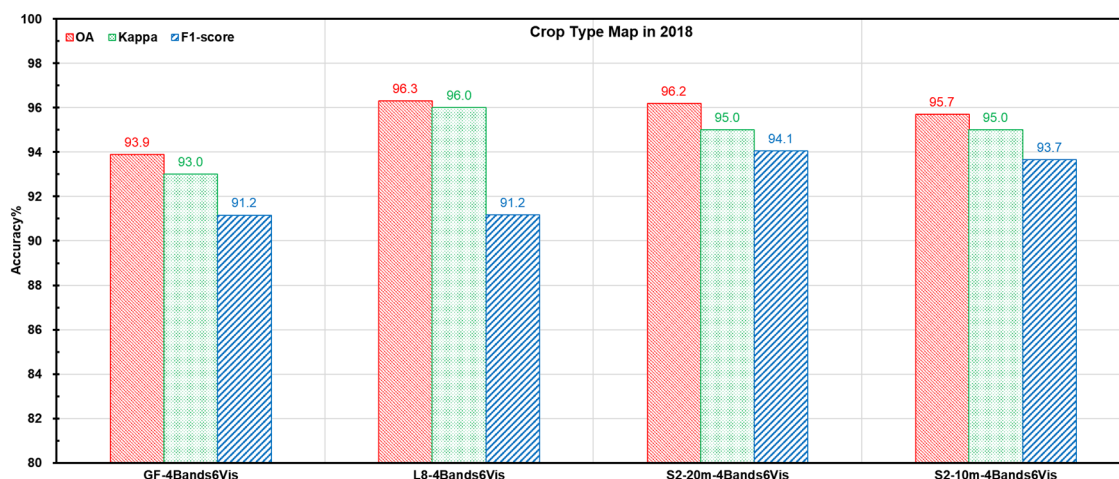

Figure 4. *Cont*.

Figure 4. The accuracies of experiment 2.

S2 and L8 have more bands available for the classification, and thus, all available bands were further used in the crop type classification. Figure 5 shows the accuracies of these three satellite data sources with the potential capability. Similar results were found in this experiment. The classification accuracy from the GF-1 WFV data always relatively has the lowest score. It is obvious that the accuracy from S2 is highest due to having the most bands available for the classification. The accuracy from 10-m S2 is a little bit better than the 20-m S2 but there is no big difference. The accuracies for S2 and L8 were increased a bit from four bands to all bands. Table 6 lists the commission and omission errors for the experiment 3.

Table 6. Commission and Omission errors for all crop types.

	GF 2017		GF 2018		L8 2017		L8 2018		S2 2017		S2 2018	
	Commission	Omission	Commission	Omission	Commission	Omission	Commission	Omission	Commission	Omission	Commission	Omission
Forage Grass	6.5	12.0	1.0	7.7	2.6	5.0	1.2	6.4	0.7	4.3	1.1	10.3
Corn	9.8	3.8	9.7	4.8	10.7	11.3	4.4	3.0	3.0	3.5	4.6	1.5
Grape	14.9	9.2	4.2	12.0	3.2	5.5	5.0	7.2	1.2	1.8	1.4	6.2
Greenhouse	2.1	1.2	2.3	0.7	2.9	1.6	2.1	0.9	2.0	0.1	1.9	0.3
Sward	36.1	39.2	0.0	24.2	2.1	6.0	25.0	52.6	3.1	10.9	38.1	8.1
Medlar	1.8	53.9	3.4	30.1	6.1	3.6	0.0	14.9	2.1	3.8	2.8	8.8
Rice	0.8	0.9	1.0	1.1	2.9	2.5	0.0	0.2	0.8	0.3	0.5	0.3
Vegetable	6.5	3.7	5.4	16.4	9.7	5.3	10.8	10.1	1.1	1.3	2.3	6.8
Wheat	32.3	23.0	17.0	8.8	7.2	11.1	5.1	2.3	9.5	3.3	7.4	5.3

Note: Table 6 lists the commission and omission errors for the experiment 3. Sward and Medlar are two problematic crop types due to being less represented in the region.

4.2. The Full Capability of Data Sources for the Crop Type Classification

As mentioned above, the highest accuracy will be achieved when all spectral bands and associated indices are involved in the classification as the input features. Figure 6 shows the best crop type maps from 15-m GF-1, 10-m S2 and 30-m L8, respectively. Table 7 shows the accuracy of OA, Kappa and F1-Score for the whole images and F1-Score for each crop type. The classified crop type maps with these three data sources are all visually acceptable, as the quantitative accuracies for the whole images are also high.

Nine types of crop were successfully classified in 2017 and 2018. The major crop types, such as rice, was classified with very high accuracy (F1-score of 98.3–100%). Forage Grass, corn, grape, greenhouse, and vegetable were always classified quite accurately (85% plus F1-score). With wheat only, Sward and Medlar obtained low accuracy in 2017 with GF data. Sward obtained low accuracy in 2018 with S2 data. Wheat was a major cereal in this region, but is now decreasing due to the cultivation cost. Sward and Medlar are cultivated in the farmland but not the major crop type. The training samples for these three types was difficult to collect and the final number of samples was low. The low accuracy is most likely related to the appearance of cloudy images and low availability of the training samples.
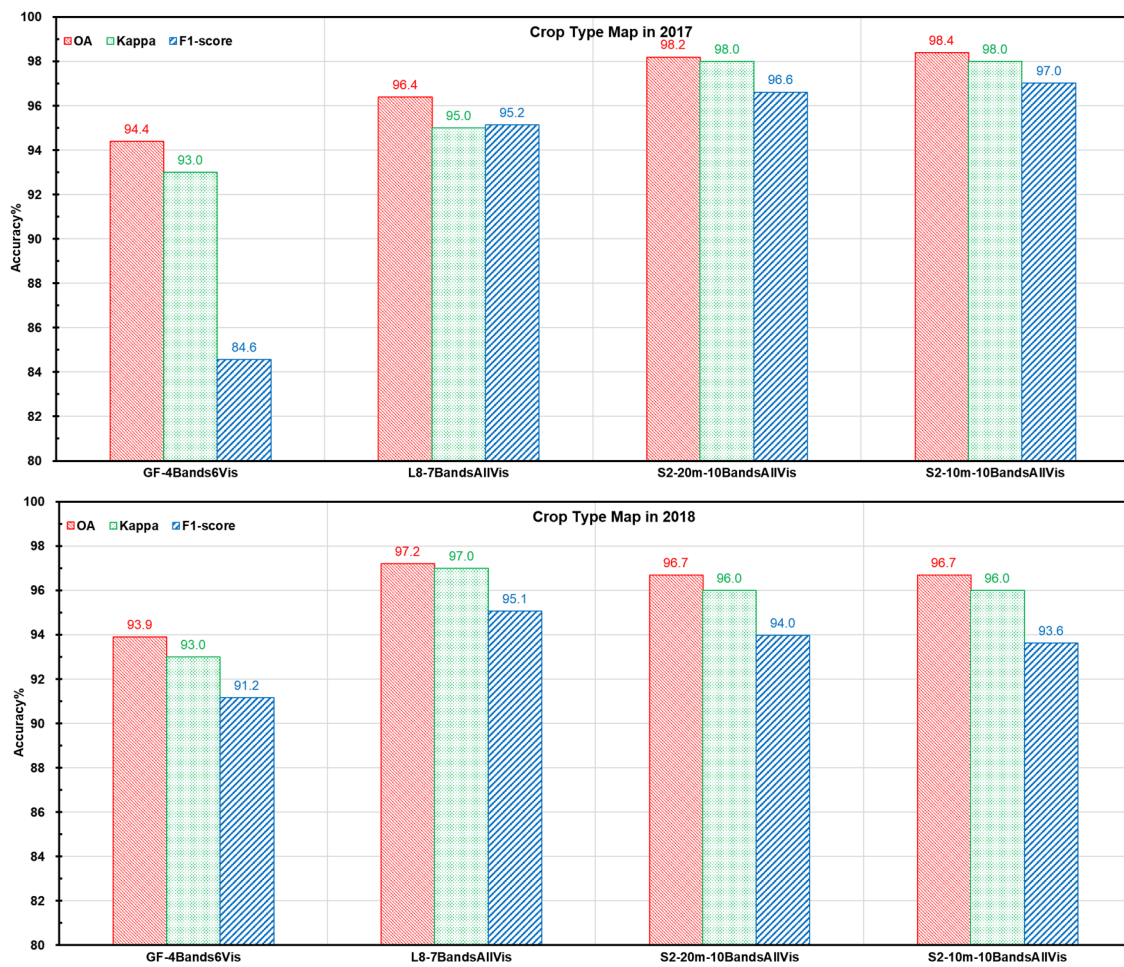
Figure 5. The accuracies of experiment 3.

Table 7. Accuracies for all crop types.

|	GF-2017	L8-2017	S2-2017	GF-2018	L8-2018	S2-2018
OA	94.4	96.4	98.4	93.9	97.2	96.7
Kappa	93.0	95.0	98.0	93.0	97.0	96.0
F1-score	84.6	95.2	97.0	91.2	95.1	94.0
Forage Grass	90.7	95.1	97.5	95.5	97.5	94.1
Corn	93.1	92.2	96.7	92.7	97.5	96.9
Grape	87.9	96.6	98.5	91.7	94.6	96.1
Greenhouse	98.3	98.2	98.9	98.5	99.5	98.9
Sward	62.3	95.0	92.8	86.2	86.6	74.0
Medlar	62.7	96.0	97.0	81.1	93.0	94.1
Rice	99.1	98.3	99.4	98.9	100.0	99.6
Vegetable	94.9	94.6	98.8	88.8	90.7	95.4
Wheat	72.1	90.3	93.5	86.9	96.2	95.4
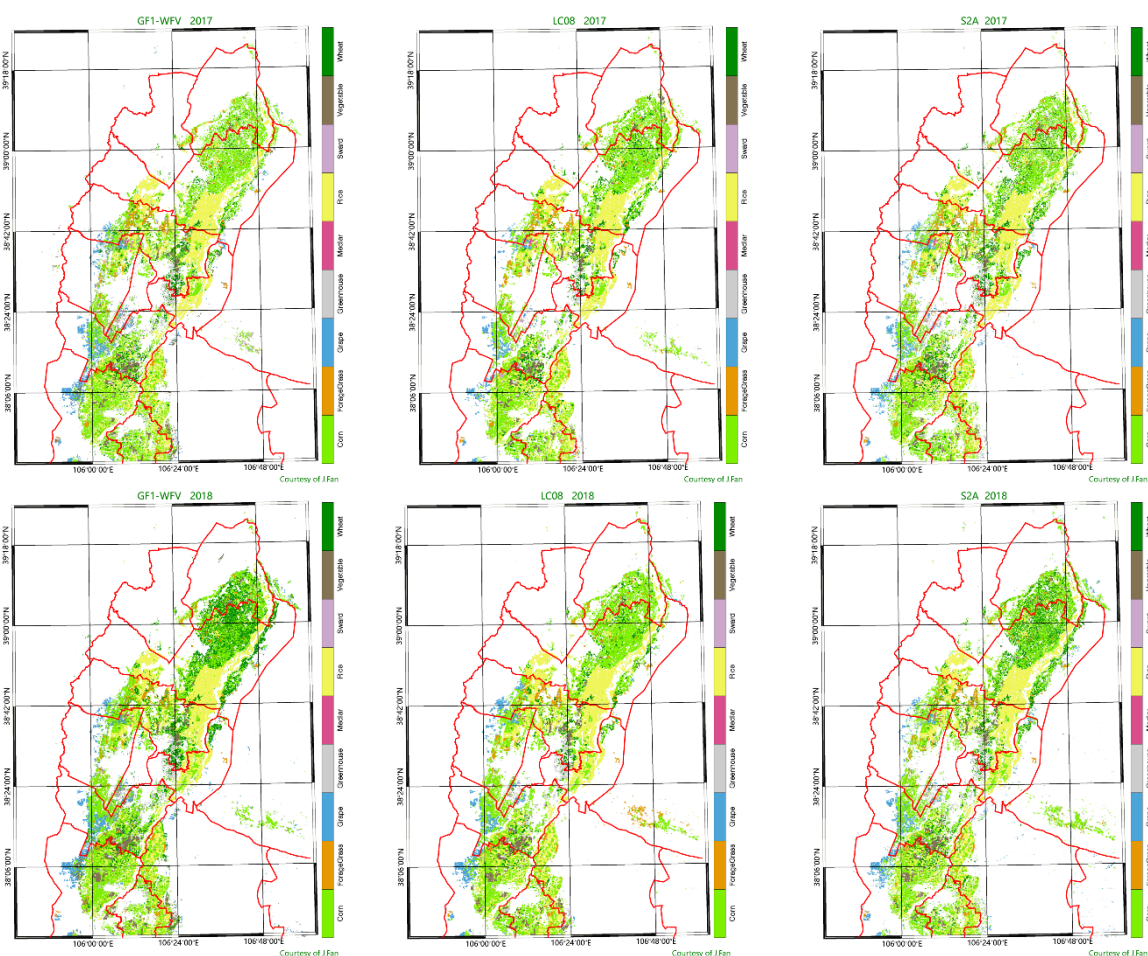
Figure 6. The full capability of crop type map of three satellite data sources.

5. Discussions

Training datasets, input data and classification algorithm are three key aspects for the satellite image classification. In comparison with the other two, the training dataset is the core and decides the quality of output result fundamentally, as the classification model is set up with the training datasets completely. The spatial distribution and the statistical distribution of the training datasets for each crop type should be taken into account. Collecting the field samples is the first step to prepare for the training datasets. The field sampling is of importance and may help the expert to understand the features on the image first and then teach the computer to identify and learn features correctly but it is not enough if only field samples are being used as the training data for the classification. In reality, it is not able to perform the field data collection completely following the route planned at home due to the unexpected road work and other disconnection problems. Therefore, adding and optimizing of the training datasets spatially and evenly with the support of images should be carried out. Adding training samples may improve it dramatically in case that a lack of training samples in some area obviously creates the incorrect classification result.

However, a set of good training datasets is also related to the field size and crop cultivation practice. In this study area, the size of each field owned by a farmer is small and normally a few meters by a few tens of meters; however, in practice, the same crop is planted in consecutive fields, which makes it easy to identify crop types on the Landsat 8 image. This advantage helps the application of 10–30-m resolution satellite data in this area. In other parts of China, the small size field hampers the wide application of these satellite data for the crop type classification. The same situation may be found in Africa due to small agricultural fields and a heterogeneous landscape. The 5-m optical and 1-m SAR data were fused for the crop classification, with an accuracy of up to 75% in west Africa [56].

One study [57] on crop classification in Madagascar and Burkina Faso also showed low performances with an accuracy of around 50% due to the presence of fields smaller than the pixel size and a mix of trees and crops in the fields. Hence, the classification system on crop group rather than crop type may improve the classification accuracy with these free satellite images. Otherwise, it had better use a higher resolution commercial satellite image.

Data volume should also be taken in account. In this study, the data volume for GF-1 and Landsat 8 was still easily handled, but the data volume for S2 was extremely large. Although the training datasets is the same for these three data sources from the point view of polygons, the training datasets used for S2 in pixel units is nine times that for Landsat 8. Thus, it has to set a threshold to select some pixels in each polygon and keep the total volume of training pixels at an acceptable level. Otherwise, it takes a very long time to build up the model. The data volume of the input data with the additional features also increases. In this study, GF-1 was increased six additional features and 10 bands were finally used as the input features for each single date image. Landsat 8 was increased 21 additional features and finally used 28 bands as the input features. S2 was increased 45 additional features and finally used 55 bands as the input features. Regarding the data volume for a single date image, the volume for 10-m S2 data is about 18 times that of the L8 data, while the volume for GF-1 data is about 1.4 times of L8 data. Therefore, the final input data volume for S2 is too large to process properly with the limited capacity of some computer as the data volume for S2 used in this study was approximately 1 Terabyte. Therefore, the reduction of input features may be carried out, applying the methods of the principal component decomposition and selecting the main component of input datasets, if the computer is not able to handle these datasets.

Undoubtedly, the GF-1 WFV still has a geometric problem after it is orthorectified with image chip matching. This is the reason why the classification accuracy from GF-1 is still 3–4% lower than that from S2 or L8, which are in very good geometric condition under the same spectral circumstances. Clouds are another big issue for crop type mapping. Sometimes it is difficult to get sufficient images even with less cloudy coverage. It will do nothing if there are no satellite data available. Optimistically, it increases the data availability if these three kinds of satellite data are combined and fused in the near future in a way like Harmonized Landsat Sentinel-2 [38].

6. Conclusions

A case study of the evaluation of individual capability for crop type mapping from difference satellite data sources, such as GF-1, S2 and L8, was carried out in the yellow river irritation area in Ningxia Hui autonomous region of Northwest China. It revealed that the crop type mapping with any of these three kinds of satellite data may achieve the acceptable accuracy since the lowest OA may reach 94%. The relatively lowest accuracy from the GF-1 WFV data is the consequence of the limited spectral bands. It proved again that the accuracy will increase as the input features increase. There is no obvious difference between the 10-m S2 data and 20-m S2 data for the crop type mapping, but, indeed, S2 may achieve the highest accuracy with 10 available bands for the crop type mapping at either 10 m or 20 m. The accuracy for L8 is in the middle, but the difference is small in comparison with S2. In order to achieve the highest accuracy, all special bands and associate indices should be jointly used as the input features for the classification. The training datasets may be tuned and improved by expert knowledge with the field samples. The evenly spatial and statistical distribution of the training sample is of importance for training datasets. Classifications based on individual image may produce the results with various accuracies due to the limited signals for differentiating the crop types on each image. The time series of satellite images used altogether as the input will bring the highest accuracy as all available information during the whole growing season are used in the classification.

Author Contributions: Conceptualization, J.F. and P.D.; Methodology, J.F., X.Z. and Z.Q.; Software, J.F. and C.Z.; Validation, M.D.V.; Data processing, C.Z.; Writing, J.F. and Z.Q.; Comments, P.D. All authors have read and agreed to the published version of the manuscript.

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

References

1. Zhao, S.H.; Wang, Q.; Yang, Y.P.; Zhu, L.; Wang, Z.T.; Jiang, D. The demonstration research of GF-1 satellite data monitoring environment application. *Satell. Appl.* **2015**, *3*, 37–40.
2. ESA Sentinels—Sentinel-2: Operations Ramp-up Phase. Available online: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/operations-ramp-up-phase (accessed on 10 December 2020).
3. USGS. *Landsat Missions*. Available online: https://www.usgs.gov/core-science-systems/nli/landsat (accessed on 28 February 2021).
4. ESA. *Sentinel-2 User Handbook*; ESA: Paris, France, 2015.
5. *Landsat 8 (L8) Data Users Handbook Version 5.0*; USGS: Sioux Falls, SD, USA, 2019.
6. Reynolds, C.A.; Yitayew, M.; Slack, D.C.; Hutchinson, C.F.; Huete, A.; Petersen, M.S. Estimating crop yields and production by integrating the FAO crop specific water balance model with real-time satellite data and ground-based ancillary data. *Int. J. Remote Sens.* **2000**, *21*, 3487–3508. [CrossRef]
7. Jinlong, F.; Mingwei, Z. Supporting Agricultural Monitoring in APEC with FengYun Satellite data. In Proceedings of the Workshop on the Application of Remote Sensing and GIS on Crops Productivity among APEC Economies, Beijing, China, 30–31 July 2012.
8. Sakamoto, T.; Gitelson, A.A.; Arkebauer, T.J. Near real-time prediction of U.S. corn yields based on time-series MODIS data. *Remote Sens. Environ.* **2014**, *147*, 219–231. [CrossRef]
9. Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* **2019**, *221*, 551–568. [CrossRef]
10. Shelestov, A.; Lavreniuk, M.; Kussul, N.; Novikov, A.; Skakun, S. Large Scale Crop Classification Using Google Earth Engine Platform. In Proceedings of the 2017 IEEE International Geoscience & Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017.
11. Xiong, J.; Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Poehnelt, J.; Congalton, R.G.; Yadav, K.; Thau, D. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 225–244. [CrossRef]
12. Lunetta, R.S.; Johnson, D.M.; Lyon, J.G.; Crotwell, J. Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sens. Environ.* **2004**, *89*, 444–454. [CrossRef]
13. Carrão, H.; Gonçalves, P.; Caetano, M. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sens. Environ.* **2008**, *112*, 986–997. [CrossRef]
14. Griffiths, P.; Nendel, C.; Hostert, P. Intra-annual reflectance composites from sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* **2019**, *220*, 135–151. [CrossRef]
15. Bauer, M.E.; Cipra, J.; Anuta, P.E.; Etheridge, J.B. Identification and Area Estimation of Agricultural Crops by Computer Classification of LANDSAT MSS Data. *Remote Sens. Environ.* **1979**, *8*, 77–92. [CrossRef]
16. Badhwar, G.B. Automatic corn-soybean classification using Landsat MSS data. II. Early season crop proportion estimation. *Remote Sens. Environ.* **1984**, *14*, 31–37. [CrossRef]
17. Conrad, C.; Machwitz, M.; Schorcht, G.; Löw, F.; Fritsch, S.; Dech, S. Potentials of Rapid Eye time series for improved classification of crop rotations in heterogeneous agricultural landscapes: experiences from irrigation systems in Central Asia. *SPIE Remote Sens.* **2011**, *8174*, 817412. [CrossRef]
18. Liu, M.W.; Ozdogan, M.; Zhu, X. Crop type classification by simultaneous use of satellite images of different resolutions. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3637–3649. [CrossRef]
19. Fan, X.S.; Xu, W.B.; Fan, J.L. Mapping winter wheat growing areas in the North China Plain with FY-3 250 m resolution data. *J. Remote Sens.* **2015**, *19*, 586–593. [CrossRef]
20. Fan, J. Updates on the FY-3 MERSI satellite data processing and delivering. In Proceedings of the SIGMA Project Annual Meeting, Brussels, Belgium, 9–13 November 2015.
21. Hao, P.; Wang, L.; Zhan, Y.; Niu, Z.; Wu, M. Using historical NDVI time series to classify crops at 30m spatial resolution: A case in Southeast Kansas. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 6316–6319. [CrossRef]

22. Walter, V. Object-based classification of remote sensing data for change detection. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 225–238. [CrossRef]

23. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]

24. Mao, D.; Wang, Z.; Du, B.; Li, L.; Tian, Y.; Jia, M.; Zeng, Y.; Song, K.; Jiang, M.; Wang, Y. A new product resulting from object-based and hierarchical classification of Landsat 8 OLI images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 11–25. [CrossRef]

25. Liu, D.; Xia, F. Assessing object-based classification: Advantages and limitations. *Remote Sens. Lett.* **2010**, *1*, 187–194. [CrossRef]

26. Lobell, D.B.; Asner, G.P. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* **2004**, *93*, 412–422. [CrossRef]

27. Verbeiren, S.; Eerens, H.; Piccard, I.; Bauwens, I.; van Orshoven, J. Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 486–497. [CrossRef]

28. Westbrook, J.; Hung, I.; Unger, D.; Zhang, Y. *Sub-Pixel Classification of Forest Cover Types in East Texas*; Faculty Publications, 2012; p. 5. Available online: https://scholarworks.sfasu.edu/spatialsci/5 (accessed on 28 February 2021).

29. Durgun, Y.Ö.; Gobin, A.; Van De Kerchove, R.; Tychon, B. Crop Area Mapping Using 100-m Proba-V Time Series. *Remote Sens.* **2016**, *8*, 585. [CrossRef]

30. Xu, X.; Li, J.; Li, S.; Plaza, A. Subpixel Component Analysis for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5564–5579. [CrossRef]

31. Osman, J.; Inglada, J.; Dejoux, J.-F. Assessment of a Markov logic model of crop rotations for early crop mapping. *Comput. Electron. Agric.* **2015**, *113*, 234–243. [CrossRef]

32. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]

33. Pelletiera, C.; Valeroa, S.; Ingladaa, J.; Championb, N.; Dedieua, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [CrossRef]

34. Fan, J.; Defourny, P.; Dong, Q.; Zhang, X.; De Vroey, M.; Belleman, N.; Xu, Q.; Li, Q.; Zhang, L.; Gao, H. Sent2Agri System Based Crop Type Mapping in Yellow River Irrigation Area. *J. Geod. Geoinf. Sci.* **2020**, *3*, 110–117. [CrossRef]

35. Blaes, X.; Vanhalle, L.; Defourny, P. Efficiency of crop identification based on optical and SAR image time series. *Remote Sens. Environ.* **2005**, *96*, 352–365. [CrossRef]

36. Zhao, W.; Qu, Y.; Chen, J.; Yuan, Z. Deeply synergistic optical and SAR time series for crop dynamic monitoring. *Remote Sens. Environ.* **2020**, *247*, 111952. [CrossRef]

37. Orynbaikyzy, A.; Gessner, U.; Mack, B.; Conrad, C. Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. *Remote Sens.* **2020**, *12*, 2779. [CrossRef]

38. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.-C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]

39. Chaves, M.E.D.; Picoli, M.C.A.; Sanches, I.D. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sens.* **2020**, *12*, 3062. [CrossRef]

40. Skakun, S.; Vermote, E.; Roger, J.C.; Franch, B. Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. *AIMS Geosci.* **2017**, *3*, 163–186. [CrossRef]

41. Zhan, Z.; Yu, L.; Li, Z.; Ren, L.; Gao, B.; Wang, L.; Luo, Y. Combining GF-2 and Sentinel-2 Images to Detect Tree Mortality Caused by Red Turpentine Beetle during the Early Outbreak Stage in North China. *Forests* **2020**, *11*, 172. [CrossRef]

42. Nguyen, M.D.; Baez-Villanueva, O.M.; Bui, D.D.; Nguyen, P.T.; Ribbe, L. Harmonization of Landsat and Sentinel 2 for Crop Monitoring in Drought Prone Areas: Case Studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon). *Remote Sens.* **2020**, *12*, 281. [CrossRef]

43. 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]

44. NERC; ESRC; DFID. *China Ecosystem Services and Poverty Alleviation Situation Analysis and Research Strategy-Ningxia Case Study Ningxia Statistics Bureau, 2017, Ningxia Statistical Yearbook*; China Statistics Press: Beijing, China, 2017.

45. Wei, J.; Guojin, H.; Huichan, L.; Tengfei, L.; Wei, W.; Shouzhu, Z.; Xiaoxiao, M. Research on China's land image mosaicking and mapping technology based on GF-1 satellite WFV data. *Remote Sens. Land Resour.* **2017**, *29*, 190–196. [CrossRef]

46. Liu, J.; Wang, L.; Yang, L.; Teng, F.; Shao, J.; Yang, F.; Fu, C. GF-1 satellite image atmospheric correction based on 6S model and its effect. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 159–168.

47. Sentinel-2 for Agriculture. Available online: http://www.esa-sen2agri.org/ (accessed on 10 December 2020).

48. Orfeo ToolBox. Available online: https://www.orfeo-toolbox.org/ (accessed on 10 December 2020).

49. Martins, V.S.; Barbosa, C.C.F.; de Carvalho, L.A.S.; Jorge, D.S.F.; Lobo, F.L.; Novo, E.M.L.M. Assessment of Atmospheric Correction Methods for Sentinel-2 MSI Images Applied to Amazon Floodplain Lakes. *Remote Sens.* **2017**, *9*, 322. [CrossRef]

50. Prieto-Amparan, J.A.; Villarreal-Guerrero, F.; Martinez-Salvador, M.; Manjarrez-Domínguez, C.; Santellano-Estrada, E.; Pinedo-Alvarez, A. Atmospheric and Radiometric Correction Algorithms for the Multitemporal Assessment of Grasslands Productivity. *Remote Sens.* **2018**, *10*, 219. [CrossRef]

51. Bégué, A.; Arvor, D.; Bellón, B.; Betbeder, J.; de Abelleyra, D.; Ferraz, R.; Lebourgeois, V.; Lelong, C.; Simoes, M.; Verón, S. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99. [CrossRef]

52. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
53. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]
54. Pal, M. Random Forests for Land Cover Classification. *Proc. Int. Geosci. Remote Sens. Symp.* **2003**, *2003*, 3510–3512.
55. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **2006**, *6*, 21–45. [CrossRef]
56. Forkuor, G.; Conrad, C.; Thiel, M.; Ullmann, T.; Zoungrana, E. Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa. *Remote Sens.* **2014**, *6*, 6472–6499. [CrossRef]
57. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* **2015**, *7*, 12356–12379. [CrossRef]