

Article

AUTS: A Novel Approach to Mapping Winter Wheat by Automatically Updating Training Samples Based on NDVI Time Series

Chunyang Wang ^{1,2} , Huan Zhang ², Xifang Wu ^{2,*}, Wei Yang ³ , Yanjun Shen ⁴, Bibo Lu ¹ and Jianlong Wang ¹ 

- ¹ College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China; wcy@hpu.edu.cn (C.W.); lubibo@hpu.edu.cn (B.L.); wangjianlong24@hpu.edu.cn (J.W.)
- ² School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; zhanghuan@home.hpu.edu.cn
- ³ Center for Environmental Remote Sensing, Chiba University, Chiba 2638522, Japan; yangwei@chiba-u.jp
- ⁴ Key Laboratory of Agricultural Water Resources, Hebei Laboratory of Water-Saving Agriculture, Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), Shijiazhuang 050021, China; yjshen@sjziam.ac.cn
- * Correspondence: wxf2020@hpu.edu.cn

Abstract: Accurate and rapid access to crop distribution information is a significant requirement for the development of modern agriculture. Improving the efficiency of remote sensing monitoring of winter wheat planting area information, a new method of automatically updating training samples (AUTS), is proposed herein. Firstly, based on the Google Earth Engine (GEE) platform, a Sentinel-2 image with a spatial resolution of 10 m was selected to extract the distribution map of winter wheat in the city of Shijiazhuang in 2017. Secondly, combined with the NDVI time series, the weighted correlation coefficients from 2017, 2018, and 2019 were calculated. Then, the 2017 winter wheat distribution map and its most significant relevant areas were used to extract sample points from 2018 and 2019 automatically. Finally, the distribution map of winter wheat in Shijiazhuang in 2018 and 2019 was generated. In addition, to test the applicability of the automatically updating training sample at different scales and regions, the proposed method was applied to Landsat 8 image data with a spatial resolution of 30 m, as well as to Handan and Baoding. The results showed that the calculated winter wheat planting area is comparable with the officially published statistics, based on Sentinel-2, extracting three years of winter wheat, the R^2 values for all three years were above 0.95. The R^2 values for 2018 and 2019, based on Landsat 8 extractions, were 0.95 and 0.90, respectively. The R^2 values extracted from Handan and Baoding in 2018 were 0.94 and 0.86, respectively. These results indicate that the proposed method has high accuracy and can provide technical support and reference for winter wheat area monitoring and yield estimation.

Keywords: winter wheat; random forests; training sample updated; Sentinel-2; Google Earth Engine (GEE)



Citation: Wang, C.; Zhang, H.; Wu, X.; Yang, W.; Shen, Y.; Lu, B.; Wang, J. AUTS: A Novel Approach to Mapping Winter Wheat by Automatically Updating Training Samples Based on NDVI Time Series. *Agriculture* **2022**, *12*, 817. <https://doi.org/10.3390/agriculture12060817>

Academic Editor: Koki Homma

Received: 12 April 2022

Accepted: 3 June 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Winter wheat is one of the most important food crops in China, and the monitoring of winter wheat planting area is an issue of great concern to the country. China is the world's largest wheat producer, accounting for 11% of the world's acreage and 17.7% of its total production [1]. With the increasing population in China, the demand for food is also increasing day-by-day. Obtaining food distribution information quickly and accurately has become a significant requirement for the development of modern agriculture. Remote sensing technology has been widely used for land use/cover on a global and regional scale [2]. However, the traditional use of remote sensing analysis software requires a large amount of data; it is slow, and the engineering volume for winter wheat area extraction based on long time series data is too large, so using the Google Earth Engine (GEE) cloud

platform is a better choice. GEE is the world's most advanced cloud platform for remote sensing big data processing, which not only has efficient computing power, but also combines a large number of public geospatial datasets, including many free image resources (such as Landsat, MODIS, and Sentinel), meteorological data, land use/cover data, and population distribution data [3]. Using GEE, it is easy to conduct studies on vegetation monitoring [4,5], land use/cover [6,7], moisture changes [8], and drought analysis [9]. In addition, many machine learning algorithms are built into GEE, and have been important methods for crop mapping and land use/cover [10–12].

Scientific researchers in remote sensing have conducted some research in the field of remote sensing based on the GEE cloud platform [13]. Zhang et al. [14] used the GEE cloud platform's random forest machine learning classifier to analyze the temporal and spatial dynamics of winter wheat planting areas in the North China Plain from 1999 to 2019. Hadi et al. [15] used Landsat data and the GEE cloud platform to estimate the yields of crops such as potatoes, wheat, corn, and more in Brazil, Spain, and elsewhere. Based on the GEE cloud platform, Fang et al. [16] carried out winter wheat identification and mapping from Sentinel-2 data, with a resolution of 10 m, using support vector machines (SVMs), random forests (RF), classification and regression trees (CART), and classification and regression trees (CART), and the sensitivity and performance of these three classification methods were further analyzed. Li et al. [17] used the GEE cloud platform, combined with synthetic-aperture radar (SAR) and Sentinel-2 optical images, to extract the planting area of winter wheat in Henan Province via a random forests algorithm, and obtained good accuracy. Han et al. [18] integrated climate data, remote sensing data, and soil data, comparing eight machine learning models for predicting winter wheat yields based on the GEE cloud platform, and found that support vector machines, Gaussian process regression, and random forests were the three best methods for predicting yields. Xu et al. [19] used the GEE cloud platform to combine Landsat 8 Operational Land Imager (OLI) and Sentinel-2 Multispectral Imager (MSI) data, in order to use random forest classification to map the distribution of winter wheat in Shandong Province, China, with an overall accuracy of 93.4%. However, most supervised classification methods rely heavily on training samples. Still, the collection of training samples is very time-consuming and laborious, which is a significant drawback of large-area classification [20,21]. Some studies recommend that existing training data be used for land-use classification, but historical training data may not be directly available, due to changing human activities and natural processes [22,23]. In previous studies, unsupervised methods did not rely on training data and had limited applications in crop mapping, due to poor performance [24]. Extracting crop distribution maps without collecting training samples is usually based on phenological algorithms [25,26]. Considering the trade-off between spatial and temporal resolution, complex multitemporal information is often obtained from sensors with coarse spatial resolution, such as moderate-resolution MODIS [27–29]. Nevertheless, small and fragmented crop fields with areas of less than 2.56 ha are widely distributed worldwide and will produce a large number of mixed pixels in coarse-resolution images [30,31]. However, few people adopt effective training sample generation strategies for regional winter wheat mapping. If high-quality training samples can also be automatically updated and reused on the GEE, crop types could be automatically and repeatedly mapped, and the mapping cost would be significantly reduced [32].

In this study, we propose a new method of automatically updating training samples (AUTS) that only needs to collect one year of training samples to automatically extract multiyear training samples and, thus, extract multiyear winter wheat distribution maps. The research goals of this article were as follows: (1) The distribution map of winter wheat in the city of Shijiazhuang in 2017 was extracted by using Sentinel-2 imagery combined with a random forest classification algorithm. AUTS is proposed in combination with weighted correlation analysis. Training samples from 2018 and 2019 were extracted. (2) The distribution map of winter wheat in Shijiazhuang in 2018 and 2019 was extracted based on AUTS combined with a random forest classification algorithm based on Sentinel-2 imagery.

(3) The accuracy was evaluated through visual interpretation, verification, and comparison with statistical data, and we assessed the feasibility of AUTS based on the GEE platform in extracting the winter wheat planting area. (4) By applying AUTS to Landsat 8 image data, with a spatial resolution of 30 m, the training samples were automatically obtained. The distribution map of winter wheat in Shijiazhuang was generated. The training samples extracted from Shijiazhuang were used as samples to extract winter wheat in Handan and Baoding to the north and south of Shijiazhuang, respectively, in order to study the applicability of the automatic sample upgrade algorithm proposed by us at different scales and in different regions.

2. Materials

2.1. Study Area

As shown in Figure 1, the research area is mainly located in the city of Shijiazhuang, Hebei Province, and the two cities of Handan and Baoding, which are to the north and south of Shijiazhuang, respectively. Shijiazhuang is located in the south–central part of Hebei Province, bordered by Hengshui to the east, Xingtai to the south, Shanxi Province to the west, and Baoding to the north and located between $37^{\circ}27'$ and $38^{\circ}47'$ N latitude and $113^{\circ}30'$ and $115^{\circ}20'$ E longitude. Shijiazhuang spans two central geomorphological units in the Taihang Mountains and the North China Plain. The western part is located in the middle of the Taihang Mountains, with an altitude of about 1000 m and amidst towering terrain. The eastern part is the Hutuo River Alluvial Plain, and the east of the Beijing–Guangzhou Railway is part of the North China Plain. The terrain is low in the southeast and high in the northwest, with significant gaps and complex landforms. Shijiazhuang often experiences heavy snow in winter; although this heavy snow brings some inconvenience to traffic, it plays a protective role in wheat overwintering, providing favorable moisture conditions for spring rejuvenation. The region has a temperate monsoon climate. The seasonal solar radiation changes are significant, high- and low-pressure activities on the ground are frequent, four seasons are distinct, winter and summer are clear, and rainfall is concentrated in summer and autumn. The total precipitation is 401.1–752.0 mm, and the average annual temperature is 13.5°C . Natural conditions, such as water and fertilizer, are good—suitable for winter wheat planting. The entire growth period of winter wheat in the study area is usually from early October to mid-June of the following year.

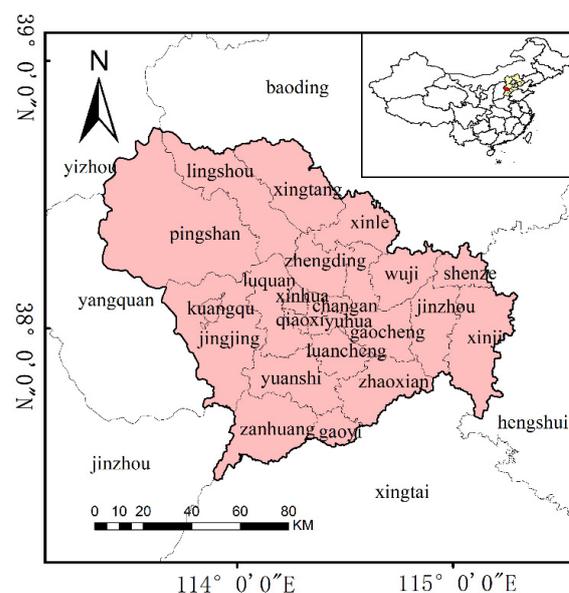


Figure 1. Geographical location and administrative division of the study area.

2.2. Data Sources

Sentinel-2 is an optical satellite developed by the European Space Agency (ESA). The Shijiazhuang Sentinel-2 remote sensing image set is directly based on the GEE cloud platform. The product level is Level-1C (L1C) class, with 13 spectral bands and a spatial resolution of 10 m. According to the official Sentinel-2 image data description, L1C grade products are images that have been processed with orthophoto correction and atmospheric apparent reflectivity. Previous studies have shown that, if the training data and images are classified at the same scale, the image classification does not require atmospheric correction [33]. This study selected all images of winter wheat growth periods in 2017, 2018, and 2019 for subsequent processing. In addition to Sentinel-2, Landsat 8 images of winter wheat growth periods in 2018 and 2019 were used to study the suitability of AUTS at different scales. In addition, the statistics of the winter wheat sowing area were derived from National Bureau of Statistics, which we downloaded from the China National Knowledge Infrastructure (CNKI).

We downloaded the wheat area statistics of the cities of Shijiazhuang, Handan, and Baoding from the Hebei statistical yearbook; since some county-level districts had no data for some years, the practical datasets for each city were used to contain totals of 21, 18, and 17 county-level administrative districts, respectively. Vector data include Shijiazhuang, Handan, Baoding, and county-level administrative boundaries within each city.

3. Methodology

The technical route of this study mainly includes five parts: the pre-processing of remote sensing images; establishing the sample set of the city of Shijiazhuang in 2017; constructing Normalized Difference Vegetation Index (NDVI) time series curves using mean filtering; and extracting the winter wheat area of Shijiazhuang in 2017 using a random forest classification algorithm; combined with correlation analysis, AUTS is proposed. Combined with AUTS and the random forest classification algorithm, the winter wheat areas of Shijiazhuang in 2018 and 2019 were extracted.

3.1. Remote Sensing Image Pre-Processing

Firstly, the QA60 band was used, and the cloud mask was used to remove the cloud information in the image, so that it would no longer participate in subsequent calculations, in order to ensure the image quality. Secondly, the Shijiazhuang vector boundary was used to clip the study area; after that, the NDVI was calculated, and the NDVI time series curve was constructed by selecting the maximum value of the NDVI half-month from October to June of the following year. The mean filtering was performed twice. The above process was written in the GEE cloud platform for code processing.

3.2. Established Shijiazhuang 2017 Sample Set

We merged the feature categories of the study area into three categories: winter wheat, non-wheat vegetation, and non-vegetation (water, roads, and buildings). The 2017 training sample set was collected in the Red, Green, Blue (RGB) map of high-resolution satellite imagery provided by the Google Earth map. The numbers of samples collected for winter wheat, non-wheat vegetation, and non-vegetation (water, roads, buildings) were 5784, 5017, and 14,445, respectively.

3.3. NDVI Time Series Feature Construction

3.3.1. NDVI Time Series

To avoid the phenomenon of homologous foreign bodies, the entire growth period of winter wheat in the study area was selected, and the NDVI half-month maximum value synthesis was carried out from October to June of the following year to construct a time series [34]. The normalized difference vegetation index (NDVI), as an index reflecting

vegetation growth status and cover, increases or decreases in value corresponding to changes in the crop growth process [35]. Its calculation formula is as follows:

$$\text{NDVI} = (\rho_{\text{NIR}} - \rho_{\text{R}}) / (\rho_{\text{NIR}} + \rho_{\text{R}}) \quad (1)$$

where ρ_{NIR} is the reflectance of the near-infrared band, and ρ_{R} is the reflectance of the red band. In the GEE, NDVI is calculated according to the NDVI index calculation formula, and the half-month maximum value of each cell's NDVI is calculated. The NDVI half-month maximum image was combined into a time series image dataset to construct the NDVI time series curve in the study area.

3.3.2. Mean Filtering

Since NDVI time series data are affected by clouds and other noise, producing jagged fluctuations, in order to remove the noise effects and smooth the time series curves, we adopted the method of mean filtering twice to effectively remove the noise effects. Mean filtering is a typical linear filtering algorithm, and the primary method is neighborhood averaging, which means that the target pixel and surrounding pixels are averaged and then filled back to the target pixel to achieve the filtering purpose. Its calculation formula is as follows:

$$\text{NDVI}_i = \frac{\text{NDVI}_{i-1} + \text{NDVI}_i + \text{NDVI}_{i+1}}{3} \quad (2)$$

where NDVI_i is the NDVI at the i th time point, and NDVI_i is the NDVI at the i th time point after mean filtering.

3.4. Random Forest Classification Algorithm

Image classification methods have been an important topic in remote sensing research, and different classification methods can be selected to obtain different classification results. The random forests algorithm is a machine learning algorithm that integrates multiple decision trees through the idea of integrated learning. The forest consists of many decision trees that are not related to one another. The decision trees each generate a classification model, learn and make predictions independently, and decide the final random forest classification result by the vote of the decision tree classifier [36]. This classification model is based on the sampling method of putting back that which was obtained from the original training dataset, in order to complete the construction process of the sub-dataset. The elements of different sub-datasets can be repeated in this process, as can the elements of the same sub-dataset [37]. In this study, the constructed NDVI time series features were added to the images in independent bands in the GEE platform. The random forest classification algorithm was used to extract the winter wheat distribution map of the city of Shijiazhuang in the year 2017. For parameter setting, considering the accuracy of winter wheat extraction and computational efficiency, 70% of the sample data were used as training samples and 30% as test samples; the number of decision trees was set as 50.

3.5. AUTS

Often, phenological changes are considered interannual temporal changes in phenology. If there is no change in land cover at a site, NDVI time series from different years can be assumed to be similar in shape. This hypothesis is reasonable for relatively stable land cover types and areas without significant human interference, where there is slight variation in species composition. Correlation can be a more intuitive reflection of how closely features interact with one another [38]. This study used the statistical method of correlation coefficients to study the correlation of NDVI time series in different years and at the same location. The maximum value of the NDVI half-month from October of the previous year to June of the following year was selected, the NDVI time series was

constructed from 18 NDVI points in 9 months, and the mean filtering was performed twice. The equation for the correlation coefficient is as follows:

$$R_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

where $n = 18$, x is the NDVI of 2017, y is the NDVI of 2018, x_i is the NDVI at the i th point in time in 2017, y_i is the NDVI at the i th point in time in 2018, and \bar{x} and \bar{y} are the average NDVI half-month maximum of 18 time points in 2017 and 2018, respectively.

Once the winter wheat growth period was determined, the next step was to use the NDVI time series curve for weighted correlation analysis. In traditional correlation analysis, all points are treated as equally important. However, phenological changes are usually very subtle and typically occur over a short period. It makes more sense to assign different weights to different points in the period, making change monitoring more sensitive to local changes in the identified winter wheat growth period. Therefore, we took the NDVI time series curve's middle time point t_9 as the reference point and set the weight W_i of each time point on the NDVI curve to become larger as the time distance D_i between the points t_9 and t_i (where circles of different radii represent different weights) decreased [39]. We found that the NDVI time series curve of winter wheat has the shape characteristics of double peaks and one valley (as shown in Figure 2), so we stipulate that the D_i values of t_5 , t_9 , and t_{15} are all 1, as shown in the following equation:

$$D_i = \begin{cases} 1, & t_i = t_5, t_9, t_{15} \\ \frac{1}{(t_i - t_9)^2}, & t_i \neq t_5, t_9, t_{15} \end{cases} \quad (4)$$

When the weighted correlation coefficients of the NDVI time series of two different years are calculated, the equation of the weighted correlation coefficients is as follows:

$$R_{xy}^w = \frac{\sum_{i=1}^n w_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n w_i (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n w_i (y_i - \bar{y})^2 \right]}} \quad (5)$$

where W_i is the weight accounted for by each time point, $n = 18$, x is the NDVI of 2017, y is the NDVI of 2018, x_i is the NDVI at the i th point in time in 2017, y_i is the NDVI at the i th point in time in 2018, \bar{x} and \bar{y} are the average NDVI half-month maximum of 18 time points in 2017 and 2018, respectively. If the weighted correlation coefficient passes the hypothesis test with a confidence level of 99.99% ($p < 0.001$), the two-year NDVI time series correlation is judged to be highly significantly correlated.

Similarly, the correlation between the 2017 and 2019 NDVI time series was studied.

In this study, the 2018 and 2019 training samples were extracted by using the distribution map of winter wheat in the city of Shijiazhuang in 2017, combined with weighted correlation analysis, to establish AUTS, since the areas where wheat has been planted for two consecutive years did not coincide completely. Based on the change monitoring of the weighted correlation coefficient of NDVI timing, the areas with a two-year NDVI sequence correlation that were highly significantly correlated were judged to be the same as the land cover type. Then, based on the 2017 classification map with highly significantly correlated areas, 4000 cells from each of the five sample classes were extracted, for a total of 20,000 cells, which constituted winter wheat, non-wheat vegetation, and non-vegetation (water, roads, and buildings). After that, the automatically extracted sample set was imported into the random forest classification algorithm; then, the distribution maps of winter wheat in Shijiazhuang in 2018 and 2019 were extracted.

4. Results

4.1. The Quality of the Updated Training Samples

4.1.1. The NDVI Time Series Curve for the Sample Points

Figure 2a,b shows the NDVI time series curves for 2018 and 2019, respectively. The curves resulted from three types of sample points extracted by AUTS, averaging the NDVI half-month maximum of 4000 points each and filtering them twice. From the figure, it can be seen that for the non-vegetation type, the NDVI half-month maximum remains stable throughout the year and is less than 0.3. For non-wheat vegetation types, the NDVI half-month maximum was relatively stable from November to March, reaching a peak in June; this was mainly because the climate is suitable for plant growth at this time. In contrast to the other two types, the NDVI half-month maximum time series curve characteristics of winter wheat are pronounced; after planting winter wheat in October, the maximum NDVI increases in the half-month, and the first peak of the tillering period occurs in November or December. With the arrival of winter and decrease in temperature, the NDVI of winter wheat begins to decline. With the arrival of spring, winter wheat enters the rising green stage; the NDVI starts to increase, and then enters the jointing period, reaching the second peak in the panicle period (around April or May). After the panicle period, the winter wheat enters the filling period and is harvested around June. This curve is entirely consistent with winter wheat's NDVI time series characteristics, so it is sufficient to prove that the sample points extracted based on our proposed AUTS are consistent with expectations.

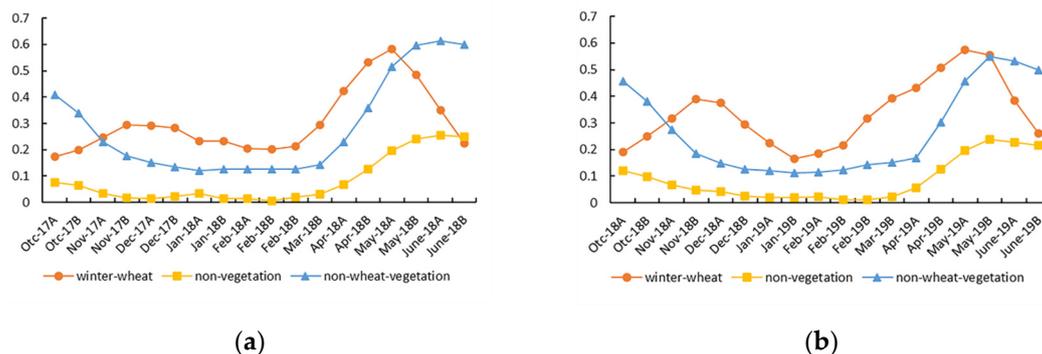


Figure 2. (a) NDVI time series curves of automatically extracted sample sites in Shijiazhuang in 2018. (b) NDVI time series curves of automatically extracted sample sites in Shijiazhuang in 2019.

4.1.2. Winter Wheat Training Samples' Distribution

The distribution maps of 4000 winter wheat samples that were automatically extracted in Shijiazhuang in 2018 and 2019, based on the AUTS we proposed, are shown in Figure 3a,b, respectively. It can be seen from the figure that most of the sample points are randomly distributed in the eastern part of Shijiazhuang, with a small number in the central region, and only sporadically distributed in the western region. The gap is large, mainly due to the low southeast and high northwest terrain of Shijiazhuang, and the landform is complex. The western part is located in the middle of the Taihang Mountains, and the eastern part is the alluvial plain of the Hutuo River, which is more suitable for growing winter wheat. The winter wheat sample points that we extracted were consistent with these topographic features.

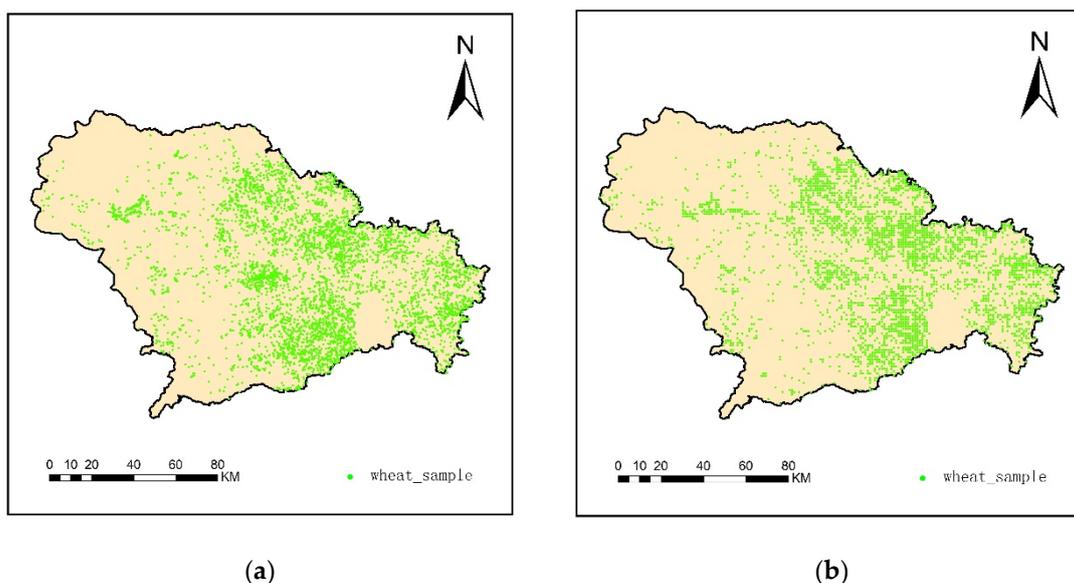


Figure 3. (a) Distribution of automatically extracted winter wheat sample sites in Shijiazhuang in 2018. (b) Distribution of automatically extracted winter wheat sample sites in Shijiazhuang in 2019.

4.2. Accuracy Assessment

4.2.1. Visual Interpretation

Firstly, ArcGIS 10.2 developed by Esri corporation in the United States was used to automatically create 2000 random points in the study area for visual interpretation. Secondly, using Sentinel-2 remote sensing images, the random points were distinguished into winter wheat, non-wheat vegetation, and non-vegetation by visual interpretation. The classification map was also used to distinguish the random points into winter wheat, non-wheat vegetation, and non-vegetation by visual interpretation. Finally, the confusion matrix was used to calculate the overall accuracy of the visual interpretation and kappa coefficient. Table 1 shows the results and accuracy of visual interpretation of winter wheat extraction in Shijiazhuang in 2017, 2018, and 2019. All the overall accuracy of three land use types is greater than 91%, and the Kappa coefficient is all greater than 0.86. Compared to Dong [40], who studied area extraction from winter wheat, our overall accuracy increased by 2%. The results show that the AUTS proposed in this study is highly accurate for winter wheat extraction.

Table 1. Visual interpretation results of winter wheat in Shijiazhuang in 2017, 2018, and 2019. PA: producer’s accuracy; UA: user’s accuracy; OA: overall accuracy; a: winter wheat; b: non-vegetation; c: non-wheat vegetation.

	2017				2018				2019		
	a	b	c	PA (%)	a	b	c	PA (%)	a	b	c
a	448	62	30	82.96	470	75	6	85.30	448	56	26
b	12	511	36	91.41	4	567	49	91.45	11	696	27
c	11	21	869	96.45	14	32	783	94.45	8	33	714
UA (%)	95.12	86.03	93.94		96.31	84.12	93.44		95.76	88.66	93.09
OA (%)	91.40				91.00				91.95		
Kappa	0.87				0.86				0.88		

4.2.2. Comparison with Statistical Data

To further verify the accuracy of winter wheat extraction, this study compared the winter wheat area extraction data of 21 county-level administrative regions in the city of Shijiazhuang with the official winter wheat area statistical data. Figure 4a–c shows the

comparative results of the statistical area and extracted area of winter wheat in Shijiazhuang in 2017, 2018, and 2019, respectively. These findings suggest that the winter wheat area data extracted based on our proposed AUTS are consistent with the statistical data, and the accuracy of winter wheat extraction in this study was very high. All points were distributed near the 1:1 line, and the R^2 values for three years were above 0.95. The Root Mean Squared Error (RMSE) values were 3953.82, 3598.11, and 3805.99 ha, respectively, while the Normalized root mean square error (NRMSE) values were 8.24%, 7.54%, and 7.96%, respectively. Compared to Xie [41], who studied area extraction from winter wheat, our R^2 increased by 3%. This further shows that the AUTS we propose is very feasible in winter wheat extraction.

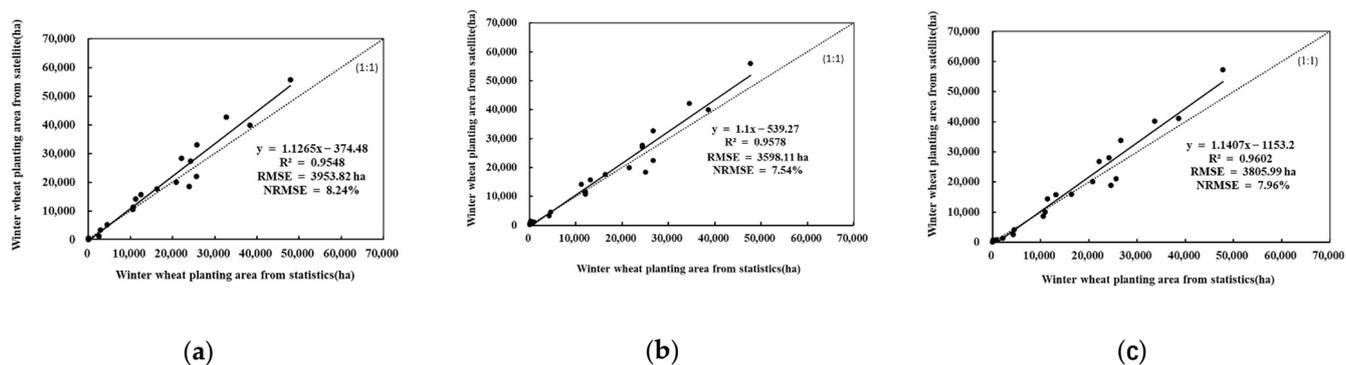


Figure 4. (a) The comparative results of the statistical and extracted areas of winter wheat in Shijiazhuang in 2017. (b) The comparative results of the statistical and extracted areas of winter wheat in Shijiazhuang in 2018. (c) The comparative results of the statistical and extracted areas of winter wheat in Shijiazhuang in 2019.

4.3. Time Variation Characteristics of Winter Wheat in Shijiazhuang

Figure 5 shows the changes in winter wheat area in various county-level administrative districts of Handan. The total areas of winter wheat planting in Handan were 338,001, 348,444, and 342,065 hectares in 2017, 2018, and 2019, respectively, with the most being in 2018. It can be seen from the figure that the winter wheat planting areas in Luquan District, Xingtang County, and Lingshou County increased significantly in 2018, and there were no noticeable changes in the winter wheat planting areas of other regions. In addition, it can be seen that there are very few areas where winter wheat is planted in the Chang’an District, Qiaoxi District, Xinhua District, Yuhua District, and other places, mainly because these areas constitute the urban area of Shijiazhuang. Most of the urban area is construction land, and the cultivated land area is small.

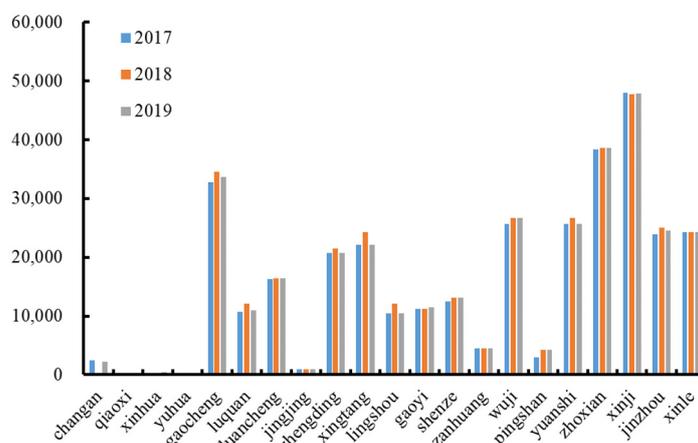


Figure 5. Changes in winter wheat area by county-level administrative region in Shijiazhuang.

4.4. Spatial Variation Characteristics of Winter Wheat in Shijiazhuang

Figure 6a–c shows the distribution maps of winter wheat in Shijiazhuang in 2017, 2018, and 2019, respectively, which we extracted using Sentinel-2 remote sensing images. It can be seen from the figure that the winter wheat growing area is mainly distributed in the east of Shijiazhuang, while very little winter wheat is produced in the west, which is primarily dominated by non-wheat-vegetation. This is mainly because the western part of Shijiazhuang is connected to—and influenced by—the Taihang Mountains. The general characteristics of Shijiazhuang’s terrain are high in the west and low in the east, with a stepped distribution from west to east. The terrain comprises mountains, hills, and plains, of which, most of the mountains are distributed to the west of Shijiazhuang, the plain area is mainly located in the east, and the hills are widely distributed in the central region. In addition, it is more evident that non-winter wheat vegetation is planted in Zhao County, bordering the city of Jinzhou to the southeast of Shijiazhuang. Therefore, the eastern plain area of Shijiazhuang is suitable for agricultural development, and the main crops grown are grain and cash crops, of which wheat and corn are the main food crops, and peanuts and cotton are auxiliary cash crops.

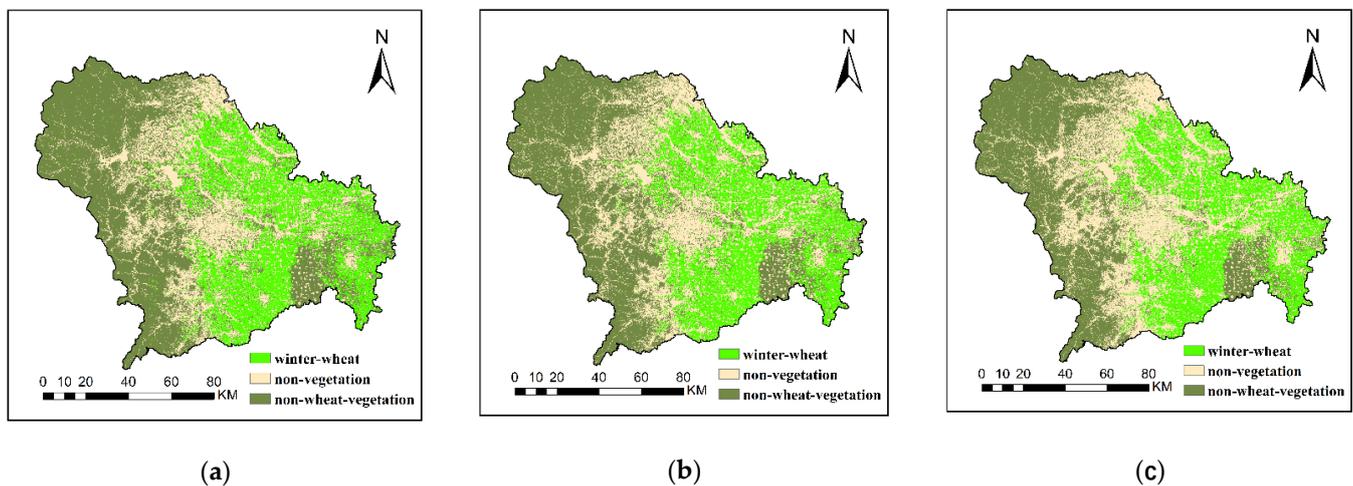


Figure 6. (a) Distribution of winter wheat in Shijiazhuang in 2017. (b) Distribution of winter wheat in Shijiazhuang in 2018. (c) Distribution of winter wheat in Shijiazhuang in 2019.

5. Discussion

5.1. Applicability of AUTS in Landsat 8 Images

To investigate whether our proposed AUTS is suitable for remote sensing satellite imagery at other scales, we used Landsat 8 image data, with a spatial resolution of 30 m, to extract training samples and winter wheat distribution maps. This experiment selected Landsat 8 remote sensing images of winter wheat in Shijiazhuang in 2018 and 2019 for NDVI time series synthesis. Since the revisit period of the Landsat 8 satellite is 16 days, we constructed a time series using NDVI monthly maximum synthesis. In addition, since images from February 2019 were missing from parts of the study area, February was removed, and the results were not affected. Then, using the 2017 winter wheat distribution map of the city of Shijiazhuang, combined with the Landsat 8 images, the training samples of Shijiazhuang in 2018 and 2019 were extracted based on the AUTS. Then, the winter wheat planting area was extracted, and the accuracy of the results was evaluated, as shown in Figure 7a,b, which are the distribution maps of winter wheat in Shijiazhuang in 2018 and 2019, respectively, based on Landsat 8 extraction. Figure 8a,b shows the results of the extraction and statistical data of winter wheat in Shijiazhuang in 2018 and 2019, respectively. The results show that, comparing the extracted area and statistical area of winter wheat in Shijiazhuang in 2018 and 2019, the R^2 were 0.95 and 0.90, respectively; the RMSE were 6617.18 and 5033.46 ha, respectively, and the NRMSE were 13.88% and 10.53%, respectively.

These results are good, and this experiment shows that our automatically updated sample algorithm is not limited to single-scale remote sensing satellite imagery.

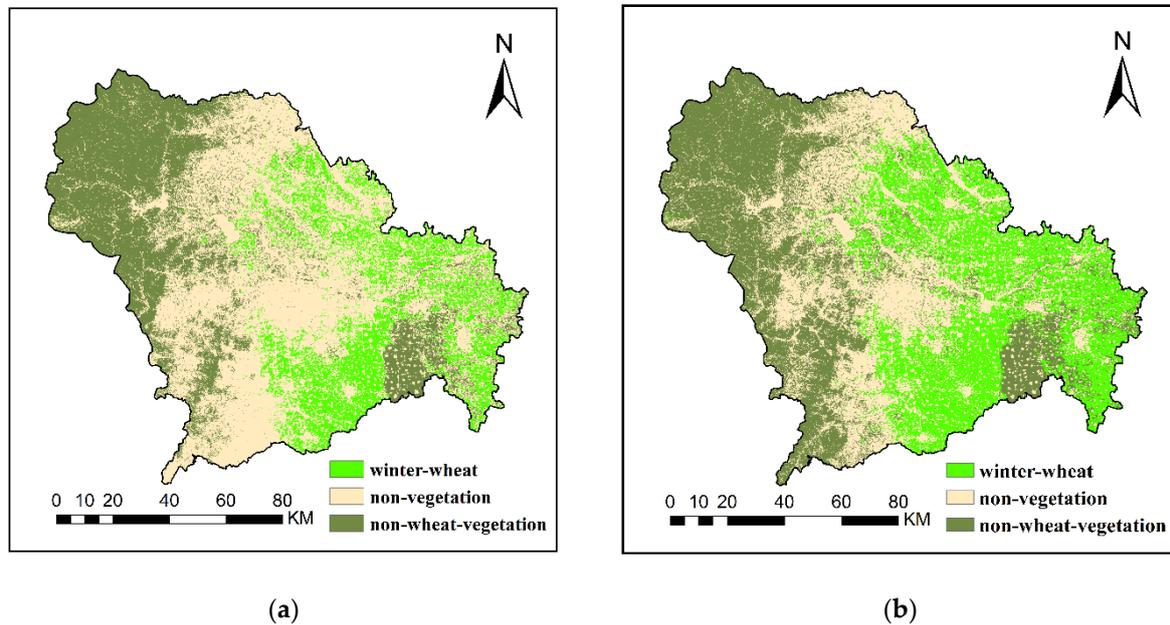


Figure 7. (a) Distribution of winter wheat in Shijiazhuang, based on Landsat 8 extraction for 2018. (b) Distribution of winter wheat in Shijiazhuang, based on Landsat 8 extraction for 2019.

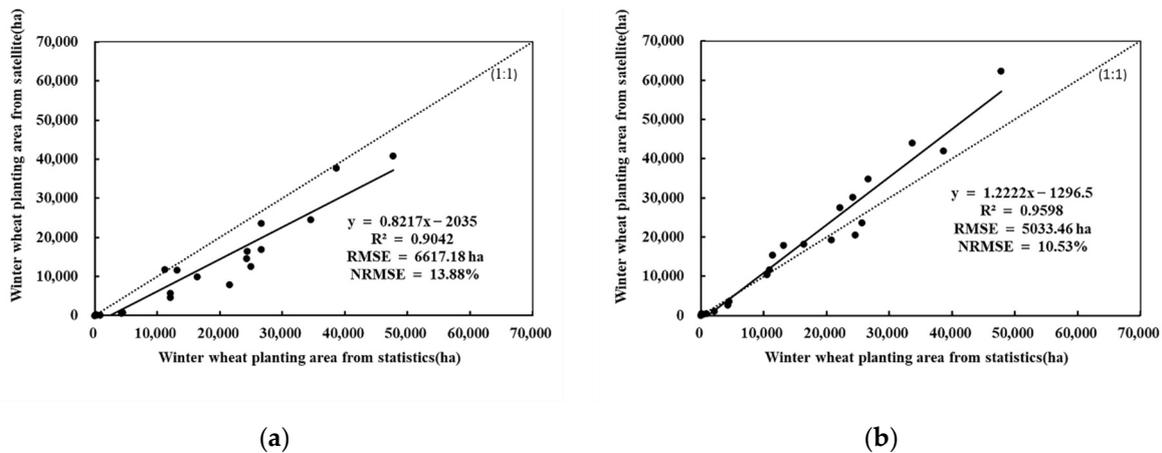


Figure 8. (a) The results of the extraction and statistical data of winter wheat in Shijiazhuang in 2018. (b) The results of the extraction and statistical data of winter wheat in Shijiazhuang in 2019.

However, the accuracy is slightly lower than that of the more delicate, high-spatial-resolution Sentinel-2 image data. This is mainly due to the presence of more mixed cells for traditional coarse satellite data, significantly hindering the classification accuracy [28]. In contrast, the 10 m resolution produces more “pure pixels” of winter wheat, without mixing with other objects (e.g., soil and non-wheat vegetation). As evidenced by the Mali cotton belt study, a map with a resolution of 10 m can effectively facilitate the management of agricultural systems [42]. Thus, the 10 m resolution Sentinel-2 satellite imagery can more accurately estimate the winter wheat planting area than the coarse Landsat 8 satellite images.

5.2. Applicability of AUTS in Different Regions

To study whether our proposed AUTS is suitable for different regions, we used the training samples extracted from Shijiazhuang to extract winter wheat in the two cities of Handan and Baoding, to the north and south of Shijiazhuang, respectively. Figure 9a,b shows the extracted winter wheat distribution in Handan and Baoding, respectively, in 2018, from which we can see that the winter wheat cultivation in Handan and Baoding is similar to that in Shijiazhuang—mainly distributed in the east. In contrast, very little winter wheat is grown in the west, which is primarily dominated by non-wheat vegetation. This is mainly due to the fact that Handan, Baoding, and Shijiazhuang are all located in the south of Hebei Province. The general characteristics of the terrain are high in the west and low in the east; the distribution is stepped from west to east, and the terrain comprises mountains, hills, and plains. Most of the mountainous areas are distributed to the west of Shijiazhuang; the plain area is mainly located in the east, and the hills are widely distributed in the central region.

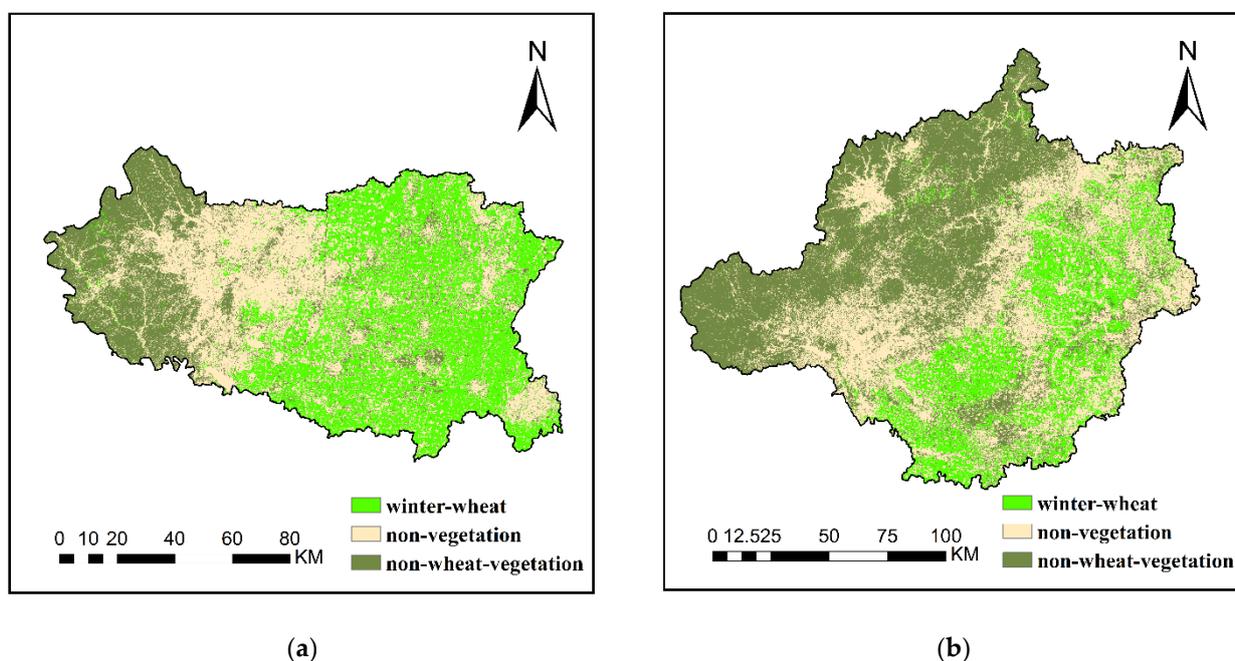


Figure 9. (a) Distribution of winter wheat in Handan in 2018. (b) Distribution of winter wheat in Baoding in 2018.

We downloaded the official statistics of the winter wheat planting area in the county-level cities of Handan and Baoding from the CNKI; there are 18 county-level cities in Handan and 17 county-level cities in Baoding. Comparing the satellite extraction area of winter wheat with the official statistical data of the winter wheat area, the R^2 values were 0.94 and 0.86, respectively; the RMSE values were 7190.21 and 3667.84 ha, respectively, and the NRMSE values were 11.13% and 11.74%, respectively. The results are shown in Figure 10. The fitting results of Handan are good enough to show that the winter wheat samples extracted by our proposed AUTS are of good quality. However, the fitting effect for Baoding is relatively low. It is well-known that winter wheat has noticeable phenological changes and distribution characteristics, which may be due to the complexity of crops grown in Baoding and influence of the phenological differences.

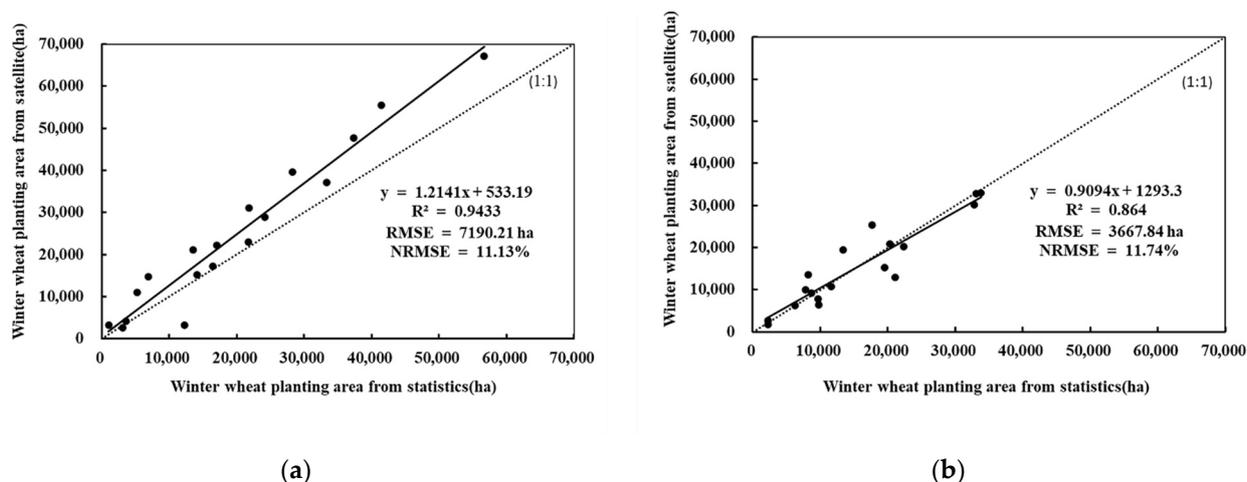


Figure 10. (a) The results of the extraction and statistical data of winter wheat in Handan. (b) The results of the extraction and statistical data of winter wheat in Baoding.

5.3. Advantages and Limitations of AUTS for Winter Wheat Extraction

Our proposed AUTS solves the high cost and difficulty of training sample collection. It is well-known that winter wheat has a long growth cycle, and its NDVI time series has prominent characteristics that distinguish it from other features. Taking this as the starting point, in this study, we calculated the correlation between NDVI time series in different years in the same region. Combined with the algorithm of automatic sample extraction of the GEE cloud platform, the AUTS was proposed. The distribution maps of winter wheat in Shijiazhuang in other years were extracted, not only providing higher accuracy, but also greatly saving time.

Although the GEE has abundant remote sensing data resources and powerful computational capabilities, it still has drawbacks. For example, the computational and storage capacity that the GEE allocates to users is limited. Our study area is a municipal administrative district in China. Further research is needed to determine whether our proposed AUTS is suitable for automatic sample extraction on a large scale. According to a recent study, it is possible to select a few representative plots, instead of the whole study area, for training sample generation, further improving efficiency and reducing the need for image input [43]. However, large-scale study areas have significant differences in terrain, climate, and other factors, and their geomorphology is complex. The task of collecting training samples for many years is costly and difficult. We expect that the GEE will have better computational power in the future, enabling us to further improve the performance of winter wheat mapping with more reliable features (e.g., digital elevation models or textures) [44].

6. Conclusions

In this study, we collected the 2017 Shijiazhuang training sample set in an RGB map based on high-resolution satellite images provided by a Google Earth map, constructed an NDVI time series using Sentinel-2 images, and extracted the distribution map of winter wheat in Shijiazhuang in 2017 by combining it with the random forest classification algorithm. The 2017 winter wheat distribution map of Shijiazhuang was combined with correlation analysis to propose AUTS to extract training samples from 2018 and 2019. The distribution maps of winter wheat in Shijiazhuang in 2018 and 2019 were extracted using Sentinel-2 images, based on AUTS and combined with a random forest classification algorithm. The accuracy was assessed by visual interpretation, verification, and comparison with statistical data, and we assessed the feasibility of AUTS in extracting winter wheat planting areas based on our GEE platform. The main conclusions obtained from the study were as follows: The NDVI time series curve plotted by the sample points extracted from our proposed AUTS was highly consistent with the winter wheat time series curve. The

distribution maps of winter wheat, derived from 2017, 2018, and 2019, were verified by visual interpretation; the overall accuracy of all three land use types was greater than 91%, and the Kappa coefficient was all greater than 0.86. Comparing the calculated winter wheat planting area with the statistics, the R^2 values for all three years were above 0.95. Finally, we further used Landsat 8 image data, with a rough spatial resolution of 30 m, for the cities of Handan and Baoding, in order to study the applicability of AUTS at different scales and in different regions. Comparing the satellite extraction area with the official statistical data of winter wheat in Shijiazhuang in 2018 and 2019, the R^2 values were 0.95 and 0.90, respectively. Comparing the satellite extraction area with the official statistical data of winter wheat in Handan and Baoding in 2018, the R^2 values were 0.94 and 0.86, respectively.

Author Contributions: Conceptualization, C.W. and H.Z.; methodology, X.W.; software, C.W., X.W. and H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, C.W., X.W., Y.S., W.Y., B.L. and J.W.; supervision, Y.S., B.L. and J.W.; project administration, X.W., C.W. and W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Henan Provincial Science and Technology Research Project (222102110038, 222102210131), Japan Society for the Promotion of Science (JSPS) KAKENHI Grant (No. 20K12146), and Henan Polytechnic University Doctoral Fund Project (B2021-19).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sentinel-2 and Landsat 8 data are openly available via the Google Earth Engine.

Acknowledgments: We appreciate the technical support of the Google Earth Engine cloud platform.

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

References

1. Porkka, M.; Kummu, M.; Siebert, S.; Varis, O. From food insufficiency towards trade dependency: A historical analysis of global food availability. *PLoS ONE* **2013**, *8*, e82714. [[CrossRef](#)] [[PubMed](#)]
2. Gong, P.; Liu, H.; Zhang, M.N.; Li, C.C.; Wang, J.; Huang, H.B.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [[CrossRef](#)]
3. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
4. Barboza Castillo, E.; Turpo Cayo, E.Y.; de Almeida, C.M.; Salas López, R.; Rojas Briceño, N.B.; Silva López, J.O.; Barrena Gurbillón, M.Á.; Oliva, M.; Espinoza-Villar, R. Monitoring Wildfires in the Northeastern Peruvian Amazon Using Landsat-8 and Sentinel-2 Imagery in the GEE Platform. *ISPRS Int. J. Geoinf.* **2020**, *9*, 564. [[CrossRef](#)]
5. Aghababaei, M.; Ebrahimi, A.; Naghipour, A.A.; Asadi, E.; Verrelst, J. Vegetation Types Mapping Using Multi-Temporal Landsat Images in the Google Earth Engine Platform. *Remote Sens.* **2021**, *13*, 4683. [[CrossRef](#)]
6. Goldblatt, R.; You, W.; Hanson, G.; Khandelwal, A.K. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. *Remote Sens.* **2016**, *8*, 634. [[CrossRef](#)]
7. Li, X.C.; Gong, P.; Liang, L.A. 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. *Remote Sens. Environ.* **2015**, *166*, 78–90. [[CrossRef](#)]
8. Busker, T.; Roo, A.; Gelati, E.; Schwatke, C.; Adamovic, M.; Bisselink, B. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrol. Earth Syst. Ences.* **2019**, *23*, 669–690. [[CrossRef](#)]
9. Sazib, N.; Mladenova, I.; Bolten, J. Leveraging the Google Earth Engine for Drought Assessment Using Global Soil Moisture Data. *Remote Sens.* **2018**, *10*, 1265. [[CrossRef](#)] [[PubMed](#)]
10. Ge, G.B.T.; Shi, Z.J.; Zhu, Y.J.; Yang, X.H.; Hao, Y.G. Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Glob. Ecol. Conserv.* **2020**, *22*, e00971. [[CrossRef](#)]
11. Shelestov, A.; Lavreniuk, M.; Kussul, N.; Novikov, A.; Skakun, S. Exploring Google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Front. Earth Sci.* **2017**, *5*, 17. [[CrossRef](#)]
12. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [[CrossRef](#)] [[PubMed](#)]

13. Wang, L.; Diao, C.; Xian, G.; Yin, D.; Lu, Y.; Zou, S.; Erickson, T.A. A summary of the special issue on remote sensing of land change science with Google earth engine. *Remote Sens. Environ.* **2020**, *248*, 112002. [[CrossRef](#)]
14. Zhang, W.M.; Brandt, M.; Prishchepov, A.V.; Li, Z.F.; Lyu, C.G.; Fensholt, R. Mapping the Dynamics of Winter Wheat in the North China Plain from Dense Landsat Time Series (1999 to 2019). *Remote Sens.* **2021**, *13*, 1170. [[CrossRef](#)]
15. Jaafar, H.; Mourad, R. GYMEE: A Global Field-Scale Crop Yield and ET Mapper in Google Earth Engine Based on Landsat, Weather, and Soil Data. *Remote Sens.* **2021**, *13*, 773. [[CrossRef](#)]
16. Fang, P.; Zhang, X.; Wei, P.; Wang, Y.; Zhang, H.; Liu, F.; Zhao, J. The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 m Resolution Imagery. *Appl. Sci.* **2020**, *10*, 5075. [[CrossRef](#)]
17. Li, C.; Chen, W.; Wang, Y.; Wang, Y.; Ma, C.; Li, Y.; Li, J.; Zhai, W. Mapping Winter Wheat with Optical and SAR Images Based on Google Earth Engine in Henan Province, China. *Remote Sens.* **2022**, *14*, 284. [[CrossRef](#)]
18. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* **2020**, *12*, 236. [[CrossRef](#)]
19. Xu, F.; Li, Z.; Zhang, S.; Huang, N.; Quan, Z.; Zhang, W.; Liu, X.; Jiang, X.; Pan, J.; Prishchepov, A.V. Mapping Winter Wheat with Combinations of Temporally Aggregated Sentinel-2 and Landsat-8 Data in Shandong Province, China. *Remote Sens.* **2020**, *12*, 2065. [[CrossRef](#)]
20. Waldner, F.; Canto, G.S.; Defourny, P. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 1–13. [[CrossRef](#)]
21. Zhong, L.H.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 1–13. [[CrossRef](#)]
22. Huang, H.B.; Wang, J.; Liu, C.X.; Liang, L.; Li, C.C.; Gong, P. The migration of training samples towards dynamic global land cover mapping. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 27–36. [[CrossRef](#)]
23. Zhang, L.; Liu, Z.; Liu, D.; Xiong, Q.; Yang, N.; Ren, T.; Zhang, C.; Zhang, X.; Li, S. Crop mapping based on historical samples and new training samples generation in Heilongjiang Province, China. *Sustainability* **2019**, *11*, 5052. [[CrossRef](#)]
24. Sharma, R.; Ghosh, A.; Joshi, P.K. Decision tree approach for classification of remotely sensed satellite data using open source support. *J. Earth Syst. Sci.* **2013**, *122*, 1237–1247. [[CrossRef](#)]
25. Lunetta, R.S.; Shao, Y.; Ediriwickrema, J.; Lyon, J.G. Monitoring agricultural cropping patterns across the Laurentian Great Lakes Basin using MODIS-NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 81–88. [[CrossRef](#)]
26. Sakamoto, T.; Yokozawa, M.; Toritani, H.; Shibayama, M.; Ishitsuka, N.; Ohno, H. A crop phenology detection method using time-series MODIS data. *Remote Sens. Environ.* **2005**, *96*, 366–374. [[CrossRef](#)]
27. Wardlow, B.D.; Egbert, S.L. A comparison of MODIS 250-m EVI and NDVI data for crop mapping: A case study for southwest Kansas. *Int. J. Remote Sens.* **2010**, *31*, 805–830. [[CrossRef](#)]
28. Pan, Y.Z.; Li, L.; Zhang, J.S.; Liang, S.L.; Zhu, X.F.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index. *Remote Sens. Environ.* **2012**, *119*, 232–242. [[CrossRef](#)]
29. Wardlow, B.D.; Egbert, S.L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116. [[CrossRef](#)]
30. Fritz, S.; See, L.; McCallum, I.; You, L.; Bun, A.; Moltchanova, E.; Duerauer, M.; Albrecht, F.; Schill, C.; Perger, C.; et al. Mapping global cropland and field size. *Glob. Chang. Biol.* **2015**, *21*, 1980–1992. [[CrossRef](#)]
31. Lobell, D.B.; Asner, G.P. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* **2004**, *93*, 412–422. [[CrossRef](#)]
32. Yang, G.; Yu, W.; Yao, X.; Zheng, H.; Cao, Q.; Zhu, Y.; Cao, W.; Cheng, T. AGTOC: A novel approach to winter wheat mapping by automatic generation of training samples and one-class classification on Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102446. [[CrossRef](#)]
33. Song, C.; Woodcock, C.E.; Seto, K.C.; Lenney, M.P.; Macomber, S.A. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sens Environ.* **2001**, *75*, 230–244. [[CrossRef](#)]
34. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *Third Earth Resources Technology Satellite-1 Symposium*; Freden, S.C., Mercanton, E.P., Becker, M.A., Eds.; NASA: Washington, DC, USA, 1973; Volume 1, pp. 309–317.
35. Ren, J.Q.; Chen, Z.X.; Zhou, Q.B.; Tang, H.J. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 403–413. [[CrossRef](#)]
36. Mantas, C.J.; Castellano, J.G.; Moral-Garcia, S.; Abellan, J. A comparison of random forest based algorithms: Random credal random forest versus oblique random forest. *Soft Comput.* **2009**, *23*, 10739–10754. [[CrossRef](#)]
37. He, Y.H.Z.; Wang, C.L.; Chen, F.; Jia, H.C.; Liang, D.; Yang, A.Q. Feature Comparison and Optimization for 30-M Winter Wheat Mapping Based on Landsat-8 and Sentinel-2 Data Using Random Forest Algorithm. *Remote Sens.* **2019**, *11*, 535. [[CrossRef](#)]
38. Rousselet, G.A.; Pernet, C.R. Improving standards in brain-behavior correlation analyses. *Front. Hum. Neurosci.* **2012**, *6*, 119. [[CrossRef](#)] [[PubMed](#)]
39. Chen, J.; Rao, Y.; Shen, M.; Wang, C.; Zhou, Y.; Ma, L.; Tang, Y.; Yang, X. A Simple Method for Detecting Phenological Change From Time Series of Vegetation Index. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3436–3449. [[CrossRef](#)]
40. Dong, J.; Fu, Y.; Wang, J.; Tian, H.; Fu, S.; Niu, Z.; Han, W.; Zheng, Y.; Huang, J.; Yuan, W. Early-season mapping of winter wheat in China based on Landsat and Sentinel images. *Earth Syst. Sci. Data.* **2020**, *12*, 3081–3095. [[CrossRef](#)]

41. Xie, Y.; Huang, J.X. Integration of a Crop Growth Model and Deep Learning Methods to Improve Satellite-Based Yield Estimation of Winter Wheat in Henan Province, China. *Remote Sens.* **2021**, *13*, 4372. [[CrossRef](#)]
42. Lambert, M.J.; Traore, P.C.S.; Blaes, X.; Baret, P.; Defourny, P. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens. Environ.* **2018**, *216*, 647–657. [[CrossRef](#)]
43. Phalke, A.R.; Ozdogan, M.; Thenkabail, P.S.; Erickson, T.; Gorelick, N.; Yadav, K.; Congalton, R.G. Mapping croplands of Europe, Middle East, Russia, and Central Asia using Landsat, Random Forest, and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 104–122. [[CrossRef](#)]
44. Zhang, C.K.; Zhang, H.Y.; Zhang, L.P. Spatial domain bridge transfer: An automated paddy rice mapping method with no training data required and decreased image inputs for the large cloudy area. *Comput. Electron. Agric.* **2021**, *181*, 105978. [[CrossRef](#)]