High-resolution distribution maps of single-season rice in China from 2017 to 2022

Ruoque Shen[1,2], Baihong Pan[3], Qiongyan Peng[1,2], Jie Dong[4], Xuebing Chen[1,2], Xi Zhang[1,2], Tao Ye[5], Jianxi Huang[6], and Wenping Yuan[1,2]

5   [1]International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China

[2]School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai 519082, Guangdong, China

[3]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, 73019, USA.

[4]College of Geomatics & Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, Zhejiang, China

[5]Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China

[6]College of Land Science and Technology, China Agricultural University, Beijing 100083, China

Correspondence to: Wenping Yuan (yuanwp3@mail.sysu.edu.cn)

15   **Abstract**. Paddy rice is the second-largest grain crop in China and plays an important role in ensuring global food security. However, there is no high-resolution map of rice covering all of China. This study developed a new rice mapping method by combining optical and synthetic aperture radar (SAR) images in cloudy areas based on the time-weighted dynamic time warping (TWDTW) method and produced distribution maps of single-season rice in 21 provincial administrative regions of China from 2017 to 2022 at 10 or 20-m resolution. The accuracy was examined by using 108195 survey samples and county-level statistical data. On average, the user's, producer's, and overall accuracy over all investigated provincial administrative regions were 73.08 %, 82.81 %, and 85.23 %, respectively. Compared with the statistical data from 2017 to 2019, the distribution map explained 83 % of the spatial variation of county-level planting areas on average. The distribution maps can be obtained at https://doi.org/10.57760/sciencedb.06963 (Shen et al., 2023).

Keywords: rice; single-season rice; time-weighted dynamic time warping; Sentinel

1. Introduction

As the fourth-largest grain crop in the world, rice contributed 8 % to world food production in 2019 (FAO, 2021). Rice is a staple food for more than half of the world's population and plays an important role in ensuring global food security (Elert, 2014; Kuenzer and Knauer, 2013). The flooding of rice paddy fields constitutes a major source of methane emissions (IPCC, 2022; Mohammadi et al., 2020). Therefore, quickly and accurately identifying the planting location of rice over a large area is very important.

Most commonly, large-scale crop mapping takes advantage of satellite data (Dong et al., 2020; Huang et al., 2022; Xiao et al., 2006, 2005). Popular crop-mapping methods are various machine learning methods, such as random forest (Boryan et al., 2011; Fiorillo et al., 2020; You et al., 2021), support vector machine (Zheng et al., 2015), and deep learning (Thorp and Drajat, 2021; Zhao et al., 2019; Zhong et al., 2019). Machine learning methods provide several advantages in crop mapping

35 but require training samples (Belgiu and Csillik, 2018), commonly in the order of hundreds or even thousands to obtain a satisfactory accuracy (Millard and Richardson, 2015; Valero et al., 2016). For example, the Cropland Data Layer (CDL) products produced by the U.S. Department of Agriculture (USDA) use tens of thousands of training samples to map the crops of a single state (Boryan et al., 2011). Therefore, such large-scale investigations are very time-consuming and labor-intensive.

Another crop mapping approach is based on the detection of specific phenological signals. Xiao et al. (2005, 2006)

40 produced a 500-m resolution rice map of Southern China, Southeast Asia, and South Asia using MODIS data by comparing the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) with the Land Surface Water Index (LSWI). In addition, Dong et al. (2016) also used the flood-detection method, producing a rice map with 30-m spatial resolution in Northeast Asia based on Landsat-8 data. Because of the short flooding period, the influence of clouds and rain in a few images will lead to missing the flooding signal and decreased accuracy, placing high requirements for image quality and

45 time resolution (Dong et al., 2016).

Additional crop mapping approaches are the dynamic time warping (DTW) and the time-weighted dynamic time warping (TWDTW) methods, which do not consider the crop characteristics within a certain time period but compare the signals over an extended period (Belgiu and Csillik, 2018; Qiu et al., 2017; Skakun et al., 2017; Zheng et al., 2022b). Guan et al. (2016) mapped rice in Vietnam using the DTW method based on MODIS (Moderate Resolution Imaging Spectroradiometer) NDVI

50 data, reaching an R^2 of 0.809. The TWDTW method, which is an improvement of the DTW method, adds a time weight to the calculation to characterize the temporal difference, improving identification accuracy (Maus et al., 2016). The TWDTW method has been used in several studies to produce high-resolution crop maps of many kinds of crops, including winter wheat, sugar cane, and maize (Dong et al., 2020; Huang et al., 2022; Zheng et al., 2022a; Shen et al., 2022). A previous study also used the TWDTW method to produce a map of double-season paddy rice in China by using the vertical-horizontal (VH) band

55 signal from the Sentinel-1 satellite (Pan et al., 2021).

Because of the flooding during rice planting, the traditional rice mapping methods use water indexes derived from optical images, such as LSWI (Xiao et al., 2002, 2005). However, optical images are greatly impacted by clouds, heavily limiting their availability in cloudy regions (Li and Chen, 2020; Sudmanns et al., 2020; Zhou et al., 2019). An alternative is the use of synthetic aperture radar (SAR) images. Compared with the optical signal, the SAR signal can penetrate through clouds,

60 completely avoiding their influence (Oguro et al., 2001; Phan et al., 2018). Several studies have demonstrated the capability

of SAR in rice identification and obtained good mapping results at the regional scale (Nguyen et al., 2016; Han et al., 2021; Pan et al., 2021). However, compared with optical data, SAR data also has more significant salt-and-pepper noises, which may affect the accuracy of the distribution map (Oliver and Quegan, 2004).

China is the world's largest rice producer, producing 209.61 million tons of rice in 2019 (*China Statistical Yearbook*, 65 2020). Except for a few provinces in Southeastern China, most of the rice-planting provinces are dominated by single-season rice. Although there are many previous studies on mapping rice in China, a high-resolution single-season rice map is still not available for the entire country. This study attempts to fill this gap and aims to: (1) develop a new phenology-based method for rice mapping; (2) produce high-resolution distribution maps of single-season rice in China from 2017 to 2022; (3) evaluate the accuracy of the identified areas using county-level statistical data and survey samples.

70 **2. Data and methods**

2.1 Study area

This study was conducted in 21 provincial administrative regions in mainland China, where the total planting area of single-season rice was 19.92 million hectares, accounting for approximately 99.01 % of the total planting area of single-season rice in mainland China according to the statistical data in 2018 (https://data.stats.gov.cn). The total production of the 75 single-season rice in the study area was 150.46 million tons, accounting for approximately 98.91 % of the total production in mainland China in 2018. As single-season rice is wildly planted in China, this study further divided the study area into four subregions (Fig. 1). Subregion I is the northern rice planting area, including Heilongjiang, Jilin, Liaoning, Inner Mongolia, and Ningxia. Because of temperature limitations, only single-season rice is planted in this subregion, with the transplanting period generally between late May and mid-June. Subregion II is the middle southern single-season rice planting area, 80 including provinces that only or mainly plant single-season rice (Jiangsu, Anhui, Hubei, Henan, Shandong, Shaanxi, and Shanghai) and provinces where single-season and double-season rice are both planted (Hunan and Jiangxi). The single-season rice in this subregion is generally transplanted between mid-late May and late June. Subregion III is the southeastern coastal single-season rice planting area, including Zhejiang, Fujian, and Guangxi. Here, single-season rice may be transplanted later than in Subregion II, generally between mid-late May and early July. Subregion IV is the southwestern rice 85 planting area, including Sichuan, Yunnan, Guizhou, and Chongqing. Single-season rice in this subregion is transplanted much earlier than in other subregions, generally between late April and mid-May.

2.2 Data

2.2.1 Satellite data and landcover data

90   The satellite data used in this study were from the Sentinel series launched by the European Space Agency (ESA). The

SAR data were obtained from the Ground Range Detected (GRD, Level-1) product of Sentinel-1A, and the optical data were

obtained from the Level-1C product of Sentinel-2. The SAR data used in this study were the VH band (dual-band cross-

polarization, vertical transmit/horizontal receive) at a spatial resolution of 10-m, and was composited into a 12-day temporal

resolution by median. Optical data included ten bands (blue, green, red, and near infrared (NIR) at a 10-m spatial resolution,

95   and red edge1 (RE1), red edge2 (RE2), red edge3 (RE3), red edge4 (RE4), shortwave infrared1 (SWIR1), and shortwave

infrared2 (SWIR2) at a 20-m spatial resolution). Additionally, two indexes, NDVI and LSWI at a 10-m spatial resolution, were

calculated with the following equations:

$$\text{NDVI} = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \tag{1}$$

$$\text{LSWI} = \frac{\rho_{NIR} - \rho_{SWIR1}}{\rho_{NIR} + \rho_{SWIR1}} \tag{2}$$

100   where ρ_{NIR}, ρ_{red}, ρ_{SWIR1} are the reflectances of the NIR, red, and SWIR1 bands of Sentinel-2, respectively.

     The Sentinel-2 Cloud Probability (S2C) product (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY) was used to eliminate the influence of clouds. The

product provides a cloud probability from 0 to 100 at a 10-m resolution, which has a higher resolution than the original QA60

band of the Sentinel-2 dataset and is more flexible and accurate. In this study, the threshold of cloud probability was set to 50;

105   pixels with a higher probability were regarded as clouds and removed. Considering the length of the transplanting period and

the number of cloud-free images of each subregion, the optical data were composited to 12-day (Subregion I) or 6-day

(Subregion II, III, and IV) temporal resolution by median. Figure 2 shows the percentage of good optical observations during

the study period of each pixel in the study area. A linear interpolation was applied to fill the gaps in the time series. To further

eliminate the noise in the time series, a Savitzky-Golay (SG) filter with the order set to two and the window size set to five

110   was applied to smooth the time series (Chen et al., 2004). All the preprocessing was completed on the Google Earth Engine

(GEE) platform (Gorelick et al., 2017). Besides, this study used the Finer Resolution Observation and Monitoring of Global

Land Cover (FROM-GLC) product as a mask to exclude non-cultivated areas (Gong et al., 2019).

2.2.2 Field data and agricultural statistical data

     The field data were obtained through several field surveys we conducted across China during 2017–2021, including 37036

115   samples of single-season rice and 71159 samples of other crops (double-season rice, maize, soybean, peanuts, etc.), forests,

built-up areas, water bodies, etc. An unmanned aerial vehicle (UAV; eBee, senseFly Ltd., Switzerland) was used in some of our surveys to take very-high-resolution images covering on average 0.1 km². The images were visually interpreted to obtain sample points at a spatial resolution of 20 m. The province-level agricultural statistical data are published in the statistical yearbook of each province, and the county-level statistical data are published sporadically in the statistical yearbook of each

120 province or city. Since the release of data usually lagged by two years or more, and the single-season rice planting areas were not published in many counties, this study only collected a total of 2748 county-level statistical single-season rice planting area data from 2017 to 2019 (Table 1). No available county-level statistical data were collected for Heilongjiang and Inner Mongolia due to the discrepancies between the administrative division and statistical caliber.

2.3   Method

125 ### 2.3.1   Time-weighted dynamic time warping method

This study generated the single-season rice distribution map by comparing the dissimilarity of the time series of each pixel with the standard time series of rice. The TWDTW method was used to calculate the dissimilarity (Petitjean et al., 2012; Dong et al., 2020). In this method, the unknown time series is non-linearly stretched or compressed to align with the standard single-season rice time series, and an accumulated distance is calculated by cumulating the distance of the alignment path. The

130 accumulated distance of all possible alignments is calculated, and the minimum accumulated distance is used to represent the dissimilarity of two time series. Considering the phenophases of crops, a penalty called time weight is added to the calculation (Maus et al., 2016). When the time series is stretched or compressed, the difference caused by the dislocation of time axes is calculated, and a function (e.g., logistic function) is used to convert the time difference into a time weight. As a result, the TWDTW measures the difference between two time series by considering both shape and phenological information.

135 Parameters of the TWDTW used in this study were suggested by Belgiu and Csillik (2018), by using a logistic function with α and β set to 0.1 and 50, respectively. Finally, the single-season rice was identified through a threshold of dissimilarity determined by the province-level statistical area. The total area of pixels with dissimilarity lower than the threshold was equal to the statistical area.

2.3.2   Optical bands selection

140 The most significant feature that distinguishes rice from other crops is the flooding during its transplanting period. Figure 3 shows the time series of all 10 optical bands or indexes of four main crops in Jilin Province in 2019. It can be seen that the time series of rice of three moisture-related bands or indexes, including SWIR1, SWIR2, and LSWI are significantly different from those of other crops during the transplanting period (DOY 133–181). LSWI is designed to characterize land

surface moisture, and its value positively correlates with land surface moisture, showing a high value during the

145 transplanting period. In contrast, the two SWIR bands show the opposite, first decreasing and then increasing during the

transplanting period, following a "V" shape. As the aim of this study was to map the distribution of single-season rice, the

time series did not necessarily need to be able to characterize a certain land surface parameter like LSWI. The priority was

whether a band or index could distinguish rice from other crops. As LSWI is calculated as the normalized difference of the

NIR and SWIR1, and the NIR of rice also decreases slightly during the transplanting period, offsetting the LSWI increase

150 caused by the decrease of SWIR1, erasing some differences and information. The change of SWIR2 was less pronounced

than that of SWIR1, so SWIR1 was selected in this study to calculate the dissimilarity.

The standard time series were generated using survey samples. Fifty single-season rice survey samples were randomly

selected from all single-season rice points in each province. The SWIR1 time series of these samples were extracted, aligned

according to the time when the minimum value appears, and averaged to obtain the standard time series of each province.

155 The standard time series of 21 provincial administrative regions in four subregions all showed a "V" shape (Fig. 4). The time

period of the standard time series was limited to the transplanting period, and the length of the standard time series were five

in Subregion I and seven in other subregions. Because the method is transferable between years, the standard time series

retrieved from one year was used in all six years.

2.3.3 TWDTW with translation and stretching

160 Although the transplanting period of rice is short, farmers may transplant single-season rice over a longer period.

Therefore, the "V" shape may appear earlier or later. In this study, the standard time series was translated and stretched along

the time axis to match any possible period in which transplant might have occurred. To reduce the computational effort as

well as to prevent overstretching, the standard time series was allowed to stretch for at most one time. Taking Subregion I as

an example, where rice is generally transplanted from late May to mid-June, the study period was set to day of year (DOY)

165 121–193, with a total of seven observations to take decreasing and increasing phases of the "V" curve into consideration.

The length of the standard time series was five, and time series within the study period with a length of five or six were

selected to calculate the distance with the standard time series using the TWDTW method (Fig. 5). The minimum of all

distance was selected to represent the dissimilarity of the pixel. Study periods of Subregion II, III, and IV were set to DOY

121–199, 121–211, and 97–163, respectively.

170 ### 2.3.4 Combine SAR images in Southern China

Compared with northern China, southern China is more heavily affected by clouds and rain, resulting in a worse quality

of optical observation (Fig. 2). This study introduced SAR observations as an auxiliary in Subregion II, III, and IV, as it can

pass through clouds. Specifically, the VH band was used because studies have shown that VH polarization is more sensitive

than VV in detecting field flooding (Nguyen et al., 2016; Wakabayashi et al., 2019). The VH time series of rice in the

175   transplanting period also shows a "V" shape (Fig. 6). Although the shape of the rice curve differs from that of the other

crops, their values partly overlap. As a coherent radar system, SAR images will inevitably carry salt-and-pepper noises

(Veloso et al., 2017). Therefore, VH was only used when the quality of optical observation was extremely poor.

First, the dissimilarity of an unknown VH time series and the standard VH time series were calculated at each pixel

using the TWDTW method. The standard VH time series was generated using the same procedure as for SWIR1 (section

180   2.3.2). The study periods of VH of Subregions II, III, and IV were set to DOY 121–193, 121–205, and 97–169, respectively

(Fig. 7). Second, the distance of the pixel was replaced by the ranking of the pixel by sorting the distance as the distances

were not comparable between different bands (i.e., SWIR1 and VH). In addition, a 20-m SWIR1 pixel covers four 10-m VH

pixels, so the rankings of the SWIR1 pixels were multiplied by four and resampled to 10-m resolution to ensure that the

dissimilarities of SWIR1 and VH were comparable. Third, the combined dissimilarity of each pixel was calculated by a

185   weighted sum of the rankings of two bands, while the weight was determined by the times of good observations of the

optical images. The weight w was calculated using a logistic function as:

$$w = \frac{1}{1 + e^{-\alpha(x-\beta)}} \tag{3}$$

where x is the times of good observations and α and β are parameters. Through a small range of tests, α and β were set to 2

and 2.5, respectively. By setting the parameters, w was close to 1 when x was greater than 3, and close to 0 when x was less

190   than 2.

The combined dissimilarity d was calculated as:

$$d = r_{SWIR1} \times w + r_{VH} \times (1 - w) \tag{4}$$

where r_{SWIR1} and r_{VH} are the ranking of SWIR1 and VH, respectively.

The distribution map was generated from the combined dissimilarity using the threshold mentioned in section 2.3.1.

195   **2.3.5   Accuracy assessment**

The study assessed the accuracy of the distribution map by using field data and county-level statistical areas. In this

study, the confusion matrix was used to show the classification of the distribution map on the survey samples, and three

accuracies were calculated, including the producer's accuracy (PA), user's accuracy, and overall accuracy (OA), calculated

as:

200                                    $$PA = \frac{TP}{TP + FP} \times 100\% \tag{5}$$

$$UA = \frac{TP}{TP + FN} \times 100\ \% \tag{6}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \times 100\ \% \tag{7}$$

where TP is the number of correctly classified single-season rice samples. TN is the number of correctly classified non-single-season rice samples. FP is the number of non-single-season rice samples classified as single-season rice. FN is the

205   number of single-season rice samples classified as non-single-season rice.

The county-level statistical planting areas from statistical year books were also used to verify the accuracy of the distribution map by comparing with the identified planting area at the county level. The relationships between the identified areas and the statistical areas were evaluated by linear regression. The coefficient of determination (R^2) and a relative error (RMAE) are calculated. The calculation equations of RMAE is as follows:

210

$$RMAE = \frac{\sum_{i=1}^{n} |SA_i - IA_i|}{\sum_{i=1}^{n} SA_i} \tag{8}$$

where SA_i and IA_i are the statistical area and identified area of the ith county, and n indicates the number of counties.

3. Results

This study generated the distribution maps of single-season rice from 2017 to 2022 in 21 provincial administrative regions in China, which well reproduced the distribution of single-season rice in China (Fig. 8). Northeast China Plain, Yangtze Plain,

215   and Sichuan Basin are three major single-season rice production areas in China, and single-season rice is planted most frequently in Northeast China Plain (Fig. 8). To show the ability of the distribution map of representing the details of rice fields, we chose three UAV sites and compared the distribution map with very-high-resolution UAV images (Fig. 8). Figures 8a and 8c were taken in July and show single-season rice fields in dark green (light green areas represent other planted vegetation). Figure 8b was taken in October, when single-season rice was about to be harvested, showing single-season rice fields in yellow-

220   green. Despite some noise, the single-season rice fields were well classified in our distribution map (Fig. 8d–f).

The distribution map shows good performance in most of the provincial administrative regions. On average, the user's, producer's, and overall accuracy over all 21 provincial administrative regions were 73.08 %, 82.81 %, and 85.23 %, respectively (Table 2). The average overall accuracies in the four subregions were 95.69 %, 81.15 %, 86.75 %, and 80.18 %, respectively (Table 2). Subregion I (i.e., the northern provinces), had higher accuracy, while the southern provinces, especially

225   the provinces in subregion IV (southwest) had poor accuracy. User's and producer's accuracies varied more between provinces than overall accuracy. The best user's and producer's accuracy all appeared in the northern provinces; the best user's accuracy was obtained for Inner Mongolia (97.67 %), and the best producer's accuracy for Liaoning (99.75 %) (Table 2). The lowest user's accuracy appeared in Guangxi (46.96 %), and the lowest producer's accuracy appeared in Jiangxi (49.22 %) (Table 2).

The county-level comparison with the statistical data showed a good performance. The identified area and statistical area

230   had a very strong correlation, and the regression line was very close to the 1:1 line over all three years (Fig. 10). The slope

ranged from 0.86 to 0.90, and R^2 ranged from 0.78 to 0.86.

Comparing on each province, the R^2 of the distribution map compared with the statistical data ranged from 0.15 to 0.94;

the slope ranged from 0.24 to 1.44, and the relative error ranged from 0.24 to 0.56. Subregion I had the highest accuracy with

an average R^2 of 0.92, followed by Subregion II with an average R^2 of 0.70, and subregions III and IV had poorer precision,

235   both with an average R^2 of 0.55. Several provinces with more mountainous areas (Fujian, Guangxi, and Guizhou) had lower

accuracies, while plain and main production provinces (Jilin, Liaoning, Jiangsu, Anhui, and Hubei) had higher accuracies.

4. Discussion

Paddy rice is the second most widely planted crop in China. Its planting area has been relatively stable for many years

(*China Statistical Yearbook*, 2020). From 1978 to 2005, the planting area of paddy rice decreased slowly, from 34.4 million

240   hectares to 28.9 million hectares, and the planting area has been maintained at about 30 million hectares after 2005 (*China

Statistical Yearbook*, 2020). The planting area of single-season rice accounted for two-thirds of all rice in China, and the

production accounted for three-quarters of all rice (https://data.stats.gov.cn). Despite the importance of the single-season rice,

rice mapping on a regional scale is still difficult. Many efforts have been made to map rice with a moderate or high spatial

resolution at the provincial and regional scale, using machine learning methods and phenology-based methods (Pan et al., 2021;

245   Xiao et al., 2006, 2005; You et al., 2021). However, these mapping methods have some limitations.

Compared to machine learning methods, the TWDTW method has the advantage of requiring fewer training samples. In

this study, the number of samples used in obtaining the standard time series was only 50. Many studies have reported that

using machine learning methods to achieve high accuracy requires a large volume of training samples while obtaining samples

is time-consuming and labor-intensive (Millard and Richardson, 2015; Valero et al., 2016). Therefore, the TWDTW method

250   can be easily extended to regions or years with limited survey data compared to the machine learning methods. For example,

You et al. (2021) mapped three crops, including rice in Northeast China from 2017 to 2020 by using a machine learning method

(Random Forest), and achieved producer's and user's accuracy of rice greater than 90 %, except for the user's accuracy in

2017 (87 %). However, their study used more than 8000 training samples per year that needed to be updated every year. In

contrast, this study achieved a similar accuracy with only 50 sample points per province in Northeast China.

255   Compared with the flood-detection method developed by Xiao et al. (2005), the TWDTW method uses signals in a certain

period before and after rice flooding, including more phenological information. Flood-detection methods are deeply affected

by clouds and rain. The accuracy of a moderate-resolution rice map based on the MODIS data can be relatively high due to

the high temporal resolution and less cloudy pixels of the MODIS data (Xiao et al., 2006, 2005). However, when based on Landsat data, the accuracy of such high-resolution product was not satisfactory due to the influence of cloudy images (Dong

260 et al., 2016). Furthermore, good observations in the southern areas of China are extremely scarce, especially in Subregion IV (southwestern), where the six-year average of the frequency of good observations is only between 25 % and 40 % during the transplanting period, making it impossible to map rice in these provinces using only optical images (Fig. 2). SAR images were introduced in this study in cloudy areas, making it possible to map rice in these areas.

The introduction of SAR has made rice identification possible in these areas. However, the quality of SAR images is

265 somewhat worse than that of optical images, which makes the accuracy of the distribution map in these areas still lower than that in less cloudy areas. The optical data for 2017 have the poorest observation quality, with the number of observations corresponding to the minimum distance only 1.16 (Fig. 12). This number of observations determines the high weight of the distances calculated from the SAR images, which explains the lowest R^2 in 2017 (Fig. 10). Comparing the R^2 of the county-level comparison with the statistical data with the frequencies of optical observations during the study period, shows a positive

270 correlation with R^2 range from 0.35 to 0.57 (Fig. 13). That is, in areas where the optical observations are heavily affected by clouds, the accuracy remains low to some extent, even if SAR images are used as auxiliaries.

Another important factor that affects identification accuracy is the fragmentation of planted areas. In mountainous provinces such as Guizhou, Chongqing, and Yunnan, there are many terraced rice fields, which are very narrow and fragmented (Cao et al., 2021; Yan et al., 2016). In these mountainous areas, rice fields may be less than 10 m wide, resulting in mixed

275 pixels at 10-m resolution. In addition, as a side-looking radar system, SAR has a terrain effect, which produces more errors in mountainous areas (Beaudoin et al., 1995). To quantify the fragmentation of the distribution map, we regarded adjacent single-season rice pixels as a patch, and counted the size of each patch. The fragmentation of the distribution maps in the same province varied little from year to year (Fig. 14). Guangxi, Guizhou, Shaanxi, Yunnan, and Fujian were the most fragmented provinces, with more than half of the pixels belonging to patches smaller than 100 pixels (about 1 hectare). The most

280 fragmented province was Guangxi, where, each year, an average of 85.45 % of the pixels belonged to patches smaller than 100 pixels (Fig. 14). Although there are plains in Guangxi, the plains are mostly planted with double-season rice, while single-season rice is mostly planted in mountainous areas, resulting in extremely fragmented single-season rice cultivation. Using the percentage of pixels belonging to patches smaller than 100 pixels as an indicator of fragmentation, and comparing it with the identification accuracy, a significant negative correlation can be found (Fig. 15). The R^2 of the fragmentation and identification

285 accuracy ranged from 0.51 to 0.72, confirming that the fragmentation of single-season rice fields has a strong negative effect on the identification accuracy and is an important source of identification error (Fig. 15).

Rice mapping strongly depends on the distinctive spectral characteristics of the flooding period. Its spectral characteristics

in the growing and harvesting periods are very similar to those of other summer crops. Therefore, previous studies all chose to capture the characteristics of the flooding period. However, optical and SAR images have their own limits during this short

290 period. In this study, we combined two sources of satellite images together to overcome the limitations of each source of satellite data. However, this combination was still relatively simple. Some recent data fusion studies use machine learning methods to reconstruct high-quality optical data with both high spatial and temporal resolutions. We hope that these kinds of reconstructed datasets will help solve the limitations of optical images and help to produce more accurate single-season rice maps.

295 **5. Data availability**

The distribution map of single-season rice of 21 provincial administrative regions in China from 2017 to 2022 is available at https://doi.org/10.57760/sciencedb.06963 (Shen et al., 2023). The file format of the product is GeoTIFF with the spatial reference of WGS84 (EPSG:4326).

6. Conclusions

300 This study proposed a new rice mapping method based on the time-weighted dynamic time warping (TWDTW) method. The TWDTW distances of the shortwave infrared1 (SWIR1) band from optical images and of the VH band from synthetic aperture radar (SAR) images were combined according to a weight, and the number of good optical observations was used to determine the weight. By using this method, this study produced distribution maps of single-season rice in China from 2017 to 2022 at 10-m or 20-m resolution. The overall accuracy over 21 provincial administrative regions averaged 85.23 % based

305 on 108195 samples; the average R^2 was 0.83 over three years compared with county-level statistical planting areas. However, the method did not fully resolve the limitations of optical and SAR images as clouds and the fragmentation of the rice fields still affected the accuracy of the distribution map. In general, this study produced high-resolution single-season rice maps of China, and the method can be easily applied to other regions and the maps can be updated annually.

Author contributions

310 WY and RS designed the research. BP, RS, QP, JD, XC, XZ, TY, and JH performed the investigation. RS, BP, JD, and WY developed the method. RS implemented the computer code, performed the formal analysis, visualized the results, and wrote the manuscript. WY and QP edited and revised the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This study was supported by the Open Research Program of the International Research Center of Big Data for Sustainable Development Goals (Grant No. CBAS2023ORP02).

References

Beaudoin, A., Stussi, N., Troufleau, D., Desbois, N., Piet, L., and Deshayes, M.: On the use of ERS-1 SAR data over hilly terrain: necessity of radiometric corrections for thematic applications, in: 1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications, 1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications, Firenze, Italy, 2179–2182, https://doi.org/10.1109/IGARSS.1995.524141, 1995.

Belgiu, M. and Csillik, O.: Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sensing of Environment, 204, 509–523, https://doi.org/10.1016/j.rse.2017.10.005, 2018.

Boryan, C., Yang, Z., Mueller, R., and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, Geocarto International, 26, 341–358, https://doi.org/10.1080/10106049.2011.562309, 2011.

Cao, B., Yu, L., Naipal, V., Ciais, P., Li, W., Zhao, Y., Wei, W., Chen, D., Liu, Z., and Gong, P.: A 30 m terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine, Earth Syst. Sci. Data, 13, 2437–2456, https://doi.org/10.5194/essd-13-2437-2021, 2021.

Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., Biradar, C., and Moore, B.: Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine, Remote Sensing of Environment, 185, 142–154, https://doi.org/10.1016/j.rse.2016.02.016, 2016.

Dong, J., Fu, Y., Wang, J., Tian, H., Fu, S., Niu, Z., Han, W., Zheng, Y., Huang, J., and Yuan, W.: Early-season mapping of winter wheat in China based on Landsat and Sentinel images, Earth Syst. Sci. Data, 12, 3081–3095, https://doi.org/10.5194/essd-12-3081-2020, 2020.

Elert, E.: Rice by the numbers: A good grain, Nature, 514, S50–S51, https://doi.org/10.1038/514S50a, 2014.

FAO: World Food and Agriculture – Statistical Yearbook 2021, FAO, https://doi.org/10.4060/cb4477en, 2021.

Fiorillo, E., Di Giuseppe, E., Fontanelli, G., and Maselli, F.: Lowland Rice Mapping in Sédhiou Region (Senegal) Using Sentinel 1 and Sentinel 2 Data and Random Forest, Remote Sensing, 12, 3403, https://doi.org/10.3390/rs12203403, 2020.

Guan, X., Huang, C., Liu, G., Meng, X., and Liu, Q.: Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance, Remote Sensing, 8, 19, https://doi.org/10.3390/rs8010019,

2016.

345 Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Cheng, F., Zhuang, H., Zhang, J., and Tao, F.: NESEA-Rice10: high-resolution annual paddy rice maps for Northeast and Southeast Asia from 2017 to 2019, Earth Syst. Sci. Data, 13, 5969–5986, https://doi.org/10.5194/essd-13-5969-2021, 2021.

Huang, X., Fu, Y., Wang, J., Dong, J., Zheng, Y., Pan, B., Skakun, S., and Yuan, W.: High-Resolution Mapping of Winter Cereals in Europe by Time Series Landsat and Sentinel Images for 2016–2020, Remote Sensing, 14, 2120, 350 https://doi.org/10.3390/rs14092120, 2022.

IPCC: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.

Kuenzer, C. and Knauer, K.: Remote sensing of rice crop areas, International Journal of Remote Sensing, 34, 2101–2139, 355 https://doi.org/10.1080/01431161.2012.738946, 2013.

Li, J. and Chen, B.: Global Revisit Interval Analysis of Landsat-8 -9 and Sentinel-2A -2B Data for Terrestrial Monitoring, Sensors, 20, 6631, https://doi.org/10.3390/s20226631, 2020.

Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., and de Queiroz, G. R.: A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping, IEEE Journal of Selected Topics in Applied Earth Observations and 360 Remote Sensing, 9, 3729–3739, https://doi.org/10.1109/JSTARS.2016.2517118, 2016.

Millard, K. and Richardson, M.: On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping, Remote Sensing, 7, 8489–8515, https://doi.org/10.3390/rs70708489, 2015.

Mohammadi, A., Khoshnevisan, B., Venkatesh, G., and Eskandari, S.: A Critical Review on Advancement and Challenges of Biochar Application in Paddy Fields: Environmental and Life Cycle Cost Analysis, Processes, 8, 1275, 365 https://doi.org/10.3390/pr8101275, 2020.

National Bureau of Statistics of China: China Statistical Yearbook, MARY MARTIN, 2020.

Nguyen, D. B., Gruber, A., and Wagner, W.: Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data, Remote Sensing Letters, 7, 1209–1218, https://doi.org/10.1080/2150704X.2016.1225172, 2016.

Oguro, Y., Suga, Y., Takeuchi, S., Ogawa, M., Konishi, T., and Tsuchiya, K.: Comparison of SAR and optical sensor data for 370 monitoring of rice plant around Hiroshima, Advances in Space Research, 28, 195–200, https://doi.org/10.1016/S0273-1177(01)00345-3, 2001.

Oliver, C. and Quegan, S. (Eds.): Understanding synthetic aperture radar images, SciTech Publishing, Raleigh, NC, 479 pp., 2004.

Pan, B., Zheng, Y., Shen, R., Ye, T., Zhao, W., Dong, J., Ma, H., and Yuan, W.: High Resolution Distribution Dataset of Double-375 Season Paddy Rice in China, Remote Sensing, 13, 4609, https://doi.org/10.3390/rs13224609, 2021.

Petitjean, F., Inglada, J., and Gancarski, P.: Satellite Image Time Series Analysis Under Time Warping, IEEE Trans. Geosci. Remote Sensing, 50, 3081–3095, https://doi.org/10.1109/TGRS.2011.2179050, 2012.

Phan, H., Le Toan, T., Bouvet, A., Nguyen, L., Pham Duy, T., and Zribi, M.: Mapping of Rice Varieties and Sowing Date Using X-Band SAR Data, Sensors, 18, 316, https://doi.org/10.3390/s18010316, 2018.

380 Qiu, B., Luo, Y., Tang, Z., Chen, C., Lu, D., Huang, H., Chen, Y., Chen, N., and Xu, W.: Winter wheat mapping combining variations before and after estimated heading dates, ISPRS Journal of Photogrammetry and Remote Sensing, 123, 35–46, https://doi.org/10.1016/j.isprsjprs.2016.09.016, 2017.

Shen, R., Dong, J., Yuan, W., Han, W., Ye, T., and Zhao, W.: A 30 m Resolution Distribution Map of Maize for China Based on Landsat and Sentinel Images, Journal of Remote Sensing, 2022, 9846712, https://doi.org/10.34133/2022/9846712, 2022.

385 Shen, R., Pan, B., Peng, Q., Dong, J., Chen, X., Zhang, X., Ye, T., Huang, J., and Yuan, W.: High-resolution distribution maps of single-season rice in China from 2017 to 2022, Science Data Bank [dataset], https://doi.org/10.57760/sciencedb.06963, 2023.

Skakun, S., Vermote, E., Roger, J.-C., Franch, B., 1 Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA, and 2 NASA Goddard Space Flight Center Code 619, 8800 Greenbelt Road, Greenbelt, MD 20771,
390 USA: Combined Use of Landsat-8 and Sentinel-2A Images for Winter Crop Mapping and Winter Wheat Yield Assessment at Regional Scale, AIMS Geosciences, 3, 163–186, https://doi.org/10.3934/geosci.2017.2.163, 2017.

Sudmanns, M., Tiede, D., Augustin, H., and Lang, S.: Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass, International Journal of Digital Earth, 13, 768–784, https://doi.org/10.1080/17538947.2019.1572799, 2020.

395 Thorp, K. R. and Drajat, D.: Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia, Remote Sensing of Environment, 265, 112679, https://doi.org/10.1016/j.rse.2021.112679, 2021.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., and Koetz, B.: Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions, Remote Sensing, 8, 55, https://doi.org/10.3390/rs8010055, 2016.

400 Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., and Ceschia, E.: Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications, Remote Sensing of Environment, 199, 415–426, https://doi.org/10.1016/j.rse.2017.07.015, 2017.

Wakabayashi, H., Motohashi, K., Kitagami, T., Tjahjono, B., Dewayani, S., Hidayat, D., and Hongo, C.: FLOODED AREA EXTRACTION OF RICE PADDY FIELD IN INDONESIA USING SENTINEL-1 SAR DATA, Int. Arch. Photogramm.
405 Remote Sens. Spatial Inf. Sci., XLII-3/W7, 73–76, https://doi.org/10.5194/isprs-archives-XLII-3-W7-73-2019, 2019.

Xiao, X., He, L., Salas, W., Li, C., Moore, B., Zhao, R., Frolking, S., and Boles, S.: Quantitative relationships between field-measured leaf area index and vegetation index derived from VEGETATION images for paddy rice fields, International Journal of Remote Sensing, 23, 3595–3604, https://doi.org/10.1080/01431160110115799, 2002.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., and Moore, B.: Mapping paddy rice agriculture in
410 southern China using multi-temporal MODIS images, Remote Sensing of Environment, 95, 480–492, https://doi.org/10.1016/j.rse.2004.12.009, 2005.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J. Y., Salas, W., and Moore, B.: Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images, Remote Sensing of Environment, 100, 95–113, https://doi.org/10.1016/j.rse.2005.10.004, 2006.

415 Yan, J., Yang, Z., Li, Z., Li, X., Xin, L., and Sun, L.: Drivers of cropland abandonment in mountainous areas: A household decision model on farming scale in Southwest China, Land Use Policy, 57, 459–469, https://doi.org/10.1016/j.landusepol.2016.06.014, 2016.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y., and Xiao, X.: The 10-m crop type maps in Northeast China during 2017–2019, Sci Data, 8, 41, https://doi.org/10.1038/s41597-021-00827-9, 2021.

420 Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., and Feng, M.: Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China, Remote Sensing, 11, 2673, https://doi.org/10.3390/rs11222673, 2019.

Zheng, B., Myint, S. W., Thenkabail, P. S., and Aggarwal, R. M.: A support vector machine to identify irrigated crop types using time-series Landsat NDVI data, International Journal of Applied Earth Observation and Geoinformation, 34, 103–112, 425 https://doi.org/10.1016/j.jag.2014.07.002, 2015.

Zheng, Y., Li, Z., Pan, B., Lin, S., Dong, J., Li, X., and Yuan, W.: Development of a Phenology-Based Method for Identifying Sugarcane Plantation Areas in China Using High-Resolution Satellite Datasets, Remote Sensing, 14, 1274, https://doi.org/10.3390/rs14051274, 2022a.

Zheng, Y., dos Santos Luciano, A. C., Dong, J., and Yuan, W.: High-resolution map of sugarcane cultivation in Brazil using a 430 phenology-based method, Earth Syst. Sci. Data, 14, 2065–2080, https://doi.org/10.5194/essd-14-2065-2022, 2022b.

Zhong, L., Hu, L., Zhou, H., and Tao, X.: Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US, Remote Sensing of Environment, 233, 111411, https://doi.org/10.1016/j.rse.2019.111411, 2019.

Zhou, Y., Dong, J., Liu, J., Metternicht, G., Shen, W., You, N., Zhao, G., and Xiao, X.: Are There Sufficient Landsat 435 Observations for Retrospective and Continuous Monitoring of Land Cover Changes in China?, Remote Sensing, 11, 1808, https://doi.org/10.3390/rs11151808, 2019.

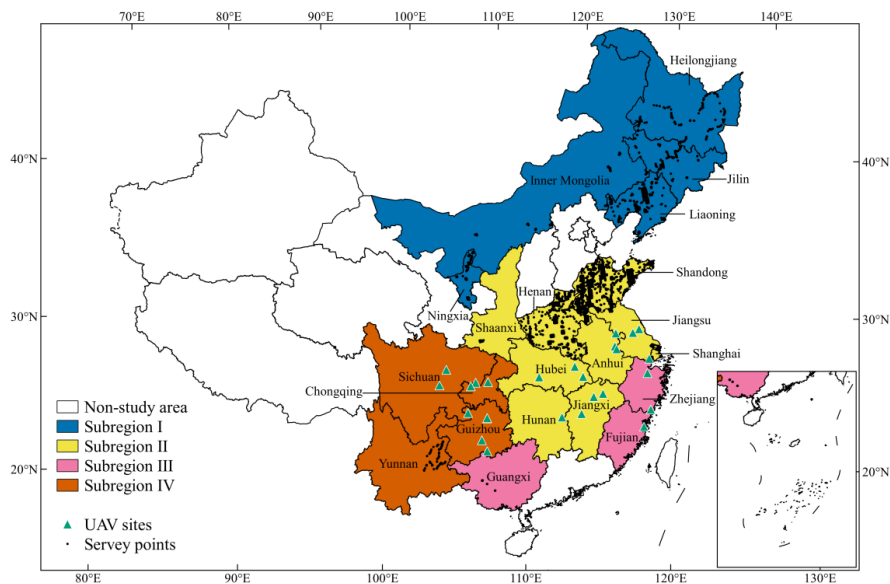



Figure 1: The study area includes 21 provincial administrative regions in China and is divided into four subregions (colored

440 areas). The black dots indicate the samples obtained from the survey, and the green triangles indicate the unmanned aerial

vehicle (UAV) survey sites.

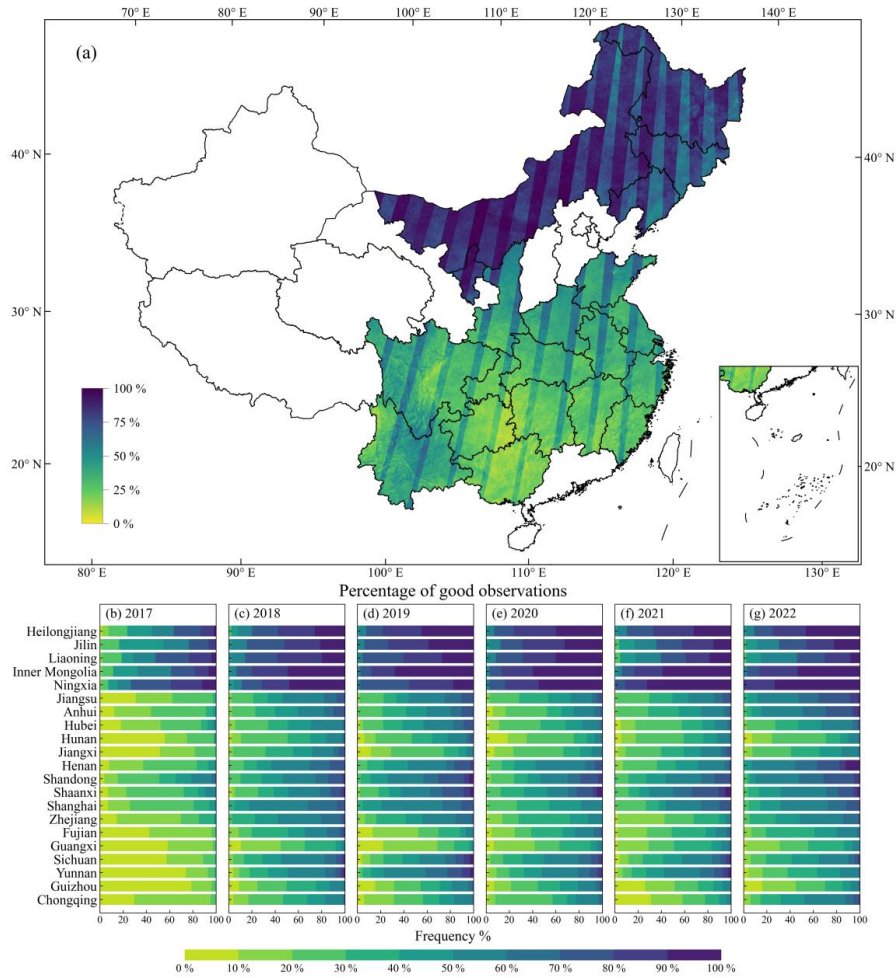



Figure 2: Percentage of good observations of Sentinel-2 during the study period of 2017–2022 (a). The bottom row (b–g)

445   shows the frequencies of percentages of good observations in each province during the study period of each year.

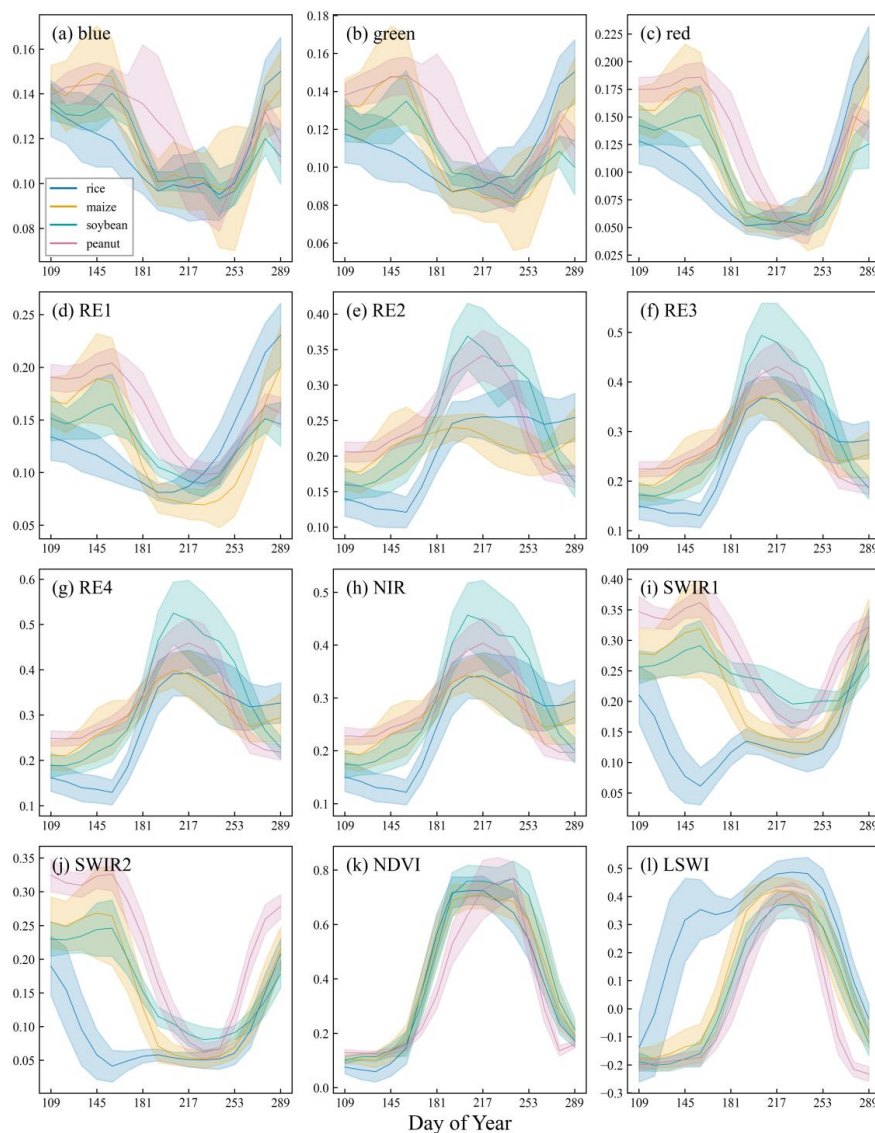



Figure 3: Time series of 10 optical bands or indexes over four main crop types in Jilin province in 2019. Solid lines indicate

the average time series, and the shaded error bands represent the standard deviations.

450

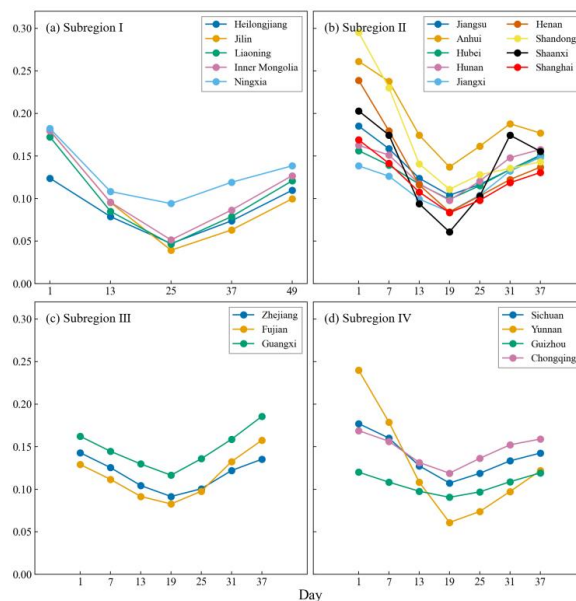
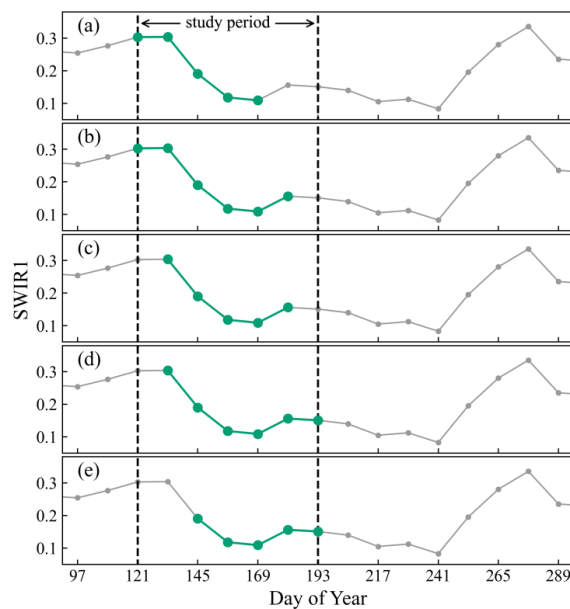



Figure 4: Standard SWIR1 time series of single-season rice in 21 provincial administrative regions in four subregions.

455   **Figure 5:** A time series of SWIR1 in Jilin Province. The period between dashed lines (DOY 121–193) is the study period.

Five green curves are the time series selected to calculate the distance with the standard time series using the TWDTW

method.

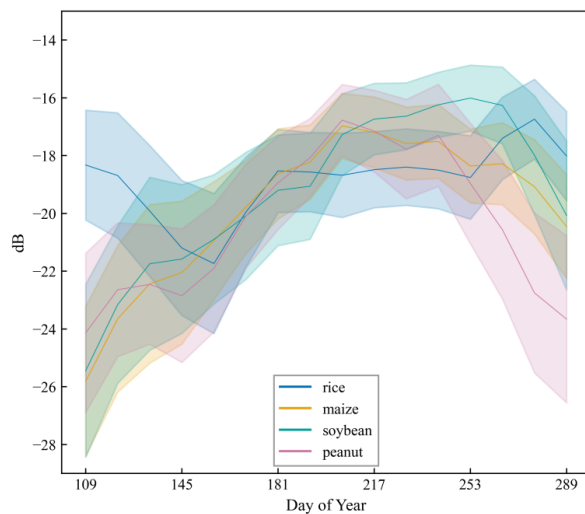



Figure 6: Time series of the VH band over four main crop types in Henan Province in 2019. Solid lines indicate the average
time series; the shaded error bands represent the standard deviations.

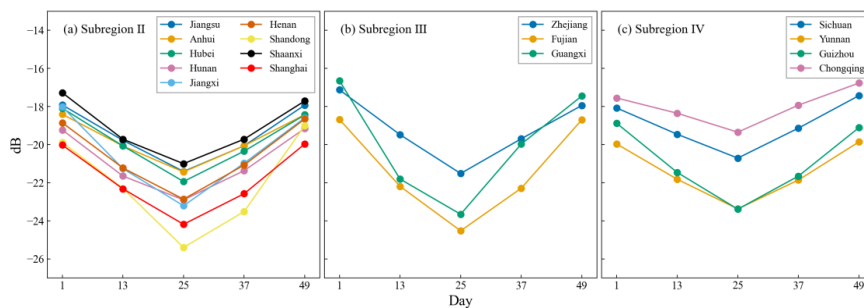



Figure 7: Standard VH time series of single-season rice in 16 provincial administrative regions in Subregions II, III, and IV.

465

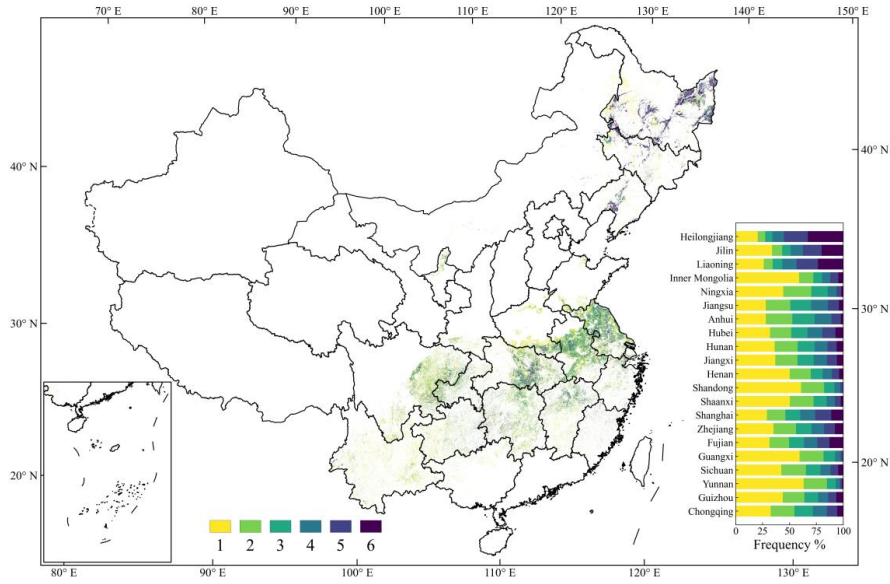



Figure 8: Planting frequency of single-season rice in China from 2017 to 2022.

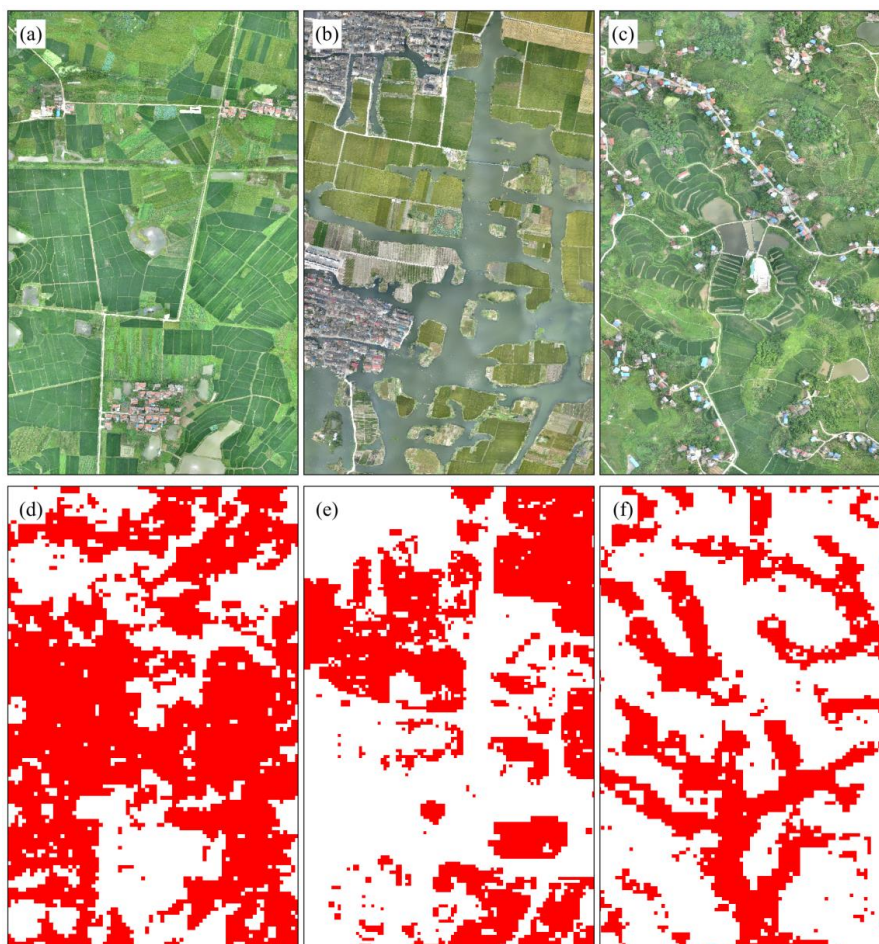



Figure 9: Distribution map in three UAV sites of Hubei (114°47′49″ E, 31°1′11″ N), Zhejiang (120°32′33″ E, 29°57′14″ N), and Sichuan (106°44′15″ E, 30°19′5″ N). (a)–(c) are very-high-resolution UAV images taken at three sites on July 8, 2018, October 12, 2018, and July 13, 2018, respectively. (d)–(f) are distribution maps with identified single-season rice pixels indicated in red.

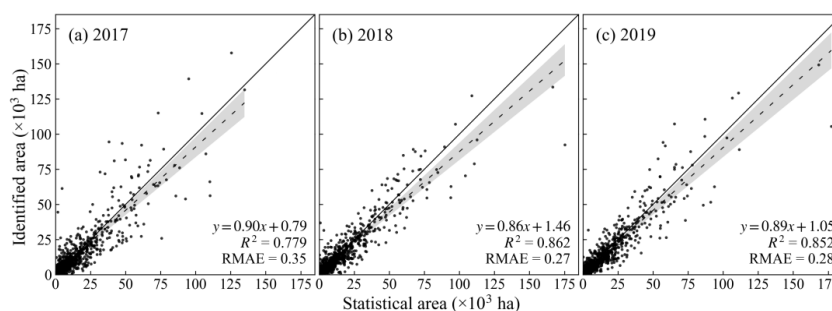
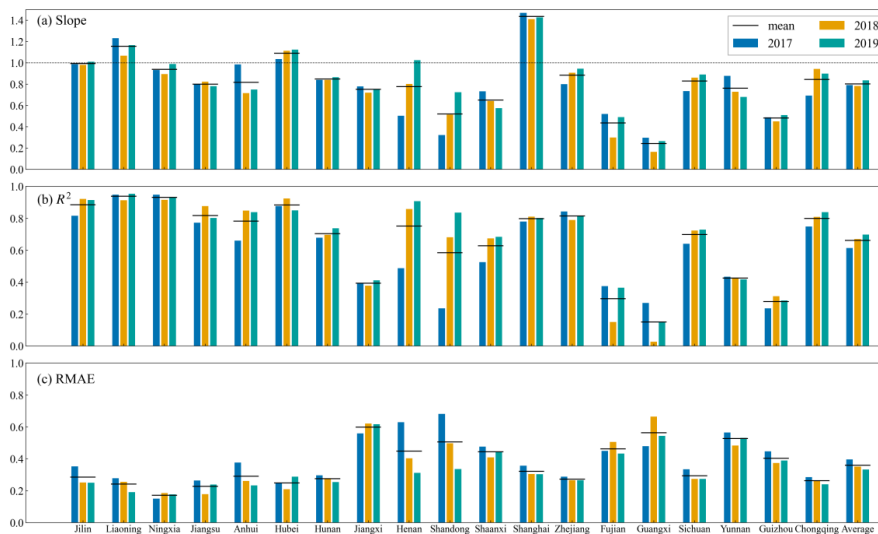



Figure 10: County-level comparison of identified and statistical planting areas of 2017–2019. Solid lines are 1:1 lines; dashed

lines are regression lines. The confidence intervals are shaded in gray.

480   **Figure 11:** Comparison between identified and statistical planting areas at the county-level of 2017–2020 in 19 provincial

administrative regions.

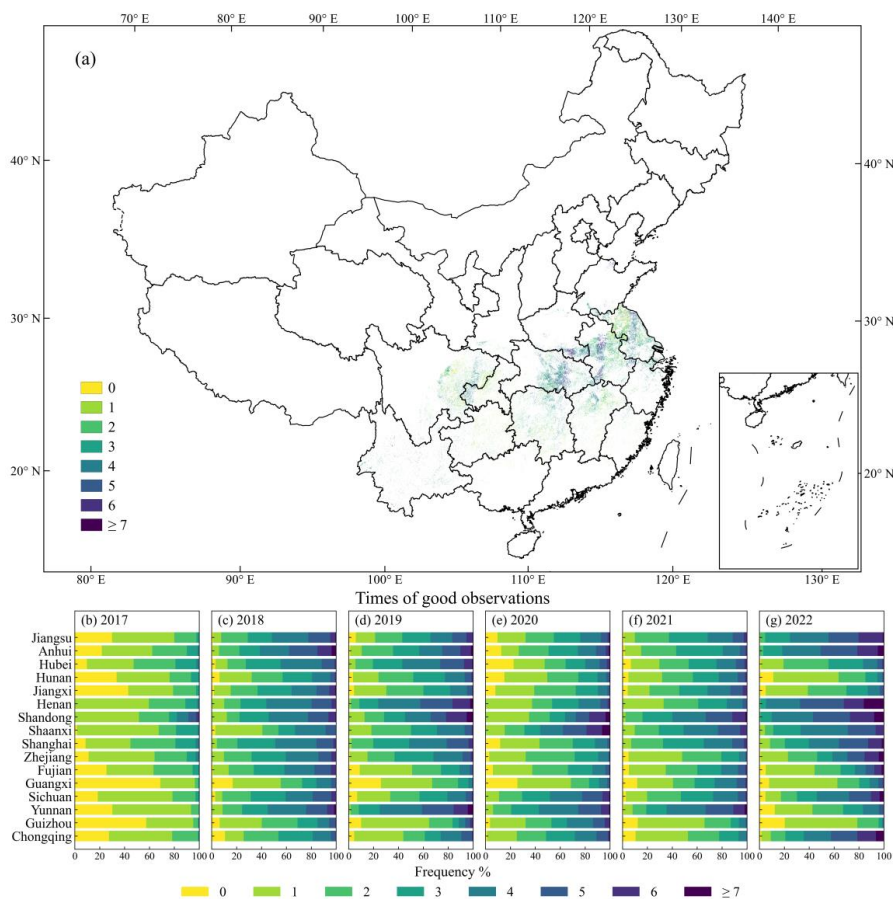
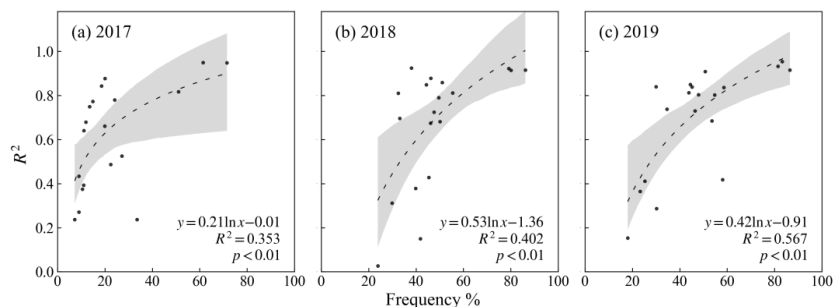



Figure 12: Times of good observations of Sentinel-2 during the time period corresponding to the minimum TWDTW

485   distance in identified single-season rice pixels in 19 provincial administrative regions in Subregion II, III, and IV in 2019.

The bottom row (b–g) shows the times of good observations in each province during the time period corresponding to the

minimum TWDTW distance of each year.

490   **Figure 13:** Relationship between identification accuracies (R^2 of county-level comparison with the statistical planting area)

and provincial mean of good observation frequencies in 19 provincial administrative regions from 2017 to 2020. Dashed lines

are regression lines; the confidence intervals are shaded in gray.

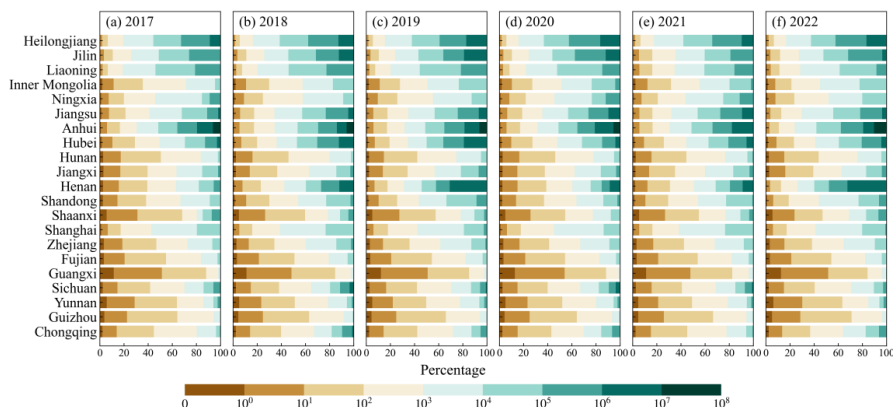



Figure 14: Distribution of the number of single patches.

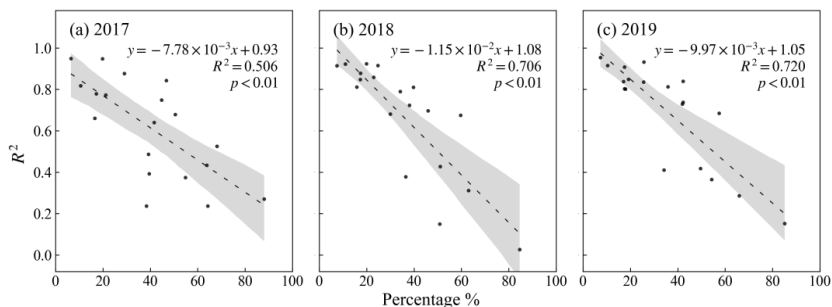



Figure 15: Relationship between identification accuracies (R^2 of county-level comparison with the statistical planting area)

and percentage of patches with a size less than or equal to 100 in 19 provincial administrative regions from 2017 to 2020.

Dashed lines are regression lines; the confidence intervals are shaded in gray.

Table 1: Number of available county-level statistical data

Province	Total number of counties	2017	2018	2019
Heilongjiang	128	0	0	0
Jilin	60	49	50	50
Liaoning	100	20	21	21
Inner Mongolia	102	0	0	0
Ningxia	22	10	10	10
Jiangsu	99	66	71	60
Anhui	104	57	90	68
Hubei	103	78	78	77
Hunan	122	111	119	117

Jiangxi	100	23	22	22
Henan	158	39	50	38
Shandong	137	18	18	18
Shaanxi	107	21	21	21
Shanghai	17	9	9	9
Zhejiang	90	22	17	22
Fujian	85	24	19	20
Guangxi	110	44	29	27
Sichuan	183	95	102	102
Yunnan	129	125	125	125
Guizhou	88	56	56	56
Chongqing	38	37	37	37

Table 2: Confusion matrices of the distribution map in 21 provincial administrative regions.

Province	Class	SR[1]	Other	UA (%)	PA (%)	OA (%)
Heilongjiang	SR[2]	164	5	89.13	97.04	95.70
	Other	20	393	98.74	95.16	
Jilin	SR	5598	16	90.32	99.71	96.77
	Other	600	12840	99.88	95.54	
Liaoning	SR	5890	15	92.41	99.75	96.87
	Other	484	9541	99.84	95.17	
Inner Mongolia	SR	84	10	97.67	89.36	98.09
	Other	2	531	98.15	99.62	
Ningxia	SR	47	5	69.12	90.38	91.03
	Other	21	217	97.75	91.18	
Jiangsu	SR	2249	58	62.42	97.49	67.14
	Other	1354	636	91.64	31.96	
Anhui	SR	1133	168	54.55	87.09	63.88
	Other	944	834	83.23	46.91	
Hubei	SR	2034	206	87.18	90.8	87.31
	Other	299	1441	87.49	82.82	
Hunan	SR	397	62	62.92	86.49	83.77
	Other	234	1131	94.80	82.86	
Jiangxi	SR	603	622	70.2	49.22	67.18
	Other	256	1194	65.75	82.34	
Henan	SR	2694	57	95.80	97.93	99.13
	Other	118	17315	99.67	99.32	
Shandong	SR	1977	241	72.47	89.13	89.60
	Other	751	6566	96.46	89.74	
Shaanxi	SR	454	43	71.50	91.35	84.18
	Other	181	738	94.49	80.30	
Shanghai	SR	128	7	83.12	94.81	88.13

	Other	26	117	94.35	81.82	
Zhejiang	SR	900	200	85.88	81.82	90.67
	Other	148	2480	92.54	94.37	
Fujian	SR	530	108	94.81	83.07	90.36
	Other	29	754	87.47	96.30	
Guangxi	SR	108	25	46.96	81.2	79.21
	Other	122	452	94.76	78.75	
Sichuan	SR	2031	353	62.51	85.19	77.16
	Other	1218	3275	90.27	72.89	
Yunnan	SR	78	72	69.64	52.00	88.87
	Other	34	768	91.43	95.76	
Guizhou	SR	1836	477	83.42	79.38	89.39
	Other	365	5257	91.68	93.51	
Chongqing	SR	486	408	54.30	54.36	71.07
	Other	409	1521	78.85	78.81	

505 [1]number of field surveyed samples. [2]number of identified samples. SR is single-season rice.