Extraction of Sugarcane Planting Area Based on Similarity of NDVI Time Series

SHIQIN DENG1,2, MAOFANG GAO1, (Member, IEEE), CHAO REN3, SHILEI LI1, AND YONGJIAN LIANG4

1Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
2Guangxi Zhuang Autonomous Region Remote Sensing Institute, Nanning 530000, China
3Guangxi Key Laboratory of Spatial Information and Surveying and Mapping, School of Surveying, Mapping and Geographic Information, Guilin University of Technology, Guilin 541004, China
4Guangxi South Subtropical Agricultural Science Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 532415, China

Corresponding authors: Maofang Gao (gaomaofang@caas.cn) and Chao Ren (renchao@glut.edu.cn)

This work was supported by the Fundamental Research Funds for Central Non-Profit Scientific Institution under Grant 1610132020015.

ABSTRACT
Timely and accurate estimation of the sugarcane planting area is of vital importance to the country’s agricultural production and sugar development. In remote sensing crop mapping based on spectral similarity, there will be a phenomenon of foreign matters with the same spectrum by obtaining an accurate crop reference curve for crop identification, limiting mapping accuracy. In this study, we improved the spectral reconstruction method based on singular value decomposition (SR-SVD). A decision tree model was established based on the similarity of the sugarcane Normalized Difference Vegetation Index (NDVI) time series curve and the fluctuation range of NDVI in different growth periods. Using the Sentinel-2 (Level-2A) image data set to extract sugarcane planting area in two regions of Chongzuo City, Guangxi, China, the overall accuracy was higher than 96%, respectively. The results show that through the spectral similarity and the determination of the threshold fluctuation range, not only high-precision mapping of sugarcane can be achieved, but the problem of “same spectrum with different objects” can also be solved. Therefore, this method can provide accurate information on the sugarcane planting areas and technical support for monitoring the structure of sugarcane planting in the region.

INDEX TERMS
Decision-tree, normalized difference vegetation index, singular value decomposition, sugarcane.

I. INTRODUCTION
Sugarcane is a perennial crop in the grass family grown [1]. It is one of the primary raw materials for sugar production and one of the essential raw materials for bioenergy [2]. It exhibits dual characteristics of an agricultural product and a product of the light industry [3]. Since sugarcane growth depends on sufficient sunlight, high temperature, and much precipitation, most sugarcane cultivation is distributed in tropical and subtropical regions [4], [5]. In China, 65% of the total sugarcane is produced in Guangxi. Its high economic value and multiple uses serve as essential sources of income and provided employment opportunities for the people of Guangxi.

Therefore, timely information on sugarcane is essential for agricultural production and structure adjustments, monitoring the growth and yield of sugarcane, and assessing sugar security [6].

So far, the acquisition of sugarcane planting area information and yield information has mainly relied on agronomic methods. However, the conventional agronomy and meteorology models have several disadvantages, such as complex calculations, heavy field workload, and a strong influence of anthropogenic factors [7]. With the rapid development of remote sensing technology, remote sensing provides an effective means for quick and accurate estimation of the spatial distribution of crops [8]. The application of high-resolution satellite data provides excellent possibilities for crop classification and crop distribution mapping.
In the past ten years, some studies have proven to be very effective in extracting sugarcane planting area [9], [10], [11]. Based on the principle of phenomenology [12], Mulianga et al. used Landsat 8 images to obtain the shape, texture, and other information of different crops in the training sample set and used the maximum likelihood classifier to extract the sugarcane planting area [6]. Chen et al. used MODIS/HJ-1 CCD for fusion and constructed a sugarcane recognition model through the Normalized Difference Vegetation Index (NDVI) change rate and the sample automatic training threshold [13]. The overall accuracy can reach 92.17%. Henry et al. used the gradient descent algorithm to determine the best parameters of the Logistic regression model and took NDVI time series data as input to extract sugarcane planting information [14]. Some object-oriented image analysis methods have also proved to be very effective in extracting sugarcane information. Vieira et al. used segmented Landsat TM and EMT+ images to obtain training data sets [15] and achieved high-precision mapping of sugarcane in three municipalities in the northwest of Sao Paulo through object-based image analysis (OBIA) and data mining (DM). In southern China, the sugarcane planting land is broken and the sugarcane extraction process is complicated, which leads to the low classification accuracy of sugarcane extraction by object-oriented analysis. The overall accuracy of sugarcane extraction is generally below 90%. Currently, great progress has been made in crop acreage extraction through the use of spectral similarity methods [16], [17], [18]. The similarity between the target and reference time series is evaluated based on a specific spectral similarity and classification rules are developed based on their similarity. However, this approach has rarely been applied to sugarcane planting area studies. An important issue in crop identification by spectral similarity is the selection of crop reference curves. To improve the crop reference curve selection problem, Li et al proposed a method based on singular value decomposition (SR-SVD) to decompose and reconstruct the NDVI time series curve of winter wheat to achieve high accuracy mapping of winter wheat [17]. In northern China, the crop species are single and the planting area is large, so SR-SVD method can effectively extract crop planting information. However, in south China, where there are many crop species and land parcels are broken, only SR-SVD method is used to obtain crop information, resulting in “same spectrum with different objects”. In order to solve this problem, we have improved the SR-SVD. We set the threshold range of NDVI time series of sugarcane by analyzing the variation pattern of NDVI time series of sugarcane training sample points. And it could provide the possibility of extracting sugarcane acreage with high accuracy.

II. STUDY AREA AND DATA

A. STUDY AREA AND SUGARCANE PHENOLOGY

The study area is located in Kongxiang, Xinhe Town, Jiangzhou District, Chongzuo City (107°16’07.33”E, 22°33’26.24”N), Guangxi Zhuang Autonomous Region, and Longzhou County Institute of Subtropical Botany (106°47’37.87”E, 22°20’41.80”N) (Figure 1) and is located south of the Tropic of Cancer. It experiences a subtropical monsoon climate with abundant rainfall. The annual number of sunshine hours is more than 1600 h, the annual average temperature is 20.8–22.4 °C, and the annual average rainfall is more than 1200 mm. Chongzuo City has abundant plantations and is a critical sugarcane production base in the country. The main cash crops are sugarcane, banana, corn, tangerine, pineapple, walnut.

FIGURE 1. Geographical location of the study area.

We conducted field research in the study area in July 2019 to understand the local habits of sugarcane cultivation and its growth patterns. Talking to local farmers who grow sugarcane, we learned that the growth pattern of sugarcane is the same throughout southwest China. Due to the abundant rainfall in the spring and summer in southwest China, sugarcane is raised through rain. Fertilizer is applied twice a year in March (375kg of N/P/K compound fertilizer and 1125kg of phosphate per hectare), and top dressing starts in May or June (150kg of urea and 225kg of potassium chloride per hectare). Therefore, the sugarcane growing cycle is the same in southwest China, growing from March and maturing in December every year. The exact time of the harvest is determined by the farmers’ planting habits. After the sugarcane has been harvested, the roots are retained and continue to grow.
TABLE 1. Sugarcane phenology calendar.

| Month | Phenological characteristics of sugarcane | NDVI fluctuation trend |
|-------|------------------------------------------|------------------------|
|       | 1-2 Sugarcane harvest ratoon sugarcane continues to grow; 3-4 Newly planted sugarcane planting (Perennials are generally retained for 3 years) | Rise |
|       | Seedling                                 | Tillering              |
|       | The stage from tiller to jointing of sugarcane seedling | Rise |
|       | Rapid growth                             | Floating up and down slightly |
|       | During this period, sugarcane grows fastest |         |
|       | Mature                                   |                       |

| January to April | May and June | July to September | October to December |
|------------------|--------------|-------------------|---------------------|
| Seeding          | Tillering    | Rapid growth      | Mature              |

Ratoon-sugarcane is usually retained for three years. To explore the pattern of change in the NDVI time series curve of sugarcane, we selected as many and good quality images as possible for one growing period (2019.3-2020.3). We did not select the 2020 image for the study area Xinhe due to the poor quality of the Sentinel-2A image between December 2019 and March 2020. We collected the phenological calendar of sugarcane growth in the study area and took photographs of sugarcane in the study area during different growth periods, as shown in Table 1. The combination of the phenological calendar and photographs can accurately reflect the growth status of sugarcane in different growing periods [19], [20].

B. SATELLITE DATA
Sentinel-2 satellite is a multispectral high-resolution satellite consisting of two polar-orbiting satellites, namely 2A and 2B. The revisit time for one Sentinel-2 satellite is 10 days, while it can be up to 5 days for both satellites. Sentinel-2 carries a multispectral instrument (MSI) with a width of up to 290km, covering 13 spectra bands. In this study, we used Google Earth Engine to obtain Sentinel-2 (Level-2A) images of the study area. We selected 19 images for Longzhou and 12 images for Xinhe. The images acquired by Longzhou are from March 2019 to March 2020, and the images acquired rather at Xinhe are from March 2019 to December 2019. The images we obtained cover the entire sugarcane growth period. In order to obtain the reference spectrum accurately, we removed the images with high cloud content and retained the high-quality images as much as possible. Level-2A is the surface reflectance, which eliminates the influence of clouds and atmospheric components. The image information is shown in Table 2.

C. SAMPLE DATA
In this study, we obtained a large number of reference data sets from high-resolution Google Earth imagery, field survey samples, and UAV images (as shown in Figure 2). Then, we divided the sugarcane points in the two study areas into training and test samples, as shown in Table 3 below.

FIGURE 2. UAV images of the study area on July 6th,2019.
decision tree from the similarity of the curves and the thresholds of the curves. Through the shape of the NDVI time series before and after the reconstruction of the image element and the threshold value corresponding to the NDVI time series, we determined whether the image element belongs to sugarcane. The experimental flowchart is shown in Figure 3.

### A. PREPARING NDVI TIME-SERIES IMAGES OF SUGARCANE

In agricultural remote sensing research, many vegetation indices have proven effective in extracting crop planting information, such as NDVI [17] and Soil Adjusted Vegetation Index (SAVI) [19], [20]. NDVI is widely used for crop identification and classification. NDVI is both an important indicator for assessing crop growth and a major basis for estimating crop yield [21], [22]. This study selects NDVI time series data to extract the spatial distribution of sugarcane. It is calculated as follows:

\[
NDVI = \frac{NIR - R}{NIR + R} \tag{1}
\]

where R is the fourth band of the Sentinel-2 images, i.e., the red band, and NIR is the eighth band of the Sentinel-2 images, i.e., near-infrared band.

Only instantaneous spectral characteristics of land cover can be acquired using single-phase remote sensing images. However, multiple time phases can conduct multiple observations of the same surface to acquire dynamic changes of the surface spectrum over a period [21], [22]. In this study, Sentinel-2 (Level-2A) images of the study areas in different time phases were collected for NDVI calculation; subsequently, the time series NDVI images were layer-stacked. Combining field survey information and NDVI time series images, we obtained NDVI time series of sugarcane and main crops in the research area. It can be seen from Figure 4 that the time series curve of sugarcane NDVI is different from that of other crops. Then, this study chooses the NDVI time series curve as the criterion to distinguish sugarcane fields from other land cover types.

### B. THE SPECTRAL RECONSTRUCTION METHOD BASED ON SINGULAR VALUE DECOMPOSITION (SR-SVD)

Singular Value Decomposition (SVD) is a data-driven signal processing method. Its essence is to decompose a large set of related variables into a small set of unrelated SVs and arrange the SVs according to the monotonic decrease in the variability of each variable [23]. Generally, the SVD of \( m \times n \) matrix \( M \) can be expressed as:

\[
M = U \Sigma V^T \tag{2}
\]

where \( U \) is an \( m \times m \) matrix and \( S \) is an \( m \times n \) non-negative real diagonal matrix with the elements on its diagonal being the singular values of matrix \( M \) arranged in descending order. \( V \) is an \( n \times n \) matrix, and \( T \) refers to transpose.

The core of the SR-SVD method is as follows. Firstly, we obtained many singular vectors by decomposing the NDVI time series and using the SVD method. Then, the NDVI time series were reconstructed by selecting different numbers of singular vectors. At last, we achieved the extraction of crop acreage data by evaluating the similarity between the original NDVI time series and the reconstructed NDVI time series for each image element. This is simultaneously the purpose of reconstructing the NDVI time series. The reconstruction of the NDVI time series can be formulated as follows [17]:

\[
NDVI_r = \sum_{i=0}^{n} \omega_i v_i \tag{3}
\]

where \( \text{NDVI}_r \) is the reconstructed NDVI time series, \( v \) is the singular vector (SV), \( \omega \) is the weight of the SV, \( n \) represents the total number of SVs used, and \( i \) is the serial number.

### C. SIMILARITY CALCULATION BETWEEN THE CURVES

We used three different indicators to evaluate the similarity between the reconstructed and original time series curves.
Spectral angle classification is also called spectral angle mapping, which is a type of optical harmonic matching technology [24], [25]. Spectral angle calculation uses the spectral curve of each pixel as a high-dimensional vector and measures the similarity between the spectra by calculating the angle between the two vectors. The smaller the angle, the higher is the similarity between the two spectra, and the higher is the possibility of the vectors belonging to the same object type. The spectral angle is calculated as follows

\[
\theta_{\text{SAM}}(\text{NDVI}_t, \text{NDVI}_r) = \arccos \frac{\sum_{i=1}^{n} \text{NDVI}_t \cdot \text{NDVI}_r}{\sqrt{\sum_{i=1}^{n} \text{NDVI}_t^2} \cdot \sqrt{\sum_{i=1}^{n} \text{NDVI}_r^2}}
\]  

where \(\theta_{\text{SAM}}\) represents the similarity, \(\text{NDVI}_t\) is the original NDVI time series, \(\text{NDVI}_r\) is the reconstructed NDVI time series, \(i\) is the serial number, and \(n\) is the total number of time series.

Coefficient of determination (R) measures the model’s fit and is widely used in evaluating the accuracy of regression equations and the similarity between the simulated values and the actual values of the samples. In this study, the R magnitude (correlation) was used to evaluate the similarity between the NDVI time series curve of the pixels in the study area and the reconstructed NDVI time series curve. The R is calculated as follows:

\[
R(\text{NDVI}_t, \text{NDVI}_r) = \frac{\sum_{i=1}^{n} (\text{NDVI}_t_i - \bar{\text{NDVI}}_t)(\text{NDVI}_r_i - \bar{\text{NDVI}}_r)}{\sqrt{\sum_{i=1}^{n} (\text{NDVI}_t_i - \bar{\text{NDVI}}_t)^2 \cdot \sum_{i=1}^{n} (\text{NDVI}_r_i - \bar{\text{NDVI}}_r)^2}}
\]

where \(R\) represents the similarity, \(\text{NDVI}_t_i\) is the original NDVI time series, \(\text{NDVI}_r_i\) is the mean of the original NDVI time series, \(\text{NDVI}_t_i\) is the reconstructed NDVI time series, \(\text{NDVI}_r_i\) is the mean of the reconstructed NDVI time series, \(i\) is the serial number, and \(n\) is the total number of time series.

Euclidean distance (ED) refers to the natural length of the actual distance between two m-dimensional spaces of vectors [26]. In this study, the Euclidean distance was used as an index to evaluate the similarity between the NDVI time series curve of all pixels in the study area and the reconstructed NDVI time series curve. The Euclidean distance is calculated as follows:

\[
D_{\text{Euclidean}} = \sqrt{\sum_{i=1}^{n} (\text{NDVI}_t_i - \text{NDVI}_r_i)^2}
\]

where \(D_{\text{Euclidean}}\) represents the ED, \(\text{NDVI}_t_i\) is the original NDVI time series, \(\text{NDVI}_r_i\) is the reconstructed NDVI time series, \(i\) is the serial number, and \(n\) is the total number of time series.

**D. DETERMINING THE NDVI FLUCTUATION THRESHOLD**

According to the number of image coverage in the two study areas, we analyzed the NDVI time series of sugarcane characteristics in different growth periods in the two study areas. Combining the ground-measured sample points and sugarcane phenology characteristics (the image acquisition time corresponds to the sugarcane NDVI value change trend during the sugarcane growth period), the NDVI value distribution of the sugarcane sample points in different growth periods was counted.
In Xinhe, the main land cover classes are sugarcane fields, hills, and a small number of other crops. We extracted three-time series points (TSP) from the Xinhe sugarcane NDVI time series, representing three different growth periods of sugarcane (seedling stage, rapid growth stage, mature stage). Through the statistics of the maximum and minimum NDVI of the sugarcane training samples in the three different growth periods, the fluctuation threshold of the sugarcane NDVI in the critical growth period is determined. In Longzhou, the main land cover classes are sugarcane, citrus, jute, pineapple, and other crops. Therefore, different growth periods NDVI TSP have been added to avoid confusing sugarcane with other crops with similar characteristics when selecting the representative growth period of the region for decision tree classification. We selected an NDVI TSP every other month, between the sugarcane seedling and the mature stage for Longzhou. Statistics of the maximum and minimum NDVI of the sugarcane training samples for each TSP. Determine the fluctuation range of the sugarcane NDVI threshold in each TSP.

Then, we found that in the middle of the sugarcane elongation period, NDVI reached its highest value and saturated. Saturation in this case means that the sugarcane NDVI values will not change any more widely over this period. Different varieties of sugarcane show differences in the maximum values during this period. Some varieties of sugarcane reach a maximum NDVI of 0.9 during this period, while others only reach around 0.8. The NDVI threshold of sugarcane fluctuates in a small range between the elongation and maturity periods. The fluctuation may be due to the influence of thin clouds or artificial pruning of branches and leaves.

In this study, we determined the range of NDVI threshold fluctuations at different time series points (different sugarcane growth periods). We used this fluctuation range as the condition for judging sugarcane pixels.

Based on the above determination and analysis of sugarcane’s NDVI time series threshold in the study area, we set two rules to determine whether a pixel is sugarcane. The rules are as follows: (1) The NDVI time series points of sugarcane pixels should be within the fluctuation range of the critical growth period threshold; (2) Between the elongation period and the maturity period, the difference between the front and back of the NDVI time series of sugarcane pixels must not exceed 0.3.

E. CONSTRUCTION OF DECISION TREE BASED ON CURVE SIMILARITY AND NDVI

The classification and regression decision tree (CART) is a primary classification and regression method for machine learning and is constructed mainly by relying on strategic choices for prediction by constructing a binary tree. The CART model was proposed by Breiman et al. and has been widely used in data mining technology and statistics [27]. It uses known multivariate data to construct the prediction criteria and predict variables based on other variables. In this study, the decision tree was constructed to judge the land attributes of the pixels. The judgment basis for the sub-nodes is the shape of the NDVI time series curve and the corresponding value of the NDVI time series curve. We evaluate the accuracy of sugarcane mapping by constructing a confusion matrix.

IV. RESULTS

A. CONSTRUCTION OF THE SUGARCANE NDVI TIME-SERIES CURVE

Important steps in the SR-SVD method are the singular vector decomposition of the spectra of the training samples and reconstruction of the spectra of the pixel to be classified. We obtained the first four singular vectors of the NDVI time series corresponding to the sugarcane training sample points in the two study areas by the SVD method, and the results are shown in Figures 5 and 6. Since we selected different numbers of images in the two study areas, there were differences in the NDVI time series of sugarcane training samples in the two study areas, and also there were significant differences in the shape of the singular vectors in the two study areas. P represents the percentage of the original time series information contained in that singular vector. Selecting different numbers of singular vectors for the reconstruction of the NDVI time series can result in large differences between the reconstructed curves. If all the singular vectors are selected for reconstruction, the reconstructed curve will be identical in shape to the original curve. If the curves before and after reconstruction are identical, similarity cannot be used for crop information extraction. So, we need to choose the right number of singular vectors for the reconstruction of the time series.

In this study, we found experimentally that if using the first two singular vectors to reconstruct the NDVI time series of sugarcane, the similarity between the NDVI time series of sugarcane before and after reconstruction was high, while the similarity of other crops was low. We used spectral angle mapping, Euclidean distance, and the coefficient of determination R to determine the similarity between the curves.
The results are shown in Figures 7 and 8. Therefore, we chose the first two singular vectors (SV1, SV2) for the reconstruction of NDVI time series in the two study areas.

### B. Sugarcane Planting Area Extraction Analysis

In this study, the spectral angle was used to evaluate the similarity between the original NDVI time series and the reconstructed NDVI time series. We obtained the maximum spectral angle of the sugarcane training samples by calculating the original sugarcane NDVI time series and the reconstructed sugarcane NDVI time series in the sugarcane training samples.

Using only the SR-SVD-SAM method to extract sugarcane acreage would result in many non-sugarcane image elements being misclassified as sugarcane. Therefore, we improved on this basis. We set the range of fluctuation of NDVI thresholds for sugarcane at different fertility periods by counting the threshold distribution of sugarcane sample points at different fertility periods. Different time series points (TSP) were selected according to the actual situation (number of species grown in the crop, number of images) in the two study areas. As it has shown in Tables 4 and 5, we counted the maximum and minimum NDVI of the sugarcane training samples at different time series points. The NDVI values of sugarcane pixels at different time series points had to be between the maximum and minimum values of the sugarcane training samples. The sugarcane pixel discriminant decision tree was constructed by the similarity of the curves and the threshold of the curves.

In order to be able to visualise the advantages of the improved SR-SVD method more clearly, three sets of controlled experiments were carried out. The first experiment was sugarcane acreage extraction by judging the similarity between the original and reconstructed NDVI time series. In the process only the SR-SVD method was used.

![Figure 6](image-url)

**TABLE 4.** Xinhe-NDVI distribution of sugarcane sample points in important growth periods.

| Image Date   | Growing-Period | Min   | Max   | Mean     |
|--------------|----------------|-------|-------|----------|
| 2019/3/12    | Seeding        | 0.10924 | 0.38123 | 0.18106 |
| 2019/5/1     | Tilling        | 0.2344  | 0.40813 | 0.3014  |
| 2019/7/20    | Rapid growth   | 0.3569  | 0.6001  | 0.5437  |
| 2019/8/9     | Rapid growth   | 0.4569  | 0.6924  | 0.62114 |
| 2019/8/14    | Rapid growth   | 0.59417 | 0.7534  | 0.6655  |
| 2019/8/24    | Rapid growth   | 0.6332  | 0.7927  | 0.70515 |
| 2019/9/13    | Rapid growth   | 0.645   | 0.8535  | 0.7211  |
| 2019/9/23    | Rapid growth   | 0.6284  | 0.8405  | 0.753   |
| 2019/9/28    | Rapid growth   | 0.571   | 0.839   | 0.749   |
| 2019/11/22   | Mature         | 0.548   | 0.7755  | 0.7294  |
| 2019/12/2    | Mature         | 0.536   | 0.8178  | 0.72279 |
| 2019/12/12   | Mature         | 0.5417  | 0.80877 | 0.6578  |

**TABLE 5.** Longzhou-NDVI distribution of sugarcane sample points in important growth periods.

| Image Date   | Growing-Period | Min   | Max   | Mean     |
|--------------|----------------|-------|-------|----------|
| 2019/3/12    | Seeding        | 0.185  | 0.4354 | 0.2888  |
| 2019/4/19    | Seeding        | 0.2511 | 0.5868 | 0.3914  |
| 2019/5/19    | Tilling        | 0.3515 | 0.6837 | 0.5101  |
| 2019/6/20    | Tilling        | 0.5012 | 0.8419 | 0.7682  |
| 2019/8/9     | Tilling        | 0.6131 | 0.8860 | 0.7615  |
| 2019/8/24    | Tilling        | 0.63145| 0.8714 | 0.7544  |
| 2019/9/6     | Tilling        | 0.6324 | 0.8720 | 0.7611  |
| 2019/9/13    | Tilling        | 0.6511 | 0.8760 | 0.7622  |
| 2019/9/21    | Tilling        | 0.6611 | 0.8940 | 0.7711  |
| 2019/9/23    | Tilling        | 0.6782 | 0.8911 | 0.7744  |
| 2019/9/28    | Tilling        | 0.6421 | 0.9011 | 0.7753  |
| 2019/10/1    | Mature         | 0.6421 | 0.9019 | 0.7738  |
| 2019/10/6    | Mature         | 0.6566 | 0.8711 | 0.7625  |
| 2019/11/5    | Mature         | 0.6413 | 0.8612 | 0.7512  |
| 2019/11/10   | Mature         | 0.6211 | 0.8511 | 0.7532  |
| 2019/11/22   | Mature         | 0.6121 | 0.8568 | 0.6832  |
| 2019/12/7    | Mature         | 0.5811 | 0.866  | 0.6842  |
| 2019/12/10   | Mature         | 0.5724 | 0.8769 | 0.6990  |
| 2020/3/9     | Seeding        | 0.1243 | 0.6012 | 0.3045  |
The second experiment was by counting the distribution pattern of NDVI at the sugarcane training samples. And then a decision tree was used to extract sugarcane image elements that matched the threshold range. The third experiment combines the two methods to extract sugarcane acreage. It was by discriminating the similarity of NDVI time series curves of image elements before and after reconstruction and the range of NDVI values corresponding to the curves.

In Figure 9 we can clearly see that the use of a threshold decision tree (DT) approach has many shortcomings, as the thresholding alone can lead to many misclassifications. Using the SR-SVD-SAM method, although it can improve the accuracy of sugarcane area extraction, the “same spectrum with different objects” phenomenon can cause some non-sugarcane pixels to be misclassified. By combining the two methods (SR-SVD-SAM-DT), not only the area of sugarcane can be extracted accurately, but the problem of misclassification caused by “same spectrum with different objects” can also be solved.

The results of the extraction of sugarcane planting information were combined with ground verification points to create a confusion matrix to evaluate the accuracy of the extraction results. Many scholars have demonstrated the drawbacks of using Kappa coefficients to represent the accuracy of classification [28]. In this study, the F-score was used instead of the Kappa coefficient to indicate the accuracy of the extraction results.

C. COMPARISON WITH OTHER INDICATORS FOR EVALUATING CURVE SIMILARITY

In order to further determine that using the spectral angle to evaluate the similarity of the original and reconstructed NDVI time series curves is the best, we also used two other indicators to evaluate the similarity of the original and reconstructed NDVI time series. By calculating the Euclidean distance and coefficient of determination between the original NDVI time series and the reconstructed NDVI time series. The smaller the value of the Euclidean distance between the curves before and after reconstruction, the more similar the curves are to each other. Therefore, we choose the maximum Euclidean distance between the NDVI time series before and after the reconstruction of the sugarcane training sample points as the threshold value for classification. If the Euclidean distance between the original NDVI time series and the reconstructed NDVI time series of the pixel is smaller than the maximum Euclidean distance of the sugarcane training sample points, the pixel is then determined to be sugarcane. R mainly evaluates the fit of the NDVI curves before and after reconstruction, with a larger R indicating a more similar curve. We set the R of the NDVI time series before and after the pixel reconstruction to be greater than the smallest R of the training sample points of sugarcane, and if it is greater than that, then the pixel is determined to be sugarcane. Determine the fluctuation range of sugarcane NDVI value at different time series points by constructing a decision tree to extract sugarcane. The results are shown in Figure 10.

It can be seen from Figure 10 and Table 7 that among the three indicators for evaluating the similarity of curves, using the spectral angle to extract sugarcane has the best effect and can accurately extract large pieces of pure sugarcane pixels. Using Euclidean distance to extract sugarcane acreage has the worst effect, and the overall accuracy is around 80%; there are many misclassifications and omissions.

In general, using three different indicators to evaluate the similarity of the original and reconstructed sugarcane NDVI time series curves, the use of spectral angles for sugarcane planting areas extraction has the best effect, and the overall accuracy is above 96%.

V. DISCUSSION

A. OBTAIN AN ACCURATE SUGARCANE REFERENCE CURVE

Spectral and phase characteristics are the key elements of crop remote sensing recognition. Each crop has its own unique canopy spectral characteristics. In 1978, Lee-Lovick et al. [29] found that the spectrum of sugarcane canopy could transmit important information. These findings provide theoretical support for the use of spectral similarity to extract sugarcane information in this study. For crop identification based on spectral similarity, the reference spectral curve needs to obtain time series points as much as possible [30], [31], [32]. Due to the cloudy and rainy spring and summer, the lack of adequate image data in Southwest China is a serious problem [33]. The lack of time series points will affect the differences between crop reference curves and cause the lack of crop growth information. Sugarcane is a crop that grows for many years. The growth characteristics of sugarcane are significantly different from other crops. This study explored the differences between the NDVI time series curves of sugarcane and other main crops and found that the NDVI time series can be used as the basis for sugarcane extraction. SR-SVD obtains the accurate NDVI time series reference curve of sugarcane by capturing the shape characteristics of the sugarcane NDVI time series. The accurate crop reference curve contains crop growth status information and phenological information. Using the reconstructed crop reference curve, based on the spectral similarity method, can significantly improve the accuracy of sugarcane mapping. Compared with the previous supervised classification methods for extracting sugarcane information, this method can significantly improve the mapping accuracy. Muliana [12] segmented medium resolution remote sensing images according to crop texture features and shape features. Although the sugarcane planting distribution map in Kenya area was obtained, there were many misclassification phenomena in the extracted result map. In this study, the difference between sugarcane and non-sugarcane can be accurately obtained by using high spatial resolution spectral time series.
When using the SR-SVD method and the principle of spectral similarity to extract sugarcane planting information, a problem that cannot be ignored is “same spectrum with different objects.” As can be seen from Figure 11, determining whether an image was a sugarcane one simply by judging the similarity of the shape of the sugarcane NDVI time series before and after reconstruction can lead to the phenomenon of “same spectrum with different objects.” In other words, there will be errors. In this study, we solve this problem by constructing a decision tree model. Decision trees can handle measurement data at different scales and can flexibly reflect the relationship between features and classes [34]. Li et al. used THE SR-SVD method to extract winter wheat with an
accuracy of more than 97% because of the characteristics of single crops and large plot area in cultivated areas in northern China [17]. In southern China, the cultivated land was broken and there were many crops. The overall accuracy of extracting sugarcane planting information by SR-SVD method was 86%, which would lead to the wrong classification of “same spectrum with different objects.”

NDVI values of sugarcane at different times can be obtained by combining NDVI time series images and sugarcane sampling points. We calculated the maximum and minimum values of sugarcane sample points at different time series points. The determination of the threshold range of the curve can not only solve the problem of same spectrum with different objects, but also solve some uncertainties in sugarcane mapping. Uncertainties mainly include two: the difference in the NDVI time series curve between the newly planted sugarcane and the ratoon sugarcane, and the other is the difference in the NDVI time series curve of different sugarcane varieties. Due to the previous season’s growth, the initial NDVI value of regenerated sugarcane is relatively high because of the high growth rate in the seedling stage. On the contrary, the leaves of the newly planted sugarcane were thin during the seedling stage, the corresponding NDVI value was low (As shown in Figure 12 (a)). The sugarcane training samples contain information about different sugarcane varieties, newly planted sugarcane, and ratoon sugarcane. In this study, we obtained the fluctuation threshold range of the NDVI time series curve of sugarcane by counting the value ranges of the NDVI time series curve of the sugarcane training samples at different growth periods. (As shown in Figure 12 (b)). Determining the fluctuation range of the sugarcane NDVI
time series curve can avoid misclassification errors caused by the above problems.

FIGURE 12. (Xinhe) NDVI Time Series Curves of Ratoon Sugarcane and Newly Planted Sugarcane (a). Sugarcane NDVI curve fluctuation range in sugarcane training samples(b). The green dashed box represents the highlights of the NDVI difference between ratoon and newly planted sugarcane at the seedling stage.

C. SELECT THE JUDGMENT INDEX OF CURVE SIMILARITY

At present, great progress has been made in extracting crop area using the principle of spectral similarity [17]. The representation and similarity judgment of time series is the basis of time series research. The selection of curve similarity indicators is a critical step in crop information extraction based on the principle of spectral similarity. Different judgment indicators have inconsistent judgments on the similarity between curves [35]. We formulate corresponding crop classification rules by judging the original spectrum and reference spectrum of crops. The selection of curve similarity index is a key step in extracting crop information based on the principle of spectral similarity. Different indicators describe the correlation of features between curves based on different emphases. Sun et al. [18] used Euclidean distance to evaluate the similarity between the original EVI1 and reference EVI2 curves. Although the winter wheat information is successfully extracted, the overall accuracy of the extraction is not high. The main reason for the above problems is that the Euclidean distance can only describe the relative position between the curves, and cannot effectively reflect the shape difference of the curves. In order to investigate the influence of different indicators of similarity between curves on the accuracy of sugarcane acreage extraction, we set up three sets of control experiments. This study selects the spectral angle SAM, Euclidean distance (ED), and coefficient of determination (R) to determine the similarity between the original and reconstructed NDVI time series of sugarcane. The results show that using the spectral angle to evaluate the similarity of sugarcane time series curves, combined with the sugarcane NDVI fluctuation threshold, the extraction effect of sugarcane is the best. Followed by R and ED. SAM mainly focuses on evaluating the similarity of spectral shapes, which can effectively describe the changing trend of sugarcane NDVI. R mainly evaluates the fitting effect of the original NDVI time series after reconstruction, and ED focuses on describing the deviation between these two curves. Using ED and R to determine the curve similarity has three primary defects [36]: (1) the curve shape similarity cannot be identified (2) it cannot reflect the similarity of the trend dynamic change amplitude (3) the calculation based on the point distance cannot reflect the difference in the frequency of change. The results of this research also prove the above conclusions.

VI. CONCLUSION

In this study, we used the SR-SVD method to perform the singular vector decomposition of the NDVI time series of sugarcane training sample points and used the singular vectors for the reconstruction of sugarcane NDVI time series. Through experiments, it was found that when using the first two singular vectors for NDVI time series reconstruction, the similarity of NDVI time series before and after reconstruction was high for sugarcane but low for other crops. Since there are more crop species in southern China than in northern China, using spectral similarity for sugarcane acreage extraction alone will result in “same spectrum with different objects” classification errors. We improved on the SR-SVD method. By analyzing the variation pattern of NDVI time series of sugarcane training samples and setting the threshold range of sugarcane NDVI time series, we constructed a decision tree for sugarcane image discrimination. The results show that the improved SR-SVD can not only extract the planting information of winter wheat, but also extract the planting area information of sugarcane accurately and efficiently. The problem of classification error caused by “same spectrum with different objects” is effectively solved. At the same time, the method also provides a new idea for extracting planting information of southern crops.

REFERENCES

[1] S. Hu, L. Shi, K. Huang, Y. Zha, X. Hu, H. Ye, and Q. Yang, “Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation,” Field Crops Res., vol. 232, pp. 49–61, Dec. 2019.
[2] I. Shield, “Sugar and starch crop supply chains,” in Biomass Supply Chains for Bioenergy and Biofuelling, J. B. Holm-Nielsen and E. A. Ehimen, Eds. Sawston, U.K.: Woodhead, 2016, ch. 11, pp. 249–269.
[3] A. L. Alarcón, L. M. Palacios, C. Osorio, P. C. Narváez, F. J. Heredia, A. Orjuela, and D. Hernanz, “Chemical characteristics and colorimetric properties of non-centrifugal cane sugar (‘panela’) obtained via different processing technologies,” Food Chem., vol. 340, Mar. 2021, Art. no. 128183.
[4] P. Heerden, G. Eggelston, and R. Donaldson, Sugarcane: Physiology, Biochemistry, and Functional Biology. Hoboken, NJ, USA: Wiley-Blackwell, 2013.
[5] R. Sindhu, E. Gnansounou, P. Binod, and A. Pandey, “Bioconversion of sugarcane crop residue for value added products—An overview,” Renew. Energy, vol. 98, pp. 203–215, Dec. 2016.
[6] B. Mulanga, A. Bégué, P. Clouvel, and P. Todoroff, “Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing,” Remote Sens., vol. 7, no. 11, pp. 14428–14444, Oct. 2015.
[7] F. Lin, S. Guo, C. Tan, X. Zhou, and D. Zhang, “Identification of rice sheath blight through spectral responses using hyperspectral images,” Sensors, vol. 20, no. 21, p. 6243, Nov. 2020.
[8] A. K. Verma, P. K. Garg, and K. S. H. Prasad, “Sugarcane crop identification from LISS IV data using ISODATA, MLC, and indices based decision tree approach,” Arabian J. Geosci., vol. 10, no. 1, pp. 1–17, Jan. 2017.
[9] M. P. Grisham, R. M. Johnson, and P. V. Zimba, “Detecting sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes,” J. Virol. Methods, vol. 167, no. 2, pp. 140–145, Aug. 2010.
[10] H. Lin, J. Chen, Z. Pei, S. Zhang, and X. Hu, “Monitoring sugarcane growth using ENVISAT ASAR data,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 8, pp. 2572–2580, Aug. 2009.
[11] C. Palaniswami, R. Viswanathan, A. Bhaskaran, P. Rakkhiyanpan, and P. Gopalasundaram, “Mapping sugarcane yellow leaf disease affected area using remote sensing technique,” J. Sugarcane Res., vol. 4, no. 1, pp. 1–7, 2014.

[12] J. M. Peña-Barragán, M. K. Ngugi, R. E. Plant, and J. Six, “Object-based crop identification using multiple vegetation indices, textural features and crop phenology,” Remote Sens. Environ., vol. 115, no. 6, pp. 1301–1316, Jun. 2011.

[13] Y. Chen, L. Feng, J. Mo, W. Mo, M. Ding, and Z. Liu, “Identification of sugarcane with NDVI time series based on HJ-1 CCD and MODIS fusion,” J. Indian Soc. Remote Sens., vol. 48, no. 2, pp. 249–262, Feb. 2020.

[14] F. Henry, D. E. Herwindiati, S. Mulyono, and J. Hendryli, “Sugarcane land classification with satellite imagery using logistic regression model,” in Proc. Int. Conf. Remote Sens. Environ. Technol., vol. 4, 2017, Art. no. 012024.

[15] M. A. Vieira, A. R. Formaggio, C. D. Rennó, C. Atzberger, D. A. Aguiar, F. Henry, D. E. Herwindiati, S. Mulyono, and J. Hendryli, “Sugarcane land classification with satellite imagery using logistic regression model,” in Proc. Int. Conf. Remote Sens. Environ. Technol., vol. 4, 2017, Art. no. 012024.

[16] Q. Dong, X. Chen, J. Chen, C. Zhang, L. Liu, X. Cao, Y. Zang, X. Zhu, and X. Cui, “Mapping winter wheat in north China using sentinel 2A/B data: A method based on phenology-time weighted dynamic time warping,” Remote Sens., vol. 12, no. 8, p. 1274, Apr. 2020.

[17] S. Li, F. Li, M. Gao, Z. Li, P. Leng, S. Duan, and J. Ren, “A new method for winter wheat mapping based on spectral reconstruction technology,” Remote Sens., vol. 13, no. 9, p. 1810, May 2021.

[18] H. Sun, A. Xu, H. Lin, L. Zhang, and Y. Mei, “Winter wheat mapping using temporal signatures of MODIS vegetation index data,” Int. J. Remote Sens., vol. 33, no. 16, pp. 5026–5042, 2012.

[19] G. Messina, L. M. Peña, M. Vizzari, and G. Modica, “A comparison of UAV and satellites multispectral imagery in monitoring onion crop. An application in the ‘Cipolla rossa di Tropea’ (Italy),” Remote Sens., vol. 12, no. 20, p. 3424, Oct. 2020.

[20] G. Messina, S. Praticó, G. Badagliacca, S. di Fazio, M. Monti, and G. Modica, “Monitoring onion crop ‘cipolla rossa di tropea Calabria IGP’ growth and yield response to varying nitrogen fertilizer application rates using UAV imagery,” Drones, vol. 5, no. 1, p. 1108, Apr. 2020.

[21] G. Duveiller, R. López-Lozano, and B. Barthu, “Enhanced processing of 1-km spatial resolution FAPAR time series for sugarcane yield forecasting and monitoring,” Remote Sens., vol. 5, no. 3, pp. 1091–1116, Mar. 2013.

[22] C. Ren, Y.-J. Liang, X.-J. Lu, and H.-B. Yan, “Research on the soil moisture sliding estimation method using the LS-SVM based on multi-satellite fusion,” Int. J. Remote Sens., vol. 40, pp. 1–16, Jun. 2018.

[23] L. Kliiber, D. Martyenko, and T. Holzer-Porp, “Thermal infrared remote sensing of mineral dust over land and ocean: A spectral SVD based retrieval approach for IASI,” Atmos. Meas. Techn., vol. 4, no. 5, pp. 757–773, May 2011.

[24] M. N. M. Sap and M. Kohram, “Spectral angle based kernels for the classification of hyperspectral images using support vector machines,” in Proc. 2nd Asia Int. Conf. Modelling Simulation, 2006, pp. 559–563.

[25] C. Yonezawa, “Maximum likelihood classification combined with spectral angle mapper algorithm for high resolution satellite imagery,” Int. J. Remote Sens., vol. 28, no. 16, pp. 3729–3737, Aug. 2007.

[26] R. Nigam, R. Tripathy, S. Dutta, N. Bhagia, R. Nagori, K. Chandrasekar, R. Kot, B. K. Bhattacharya, and S. Ustin, “Crop type discrimination and health assessment using hyperspectral imaging,” Current Sci., vol. 116, no. 7, p. 1108, Apr. 2019.

[27] L. Breiman et al., “Classification and regression trees,” Encyclopedia Ecol., vol. 57, no. 1, pp. 582–588, 2015.

[28] G. M. Foody, “Explaining the unsuitability of the Kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification,” Remote Sens. Environ., vol. 239, Mar. 2020, Art. no. 111630.

[29] H. S. Nzioki, J. E. Jamoza, C. O. Olweny, and J. K. Rono, “Characterization of physiologic races of sugarcane smut (Ustilago scitaminea) in Kenya,” Afr. J. Microbiol. Res., vol. 4, no. 16, pp. 1694–1697, 2010.

[30] L. Liu, X. Xiao, J. Qin, Y. Jing, W. Xu, Y. Hu, and Z. Qiao, “Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth engine,” Remote Sens. Environ., vol. 239, Mar. 2020, Art. no. 111624.

[31] L. Sun, F. Gao, D. Xie, M. Anderson, R. Chen, Y. Yang, Y. Yang, and Z. Chen, “Reconstructing daily 30 m NDVI over complex agricultural landscapes using a crop reference curve approach,” Remote Sens. Environ., vol. 253, Feb. 2021, Art. no. 112156.

[32] J. Wang, X. Xiao, L. Liu, X. Wu, Y. Qin, J. L. Steiner, and J. Dong, “Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images,” Remote Sens. Environ., vol. 247, Sep. 2020, Art. no. 111951.

[33] F. Ullah, R. Moore, and A. Fung, Microwave Remote Sensing Active and Passive: Microwave Remote Sensing Fundamentals and Radiometry. Boston, MA, USA: Artech House, 1981.

[34] M. A. Friedli and C. E. Brodley, “Decision tree classification of land cover from remotely sensed data,” Remote Sens. Environ., vol. 61, pp. 399–409, Sep. 1997.

[35] M. Silva, O. de Carvalho, R. Guimarães, R. Gomes, and C. Silva, “Wheat planted area detection from the MODIS NDVI time series classification using the nearest neighbor method calculated by the Euclidean distance and cosine similarity measures,” Geocarto Int., vol. 35, pp. 1–15, Feb. 2019.

[36] X.-L. Dong, C.-K. Gu, and Z.-O. Wang, “Research on shape-based time series similarity measure,” in Proc. Int. Conf. Mach. Learn. Cybern., 2006, pp. 1253–1258.