

A new automated approach for remote sensing recognition and yield estimation of cultivated alfalfa crop based on Sentinel-2 NDVI time-series data: A case study of Hexi Corridor, China

jie liu

Lanzhou University

Xuying Bao

Yongchang county vocational secondary school

Qisheng Feng

Lanzhou University

Shuai Fu

Lanzhou University

Chunli Miao

Lanzhou University

Yunhao Li

Lanzhou University

Tiangang Liang (✉ tgliang@lzu.edu.cn)

Lanzhou University

Research Article

Keywords: Remote sensing, image recognition, time series, cultivated alfalfa, yield estimation, information extraction

Posted Date: July 26th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3191538/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **A new automated approach for remote sensing recognition and**
2 **yield estimation of cultivated alfalfa crop based on Sentinel-2**
3 **NDVI time-series data: A case study of Hexi Corridor, China**
4

5 Jie Liu^a, Xuying Bao^b, Qisheng Feng^a, Shuai Fu^a, Chunli Miao^a, Yunhao Li^a, Tiangang Liang^{a*}
6

7 ^a State Key Laboratory of Grassland Agro-ecosystems; Key Laboratory of Grassland Livestock Industry Innovation,
8 Ministry of Agriculture and Rural Affairs; Engineering Research Center of Grassland Industry, Ministry of
9 Education; College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, 730000, China;

10 ^b Yongchang county vocational secondary school, Jinchang, Gansu, 737200, China;

11 * Corresponding author. E-mail address: tgliang@lzu.edu.cn (T. Liang).

Abstract

Alfalfa (*Medicago sativa*) is an important forage source for grassland agricultural development, so it would be worthwhile to explore accurate and fast methods of alfalfa remote sensing identification and yield estimation. However, the traditional methods of identifying large areas of crops and yield estimation have some problems, such as the limited spatial resolution of remote sensing data and heavy reliance training data. In this study, based on Sentinel-2 high-resolution images and the Google Earth Engine (GEE) platform to establish a cloud-free normalized difference vegetation index (NDVI) time-series dataset, we proposed an effective method for alfalfa feature extraction and yield estimation method. The results show that (1) The producer's accuracy, user's accuracy, overall accuracy, and Kappa coefficient of alfalfa identification using a trough recognition algorithm were 98.51%, 91.67%, 94.26%, and 0.88, respectively. The total area of cultivated alfalfa identified in the study area in 2020 was estimated at 46,793.21 hm², which was mainly distributed in areas in north of the Qilian Mountains; (2) NDVI had a highly significant correlation with alfalfa hay yield, and the power function regression model was the greatest, with an R² greater than 0.65; (3) The annual unit hay yield of four alfalfa cuttings was estimated at 17,497.55-32,962.10 kg/hm², with a total hay yield of 48.38×10⁷ kg and an average hay yield of 4,464.95 kg/hm². The method proposed has important application potential for automatic and rapid remote sensing identification and yield evaluation of large-scale cultivated alfalfa.

Keywords: Remote sensing; image recognition; time series; cultivated alfalfa; yield estimation; information extraction.

1. Introduction

Alfalfa (*Medicago sativa*) is a forage species of global importance, owing to its high productivity, nutritional value, palatability and digestibility for livestock (1, 2). It is widely cultivated in many regions and plays a vital role in ensuring ecological and food security. In China, alfalfa has been extensively planted in recent years as part of agricultural structural adjustment and animal husbandry development, which has greatly promoted the development of the alfalfa industry (3, 4). However, there is a lack of efficient and accurate methods for large-scale monitoring of alfalfa grasslands, which hampers the availability of reliable data for alfalfa production and management.

Remote sensing technology enables large-scale, real-time monitoring of crop growth and yield prediction (5-8). This technology avoids contacting or damaging to the research objects and provides timely, fast, dynamic and macroscopic information (9). Therefore, it can help predict the growth of crops in advance so that government departments take effective measures for rational deployment (10). Yield estimation through remote sensing includes two important steps: (1) crop identification and area extraction and (2) growth monitoring and yield forecasting (11). The choice of remote sensing data

sources (GF-1, Landsat 8 and Sentinel-2) is crucial for crop identification (12, 13). And the Sentinel-2 data sources are suitable for high-temporospatial-resolution analysis of crop planting areas (14-18). Previous studies have used different methods to extract vegetation and crop areas. Kobayashi *et al.* (14) applied a random forest (RF) algorithm to 91 spectral indices computed from Sentinel-2A data to classify crop types in Turkey. Comparing the classification accuracy of different indices for different crops, the result showed that the classification based on the reflectance including Bands 2, 4, 11 and 12, and 8 spectral indices from Sentinel-2A data was useful for identification of each crop and achieved an overall accuracy (OA) of 93.1%. And their findings demonstrated the utility of Sentinel-2A data for generating high-quality crop cover maps. Using multi-temporal Sentinel-1 and Sentinel-2 image data and applying a RF classifier, Chen *et al.* (15) developed a method to extract vegetation cover and corn crop distribution within it. The results showed that it was more accurate and robust than other methods using single-sensor or single-date images. Belgiu and Csillik (16) applied a time-weighted dynamic time warping (TWDTW) method for crop classification using Sentinel-2 data. They compared pixel-based and object-based approaches of TWDTW in three different regions with diverse agrosystems. They found that object-based TWDTW performed better than pixel-based TWDTW. Based on RF classifier, Huang *et al.* (17) proposed a framework for early season mapping of winter wheat using spectral and temporal information of Sentinel-2 images. The framework consisted of three steps: image preprocessing, feature extraction, and classification. Applying the framework to Henan province of China, which is a major winter wheat growing region, it achieved an OA of 91.6%. Zhang *et al.* (18) proposed an automated early-season method to map winter wheat using time-series Sentinel-2 data (AEMMS). Applying it to Shandong province of China, it achieved an OA of 94.5% and a kappa coefficient of 0.89 for the winter wheat map. They also compared their method with other existing methods based on different data sources and time periods, and demonstrated that AEMMS had advantages in terms of accuracy, timeliness, simplicity, and applicability. According to the above research findings, one of the key steps in crop remote-sensing monitoring is to select “critical time nodes” (e.g., the maximum NDVI value before alfalfa mowing) or “the best phase” that capture the most essential and characteristic information about the temporal variations of crop. Remote sensing images at these nodes should have high spatial resolution and can provide spectral data that can differentiate the target crop from the others (19). However, existing researches for extracting “critical time nodes” based on remote sensing features (e.g., difference method (20), threshold setting method (21), etc.), only focus on the differences in remote sensing VIs of local crops, are limited by uncertain identification results and poor universality.

Spectral vegetation indices (VIs) can reflect potential crop yield. And the normalized difference vegetation index (NDVI) derived from red band and near-infrared band (NIR) information is the most commonly used VI (22, 23). Casa *et al.* (24) integrated phenological data of corn crop with NDVI estimations from National Oceanic and Atmospheric

Administration-Advanced Very High-Resolution Radiometer (NOAA-AVHRR) system to estimate county yield of corn in Córdoba, Argentina. The result showed that the NDVI value during the reproductive stage expressed a high association with corn yield for any sowing date. Lopresti *et al.* (25) estimated wheat yield using an empirical model during flowering period based on MODIS-NDVI and field observation data at department level in northern Buenos Aires province, Argentina, and found a good correlation between the estimation data of the model and official yields (a coefficient of determination (R^2) of 0.75). Arab *et al.* (26) proposed a novel method to predict table grape yield using artificial neural network (ANN) method and time-series VIs derived from Landsat 8 data. The results showed that among all VIs, NDVI had the yield the highest accuracy, and the correlation coefficients (R) for 2017, 2018, and 2019 were 0.94, 0.95, and 0.92, respectively. Marshall and Thenkabail *et al.* (27) compared the performance of multispectral broadband (MSBB) and hyperspectral narrowband (HNB) VIs in estimating crop biomass (rice, alfalfa, cotton and maize). They found that HNB VIs derived from EO-1 Hyperion sensor performed better than MSBB VIs derived from IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS sensors, and identified the best HNB VIs for different crop types based on their spectral characteristics. And the R^2 of the estimation model for rice, alfalfa, cotton and maize was 0.91, 0.81, 0.97 and 0.94, respectively. Kayad *et al.* (28) conducted a study on a center pivot irrigated alfalfa field in Saudi Arabia to assess the spatial variability of alfalfa yield using Landsat-8 imagery and a hay yield monitor data. They found that both satellite and ground-based data showed a clear spatial variation in alfalfa yield across the field for four harvests, and that the NDVI derived from Landsat-8 was highly correlated with alfalfa yield. Most of the above studies on yield are limited to small scale areas, and the premise of large-scale alfalfa yield estimation is to extract the spatial distribution of large-scale alfalfa, the traditional large-scale monitoring of alfalfa distribution and yield estimation has several shortcomings such as limited spatial resolution of remote sensing data, difficulty in processing a large amount of remote sensing data, and heavy dependence on training data. These limitations hinder their application in large-scale remote sensing identification and yield estimation of cultivated alfalfa crop.

In addition, due to factors such as time, space, and image quality, remote sensing data suitable for certain study areas are often limited in both quality and quantity, while large-scale crop surveys relying on limited remote sensing data over a short time can be problematic (29). Firstly, downloading and preprocessing high-resolution remote sensing data for large-area crop identification is time-consuming and inefficient, resulting in poor timeliness and accuracy of research results and difficulty of applications on a large scale. Google Earth Engine (GEE) is a cloud-based geospatial data analysis platform that provides access to massive global-scale Earth science data, especially satellite data, and offers technical support for a wide range of users for data processing (30, 31). Coding on GEE allows fast data preprocessing and access to large-area satellite data, showing potential for large-scale crop identification and yield prediction inversion.

Considering above background, we used Sentinel-2 high-resolution remote sensing images with high temporal and spatial resolutions as the data source and created a cloud-free NDVI dataset on the GEE platform in this study, and developed an NDVI-based algorithm for alfalfa identification that can automatically search for “critical time nodes” by integrating alfalfa phenology, growth characteristics, ecological environment, sampling site information and other field survey data to define the spatial distribution of alfalfa. Then, a remote sensing yield estimation model was constructed to monitor cultivated alfalfa yield in the study area rapidly and accurately.

2. Data and methods

2.1 Study area

The Hexi Corridor in northern China is a major source of high-quality alfalfa for domestic markets. Within this region, three cities - Zhangye, Jinchang and Wuwei - produce most of the commercial alfalfa production and form a key high-quality forage base in China (32). These cities span 83,008 km² in the central-eastern segment of the Hexi Corridor (97°21'21"–104°13'35"E, 36°44'25"–39°53'16"N), bordering Lanzhou to the east, Jiayuguan and Jiuquan to the west, Qinghai Province to the south and Inner Mongolia and Ningxia to the north (Fig. 1a). The landscape is characterized by mountains and plains, interspersed with Gobi Desert and oases. The climate is continental temperate arid with large seasonal and diurnal temperature variations, low and uneven precipitation, high evaporation and solar radiation and sufficient sunlight, with the average annual temperature from 4.1 to 9.2 °C, the annual sunshine rate from 51% to 66%, the average annual rainfall from 112.3 to 354.0 mm, and the frost-free period from 96 to 212 d. These conditions are favorable for alfalfa growth. The main land cover types are grassland and forest, followed by farmland, shrubs and non-vegetated areas (Fig. 1b). Agriculture dominates the local economy with farming, animal husbandry and breeding as common activities. Alfalfa planting has expanded rapidly in recent years with government support.

The study area's alfalfa planting companies have mechanized most production processes, including sowing, harvesting, picking, turning and bundling. They have also adopted high-yielding integrated techniques such as seed pellet coating, rhizobia inoculation, mulching precision hole seeding and integrated control of diseases, insects and weeds, which have enhanced alfalfa yield and quality (33). The alfalfa hay has a crude protein content of 16%–18%, making it a palatable and nutritious fodder for cattle and sheep and near the top among domestic high-quality alfalfa (34).

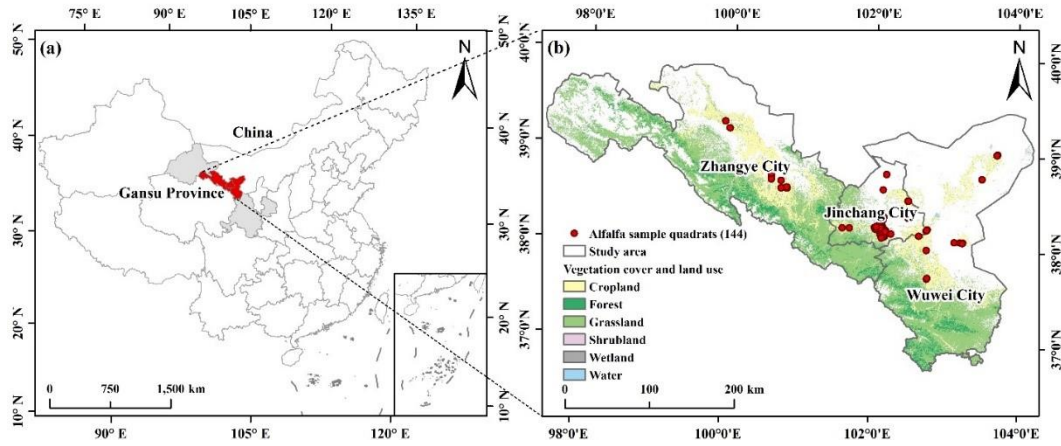


Figure 1. (a) Geographical location of the study area, (b) distribution of sample plots and land cover type.

2.2 Ground-measured data

The field investigation mainly focused on growing season of the alfalfa grassland with distinct features (May to September) in 2020. We surveyed 81 alfalfa grassland sample plots (sizes of 100 m × 100 m or larger) in the first survey (from May 27 to June 6), 51 sample plots in the second survey (from July 17 to July 26), and 12 sample plots in the third survey (from September 20 to September 24), and three subplots (0.5 m × 0.5 m) were randomly set up in each plot. A total of 144 plots and 432 subplots were collected in the field. In every plot, the information of alfalfa variety, annual cutting frequency, current cutting situation, planting years, irrigation and fertilization, business entities were collected. Meanwhile, photos of each plot were taken vertically at altitudes of 20 m and 50 m, respectively, and a panoramic photo was taken at an altitude of 100 m using a Dajiang unmanned aerial vehicle (DJI UAV Phantom 4 Pro). In every subplot, latitude, longitude, alfalfa height, leaf area and chlorophyll content were recorded. The above-ground part of alfalfa with a 5 cm stubble was harvested and measured as fresh weight on site and its dry weight after drying it in a constant-temperature oven at 65 °C for 72 hours until reaching constant weight. The dry weight was used as the basic input for alfalfa remote sensing yield estimation modeling. To verify the accuracy of the recognition results, the geographical locations of 100 non-alfalfa plots (corn, chili peppers, onions, oats, etc.) were recorded. Then using ArcGIS software to establish the attributes and spatial database of the field survey plots (Fig. 1).

2.3 Cultivated land data

The Finer Resolution Observation and Monitoring–Global Land Cover 10 (FROM-GLC10) product with a resolution of 10 m was used for the cultivated land data in this study (<http://data.ess.tsinghua.edu.cn/>). It was developed by Professor Gong Peng using the RF method with the world's first set of multi-season sample databases from Sentinel-2A global images and the GEE cloud computing platform. And it classified the world's land coverage into 10 categories: crop, forest, grass,

shrub, wetland, water, tundra, impervious, bare land and snow/ice (35).

2.4 Sentinel-2 data

The Sentinel series satellites provide free remote sensing data with the highest spatial and temporal resolution. They also offer multispectral remote sensing data at a global scale for monitoring the earth's environmental conditions and responding promptly to sudden emergencies such as floods, forest fires and landslides (36).

The European Copernicus Mission operates two polar-orbiting multispectral high resolution optical satellites, Sentinel-2A and Sentinel-2B, launched on 22 June 2015 and 7 March 2017, respectively. These satellites are phased at 180° to each other with a complementary revisit time of 5 days and carry a multispectral instrument that covers 13 bands: four bands (three visible and one NIR) at 10 m resolution; six bands (three red edge, one NIR and two shortwave infrared) at 20 m resolution; and three bands (coastal aerosol, water vapor and cirrus) at 60 m resolution (37). The Level-1C product of Sentinel-2 data used in this study has been corrected using a digital elevation model and provides the top of the atmosphere (TOA) reflectance values for each pixel in 13 unsigned integer spectral bands and three quality assessment bands (QA10, QA20 and QA60). The QA60 band contains cloud mask information that allows us to remove opaque and cirrus clouds from the image.

2.5 GEE platform data download and processing

GEE is a new-generation platform that enables online visualization, computation, analysis and processing of massive global-scale Earth science data, especially satellite data (38). The Sentinel-2 image data used in this study were stored on the GEE cloud platform and processed entirely through online coding.

Phase and range selection: To obtain alfalfa plant data for the entire phenological period (seedling, budding, flowering and maturing stages), the image phase was set to cover April 1 to November 30, 2020, ensuring the completeness and suitability of the data.

NDVI calculation: A program was compiled to calculate NDVI from Sentinel-2 images using the vegetation index formula on the GEE platform, which extracted data from the NIR (B8) and red (B4) bands. When outputting the result, it assigned each image a specific number and track number for identification. The formula is as follows:

$$NDVI = (B8 - B4) / (B8 + B4) \quad (1)$$

Cloud removal: Applying the C Function Mask S2 Clouds (CFMASK) algorithm to Sentinel-2 QA60 band data to detected opaque and cirrus clouds, a threshold was set for Bit10 of the QA60 band to eliminate cloud shadows: where pixel with Bit10=0 was masked as clouds. Similarly, where pixel with Bit11=0 was masked as cirrus clouds. Then the cloud-free images were selected using the update mask tool and the labelled cloud, cloud shadow, and cirrus range were assigned with

no data.

Interpolation optimization: A three-sliding-window algorithm was applied to interpolate pixels with no data. Given the image pixel scale (10 m) and revisit period (5d) of Sentinel-2 data, three consecutive images were respectively designated N1, N2, and N3 (for example, N1 is the NDVI value on April 10, N2 is the NDVI value on April 15, N3 is the NDVI value on April 20) and the N2 was assigned values based on different conditions: if $N2 > 0$, original value; if $N2 < 0$ and $N1 < 0$ and $N3 < 0$, original value; if $N2 < 0$ and $N1 < 0$ and $N3 > 0$, value of $N3$; if $N2 < 0$ and $N1 > 0$ and $N3 < 0$, value of $N1$; if $N2 < 0$ and $N1 > 0$ and $N3 > 0$, average of $(N1+N3)$. This process was repeated for subsequent every three images (N2, N3, N4...) until covering the selected time range. A cloud-free NDVI time-series image dataset for the study area using this method was created. The data processing flow chart in Fig. 2.

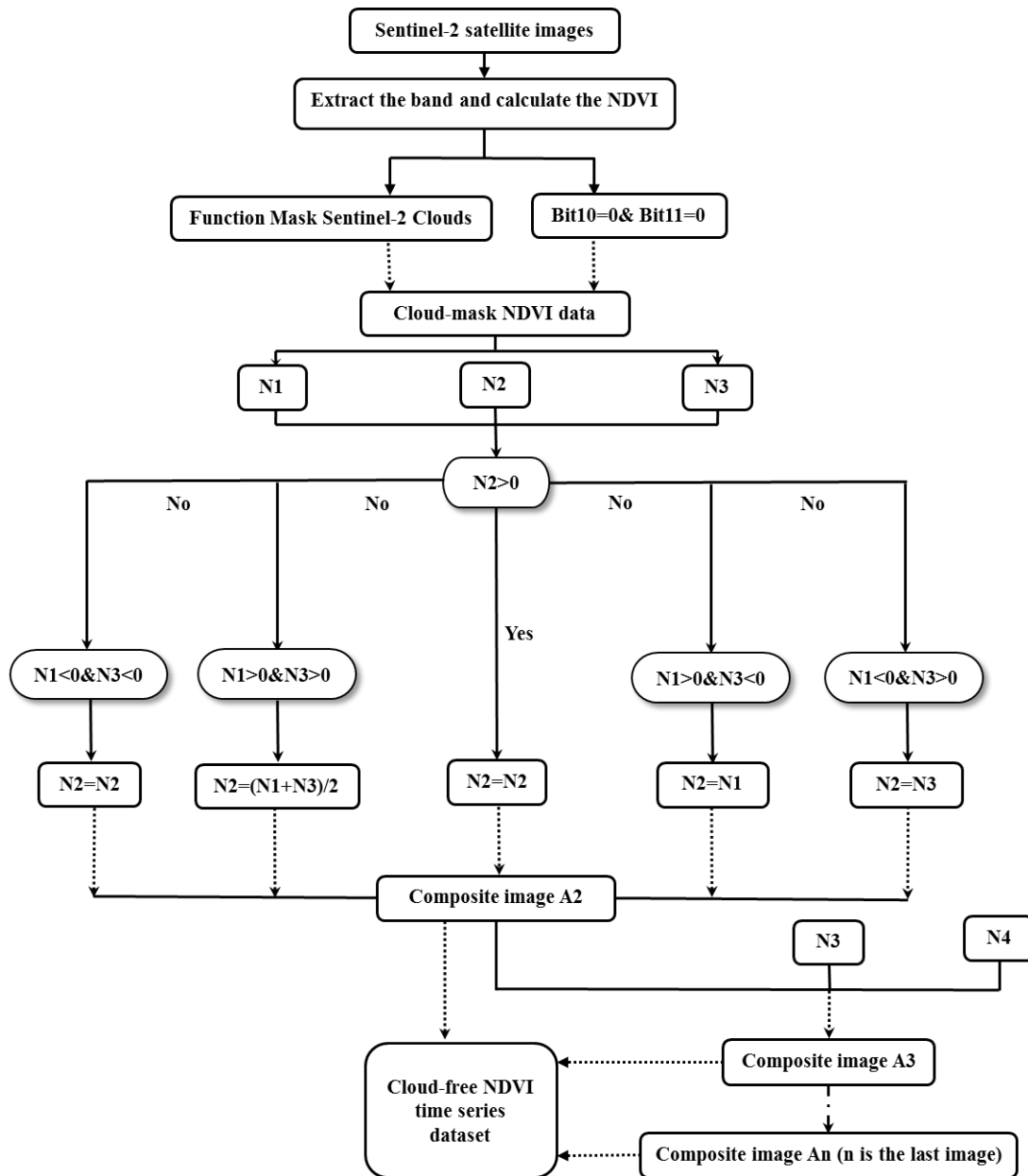


Figure 2. Data processing flow chart

2.6 Alfalfa feature extraction

The study area has adopted the one-crop-per-year cropping system with a variety of crops, such as spring and winter wheat, spring corn, cotton, onions, potatoes, rapeseed and Chinese kale. The phenological periods of alfalfa and some of these crops overlaps substantially (Fig. 3), which complicates discrimination of alfalfa from other crops using vegetation index thresholds at a specific imaging time. By analysis and comparison, we observed that alfalfa is significantly different from other crops in multiple harvests during the growth period in a year, resulting in dramatic characteristic changes after cutting. This feature was utilized to enhance the remote sensing identification of alfalfa.

The NDVI time-series curves capture the variation characteristics of crop NDVI throughout the growing season (39). The NDVI of alfalfa grassland exhibits multiple NDVI fluctuations with a “rise-peak-fall” pattern in a year, resulting in several peaks and troughs in the NDVI time-series curve. A peak indicates the higher NDVI value and vigorous growth period of alfalfa, while a trough represents the state of alfalfa after harvest. The NDVI drops sharply after each harvest, creating abrupt changes in the curve. The number and timing of the troughs reflect the number and timing of the harvests. In the study area, there were 2-4 troughs. Making full use of this feature, an algorithm was developed for alfalfa recognition, classification and yield estimation based on the NDVI time-series dataset, by comparing and analyzing the spatiotemporal variation characteristics of the NDVI data.

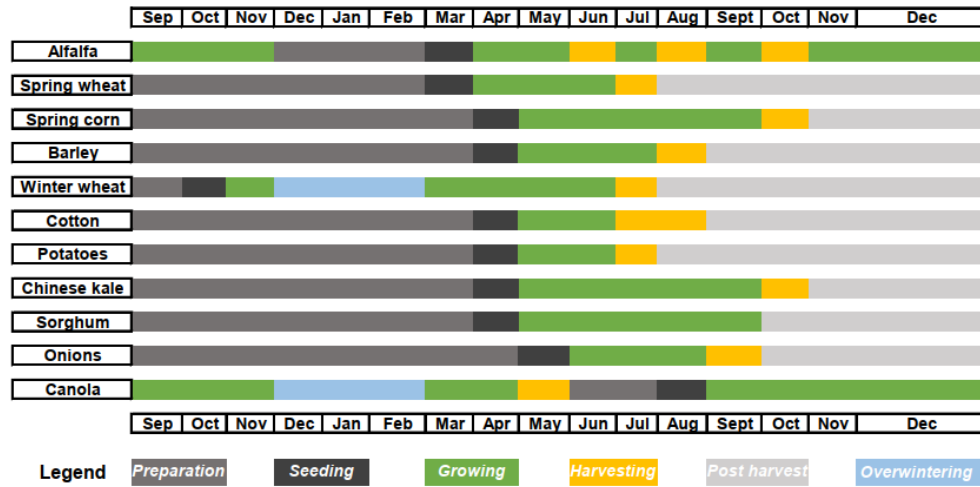


Figure 3. Calendar and type map of main crops in the study area

2.7 Area extraction and yield estimation

The periodic harvest of alfalfa provides a unique opportunity to discriminate it from other crops using remote sensing time-series data. Therefore, the trough recognition algorithm was devised based on detecting the troughs caused by cutting in NDVI time-series data.

The trough recognition algorithm was implemented using MATLAB's Findpeaks function (Fig. 4), which returned local maximum peak value (Peak) of the input signal and allowed found the peak value of interest by setting thresholds for parameters, i.e., the value of a certain element was greater than that of either of two adjacent elements. Among them, the peak prominence parameter measures the minimum vertical distance from a peak to its two sides, and the minimum peak prominence (MPP) in Findpeaks enables fast searches for relatively important peaks. In our automatic identification algorithm, MPP corresponds to the peak-trough difference of NDVI time series images. Thus, when MPP reaches a certain value, the algorithm automatically identifies a peak and its associated trough based on peak-trough association analysis. On the basis of previous studies (40), the MPP was set to 0.4 for alfalfa extraction in our study area. Findpeaks then automatically classified signals with peak-trough differences greater than 0.4 as alfalfa, extracted their spatial distribution and recorded their dates, which provided the basic information for extraction and yield estimation of the alfalfa.

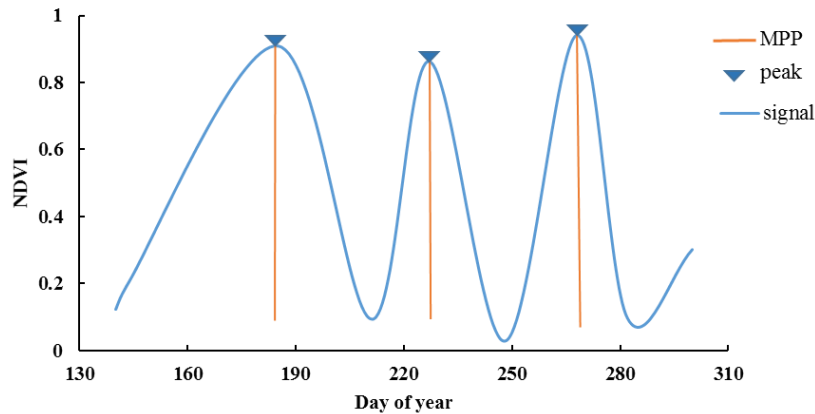


Figure 4. NDVI time series curve and concept map of Findpeaks function

While using the trough identification method to identify alfalfa, we also selected the pixel-based satellite data that were collected at the time closest to the cutting time point for yield estimation and examined the number of troughs, their dates and the peak NDVI values of each pixel from cloud-free Sentinel-2 NDVI time-series data during the growing season. This established some characteristic datasets pixel by pixel: number of times (NOT), day of year (DOY) and peak value of NDVI (PVON), which used to determine the following important parameters:

1) The alfalfa distribution range was identified by restricting NOT and MPP values. Based on the local alfalfa production pattern, the NOT value was 2-4, i.e., the value set of NOT was {2, 3, 4}.

2) The dates of peak occurrence corresponding to the NOT values (1-4) were determined, i.e., DOY1, DOY2, DOY3 and DOY4.

3) The peak NDVI values corresponding to cutting times (NOT values 1-4), i.e., PVON1, PVON2, PVON3 and PVON4, were obtained for alfalfa yield estimation.

Finally, masking was performed on the cultivated land in the study area by combining the terrain features and farmland spatial distribution data, and the alfalfa spatial distribution was established through post processing procedures such as mode filtering and debris elimination using ENVI version 5.3 software. The spatial distribution features of the first, second, third, and fourth cuttings of alfalfa in the study area as well as the total annual alfalfa yield in 2020 were analyzed. (Fig. 5).

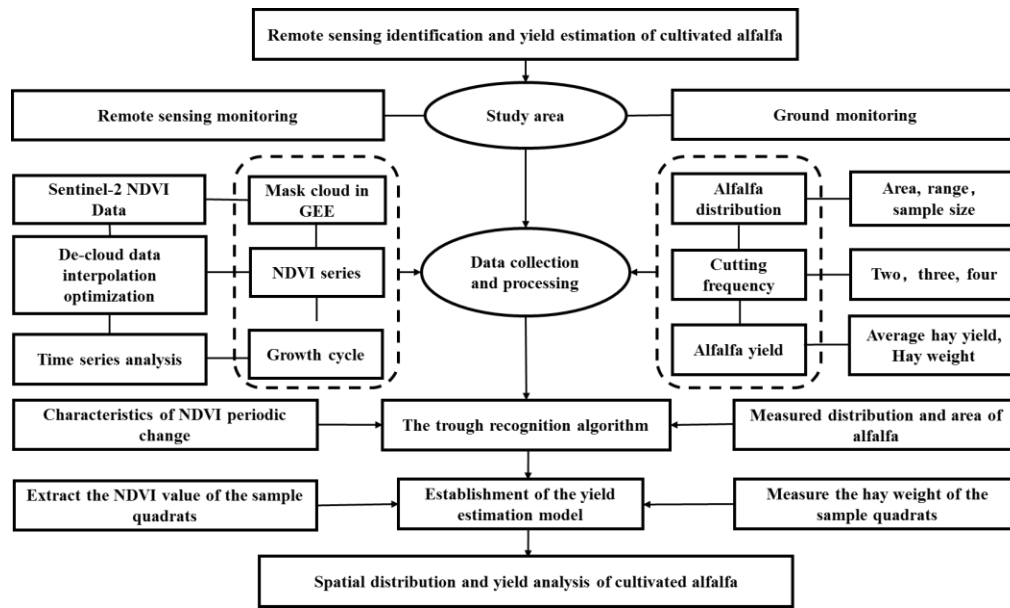


Figure 5. Methods and technical flowchart

2.8 Estimated yield model

Based on the coordinates of the sample plots collected during our field survey, the NDVI of each plot was extracted from GEE platform, and the extraction dates were as close as possible to the field measurement to ensure the validity and reliability of our yield estimation model. 17 plots affected by cloud or cloud shadow were excluded out of the total 144 plots, and the rest 127 plots were used to construct the yield estimation model (64 from the first cutting, 51 from the second cutting and 12 from the third cutting). After analyzing the correlation of NDVI and hay yield using SPSS 26.0 software, the yield estimation models were constructed for three cuttings and annual production using linear ($y = Ax + B$), logarithmic ($y = A\ln(x) + B$), exponential ($y = Ae^{xD}$), power ($y = Ax^F$) and quadratic polynomial ($y = Ax^2 + Bx + C$) functions, in which y represents the alfalfa hay yield (kg/hm^2); x represents the NDVI value; and A , B , C , D , and F are parameters.

2.8 Evaluation of extraction accuracy and yield estimation model

Specific evaluation indicators include OA, producer's accuracy (PA), user's precision (UA), and kappa coefficient, reflecting the accuracy of image classification from different perspectives (41, 42). And the R^2 , root mean square error (RMSE) and mean relative error (MRE) are used as indicators for estimating the performance of the model. Their formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - E_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - E_i)^2}{n}} \quad (3)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - E_i}{Y_i} \right| \quad (4)$$

where i represents the data of the i^{th} sampling point; Y_i is the measured crop yield of the i^{th} sampling plot (kg/hm^2); E_i is the crop yield estimate of the i^{th} sampling plot calculated by the model (kg/hm^2); and $\bar{Y}$ is the measured mean yield (kg/hm^2).

3 Results

3.1 Area extraction results and analysis

3.1.1 NDVI time-series data optimization results

The basic data of this study was the Sentinel-2 NDVI time-series data, which was screened to meet the criteria during the alfalfa growth cycle. It is essential for accurately extracting alfalfa spatial distribution information (Fig. 6). However, Sentinel-2 images are often affected by cirrus clouds, resulting in NDVI value deviations in some areas and hindering the identification process. Therefore, the NDVI time-series data should be interpolated and optimized in the study area. Fig. 6 shows an example of unoptimized and optimized NDVI time-series curves using Sentinel-2 data from field survey sampling plots in Wuwei city in 2020. The unoptimized NDVI time-series curve had a clear discontinuity in the time range of 240-260 d, making it impossible to connect the data for classification. Moreover, in the time range of 160-240 d, NDVI value frequently dropped to 0. It was caused by clouds and cloud shadows affecting this area at this time, resulting in abnormal data that impaired the performance of the algorithm based on the troughs caused by cuttings during the alfalfa growth cycle and increased misclassification rates.

After optimization, the faults and outliers on the NDVI time-series curve was eliminated and the curve was smooth with obvious peaks and troughs. Therefore, the improved Sentinel data provided more reasonable NDVI time-series data that met the recognition needs of the trough-based algorithm and were an important prerequisite for implementing the algorithm.

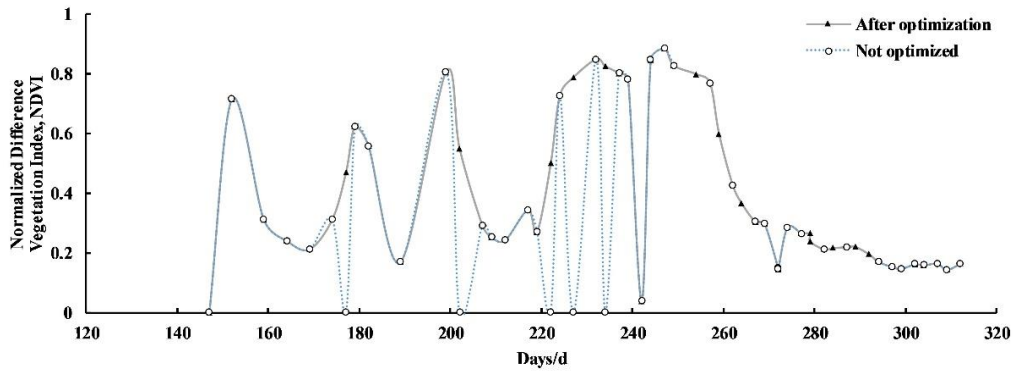


Figure 6. NDVI timing data optimization results

3.1.2 Alfalfa remote sensing extraction and accuracy verification

The total alfalfa area identified in the study area in 2020 was 46,793.21 hm^2 . Table 1 shows alfalfa areas and proportions of local cultivated land in different regions. At the city level, Zhangye had the largest alfalfa acreage (19,268.77 hm^2), followed by Jinchang, Wuwei had the smallest (11,638.18 hm^2). At the county level, Yongchang County of Jinchang city had the largest alfalfa area of 14,568.70 hm^2 , accounting for 1.76% of the county's total cultivated area and making it the most concentrated alfalfa planting area in the study area. The alfalfa areas in all counties and districts of Zhangye were rather uniform, and all were above 2,000 hm^2 , with Gaotai County having the largest area at 4,374.48 hm^2 , which accounted for 0.60% of the county's total cultivated area. The alfalfa planting area of Wuwei was mainly in Minqin County and Gulang County, at 5,375.59 hm^2 and 4,368.00 hm^2 , respectively. These two areas accounted for 83.72% of the city's total alfalfa area, while the alfalfa acreage of Tianzhu Tibetan Autonomous County was only 10.60 hm^2 , which was spread sporadically and accounted for 0.01% of the total cultivated area, representing the smallest alfalfa acreage in the study area.

Table 1. Areas of alfalfa in different regions

City	District and county	Alfalfa area (hm^2)	Percentage of cropland (%)
Zhangye City	Yugur Autonomous County of Sunan	3,325.24	0.95
	Gaotai County	4,374.48	0.60
	Linze County	2,986.35	0.42
	Ganzhou District	3,650.71	0.22
	Minle County	2,697.99	0.25
	Shandan County	2,234.00	0.3
Jinchang City	JinChuan District	1,317.56	0.39
	Yongchang County	14,568.70	1.76
Wuwei City	Minqin County	5,375.59	0.39

Liangzhou District	1,883.99	0.11
Gulang County	4,368.00	0.61
Tianzhu Tibetan Autonomous County	10.60	0.01

Fig. 7 shows the alfalfa spatial distribution in the study area in 2020 based on remote sensing information extraction. Obviously, the Alfalfa was concentrated in north of the Qilian Mountains. According to the field investigation, (a), (b), (c) and (d) are taken as the typical areas of alfalfa recognition results in the study area, and it can be seen that the alfalfa plots identified in the four regions have obvious shape and high fullness. Among them, the alfalfa fields in the typical areas (a), (c) and (d) are rectangular in shape and distributed in a large area. The shape of the alfalfa field in (b) is round, and the overall distribution is relatively scattered, which is related to the circular nozzle facilities of the local management unit. Based on the data collected from 144 sampling points of fields and 100 sampling points of other crops (corn, chili peppers, onions, oats, etc.), we verified the accuracy of the alfalfa distribution and found that the PA, UA, OA, and kappa coefficient of validation set were 98.51%, 91.67%, 94.26%, and 0.88, respectively.

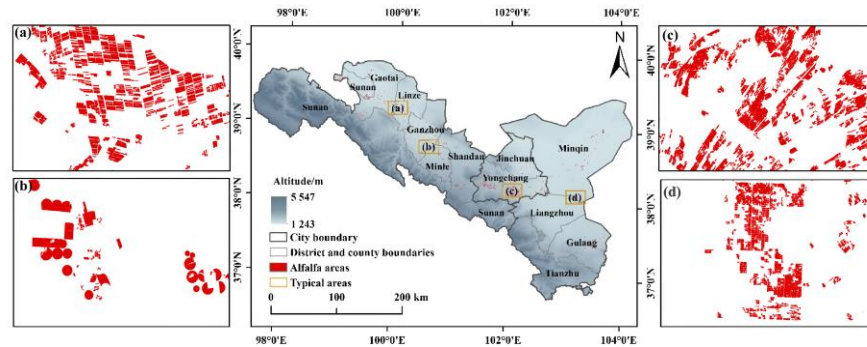


Figure 7. Spatial distribution of alfalfa in study area in 2020; (a), (b), (c) and (d) are the typical areas of alfalfa recognition results

3.2 Yield estimation results and analysis

3.2.1 Establishment of the yield estimation model

The results of alfalfa yield estimation model are shown in the Fig. 8. The NDVI value was correlated positively with the actual yield, with R^2 values above 0.55 for all models. Among the models, the R^2 values of yield estimation models for the first cutting were all above 0.5, the power function model having the highest R^2 at 0.652, making it the best yield estimation model ($y = 8484.2x^{0.7092}$). For the second cutting, the R^2 values were generally higher than those for the first cutting, in the range of 0.64-0.71, and the power function model was again the best, with an R^2 value of 0.707 ($y = 9700x^{0.8408}$). For the third cutting, the R^2 values for the third cutting were all approximately 0.8, but due to a small sample

size, the fitted model was not representative, so when estimating the yield, the annual yield estimation model should be referenced. And the R^2 values of the annual yield estimation models were all approximately 0.7, and the power function model, given as $y = 9043.9x^{0.789}$, was again the best, with an R^2 value of 0.756. Some parts in the study area cut alfalfa four times one year, so the alfalfa yield of the fourth cutting should be estimated by referencing the annual yield estimation model.

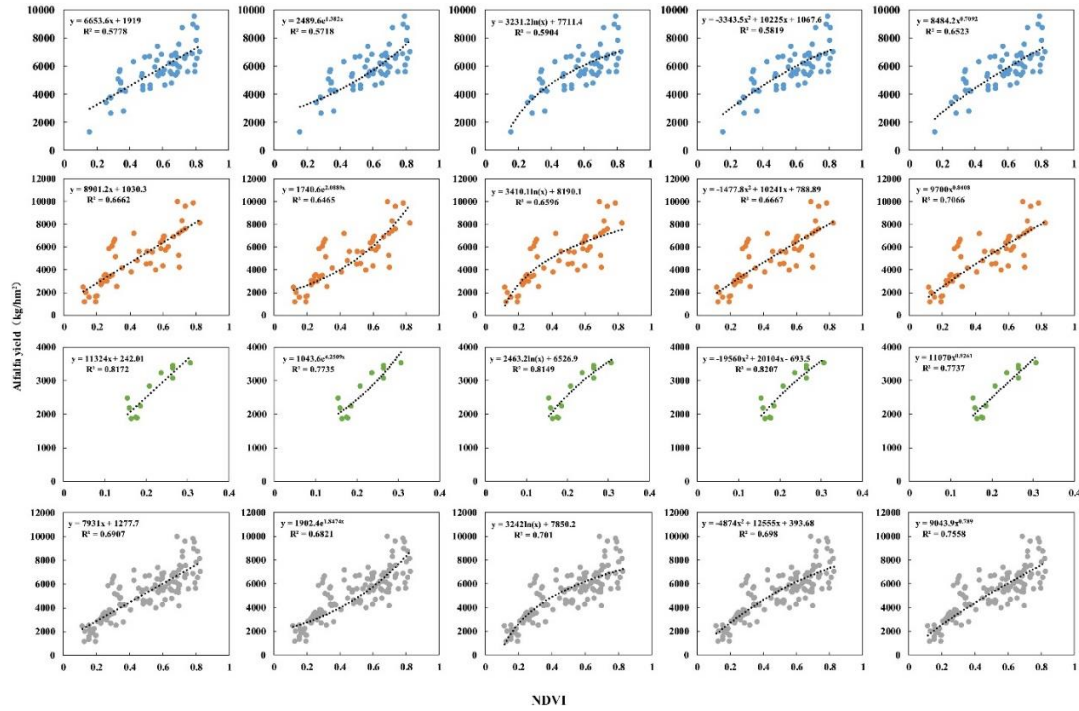


Figure 8. Model diagram of alfalfa NDVI and yield estimation.

3.2.2 Analysis of the simulation error of the yield estimation model

Comparing the estimated yield based on the best yield estimation model with the measured yield, the models for estimating the first, second, and third cuttings and the annual yield showed certain errors, with varying degrees of underestimation or overestimation. The estimated yield of the first cutting had errors within 2000 kg/hm² and mostly matched the measured yield, indicating a good model performance (Fig. 9a). The estimated yield of second cutting had larger discrepancies at some sites, with errors exceeding 2000 kg/hm² (Fig. 9b). For the third cutting, the differences between the estimated and actual yields were small, all having an error below 1000 kg/hm² (Fig. 9c), but the model was not representative due to the small sample size. Therefore, it is more appropriate to use the annual yield estimation model to estimate the yield of the third cutting.

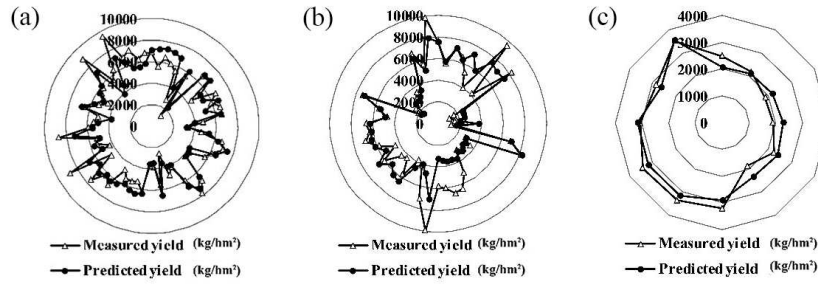


Figure 9. Model forecast yield and statistical yield differences

Regression analysis was performed using the estimated and measured alfalfa yields of the first, second and third cuttings and the annual yield to validate that the estimated yield was within the acceptable error range. The regression equation and the corresponding R^2 , RMSE and MRE values are shown in Fig. 10 a-d. The models had low errors, as indicated by the R^2 values above 0.6 and the MRE values below 20%. The estimation data were uniformly dispersed, suggesting that the alfalfa yield estimation algorithm based on peak NDVI value performs well and shows a relatively consistent error level under different cutting times and thus can be reliably applied to yield estimations. The results indicate that the second cutting model had higher errors, with a RMSE of 1301.21 kg/hm² and a MRE of 18.82%, and the third cutting model had the best performance, with a RMSE of 274.11 kg/hm², a MRE of 10.17% and a R^2 value of 0.81. The models of first cutting and the annual yield also performed well, with R^2 values of 0.62 and 0.70, respectively, RMSE values of 966.79 kg/hm² and 1104.54 kg/hm², and MRE values of 14.70% and 17.64%.

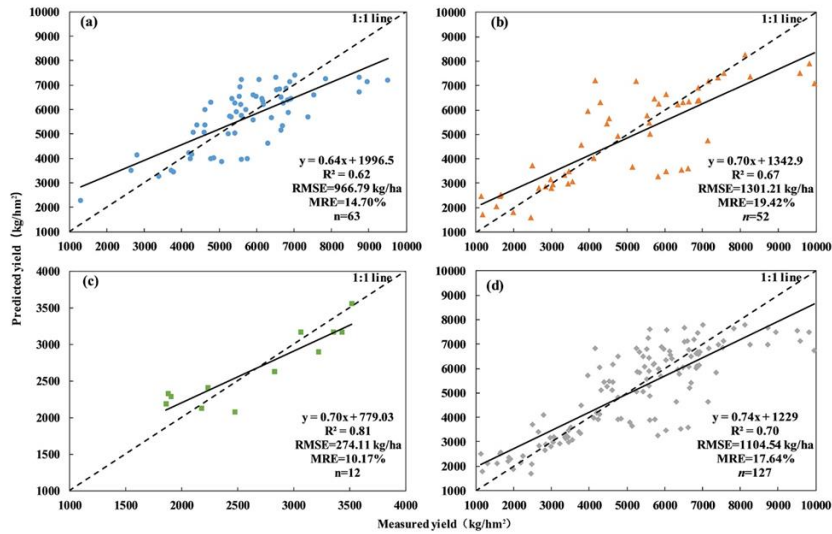


Figure 10. Accuracy of yield estimation model, (a) The first cutting, (b) The second cutting, (c) The third cutting, (d) The annual yield.

3.2.3 Results and analysis of alfalfa yield estimation

Based on the PVON database obtained through the trough recognition algorithm, The alfalfa hay yields of the four

cuttings and the annual total hay yield were calculated using Raster calculator tool in ArcGIS software. Fig. 11 shows that the alfalfa yield of the first cutting was the highest, followed by that of the second cutting, the third cutting, and the fourth cutting. And the yield distribution trends of the four cuttings were similar with single-peak curves, but with varying curve widths and heights, reflecting significant variations in the yields of the four cuttings.

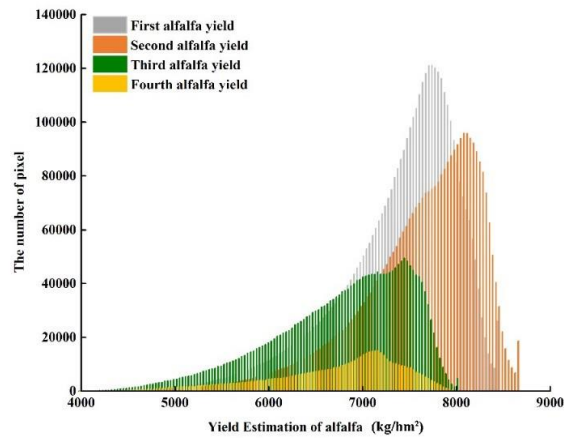


Figure 11. Statistical histogram of alfalfa yield distribution

According to the results of yield estimation, the per unit yield of alfalfa was divided into three grades: low yield ($< 5000\text{kg/hm}^2$), middle yield ($5000\text{--}7000\text{kg/hm}^2$) and high yield ($> 7000\text{kg/hm}^2$). The alfalfa yield distribution in the study area exhibited obvious characteristics (Fig. 12). The area of high-yield and middle-yield alfalfa in the first and second cutting accounted for a large proportion, while the area of low-yield grade increased in the third and fourth cutting. Overall, the per unit yield of the first two cutting alfalfa was significantly higher than that of the latter two. Through the yield distribution of typical alfalfa areas, it can be seen that the yield grades of most alfalfa parts have changed among the four cutting. Therefore, low-yield alfalfa should improve field management to enhance its yield.

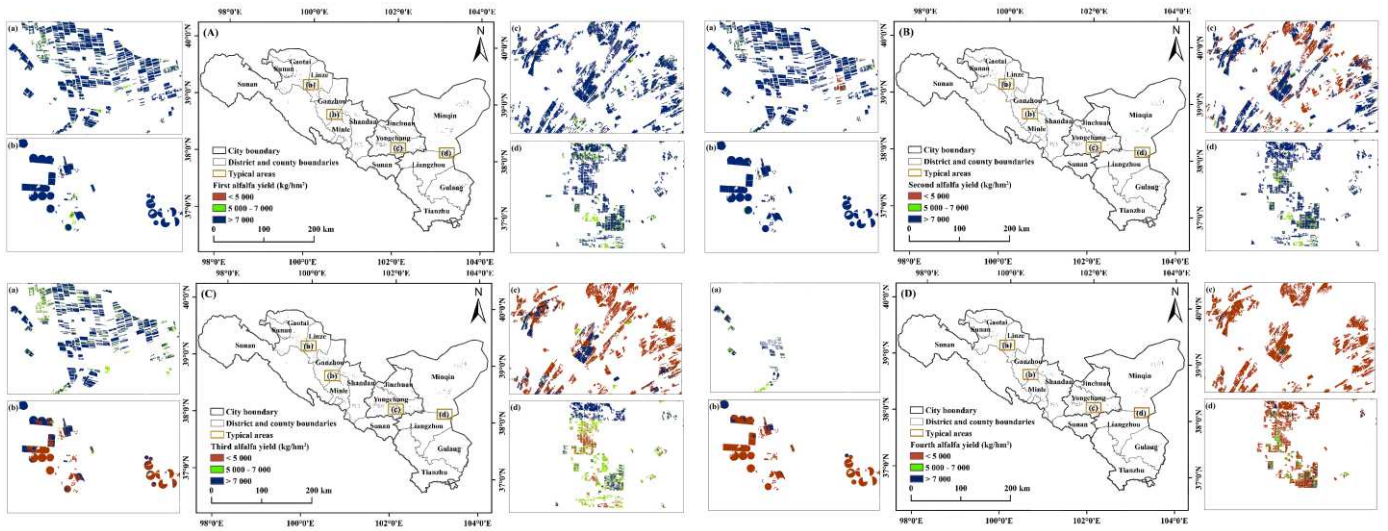


Figure 12. Alfalfa yield distribution maps of the first (A), second (B), third (C) and fourth times (D) in the study area in 2020; (a), (b), (c) and (d) are the typical areas of alfalfa.

The yield differences were mainly attributed to the following factors: (1) Alfalfa harvest time is influenced by many factors, such as planting time, production year, variety, harvesting machinery availability and weather (43), which result in profound variations in the statistical characteristics of alfalfa fields. This affects the remote sensing features of adjacent alfalfa fields in the same phase and consequently the alfalfa hay yield of each cutting. (2) Alfalfa yield depends on local production management measures, cutting time and method, stubble height and other factors. To ensure nutrient accumulation and regrowth after cutting, the last two cuttings have higher stubble heights than the previous ones. For instance, in the last cutting before winter, the stubble height is usually above 8 cm to facilitate overwintering and growth in the next year (44), which led to lower alfalfa yields of the third and fourth cuttings in the study area.

The alfalfa yield in the study area was calculated using Zonal Statistics as Table tool of ArcGIS software (Table 2). In 2020, for the first cutting, the alfalfa hay yield of the study area was 4,229.88–8,482.51 kg/hm², with a total hay yield of 30.68×10^7 kg and an average hay yield of 7,164.68 kg/hm². For the second cutting, the alfalfa hay yield was 4,489.35–8,698.17 kg/hm², with a total hay yield of 21.47×10^7 kg and an average hay yield of 6,080.62 kg/hm². For the third cutting, the alfalfa hay yield of the study area was 4,389.16–8,042.09 kg/hm², with a total hay yield of 8.95×10^7 kg and an average hay yield of 3,702.17 kg/hm². For the fourth cutting, the alfalfa hay yield of the study area was 4,389.16–8,041.90 kg/hm², with a total hay yield of 0.55×10^7 kg and an average hay yield of 912.33 kg/hm². For the entire year, the total alfalfa hay yield of the study area was between 17,497.55 kg/hm² and 32,962.10 kg/hm², with a total hay yield of 48.38×10^7 kg and an average hay yield of 4,464.95 kg/hm².

Table 2. Statistical result of alfalfa yield in the study area

Indicators	The first cutting	The second cutting	The third cutting	The fourth cutting	Total annual yield
Maximum hay yield (kg/hm ²)	8,482.51	8,698.17	8,698.17	8,041.9	32,962.1
Minimum hay yield (kg/hm ²)	4,229.88	4,489.35	4,389.16	4,389.16	17,497.55
Average hay yield (kg/hm ²)	7,164.68	6,080.62	3,702.17	912.33	4,464.95
Hay weight (10 ⁷ kg)	30.68	21.47	8.95	0.55	48.38

4 Discussion

4.1 Alfalfa remote sensing recognition and area extraction

Remote sensing recognition of alfalfa remains a challenge. The complex planting and growth characteristics of cultivated alfalfa combines the features of natural grassland and field crops, such as sowing and harvesting patterns of crops, phenological phases of natural grassland. Therefore, it is necessary to comprehensively consider various factors, such as farming method, sowing time, harvest time, growth period, phenological characteristics, and cutting time to develop simple and easy-to-implement algorithms to efficiently identify and classify alfalfa on a large scale.

This study presented a novel automated method for large-scale alfalfa identification using optimized Sentinel-2 data and based on the growth characteristics of cultivated alfalfa. The method, called the “trough recognition algorithm”, achieved high-accuracy alfalfa identification and obtained a 10-m spatial resolution alfalfa map. The method has the following advantages: (1). The 10-m alfalfa map was derived from the 10-m Sentinel-2 images, which may facilitate the mapping of smallholder alfalfa fields that are not achievable from MODIS or Landsat images due to their coarseness and lack of high spatial resolution. Moreover, we interpolated the Sentinel-2 data to obtain a complete time series data set, filling in the abnormal values affected by clouds and thereby improving the identification accuracy; (2). Using GEE platform to process large-scale remote sensing data, it overcome the problems of cloud cover and insufficient computing resources in traditional remote sensing methods, and enhanced the processing efficiency and quality of remote sensing data; (3). The method was a fully automated method that did not rely on manually labeled sample data (training data), reflecting the low-cost, convenient and fast mapping process with strong generality and scalability; (4). It can not only accurately extract alfalfa spatial distribution information from complex and diverse crop types, but also eliminate the variability between different alfalfa fields in the same area and the interference from other crops, achieving high accuracy. Thus, this method has great application potential for the monitoring of large-scale cultivated alfalfa grassland.

Moreover, few studies have made results on large area alfalfa recognition, but there are studies on other crops recognition. For example, Zhang *et al.* (18) achieved fine-scale remote sensing monitoring and mapping of winter wheat in Shandong Province, China, using AEMMS method. They showed that the AEMMS method could generate high-accuracy (the OA of 95.5%, Kappa coefficient of 0.91) and high-resolution (10 m) winter wheat mapping products with strong stability and adaptability. It is similar to ours, as both rely on NDVI time series and crop growth characteristics for identification without training data, demonstrating its potential for large-scale crop recognition.

Nevertheless, there are some limitations of trough recognition algorithm. The alfalfa VIs depends on the environmental conditions in each year and area. Climatic factors such as precipitation or temperature can affect the alfalfa growth and cutting frequency. Thus, it needs to be adjusted for different areas based on local field conditions and to set reasonable MPP values and peak-trough numbers for accurate extraction of cultivated alfalfa spatial distribution. The result of the analysis on the MPP values of 144 sample plots in the study area (Table 3) showed that the values varied depending on the geographical location and cutting time of alfalfa fields. The reasonableness of the MPP value directly influences the accuracy of alfalfa identification and thus the estimation of cultivated alfalfa area and yield.

Table 3. Minimum Peak Prominence value

Indicators	The first cutting	The second cutting	The third cutting
Maximum value	0.8731	0.7735	0.6714
Minimum value	0.4052	0.3609	0.3252
Average value	0.5823	0.4752	0.4142
Variance	0.0128	0.0240	0.0162
Standard deviation	0.1129	0.1551	0.1271

In addition, due to the high cloud cover for wet areas or mountainous areas, there was no available cloud-free pixel in the adjacent time series, resulting in incomplete NDVI time series that may hamper subsequent analysis. Therefore, in the future research, the method of fusing multiple optical remote sensing data (e.g., Sentinel-2, Landsat and other high-resolution images), and Synthetic-Aperture Radar (SAR) data (e.g., Sentinel-1), which was less affected by cloud cover. Combining the advantages of various remote sensing can enhance the image quality and extract diverse crop features, which can improve crop identification accuracy in wet areas.

The cultivated land boundary used in this study was based on the 10-m resolution land-cover data, which only distinguishes large areas of cultivated land from other features (e.g., residential areas, woodland, grassland, etc.). It does not

capture the detailed boundaries of fields owned by different farmers or companies. Therefore, the alfalfa pixels identified do not allow more fine-grained analysis based on field units in the subsequent integrated analysis, resulting in “empty windows” (some pixels in the same alfalfa field are not recognized as alfalfa due to different growth conditions) and underestimation of alfalfa area. New methods for extracting farmland boundaries are needed to address this issue. Future research on alfalfa remote sensing recognition can also incorporate more factors (such as agricultural operation time nodes, crop growth periods, field vegetable growth and harvest characteristics, meteorological factors) combining machine learning algorithms to enhance the accuracy of alfalfa remote sensing recognition and provide more reliable data support for rational alfalfa management.

4.2 Alfalfa yield estimation and analysis

Satellite VIs can provide real-time data for dynamic macroscopic monitoring of crop growth and yield estimation, which can support the construction of crop yield estimation models and complement the limitations of traditional methods in real-time crop growth monitoring. This study applied the trough recognition algorithm to identify alfalfa over a large area and obtained pre-harvest NDVI values for the entire study area. Based on the NDVI value, the yield model was constructed, and the results showed that the power function yield estimation model was the best, with an accuracy of 76.5% and a R^2 of 0.7 between predicted and measured yield. The method and results are similar to those of previous studies. For example, Kayad *et al.* (28) studied the correlation between alfalfa NDVI obtained from landsat8 data and actual yield, with R^2 of 0.63, indicating that NDVI is a reliable indicator of alfalfa yield.

However, relying solely on vegetation indices for alfalfa yield estimation can be problematic (44). China has a large and diverse arable land area with high levels of mixed cropping and low levels of large-scale planting, which constrain the alfalfa yield estimations based on remote sensing VIs data. Especially in the Loess Plateau of China, where terraced fields with fragmented blocks are common, making it even more difficult to estimate its yield. Other issues include NDVI saturation, remote sensing satellite data accuracy, atmospheric correction effects (45), which affect the accuracy to different extents and need to be considered in future development of remote sensing-based yield estimations. previous studies have suggested that incorporating multiple VIs and environmental factors can improve yield prediction for different crops. Iniyen *et al.* (46) proposed a method to predict wheat and rice yield based on multiple VIs and environmental factors derived from MODIS and Sentinel-2 image data in India, and used support vector machine (SVM) to for model building and prediction. The prediction results reached more than 90%. Therefore, the future studies should Compare different VIs models for alfalfa yield estimation, and incorporate more variables, such as altitude, slope, evapotranspiration, and precipitation into the yield estimation model. This would enhance the model’s generalizability and applicability across regions and provide important

scientific and technological support for the integrated management and planning of alfalfa grassland and other crops.

5 Conclusions

In this study, Zhangye, Jinchang, and Wuwei cities of the Hexi region in Gansu Province were selected as the study area. The GEE platform was used to optimize Sentinel-2 data for the study area to generate a cloud-free NDVI time-series dataset. And the alfalfa recognition algorithm was developed to automatically identify the peak-trough features of the NDVI time-series data to extract alfalfa spatial distribution information. And the macroscopic monitoring of alfalfa yield was achieved in the study area by constructing yield estimation models based on the field data. The conclusions are as following:

1) The interpolation optimization of the NDVI time-series data of cultivated alfalfa crop was significantly effective and resolved the faults and outliers of the original NDVI time-series curve.

2) Using the optimized Sentinel-2 data and the proposed peak-trough recognition algorithm, it achieved high recognition accuracy with PA of 98.51%, UA of 91.67%, OA of 94.26% and kappa coefficient of 0.88. The remote sensing identified alfalfa cultivation area in the region in 2020 was 46,793.21 hm², mainly distributed in areas north of the Qilian Mountains.

3) NDVI had a strong correlation with alfalfa hay yield, and the power function regression models had the best performance with all R² values above 0.65 and were suitable for alfalfa hay yield estimation in the study area.

4) The regression calibration error between the predicted and measured yield using the models was relatively low, with R² above 0.6, stable MREs within 20% and uniform dispersion of predicted data, indicating the alfalfa yield estimation algorithm based on searching for peak NDVI values performed well and had good potential for future applications.

5) The annual unit hay yield of four alfalfa cuttings in the study area was estimated at 17,497.55-32,962.10 kg/hm², with a total hay yield of 48.38×10⁷ kg and an average hay yield of 4,464.95 kg/hm².

Declarations

- Ethics approval and consent to participate: Not applicable.
- Consent for publication: Not applicable.
- Availability of data and materials: The satellite data (i.e., Sentinel-2 data) are available from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). The Finer Resolution Observation and Monitoring–Global Land Cover 10 (FROM-GLC10) product is freely and publicly available from Tsinghua University (<http://data.ess.tsinghua.edu.cn/>).

- Competing interests: The authors declare that they have no competing interests.
- Funding: This study was supported by the earmarked fund for the China Agricultural Research System (CARS-34); Key Consulting and Research projects of the Chinese Academy of Engineering (2022-HZ-09, 2022-XY-139, 2021-HZ-5); Special expenses for basic Scientific Research Operations of Central Colleges and Universities of Lanzhou University (lzujbky-2022-sp13); the Science & Technology project “study and application of photovoltaic power stations on ecological environment and ecological restoration technology in arid and semi-arid Areas (HNKJ21-H76)” from China Huaneng Group Co. LTD; “Innovation Star” project of outstanding graduate students in Gansu Province, China (2022CXZX-087).
- Authors’ contributions: Tiangang Liang, and Qisheng Feng conceived the ideas and designed methodology; Jie Liu, Xuying Bao, Shuai Fu, Chunli Miao, Yunhao Li collected the data; Jie Liu, Xuying Bao led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication. We claim that all authors on the manuscript have seen and approved the submitted version of the manuscript, that all authors have substantially contributed to the work, and that all persons entitled to co-authorship have been included.
- Acknowledgements: Not applicable.

References

1. Zhang C, Wang C, Hu X, Liu Q, He Y. The nutritional value of alfalfa and its application in feedstuffs. *China Feed*. 2005; 1: 15-17.
2. Atumo T, Kauffman R, Gemiyo Talore D, Abera M, Tesfaye T, Tunkala B, et al. Adaptability, forage yield and nutritional quality of alfalfa (*Medicago sativa*) genotypes. *Sustainable Environment*. 2021; 7(1): 0-7.
3. Li X, Wang L. Research progress on *Medicago sativa* silage technology. *Acta Prataculturae Sinica*. 2005; 2: 9-15.
4. Liu Y, Zhang Y. National alfalfa industry development plan (2016-2020). *China Animal Industry*. 2017; 11:32
5. Brown J, Jepson W, Kastens J, Wardllow B, Lomas J, Price K. Multitemporal, Moderate-Spatial-Resolution Remote Sensing of Modern Agricultural Production and Land Modification in the Brazilian Amazon. *Mapping Sciences & Remote Sensing*. 2007; 44(2): 117-148.
6. Destro D, Carpentieri-Pipolo V, Kiihl R, Almeida L. Photoperiodism and genetic control of the long juvenile period in soybean: a review. *Crop Breeding and Applied Biotechnology*. 2001; 1(1): 72-92.
7. Hugh-Jones M. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitology Today*. 1989; 5(8): 244-251.
8. Sánchez N, González-Piqueras J, González-Dugo M, Baroncini-Turricchia G, Torres E, Calera A, et al. Water balance at plot scale for soil moisture estimation using vegetation parameters. *Agricultural and Forest Meteorology*. 2012; 166-167: 1-9.
9. Ran L, Lu X. Delineation of reservoirs using remote sensing and their storage estimate: an example of the Yellow River basin, China. *Hydrological Processes*. 2012; 26(8): 1215-1219.
10. Vikas K. Identification of Crop Pests using Machine Learning Classifier System. *International Journal for Research in Applied Science and Engineering Technology*. 2020; 8:1020-1025.
11. Vörösmarty C, Federer C, Schloss A. Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. *Journal of Hydrology*. 1998; 207: 3-4.

12. Tennakoon S, Murty V, Eiumnoh A. Estimation of cropped area and grain yield of rice using remote sensing data. *International Journal of Remote Sensing*. 1992; 13(3): 427-439.
13. Vafeidis A, Drake N, Wainwright J. A proposed method for modelling the hydrologic response of catchments to burning with the use of remote sensing and GIS. *Catena*. 2007; 70(3): 396-409.
14. Kobayashi N, Tani H, Wang F, Sonobe R. Crop classification using spectral indices derived from Sentinel-2A imagery. *Journal of Information and Telecommunication*. 2019; 4(1): 67-90.
15. Chen Y, Hou J, Huang C, Zhang Y, Li X. Mapping Maize Area in Heterogeneous Agricultural Landscape with Multi-Temporal Sentinel-1 and Sentinel-2 Images Based on Random Forest. 2021; 13(15): 2988.
16. Belgiu M, Csillik O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*. 2018; 204: 509-523.
17. Huang X, Huang J, Li X, Shen Q, Chen Z. Early Mapping of Winter Wheat in Henan Province of China Using Time Series of Sentinel-2 Data. *GIScience & Remote Sensing*. 2022; 59(1): 1534-1549.
18. Zhang H, Du H, Zhang C, Zhang L. An automated early-season method to map winter wheat using time-series Sentinel-2 data: A case study of Shandong, China. *Computers, Environment and Urban Systems*. 2021; 182: 748-762.
19. Almeida C, Batty M, Monteiro A, Câmara G, Soares-Filho B, Cerqueira G, et al. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*. 2003; 27(5): 481-509.
20. Zhang R, Wang S, Gao W, Sun W, Wang J, Niu L. Study on remote Sensing Classification of County crops based on time Series Vegetation Index. *Journal of Agricultural Machinery*. 2015; 46: 251-257.
21. Wei P, Xu X, Yang G, Li Z, Wang J, Chen G. Crop remote sensing classification based on the variation characteristics of vegetation index in multi-temporal images. *China Agricultural Science and Technology Guide*. 2019; 21(2): 60-67.
22. Mitran T, Meena R, Chakraborty A. *Geospatial Technologies for Crops and Soils: An Overview*. Springer. 2020: 1-48.
23. Benedetti R, Rossini P. On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment*. 1993; 46(3): 311-326.
24. Casa A, Ovando G. Normalized Difference Vegetation Index (NDVI) and phenological data integration to estimate county yield of corn in Córdoba, Argentina. *Agricultura Técnica*. 2013; 34: 362-371.
25. Lopresti M, Bella C, Degioanni A. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. *Information Processing in Agriculture*. 2015; 2(2): 73-84.
26. Arab S, Noguchi R, Matsushita S, Ahamed T. Prediction of Grape Yields from Time-Series Vegetation Indices Using Satellite Remote Sensing and a Machine-Learning Approach. *Remote Sensing Applications Society and Environment*. 2021; 22: 100485.
27. Marshall M, Thenkabail P. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2015; 108: 205-218.
28. Kayad A, Al-Gaadi K, Tola E, Madugundu R, Zeyada A, Kalaitzidis C. Assessing the Spatial Variability of Alfalfa Yield Using Satellite Imagery and Ground-Based Data. *PLoS ONE*. 2016;11(6).
29. Huang J, Wang H, Dai Q, Han D. Analysis of NDVI Data for Crop Identification and Yield Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2017; 7(11): 4374-4384.
30. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google earth engine: planetary- scale geospatial analysis for everyone. *Remote Sensing of Environment*. 2017; 202(1): 18-27.
31. Kobayashi H. Geospatial Data Analysis on Google Earth Engine. *Journal of The Remote Sensing Society of Japan*. 2018; 38(2): 125-130.
32. Liu F, Yu Z, Zhang Z, Zhao J, Wu Y, Sun Q. Study on the Curing Technology of Alfalfa Hay in Hexi Corridor Area. *Acta Agrestia Sinica*. 2009; 17(4): 485-489.
33. Qin W. The achievement transformation of "demonstration and Application of standardized production Technology of Alfalfa" was carried out smoothly. *Modern Rural Science and Technology*. 2015; 1: 18-77.
34. Lamb J, Jung H, Riday H. Growth environment, harvest management and germplasm impacts on potential ethanol and crude protein yield in alfalfa. *Biomass and Bioenergy*. 2014; 63: 114-125.
35. Gong P, Liu H, Zhang M, Li C. Stable classification with limited sample: Transferring a 30m resolution sample set collected in 2015 to mapping 10m resolution global land cover in 2017. *Science Bulletin*. 2019; 64(6): 370-373.
36. Xie B, Tang Y. Europe accelerated the promotion of the "Copernicus. *Space Exploration*. 2015; 6: 36-39.
37. He Y, Huang C, Li H, Liu Q, Liu G, Zhou Z, et al. Land-cover classification of random forest based on Sentinel-2A image feature optimization. *Resources Science*. 2019; 41(5): 992-1001.
38. Zhong H, Liu J, Wang W. The present situation and trend of Scientific big data Intelligent Analysis Software. *Journal of the Chinese Academy of Sciences*. 2018; 33(8): 812-817.
39. Zhang H, Huang H, Wei K, Zhong H. Land coverage classification of Sentinel-2A image based on object-oriented and rules: A case study of Duchang County, Jiangxi Province. *Bulletin of Surveying and Mapping*. 2020; 0(6): 111-117.
40. Bao X, Wang Y, Feng Q, Ge J, Hou M, Liu C, et al. Spatial distribution extraction of alfalfa based on Sentinel-2 and GF-1 images. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*. 2021; 37(16): 153-160.
41. Ma L, Gu X, Xu X, Huang W, Jia J. Remote sensing measurement of corn planting area based on field-data. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*. 2009; 25(8): 147-151.

42. Hou X, Di X, Hou W, Wu L, Liu J, Wang J, et al. Accuracy evaluation of land use mapping using remote sensing techniques in coastal zone of China. *Journal of Geo-Information Science*. 2018; 20(10): 1478-1488.
43. Huang D, Xu S, Sun J, Liang S, Song W, Wang Z. Accuracy assessment model for classification result of remote sensing image based on spatial sampling. *Journal of Applied Remote Sensing*. 2017; 11(4): 1-13.
44. Yuan G, Zhang S. Overwintering cutting time and stubble height of alfalfa. *China Rural Science and Technology*. 2001; 10(15):14.
45. Noland R, Coulter M, Tiede T, Baker J, Martinson K, Sheaffer C. Estimating alfalfa yield and nutritive value using remote sensing and air temperature. *Field Crops Research*. 2018; 222(1): 189-196.
46. Iniyan S, Jebakumar R. Mutual Information Feature Selection (MIFS) Based Crop Yield Prediction on Corn and Soybean Crops Using Multilayer Stacked Ensemble Regression (MSER). *Wireless Personal Communications*. 2022; 126(3): 1935-1964.