

Article

Mapping and Yield Estimation of Cultivated Alfalfa Using Cutting-Induced NDVI Peak–Trough Features from Sentinel-2 Time Series

Jie Liu ^{1,2}, Qisheng Feng ^{3,*}, Shuai Fu ³, Tiangang Liang ³, Jinlong Gao ³ and Wei Sun ^{3,4}

¹ College of Environment and Life Sciences, Weinan Normal University, Weinan 714099, China; lj2025joy@163.com

² Key Laboratory for Ecology and Environment of River Wetlands in Shaanxi Province, Weinan 714099, China

³ State Key Laboratory of Grassland Agro-Ecosystems, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China; fush2023@lzu.edu.cn (S.F.); tgliang@lzu.edu.cn (T.L.); gaojinlong@lzu.edu.cn (J.G.); swlymyy@163.com (W.S.)

⁴ Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 750002, China

* Correspondence: fengqsh@lzu.edu.cn

Abstract

Alfalfa (*Medicago sativa*) is an important forage source for grassland agricultural development; developing accurate and efficient methods for alfalfa identification and yield estimation using remote sensing is of considerable interest. 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 a strong dependence on training data. In this study, using Sentinel-2 high-resolution imagery and the Google Earth Engine (GEE) platform, we constructed a cloud-free normalized difference vegetation index (NDVI) time-series dataset and proposed an effective method for alfalfa feature extraction and yield estimation. The results show that: (1) the producer's accuracy, user's accuracy, overall accuracy, and Kappa coefficient of alfalfa identification using the 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, and was mainly distributed in the northern region of the Qilian Mountains. (2) NDVI showed a highly significant correlation with alfalfa hay yield, and the power function regression model performed best, 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 4.838 × 10⁸ kg and an average hay yield of 4464.95 kg/hm². The proposed method has significant application potential for automated and rapid remote sensing-based identification and yield estimation of large-scale alfalfa cultivation.

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



Academic Editor: Jayme Garcia Arnal Barbedo

Received: 20 May 2026

Revised: 23 June 2026

Accepted: 23 June 2026

Published: 29 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Alfalfa (*Medicago sativa*) is a forage species of global importance due 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 livestock industry development, thereby promoting the rapid development of the alfalfa industry [3,4]. However, there is a lack of efficient and accurate methods for large-scale monitoring of alfalfa grasslands, limiting the availability of reliable data for production and management.

Remote sensing technology enables large-scale, real-time monitoring of crop growth and yield estimation [5–8]. This technology is non-contact and non-destructive, and provides timely, fast, dynamic, and large-scale spatial information [9]. Therefore, it can help forecast crop growth in advance, enabling government agencies to implement timely and effective management strategies [10]. Yield estimation using remote sensing typically includes two key steps: (1) crop identification and area extraction and (2) growth monitoring and yield forecasting [11]. The selection of remote sensing data sources (e.g., GF-1, Landsat 8 and Sentinel-2) is crucial for crop identification [12,13]. Sentinel-2 data are particularly suitable for high spatial and temporal resolution analysis of crop distribution [14–18]. Previous studies have employed various methods to extract vegetation and crop distribution. Kobayashi et al. [14] applied a random forest (RF) algorithm to 91 spectral indices derived 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%. Their findings demonstrated the utility of Sentinel-2A data for generating high-quality crop cover maps. Using multi-temporal Sentinel-1 and Sentinel-2 data and applying the 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 the 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. Based on the above studies, 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 in crops. 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 studies for extracting “critical time nodes” based on remote sensing features (e.g., difference method [20], threshold setting method [21], etc.) focus only on the differences in remote sensing VIs of local crops, and are limited by uncertain identification results and poor universality. Previous studies, including Bao et al. [22], have demonstrated that NDVI time-series peak and trough features are effective for alfalfa field extraction. However, these methods are generally developed and validated in relatively homogeneous and limited study areas, which restricts their ability to generalize to more complex agricultural landscapes. In addi-

tion, such methods often rely on semi-manual parameter calibration and region-specific threshold tuning, which reduces their level of automation and increases the dependency on local prior knowledge. Moreover, the use of optical remote sensing data is still affected by cloud contamination and temporal gaps, which may lead to incomplete NDVI time-series reconstruction and further influence the robustness of feature extraction. As a result, these methods may face challenges when applied to large-scale and ecologically diverse regions, where crop phenology and management practices vary significantly across space, limiting their cross-regional transferability.

Spectral vegetation indices (VIs) can reflect potential crop yield. The normalized difference vegetation index (NDVI), derived from red and near-infrared (NIR) bands, is the most commonly used VI [23,24]. Casa et al. [25] integrated phenological data of corn with NDVI estimates from National Oceanic and Atmospheric Administration Advanced Very High-Resolution Radiometer (NOAA-AVHRR) system to estimate county-level corn yield in Córdoba, Argentina. The result showed that NDVI during the reproductive stage exhibited a strong correlation with corn yield across all sowing dates. Lopresti et al. [26] estimated wheat yield using an empirical model during the flowering period based on MODIS NDVI and field observations at the departmental level in northern Buenos Aires Province, Argentina, and found a strong agreement between model estimates and official yields ($R^2 = 0.75$). Arab et al. [27] proposed a method to predict table grape yield using an artificial neural network (ANN) and time-series VIs derived from Landsat 8 data. The results showed that NDVI achieved the highest accuracy among all VIs, and the correlation coefficients (R) for 2017, 2018, and 2019 were 0.94, 0.95, and 0.92, respectively. Marshall and Thenkabail [28] 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. The R^2 values of the estimation models for rice, alfalfa, cotton and maize were 0.91, 0.81, 0.97 and 0.94, respectively. Kayad et al. [29] conducted a study in a center pivot irrigated alfalfa field in Saudi Arabia to assess the spatial variability of alfalfa yield using Landsat-8 imagery and 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 previous yield estimation studies have been conducted at relatively small spatial scales. Large-scale alfalfa yield estimation requires accurate mapping of its spatial distribution as a prerequisite. However, traditional large-scale monitoring methods for alfalfa distribution and yield estimation suffer from several limitations, including limited spatial resolution of remote sensing data, high computational burden in processing large volumes of data, and strong dependence on training samples. These limitations hinder their application in large-scale remote sensing-based identification and yield estimation of cultivated alfalfa crops.

In addition, due to factors such as temporal coverage, spatial variability, and image quality, remote sensing data suitable for specific study areas are often limited in both quality and quantity. As a result, large-scale crop surveys relying on limited remote sensing data over a short time can be challenging [30]. Firstly, downloading and preprocessing high-resolution remote sensing data for large-area crop identification is time-consuming and inefficient, leading to low timeliness, reduced accuracy, and limited scalability of the results. Google Earth Engine (GEE) is a cloud-based geospatial data analysis platform that provides access to massive global-scale Earth science data, particularly satellite imagery, and offers powerful support for data processing and analysis [31,32]. Coding on GEE

enables efficient data preprocessing and large-area satellite data access, demonstrating strong potential for large-scale crop identification and yield inversion.

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

2. Materials and Methods

2.1. Study Area

The Hexi Corridor in northern China is a major production region of high-quality alfalfa for domestic markets. Within this region, three cities—Zhangye, Jinchang and Wuwei—account for most commercial alfalfa production and constitute a key high-quality forage base in China [33]. These cities span an area of 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 (Figure 1a). The landscape is characterized by mountains and plains, interspersed with Gobi Desert and oasis systems. The climate is a continental temperate arid climate 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 highly favorable for alfalfa cultivation. The main land cover types include grassland and forest, followed by cropland, shrubs and non-vegetated areas (Figure 1b). Agriculture plays a dominant role in the local economy, with farming, animal husbandry and breeding activities being common. Alfalfa cultivation has expanded rapidly in recent years with strong government support.

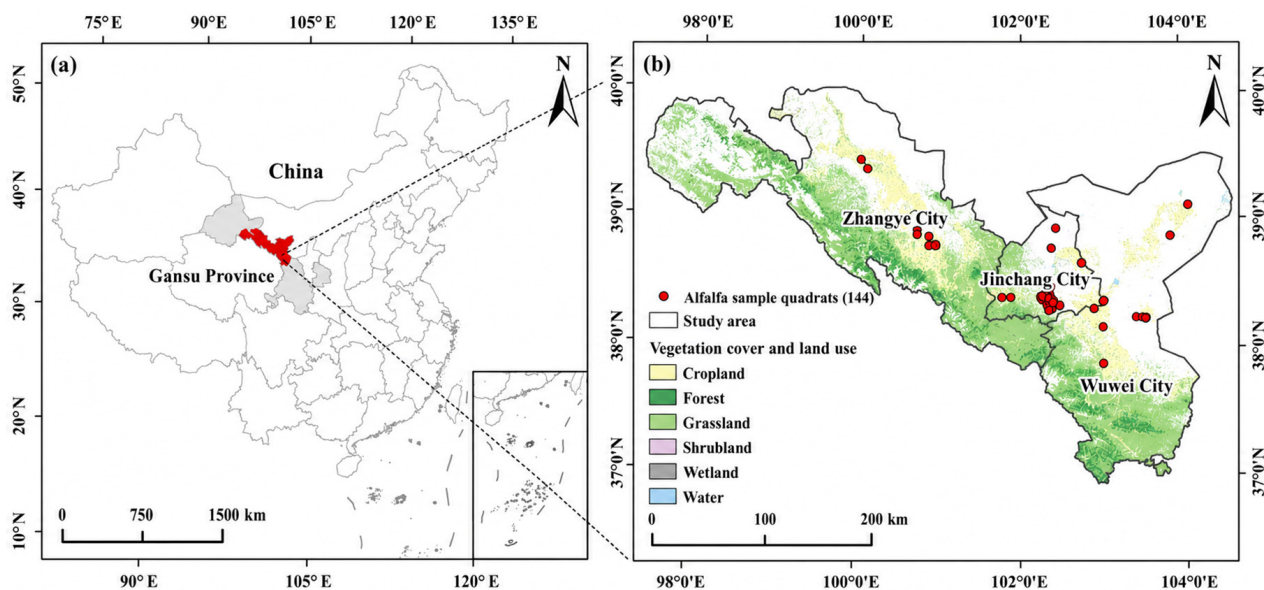


Figure 1. (a) Geographical location of the study area, (b) distribution of sample plots and land cover type.

The study area's alfalfa planting companies have mechanized most production processes, including sowing, harvesting, cutting, turning and bundling. They have also adopted high-yielding advanced integrated agronomic techniques such as seed coating, rhizobia inoculation, plastic mulching, precision seeding and integrated control of diseases, insects and weeds, which have enhanced alfalfa yield and quality [34]. Alfalfa hay contains 16–18% crude protein, making it a highly palatable and nutritious fodder for cattle and sheep and near the top among domestic high-quality alfalfa [35].

2.2. Ground-Measured Data

The field investigation was mainly conducted during the alfalfa growing season (May to September) in 2020. To minimize spatial autocorrelation and ensure sample independence, most sampling plots were spaced more than 10 km apart, and sampling locations were deliberately distributed across different agricultural zones rather than clustered in contiguous fields. We surveyed 81 alfalfa grassland sample plots (sizes of 100 m × 100 m or larger) in the first survey (from 27 May to 6 June), 51 sample plots in the second survey (from 17 July to 26 July), and 12 sample plots in the third survey (from 20 September to 24 September), and three subplots (0.5 m × 0.5 m) were randomly established within each plot. A total of 144 plots and 432 subplots were collected in the field. In each plot, information on alfalfa variety, annual cutting frequency, current cutting status, planting years, irrigation and fertilization practices, and management entities was collected. Meanwhile, UAV images were acquired at flight altitudes of 20 m and 50 m, and a panoramic photo was taken at an altitude of 100 m using a DJI Phantom 4 Pro unmanned aerial vehicle (SZ DJI Technology Co., Ltd., Shenzhen, China). In each subplot, latitude, longitude, alfalfa height, leaf area and chlorophyll content were recorded. The aboveground biomass was harvested with a 5 cm stubble height and weighed as fresh biomass in the field, and then dried at 65 °C for 72 h until constant weight. The dry weight was used as the primary input for alfalfa yield estimation modeling. To validate classification accuracy, the geographical locations of 100 non-alfalfa plots (corn, chili peppers, onions, oats, etc.) were also recorded. ArcGIS 10.8 was used to establish an attribute database and spatial database of field survey plots (Figure 1).

2.3. Cultivated Land Data

The Finer Resolution Observation and Monitoring–Global Land Cover 10 (FROM-GLC10) with a spatial resolution of 10 m was used as the cultivated land dataset in this study (<http://data.ess.tsinghua.edu.cn/>). It was developed by Professor Peng Gong using the RF method based on the world's first multi-season sample database derived from Sentinel-2A global imagery and the GEE cloud computing platform. It classifies global land cover into ten categories: crop, forest, grass, shrub, wetland, water, tundra, impervious, bare land and snow/ice [36].

2.4. Sentinel-2 Data

The Sentinel series satellites provide free and high-quality remote sensing data with high spatial and temporal resolution. They also provide multispectral imagery at a global scale for monitoring Earth's environmental conditions and responding promptly to sudden emergencies such as floods, forest fires and landslides [37].

The European Copernicus Mission operates two polar-orbiting high-resolution multispectral 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 with 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 [38]. The Sentinel-2 Level-1C product used in this study is radiometrically and geometrically corrected using a digital elevation model and provides top-of-atmosphere (TOA) reflectance values for each pixel across 13 spectral bands and three quality assessment bands (QA10, QA20 and QA60). The QA60 band contains cloud mask information, which is used to remove opaque and cirrus clouds from the image.

2.5. GEE Platform Data Download and Processing

GEE is a cloud-based platform that enables online visualization, computation, analysis and processing of massive global-scale Earth science datasets, particularly satellite imagery [39]. Sentinel-2 imagery used in this study was stored and fully processed on the GEE cloud platform using online scripting.

Temporal range definition: To obtain alfalfa plant data for the entire phenological period (seedling, budding, flowering and maturing stages), the temporal range was defined as 1 April to 30 November 2020, ensuring the completeness and suitability of the data.

NDVI calculation: NDVI was calculated using Sentinel-2 red (B4) and NIR (B8) bands on the GEE platform. Each image was assigned a unique identifier for temporal tracking and processing. The formula is as follows:

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

Cloud removal: Sentinel-2 QA60 band was used for cloud detection based on bitwise information. Bit 10 represents opaque clouds and Bit 11 represents cirrus clouds. Pixels flagged as clouds or cirrus were masked and removed. Cloud-free observations were retained for subsequent analysis.

Interpolation optimization: Sentinel-2 imagery is frequently affected by cirrus clouds and cloud shadows, resulting in abnormal NDVI values and discontinuities in the NDVI time series that may hinder alfalfa identification. Therefore, interpolation optimization was applied to reconstruct a continuous and high-quality NDVI time-series dataset for the study area. A three-image sliding-window interpolation algorithm was implemented on a pixel-by-pixel basis (Figure 2). Given the 10 m spatial resolution and 5-day revisit period of Sentinel-2 imagery, three consecutive observations were defined as N1, N2, and N3. For example, N1, N2, and N3 may represent NDVI values acquired on April 10, April 15, and April 20, respectively. The value of the central observation (N2) was updated according to the following rules: (1) if $N2 > 0$, original value; (2) if $N2 < 0$ and $N1 < 0$ and $N3 < 0$, original value; (3) if $N2 < 0$ and $N1 < 0$ and $N3 > 0$, value of N3; (4) if $N2 < 0$ and $N1 > 0$ and $N3 < 0$, value of N1; and (5) if $N2 < 0$ and $N1 > 0$ and $N3 > 0$, average of $(N1 + N3)$. This process was repeated for all consecutive time steps (N2, N3, N4. . .) until all images within the selected time period had been processed. This procedure effectively reduced abnormal fluctuations and discontinuities in the NDVI time series, resulting in a temporally continuous dataset for subsequent alfalfa identification and yield estimation.

2.6. Alfalfa Feature Extraction

The study area follows a one-crop-per-year cropping system, with major crops including spring and winter wheat, spring corn, cotton, onions, potatoes, rapeseed and Chinese kale. The phenological periods of alfalfa and several of these crops overlap considerably (Figure 3), which complicates discrimination of alfalfa from other crops using vegetation index thresholds at a specific imaging time. Through analysis, we found that alfalfa exhibits distinct temporal dynamics compared with other crops during multiple harvest cycles. This results in abrupt changes in spectral-temporal signatures after cutting events. This characteristic was exploited to improve remote sensing-based alfalfa identification.

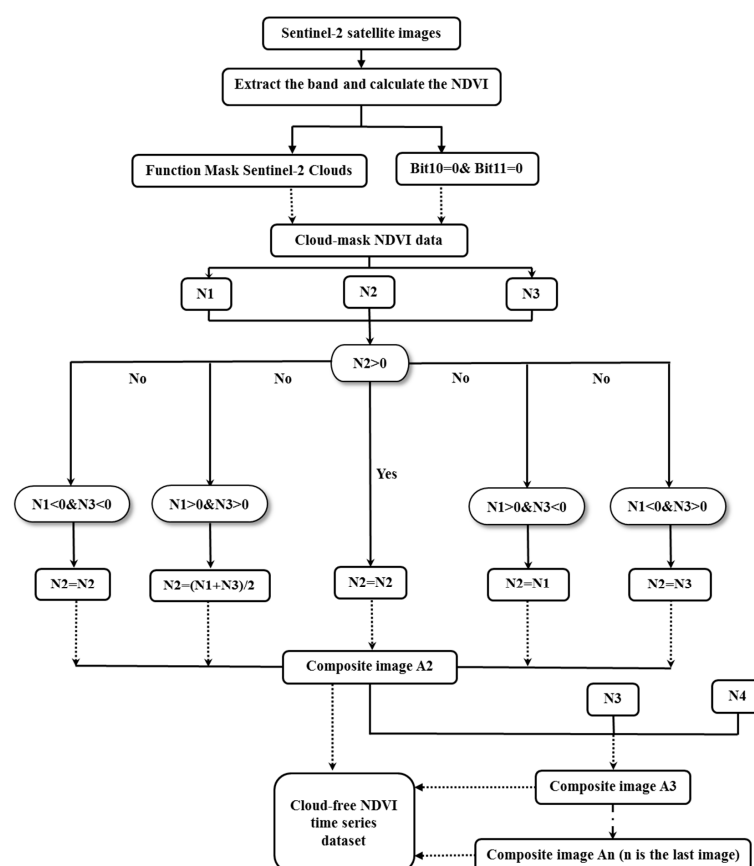


Figure 2. Workflow of Sentinel-2 NDVI time-series processing and interpolation optimization. N1, N2, and N3 represent three consecutive Sentinel-2 observations in chronological order.

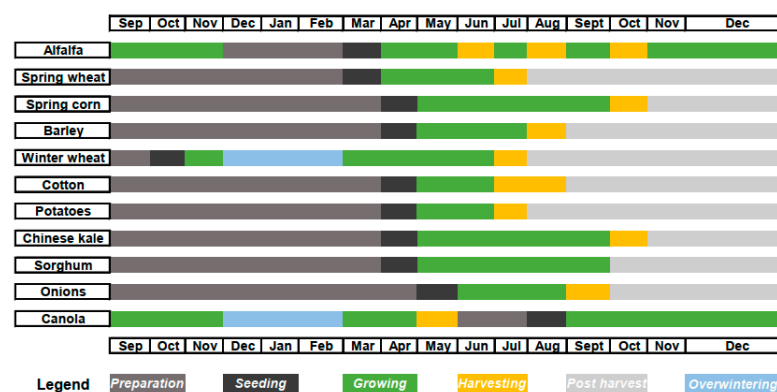


Figure 3. Calendar and type map of main crops in the study area.

NDVI time-series curves capture temporal variation patterns of vegetation growth during the growing season [40]. The NDVI of alfalfa grassland exhibits multiple NDVI fluctuations characterized by 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 discontinuities in the time series. The number and timing of the troughs reflect the number and timing of the harvests. In the study area, there were 2–4 troughs. Based on this characteristic, an algorithm was developed for alfalfa identification and yield estimation using NDVI time-series data by analyzing spatiotemporal variations.

2.7. Area Extraction and Yield Estimation

The periodic harvesting of alfalfa provides a unique opportunity to discriminate it from other crops using remote sensing time-series data. Therefore, a trough recognition algorithm was developed to detect cutting-induced troughs in NDVI time-series data.

The trough recognition algorithm was implemented using MATLAB R2022a's findpeaks function (Figure 4), which identifies local maxima (Peaks) in the input signal and allows detection of peaks 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 amplitude between peak and trough in the NDVI time series. 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 [22] and the sensitivity analysis (in the supplementary material), the MPP was set to 0.4 for alfalfa extraction in our study area. To briefly summarize the sensitivity analysis, the MPP threshold was varied from 0.35 to 0.50 at an interval of 0.05, and the corresponding overall accuracy values were 82.26%, 94.26%, 90.35%, and 89.41%, respectively. The highest overall accuracy was obtained when MPP = 0.40 (Figure S1). Therefore, MPP = 0.40 was selected as the optimal threshold, as it provided the best balance between detecting true alfalfa cutting signals and suppressing false detections. 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.

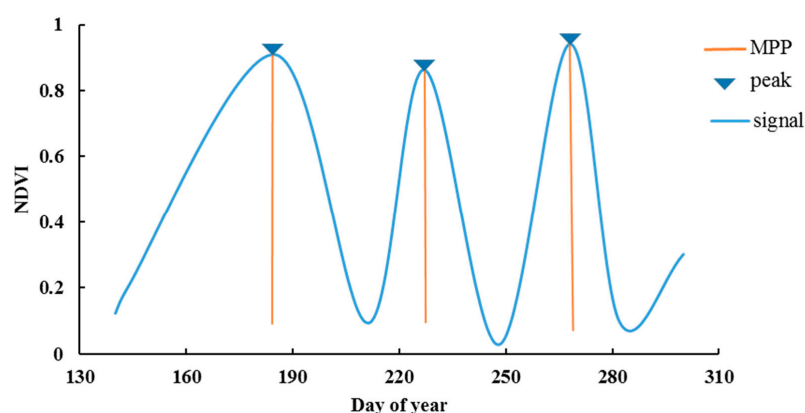


Figure 4. NDVI time-series curve and concept map of Findpeaks function.

While applying 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 process generated pixel-wise feature datasets, including number of times (NOT), day of year (DOY) and peak value of NDVI (PVON), which were used to derive the following parameters:

(1) The spatial distribution of alfalfa was determined by constraining 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 operations, including majority filtering and noise removal, 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 (Figure 5).

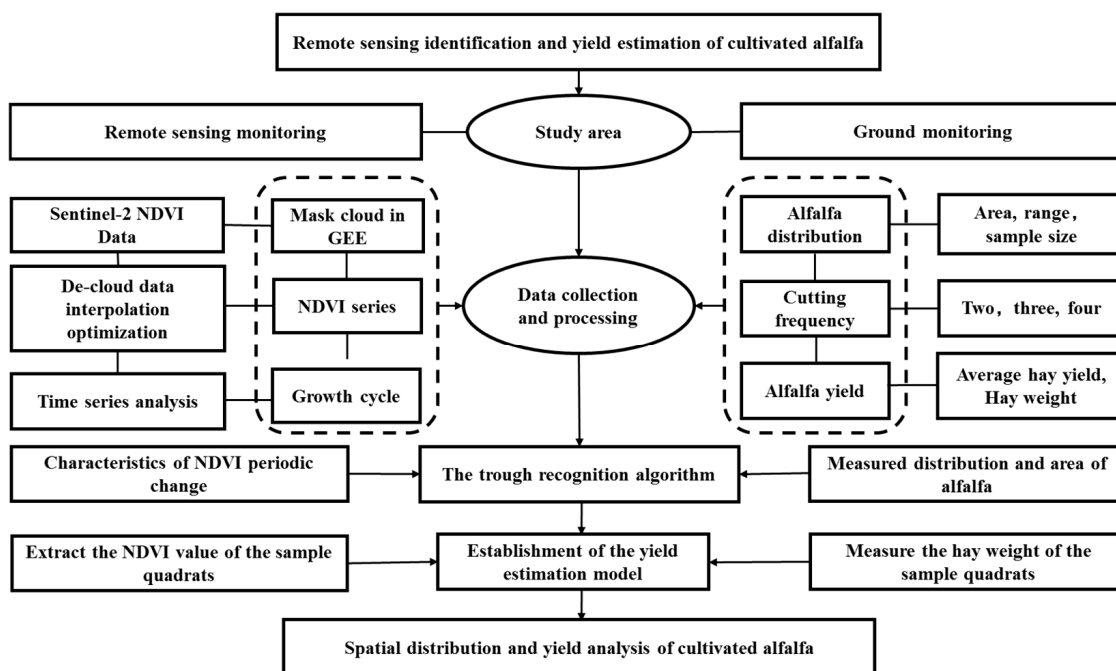


Figure 5. Methods and technical flowchart.

2.8. Estimated Yield Model

Based on the geographic coordinates of the field sample plots, NDVI values for each plot were extracted from the GEE platform, and extraction dates were selected to be as close as possible to field measurement dates to ensure the validity and reliability of our yield estimation model. Then, 17 plots contaminated by cloud or cloud shadow were excluded from the total 144 plots, and the remaining 127 plots were used to construct the yield estimation models (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, where y represents alfalfa hay yield (kg/hm^2); x represents the NDVI value; and A , B , C , D , and F are parameters.

2.9. Accuracy Assessment and Validation Strategy

To ensure an unbiased evaluation, the field samples were divided according to their roles in MPP threshold calibration and accuracy validation. Among the 144 alfalfa samples, 30% were randomly selected as the calibration subset for determining the MPP threshold, while the remaining 70% were reserved as independent alfalfa validation samples. The calibration samples were not used in the subsequent accuracy assessment.

The evaluation metrics include OA, producer's accuracy (PA), user's accuracy (UA), and Kappa coefficient, reflecting the accuracy of image classification from different perspectives [41,42]. In addition, R^2 , root mean square error (RMSE) and mean relative error (MRE) are used to evaluate 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

Figure 6 illustrates unoptimized and optimized NDVI time-series curves derived from Sentinel-2 data collected in field sampling plots in Wuwei City in 2020. The unoptimized NDVI time series exhibits a clear discontinuity in the time range of 240–260 d, which disrupts temporal continuity and degrades classification performance. Moreover, in the time range of 160–240 d, NDVI values frequently dropped to 0. This was caused by cloud 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.

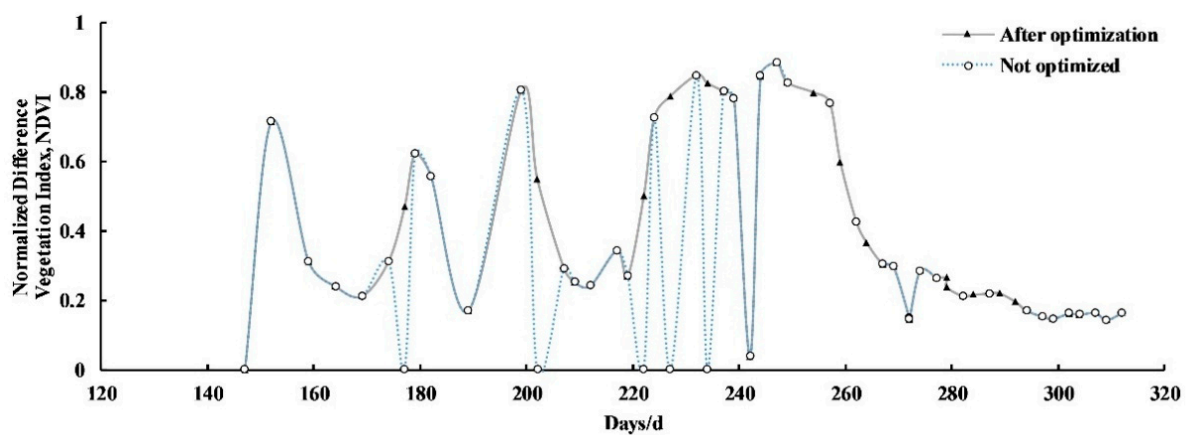


Figure 6. NDVI timing data optimization results.

After optimization, the faults and outliers on the NDVI time-series curve were eliminated and the curve was smooth with obvious peaks and troughs. Importantly, the interpolation process was only used to reconstruct missing observations, thereby preserving the original NDVI dynamics. Therefore, the improved Sentinel data provides more reasonable NDVI time-series data that met the recognition needs of the trough-based algorithm, which is essential for implementing the trough-based algorithm.

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 hm². Table 1 shows alfalfa areas and their proportions of local cultivated land in different regions. At the city level, Zhangye had the largest alfalfa area (19,268.77 hm²), followed by Jinchang, whereas Wuwei had the smallest (11,638.18 hm²). At the county level, Yongchang County in Jinchang city had the largest alfalfa area of 14,568.70 hm², accounting for 1.76% of the county's total cultivated area and making it the most concentrated alfalfa planting area in the study area. Alfalfa distribution in Zhangye was relatively uniform across counties, and all were above 2000 hm², with Gaotai County having the largest area at 4374.48 hm², accounting 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 5375.59 hm² and 4368.00 hm², respectively. These two counties accounted for 83.72% of the city's total alfalfa area, while Tianzhu Tibetan Autonomous County had only 10.60 hm² of alfalfa, 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	3325.24	0.95
	Gaotai County	4374.48	0.60
	Linze County	2986.35	0.42
	Ganzhou District	3650.71	0.22
	Minle County	2697.99	0.25
	Shandan County	2234.00	0.3
Jinchang City	JinChuan District	1317.56	0.39
	Yongchang County	14,568.70	1.76
	Minqin County	5375.59	0.39
Wuwei City	Liangzhou District	1883.99	0.11
	Gulang County	4368.00	0.61
	Tianzhu Tibetan Autonomous County	10.60	0.01

Figure 7 illustrates the alfalfa spatial distribution in the study area in 2020 based on remote sensing information extraction. The results indicate that alfalfa is mainly concentrated north of the Qilian Mountains. Based on the field investigation, Figure 7a–d are taken as the typical areas of alfalfa recognition results in the study area, and the results show that the identified alfalfa plots in the four regions exhibit clear spatial patterns and high spatial compactness. Among them, the alfalfa fields in the typical areas (a), (c) and (d) are rectangular fields distributed over large continuous areas. The shape of the alfalfa field in (b) is round, and the overall distribution is relatively scattered, which is associated with center-pivot irrigation systems used in local management practices. A total of 144 alfalfa samples and 100 non-alfalfa samples, including corn, chili pepper, onion, oat, and other crop samples, were collected through field investigation. As described in Section 2.9, 30% of the alfalfa samples were used only for MPP threshold calibration and were excluded from the accuracy assessment. The remaining 70% of the alfalfa samples, together with all 100 non-alfalfa samples, were used as the independent validation dataset. Based on this validation dataset, the PA, UA, OA, and Kappa coefficient were 98.51%, 91.67%, 94.26%, and 0.88, respectively.

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 were 98.51%, 91.67%, 94.26%, and 0.88, respectively.

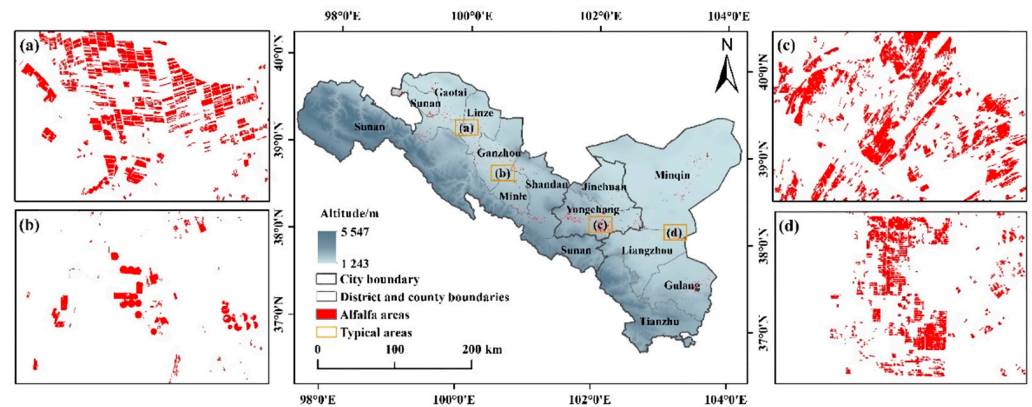


Figure 7. Spatial distribution of alfalfa in the study area in 2020; (a–d) are the typical areas of alfalfa recognition results.

3.2. Yield Estimation Results and Analysis

3.2.1. Establishment of the Yield Estimation Model

The results of alfalfa yield estimation models are shown in Figure 8. NDVI showed a positive correlation with actual yield, with R^2 values exceeding 0.55 across all models. Among the models, the R^2 values of yield estimation models for the first cutting were all above 0.5, with the power function model achieving the highest R^2 at 0.652, making it the best-performing model ($y = 8484.2x^{0.7092}$). For the second cutting, 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, R^2 values were relatively high (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-performing model, with an R^2 value of 0.756. Some parts of the study area cut alfalfa four times a year, so the alfalfa yield of the fourth cutting should also be estimated using the annual model.

3.2.2. Analysis of the Simulation Error of the Yield Estimation Model

By comparing estimated yields with measured yields, the models for estimating the first, second, and third cuttings and the annual yield showed certain errors, with both underestimation and overestimation observed. The estimated yield of the first cutting had errors within 2000 kg/hm² and showed good agreement with measured values, indicating a good model performance (Figure 9a). The estimated yield of the second cutting showed larger deviations at several sampling sites, with errors exceeding 2000 kg/hm² (Figure 9b). For the third cutting, the differences between the estimated and actual yields were relatively small, with errors below 1000 kg/hm² (Figure 9c), but the model was not representative due to the small sample size. Therefore, the annual yield model is more appropriate for estimating the third-cutting yield.

Regression analysis was conducted between estimated and measured alfalfa yields for the first, second, and third cuttings, as well as annual yield, to evaluate model performance. The regression equations and corresponding R^2 , RMSE and MRE values are shown in Figure 10a–d. The models exhibited relatively 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 an R² value of 0.81. The models of the first cutting and the annual yield also performed well, with R² 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%.

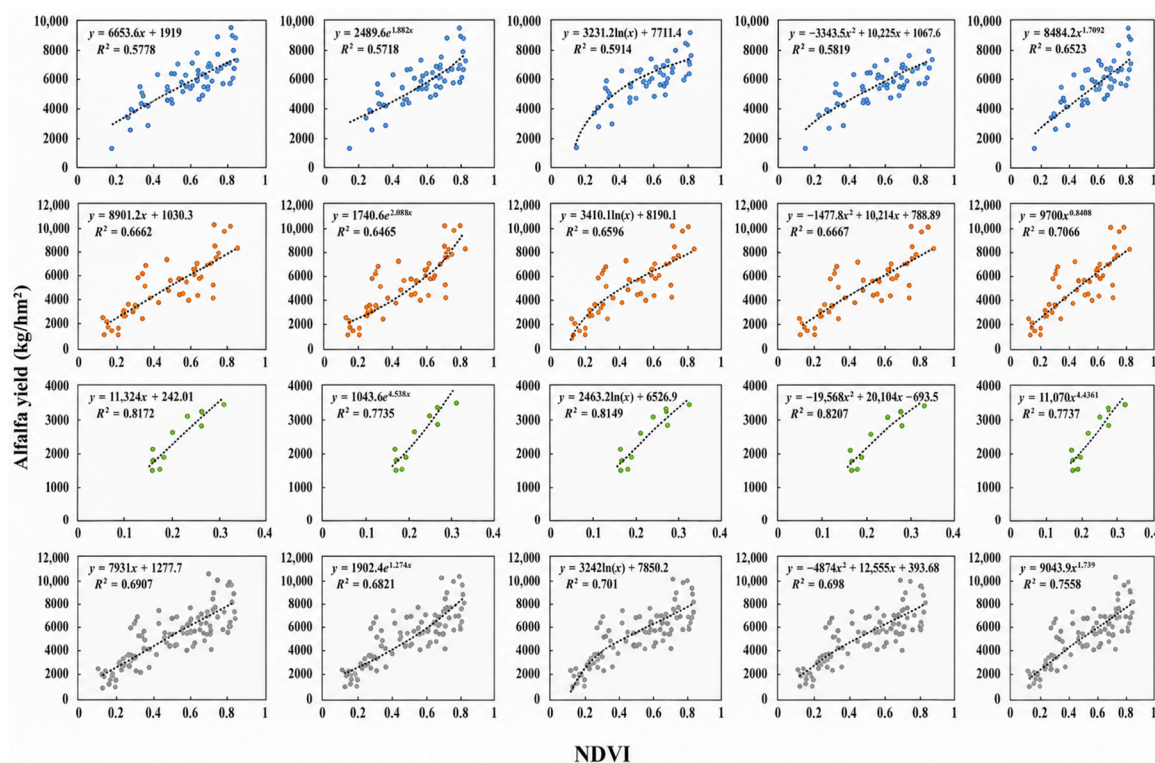


Figure 8. Relationships between NDVI and alfalfa hay yield for the first, second, and third cuttings and annual yield. The X-axis and Y-axis denote the NDVI values and alfalfa yield (kg/hm²), respectively.

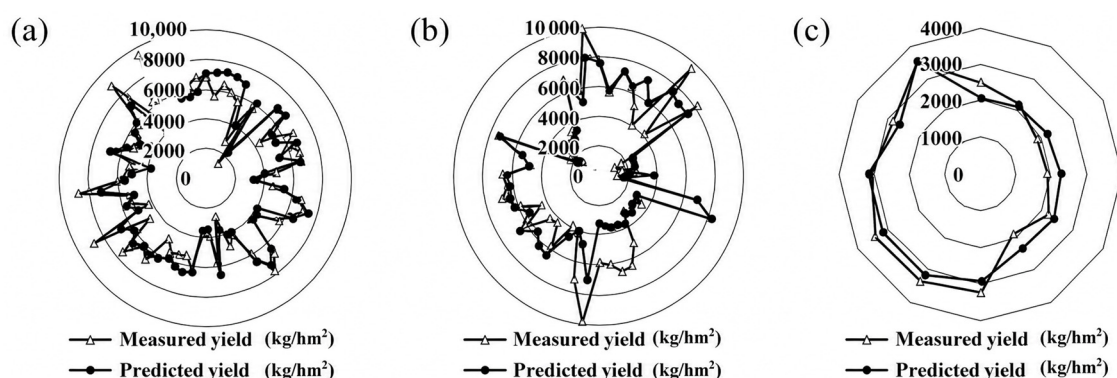


Figure 9. Comparison between measured and predicted alfalfa yields across different cuttings: (a) the first cutting, (b) the second cutting, and (c) the third cutting. The two curves indicate measured and predicted yields, respectively.

3.2.3. Results and Analysis of Alfalfa Yield Estimation

Based on the PVON database obtained through the trough recognition algorithm, alfalfa hay yields for the four cuttings and annual total yield were estimated using the Raster Calculator tool in ArcGIS 10.8 software. Figure 11 shows that the first cutting exhibited the highest yield, followed by the second, third, and fourth cuttings. 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.

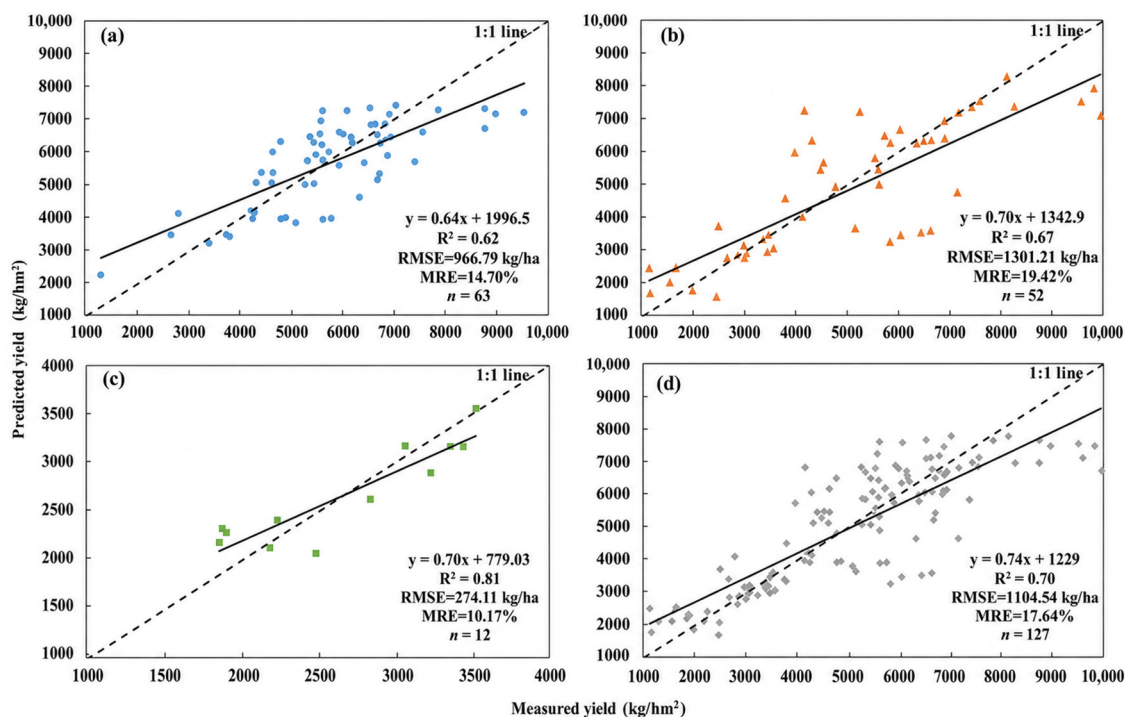


Figure 10. Validation of alfalfa yield estimation models using measured and predicted yields: (a) the first cutting, (b) the second cutting, (c) the third cutting, and (d) the annual yield. Each point represents an individual field sampling plot. The solid line indicates the fitted regression relationship between measured and predicted yields, while the dashed line represents the 1:1 relationship.

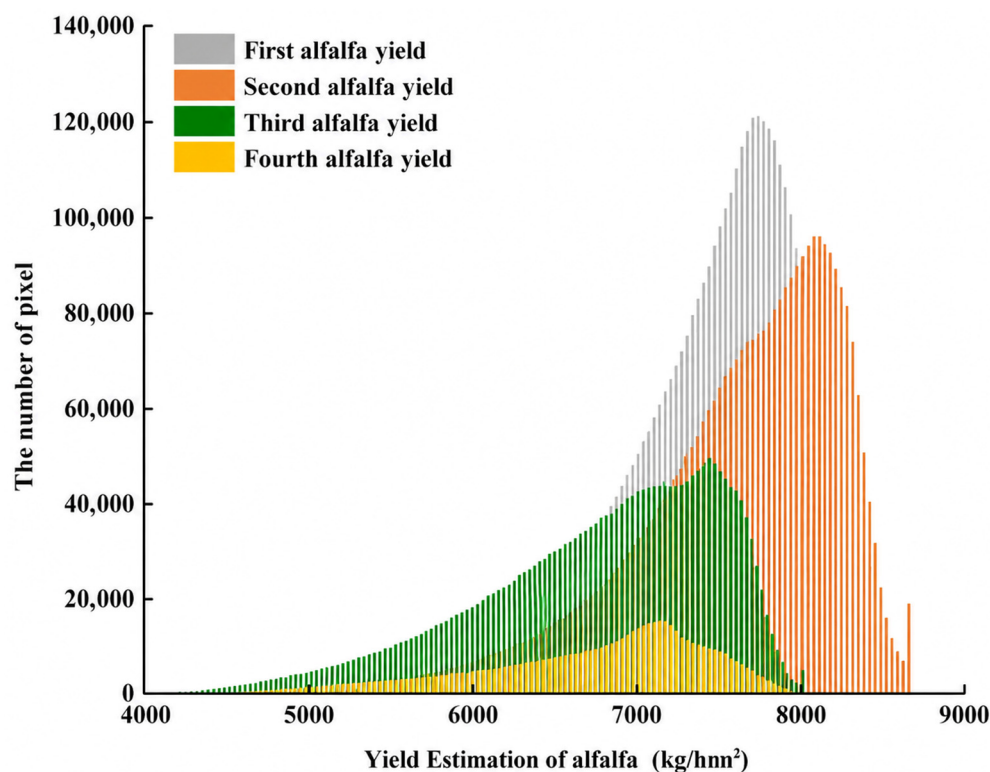


Figure 11. Statistical histogram of alfalfa yield distribution.

According to the results of yield estimation, alfalfa yield per unit area was classified into three levels: 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 spatial distribution of alfalfa yield showed clear spatial heterogeneity (Figure 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 during the third and fourth cutting periods. Overall, the per unit yield of the first two cutting of alfalfa was significantly higher than that of the latter two. Through the yield distribution of typical alfalfa areas, the results indicate that the yield grades of most alfalfa parts have changed among the four cuttings. Therefore, improving field management practices is recommended for low-yield areas to enhance productivity.

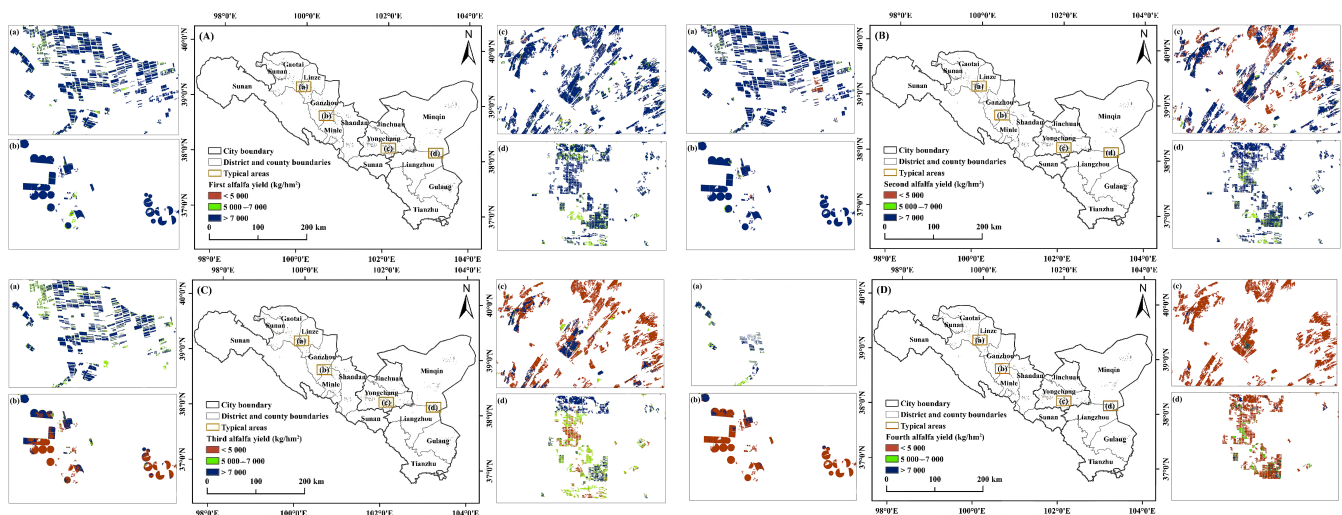


Figure 12. Alfalfa yield distribution maps of the first (A), second (B), third (C) and fourth times (D) in the study area in 2020; (a–d) are the typical areas of alfalfa.

The yield variations can be mainly attributed to the following factors: (1) Alfalfa harvesting time is influenced by many factors, such as planting time, production year, variety, harvesting machinery availability and weather [43], which leads to substantial spatial and temporal variability in alfalfa fields. This affects the remote sensing features of adjacent alfalfa fields in the same phase and thus affects the estimated alfalfa yield of each cutting. (2) Alfalfa yield is also affected by field management practices, 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. 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 may contribute to reduced yield in later cuttings in the study area.

The alfalfa yield in the study area was calculated using the 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 $4229.88\text{--}8482.51 \text{ kg/hm}^2$, with a total hay yield of $30.68 \times 10^7 \text{ kg}$ and an average hay yield of 7164.68 kg/hm^2 . For the second cutting, the alfalfa hay yield was $4489.35\text{--}8763.51 \text{ kg/hm}^2$, with a total hay yield of $21.47 \times 10^7 \text{ kg}$ and an average hay yield of 6080.62 kg/hm^2 . For the third cutting, the alfalfa hay yield of the study area was $4399.28\text{--}8698.17 \text{ kg/hm}^2$, with a total hay yield of $8.95 \times 10^7 \text{ kg}$ and an average hay yield of 3702.17 kg/hm^2 . For the fourth cutting, the alfalfa hay yield of the study area was $4389.16\text{--}8041.90 \text{ kg/hm}^2$, with a total hay yield of $0.55 \times 10^7 \text{ kg}$ and an average hay yield of 912.33 kg/hm^2 . For the entire year, the total alfalfa hay yield of the study area was between $17,497.55 \text{ kg/hm}^2$ and $32,962.10 \text{ kg/hm}^2$, with a total hay yield of $48.38 \times 10^7 \text{ kg}$ and an average hay yield of 4464.95 kg/hm^2 .

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)	8482.51	8763.51	8698.17	8041.9	32,962.1
Minimum hay yield (kg/hm ²)	4229.88	4489.35	4399.28	4389.16	17,497.55
Average hay yield (kg/hm ²)	7164.68	6080.62	3702.17	912.33	4464.95
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-based alfalfa recognition remains challenging. The complex planting and growth characteristics of cultivated alfalfa combine characteristics of both natural grassland and cropland systems, such as sowing and harvesting patterns of crops, and 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 yet effective 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 is derived from Sentinel-2 imagery, which facilitates the mapping of smallholder fields that cannot be effectively captured using MODIS or Landsat data due to their coarser spatial resolution; moreover, the interpolation procedure generates complete NDVI time-series data by filling gaps caused by cloud contamination, thereby improving identification accuracy; (2) the use of the Google Earth Engine (GEE) platform enables efficient processing of large-scale remote sensing data, overcoming limitations of computational resources and cloud coverage in traditional methods, and improving both processing efficiency and data quality; (3) the proposed method is fully automated and does not require manually labeled training data, demonstrating low cost, convenience, and scalability; and (4) it effectively extracts alfalfa spatial distribution from complex crop systems and reduces variability among fields and interference from other crops, achieving satisfactory classification accuracy. Therefore, this method shows strong potential for large-scale monitoring of cultivated alfalfa grassland. Moreover, few studies have been conducted on large-scale alfalfa recognition, but there are studies on other crops’ recognition. For example, Zhang et al. [18] achieved high-resolution remote sensing mapping of winter wheat in Shandong Province, China, using the AEMMS method. Their results demonstrated 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. This is consistent with our study, as both rely on NDVI time-series analysis and crop growth characteristics for identification without requiring training samples, demonstrating its potential for large-scale crop recognition.

Nevertheless, there are still some limitations to the trough recognition algorithm. Alfalfa VIs are affected by interannual and spatial variations in environmental conditions, and climatic factors such as precipitation and temperature can influence alfalfa growth, phenological development, and cutting frequency. Therefore, when the proposed method is applied to different regions, local calibration is required to determine appropriate MPP thresholds and peak–trough number constraints for accurate extraction of cultivated alfalfa. The analysis of MPP values from the 144 sample plots in the study area showed that

MPP varied across geographic locations and cutting stages (Table 3), indicating that the reasonableness of the selected MPP threshold directly affects alfalfa identification accuracy and, consequently, area and yield estimation.

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

The transferability of the selected MPP threshold should therefore be interpreted with caution. In the present study, an MPP value of 0.4 was suitable for irrigated alfalfa systems in the arid and semi-arid Hexi Corridor, where alfalfa fields generally showed regular cutting cycles and clear NDVI peak–trough signals. However, under different climatic conditions, the amplitude and timing of NDVI fluctuations may change substantially. For example, wetter or warmer environments may promote rapid regrowth and increase cutting frequency, whereas drought, low temperature, or shortened growing seasons may reduce peak NDVI values and weaken peak–trough contrasts. Crop management practices, including irrigation regime, fertilization level, cutting date, cutting frequency, and stubble height, may also affect the depth of post-cutting NDVI troughs and the subsequent recovery rate. Therefore, when the algorithm is applied to regions with different environmental conditions or management systems, the MPP threshold and the number-of-trough constraint should be recalibrated using local samples or adaptive threshold selection methods.

In addition, high cloud cover in humid and mountainous regions may lead to the absence of cloud-free observations in adjacent time-series images, resulting in incomplete NDVI time-series data and potentially hampering subsequent peak–trough detection. Therefore, future research should consider fusing multi-source optical remote sensing data, such as Sentinel-2, Landsat, and other high-resolution imagery, with Synthetic Aperture Radar (SAR) data, such as Sentinel-1, which is less affected by cloud contamination. The integration of multi-source remote sensing data may improve image availability, enhance crop feature extraction, and further improve alfalfa identification accuracy in humid or cloudy regions.

The cultivated land boundaries used in this study were based on the 10 m resolution land cover data, which only provides coarse classification 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 support fine-scale field-level analysis based on field units in the subsequent integrated analysis, resulting in “unclassified pixels within homogeneous fields” (some pixels in the same alfalfa field are not recognized as alfalfa due to different growth conditions) and leading to potential 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) combined with machine learning algorithms to enhance the accuracy of alfalfa remote sensing recognition and provide more reliable 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 limitations of traditional methods in real-time crop growth monitoring. This study applied the proposed 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 values, 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 an R^2 of 0.7 between predicted and measured yield. These results are consistent with previous studies. For example, Kayad et al. [29] studied the correlation between alfalfa NDVI obtained from landsat8 data and actual yield, with an R^2 of 0.63, indicating that NDVI is a reliable indicator of alfalfa yield. Although the overall performance of the annual yield model was satisfactory, uncertainties remain in the estimation of individual cutting stages. In particular, the third-cutting yield model was developed using only a limited number of field samples ($n = 12$). Therefore, although the fitted relationship showed a relatively high R^2 value, this result should be interpreted with caution because the small sample size may not fully capture the spatial variability of alfalfa growth, management practices, and yield conditions during the third cutting. For this reason, the third-cutting yield estimates in this study should be regarded as preliminary, and additional field samples are needed to improve the reliability and representativeness of cutting-specific yield estimation.

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 effects, remote sensing satellite data accuracy, and 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 demonstrated 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) for model development and prediction. The prediction results reached more than 90%. Therefore, future studies should compare different VI-based models for alfalfa yield estimation and incorporate more variables, such as altitude, slope, evapotranspiration, and precipitation, into the yield estimation model. This approach enhances the model's generalizability and applicability across regions and provides important scientific and technological support for the integrated management and planning of alfalfa grassland and other crops.

4.3. Limitations and Future Work

This study demonstrates the effectiveness and potential of Sentinel-2 NDVI time-series data for automated alfalfa mapping and yield estimation. However, a key limitation is that the proposed method was evaluated using a single-year dataset (2020), which limits assessment of its temporal robustness under inter-annual climatic variability and constrains evaluation of its broader spatial transferability.

The MPP threshold and the number-of-trough constraint are closely related to local phenological patterns, environmental conditions, and crop management practices. Therefore, although the selected MPP threshold performed well in the irrigated alfalfa systems of the Hexi Corridor, its applicability to other regions or years should be further tested. In par-

ticular, areas with different climatic backgrounds, rainfed production systems, or irregular cutting schedules may require local recalibration or adaptive parameter selection before the algorithm is applied. Therefore, multi-year validation under different climatic conditions is essential to fully evaluate the robustness and transferability of the proposed framework. Future work should focus on extending the temporal coverage of the dataset and developing adaptive or data-driven parameterization strategies for MPP selection. In addition, integrating multi-source remote sensing data (e.g., Sentinel-1 SAR) and environmental variables such as precipitation and evapotranspiration may further enhance the stability and generalizability of phenology-based classification under variable climatic conditions.

5. Conclusions

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

(1) The interpolation optimization significantly improved NDVI time-series quality of the 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, high classification accuracy was achieved 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 as 46,793.21 hm², mainly distributed in areas north of the Qilian Mountains.

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

(4) The validation results showed that the predicted and measured yields were generally consistent, with R² values exceeding 0.6 and MRE values within 20%, suggesting that the proposed NDVI-based yield estimation approach has good application potential. However, the third-cutting yield model was developed using a limited number of field samples ($n = 12$). Therefore, the third-cutting model and corresponding yield estimates should be interpreted with caution and regarded as preliminary. Additional multi-year and multi-site field samples are needed to further improve the robustness and representativeness of cutting-specific yield estimation.

(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 48.38×10^7 kg and an average hay yield of 4464.95 kg/hm².

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy16131255/s1>, Figure S1. Sensitivity analysis of classification accuracy under different MPP values. Figure S2. Comparison of classification performance between the proposed trough recognition algorithm and three machine learning classifiers, including RF, SVM, and CatBoost. Figure S3. Misclassification sources of non-alfalfa crops in the alfalfa classification results. Table S1. Confusion matrix of alfalfa classification results.

Author Contributions: Q.F. and T.L. conceived the study design, developed the research methodology, participated in data analysis, drafted and critically revised the manuscript for important intellectual content, and approved the final version. J.L. performed the experimental work, contributed to data collection and analysis, drafted the primary manuscript, and revised the full text. S.F., J.G. and W.S. participated in data collection, contributed to the experimental work, assisted in writing and

revising the manuscript, and reviewed the final text. All authors have made substantial intellectual contributions to this study, agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved, and approved the final manuscript for publication. All individuals who meet the authorship criteria are included in the author list. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by earmarked fund for the China Agricultural Research System (CARS-34); the Agricultural Science and Technology Independent Innovation Fund, Science and Technology Innovation Guidance Project of Ningxia Hui Autonomous Region (NKYGG-2026-12).

Data Availability Statement: The satellite data used in this study, i.e., Sentinel-2 data, are freely available from the Copernicus Data Space Ecosystem (<https://dataspace.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.starcloud.pcl.ac.cn/>).

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

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.; Zeleke, M.; Kalsa, G.K. Adaptability, forage yield and nutritional quality of alfalfa (*Medicago sativa*) genotypes. *Sustain. Environ.* **2021**, *7*, 1895475. [[CrossRef](#)]
3. Li, X.; Wang, L. Research progress on *Medicago sativa* silage technology. *Acta Prataculturae Sin.* **2005**, *2*, 9–15.
4. Qing, W.; Yang, Z. China's alfalfa market and imports: Development, trends, and potential impacts of the U.S.–China trade dispute and retaliations. *J. Integr. Agric.* **2020**, *19*, 1149–1158.
5. Brown, J.; Jepson, W.; Kastens, J.; Wardlaw, B.; Lomas, J.; Price, K. Multitemporal, Moderate-Spatial-Resolution Remote Sensing of Modern Agricultural Production and Land Modification in the Brazilian Amazon. *Mapp. Sci. Remote Sens.* **2007**, *44*, 117–148.
6. Destro, D.; Carpentieri-Pipolo, V.; Kihl, R.; Almeida, L. Photoperiodism and genetic control of the long juvenile period in soybean: A review. *Crop Breed. Appl. Biotechnol.* **2001**, *1*, 72–92. [[CrossRef](#)]
7. Hugh-Jones, M. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitol. Today* **1989**, *5*, 244–251. [[CrossRef](#)] [[PubMed](#)]
8. Sánchez, N.; González-Piqueras, J.; González-Dugo, M.; Baroncini-Turricchia, G.; Torres, E.; Calera, A.; Pérez-Gutiérrez, C. Water balance at plot scale for soil moisture estimation using vegetation parameters. *Agric. For. Meteorol.* **2012**, *166–167*, 1–9. [[CrossRef](#)]
9. Ran, L.; Lu, X. Delineation of reservoirs using remote sensing and their storage estimate: An example of the Yellow River basin, China. *Hydrol. Process.* **2012**, *26*, 1215–1219.
10. Vikas, K. Identification of Crop Pests using Machine Learning Classifier System. *Int. J. Res. Appl. Sci. Eng. Technol.* **2020**, *8*, 1020–1025. [[CrossRef](#)]
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. *J. Hydrol.* **1998**, *207*, 3–4. [[CrossRef](#)]
12. Tennakoon, S.; Murty, V.; Eiumnoh, A. Estimation of cropped area and grain yield of rice using remote sensing data. *Int. J. Remote Sens.* **1992**, *13*, 427–439. [[CrossRef](#)]
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*, 396–409. [[CrossRef](#)]
14. Kobayashi, N.; Tani, H.; Wang, F.; Sonobe, R. Crop classification using spectral indices derived from Sentinel-2A imagery. *J. Inf. Telecommun.* **2019**, *4*, 67–90. [[CrossRef](#)]
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. *Remote Sens.* **2021**, *13*, 2988. [[CrossRef](#)]
16. Belgii, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [[CrossRef](#)]
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 Sens.* **2022**, *59*, 1534–1549. [[CrossRef](#)]
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. *Comput. Environ. Urban Syst.* **2021**, *182*, 748–762. [[CrossRef](#)]

19. Almeida, C.; Batty, M.; Monteiro, A.; Câmara, G.; Soares-Filho, B.; Cerqueira, G.; Pennachin, C.L. Stochastic cellular automata modeling of urban land use dynamics: Empirical development and estimation. *Comput. Environ. Urban Syst.* **2003**, *27*, 481–509. [\[CrossRef\]](#)
20. Zhang, R.; Wang, S.; Gao, W.; Sun, W.; Wang, J.; Niu, L. Remote-sensing Classification Method of County-level Agricultural Crops Using Time-series NDVI. *Trans. Chin. Soc. Agric. Mach.* **2015**, *46*, 246–252.
21. Mazzia, V.; Khaliq, A.; Chiaberge, M. Improvement in land cover and crop classification based on temporal features learning from Sentinel2 data using recurrent-convolutional neural network (R-CNN). *Appl. Sci.* **2019**, *10*, 238.
22. Bao, X.; Wang, Y.; Feng, Q.; Ge, J.; Hou, M.; Liu, C.; Gao, X.; Liang, T. Spatial distribution extraction of alfalfa based on Sentinel-2 and GF-1 images. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)* **2021**, *37*, 153–160.
23. Mitran, T.; Meena, R.; Chakraborty, A. *Geospatial Technologies for Crops and Soils: An Overview*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–48.
24. 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 Sens. Environ.* **1993**, *46*, 311–326. [\[CrossRef\]](#)
25. De la Casa, A.; Ovando, G. Normalized Difference Vegetation Index (NDVI) and phenological data integration to estimate county yield of corn in Córdoba, Argentina. *Agric. Tec.* **2013**, *34*, 362–371.
26. Lopresti, M.; Bella, C.; Degioanni, A. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. *Inf. Process. Agric.* **2015**, *2*, 73–84. [\[CrossRef\]](#)
27. 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 Sens. Appl. Soc. Environ.* **2021**, *22*, 100485. [\[CrossRef\]](#)
28. 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 J. Photogramm. Remote Sens.* **2015**, *108*, 205–218. [\[CrossRef\]](#)
29. 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*, e0157166. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Huang, J.; Wang, H.; Dai, Q.; Han, D. Analysis of NDVI Data for Crop Identification and Yield Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *7*, 4374–4384. [\[CrossRef\]](#)
31. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google earth engine: Planetary- scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [\[CrossRef\]](#)
32. Kobayashi, H. Geospatial Data Analysis on Google Earth Engine. *J. Remote Sens. Soc. Jpn.* **2018**, *38*, 125–130.
33. He, Y.; Yuan, Y.; Wu, L.; Chen, C.; Zhou, Q.; Chen, Y. Study on fertilization and yield increase of high yield and high-quality alfalfa in the desertification irrigated area of the Hexi corridor. *Arch. Agron. Soil Sci.* **2026**, *72*, 1–16. [\[CrossRef\]](#)
34. Qin, W. The achievement transformation of “demonstration and Application of standardized production Technology of Alfalfa” was carried out smoothly. *Mod. Rural Sci. Technol.* **2015**, *1*, 18–77.
35. Lamb, J.; Jung, H.; Riday, H. Growth environment, harvest management and germplasm impacts on potential ethanol and crude protein yield in alfalfa. *Biomass Bioenergy* **2014**, *63*, 114–125. [\[CrossRef\]](#)
36. 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. *Sci. Bull.* **2019**, *64*, 370–373.
37. Xie, B.; Tang, Y. Europe accelerated the promotion of the “Copernicus”. *Space Explor.* **2015**, *6*, 36–39.
38. He, Y.; Huang, C.; Li, H.; Liu, Q.; Liu, G.; Zhou, Z.; Zhang, C. Land-cover classification of random forest based on Sentinel-2A image feature optimization. *Resour. Sci.* **2019**, *41*, 992–1001.
39. Velastegui-Montoya, A.; Montalván-Burbano, N.; Carrión-Mero, P.; Rivera-Torres, H.; Sadeck, L.; Adami, M. Google Earth Engine: A Global Analysis and Future Trends. *Remote Sens.* **2023**, *15*, 3675. [\[CrossRef\]](#)
40. 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. *Bull. Surv. Mapp.* **2020**, *6*, 111–117.
41. Ma, L.; Gu, X.; Xu, X.; Huang, W.; Jia, J. Remote sensing measurement of corn planting area based on field-data. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)* **2009**, *25*, 147–151.
42. Hou, X.; Di, X.; Hou, W.; Wu, L.; Liu, J.; Wang, J.; Su, H.; Lu, X.; Ying, L.; Yu, X.; et al. Accuracy evaluation of land use mapping using remote sensing techniques in coastal zone of China. *J. Geo-Inf. Sci.* **2018**, *20*, 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. *J. Appl. Remote Sens.* **2017**, *11*, 046023. [\[CrossRef\]](#)
44. Yuan, G.; Zhang, S. Overwintering cutting time and stubble height of alfalfa. *China Rural Sci. Technol.* **2001**, *10*, 14. [\[CrossRef\]](#)

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 Crop. Res.* **2018**, *222*, 189–196. [[CrossRef](#)]
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). *Wirel. Pers. Commun.* **2022**, *126*, 1935–1964.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.