Postprint of Crop Planting Structure Extraction in the Tianshan North Slope Economic Belt Based on [WTHX]NDVI[WTBZ] Time Series Imagery

Authors: YuanKang Xiong, Zhang Qingling, Zhang Qingling

Date: 2019-09-11T00:00:00+00:00

Abstract

Water scarcity is the greatest obstacle to achieving sustainable development in arid regions. Agricultural irrigation in arid regions consumes substantial amounts of water resources, and different crops exhibit significant differences in irrigation water requirements during their growth periods. Therefore, rapidly and accurately understanding the agricultural planting structure in arid regions can provide an important basis for optimizing water-saving agricultural planting structures. Taking the Economic Belt on the Northern Slope of the Tianshan Mountains as the study area, supported by the Google Earth Engine (GEE) cloud platform, and using Sentinel-2 and Landsat 7/8 data as remote sensing data sources, the agricultural planting structure of the study area was extracted through the following steps: First, to streamline the extraction process, a cropland mask layer was constructed using annual maximum NDVI values and slope information; Then, based on the phenological calendar of major crops in the study area, time series data of maximum NDVI values for different periods and the dates when crops reached their maximum NDVI values during the year were obtained, and a 10-band feature image was constructed based on this; Finally, combining valid sample points obtained from field surveys with the 10-band feature image masked by the cropland mask layer, a Random Forest classifier was employed to extract the agricultural planting structure of the study area. Classification results indicate that the overall classification accuracy for cotton, maize, and wheat in the study area in 2018 was 92.19%, with a Kappa coefficient of 0.883. To further compare the classification results with statistical data, the trained classifier was applied to 2017 remote sensing imagery to extract agricultural planting structure information for the study area in 2017. The results show that the planting areas of cotton, maize, and wheat in the study area in 2017 were 5,270 km^2, 2,000 km^2, and 2,340 km^2, respectively, with relative

accuracies of 86.53%, 77.54%, and 86.19%.

Full Text

Preamble

This study presents a method for extracting cropping structure in arid regions using multi-source remote sensing data on the Google Earth Engine (GEE) cloud platform. Sentinel-2 and Landsat 7-8 data were utilized to derive NDVI time series for crop classification. The classification approach achieved an overall accuracy of 92.19% and a Kappa coefficient of 0.883 for the 2018 growing season. Validation against statistical yearbook data for 2017 showed relative accuracies of 86.53%, 77.54%, and 86.19% for cotton, corn, and wheat planted areas, respectively.

1 Study Area and Data

1.1 Study Area

The Northern Tianshan Economic Belt in Xinjiang, China, located between 79°52 ~91°32 E and 43°01 ~46°13 N, was selected as the study area. This region encompasses 17 major river systems and represents a typical oasis agricultural zone where water resources are predominantly used for irrigation. Agricultural development is concentrated in alluvial plains and piedmont zones, with cultivated land accounting for approximately 50% of the total oasis area. The region' s crop cultivation patterns are characterized by distinct phenological calendars for major crops including cotton, corn, and wheat [17-18]. According to the Xinjiang Statistical Yearbook, the planted area proportions of these three crops reached 76.9% of the total cultivated land in 2017 [21].

1.2 Data Sources

1.2.1 Remote Sensing Data Acquisition All remote sensing data were processed on the GEE platform. Sentinel-2 MSI and Landsat 7-8 imagery served as primary data sources, with spectral response functions (SRF) applied to ensure NDVI continuity across sensors. The study period covered March 1 to December 1, 2018. Sentinel-2 data (10 m resolution) were used for their high spatial resolution, while Landsat 7-8 data (30 m resolution) provided temporal continuity. The JavaScript and Python APIs available on GEE facilitated efficient data processing [12]. SRTM digital elevation data were employed to derive slope information for crop mask construction.

1.2.2 NDVI Calculation and Normalization The Normalized Difference Vegetation Index (NDVI) was calculated using the standard formula:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$$

where ρ_{NIR} and ρ_{Red} represent near-infrared and red band reflectance, respectively [34]. To ensure comparability between Landsat and Sentinel-2 NDVI values, linear normalization equations were applied based on spectral response function characterization [36-37]:

$$NDVI_{Landsat-8} = 0.0235 + 0.9723 \times NDVI_{Landsat-7}$$

$$NDVI_{Landsat-8} = -0.0046 + 1.0183 \times NDVI_{Sentinel-2}$$

A threshold of Date > 0.3 was used to identify valid NDVI observations, with Normalize_Date representing the day-of-year for maximum NDVI values.

1.2.3 Crop Mask Development The FMASK algorithm [29] was implemented on GEE for cloud and cloud shadow detection. Pixels with FMASK values indicating clear conditions were retained for analysis. The final crop mask was constructed using two criteria: NDVI Max > 0.4 and Slope $< 7°$ [30], effectively excluding non-agricultural vegetation and steep terrain.

2 Methods

2.1 Crop Mask Construction

The cropland mask was generated by applying thresholds to annual maximum NDVI values and topographic slope data. All pixels satisfying NDVI Max > 0.4 and Slope $< 7°$ were classified as potential cropland. This mask was subsequently applied to all classification procedures to restrict analysis to agricultural areas.

2.2 Feature Extraction

Ten phenological features were extracted from the NDVI time series:

- F1: Max_NDVI_Mar
- F2: Max_NDVI_Apr
- F3: Max_NDVI_May
- F4: Max_NDVI_Jun
- F5: Max_NDVI_Jul
- F6: Max_NDVI_Aug
- F7: Max_NDVI_Sep
- F8: Max_NDVI_Oct
- F9: Max_NDVI_Nov
- F10: Normalize_Date

These features were calculated from monthly maximum NDVI composites spanning March 1 to December 1, 2018, and masked using the cropland mask.

2.3 Random Forest Classification

A random forest classifier was implemented using the CART algorithm [38-41]. Training samples were collected from field surveys conducted in 2017 and high-resolution Google Earth imagery. The classifier was trained on the 10 phenological features to discriminate between cotton, corn, wheat, and other land cover types.

3 Results

3.1 Classification Accuracy

The classification accuracy was assessed using independent validation samples. The overall accuracy reached 92.19% with a Kappa coefficient of 0.883. User' s accuracy (UA) and producer' s accuracy (PA) for individual crops exceeded 85% for cotton and corn, while wheat showed a user' s accuracy of 64.44% due to spectral confusion with other crops.

To further validate the approach, the 2017 cropping structure was extracted and compared with statistical yearbook data. The planted areas for cotton, corn, and wheat were 5270 km^2, 2000 km^2, and 2340 km^2, respectively. The relative accuracies compared to official statistics were 86.53% for cotton, 77.54% for corn, and 86.19% for wheat.

3.2 Spatial Distribution of Crops

The spatial distribution of crops in 2018 revealed distinct patterns: cotton was predominantly cultivated in alluvial plains, corn in piedmont zones, and wheat in areas with better water availability. The classification captured the typical oasis agriculture mosaic characteristic of the region.

[TABLE 1]
[TABLE 2]
[FIGURE 1]
[FIGURE 2]
[FIGURE 3]
[FIGURE 4]
[FIGURE 5]

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.