https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Earth System Science Data Discussions
Open Access

ChinaSoyArea10m: a dataset of soybean planting areas with a spatial resolution of 10 m across China from 2017 to 2021

Qinghang Mei[1], Zhao Zhang[1], Jichong Han[1,2], Jie Song[1,2], Jinwei Dong[3,4], Huaqing Wu[1], Jialu Xu[1], Fulu Tao[3,4]

[1] School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875 / Zhuhai 519087, People's Republic of China
[2] School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China
[3] Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, People's Republic of China
[4] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

Correspondence to: Zhao Zhang (zhangzhao@bnu.edu.cn)

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Abstract

Soybean, an essential food crop, has witnessed a steady rise in demand in recent years. There is a lack of high-resolution annual maps depicting soybean planting areas in China, despite China being the world's largest consumer and fourth largest producer of soybeans. To address this gap, we developed a novel method called phenological- and pixel-based soybean area mapping (PPS) based on Sentinel-2 remote sensing images from the Google Earth Engine (GEE) platform. We utilized various auxiliary data (e.g., cropland layer, detailed phenology observations) to select the distinct features that differentiate soybeans most effectively from other crops across various regions. These features were then input for an unsupervised classifier (K-means), and the most likely type was determined by a post-classification method based on dynamic time warping (DTW). For the first time, we generated a dataset of soybean planting areas across China, with a high spatial resolution of 10 meters, spanning from 2017 to 2021 (ChinaSoyArea10m). The R^2 values between the mapping results and the census data at both county- and prefecture-level were consistently around 0.85 in 2017-2020. Moreover, the overall accuracy of mapping results at the field level in 2017, 2018, and 2019 were 77%, 84% and 88%, respectively. Compared with the existing 10-m crop-type maps in Northeast China (Cropland Data Layer, CDL) based on field samples and supervised classification methods, the mapping accuracy is significantly improved by 31% (R^2 increases from 0.53 to 0.84) according to their consistency with census data, particularly at the county level. ChinaSoyArea10m is spatially consistent well with the two existing datasets (CDL and GLAD maize-soybean map). ChinaSoyArea10m provides important information for sustainable soybean production and management, as well as agricultural system modeling and optimization. ChinaSoyArea10m can be downloaded from an open-data repository (DOI: https://zenodo.org/doi/10.5281/zenodo.10071426, Mei et al., 2023).

1 Introduction

Soybean, one of the most important crops around the world, plays an important role in diet and livestock breeding (Hartman et al., 2011). As the global demand for protein and meat increases, China's demand for soybeans has been keeping rising nowadays. In the past decade, China has averagely accounted for

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

over 30% of the world's total soybean consumption (Liu and Fan, 2021). Despite being the fourth-largest producer of soybeans after Brazil, the United States, and Argentina, China's self-sufficiency rate is low (FAOSTAT, 2023; Wang et al., 2023). Given the rapid growth of demand and the shortcomings of domestic supply, mapping soybean planting areas across China is crucial for sustainable soybean production and management.

Soybean planting area in some regions of China was mapped in previous studies, but long-term soybean maps over all major producing areas in China have not been available. A decision tree method based on phenological and near-infrared reflectance differences was applied in the state of Parana in Brazil to produce corn-soybean maps with a resolution of 500 m (Zhong et al., 2016). However, this study was limited to one state and a simple planting pattern (including soybeans and corn only) at a medium resolution. The field size in China is generally small, and 500 m-resolution maps will inevitably bring pixel mixing problem (Lowder et al., 2016). More recently, 20-year soybean-corn maps with 30 m resolution have been generated by collecting a large number of samples and using green chlorophyll vegetation index (GCVI) time series features, which is a large-scale, high-precision soybean mapping attempt (Wang et al., 2020). Similarly, high-precision soybean maps in China were also made by collecting major crop samples and utilizing spectral reflectance and vegetation index characteristics, for 2017-2019 in Northeast China (You et al., 2021). Some studies have also utilized unique canopy water content and chlorophyll content to produce soybean maps in the same areas from 2017 to 2021 (Huang et al., 2022). A particularly encouraging attempt was made to develop a national maize-soybean map for 2019 in China by combining field sample data and regression estimators (Li et al., 2023). However, these studies focused only on the limited areas or a single year despite of good attempts for accurately mapping soybean areas. Long-term annual soybean maps over mainly planting areas in China with a higher spatial resolution have not been available so far.

Mapping crops by remote sensing can generally be categorized into five methods : (1) Supervision classification based on a large number of field samples (You et al., 2021; Shangguan et al., 2022); (2) Decision tree method based on prior knowledge and appropriate classification rules (Zhong et al., 2016); (3) Developing some composite indexes and using threshold segmentation (Huang et al., 2022; Chen et al., 2023); (4) Matching satellite-based crop classification with sample-based area estimation (Song et al., 2017; Li et al., 2023); (5) Combining unsupervised classification and post-classification methods

(Wang et al., 2019; You et al., 2023). The methods (1) and (4) are both relied on ground samples, while the methods (2) and (3) are both based on thresholds. However, mapping soybean by these methods was mainly applied in small areas, and maps over a larger region are very few. The methods (1) and (4) based on sufficient field samples over larger region are relatively mature and can achieve maps with a higher accuracy. However, collecting sufficient field samples for consecutive years in a large region is extremely time, money, and labor costly, and unsuitable for long-term years and over larger areas (Luo et al., 2022). Moreover, the threshold-based methods (2) and (3) have been applied into large areas, but determining the thresholds will inevitably bring large uncertainty, especially for the areas with high heterogeneity in climate, environment, and planting patterns. As for mapping soybean, it is a big challenge due to their similar growth characteristics with many other summer crops. The thresholds that work well in some areas are not necessary to work well in other areas (Graesser and Ramankutty, 2017; Guo et al., 2018). These limitations of the methods restrict the availability of accurate soybean maps, especially over large regions in China.

Along this line, the adaptive classification approach tailored to distinct areas, i.e., method (5), is a highly effective for accurately mapping crops over a larger region. Such unsupervised classification can effectively address the above issues such as insufficient samples and limited spatial scalability by training classifiers separately in different areas (Ma et al., 2020; Wang et al., 2022). Remarkable successes have been achieved when applying the approach into the United States in mapping soybean and maize, substantiated by Wang et al. (Wang et al., 2019). Due to the different climatic and environmental conditions, together with huge differences in cultivating patterns over different areas, crop phenological information has become an important reference for crop classification. For example, the phenological observations at the agricultural meteorological stations were employed as a reference to identify the critical phenological dates of pixels, thereby generating planting areas for three major crops in China (Luo et al., 2020). The dynamic weighting approach based on the similarity of phenological curves of Normalized Difference Vegetation Index (NDVI) has successfully estimated the planting area of maize in China (Shen et al., 2022). Phenological-based Vertical transmit Horizontal receive (VH) polarized time series were used for an unsupervised classifier to map the seasonal soybeans (Kumari et al., 2019). By fully using phenological information in different areas, local classifiers can be well trained to solve the spatial generalization limitation in mapping crops over a larger region.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

The main objectives of this study are: 1) to develop a novel framework to map soybean planting area over a larger region; 2) to test the generalization ability of the framework and assess the accuracy of maps at different levels; and 3) to provide a new data product of soybean planting area across mainly planting areas in China, for multi-years with a high spatial resolution.

2 Materials and methods

2.1 Study area

We selected 14 major soybean producing provinces (including Chongqing Municipality) as study area, which cover over 90% of the total planting area in China (National Bureau of Statistics of China, 2023) (Fig. 1). The soybean planting areas were classified into four agro-ecological zones (AEZs) based on their diverse geographical environment and planting habits, including Northeast single cropping eco-region (NE, Zone I), Huang-Huai-Hai double cropping eco-region (HH, Zone II), Middle-Lower Yangtze River double cropping eco-region (MLY, Zone III) and Southwest double cropping eco-region (SW, Zone IV) (Wang and Gai, 2002). In particular, Zone I and Zone II are the main soybean producer in China, accounting for more than 70% of the national soybean planting area.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

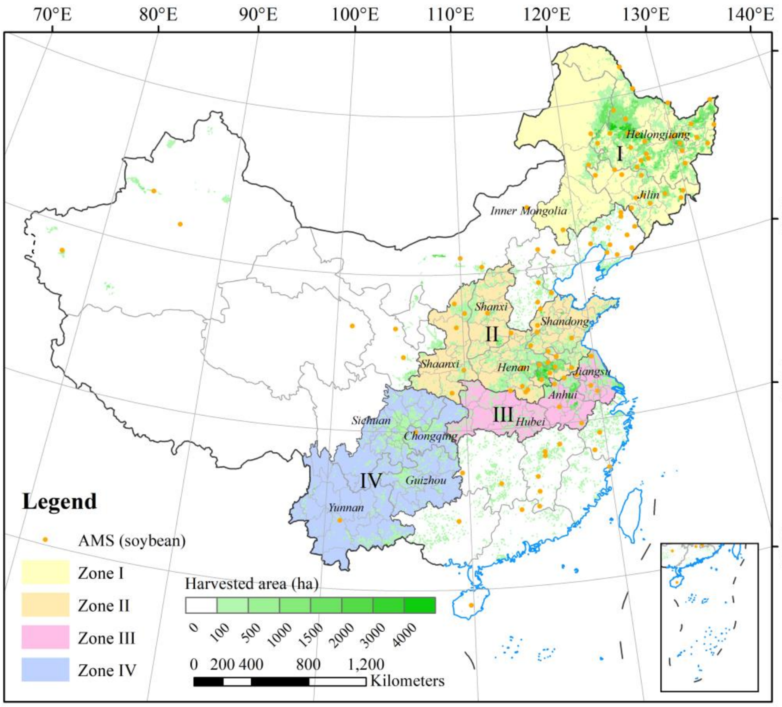


Figure 1. The study area including 14 provinces (including Chongqing Municipality) in China. The 14 provinces include Heilongjiang, eastern Inner Mongolia, Anhui, Henan, eastern Sichuan, Jilin, Hubei, Guizhou, Jiangsu, Yunnan, Shandong, Shaanxi, Shanxi, and Chongqing. The dots represent the location of soybean agricultural meteorological stations (AMSs). The harvested area was obtained from SPAM2010 0.5° ×0.5° grid data (Yu et al., 2020).

2.2 Data

2.2.1 Remote sensing data

We used Sentinel-2A/B Multi-Spectral Instrument (MSI) Level-1C top-of-atmosphere reflectance data during 2017-2021 (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2, last access: September 2023). Sentinel-2 sensors provide observations in 13 spectral bands at 10 m or 20 m resolution. In addition to the traditional bands (i.e., the visible and near-infrared bands), the red-edge bands and shortwave infrared bands equipped with sentinel-2 play a great role in enhancing the accuracy

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

of crop classification (Luo et al., 2021; Marshall et al., 2022). In addition, the S2 cloud probability provided by the official can identify cloud pollution areas and be used as cloud removal processing.

2.2.2 In-situ phenological observations

The soybean phenology observations from 2017 to 2020 were obtained from 115 agricultural meteorological stations (AMSs) governed by the CMA (https://data.cma.cn/, last access: May 2022). Phenology information of each AMS is observed on alternate days or once a day, and key phenological events such as sowing, emergence, three-true-leaves, branching, flowering, podding, full-seeding, and maturity are noted by technicians to ensure accuracy. We defined the period from sowing to flowering as the vegetative growth period (VGP), and the period from flowering to maturity as the reproductive growth period (RGP) of soybeans (Gong et al., 2021). Additionally, we used the average of observed dates in adjacent years to fill gaps for the missing data.

2.2.3 Cropland data

GLAD cropland product with a 30-m resolution in China was used as cropland masks (https://glad.umd.edu/dataset/croplands, last access: September 2023) (Potapov et al., 2022). The crop layer was conducted every four years from 2000 to 2019. We used the file for the 2016-2019 interval which is closest to the study years. GLAD's overall accuracy of pixel-wise validation is 0.88 in China, consistent well with the census data. The accuracy of the product is higher than that of similar products, making it a reliable for crop mapping (Zhang et al., 2022).

2.2.4 Census data and ground samples

To determine the number of clusters at prefecture-level and validate the accuracy of the soybean maps at county (2017-2018) or prefecture (2019-2020) level, we utilized agricultural census data obtained from the National Bureau of Statistics of China (http://www.stats.gov.cn/, last accessed: June 2023). To assess the reliability of the soybean maps, we collected ground samples from field surveys from 2017 to 2019. Since the soybean planting area in Heilongjiang (HLJ), Inner Mongolia (NMG), Anhui (AH), Henan (HN), and Jilin (JL) accounted for more than 70% of the country's total area, we used ground samples from these five provinces (Table 1). Crop types (soybean, maize, rice, wheat, others) and other land cover

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Open Access Earth System Science Data Discussions

types were recorded. To ensure the impartiality of verification results, we only selected crop samples for validation.

Table 1. Summary of ground samples for validation.

		HLJ	NMG	AH	HN	JL
2017	Soybean	1013	451	-	-	0
	Maize	1061	146	-	-	11
	Rice	513	38	-	-	13
	Other crops	124	459	-	-	0
2018	Soybean	525	746	72	15	117
	Maize	764	479	73	20	217
	Rice	587	42	0	0	71
	Wheat	10	141	0	0	0
	Other crops	70	1069	0	0	0
2019	Soybean	901	562	51	-	26
	Maize	468	463	53	-	197
	Rice	392	36	0	-	148
	Other crops	62	445	0	-	36

2.2.5 Existing products

We utilized the crop map CDL of Northeast China from 2017 to 2019 (You et al., 2020) for consistency comparison with census data, and the 2019 GLAD maize-soybean map (https://glad.earthengine.app/view/china-crop-map, last access: September 2023) as a reference for spatial detail comparison with CDL. CDL is a 10m resolution crop map dataset of Northeast China from 2017 to 2019 that was created using Sentinel-2 key spectral bands and vegetation indices, multi-year field samples, and a random forest classifier (You et al., 2021). The maps include three crop types: rice, maize, and soybeans. The GLAD maize-soybean Map is a national classification map for 2019 that was produced using random forests, based on field surveys and area estimates (Li et al., 2023). The agreement (R^2) between GLAD and the statistics is higher than 0.9, and the overall mapping accuracy is greater than 90%, making it a reliable reference for comparing spatial details. We extracted the soybean layers from all the existing products.

2.3 Methods

Mapping soybean consists of three main steps (Fig.2): data processing, soybean mapping, and accuracy assessment. It is important to note that the soybean mapping strategy involves several key steps, including

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

potential area identification, feature selection, unsupervised learning, and cluster assignment. Specifically, by exploring the spectral characteristics of crop field samples, we firstly identified reflectance bands and vegetation indices that are significantly associated with soybeans but different from other crops. We then utilized the detailed phenological records at AMSs and selected the key phenological information as input features. Unsupervised classifiers were trained separately in each prefecture administrative area, and the category most likely to represent soybeans was determined through a DTW-based post-classification method. We are able to map the soybean planting areas across mainly planting areas in China with a spatial resolution of 10 m during the studied period. Finally, we conducted multi-comparisons between our soybean products with others, including census data, ground samples, and existing datasets, to evaluate the accuracy of our data product.

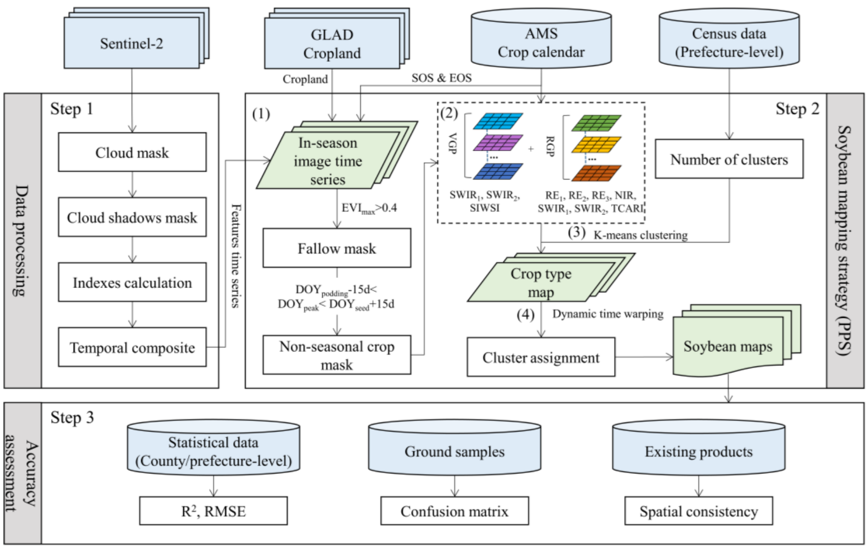


Figure 2. The methodology for mapping soybean planting area. AMS, agricultural meteorological station; $DOY_{podding}$, the podding date recorded by the nearest AMS; EVI: Enhanced Vegetation Index; DOY_{peak}, the date when EVI reached peak; DOY_{seed}, the full-seed date recorded by the nearest AMS; SOS, start of growing season; EOS, end of growing season; $SWIR_1$, Short Wave Infrared band 1; $SWIR_2$, Short Wave Infrared band 2; SIWSI, shortwave Infrared Water Stress Index; RE_1, Red Edge band 1; RE_2, Red Edge band 2; RE_3, Red Edge band 3; NIR, Near-infrared band; TCARI, Transformed Chlorophyll Absorption in Reflectance Index; VGP: vegetative growing period; RGP: reproductive growing season.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

2.3.1 Data processing

We employed the simple cloud score algorithm (Oreopoulos et al., 2011), QA60 band, cirrus band, and cloud probability dataset to identify cloud masks. The following isolated cloud masks are created: (1) Cloud and cirrus identified by QA60 band; (2) Cirrus identified by cirrus band in Level-1C products; (3) Pixels with cloud score less than 0.9; and (4) Pixels with cloud probability more than 70. Each algorithm has its own strengths and limitations. For example, QA60 band removes a large number of thin cirrus clouds while ignoring small clouds with thicker resolution, and the fixed threshold values of cloud score and cloud probability may introduce uncertainties. Therefore, we masked the pixels identified as clouds by at least two methods to achieve better cloud removal effects. Then, we used Temporal Dark Outlier Mask (TDOM) method to eliminate cloud shadows (Housman et al., 2018). We calculated the SIWSI and TCARI indices based on the Sentinel-2 image set processed above (see 2.3.2(2)). To avoid noise in image time series, Sentinel-2 time series was reconstructed by moving median composite method, resulting in a 10-day interval composite time series.

2.3.2 Phenological- and Pixel-based Soybean mapping strategy (PPS)

(1) Potential area identification

To minimize the impact from no-croplands, we firstly determine the potential cropping areas by masking GLAD cropland layer over study area. By obtaining the starting and ending dates of the growing season at the nearest agricultural meteorological station (AMS), we then extracted Sentinel-2 images within the growing season. Based on the cropland extracted, we filtered the pixels exhibiting an EVI maximum value during the growing season greater than 0.4 to remove fallow land. EVI is a vegetation index with high sensitivity in biomass:

$$EVI = G \times \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + C_1 \times \rho_{Red} - C_2 \times \rho_{Blue} + L} \quad (1)$$

Where ρ_{NIR}, ρ_{Red}, and ρ_{Blue} represented the reflectance of the Near-infrared (835.1nm (S2A) / 833nm (S2B)), Red (664.5nm (S2A) / 665nm (S2B)), Blue (496.6nm (S2A) / 492.1nm (S2B)), respectively.

The greenest period of soybean typically occurs between the podding date and the full-seed date, with a difference of more than a month from the peak date of non-seasonal crops, such as wheat (Fig. 3a). We obtained the phenological observations recorded by the nearest AMS as reference and set the restricted

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

time window from 15 days before the podding date ($DOY_{podding}$) to 15 days after the full-seed date (DOY_{seed}). We generated the potential area by eliminating pixels whose EVI maximum occurs outside the given time window because the phenological difference of soybeans in adjacent areas generally does not exceed one month.

(2) Feature selection

We selected six bands and two spectral indices for crop mapping, including Near-infrared (NIR) band, Red edge band 1 (RE1), Red edge band 2 (RE2), Red edge band 3 (RE3), Short Wave Infrared band 1 (SWIR1), Short Wave Infrared band 2 (SWIR2), Shortwave Infrared Water Stress Index (SIWSI), Transformed Chlorophyll Absorption in Reflectance Index (TCARI). SIWSI is an indicator of canopy water content that reflects soil moisture variations and canopy water stress better than Normalized Difference Vegetation Index (NDVI) (Fensholt and Sandholt, 2003; Olsen et al., 2015). TCARI is an indicator which is sensitive to chlorophyll concentration (Sobejano-Paz et al., 2020). The two spectral indices were calculated as follows:

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

$$TCARI = 3 \times ((\rho_{VRE1} - \rho_{Red}) - 0.2 \times (\rho_{VRE1} - \rho_{Green}) \times \rho_{VRE1}/\rho_{Red}) \tag{3}$$

Where ρ_{SWIR1}, ρ_{NIR}, ρ_{VRE1}, ρ_{Red} and ρ_{Green} represented the reflectance of the Short Wave Infrared band1 (SWIR1, 1613.7nm (S2A) / 1610.4nm (S2B)), Near-infrared (835.1nm (S2A) / 833nm (S2B)), Red Edge1 (VRE1, 703.9nm (S2A) / 703.8nm (S2B)), Red (664.5nm (S2A) / 665nm (S2B)), Green (560nm (S2A) / 559nm (S2B)), respectively.

During early growing season of soybean (~DOY 120-190), the flooding signal of rice was obvious due to the transplanting period. This resulted in a significantly lower SWIR reflectance and SIWSI index for rice compared to those of soybean (Fig 3f-h). SWIR bands and SIWSI index during the vegetative growing period (VGP) of soybean can effectively distinguish dryland crops (such as soybean, maize) from paddy crops (such as rice).

Soybean has a lower water content during the middle and later growing season (~DOY 190-220) than maize, resulting in higher reflectivity in SWIR bands (Fig. 3b, 3f, 3g) (Chen et al., 2005). It has been demonstrated that SWIR and red-edge bands can effectively differentiate soybean and maize (Zhong et al., 2016; You and Dong, 2020; Liu et al., 2018). Additionally, the chlorophyll content of soybean in the middle and late growth period was lower than that of maize, leading to significantly higher TCARI values.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Open Access Earth System Science Data Discussions

Meanwhile, the date at which TCARI reaches saturation for soybean, rice, and wheat varies greatly (Fig 3i). Based on these findings, we selected NIR, red-edge bands (RE2 and RE3), short-wave infrared bands, and TCARI index during soybean reproductive growing season (RGP) as key features.

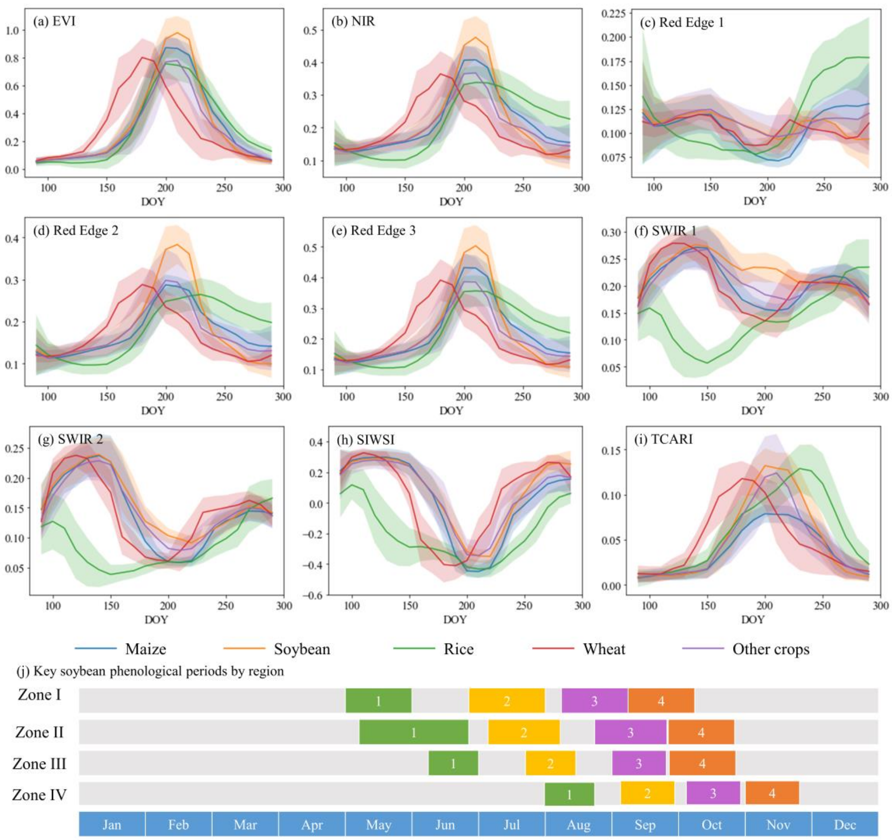


Figure 3. Temporal profiles of (a-i) for major crops in Northeast China and (j) key soybean phenological periods by region based on ground samples. Lines depict the mean values of different crops and shaded areas depict error bars with one positive/negative standard deviation. The number at the bottom represents the key phenological periods of soybean: 1 – Sowing, 2 – Flowering, 3 – Seed fulling, 4 – Maturity.

(3) Unsupervised learning

We utilized K-means algorithm to group potential area data by using the wekaKMeans Clusterer provided by Google Earth Engine (GEE). The *m* samples are divided into *k* clusters by alternately assigning the samples to the nearest cluster centroid measured by Euclidean distance or the Manhattan distance and updating the cluster centroid to the mean of the samples assigned to the cluster. This approach had been

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

widely used in land-cover classification and crop mapping (Xiong et al., 2017; Wang et al., 2019). The classifier was trained individually on each prefecture. According to the prefecture-level statistics, we determined the number of major crop types planted in the same season as soybean in each prefecture. The number of types was taken as the number of clusters for the classifier.

(4) Cluster assignment

To identify the most likely cluster that represents soybean, we randomly selected 100 points from each cluster and extracted features of key phenological dates. We then used dynamic time warping (DTW) method to calculate the distance between the features of each cluster and soybean ground samples. The cluster closest to the samples was identified as the soybean cluster. DTW is a flexible algorithm that allows for deviations in time between two sequences, and it calculates the minimum distance between them by finding misalignment matches between elements. To further reduce the possibility of misclassification due to regional differences in phenology, we only selected features corresponding to the reference phenological periods. This approach is widely used in land cover and crop identification due to its ability to handle time distortions associated with seasonal changes (Guan et al., 2016; Dong et al., 2020).

2.3.3 Accuracy assessment

To assess the accuracy of the soybean maps we generated, we validated and compared the results using 1) county- and prefecture-level census data, 2) ground samples, and 3) existing products. Since the county-level statistics after 2019 were not fully collected, we used the county-level statistics for 2017-2018 and the prefecture-level statistics for 2019-2020 to calculate the R^2 and RMSE of the mapped area with the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(s_i - y_i)^2}{\sum_{i=1}^{n}(s_i - \bar{s})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(s_i - y_i)^2}{n}} \tag{5}$$

where s_i and y_i are the statistical and mapped soybean area for county (prefecture) i, $\bar{s}$ is the average statistical area, and n represents the total number of counties (prefectures). We calculate the local crop mapping area based on the Universal Transverse Mercator (UTM) projection corresponding to the location of the province.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

We also used ground samples during 2017-2019 to verify the authenticity of the soybean maps. Confusion matrices were calculated as follows:

$$PA = \frac{N_i}{R_i} \tag{6}$$

$$UA = \frac{N_i}{C_i} \tag{7}$$

$$OA = \frac{N_c}{A} \tag{8}$$

$$F1 = 2 \times \frac{UA \times PA}{UA + PA} \tag{9}$$

where N_i is the number of correctly identified validation samples of class i, R_i is the number of ground validation samples of class i, C_i is the number of validation samples classified as class i, C_i is the number of validation samples classified as class i, N_c is the total number of correctly identified validation samples, A is the total number validation samples. PA, UA, and OA represent producer's accuracy, user's accuracy, and overall accuracy, respectively.

To ensure that the products are accurate not only in quantity but also in space, we further compared the ChinaSoyArea10m with existing products in detail space.

3 Results

3.1 Accuracy assessment

We utilized the available census data from 2017-2020 (at county-level in 2017-2018 and prefecture-level in 2019-2020) to verify the accuracy of the soybean maps across the entire studied area. Annual ChinaSoyArea10m is consistent well with the census data ($R^2 > 0.8$), with an R^2 value of 0.84, 0.85, 0.82, and 0.86 for 2017, 2018, 2019, and 2020, respectively (Fig.4). These results demonstrate that our PPS method is inter-annual robustness and can accurately capture annual dynamics of soybean planting areas. The scattered points are generally distributed around 1:1 line, without large overestimations or underestimations. However, the areas are overestimated for counties with planting area < 20 kha, or prefectures with planting area <100 kha (Fig. 4).

ChinaSoyArea10m is consistent well with census data compared to the existing product (CDL) (You et al., 2021), using both the county level in 2018 and prefecture level in 2019 (Fig. 5). CDL's results are consistent with census data at the prefecture scale, with more overestimations at the county level (Fig.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Earth System Science Data
Open Access
Discussions

5), implying the comparison at finer resolution would reveal more details. ChinaSoyArea10m is consistent with statistics at the both levels (R^2 ~0.85), while R^2 decreases by 30 % for CDL in county level (Fig. 5a).

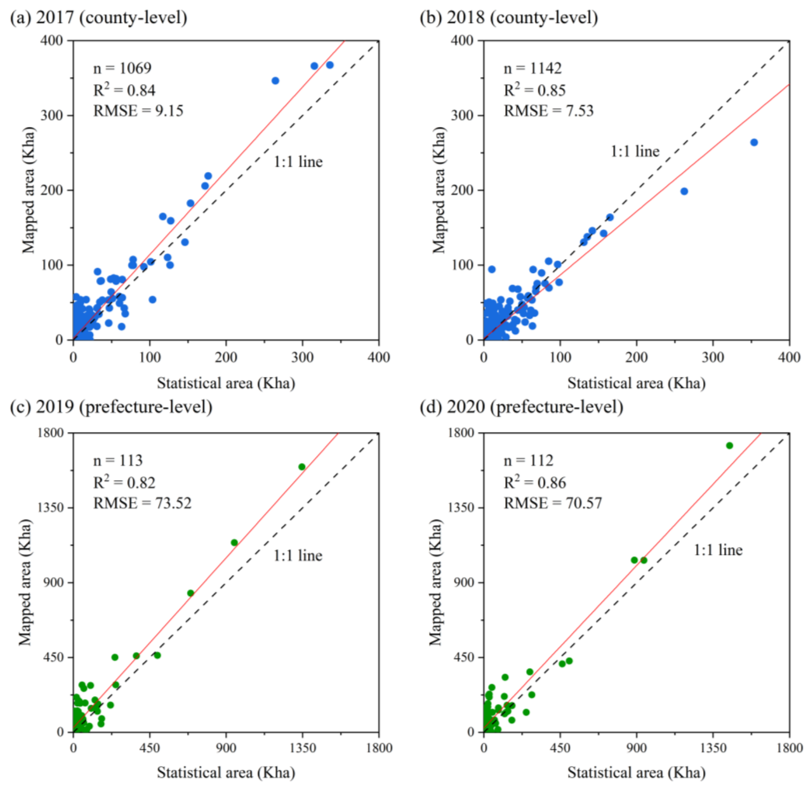
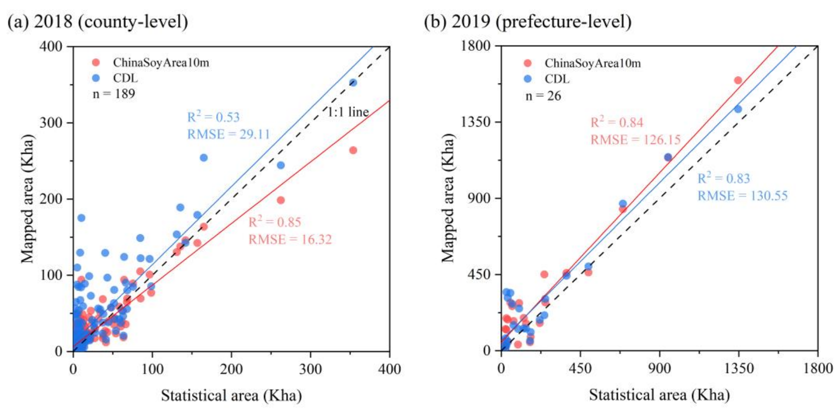


Figure 4. Comparison of soybean areas with statistics in (a) 2017 at county-level, (b) 2018 at county-level, (c) 2019 at prefecture-level, (d) 2020 at prefecture-level.



https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Figure 5. Comparison of soybean areas of ChinaSoyArea10m and CDL with statistics in (a) 2018 at county-level, (b) 2019 at prefecture-level.

Furthermore, we used ground samples in 2017-2019 to validate the reliability of the soybean maps. Since the soybean planting area maps are 0-1 binary images, we categorized the ground samples into soybean and non-soybean (maize, rice, wheat, and other crops). The verification results based on ground samples indicated that the overall accuracy of soybean maps during 2017-2019 was in the range of 0.77 to 0.88. From 2017 to 2019, the F1 scores of soybeans became higher and higher (0.68, 0.74 and 0.85, respectively) (Table 2). The variance in accuracy among years could be attributed to the quality of Sentinel-2 images, which had been indicated in previous studies (Liu et al., 2020; Han et al., 2021).

Table 2. Confusion matrix of the soybean maps during 2017-2019.

Year	Reference	Map		Producer's	User's	F1	Overall
		Soybean	Non-Soybean	Accuracy	Accuracy	Score	Accuracy
2017	Soybean	960	379	65.57%	71.70%	0.68	76.94%
	Non-Soybean	504	1986	83.97%	79.76%	0.82	
2018	Soybean	1112	426	75.39%	72.30%	0.74	84.28%
	Non-Soybean	363	3117	87.98%	89.57%	0.89	
2019	Soybean	1365	303	88.64%	81.83%	0.85	87.55%
	Non-Soybean	175	1997	86.83%	91.94%	0.89	

3.2 Spatial distributions of soybean planting areas

Based on the soybean maps, we further analyzed the spatial patterns of soybean distribution in China during 2017-2021. There were small changes in the spatial distribution of soybean in China in recent years (Fig.6-7). Several hot spots were obviously observed in Heilongjiang Province, eastern Inner Mongolia, and northern Anhui, especially for eastern Inner Mongolia and western Heilongjiang, extensively and densely distributed by soybean fields (Fig.7b-c). In Region II, soybean was planted at a larger scale, mainly concentrated in northern Anhui (Fig.7d), and extensively distributed in Henan and Shandong (Fig.7e). Soybeans in other provinces of Region II, III, and IV were scattered distribution, especially in the southwestern mountainous region (Fig.7f-h).

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

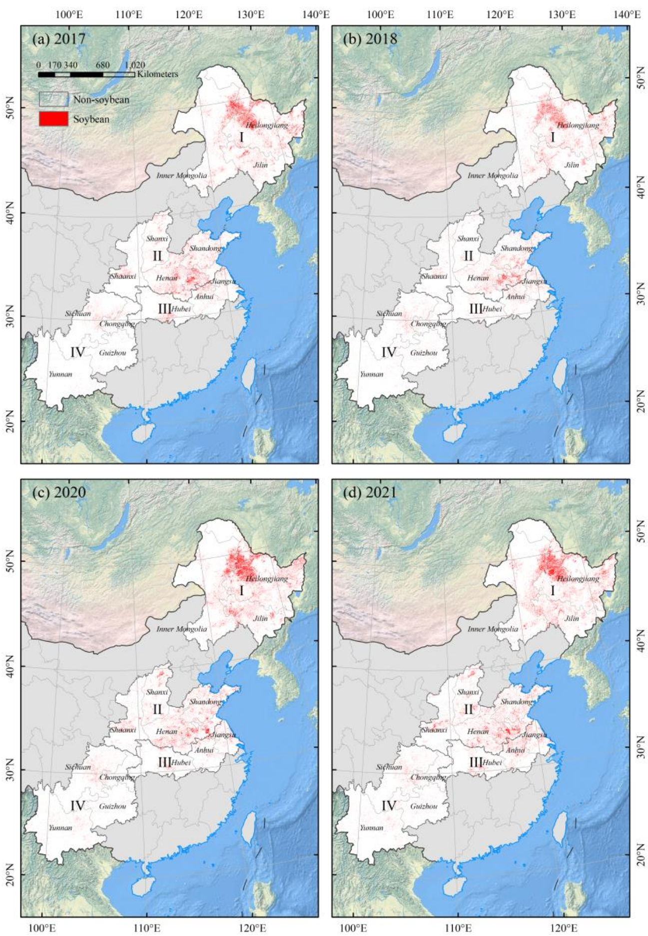


Figure 6. Spatial distribution of soybean areas at 10 m resolution across China in (a) 2017, (b) 2018, (c) 2020 and (d) 2021.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

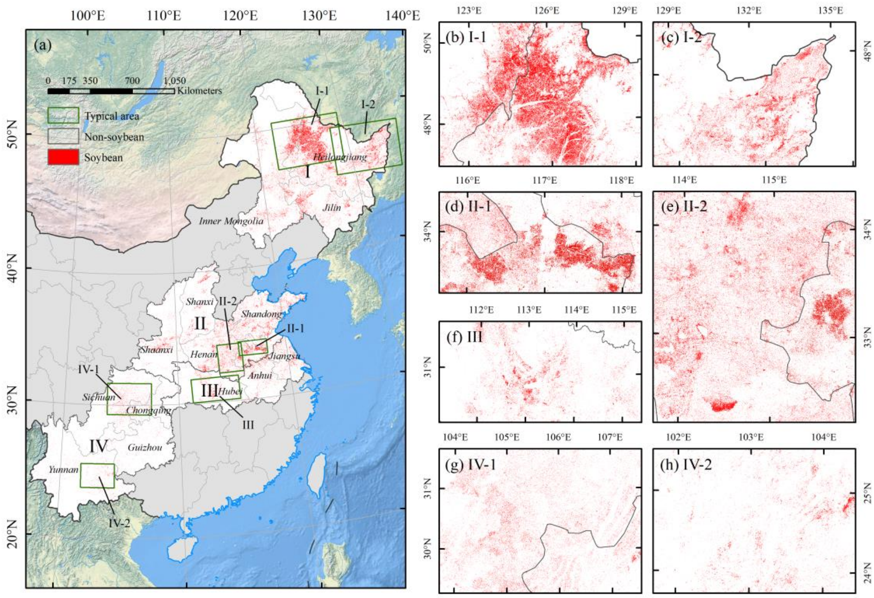


Figure 7. Spatial distribution of soybean areas at 10 m resolution across China (a) and zoom-in maps of each region (b-h) in 2019.

To further compare soybean maps in detail, we compared ChinaSoyArea10m with GLAD maize-soybean map and CDL data products in space. The GLAD product is a 10-m resolution maize-soybean map of China in 2019, and their R^2 values with provincial and prefecture statistics were reported by 0.93 and 0.94 (Li et al., 2023). Arable land near waterbodies is often misclassified as soybean plots by CDL, which has not occurred by GLAD and ChinaSoyArea10m, implying other crop types are possibly misclassified as soybeans by CDL (Fig.8 a1-d1). As for the second case (Fig.8 a2), our extraction results are similar to those of GLAD, while small plots failed to be identified by CDL (Fig.8 a2-d2). In areas where banded soybeans are planted less concentrated, CDL tended to overestimate the soybean area (Fig.8 a3-d3), further substantiating the above limitations (Fig.5). Conversely, our mapping results behaved similarly as GLAD did (Fig.8 a3-d3). The overall accuracy of GLAD map based on pure samples reaches 95.4% (Li et al., 2023), so GLAD can be regarded as a reliable reference. From the three cases, therefore, ChinaSoyArea10m has behaved more similarly with GLAD than CDL does, indicated by less underestimation, less overestimation, and higher accuracy in details.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

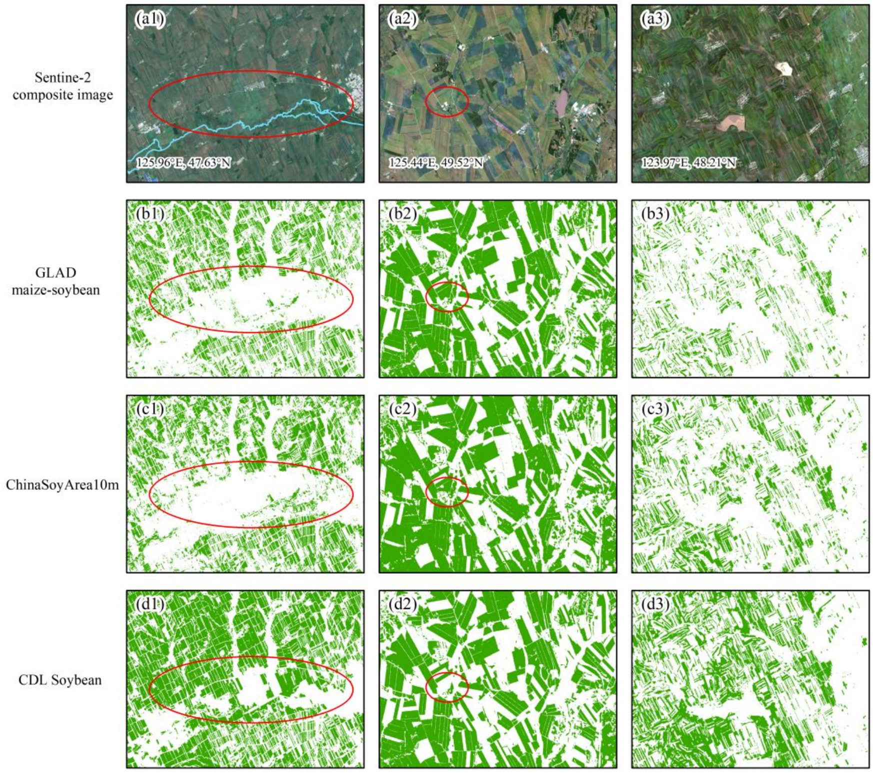


Figure 8. Visual comparison of our soybean maps and existing products in typical regions in 2019: (a1-a3) RGB composite images comprise red (Band 4), green (Band 3), and blue (Band 2) bands from Sentinel-2 median composite images during the middle and late growth of soybean; (b1-b3) soybean layer extracted from GLAD maize-soybean map; (c1-c3) ChinaSoyArea10m map; (d1-d3) soybean layer extracted from CDL.

4 Discussion

We proposed a new framework (PPS) to identify annual dynamic of soybean planting areas over larger regions. We produced firstly the longer-term series of soybean maps (ChinaSoyArea10m) across mainly planting areas in China from 2017 to 2021. The accuracy of ChinaSoyArea10m is acceptable (R^2 ~ 0.85) at both county- and prefecture-level, relatively less than GLAD (R^2 = 0.93 at prefecture-level), but higher than CDL (R^2 = 0.53 at county-level). The PPS proposed does not require quantities of field samples and can self-adopt to different environments by considering phenology information. However, there are still some limitations in our study.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Open Access Earth System Science Data Discussions

4.1 The uncertainty from image quality

The method we proposed (PPS) is strongly dependent on remote sensing images and subregional unsupervised classification by considering the bands and vegetation indices, which are all sensitive to the unique characteristics of soybeans. Therefore, the accuracy of soybean maps inevitably is associated with the quality of remote sensing images. By using ground samples to validate the mapping results, we found that the accuracy of 2017 is lower than that of 2018 and 2019, with an overall accuracy is less than 0.8 (Table 2).

We extracted cloud-free images in different regions during the soybean growing season and calculated the monthly average number of clear observations. In general, the monthly averages of clear observations in Northeast region and Huang-Huai-Hai region (Zone I and Zone II) are relatively higher than the southern zones (Zone III and IV) (Fig.9a2-e2). Obviously, the total number of images available in 2017 over the study areas was significantly fewer than those of other years (Fig.9a1-e1). Removing the cloudy pixels has left ever fewer clear images available (upper vs. down layer in Fig.9). Even in the northeast region during the growing season, the average number of clear observations per month was 1-2, lower than the requirements of 10-day time series composite we mentioned in 2.3.1. This might explain the low accuracy based on sample verification in 2017 (Table 2).

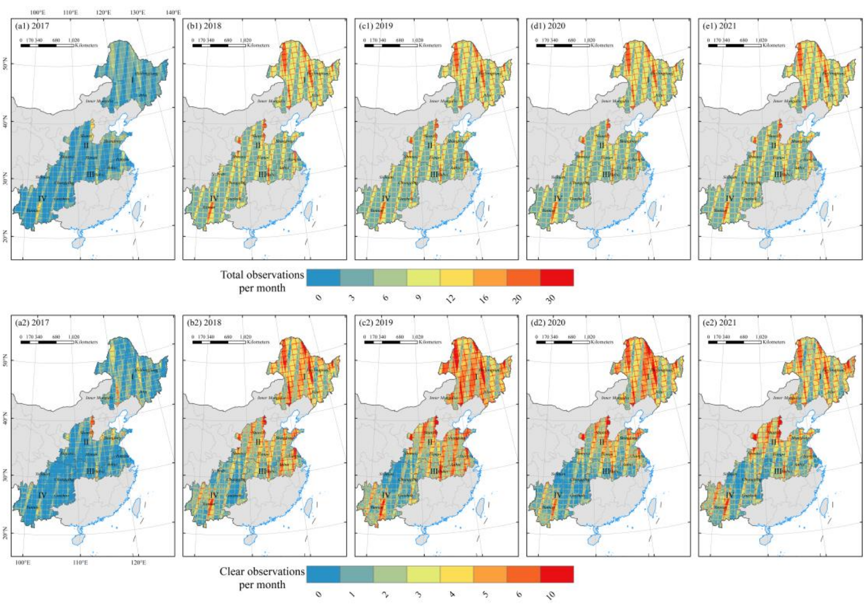


Figure 9. Total (a1-e1) and clear (a2-e2) observations per month during soybean growing season.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

4.2 Limitations in small-scale planting areas

Validation based on statistical results shows that ChinaSoyArea10m reached a high consistency (R^2 ~ 0.85) across China. However, in areas with soybean sparsely planted, the consistency is lower than that in densely planted areas, with more overestimations observed in the sparse areas. Such overestimations are caused by the limitations of unsupervised classification. Unsupervised classifier is difficult to accurately capture small plots of crops in a complex cropping system, although it can make up for the shortage of crop mapping in some areas with limited training samples (Kwak and Park, 2022). Studies have proved that the classifier performs inferiorly where dominant crop phenotypes are similar, and crop diversity is higher (Wang et al., 2019; Konduri et al., 2020). Therefore, the classifier is challenged in areas where soybean is not the dominant type due to the small plot size and spectral overlap between different crops (Chabalala et al., 2022). In southern China, cropland plots are typically small (<0.04 ha in most regions) and the crop diversity is high. The growth periods of soybean, peanut, potato, and maize are similar, dominantly indicated by a mixed planting pattern, which has contributed to the low accuracy of non-main soybean producing areas in southern China (Liu et al., 2020).

5 Data availability

The soybean planting area product for China during 2017-2021 (ChinaSoyArea10m) is available at https://zenodo.org/doi/10.5281/zenodo.10071426 (Mei et al., 2023). We encourage users to independently verify data products for special study areas before using them.

6 Conclusions

In this study, a large-scale soybean mapping method was developed and utilized to generate soybean planting area maps for major producing regions in China from 2017 to 2021. By utilizing Sentinel-2 images, spectral features and vegetation indices that best distinguish soybeans were extracted and input into an unsupervised classifier in each prefecture. The DTW method was then employed to identify the soybean distribution. PPS does not rely on many ground samples and considers the soybean phenology in various planting areas, suggesting a potential way for long-term crop mapping over larger regions.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Verification results demonstrated a high consistency between the mapping results and census data at county or prefecture level (all > 0.82), with overall accuracies of field samples reaching 0.77~0.88. These findings confirm the reliability of ChinaSoyArea10m. Our data products fill the gap in regional long-term soybean maps in China, and provide important information for sustainable soybean production and management, agricultural system modeling, and optimization.

Author contributions.

ZZ and FT conceive this study. QM, JH, and JD collected datasets. QM implemented the research and wrote the original draft of the paper. All authors discussed the results and revised the manuscript.

Competing interests.

The contact author has declared that neither they nor their co-authors have any competing interests.

Financial support.

This research was funded by the National Key Research and Development Program of China (2020YFA0608201) and National Natural Science Foundation of China (42061144003, 41977405).

References

Chabalala, Y., Adam, E., and Ali, K. A.: Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data towards Mapping Fruit Plantations in Highly Heterogenous Landscapes, Remote Sens., 14, 2621, https://doi.org/10.3390/rs14112621, 2022.

Chen, D., Huang, J., and Jackson, T. J.: Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands, Remote Sens. Environ., 98, 225–236, https://doi.org/10.1016/j.rse.2005.07.008, 2005.

Chen, H., Li, H., Liu, Z., Zhang, C., Zhang, S., and Atkinson, P. M.: A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images,

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Remote Sens. Environ., 295, 113679, https://doi.org/10.1016/j.rse.2023.113679, 2023.

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

FAOSTAT: https://www.fao.org/faostat/en/#rankings/countries_by_commodity, last access: 10 October 2023.

Fensholt, R. and Sandholt, I.: Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment, Remote Sens. Environ., 87, 111–121, https://doi.org/10.1016/j.rse.2003.07.002, 2003.

Gong, L., Tian, B., Li, Y., and Wu, S.: Phenological Changes of Soybean in Response to Climate Conditions in Frigid Region in China over the Past Decades, Int. J. Plant Prod., 15, 363–375, https://doi.org/10.1007/s42106-021-00145-5, 2021.

Graesser, J. and Ramankutty, N.: Detection of cropland field parcels from Landsat imagery, Remote Sens. Environ., 201, 165–180, https://doi.org/10.1016/j.rse.2017.08.027, 2017.

Guan, X., Huang, C., Liu, G., Meng, X., and Liu, Q.: Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance, Remote Sens., 8, 19, https://doi.org/10.3390/rs8010019, 2016.

Guo, W., Ren, J., Liu, X., Chen, Z., Wu, S., and Pan, H.: Winter wheat mapping with globally optimized threshold under total quantity constraint of statistical data, Journal of Remote Sensing, 22, 1023–1041, 2018.

Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Zhang, J., and Li, Z.: The RapeseedMap10 database: annual maps of rapeseed at a spatial resolution of 10 m based on multi-source data, Earth Syst. Sci. Data, 13, 2857–2874, https://doi.org/10.5194/essd-13-2857-2021, 2021.

Hartman, G. L., West, E. D., and Herman, T. K.: Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests, Food Secur., 3, 5–17, https://doi.org/10.1007/s12571-010-0108-x, 2011.

Housman, I. W., Chastain, R. A., and Finco, M. V.: An Evaluation of Forest Health Insect and Disease Survey Data and Satellite-Based Remote Sensing Forest Change Detection Methods: Case Studies in the United States, Remote Sens., 10, 1184, https://doi.org/10.3390/rs10081184, 2018.

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Huang, Y., Qiu, B., Chen, C., Zhu, X., Wu, W., Jiang, F., Lin, D., and Peng, Y.: Automated soybean mapping based on canopy water content and chlorophyll content using Sentinel-2 images, Int. J. Appl. Earth Obs., 109, 102801, https://doi.org/10.1016/j.jag.2022.102801, 2022.

Konduri, V. S., Kumar, J., Hargrove, W. W., Hoffman, F. M., and Ganguly, A. R.: Mapping crops within the growing season across the United States, Remote Sensing of Environment, 251, 112048, https://doi.org/10.1016/j.rse.2020.112048, 2020.

Kumari, M., Murthy, C. S., Pandey, V., and Bairagi, G. D.: Soybean Cropland Mapping Using Multi-Temporal Sentinel-1 Data, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-3-W6, 109–114, https://doi.org/10.5194/isprs-archives-XLII-3-W6-109-2019, 2019.

Kwak, G.-H. and Park, N.-W.: Unsupervised Domain Adaptation with Adversarial Self-Training for Crop Classification Using Remote Sensing Images, Remote Sensing, 14, 4639, https://doi.org/10.3390/rs14184639, 2022.

Li, H., Song, X.-P., Hansen, M. C., Becker-Reshef, I., Adusei, B., Pickering, J., Wang, L., Wang, L., Lin, Z., Zalles, V., Potapov, P., Stehman, S. V., and Justice, C.: Development of a 10-m resolution maize and soybean map over China: Matching satellite-based crop classification with sample-based area estimation, Remote Sens. Environ., 294, 113623, https://doi.org/10.1016/j.rse.2023.113623, 2023.

Liu, J., Wang, L., Yang, F., Yao, B., and Yang, L.: Recognition ability of red edge and short wave infrared spectrum on maize and soybean, Chinese Agricultural Science Bulletin, 34, 120–129, 2018.

Liu, L., Xiao, X., Qin, Y., Wang, J., Xu, X., Hu, Y., and Qiao, Z.: Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine, Remote Sens. Environ., 239, 111624, https://doi.org/10.1016/j.rse.2019.111624, 2020.

Liu, M. and Fan, Q.: Study on the Current Situation and Problems of Soybean Consumption, Production and Import in China, Grain Science And Technology And Economy, 46, 28–35, https://doi.org/10.16465/j.gste.cn431252ts.20210606, 2021.

Lowder, S. K., Skoet, J., and Raney, T.: The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide, World Development, 87, 16–29, https://doi.org/10.1016/j.worlddev.2015.10.041, 2016.

Luo, C., Liu, H., Lu, L., Liu, Z., Kong, F., and Zhang, X.: Monthly composites from Sentinel-1 and

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

Sentinel-2 images for regional major crop mapping with Google Earth Engine, J. Integr. Agr., 20, 1944–1957, https://doi.org/10.1016/S2095-3119(20)63329-9, 2021.

Luo, Y., Zhang, Z., Li, Z., Chen, Y., Zhang, L., Cao, J., and Tao, F.: Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources, Environ. Res. Lett., 15, 074003, https://doi.org/10.1088/1748-9326/ab80f0, 2020.

Luo, Y., Zhang, Z., Zhang, L., Han, J., Cao, J., and Zhang, J.: Developing High-Resolution Crop Maps for Major Crops in the European Union Based on Transductive Transfer Learning and Limited Ground Data, Remote Sens., 14, 1809, https://doi.org/10.3390/rs14081809, 2022.

Ma, Z., Liu, Z., Zhao, Y., Zhang, L., Liu, D., Ren, T., Zhang, X., and Li, S.: An Unsupervised Crop Classification Method Based on Principal Components Isometric Binning, ISPRS Int. J. Geo-Inf., 9, 648, https://doi.org/10.3390/ijgi9110648, 2020.

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., and Nelson, A.: Field-level crop yield estimation with PRISMA and Sentinel-2, ISPRS J. Photogramm. Remote Sens., 187, 191–210, https://doi.org/10.1016/j.isprsjprs.2022.03.008, 2022.

Mei, Q., Zhang, Z., Han, J., Song, J., Dong, J., Wu, H., Xu, J., and Tao, F.: ChinaSoyArea10m: a dataset of soybean planting areas with a spatial resolution of 10 m across China from 2017 to 2021 (V1), Zenodo [data set], https://doi.org/10.5281/zenodo.10071427, 2023.

National Bureau of Statistics of China: http://www.stats.gov.cn/english/, last access: 23 October 2023.

Olsen, J. L., Stisen, S., Proud, S. R., and Fensholt, R.: Evaluating EO-based canopy water stress from seasonally detrended NDVI and SIWSI with modeled evapotranspiration in the Senegal River Basin, Remote Sens. Environ., 159, 57–69, https://doi.org/10.1016/j.rse.2014.11.029, 2015.

Oreopoulos, L., Wilson, M. J., and Várnai, T.: Implementation on Landsat Data of a Simple Cloud-Mask Algorithm Developed for MODIS Land Bands, IEEE Geosci. Remote. Sens. Lett., 8, 597–601, https://doi.org/10.1109/LGRS.2010.2095409, 2011.

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., and Cortez, J.: Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century, Nat. Food, 3, 19–28, https://doi.org/10.1038/s43016-021-00429-z, 2022.

Shangguan, Y., Li, X., Lin, Y., Deng, J., and Yu, L.: Mapping spatial-temporal nationwide soybean planting area in Argentina using Google Earth Engine, Int. J. Remote Sens., 43, 1724–1748,

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

https://doi.org/10.1080/01431161.2022.2049913, 2022.

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

Sobejano-Paz, V., Mikkelsen, T. N., Baum, A., Mo, X., Liu, S., Köppl, C. J., Johnson, M. S., Gulyas, L., and García, M.: Hyperspectral and Thermal Sensing of Stomatal Conductance, Transpiration, and Photosynthesis for Soybean and Maize under Drought, Remote Sens., 12, 3182, https://doi.org/10.3390/rs12193182, 2020.

Song, X.-P., Potapov, P. V., Krylov, A., King, L., Di Bella, C. M., Hudson, A., Khan, A., Adusei, B., Stehman, S. V., and Hansen, M. C.: National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey, Remote Sens. Environ., 190, 383–395, https://doi.org/10.1016/j.rse.2017.01.008, 2017.

Wang, S., Azzari, G., and Lobell, D. B.: Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques, Remote Sens. Environ., 222, 303–317, https://doi.org/10.1016/j.rse.2018.12.026, 2019.

Wang, S., Di Tommaso, S., Deines, J. M., and Lobell, D. B.: Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive, Sci. Data, 7, 307, https://doi.org/10.1038/s41597-020-00646-4, 2020.

Wang, Y. and Gai, J.: Study on the ecological regions of soybean in China II · Ecological environment and representative varieties, Chinese Journal of Applied Ecology, 71–75, 2002.

Wang, Y., Feng, L., Sun, W., Zhang, Z., Zhang, H., Yang, G., and Meng, X.: Exploring the potential of multi-source unsupervised domain adaptation in crop mapping using Sentinel-2 images, Gisci. Remote Sens., 59, 2247–2265, https://doi.org/10.1080/15481603.2022.2156123, 2022.

Wang, Y., Ling, X., Ma, C., Liu, C., Zhang, W., Huang, J., Peng, S., and Deng, N.: Can China get out of soy dilemma? A yield gap analysis of soybean in China, Agron. Sustain. Dev., 43, 47, https://doi.org/10.1007/s13593-023-00897-6, 2023.

Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., Yadav, K., and Thau, D.: Automated cropland mapping of continental Africa using Google Earth Engine cloud computing, ISPRS J. Photogramm. Remote Sens., 126, 225–244,

https://doi.org/10.5194/essd-2023-467
Preprint. Discussion started: 18 December 2023

https://doi.org/10.1016/j.isprsjprs.2017.01.019, 2017.

You, N. and Dong, J.: Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine, ISPRS J. Photogramm. Remote Sens., 161, 109–123, https://doi.org/10.1016/j.isprsjprs.2020.01.001, 2020.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y., and Xiao, X.: The 10-m crop type maps in Northeast China during 2017-2019, Figshare [data set], https://doi.org/10.6084/m9.figshare.13090442, 2020.

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

You, N., Dong, J., Li, J., Huang, J., and Jin, Z.: Rapid early-season maize mapping without crop labels, Remote Sens. Environ., 290, 113496, https://doi.org/10.1016/j.rse.2023.113496, 2023.

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., and Yang, P.: A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps, Earth Syst. Sci. Data, 12, 3545–3572, https://doi.org/10.5194/essd-12-3545-2020, 2020.

Zhang, C., Dong, J., and Ge, Q.: Quantifying the accuracies of six 30-m cropland datasets over China: A comparison and evaluation analysis, Comput. Electron. Agr., 197, 106946, https://doi.org/10.1016/j.compag.2022.106946, 2022.

Zhong, L., Hu, L., Yu, L., Gong, P., and Biging, G. S.: Automated mapping of soybean and corn using phenology, ISPRS J. Photogramm. Remote Sens., 119, 151–164, https://doi.org/10.1016/j.isprsjprs.2016.05.014, 2016.