



Development of a 10-m resolution maize and soybean map over China: Matching satellite-based crop classification with sample-based area estimation

Haijun Li^{a,b}, Xiao-Peng Song^{a,b,*}, Matthew C. Hansen^a, Inbal Becker-Reshef^a, Bernard Adusei^a, Jeffrey Pickering^a, Li Wang^c, Lei Wang^c, Zhengyang Lin^d, Viviana Zalles^a, Peter Potapov^a, Stephen V. Stehman^e, Chris Justice^a

^a Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA

^b Department of Geosciences, Texas Tech University, Lubbock, TX 79415, USA

^c State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

^d Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

^e Department of Sustainable Resources Management, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210, United States

ARTICLE INFO

Edited by Dr. Marie Weiss

Keywords:

Sentinel-2

Planet

Crop area estimate

Random forests

Operational mapping

ABSTRACT

Spatially explicit information on crop distribution is essential for market information, food security, and agricultural sustainability. However, high-resolution crop maps are unavailable for most countries of the world. In this study, we developed an operational workflow and produced the first openly-available 10-m resolution maize and soybean map over China. We also derived area estimates for maize and soybean extent for 2019 using a stratified, two-stage, cluster sampling design and ground data collected for the entire country. We developed a multi-scale, multi-temporal procedure for mapping, in which field data were used as training to map maize and soybean over the first-stage sample of 10 km × 10 km equal-area blocks with PlanetScope and Sentinel-2 data. Then, the classified blocks were used as training to map maize and soybean for the country with wall-to-wall Sentinel-2 data using a random forests approach. We used all available Sentinel-2 surface reflectance data acquired between April and October 2019, applied quality assurance, including cloud and shadow masking, and created monthly image composites as inputs for the random forests analysis. We derived maize and soybean area estimates using the field sample data and a regression estimator. Utilizing the probability output layer of the random forests models, we found and applied empirical probability thresholds that matched map-based crop area estimates with sample-based area estimates. Maize area in China in 2019 was estimated to be 330,609 ± 34,109 km² (± value is the standard error), and soybean area was estimated to be 78,107 ± 12,969 km². Validated using the field sample data as reference, our crop map had an overall accuracy of 91.8 ± 1.2%. The user's and producer's accuracies for the maize class were 93.9 ± 2.5% and 79.2 ± 3.6%, and for the soybean class were 63.6 ± 12.1% and 61.9 ± 11.8%. Our map-based maize and soybean area estimates had close agreement with government reports at the provincial and prefectural levels, with r^2 of 0.90 and 0.92 for maize, and 0.93 and 0.94 for soybean, respectively. Our workflow can generate internally consistent results for crop area estimation and crop mapping simultaneously. As Sentinel-2 data are being acquired consistently and very-high-resolution commercial satellite data are increasingly available, our established workflow may be applied in an operational setting for annual crop mapping in China and other countries.

1. Introduction

National-scale crop production information is traditionally acquired using statistical ground surveys. Although survey-based methods can be

used to collect a large variety of agricultural data such as crop type, livestock, irrigation, field and soil conditions, they cannot provide spatially explicit information at high resolution. Survey-based methods are generally time consuming, labor intensive, and vary from country to

* Corresponding author at: Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA.

E-mail address: xpsong@umd.edu (X.-P. Song).

<https://doi.org/10.1016/j.rse.2023.113623>

Received 7 February 2023; Received in revised form 4 May 2023; Accepted 8 May 2023

Available online 16 May 2023

0034-4257/© 2023 Elsevier Inc. All rights reserved.

country (Benedetti et al., 2010). Freely available satellite Earth observations offer a complementary method to obtain crop information at broad scales consistently across space and time and at low cost (Bégué et al., 2018; Weiss et al., 2020). However, distinguishing specific crop types (such as soybean, maize, wheat, etc.) accurately with satellite data is challenging, particularly over large areas with diverse crop types, heterogeneous field sizes and varying farming practices. Generating crop maps at national to continental scales requires high-quality satellite data, robust data processing, representative field data for training, field data from a probability sample for validation, as well as high-performance computing infrastructure (Song et al., 2021a).

Satellite data with complete spatial coverage, adequate spatial resolution, frequent temporal revisit and low cost are essential for crop mapping at broad scales. Currently available datasets that fit these criteria include those acquired by the Landsat and Sentinel series of satellites. Since the opening of the Landsat data archive in 2008 (Woodcock et al., 2008), researchers and practitioners have explored Landsat data for crop mapping at 30-m resolution for many countries and regions, including the US (Boryan et al., 2011; Johnson, 2019; Song et al., 2017; Wang et al., 2020), Pakistan (Khan et al., 2021), Turkey (Rufin et al., 2019), Germany (Blickensdörfer et al., 2022) and South America (Song et al., 2021a). The 16-day revisit frequency of Landsat is sparse considering the relatively short growing season of annual crops, especially when persistent clouds exist. Thus, data from the Moderate Resolution Imaging Spectroradiometer (MODIS) with daily revisit frequency are often used in conjunction with Landsat for mapping (Li et al., 2022; Song et al., 2021a; Song et al., 2017; Zalles et al., 2019).

The spatial resolution of MODIS and Landsat are coarse when applied to heterogeneous landscapes of small crop fields. As with Landsat, Sentinel-2 has complete global coverage and free data access. With twin satellites and a broader swath, Sentinel-2 has a higher revisit frequency of 5 days, finer spatial resolution of 10 m and 20 m, and more spectral bands, making it the best freely-available satellite data source for large-area crop mapping. Studies have shown that the red-edge bands of Sentinel-2 are particularly useful for crop mapping (Immitzer et al., 2016; Song et al., 2021b) and that the 10-m resolution is capable of mapping smallholder systems in Tanzania and Kenya (Jin et al., 2019). Sentinel-2 data have been used to generate crop maps in recent years over Germany (Blickensdörfer et al., 2022; Preidl et al., 2020), and other European Union countries (Luo et al., 2022). Sentinel-1 Synthetic Aperture Radar (SAR) are also freely available, but large-area applications of Sentinel-1 alone for crop mapping are rare, although d'Andrimont et al. (2021) provide a recent example over Europe. However, the combination of Sentinel-1 and optical data has been found useful in crop classification (Blickensdörfer et al., 2022; Kussul et al., 2020; Van Tricht et al., 2018).

Commercial satellite data provide even higher resolution of 0.5–5 m (e.g., PlanetScope, WorldView, RapidEye, etc.), but they are often expensive, not openly accessible, or only publicly available in some regions such as the world's tropics (NICFI, 2022). Using such datasets for large-area wall-to-wall crop mapping is challenging due to the large data amount and data-intensive computation. However, a sample of these very-high-resolution images could be utilized for area estimation when the full coverage is not affordable (Khan et al., 2018).

Crop diversity is a primary challenge for crop mapping at national scales, especially for large countries such as China. Thus, ground data are crucial for mapping. Ground surveys are traditionally used to collect field data for statistical estimation based on sampling frames such as list frames, area frames and combination of multiple frames (FAO, 2015). List sampling frames are lists of farms derived from agricultural censuses whereas area sampling frames are sets of land segments (FAO, 2015). The sample area units or points in an area frame are inspected by enumerators through direct interviews with stakeholders, sometimes supplemented with cartographic materials such as satellite imagery and aerial photography (FAO, 2018). In recent years, land cover products have been employed to automatically generate area frame strata to

improve the sampling efficiency (Boryan et al., 2014). Stratified probability sampling is highly effective to collect reference data for producing unbiased and precise crop area and map accuracy estimators (Gallego, 2004; Olofsson et al., 2014; Stehman, 2000). Currently, only some countries have used representative field data for operational large-area crop mapping such as the Cropland Data Layer (CDL) in the US (Boryan et al., 2011) and the Annual Crop Inventory (ACI) in Canada (Fisette et al., 2013). These programs are built upon enormous ground surveys and the associated 30-m resolution maps typically have overall accuracies higher than 85% for major crop categories (Boryan et al., 2011; Fisette et al., 2013). In Europe, the Land Use/Cover Area frame Survey (LUCAS) has been conducted for several years since 2006 to collect in-situ points for high-resolution land cover mapping over the European Union (d'Andrimont et al., 2021; d'Andrimont et al., 2020). The LUCAS tri-annual surveyed samples are freely accessible and have enabled large-area crop mapping at 30-m resolution (e.g., Pflugmacher et al., 2019) as well as 10-m resolution (e.g., d'Andrimont et al., 2021; Weigand et al., 2020). These maps reported average accuracies over 75% by utilizing LUCAS as ground reference data.

Broad-scale crop mapping projects only exist in limited countries or regions, though are rapidly increasing. Since 2015, the Sentinel-2 for Agriculture (Sen2-Agri) program launched by the European Space Agency (ESA) has produced 10-m resolution crop maps for 12 test sites distributed in Europe, Asia and North America (Defourny et al., 2019; Inglada et al., 2015). The ESA WorldCereal is generating global cropland products as well as crop type maps (<https://esa-worldcereal.org/en>). In some Southeast and Northeast Asian countries, single-crop maps were produced, such as the NESEA-Rice10 (Han et al., 2021a), the Rape-seedMap10 (Han et al., 2021b), and the rice-cropping and cultural type dataset (Manjunath et al., 2015). In China, the CropWatch system has been developed since 1998 to monitor crop growth with spatial resolutions ranging from 30 m to 1 km (Wu et al., 2013). CropWatch focuses on crop growth condition, yield and production modeling by using indicators derived from satellite imagery. National-scale crop distribution maps have been produced over China, but the coarse spatial resolution (e.g., 250 m in Clauss et al. (2016) and 0.5° grid in Frohking et al. (2002)) hindered further analyses for smallholder farms (Wu et al., 2018). Although 10-m regional crop maps have been generated in recent years (Jiang et al., 2020; Xiao et al., 2021; You et al., 2021; Zhang et al., 2018), national-scale maps at 10-m resolution are still publicly unavailable in China. Considering that China is the second largest maize producer and the largest maize and soybean consumer globally (FAO-Stat, 2021), it can have a significant impact on global agricultural markets such as in 2020 when international market prices increased substantially in large part due to increased imports by China. As such, having reliable and transparent information on China's maize and soybean production is critical.

Statistical area estimation and satellite-based mapping are often two separate procedures. Crop maps, even with high accuracies, are not suitable for direct area estimation by "pixel counting" due to bias arising from misclassifications and mixed pixels (Gallego, 2004). In contrast, using a probability sample of reference data enables us to produce unbiased area estimates with associated uncertainty measures. Recent research has demonstrated the unique advantages of combining sample-based area estimation with satellite-based mapping to produce internally consistent area estimates and crop maps simultaneously at regional, national and continental scales, with demonstrated applications for soybean in the US (Song et al., 2017), Argentina (King et al., 2017), and South America (Song et al., 2021a), wheat in Pakistan (Khan et al., 2021), and cropland in Brazil (Zalles et al., 2019) and South America (Zalles et al., 2021).

The objective of this study was to estimate the area and map the distribution of maize and soybean over China for the 2019 growing season. We implemented a stratified random sampling design to collect field data over the country for area estimation, model training and map validation. We developed an operational workflow to process Sentinel-2

Multispectral Instrument (MSI) surface reflectance at 10-m resolution, processed all MSI data for the growing season of 2019, and generated a 10-m maize and soybean map that matched sample-based area estimates at the national scale.

2. Data and methods

Our method was prototyped for the US (Song et al., 2017), and tested in Argentina (King et al., 2017), Pakistan (Khan et al., 2021) and at the continental scale for South America (Song et al., 2021a). Here we

advance the method through its application in China using Sentinel-2 data. The overall method consists of six major modules (Fig. 1): (1) sampling design and field survey; (2) block-level crop mapping with high-resolution data; (3) sample-based crop area estimation; (4) wall-to-wall satellite data processing; (5) national-scale crop classification; and (6) crop map evaluation. Sample-based area estimation and national-scale satellite-based mapping interact at various stages throughout the workflow: a) national mapping relies on field data for model training; b) applying a regression estimator to improve precision of the sample-based area estimates relies on an intermediate output of national

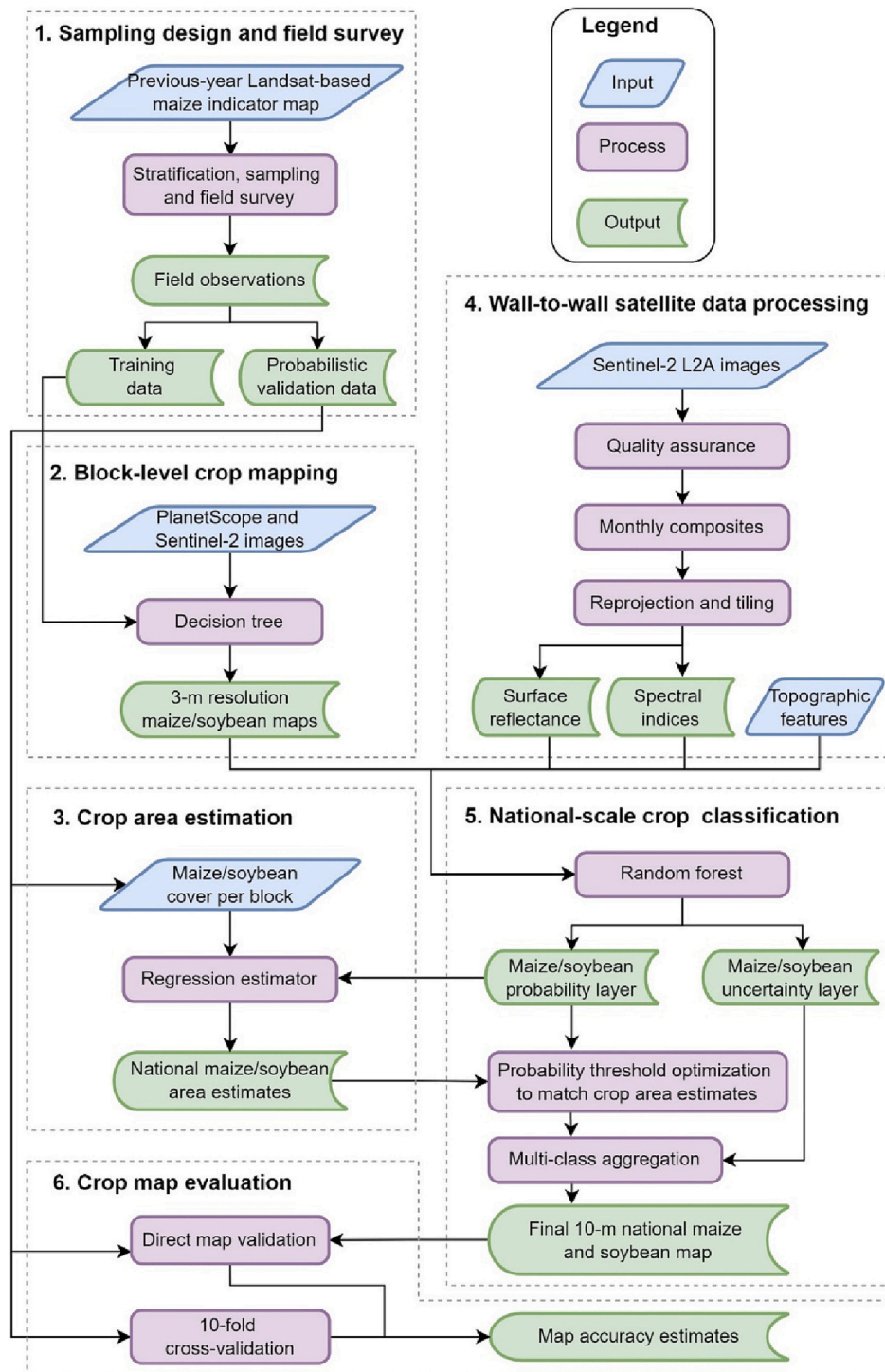


Fig. 1. An overview of the workflow for crop area estimation and mapping.

mapping as the auxiliary variable; c) final sample-based area estimates are used as a constraint for national mapping; and d) field data from the probability sample are used as reference for map validation. We describe these technical details in sections 2.1 through 2.6 below, corresponding to steps 1 through 6, respectively (Fig. 1).

2.1. Sampling design and field survey

The goal of the field survey was to collect statistical data for crop area estimation as well as training for classification models. To select field sites to visit, we implemented a stratified, two-stage cluster sample over China. The sampling approach was modified from the original method proposed by Song et al. (2017). The land extent of China was divided into 10 km × 10 km equal-area blocks. Our primary target crop was maize, and thus, we developed a Landsat-based maize indicator map for 2018 for the stratification used in the probability sampling design to collect the field data. The maize indicator map was generated using a decision tree classifier trained with expert-selected training data and growing-season Landsat imagery (Potapov et al., 2020). We calculated per-block maize area fraction using the indicator map, ranked all blocks from high to low maize fraction, and selected a total of 16,931 blocks that included 95% of all maize pixels in the indicator map. These blocks defined the statistical population (i.e., study area) represented by the sample-based area estimates. Based on the sorted blocks with maize area fraction from the highest to the lowest, we grouped the blocks into high (2264 blocks), medium (4464 blocks) and low (10,203 blocks) strata, with the high stratum covering a cumulative area of 50% of all mapped maize pixels, the medium stratum covering an additional 30% of maize pixels and the low stratum covering the remaining 15% of maize pixels (Fig. 2). We randomly selected 15 blocks in each stratum as the primary sampling units (PSUs). For each PSU, we randomly selected 10 10 m × 10 m pixels as the secondary sampling units (SSUs) (Fig. 2b). The 10-m SSUs were constructed to match the pixel size of Sentinel-2. Although the stratification was primarily designed for maize, due to the large spatial overlap between maize and soybean, we also used the sample for soybean area estimation and mapping. We manually selected five extra blocks to collect ground data, targeting the soybean growing regions outside of the maize population. In total, 45 PSUs and 450 SSUs were selected for field survey (Table 1). These 5 blocks were only used for model training, not for area estimation or map accuracy assessment.

The collected field data consisted of two subsets: 1) a probability sample to obtain the field data for area estimation and map validation and 2) data from a non-probability sample used for model training. The primary goal of the field survey was to collect crop information over 450 SSUs for area estimation (Section 2.3) and map validation (Section 2.6). We visited every SSU on site, recorded the crop type and took GPS-tagged photos (see Table S1 for detailed information). Separate from the probability sample of SSUs, we also implemented a “windshield survey” in which we recorded a large number of opportunistic field observations along the roads while we traveled to the SSU sites (Fig. 2b). These windshield survey data were used solely as training data for block-level maize and soybean classification (Section 2.2). We collected 17,858 training points in total from July 21st to August 8th in 2019 (see Fig. S1).

2.2. Block-level mapping with high-resolution satellite data and windshield survey

The goal of block-level mapping was to generate representative training data sufficient for national-scale mapping. For every PSU block, we downloaded all PlanetScope images as well as all Sentinel-2A/B images acquired between May 1st and September 30th, 2019 (Fig. 3). We used all available PlanetScope data within the growing window without cloud masking or compositing. We resampled Sentinel-2 images to 3 m using nearest neighbor to match the spatial resolution of PlanetScope data. For each PSU, we stacked PlanetScope, Sentinel-2A/B

images, employed the “windshield survey” data as training, and trained two decision tree models for maize and soybean classification. All the bands and normalized ratios of any two bands, for both PlanetScope and Sentinel-2A/B images, were used in classification. We applied the trained models to the image stack and generated a binary maize/non-maize map and a binary soybean/non-soybean map at 3-m resolution. Model training and mapping was conducted block by block. These mapped PSU blocks were used for the national-scale wall-to-wall mapping as discussed in Section 2.5 below.

2.3. Sample-based crop area estimation

The goal of sample-based crop area estimation was to derive unbiased area and uncertainty estimates at the national scale for both maize and soybean. We employed regression estimators, which have been demonstrated to be effective in reducing the variance (Bellow, 1994; Carfagna and Gallego, 2006; Gallego, 2004), to estimate the area of maize and soybean in China. We derived maize and soybean fractions for each PSU based on the crop information collected over SSUs. Following the method described in King et al. (2017), we used the SSU-based maize/soybean fractions as the dependent variable (y_i), and used a satellite-based crop probability layer of the same season as the auxiliary variable (x_i) for the regression estimator. The method used to derive the crop probability layer using Sentinel-2 data is described in detail in Sections 2.4 and 2.5 below. Here, for each stratum i , a regression estimator is constructed as the following:

$$\bar{y}_{ri} = \bar{y}_i + b_i^*(\bar{X}_i - \bar{x}_i) \quad (1)$$

where the \bar{y}_{ri} denotes the estimated crop fraction for stratum i ; \bar{y}_i is the sample mean of crop fraction derived from the reference labels applied to the SSUs in stratum i ; \bar{x}_i is the sample mean of crop fraction derived from satellite-based maps to the SSUs in stratum i ; \bar{X}_i is the population mean for stratum i derived from the satellite-based crop probability layer; b_i is the ordinary least squares estimate of the slope for stratum i . The national crop area (\bar{y}_{total}) is produced by the combination of per-stratum regression estimate:

$$\bar{y}_{total} = \frac{1}{N} \sum_i N_i \bar{y}_{ri} \quad (2)$$

where N is the number of total blocks and N_i is the number of blocks in stratum i . The estimated variance ($\hat{V}(\bar{y}_{total})$) is calculated as follows:

$$\hat{V}(\bar{y}_{total}) = \frac{1}{N^2} \sum_i \frac{N_i^2 \left(1 - \frac{n_i}{N_i}\right)}{n_i} \left[s_{yi}^2 - 2b_i s_{yxi} + b_i^2 s_{xi}^2 \right] \quad (3)$$

where n_i is the number of blocks sampled in stratum i (15 per stratum in this study); s_{yi}^2 and s_{xi}^2 are the sample variances of y and x in stratum i ; s_{yxi} is the sample covariance between x and y in stratum i .

2.4. Wall-to-wall Sentinel-2 data processing

The goal of this step was to pre-process Sentinel-2 data into Analysis Ready Data (ARD) format that can be directly used in national crop classification. We downloaded Sentinel 2A and 2B Level-2A Bottom of the Atmosphere reflectance (S2 L2A) images from Google Cloud (<https://cloud.google.com/storage/docs/public-datasets/sentinel-2>, accessed on May 09, 2022), in the format of Sentinel-2 Standard Archive Format for Europe (SAFE). We used an 80% cloud cover threshold to filter images acquired from April 1st to October 31st, 2019, resulting in 66,750 SAFE folders (Fig. 4).

2.4.1. Quality assurance

Based on the S2 scene classification (SCL) map available with L2A,

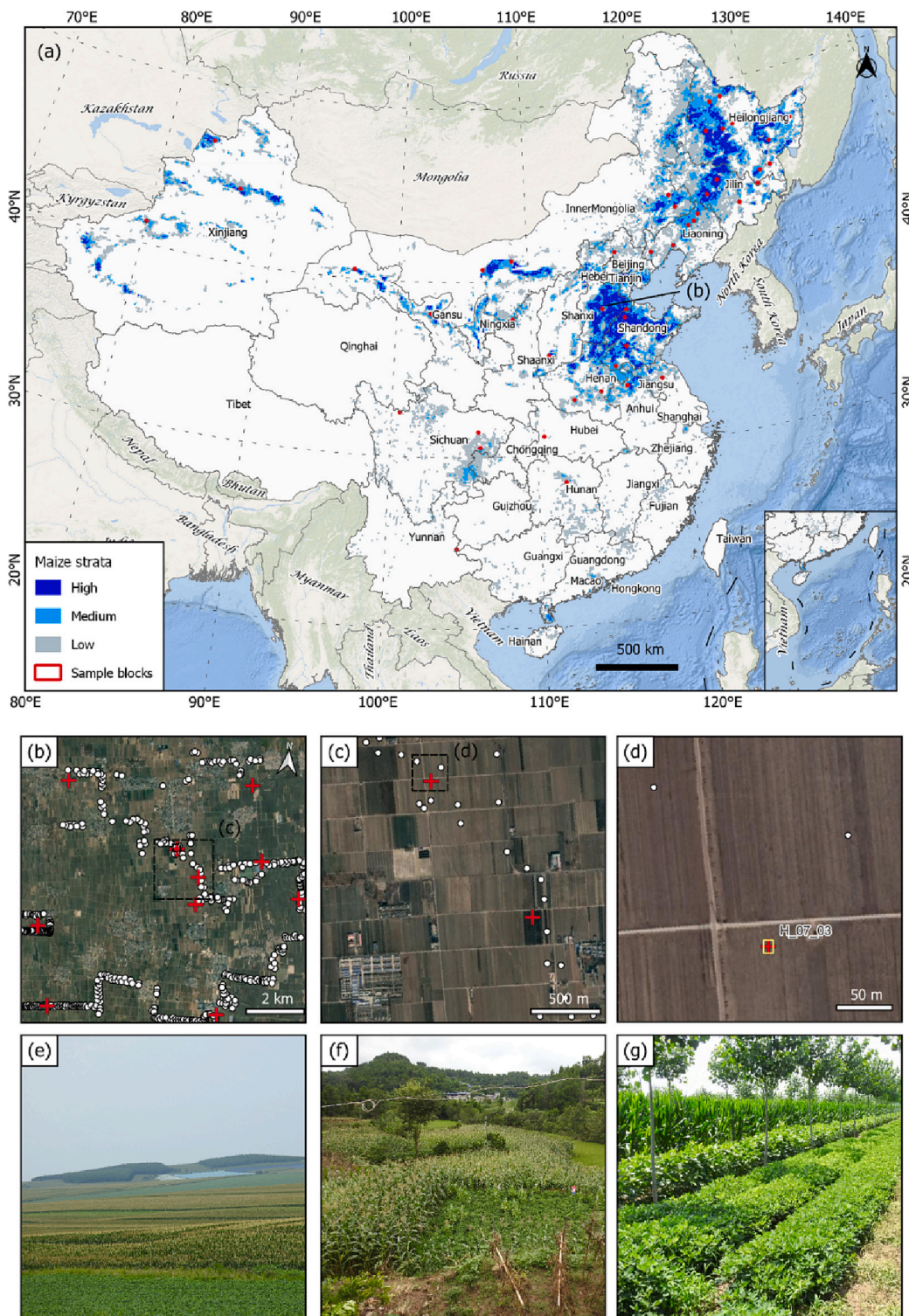


Fig. 2. Probability sampling design and field data collection for crop mapping in China. (a) maize strata and first-stage sample blocks (primary sampling units, PSUs). (b) a representative sample block in Hebei with center coordinates (126.676° E, 43.965° N). Second-stage sample pixels (secondary sampling units, SSUs) within the block are shown as red crosses, and “windshield survey” points along the roads are shown as white dots, with Google Earth image as background. (c-d) zoom-in panels for SSUs and field points at field scales, the pixel boundary is shown as the yellow box. (e-g) Selected field photos showing the large diversity of agricultural practices in China: (e) large-field maize, (f) smallholder maize mixed with other crops in irregular fields, (g) inter-cropping field with maize, soybean, peanut and trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we generated a SCL-derived quality assurance (QA) mask by merging cloud shadows, cloud with low, medium, and high probability, thin cirrus, and snow classes. To improve cloud identification over bright targets such as bare ground, reflective rocks and built-up areas, which could be committed or omitted by Sen2Cor (Coluzzi et al., 2018), we applied additional spectral tests (Bolton et al., 2020). Given that white clouds appear to have “flat” reflectance in visible bands, the blue, green, red bands were used to calculate the mean surface reflectance (SR) value (Eq. 4) to conduct the whiteness test (Eq. 5) (Gomez-Chova et al., 2007;

Zhu and Woodcock, 2012). To separate bright objects from clouds, the band ratio of near-infrared (NIR) and shortwave-infrared (SWIR) was used to compensate the SCL-derived cloud mask (Eq. 6) (Irish, 2000). By visually interpreting the quality of cloud masks, the empirical threshold of 0.3 for whiteness and 1.4 for the band ratio of NIR and SWIR (1610 nm) were used. A pixel was identified as cloud using the whiteness test <0.3 and the NIR-SWIR (1610 nm) ratio >1.4 . These spectral tests resulted in a supplementary cloud mask, which was combined with the SCL-derived mask. We then applied additional morphological

Table 1

Maize strata and sample size for area estimation. National maize coverage is derived from 2018 Landsat-based maize indicator map and the block size is 10 km × 10 km.

Stratum	National maize coverage	Maize fraction per block	Number of blocks in stratum	Number of PSU in stratum	Number of SSU in blocks
High	50%	9.7% ~ 52.8%	2264	15	150
Medium	30%	2.7% ~ 9.7%	4464	15	150
Low	15%	< 2.7%	10,203	15	150
Total	95%	–	16,931	45	450

operations—closing followed by opening using a 3 × 3 kernel—to remove small scattered pixels and to fill the holes in large clouds. 45.6% observations were flagged cloudy and removed from subsequent analysis.

$$MeanSR = \frac{Band_1 + Band_2 + Band_3}{3} \quad (4)$$

$$Whiteness\ Test = \sum_{i=1}^3 \left| \frac{Band_i - MeanSR}{MeanSR} \right| \quad (5)$$

$$Ratio_{NIR}^{SWIR} = \frac{NIR}{SWIR} \quad (6)$$

2.4.2. Monthly image composition

We resampled all the 20-m bands to 10-m resolution using nearest neighbor resampling, including Red-Edge-1 (RE1), Red-Edge-2 (RE2), Red-Edge-3 (RE3), narrow near-infrared (NNIR), SWIR1 and SWIR2. We then applied the QA mask and created monthly image composites to reduce the data volume. For all clear-view pixels in a given month, the median NIR and the corresponding day of year (DOY) information was retrieved; surface reflectance values of all bands acquired on the DOY were selected to create the composited image. The median band value approach has shown advantages for timeseries imagery compositing

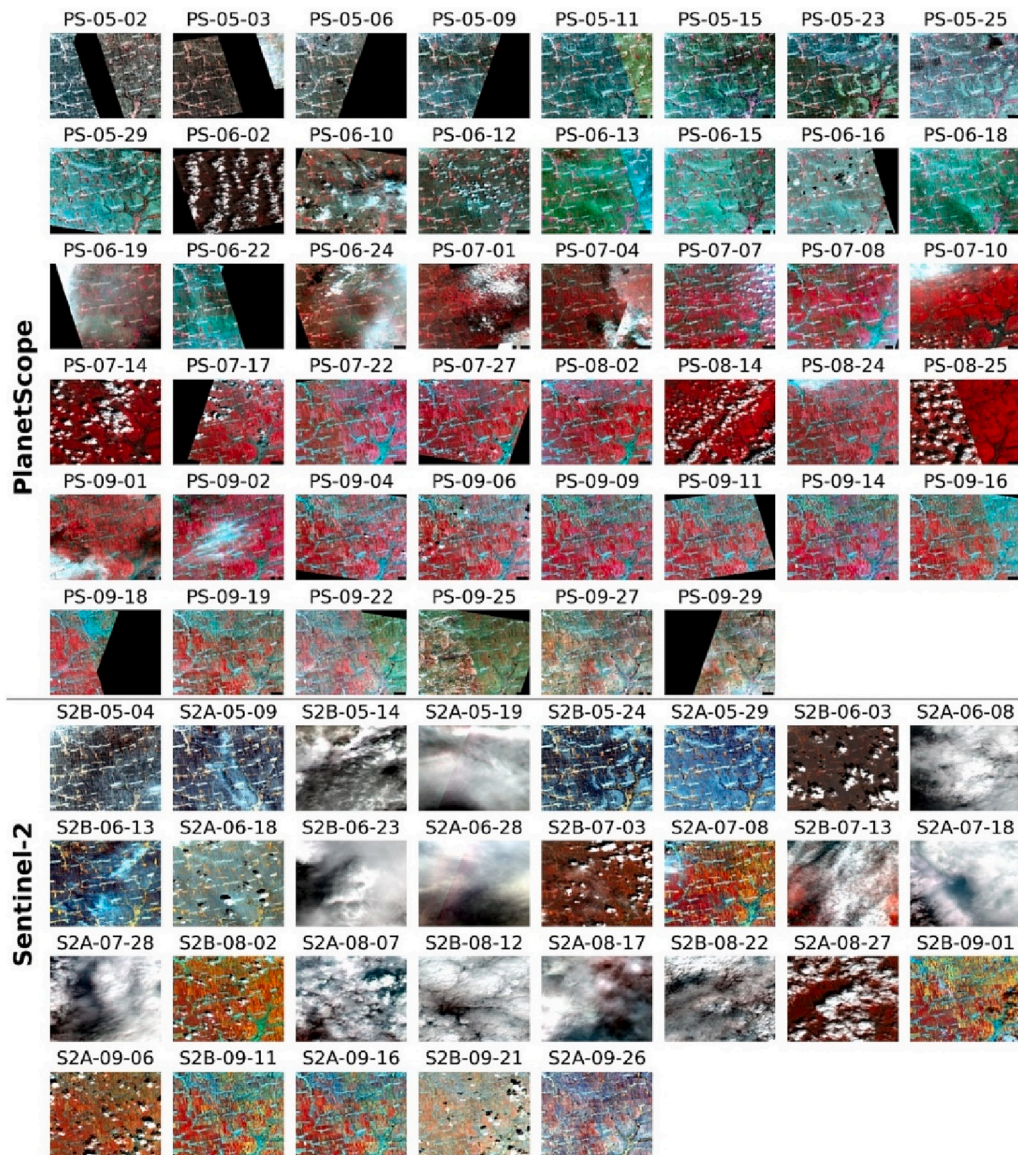


Fig. 3. PlanetScope (PS) and Sentinel-2 (S2) images acquired between May 1st and September 30th, 2019 over a representative 10 km × 10 km sample block. All panels are shown in the same band combination of NIR/Red/Green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

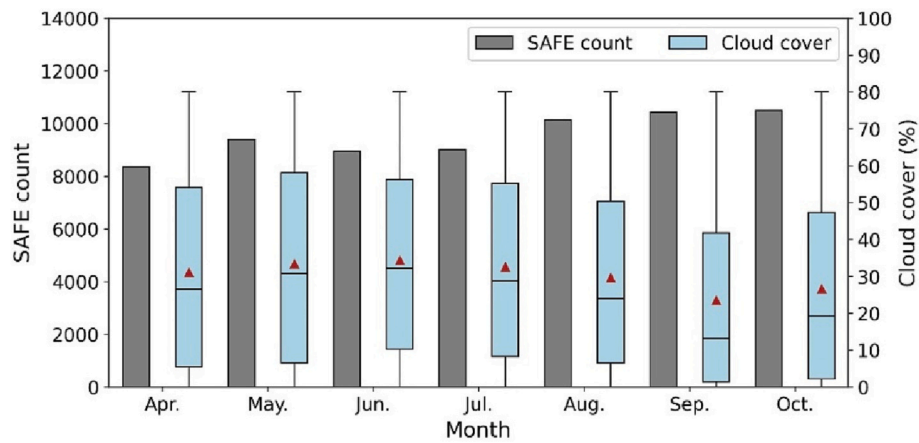


Fig. 4. Monthly Sentinel-2 data in Standard Archive Format for Europe (SAFE) format and cloud cover statistics over China from April to October 2019. For cloud cover, the median and mean values are shown as black lines and red triangles in boxplots, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

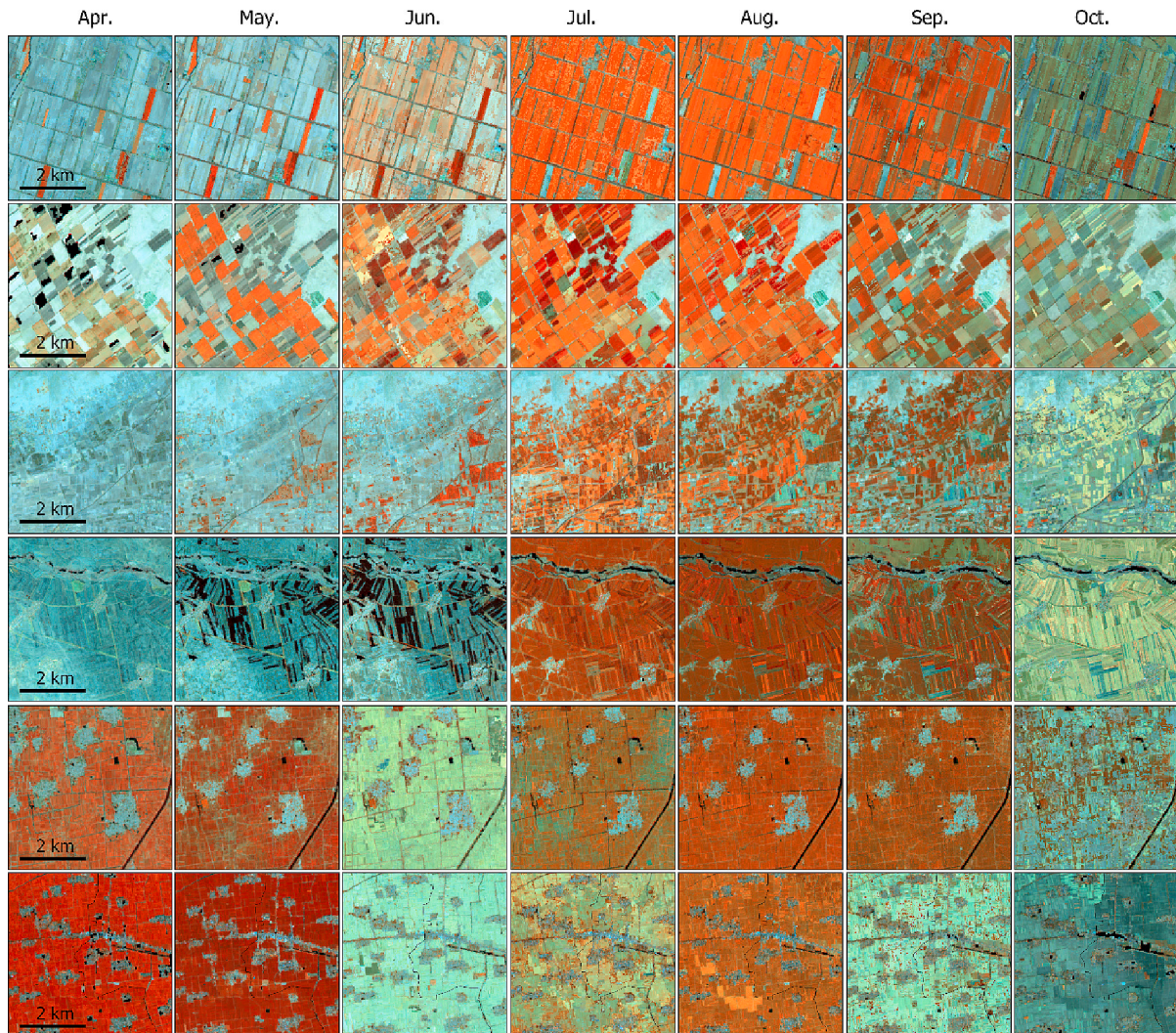


Fig. 5. Phenological variations over agricultural regions are visible from the 10-m monthly Sentinel-2 composites. All panels are shown in NIR/SWIR1/SWIR2 band combination within 5 km \times 5 km blocks. From top to bottom, the images are located in Xinjiang (85.598° E, 44.667° N), Gansu (102.481° E, 38.553° N), Inner Mongolia (107.564° E, 41.246° N), Jilin (122.261° E, 45.660° N), Shandong (117.250° E, 37.561° N) and Henan (114.894° E, 33.095° N), respectively. Single-cropping patterns in Xinjiang, Gansu, Inner Mongolia and Jilin, and double-cropping patterns in Shandong and Henan are both shown by the monthly composites.

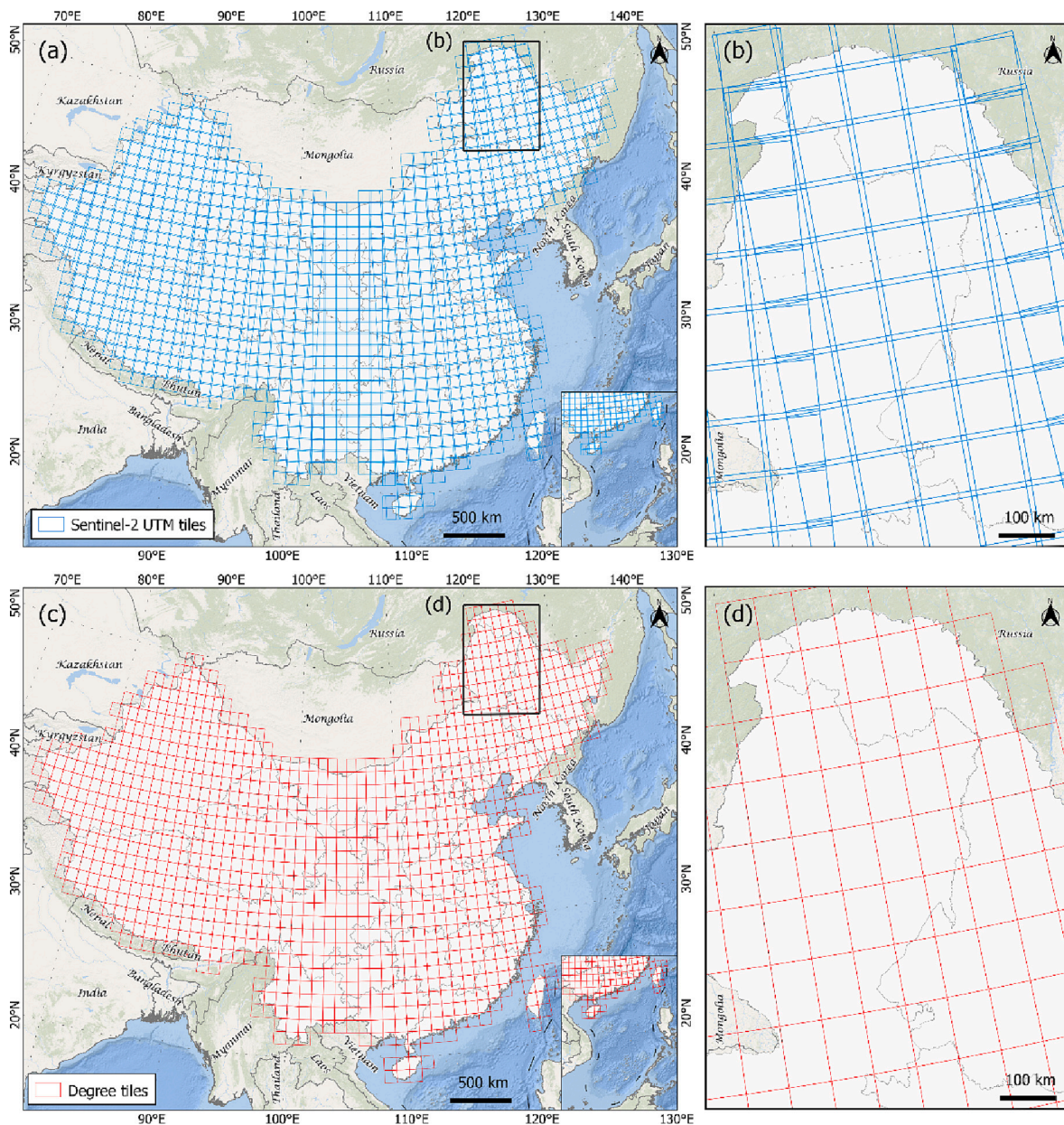


Fig. 6. The geographic tiling system for Sentinel 2 data processing over China. (a-b) original UTM tiles of Sentinel-2 data. (c-d) $1^{\circ} \times 1^{\circ}$ degree geographic tiles in WGS84 datum. Compared to the original UTM tiles, our degree tiles do not have overlap between neighboring orbital tracks (b and d).

compared to the maximum NDVI approach (Potapov et al., 2011). This median NIR-based composition strategy was chosen as a result of visual assessments of multiple composition methods, including the widely used NDVI-based compositing. We also recorded the per-pixel DOY as meta-data. Although S2 data have a 5-day revisit frequency, data gaps still exist due to atmospheric conditions. To fill the data gaps, we applied a per-pixel temporal interpolation method following Griffiths et al. (2019). For a given month, if the pixel value was missing due to cloud, reflectance values of the preceding and subsequent months were used to calculate the interpolated reflectance value. For the rare cases when clouds are present persistently for two months, we did not conduct interpolation. Our monthly compositing generated spatially coherent data with clearly revealed phenological variations (Fig. 5).

2.4.3. Reprojection and tiling

We designed a geographic degree tiling system to efficiently manage the large spatial datasets. The study area was divided into seamless $1^{\circ} \times$

1° tiles in geographic latitude/longitude projection with WGS84 datum. Each tile was named by the latitude and longitude coordinates of the lower-left corner. To approximate the 10-m resolution of Sentinel-2 pixels, we chose $0.0001^{\circ} \times 0.0001^{\circ}$ (approximately 11 m at the Equator) as the spatial resolution of the degree tile. After removing small islands and sea areas, we conducted reprojection for the 1223 Sentinel-2 Universal Transverse Mercator (UTM) tiles, resulting in 1119 degree tiles (Fig. 6).

2.4.4. Spectral indices and other inputs for classification

To prepare data for crop mapping, we derived spectral indices in addition to surface reflectance (Table 2). We also downloaded NASADEM topographic data from the EarthData Search as inputs for classification (<https://search.earthdata.nasa.gov/>, accessed on July 05, 2021). Organized in $1^{\circ} \times 1^{\circ}$ tiles at 1 arc sec resolution, NASADEM is a collection of Digital Elevation Model (DEM) and associated products derived from the Shuttle Radar Topography Mission (SRTM), the Ice,

Table 2
Spectral indices used for crop classification.

Spectral Indices	Formula	Sentinel-2 bands	Reference
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	$(\text{B8} - \text{B4}) / (\text{B8} + \text{B4})$	Tucker (1979)
Normalized Difference Water Index (NDWI)	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$	$(\text{B8} - \text{B11}) / (\text{B8} + \text{B11})$	Gao (1996)
Normalized Difference Built-up Index (NDBI)	$(\text{SWIR1} - \text{NIR}) / (\text{SWIR1} + \text{NIR})$	$(\text{B11} - \text{B8}) / (\text{B11} + \text{B8})$	Zha et al. (2003)
Normalized Difference Snow Index (NDSI)	$(\text{Green} - \text{SWIR1}) / (\text{Green} + \text{SWIR1})$	$(\text{B3} - \text{B11}) / (\text{B3} + \text{B11})$	Salomonson and Appel (2004)
Normalized Difference Tillage Index (NDTI)	$(\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$	$(\text{B11} - \text{B12}) / (\text{B11} + \text{B12})$	Deventer et al. (1997)
Normalized Burn Ratio (NBR)	$(\text{NIR} - \text{SWIR2}) / (\text{NIR} + \text{SWIR2})$	$(\text{B8} - \text{B12}) / (\text{B8} + \text{B12})$	Key and Benson (2006)
Enhanced Vegetation Index (EVI)	$2.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + 6 * \text{Red} - 7.5 * \text{Blue} + 1)$	$2.5 * (\text{B8} - \text{B4}) / (\text{B8} + 6 * \text{B4} - 7.5 * \text{B2} + 1)$	Huete et al. (2002)
Soil-Adjusted Vegetation Index (SAVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.5) * (1 + 0.5)$	$(\text{B8} - \text{B4}) / (\text{B8} + \text{B4} + 0.5) * (1 + 0.5)$	Huete (1988)
Red-Edge Position (REP)	$705 + 35 * (((\text{RE3} + \text{Red}) / 2) - \text{RE1}) / (\text{RE2} - \text{RE1}))$	$705 + 35 * (((\text{B7} + \text{B4}) / 2) - \text{B5}) / (\text{B6} - \text{B5}))$	Frampton et al. (2013)
Inverted Red-Edge Chlorophyll Index (IRECI)	$(\text{RE3} - \text{Red}) / (\text{RE1} / \text{RE2})$	$(\text{B7} - \text{B4}) / (\text{B5} / \text{B6})$	Frampton et al. (2013)

Cloud, and land Elevation Satellite (ICESat) and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Buckley et al., 2020). We used the elevation, slope and aspect layers and resampled them to $0.0001^\circ \times 0.0001^\circ$ to match our S2 degree tiles. In the end, input data for crop classification included 70 monthly spectral reflectance (i.e., 10 bands over 7 months), 70 monthly spectral indices and 3 topographic metrics.

2.5. National-scale crop classification

The goal of national crop classification was to produce a crop map for which the area of maize and soybean derived by pixel counting matched the sample-based area estimates at the national scale.

2.5.1. Training data generation

We generated training labels from the 3-m maize and soybean block maps (see section 2.2 above). We spatially aggregated the maps from 3-m to 10-m resolution and derived percent maize and soybean cover per 10-m pixel. We then applied a 50% cover threshold to derive binary maize/non-maize and soybean/non-soybean classes at 10-m resolution. We randomly selected 1% of the pixels from each block as the national training data.

2.5.2. Binary classification with Random Forests

Random Forests (RF) is an ensemble machine learning algorithm that works by constructing multiple trees through random training and variable selection (Breiman, 2001). Due to its high accuracy, effective computation, robustness to noise, and ability to deal with high-dimensional, non-linear data, the RF classifier has been widely used in remote sensing (Belgiu and Drăguț, 2016; Pal, 2007). Binarization is an efficient technique to approach multi-class problems by decomposing a multi-class problem into binary classification cases with binary classifiers trained and predicted separately, and combining binary results for final classification (Galar et al., 2011; Lorena et al., 2009). In this study, we used RF classifiers for binary classifications for maize (RF-Maize) and soybean (RF-Soybean) independently with the one-vs-all decomposition strategy (Adnan and Islam, 2015), and subsequently aggregated the results to a multi-class map.

Major parameters of a RF classifier include: (1) `n_estimators` (the number of trees in a forest); (2) `max_depth` (the maximum depth of a tree); (3) `min_samples_split` (the minimum number of samples required to split an internal node); (4) `min_samples_leaf` (the minimum number of

samples required to be at a leaf node); (5) `max_features` (the number of features to consider when looking for the best split) (Pedregosa et al., 2011). We fine-tuned the model through random search followed by a grid search using 70% of the national training data as the training dataset (Probst et al., 2019). 30% of the data was used as the test dataset, and the out-of-bag error was used as the scoring function. After obtaining optimal hyper-parameters, we retrained the model with all training data to make predictions. All the training, test, fine-tuning and prediction were implemented for RF-Maize and RF-Soybean independently, using the scikit-learn Python package (Pedregosa et al., 2011).

The prediction of a RF classifier is determined by the per-pixel class probability, which is assigned as the mean value across all trees. We also calculated the standard deviation of the class probability across all trees to represent the per-pixel prediction uncertainty using the following equation:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (7)$$

where σ is the standard deviation of class probability; N is the number of trees in a forest; x_i is the predicted probability from the i^{th} tree; μ is the mean probability across all trees.

2.5.3. Probability threshold optimization to match sample-based area estimates

Since RF-Maize and RF-Soybean produced per-pixel class probabilities independently, we generated a combined probability layer by assigning each pixel to the class with the highest probability. For pixels with equal probabilities for maize and soybean, the one with the smaller standard deviation was chosen. We also created a crop mask (maize or soybean) and recorded the crop information and the probability uncertainty information associated with the combined probability layer. For all maize pixels, following the method in Song et al. (2017) and Tyukavina et al. (2022), we found the optimal probability threshold across all strata such that the map-based maize area through pixel counting matched the sample-based area estimate (see Section 2.3). Likewise, the optimal probability threshold for soybean was also found. Then, these two thresholds were applied to the combined probability and the associated crop mask to generate the final classification map with both maize and soybean classes.

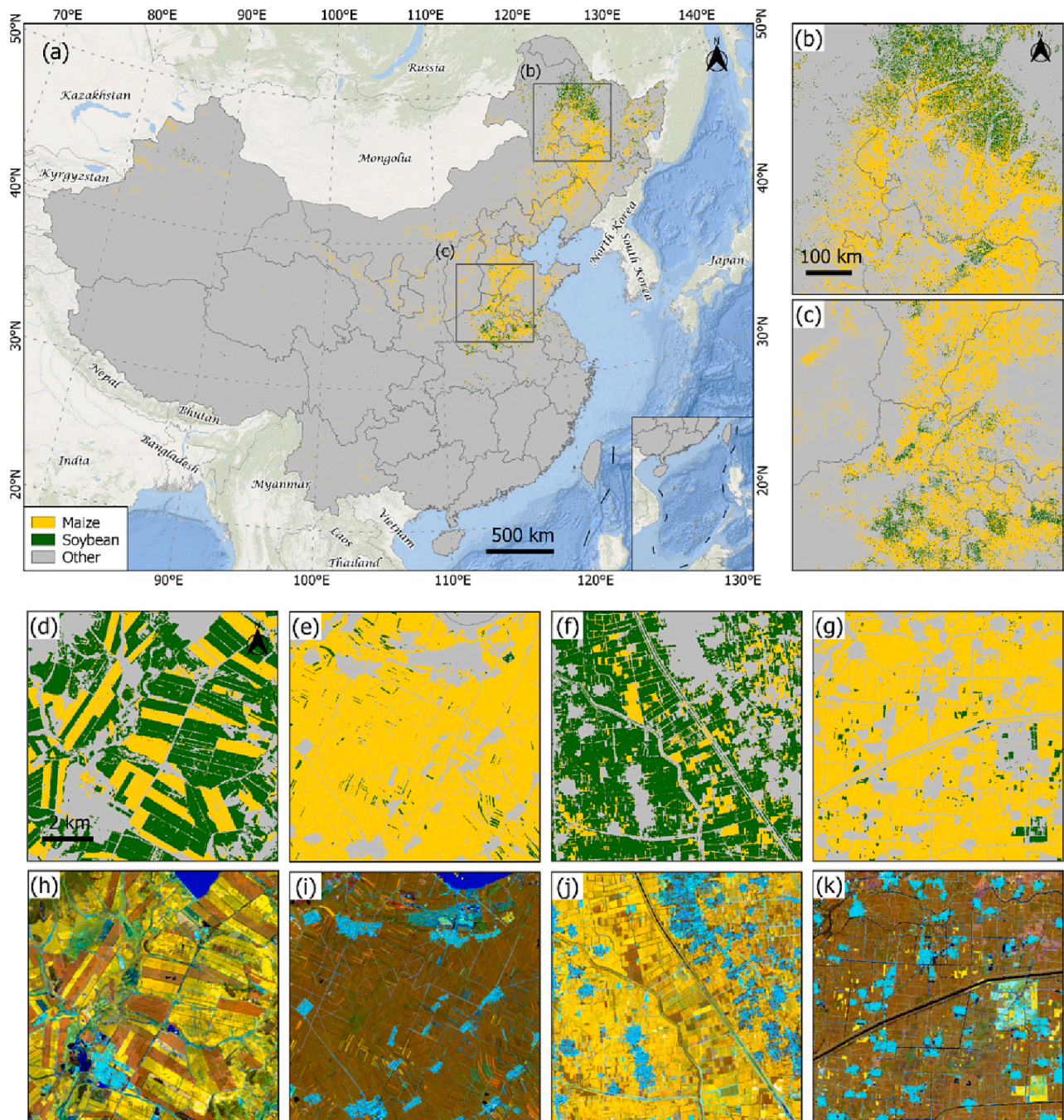


Fig. 7. National crop map at 10-m resolution over China in 2019. (a) national crop map for maize and soybean. (b–c) crop classifications in two major agricultural regions of Northeast China Plain and North China Plain displayed at the same scale (600 km \times 600 km). (d–g) crop classification displayed at a finer scale (10 km \times 10 km) in different regions. (h–k) Sentinel-2 monthly composites in August (R: NNIR, G: SWIR1, B: Red) corresponding to panels (d–g), respectively. The center coordinates of (b–g) are (124.540° E, 47.117° N), (114.834° E, 35.571° N), (125.541° E, 48.741° N), (125.035° E, 45.342° N), (113.850° E, 33.758° N), (116.674° E, 37.451° N), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. Crop map evaluation

2.6.1. Direct map validation

We used the 450 pixels (i.e., SSUs) from the probability sample to assess the accuracy of the national map. Accuracy metrics including overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) were estimated. Of the 450 SSUs, 406 were pure pixels and 44 pixels were mixed pixels, according to the per-pixel crop cover fractions derived from in-situ assessment (an example is shown in Fig. S2). We estimated two sets of accuracy metrics, one using the 406 pure pixels and one using all 450 pixels to investigate the impact of mixed pixels on

classification accuracy. For the mixed pixels, we labeled a pixel as maize if the maize proportion within the pixel exceeded 50%.

2.6.2. 10-fold cross-validation

Although only 1% of the area within mapped PSU blocks was used for training, the training data could still overlap, at a very low likelihood, with the validation data (10 pixels per PSU), as the validation data were also located within the first-stage sample PSUs. To supplement the direct map validation, we conducted a 10-fold cross-validation with the objective of keeping the validation data completely separate from the training data. We randomly divided the 50 blocks into 10 equal-size

Table 3

Accuracy metrics for the national crop map using field sample data as reference. A total of 450 pixels from the two-stage cluster sample were visited in the field, consisting of 406 pure pixels and 44 mixed pixels. SE: standard error.

Field sample	Class	Users' accuracy % (SE)	Producers' accuracy % (SE)	Overall accuracy % (SE)	Validation method
Pure pixels (n = 406)	Maize	97.2 (1.5)	86.7 (3.2)	95.4 (0.9)	Direct validation
	Soybean	84.4 (9.3)	64.1 (13.5)		
	Others	95.4 (1.1)	98.9 (0.5)		
All pixels (n = 450)	Maize	93.9 (2.5)	79.2 (3.6)	91.8 (1.2)	
	Soybean	63.6 (12.1)	61.9 (11.8)		
	Others	92.6 (1.3)	96.9 (1.0)		
Pure pixels	Maize	97.7 (1.3)	85.5 (3.6)	94.6 (1.1)	
	Soybean	71.6 (12.2)	44.6 (23.4)		
	Others	94.5 (1.3)	98.9 (0.5)		
All pixels	Maize	97.9 (1.1)	77.4 (3.8)	91.2 (1.3)	10-fold cross-validation
	Soybean	47.0 (11.6)	40.4 (10.5)		
	Others	91.4 (1.4)	97.5 (0.8)		

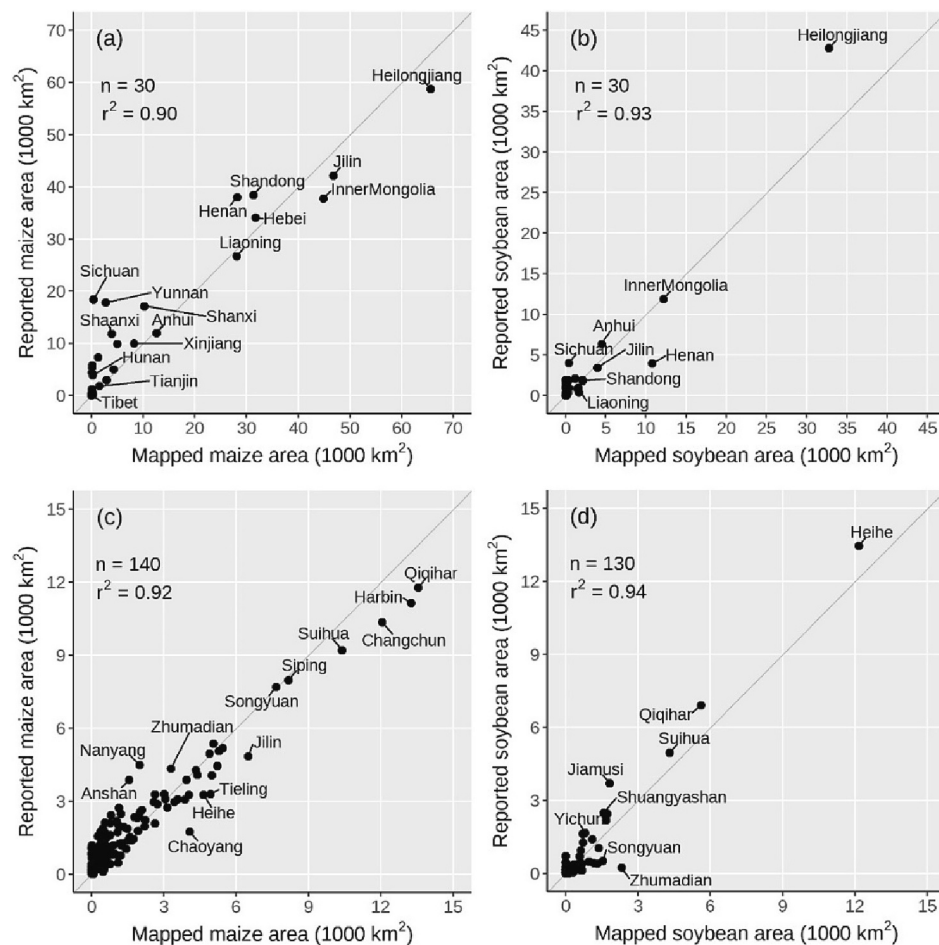


Fig. 8. Comparison between mapped crop area and planted crop area reported by official statistics at provincial and prefectural levels. (a) maize at provincial level; (b) soybean at provincial level; (c) maize at prefectural level; (d) soybean at prefectural level.

groups. For each of the 10 iterations, we used 9 groups of data (45 blocks) to train a RF classifier, and used the SSU pixel locations of the remaining group (50 pixels in 5 blocks) to extract the RF predicted classes as the validation data. After the 10 iterations, we created a dataset consisting of 500 pixels with paired RF predictions and field labels. 450 of the 500 pixels were selected from the probability sampling design, and thus, they were used to derive accuracy estimates using the stratified estimators applied to the direct map validation as described above.

3. Results

3.1. National crop area estimates

Using field data from the probability sample and a regression estimator, we estimated the national maize area in China in 2019 to be 330,609 km² with a standard error of 34,109 km², and the national soybean area to be 78,107 km², with a standard error of 12,969 km².

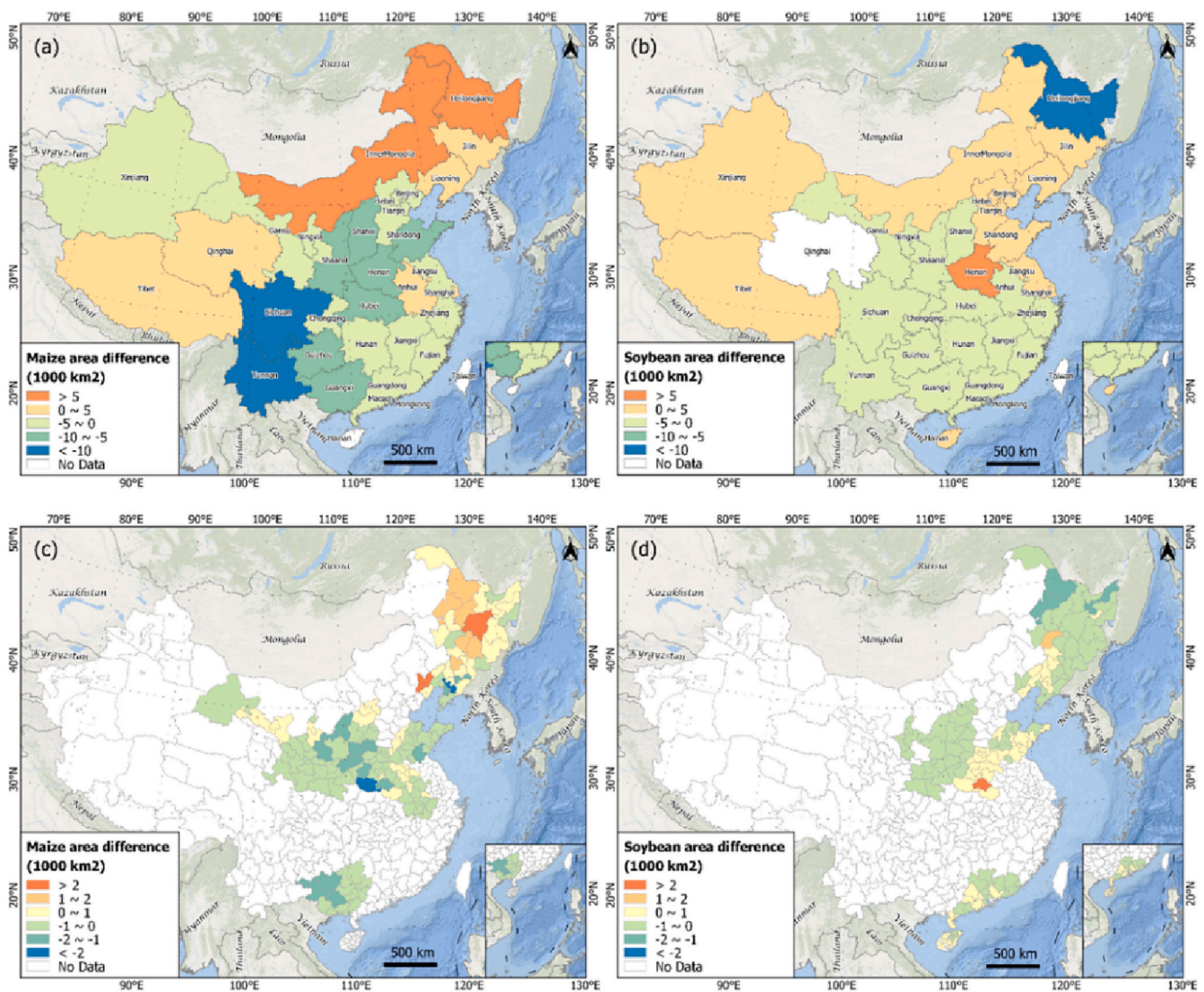


Fig. 9. Difference between mapped crop area and planted crop area reported by official statistics at provincial and prefectural level. (a) maize at provincial level; (b) soybean at provincial level; (c) maize at prefectural level; (d) soybean at prefectural level.

3.2. National crop map and accuracy

Spatial patterns of maize and soybean are shown on the 10-m crop map (Fig. 7). Maize has a much wider distribution than soybean. The dominant maize cultivation locations include the Northeast China Plain (Fig. 7b), including Heilongjiang, Jilin, Liaoning and several municipalities in eastern Inner Mongolia, and the North China Plain (Fig. 7c), including Hebei, Shandong and Henan. Soybeans were mainly clustered in Heilongjiang and Henan. The 10-m resolution of the map allowed visualization of large homogeneous fields (Fig. 7d), as well as small fragmented fields (Fig. 7e, f, and g).

Our crop map has high classification accuracies (Table 3). For the direct validation based on the probability sample field data, the map achieved OA of 91.8%, and UA of 93.9%, PA of 79.2% for maize, all with small standard errors. UA (63.6%) and PA (61.9%) for soybean were both lower than those of maize and with larger standard errors. The largest change resulting from including the mixed pixels was a decrease in UA of soybean from 84% to 64%, indicating that much of the commission error of soybean was associated with mixed pixels. For maize, the 10-fold cross-validation generated comparable accuracy results as the direct map validation. For soybean, the 10-fold cross-validation accuracies were lower than their counterparts from direct validation.

For example, for the “all pixels” case, UA decreased from 64% for the direct validation to 47% for the cross-validation, and PA decreased from 62% to 40%. The larger difference for soybean can be attributed to two factors. First, separating training and accuracy assessment sample pixels by sample blocks has been demonstrated to diminish the impact of having training and validation data in close proximity (Friedl et al., 2010; Zhen et al., 2013). Second, soybeans were mainly clustered in Heilongjiang and Henan (Fig. 7). Due to the highly clustered pattern and that our stratification was not specifically tailored for soybean, training data used in creating the random forests models in the cross-validation could become unrepresentative when the blocks in Heilongjiang and Henan were excluded. Thus, accuracies from cross-validation were lower than direct validation, where a more representative training set from all the blocks were used.

3.3. Comparison with government reports

We compared our area estimates with official statistics at national, provincial and prefectural levels. For the national and provincial levels, we collected the 2019 official agricultural statistics from the National Bureau of Statistics website. This dataset was compiled by provincial bureau of statistics, with crop information including planted and

harvested area, yield and production for major food crops. We also downloaded the crop statistics at the prefectural level from yearbooks in 2019 from the China Economic and Social Data Platform on National Knowledge Infrastructure (CNKI). The reported national maize harvested area was 412,800 km² and soybean harvested area was 84,230 km² in 2019. Compared to the official statistics, our estimate of maize area (330,609 km² ± 34,109 km²) was about 20% lower and our estimate of soybean area (78,107 km² ± 12,969 km²) was about 7% lower. Correspondence between our estimates and official statistics is improved if we use the upper bounds of our estimates. That is, if we incorporate the uncertainty of the area estimates via either a 68% (±SE) or 95% (±2SE) confidence interval, the upper bound of the confidence interval is closer to the official statistics.

We calculated the maize and soybean areas at provincial and prefectural scales using our map, and compared them with official statistics. Note that the reported area at the provincial and prefectural scales is planted area rather than the harvested area at the national scale. For maize, 30 provincial and 140 prefectural records were available, and for soybean 30 provincial and 130 prefectural records were available. Quantitative analyses were conducted by calculating the r^2 , the root-mean-square-difference (RMSD) and the mean difference (MD). At the provincial level, the area comparison showed close relationships between mapped and reported areas for both maize ($r^2 = 0.90$, RMSD = 6024 km², MD = 2714 km²) and soybean ($r^2 = 0.93$, RMSD = 2461 km², MD = 564 km²) (Fig. 8a, b). However, larger differences were observed in southwestern provinces, particularly Sichuan and Yunnan, with our map-based area lower than the reported area (Fig. 8a, Fig. 9a). This could be caused by the predominant smallholder cropping systems in mountainous areas with considerable field-level heterogeneity (Song et al., 2016). Maize in southwest China is often one of multiple mixed crops grown on small tracts of land for self-sufficiency. Remote sensing data of 10-m resolution are capable of mapping accurately crop versus no crop in such mixed cropping systems but have limited success when mapping specific crop types within the mixture. Besides, for some areas in southwestern China (Sichuan and Yunnan), cloud-free Sentinel-2 observations in the peak growing season are reduced due to continuous

cloudy conditions, which can affect mapping accuracy (see Fig. S3). Maize area in major growing provinces in the Northeast and North China Plains generally agreed well with official statistics. Our mapped soybean areas also showed close agreement with official statistics for most provinces including the southern provinces, although a notable difference was observed in Heilongjiang (Fig. 8b, Fig. 9b). At the prefectural level, for both maize ($r^2 = 0.92$, RMSD = 765 km², MD = 181 km²) and soybean ($r^2 = 0.94$, RMSD = 395 km², MD = 38 km²), our mapped areas and official statistics showed higher agreement than the provincial-level results (Fig. 8c, d), although our mapped areas were generally higher in the Northeast China Plain and lower in the North China Plain (Fig. 9c, d).

4. Discussion

4.1. Feature importance in random forests models

The trained RF-Maize model and RF-Soybean model had nearly identical structure (Table 4). The fine-tuned hyper-parameters all have the same or similar values.

We evaluated the feature importance based on the mean decrease in impurity in RF models. The ten most important features accounted for 56.4% and 49.8% of cumulative importance, for RF-Maize and RF-Soybean, respectively (Fig. 10). For maize classification, the most heavily used feature was the Red-Edge-Position (REP) in August, followed by Red-Edge-1 (RE1, 705 nm) in August and SWIR2 (2190 nm) in June. For soybean classification, the most important features were Red-Edge-2 (RE2, 740 nm) in August and September, and NIR (842 nm) in August. Overall, the red-edge (RE) bands during peak growing season (i.e., August) were more critical (Fig. 11), and the NIR and SWIR2 bands also showed great value for classification. These findings are consistent with a recent study that revealed the comparative utility of various spectral bands from Landsat and Sentinel-2 for maize and soybean mapping in the US (Song et al., 2021b).

As an intermediate output of RF, the probability layer represents the predicted likelihood of a pixel being a crop type. The mapped crop area is a function of the probability threshold applied (Fig. 12). The default threshold is 0.5. Adjusting the threshold provided us with flexibility to produce maps to match a desirable area target, which could be the sample-based area estimate, or the lower or upper confidence bounds. Here, by finding the empirical thresholds that minimized the difference between map-based area and sample-based area (0.50 for maize and 0.37 for soybean), we produced the national map that could be used to derive unbiased area estimates at the national scale through pixel counting.

Table 4
Fine-tuned hyper-parameters for RF-Maize and RF-Soybean models.

Parameter	RF-Maize	RF-Soybean
Max_depth	46	46
Max_features	91	89
Min_samples_leaf	8	8
Min_samples_split	18	19
N_estimators	100	100

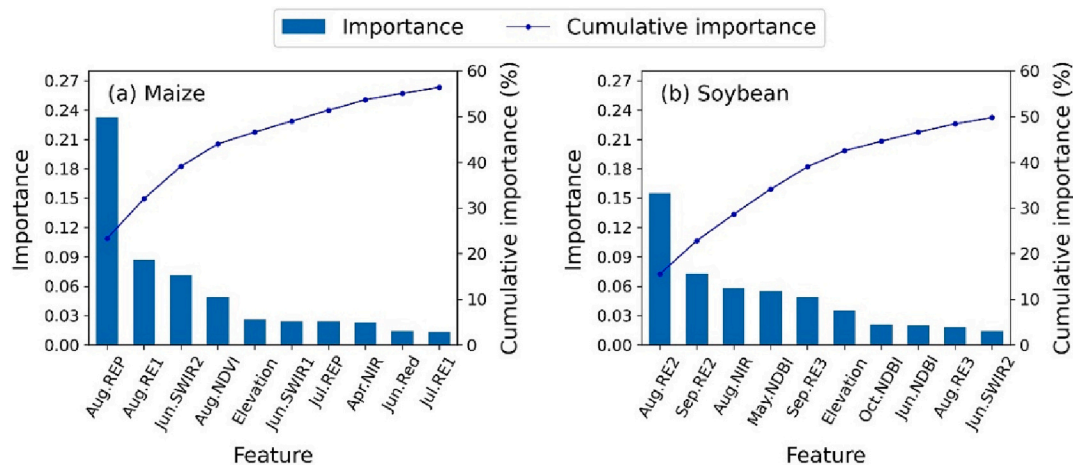


Fig. 10. Ten most important features for RF-Maize (a) and RF-Soybean (b) models. Features are named by corresponding months followed by Sentinel-2 bands or spectral indices.

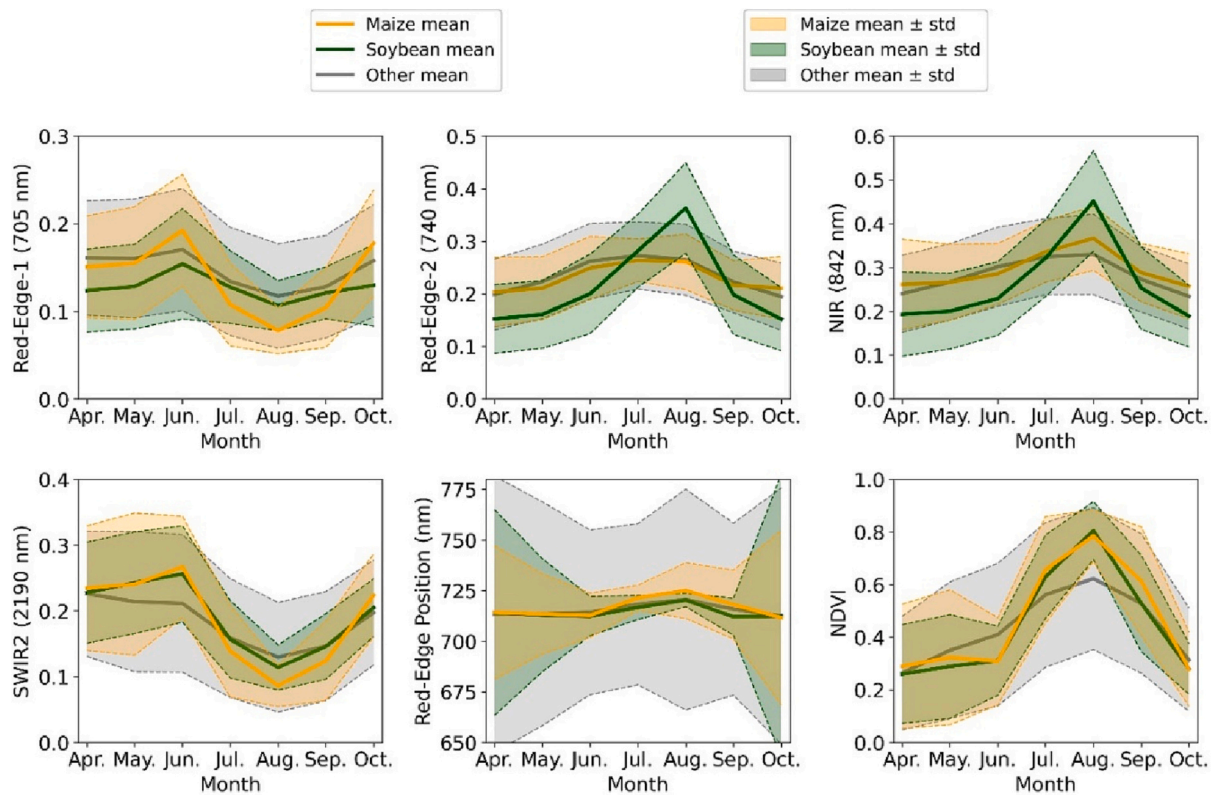


Fig. 11. Temporal profiles of important features for maize, soybean and other classes. The plots were made using training pixels of the RF models. Thick lines represent the mean value and shaded areas represent one standard deviation.

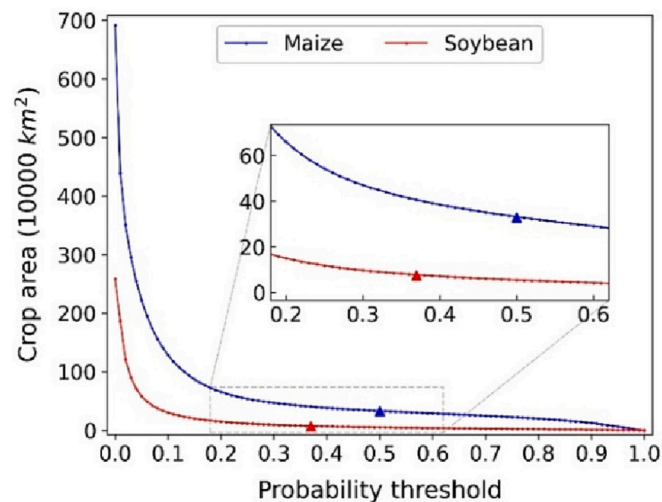


Fig. 12. Map-based maize and soybean areas as a function of thresholds of random forest-derived class probability. The empirical thresholds for which map-based areas match sample-based area estimates are symbolized by triangles.

4.2. Stratified cluster sampling for national crop area estimation and mapping

We adopted a stratified, two-stage, cluster sampling design to collect ground reference data for maize and soybean area estimation and mapping in China. Cluster sampling is a cost-effective method for regional (Khan et al., 2018), national (King et al., 2017; Song et al., 2017) and continental (Song et al., 2021a) scale applications when reference data are collected by ground visit (i.e., field data). One of the critical parameters of the sampling design is the size of the primary sampling unit, which should be adaptive to specific geographical

context in practice. For industrial agricultural regions with large field sizes (e.g., the US and South America), a PSU of 20 km × 20 km has proven to be highly effective (King et al., 2017; Song et al., 2021a; Song et al., 2017). In China, where smallholder agriculture is the dominant form and logistical constraints are greater, we have demonstrated in this study that the PSU size of 10 km × 10 km is effective. However, choosing the PSU size for any given application involves consideration of trade-offs between cost and precision (Kish, 1965, sec. 8.3). This PSU size is consistent with the study of Khan et al. (2018) which focused on smallholder wheat mapping in Punjab, Pakistan. Our sampling procedure shows great potential to be applicable in other regions dominated by

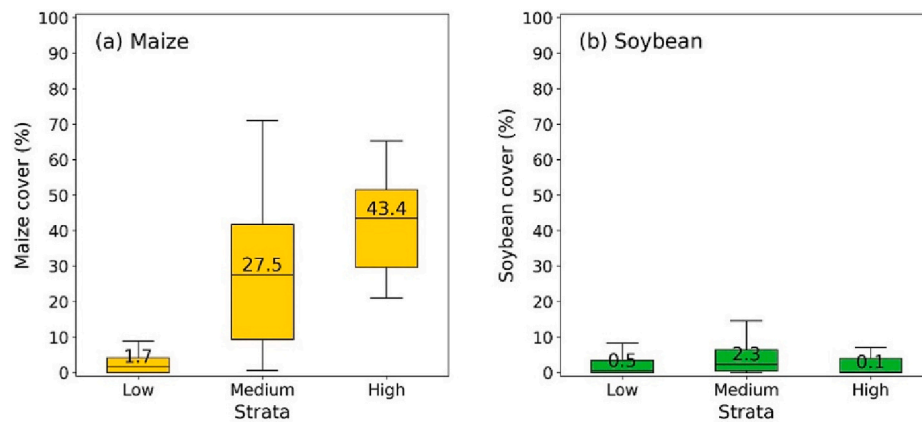


Fig. 13. Per-stratum maize and soybean cover statistics. Crop cover statistics were derived using the 3-m resolution block maps. The median values are labeled in the boxplot.

smallholder agriculture, although the PSU size might need to be adapted accordingly.

Stratified random sampling can generate smaller standard errors for land cover/land use area estimates than simple random sampling and systematic random sampling of the same sample size (Broich et al., 2009). Stratification allows us to target crops that are unevenly distributed across a country more efficiently. In this study, we generated a Landsat-based maize indicator map for stratification, as maize was our primary interest. Using strata not specifically tailored to soybean resulted in the standard errors of soybean area and accuracy estimates being larger than what might have been achieved with strata tailored to soybean (Fig. 13). These results reveal the challenge of how to choose strata and the sample size allocation to strata to achieve precise estimates for multiple crop types if their spatial distributions do not substantially overlap within the study area.

4.3. Strength of high-performance computing

Developing a 10-m resolution crop map for a large country such as China involves processing a large volume of satellite data, which requires considerable resources for computing and data storage. One of the options is cloud computing platforms such as Google Cloud, Amazon Web Service and Microsoft Azure. Google Earth Engine (GEE) is a particularly popular tool in the remote sensing research community (Gorelick et al., 2017). An alternative option is high-performance computing (HPC) clusters that are increasingly available in institutions. Our 10-m wall-to-wall mapping was conducted on the HPC cluster at Texas Tech University. Combining supercomputing resources and Message Passing Interface (MPI) programming, the major steps of our mapping approach (see Fig. 1), including satellite data pre-processing, random forests training and prediction were implemented with high efficiency on the HPC clusters. The production of our 10-m national crop map consumed about 684,367 CPU hours in total. The 66,750 original SAFE folders in 1223 UTM tiles, the reprojected data in $1119\ 1^\circ \times 1^\circ$ tiles, and the mapping results accounted for about 285 terabytes in storage. The parallel environment of the HPC clusters enabled us to carry out national-scale crop mapping within a manageable time frame. Our established workflow can be potentially implemented in an operational setting for repeated mapping, which is critical for agricultural monitoring.

5. Conclusions

We developed an operational workflow to generate a 10-m resolution maize and soybean map over China. The workflow consists of two parallel yet interactive streams of activities – field data from a probability sample and wall-to-wall satellite mapping. Their interactions are:

(1) the historical satellite-based crop map is used to construct strata for the sampling design implemented to collect field (ground) reference data; (2) collected ground data are used to train machine learning models for mapping; (3) intermediate mapping outputs are used as the auxiliary variable in the unbiased sample-based regression estimator; (4) sample-based area estimates are used as constraints to final mapping so that national estimates based on the map match the area estimates from the sample; and (5) field data from the probability sample are used to validate the satellite-based crop maps. This interactive workflow can generate two sets of internally consistent results: unbiased sample-based crop area estimates with associated uncertainty estimates, and accurate maps with area matching that of the sample-based estimates. Our national-scale maize and soybean area estimates can be produced at the end of the growing season, independent of government statistics. Given that Sentinel-2 data are being acquired continuously and commercial high-resolution imagery is becoming increasingly abundant, our established workflow could be operationally applied to generate crop maps in China and other countries.

CRediT authorship contribution statement

Haijun Li: Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Xiao-Peng Song:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Supervision, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing. **Matthew C. Hansen:** Conceptualization, Methodology, Resources, Writing – review & editing, Funding acquisition. **Inbal Becker-Reshef:** Conceptualization, Writing – review & editing, Funding acquisition. **Bernard Adusei:** Formal analysis. **Jeffrey Pickering:** Formal analysis, Data curation. **Li Wang:** Investigation. **Lei Wang:** Investigation. **Zhengyang Lin:** Investigation. **Viviana Zalles:** Methodology. **Peter Potapov:** Methodology, Software. **Stephen V. Stehman:** Methodology, Writing – review & editing. **Chris Justice:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The satellite data used in this study are accessible online: 1) the Landsat data were derived from GLAD Landsat ARD (<https://glad.umd.edu/ard/home>); 2) the Sentinel-2 data were downloaded from Google

Cloud Platform (<https://cloud.google.com/storage/docs/public-datasets/sentinel-2>); 3) the PlanetScope data were acquired from Planet (<https://www.planet.com/>). The 10-m maize and soybean map can be viewed and downloaded from the Google Earth Engine App (<https://glad.earthengine.app/view/china-crop-map>).

Acknowledgements

This research was supported by the University of Maryland and Texas Tech University. We thank Qing Ying, Yuelin Zhang, Wensheng Duan, and Ke Yang for assisting field data collection. The authors acknowledge the High Performance Computing Center (HPCC) at Texas Tech University for providing computational resources that have contributed to the research results reported within this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2023.113623>.

References

- Adnan, M.N., Islam, M.Z., 2015. One-vs-all binarization technique in the context of random forest. In: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium, pp. 385–390.
- Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleira, D., Ferraz Rodrigo, P.D., Lebourgeois, V., Lelong, C., Simões, M., Verón Santiago, R., 2018. Remote sensing and cropping practices: a review. *Remote Sens.* 10, 99.
- Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Bellow M.E., 1994. Application of satellite data to crop area estimation at the county level. US Department of Agriculture, National Agricultural Statistics Service, Research Division, STB Research Report Number STB-94-02.
- Benedetti, R., Bee, M., Espa, G., Piersimoni, F., 2010. *Agricultural Survey Methods*, first ed. John Wiley & Sons Ltd, Chichester.
- Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasm, S., Hostert, P., 2022. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and landsat 8 data for Germany. *Remote Sens. Environ.* 269, 112831.
- Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens. Environ.* 240, 111685.
- Boryan, C., Yang, Z., Di, L., Hunt, K., 2014. A new automatic stratification method for U.S. Agricultural area sampling frame construction based on the cropland data layer. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 7, 4317–4327.
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, cropland data layer program. *Geocarto Int.* 26, 341–358.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Broich, M., Stehman, S.V., Hansen, M.C., Potapov, P., Shimabukuro, Y.E., 2009. A comparison of sampling designs for estimating deforestation from Landsat imagery: a case study of the Brazilian legal Amazon. *Remote Sens. Environ.* 113, 2448–2454.
- Buckley, S.M., Agram, P.S., Belz, J.E., Crippen, R.E., Gurrola, E.M., Hensley, S., 2020. *NASADEM* (accessed on October 14, 2022). https://lpdaac.usgs.gov/documents/592/NASADEM_User_Guide_V1.pdf.
- Carfagna, E., Gallego, F.J., 2006. Using remote sensing for agricultural statistics. *Int. Stat. Rev.* 73, 389–404.
- Clauss, K., Yan, H., Kuenzer, C., 2016. Mapping paddy rice in China in 2002, 2005, 2010 and 2014 with MODIS time series. *Remote Sens.* 8, 434.
- Coluzzi, R., Imbrenda, V., Lanfredi, M., Simoniello, T., 2018. A first assessment of the Sentinel-2 level 1-C cloud mask product to support informed surface analyses. *Remote Sens. Environ.* 217, 426–443.
- d'Andrimont, R., Verhegghen, A., Meroni, M., Lemoine, G., Strobl, P., Eiselt, B., Yordanov, M., Martínez-Sánchez, L., van der Velde, M., 2021. LUCAS copernicus 2018: earth-observation-relevant in situ data on land cover and use throughout the European Union. *Earth Syst. Sci. Data* 13, 1119–1133.
- d'Andrimont, R., Yordanov, M., Martínez-Sánchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Jobges, C., Lemoine, G., van der Velde, M., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. *Sci. Data* 7, 352.
- d'Andrimont, R., Verhegghen, A., Lemoine, G., Kempeneers, P., Meroni, M., van der Velde, M., 2021. From parcel to continental scale – a first European crop type map based on Sentinel-1 and LUCAS copernicus in-situ observations. *Remote Sens. Environ.* 266, 112708.
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaut, T., Savinaud, M., Udrou, C., Valero, S., Bégué, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 221, 551–568.
- Deventer, A.P.V., Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogramm. Eng. Remote Sens.* 63, 87–93.
- Fao, 2015. *Handbook on Master Sampling Frames for Agricultural Statistics*. In: FAO, Rome, pp. 43–106.
- FAO, 2018. In: *World Programme for the Census of Agriculture 2020*, volume 2 Operational Guidelines. FAO Statistical Development Series 16. FAO, Rome, p. 145.
- FAOSTat, 2021. FAO Stat. FAO, Rome (accessed on January 29, 2023). <http://www.fao.org/faostat>.
- Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., Smith, A., Davidson, A., Shang, J., Jarvis, I., 2013. In: *AAFC annual crop inventory. 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, IEEE, Fairfax, VA, USA, pp. 270–274.
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92.
- Friedl, M.A., Woodcock, C., Gopal, S., Muchoney, D., Strahler, A.H., Barker-Schaaf, C., 2010. A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data. *Int. J. Remote Sens.* 21, 1073–1077.
- Frolking, S., Qiu, J., Boles, S., Xiao, X., Liu, J., Zhuang, Y., Li, C., Qin, X., 2002. Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China. *Glob. Biogeochem. Cycles* 16, 1091.
- Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F., 2011. An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* 44, 1761–1776.
- Gallego, F.J., 2004. Remote sensing and land cover area estimation. *Int. J. Remote Sens.* 25, 3019–3047.
- Gao, B.-C., 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266.
- Gomez-Chova, L., Camps-Valls, G., Calpe-Maravilla, J., Guanter, L., Moreno, J., 2007. Cloud-screening algorithm for ENVISAT/MERIS multispectral images. *IEEE Trans. Geosci. Remote Sens.* 45, 4105–4118.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.
- Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* 220, 135–151.
- Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Cheng, F., Zhuang, H., Zhang, J., Tao, F., 2021a. NESEA-Rice10: high-resolution annual paddy rice maps for northeast and Southeast Asia from 2017 to 2019. *Earth Syst. Sci. Data* 13, 5969–5986.
- Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Zhang, J., Li, Z., 2021b. 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.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309.
- Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sens.* 8, 166.
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sens.* 7, 12356–12379.
- Irish, R.R., 2000. Landsat-7 automatic cloud cover assessment algorithms for multispectral, hyperspectral, and ultraspectral imagery. *Int. Soc. Opt. Eng.* 4049, 348–355.
- Jiang, Y., Lu, Z., Li, S., Lei, Y., Chu, Q., Yin, X., Chen, F., 2020. Large-scale and high-resolution crop mapping in China using Sentinel-2 satellite imagery. *Agriculture* 10, 433.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google earth engine. *Remote Sens. Environ.* 228, 115–128.
- Johnson, D.M., 2019. Using the Landsat archive to map crop cover history across the United States. *Remote Sens. Environ.* 232, 111286.
- Key, C.H., Benson, N., 2006. Landscape assessment (LA). Sampling and analysis methods. In: Lutes, D.C., Keane, R.E., Caratti, J.F., Key, C.H., Benson, N.C., Sutherland, S., Gangi, L.J. (Eds.), *FIREMON: Fire effects monitoring and inventory system. Integration of standardized field data collection techniques and sampling design with remote sensing to assess fire effects*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. LA1–LA51.
- Khan, A., Hansen, M., Potapov, P., Adusei, B., Pickens, A., Krylov, A., Stehman, S., 2018. Evaluating Landsat and RapidEye data for winter wheat mapping and area estimation in Punjab, Pakistan. *Remote Sens.* 10, 489.
- Khan, A., Hansen, M.C., Potapov, P., Adusei, B., Stehman, S.V., Steininger, M.K., 2021. An operational automated mapping algorithm for in-season estimation of wheat area for Punjab, Pakistan. *Int. J. Remote Sens.* 42, 3833–3849.
- King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.-P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M., Hansen, M.C., 2017. A multi-resolution approach to

- national-scale cultivated area estimation of soybean. *Remote Sens. Environ.* 195, 13–29.
- Kish, L., 1965. *Survey Sampling*. In: John Wiley & Sons, New York, p. 644.
- Kussul, N., Lavreniuk, M., Shumilo, L., 2020. Deep recurrent neural network for crop classification task based on Sentinel-1 and Sentinel-2 imagery. In: *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa, HI, USA, pp. 6914–6917.
- Li, X.-Y., Li, X., Fan, Z., Mi, L., Kandakji, T., Song, Z., Li, D., Song, X.-P., 2022. Civil war hinders crop production and threatens food security in Syria. *Nat. Food.* 3, 38–46.
- Lorena, A.C., de Carvalho, A.C.P.L.F., Gama, J.M.P., 2009. A review on the combination of binary classifiers in multiclass problems. *Artif. Intell. Rev.* 30, 19–37.
- Luo, Y., Zhang, Z., Zhang, L., Han, J., Cao, J., Zhang, J., 2022. 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.
- Manjunath, K.R., More, R.S., Jain, N.K., Panigrahy, S., Parihar, J.S., 2015. Mapping of rice-cropping pattern and cultural type using remote-sensing and ancillary data: a case study for south and southeast Asian countries. *Int. J. Remote Sens.* 36, 6008–6030.
- NICFI, 2022. *Norway's International Climate and Forests Initiative satellite data program* (accessed on October 20, 2022). <https://www.planet.com/nicfi/>.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57.
- Pal, M., 2007. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26, 217–222.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pflugmacher, D., Rabe, A., Peters, M., Hostert, P., 2019. Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. *Remote Sens. Environ.* 221, 583–595.
- Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A., Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sens.* 12, 426.
- Potapov, P., Turubanova, S., Hansen, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens. Environ.* 115, 548–561.
- Preidl, S., Lange, M., Doktor, D., 2020. Introducing APiC for regionalised land cover mapping on the national scale using sentinel-2A imagery. *Remote Sens. Environ.* 240, 111673.
- Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and tuning strategies for random forest. *Wiley Interdisc. Rev. Data Min. Knowl. Discov.* 9, e1301.
- Rufin, P., Frantz, D., Ernst, S., Rabe, A., Griffiths, P., Özdoğan, M., Hostert, P., 2019. Mapping cropping practices on a national scale using intra-annual Landsat time series binning. *Remote Sens.* 11, 232.
- Salomonson, V.V., Appel, I., 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sens. Environ.* 89, 351–360.
- Song, X.-P., Hansen, M.C., Potapov, P., Adusei, B., Pickering, J., Adami, M., Lima, A., Zalles, V., Stehman, S.V., Di Bella, C.M., Conde, M.C., Copati, E.J., Fernandes, L.B., Hernandez-Serna, A., Jantz, S.M., Pickens, A.H., Turubanova, S., Tyukavina, A., 2021a. Massive soybean expansion in South America since 2000 and implications for conservation. *Nat. Sustain.* 4, 784–792.
- Song, X.-P., Huang, W., Hansen, M.C., Potapov, P., 2021b. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* 3, 100018.
- Song, X.-P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. 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.
- Song, Y., Zhang, Y., Song, X., Swiderska, K., 2016. Smallholder farming systems in Southwest China: Exploring key trends and innovations for resilience. IIED, London (accessed on November 14, 2022). <https://www.iied.org/14664iied>.
- Stehman, S.V., 2000. Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sens. Environ.* 72, 35–45.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150.
- Tyukavina, A., Potapov, P., Hansen, M.C., Pickens, A.H., Stehman, S.V., Turubanova, S., Parker, D., Zalles, V., Lima, A., Kommareddy, I., Song, X.-P., Wang, L., Harris, N., 2022. Global trends of forest loss due to fire from 2001 to 2019. *Front. Remote Sens.* 3, 825190.
- Van Tricht, K., Gobin, A., Gilliams, S., Piccard, I., 2018. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: a case study for Belgium. *Remote Sens.* 10, 1642.
- Wang, S., Di Tommaso, S., Deines, J.M., Lobell, D.B., 2020. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Sci. Data.* 7, 307.
- Weigand, M., Staab, J., Wurm, M., Taubenböck, H., 2020. Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data. *Int. J. Appl. Earth Obs. Geoinf.* 88, 102065.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. *Remote Sens. Environ.* 236, 111402.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free access to Landsat imagery. *Science* 320, 1011.
- Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., 2013. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. *Int. J. Digital Earth.* 7, 113–137.
- Wu, Y., Xi, X., Tang, X., Luo, D., Gu, B., Lam, S.K., Vitousek, P.M., Chen, D., 2018. Policy distortions, farm size, and the overuse of agricultural chemicals in China. *Proc. Natl. Acad. Sci.* 115, 7010–7015.
- Xiao, W., Xu, S., He, T., 2021. Mapping paddy rice with Sentinel-1/2 and phenology-, object-based algorithm—a implementation in Hangjiahui Plain in China using GEE platform. *Remote Sens.* 13, 990.
- You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y., Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Sci. Data.* 8, 41.
- Zalles, V., Hansen, M.C., Potapov, P.V., Parker, D., Stehman, S.V., Pickens, A.H., Parente, L.L., Ferreira, L.G., Song, X.-P., Hernandez-Serna, A., Kommareddy, I., 2021. Rapid expansion of human impact on natural land in South America since 1985. *Sci. Adv.* 7, eabg1620.
- Zalles, V., Hansen, M.C., Potapov, P.V., Stehman, S.V., Tyukavina, A., Pickens, A., Song, X.P., Adusei, B., Okpa, C., Aguilar, R., John, N., Chavez, S., 2019. Near doubling of Brazil's intensive row crop area since 2000. *Proc. Natl. Acad. Sci.* 116, 428–435.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24, 583–594.
- Zhang, X., Wu, B., Ponce-Campos, G., Zhang, M., Chang, S., Tian, F., 2018. Mapping up-to-date paddy rice extent at 10 m resolution in China through the integration of optical and synthetic aperture radar images. *Remote Sens.* 10, 1200.
- Zhen, Z., Quackenbush, L.J., Stehman, S.V., Zhang, L., 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *Int. J. Remote Sens.* 34, 6914–6930.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94.