CropLayer: A high-accuracy 2-meter resolution cropland mapping dataset for China in 2020 derived from Mapbox and Google satellite imagery using data-driven approaches

Hao Jiang [1], Mengjun Ku [1], Xia Zhou [1], Qiong Zheng [2], Yangxiaoyue Liu [3], Jianhui Xu [1], Dan Li [1],
5   Chongyang Wang [1], Jiayi Wei [1], Jing Zhang [1], Shuisen Chen [1], Jianxi Huang [4, 5]

[1] Guangdong Engineering Technology Research Center of Remote Sensing Big Data Application, Key Lab of Guangdong for Utilization of Remote Sensing and Geographical Information System, Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou 510070, China
[2] Department of Geomatics Engineering, School of Traffic & Transportation Engineering, Changsha University of Science &
10   Technology, Changsha 410114, China
[3] State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[4] Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, China
[5] College of Land Science and Technology, China Agricultural University, Beijing, China

15   *Correspondence to*: Xia Zhou (zhouxia@gdas.ac.cn)

Abstract. Accurate and detailed cropland maps are essential for agricultural planning, resource management, and food security, particularly in countries like China, where agricultural productivity is high but resources are limited. Despite the availability of several medium-to-high-resolution satellite-based cropland maps, significant discrepancies in area estimates and spatial distribution persist, limiting their utility. This study proposes a data-driven framework for cropland mapping that

20   leverages 2 m High Resolution (HR) imagery from Mapbox and Google. The framework consists of three main stages: First, national imagery is partitioned into 0.05 °×0.05 ° blocks for efficient parallel computation. An Image Quality Assessment (IQA) using ResNet models is performed on both sources to address the challenge of missing image acquisition metadata. Second, a robust cropland identification model integrates Mask2Former for precise segmentation and XGBoost for error assessment, facilitating iterative improvements through active learning. Finally, a novel integration strategy combines four

25   feature groups—Geography, IQA, Region Property, and Consistency—using XGBoost to merge the datasets into a unified cropland layer, named Croplayer. The Croplayer dataset achieves an overall mapping accuracy of 88.73%, with 30 out of 32 provincial units reporting area estimates within ±10% of official statistics. In contrast, only 1 to 9 provinces from seven other existing datasets meet the same accuracy standard. The results highlight Croplayer's potential for applications such as crop yield estimation and agricultural structure analysis, offering a reliable tool for addressing agricultural and food security

30   challenges.

1 Introduction

China, with its extensive agricultural resources and long-standing tradition of intensive farming, occupies a pivotal role in global agriculture. Despite possessing less than 7% of the world's cropland and only 5% of utilizable freshwater resources, China produces a quarter of the world's food, sustaining more than 22% of the global population (Kuang et al., 2022). This significant contribution underscores China's importance to both domestic and international food and nutritional security (Zhang et al., 2022). Currently, China is the world's leading producer of grains, cotton, and vegetables. However, recent uncertainties in global food security including adverse climate changes, catastrophic events, and increasing competition and trade tensions, have amplified the need for enhanced agricultural monitoring (Piao et al., 2010; Kang and Eltahir, 2018).

Accurate cropland maps are indispensable for estimating crop areas and yields, as well as for assessing losses from disasters. Remote sensing serves as a highly effective method for producing cropland maps over large geographic areas (Wu et al., 2023). Over the years, researchers have developed hundreds of cropland maps at both national and global scales, many of which are derived from land use and land cover (LULC) products (Cui et al., 2024). The resolution and accuracy of these maps have progressively advanced, corresponding to improvements in satellite capabilities. Spatial resolutions have evolved from coarse (1 km, 500 m) to medium (30 m, 10 m) (Cui et al., 2024a), and more recently to high resolutions approaching 1 m (Li et al., 2023).

Except for the early coarse-resolution data, medium-resolution data is currently widely utilized. Products based on 30-m Landsat data include: 1) the China land cover dataset (CLCD) (Yang and Huang, 2021), 2) Finer Resolution Observation and Monitoring of global land cover (FROM) (Gong et al., 2019), 3) Global land-cover product with Fine Classification System 2020 (FCS30) (Zhang et al., 2021), and 4) Globeland30-2020 (GL30) (Jun et al., 2014). Products based on 10 m Sentinel data include: 5) WorldCover from European Space Agency (ESA) (Zanaga et al., 2021) and 6) Land cover v2 from the Environmental Systems Research Institute (ESRI) (Karra et al., 2021). High-resolution products based on Google Earth imagery is relatively new, i.e. 7) National-scale land-cover map of China (SinoLC) (Li et al., 2023).

Nevertheless, existing cropland datasets for China exhibit certain limitations, including:

1) Challenges in cropland area estimation: Existing cropland maps for China exhibit significant discrepancies when compared with official statistics. For instance, our preliminary analysis indicates that in Guangdong Province, the reported cropland area in certain datasets is up to three times larger than the official statistics (Jiang et al., 2024). Such substantial discrepancies undermine the reliability of these datasets and hinder their application in agricultural planning and policy-making (Cui et al., 2024b).

2) Inadequate representation of cropland diversity: Current cropland datasets for China reveal significant discrepancies in their ability to accurately represent cropland diversity. These discrepancies manifest as overestimations in some provinces and underestimations in others, largely influenced by the diverse climatic conditions, terrain, and agricultural practices across China's vast landscape. The country's cropland includes a wide range of field types, such as terraced fields, barrage fields, flatland fields, gully fields, weir fields, and strip fields, each of which requires tailored sampling strategies to ensure

accurate representation (Zhang et al., 2020). However, the lack of comprehensive coverage for these diverse field types often
65   results in biased estimates, leading to significant overestimations or underestimations of specific cropland categories.

3) Limitations in spatial details for complex terrains: The spatial resolution of commonly utilized satellite data sources, such as Landsat (Hansen and Loveland, 2012) and Sentinel (Bontemps et al., 2015; Qiu et al., 2022), presents significant limitations when applied to regions with complex terrains. These data sources, with resolutions suited to large, contiguous agricultural fields in regions like the United States, Europe, and Northern China, struggle to capture the intricate spatial
70   details of Southern China's fragmented and rugged landscapes (Qiu et al., 2024). In these areas, characterized by mountainous terrain and small, irregular crop fields typically 0.04 ha, the resolution is insufficient to distinguish between croplands and non-agricultural features, such as ridges, roads, or highways (Liu et al., 2020). This lack of spatial detail not only leads to misclassification but also undermines the overall accuracy of cropland datasets. To address these challenges, higher-resolution imagery and more advanced classification algorithms are essential for accurately mapping such complex
75   agricultural environments.

Toward the above three requirements, we improve the cropland identification method from the following three ideas:

1) Leveraging Open Access High-Resolution Imagery

Accessing national-scale high-resolution (HR) data often incurs substantial costs, prompting the academic community to turn to more accessible alternatives. Freely available Google Satellite imagery offers a resolution of at least 0.3 m in most
80   global regions, while Mapbox provides a cost-effective option with moderate fees. These HR datasets have proven effective in applications such as land cover analysis and object detection, offering spatial detail that far surpasses Sentinel and Landsat data. However, they come with notable limitations, including missing metadata on satellite sources, acquisition dates, and the lack of temporal dynamics essential for crop phenology monitoring. Furthermore, they are limited to RGB bands, restricting their spectral information content.
85   To address these shortcomings, this study proposes an image quality assessment method to infer missing metadata and leverage human interpretation to gain prior knowledge of the imagery. By integrating this information, the study aims to selectively utilize Google Satellite and Mapbox imagery to enhance cropland extraction. This approach balances the superior spatial detail of HR imagery with the need for ancillary information to ensure reliable analysis.

2) Advancing Cropland Mapping with Deep Learning-Based Semantic Segmentation
90   Traditional pixel-wise classification methods relying on shallow machine learning algorithms struggle to capture the complexity of cropland patterns. In contrast, deep learning-based semantic segmentation methods excel by leveraging hierarchical feature extraction to analyze shape, texture, color, and contextual information. Early approaches primarily employed Convolutional Neural Networks (CNNs), which effectively capture local patterns. However, recent advancements favor Transformer-based architectures, which are adept at modeling long-range dependencies and extracting complex
95   features, making them particularly suitable for cropland mapping.

This study will evaluate the performance of widely used models, including PSPNet (Zhao et al., 2017), PIDNet(Xu et al., 2023), Segformer (Xie et al., 2021), and Mask2Former (Cheng et al., 2022), to determine the most accurate method for

cropland identification. By comparing these models, the research aims to capitalize on the strengths of Transformers while addressing the diverse and intricate characteristics of croplands, thereby advancing the precision and scalability of cropland

100   mapping.

3) Optimizing Sample Selection through Active Learning

Traditional sampling techniques, such as regular or stratified random sampling, often rely on cropland area distribution for stratification. While effective for general land cover classification, these methods are less suited for semantic segmentation, where capturing object diversity is more critical than sampling similar objects repeatedly.

105   Active Learning has emerged as an effective approach to significantly reduce the number of samples that need to be labelled, often by several orders of magnitude (Kaijage et al., 2024). Active learning provides a more efficient alternative by selectively labeling the most informative samples, reducing redundant labeling and minimizing manual workloads (Mittal et al., 2023). This study aims to establish criteria for sample selection and determine iteration termination within the active learning process. Using provincial cropland statistics as a benchmark, an active learning framework will be constructed to

110   enhance model performance while minimizing labeling costs, ensuring diversity in sample selection.

In this study, we propose the use of 2.4 m resolution (level-16) Google and Mapbox satellite imagery to generate a nationwide cropland distribution map for China. The methodology comprises 3 key steps: image quality assessment, cropland extraction, and results enhancement, all guided by data-driven strategies. The resulting cropland data is then evaluated against 7 publicly available cropland datasets, focusing on metrics such as accuracy, area, and consistency.

115   Additionally, we analyze the spatial distribution of cropland in China, offer recommendations for practical applications of this dataset, and discuss potential directions for future research.

2 Study area and data

2.1 Study area

The study area encompasses the entirety of China, located in the eastern part of the Eurasian continent, with a total land area

120   of approximately 9.6 million km². Cropland in China covers about 127.87 million ha, ranking it as the third-largest country globally in terms of cropland area. The country's topography is highly varied, with mountains, plateaus, and hills accounting for approximately 67% of the land area, while basins and plains make up the remaining 33% (Liu et al., 2024). This complex terrain, combined with diverse climatic conditions, results in substantial regional variability in agricultural practices and cropland types (Figure 1).

125   China's primary cropland types include dry land, paddy fields, and irrigated land (Zhang et al., 2024). Flat terrains in basins and plains, such as the Northeast Plain, North China Plain, the middle and lower Yangtze River Plains, and the Chengdu Basin, are well-suited for large-scale mechanized farming due to concentrated land resources. Conversely, mountainous, plateau, and hilly regions, particularly in Southwest China, feature rugged terrain and fragmented arable land, posing significant challenges to large-scale agricultural development (Xinyu et al., 2022).

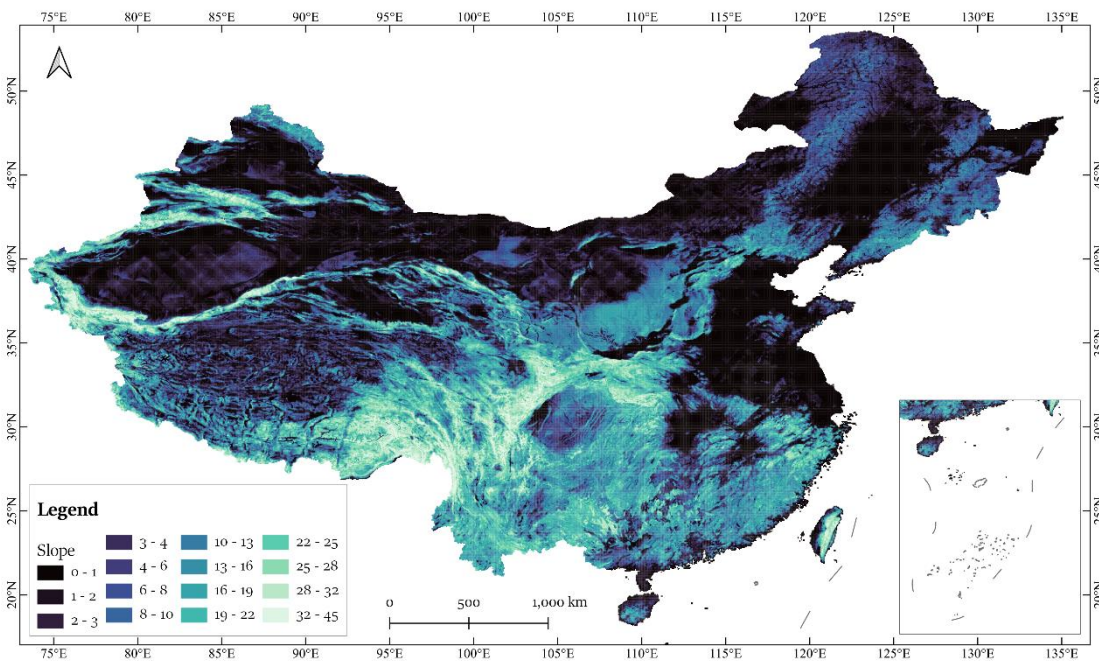



Figure 1: Slope distribution in the Study Area of China.

2.2 Data

2.2.1 High Resolution Remote Sensing Imagery and Data Structure

The Croplayer dataset was generated using HR satellite imagery from Mapbox and Google. These imagery datasets are continuously updated from various sources, including commercial providers, NASA, and USGS. According to the Google (http://mts0.googleapis.com) and Mapbox (https://api.mapbox.com/v4/mapbox.satellite) satellite services, global HR satellite imagery is provided as RGB color patches with a resolution of 256×256 pixels and is accessible via a Web Map Service (WMS) API. The imagery used in this study was accessed between August 2022 and December 2023.

Although the datasets lack key metadata, such as sensor type, viewing angle, and atmospheric conditions, they undergo extensive radiometric and geometric corrections before being made publicly available. These corrections ensure data reliability and suitability for various applications, including cropland mapping.

In this study, the HR imagery was organized into image blocks measuring $0.05\degree \times 0.05\degree$ in WGS 1984 geographic coordinates, with each block comprising mosaicked image patches. A total of 389,777 image blocks were generated, covering the entire extent of China for both datasets. The imagery was utilized at level-16, corresponding to a spatial resolution of approximately 2.4 m per pixel, which is sufficient for capturing the geometry and structure of cropland.

2.2.2 DEM Data

Given the strong correlation between cropland distribution and topographic features, this study utilized Digital Elevation Models (DEM) to improve cropland mapping accuracy. Due to the lack of high-resolution DEM data, it was not included in the initial cropland extraction phase. Instead, 30 m resolution DEM data from the Shuttle Radar Topography Mission

150 (SRTM) was incorporated during the post-processing stage. Key topographic features including slope, ruggedness, and roughness were derived from the DEM using the gdaldem tool (https://gdal.org/en/latest/programs/gdaldem.html).

2.2.3 Samples 1: Coverage type classification

All image blocks were categorized into five distinct coverage types: Planting, Non-Planting, Cloudy, Snow/Ice, and Nodata, represented by the color codes green, yellow, white, light grey, and black, respectively. Due to the relative scarcity of blocks

155 classified as Cloudy, Snow/Ice, and Nodata, an active learning strategy was adopted for sampling, enabling the selection of blocks with more distinct features. This approach improved class discriminability compared to systematic sampling methods.

Definitions for each coverage type are outlined in Table 2. From the datasets, a total of 8,084 image blocks were curated from various regions of China, ensuring comprehensive coverage. These blocks were further divided into 6,465 training samples and 1,619 validation samples to provide a robust and diverse dataset for model training and evaluation.

160 **Table 1:** Definition of image coverage type.

Coverage type	Image color	Explanation
Planting	Green	Cropping season
Non-Planting	Yellow	Areas not currently cultivated i.e. fallow cropland or the Gobi Desert
Cloudy	White	Thick cloud coverage > 10%
Snow/Ice	Light Grey	Snow/Ice coverage > 20%
Nodata	Black	No data coverage > 10%

<table>
<tr><td>Planting</td><td>Non-Planting</td><td>Cloudy</td><td>Snow/Ice</td><td>Nodata</td></tr>
</table>

a) b) c) d) e) f) g) h) i) j)

Figure 2: Examples of 5 Coverage type. The HR remote-sensing images in the figure are from © Mapbox 2023 and © Google Maps 2023.

2.2.4 Samples 2: Cropland extraction

165 Cropland samples were manually labelled using HR imagery in shapefile format via QGIS software. These samples were not collected in a single phase but were gradually acquired through an active learning strategy (Mittal et al., 2023). Initially, the model was trained on a preliminary set of samples. The predicted results were then visually inspected, and additional samples were collected from areas with the most significant errors to improve model accuracy.

The spatial distribution of the labelled samples is illustrated in Figure 3, while examples of the labeled data are provided

170 in Table 2.

Table 2: Typical representatives of various types of cropland and non-cropland. The HR remote-sensing images in the figure are from © Mapbox 2023 and © Google Maps 2023.

		Definition	South region	Central and Northeast region	Northwest region
Cropland	Paddy cropland	Cropland used for growing aquatic crops such as rice and lotus root. (Including arable land where aquatic and dry crops are rotated.			
	Irrigated cropland &Dry cropland	Artificially watered and cropland without irrigation facilities (including bananas and pineapples).			
	Terraces	Arable land with a strip terrace or wavy section built in the contour direction on hilly hillside land.			

	Fallow land	Cropland that is not planted with crops for the time being.			
	Green house	Land on which plastic greenhouses or mulch, etc., with no substantial impact on the tillage layer.			
	Cropland with scattered trees	Cropland with scattered trees at intervals, including fruit trees, mulberry trees or other trees.			
Confusing types of non-cropland	Orchards	Land planted with fruit trees with more than 50 per cent cover.			
	Woodland	Land on which trees, bamboos and shrubs grow.			
	Bare land	Land with an earthy surface and essentially no vegetation cover; or land with a rocky, gravelly surface, which covers ≥ 70 % of the land.			

Grassland	Land dominated by herbaceous plants.	

To enhance model robustness, the cropland samples were designed to encompass diverse cropping practices and crop types. The dataset consists of 157,395 polygons, most of them representing an individual cropland, distributed across 366 image blocks (0.05 ° × 0.05 ° each), collectively referred to as positive cropland samples (Figure 3).

In addition to these positive samples, the dataset includes 761 negative cropland samples blocks devoid of cropland polygons. These negative samples are particularly important as they often feature areas resembling cropland but are not actual cropland. Their inclusion strengthens the model's ability to differentiate between cropland and non-cropland regions, thereby improving classification accuracy.

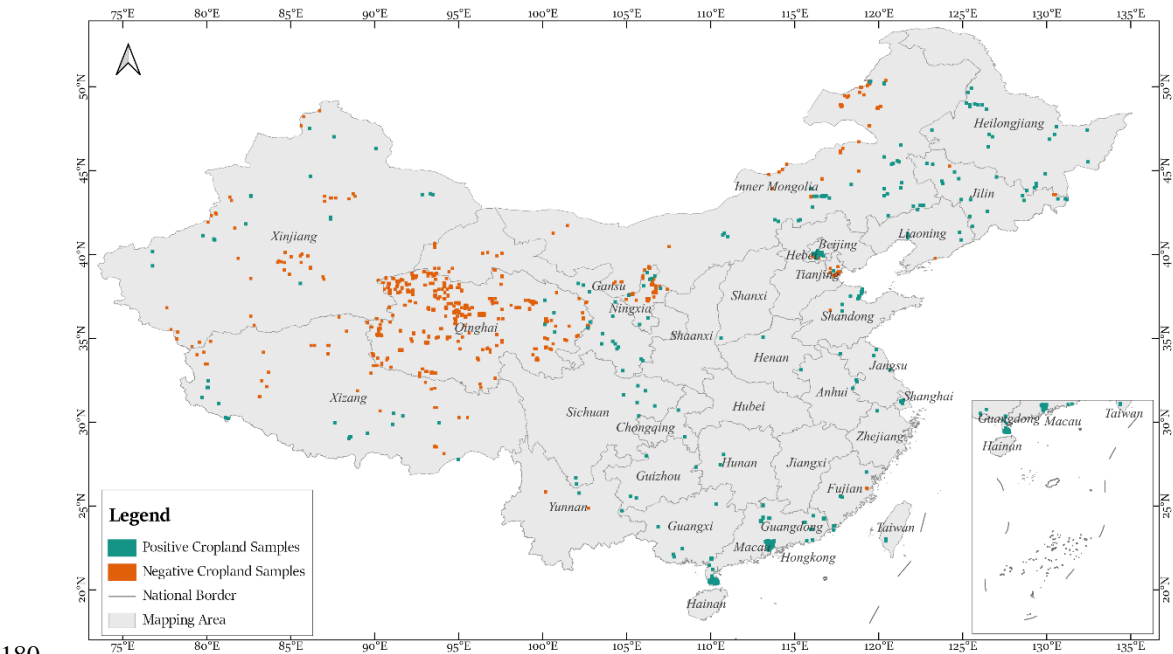



Figure 3: Distribution of positive and negative cropland samples.

2.2.5 Samples 3: Error assessment and results integration

A data-driven approach is adopted for both error assessment and results integration during the cropland extraction process. Systematic grid sampling at intervals of 0.5 ° longitude and latitude is applied, generating a total of 3,891 image blocks. The process involves two key steps:

1) Error Assessment:

Independent sample interpretation was conducted for cropland results derived from Mapbox and Google datasets, yielding 7,782 samples categorized into four groups:

a) True Positive (TP): Correctly identified croplands.

b) False Positive (FP): Non-cropland areas mistakenly identified as cropland.

c) Noise: Very small extraction errors.

d) Artifacts: Misclassified areas due to imagery inconsistencies or processing errors.

2) Results Integration:

The same set of 3,891 blocks was classified into four integration categories:

a) Neither Source: Blocks where neither dataset identifies cropland.

b) Only Mapbox: Blocks with cropland identified exclusively by Mapbox.

c) Only Google: Blocks with cropland identified exclusively by Google.

d) Merging Both: Blocks where cropland is identified in both datasets, integrated for improved accuracy.

To ensure comprehensive evaluation and reliable result integration, a systematic sampling and categorization approach was adopted. For model classification, the total sample set was stratified and randomly split, allocating 80% for training and 20% for testing. This method enhanced model accuracy and improved the reliability of the final cropland dataset.

2.2.6 Statistical Cropland Area

Provincial-level cropland area statistics for China are sourced from the Third National Land Survey (TNLS), a comprehensive initiative that involved detailed assessment and verification of land resources, including the area and distribution of cultivated land. The project employed professional investigators and a hierarchical inspection system to ensure data accuracy and reliability, making the statistics a robust reference for comparative analysis and validation. This study focuses on the 32 provincial units in China with cropland areas exceeding 100 km², excluding Hong Kong and Macau. Additionally, Taiwan's cropland statistics are not part of the TNLS and are instead derived from alternative sources (https://agrstat.moa.gov.tw/sdweb/public/indicator/Indicator.aspx). Notably, the definition of cropland in Taiwan differs significantly from TNLS, as it includes orchards.

2.2.7 Existing cropland data

To evaluate the performance of the proposed method and the public cropland datasets in identifying cropland inundation in mountainous regions, this study involves seven land cover (LC) datasets including:

1) China land cover dataset (CLCD) (Yang and Huang, 2021),

2) WorldCover from European Space Agency (ESA) (Zanaga et al., 2021),

3) Land cover v2 from the Environmental Systems Research Institute (ESRI) (Karra et al., 2021),

4) Finer Resolution Observation and Monitoring of global land cover from Tsinghua University (FROM) (Gong et al., 2019),

5) Global land-cover product with Fine Classification System 2020 from Aerospace Information Innovation Institute,
220 Chinese Academy of Sciences (FCS30) (Zhang et al., 2021),

6) Globeland30-2020 from National Geomatics Center of China (GL30) (Jun et al., 2014),

7) National-scale land-cover map of China (SinoLC) (Li et al., 2023).

The 7 LC data are all produced based on Sentinel or Landsat data, human-interpreted samples and machine learning methods.

225 These datasets are derived from Sentinel or Landsat data, combined with human-interpreted samples and machine learning methods. Table 1 summarizes the datasets. All datasets, except FROM (representing data from 2017), correspond to the year 2020. Although there is a three-year gap, differences in provincial cropland changes may still be evident. The declared accuracy for the 7 datasets ranging from 72% to 85.72%.

Table 3: Summary of 7 Land cover data

Cropland data name	Issuer	Data source	Spatial resolution	Classification method	Cropland definition	Declared accuracy
CLCD 2020	Wuhan University	Landsat	30m	Random Forest	Cropland	79.31%
ESA WorldCover 2020	European Space Agency	Sentinel-1 Sentinel-2	10m	Gradient boosting decision tree algorithm (CatBoost)	Cropland	74.40%
ESRI Land Cover 2020	Impact Observatory, Inc., Washington, D.C.	Sentinel-2	10m	Deep learning (UNet)	Crops	85%
FROM_GLC 10 2017	Tsinghua University	Landsat Sentinel-2	10m	Maximum Likelihood Classification,Decision Tree,Random Forest,SVM	Cropland	72.43%
GLC_FCS30 2020	Aerospace Information Research Institute, Chinese Academy of Sciences	Landsat	30m	Random Forest	Rainfed cropland, Irrigated cropland	82.5%,

| Globeland30-2020 | China National Geographic Information Centre | Landsat HJ-1 GF-1 | 30m | Maximum Likelihood Classification, SVM, Thresholds method | Cultivated Land | 85.72% |
| SinoLC | China university of geosciences | Google Earth imagery | 1m | L2HNet | Cropland | 73.61 % |

230

3 Methods

This study proposes a data-driven framework for cropland mapping using 2 m HR imagery from Mapbox and Google. The framework consists of three key stages (Figure 4): 1) Pre-processing: The national imagery is divided into 0.05 °×0.05 ° blocks to facilitate efficient parallel processing (Yang et al., 2019). Image quality assessment (IQA) is conducted on both

235 imagery sources using ResNet models (He et al., 2016), despite the lack of acquisition metadata. 2) Cropland Extraction: An active learning-based model for cropland identification is developed, utilizing Segformer for cropland segmentation and XGBoost (Chen and Guestrin, 2016) for error assessment. 3) Post-processing: A merging strategy is applied using XGBoost, combining four feature groups including Geography, IQA, Region Property, and Consistency to integrate the two datasets into the final cropland data, referred to as Croplayer. The performance of Croplayer is evaluated by comparing it with seven

240 publicly available cropland datasets and statistical area information.

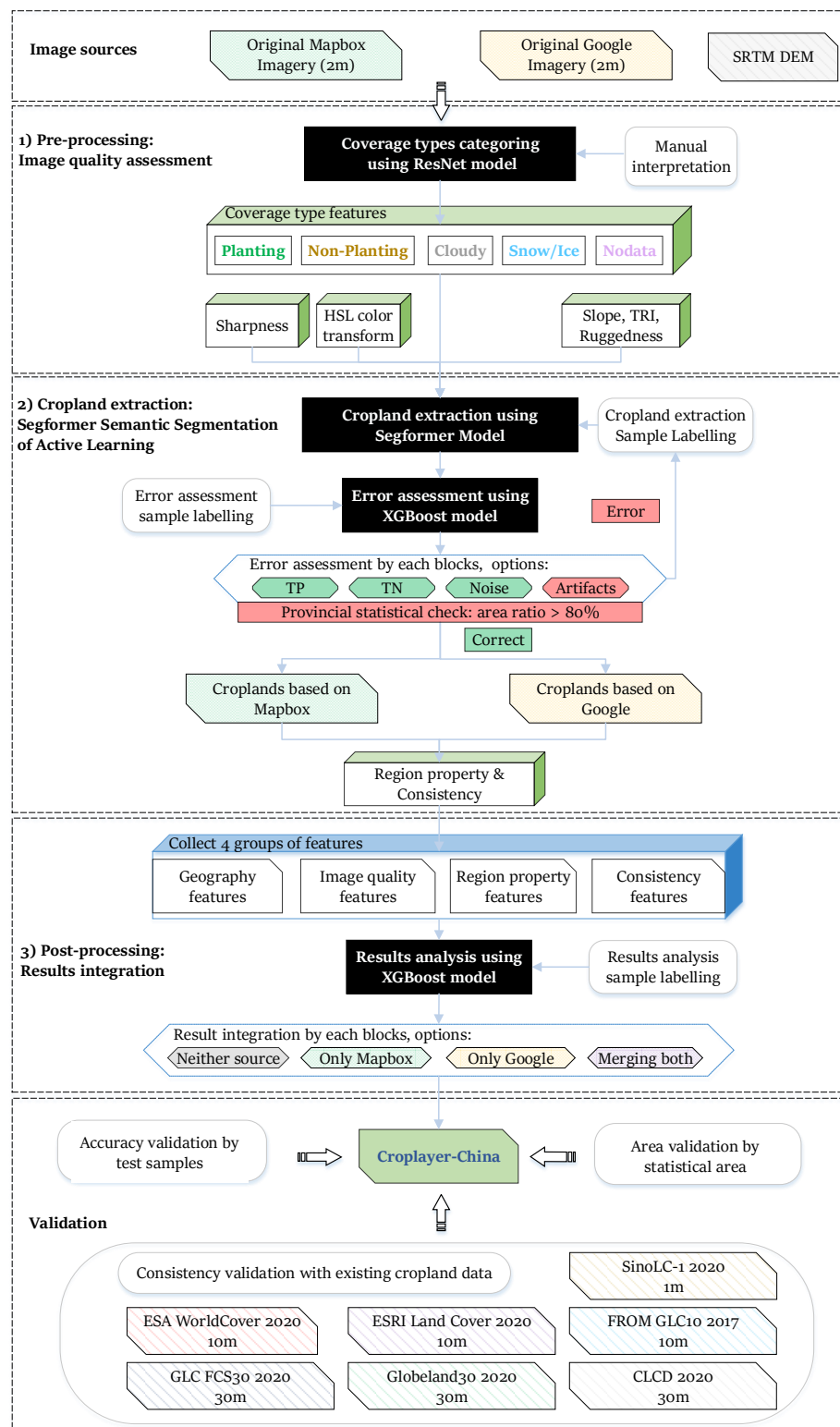



Figure 4: Flowchart of cropland mapping.

3.1 Pre-processing: image quality assessment

The identification of croplands using satellite imagery from Mapbox and Google is often hindered by low-quality images,
245 including those affected by no-data regions, cloud cover (thick and thin), snow/ice, and land fallow periods. To address these
challenges, an image quality assessment (IQA) method was developed to guide the selection of optimal datasets.

IQA replicates human perception of image quality and serves two key purposes: filtering out poor-quality images to
improve downstream tasks such as alignment, fusion, and recognition, and evaluating the performance of post-processing
algorithms (Zhu et al., 2020). IQA methods are broadly classified into with-reference and no-reference approaches (Gao et
250 al., 2024). In this study, both were utilized: with-reference IQA for classifying easily identifiable features and no-reference
IQA for detecting thin clouds through specialized feature calculations. The IQA outputs were integrated during post-
processing to enhance cropland identification results.

1) Coverage Type Classification: To classify image coverage types, we trained a ResNet-based image classification model
using the samples described in Section 2.2.3. ResNet's deep residual architecture enhances feature extraction by addressing
255 the vanishing gradient problem with skip connections, enabling the network to capture multilevel abstract features. This
capability allows ResNet to excel in identifying complex image attributes such as color, texture, shape, and spatial context,
thereby improving classification accuracy. Fine-tuning the pre-trained ResNet model on the dataset further enhances
generalization and efficiency in coverage type classification.

2) Thin cloud features calculation: Thin clouds pose a unique challenge, as they are difficult to detect through standard
260 image classification methods. To address this, two features were calculated: the Gradient (Eq.1) represents the sharpness of
land cover boundaries, serving as an indicator of image clarity. The HSL (Hue, Saturation, and Lightness) (Eq.2-4) Captures
color and saturation properties, aiding in the detection of subtle cloud contamination. These features collectively contribute
to assessing the impact of thin clouds on image quality, enabling more accurate cropland identification during post-
processing.

265 $Gradient = Sobel(R, G, B)$ (1)

$$Hue = \begin{cases} 0°, if\ max = min \\ 60° \times \frac{G-B}{max+min} + 0°, if\ max = R\ and\ G \geq B \\ 60° \times \frac{G-B}{max+min} + 360°, if\ max = R\ and\ G < B \\ 60° \times \frac{B-R}{max+min} + 120°, if\ max = G \\ 60° \times \frac{G-B}{max+min} + 240°, if\ max = B \end{cases}$$ (2)

$$Saturation = \begin{cases} 0 & if\ max = 0 \\ \frac{max-min}{max} & otherwise \end{cases}$$ (3)

$Lightness = \frac{1}{2}(max + min)$ (4)

Here, the R, G, B represent the Red, Green, and Blue bands of the RGB image. The max and min referred to $max(R, G, B)$
270 and $min(R, G, B)$. The Sobel operator is applied to each of the three bands of the RGB image to calculate the gradient, which
is a measure of the intensity of the edges in the image.

3.2 Cropland extraction: semantic segmentation of active learning

3.2.1 Cropland extraction via active learning strategy

The active learning process for semantic segmentation begins by labeling an initial set of image blocks, which is then used to
275 train the segmentation model. The trained model predicts cropland areas beyond the initial blocks, and these predictions are
reviewed. Blocks with significant commission or omission errors are labeled and added to the sample set, iterating this
process until a stopping criterion is met. The stopping criterion is defined as the absence of substantial artifacts or
underestimation errors.

A significant challenge in the iterative process is evaluating the quality of predictions and determining when to terminate
280 the iteration. During the initial iterations, errors such as overestimation and underestimation are readily apparent through
manual visual inspection. However, as the model's predictions become more refined in later iterations, these errors become
less noticeable, making manual review of the entire area increasingly impractical.

To overcome this challenge, an independent error assessment model is utilized to automatically evaluate the predictions of
the cropland extraction model. By leveraging several image features, the error assessment model is trained to estimate
285 extraction accuracy, enabling it to detect errors without relying solely on manual intervention. This automated evaluation
ensures efficient and precise identification of cropland areas while significantly reducing the need for extensive human
review.

3.2.2 Mask2Former model

Mask2Former is a universal image segmentation framework capable of handling semantic, instance, and panoptic
290 segmentation tasks through a Transformer-based decoder (Cheng et al., 2022). Unlike traditional Transformer decoders, it
introduces a masked attention mechanism that restricts cross-attention operations to the foreground regions defined by
predicted masks. This approach accelerates convergence and improves performance by focusing attention on local areas
rather than the entire image, making it particularly effective for segmenting small objects. The model incorporates multi-
scale, high-resolution feature inputs, allowing it to handle objects and regions of various sizes while iteratively refining mask
295 predictions through multiple layers.

To enhance efficiency, Mask2Former optimizes the Transformer decoder by adjusting the order of self-attention and
cross-attention for more effective feature learning. It also uses learnable query features, which provide more expressive
initial queries without relying on random initialization. Additionally, the model reduces memory consumption by computing
mask loss through random sampling, lowering memory usage by threefold while maintaining segmentation performance.

300 These improvements enable Mask2Former to achieve high segmentation accuracy while significantly increasing training efficiency and reducing computational demands.

3.2.3 Error assessment features

To account for multiple features in error assessment, a data-driven approach was employed to evaluate each block. Region properties (Burt et al., 1981), including Area Ratio, Solidity, Euler Number, and Consistency, were calculated (Maryada et

305 al., 2024; Bhosale et al., 2023). Using systematic grid sampling (Section 2.2.5), manually interpreted samples were used to train an XGBoost model to predict errors across the study area. The defined features are as follows:

1) Area Ratio: The proportion of cropland area within each block, calculated as (Eq.5):

$$Area\ Ratio = \frac{PN_{cf}}{PN_{Block}} \tag{5}$$

where PN_{cf} and PN_{Block} are the pixel number for crop fields and entire region of each block.

310 2) Solidity: Solidity represents the compactness of cropland regions, measured as (Eq.6):

$$Solidity = \frac{Area}{Area(Convex\ Hull)} \tag{6}$$

where $Area$ is the pixel count in the cropland region, and $Area(Convex\ Hull)$ is the pixel count in its convex hull. Higher solidity indicates fewer irregularities in shape.

315 3) Euler number: A topological metric evaluating the structure of cropland extraction results, defined as (Eq.7):

$$Euler\ Number = C - H \tag{7}$$

where C represents the number of connected components (objects), and H is the number of holes within those components. A higher Euler number signifies fewer holes and better segmentation quality.

4) Consistency: Consistency quantifies similarity between binary images and reference data, using the Intersection over

320 Union (IoU) metric (Eq.8):

$$IoU = \frac{TP}{TP+FP+FN} \tag{8}$$

where TP (true positives) and TN (true negatives) are correct predictions, FP (false positives) are incorrect identifications, and FN (false negatives) are missed identifications. IoU was applied to compare results from Mapbox and Google datasets, extracted cropland samples, and the final Croplayer product with other datasets.

325 ## 3.3 Post-processing: cropland results integration

Using imagery from Mapbox and Google, two distinct cropland identification results were generated for each block across the study area. To create a final, higher-quality cropland map, an integration strategy was developed to combine the strengths of both datasets. Each block was assigned one of four integration options: Neither Source, Only Mapbox, Only Google, or Merging both, as detailed in Table 4.

330 **Table 4:** Results refining options and explanations

Option	Major Region	Explanation
Neither source	Western	Applied primarily for noise-prone areas, characterized by small, isolated features with low consistency between Mapbox and Google results (e.g., dunes resembling fallow fields under certain lighting).
Only Mapbox / Only Google	Central and Eastern	Suitable for regions with larger, regular crop fields that are easily identifiable. The source with higher image quality (e.g., seasonal clarity) is preferred, avoiding merged results that may capture fallow orchards.
Merging both	Southern	Applied in areas with smaller crop fields in mountainous or hilly terrain. Complex landscapes with dense vegetation and forested croplands benefit from merging sources, as this provides additional phenological information to aid cropland boundary delineation.

To determine the optimal integration strategy for each block, a data-driven approach was employed. Four groups of features—Geography, Image Quality, Region Properties, and Consistency—were derived and summarized in Table 5. Samples collected through systematic grid sampling (Section 2.2.5) were manually interpreted and used to train an XGBoost model, which predicted the most suitable integration strategy for each block.

335 **Table 5:** Evaluation metrics for each block

Groups	Features	Explanation
Geography	Elevation	Median of SRTM DEM
	Slope	Median slope value
	Ruggedness	Median Terrain Ruggedness Index (TRI)
	Roughness	Median Roughness value
Image quality	Coverage type	Image quality: Planting, Non-Planting, Cloudy, Snow/Ice and Nodata
	Sharpness	Median Sobel gradient value
	HSL	Median Hue, Saturation, Lightness values
Region property of results	Area	Total cropland area within a block
	Area ratio	Ratio of cropland area to block area.
	$Solidity_{mean}$	Mean Solidity of binary cropland images
	$Solidity_{total}$	Total Solidity values across all cropland regions in a block
	EN_{mean}	Mean Euler Number of binary cropland images
	EN_{total}	Total Euler Number across all cropland regions in a block
Consistency	Consistency	IoU between Mapbox and Google cropland extractions

3.4 Cropland results assessment

The validation of Croplayer data involves three objective metrics: 1) Accuracy assessment: The IoUscore is used to evaluate the performance of the cropland extraction model with the test sample set. 2) Area assessment: Validation is conducted at the provincial level, where the extracted cropland area is compared to the statistical cropland area to assess proportionality. 3)

340 Consistency assessment: The IoU score is also employed to compare Croplayer with seven existing cropland datasets, evaluating consistency.

In addition to the quantitative assessments, a visual analysis was conducted to compare Croplayer data with other datasets. This analysis focused on typical details to better understand the reasons for any discrepancies between the datasets.

4 Result

345 ## 4.1 Image quality assessment

The Coverage Types classification model, based on ResNet, achieved a top-1 accuracy of 95.6%. Figure 5 presents the statistical values for the different Coverage types. Overall, the combined proportion of Planting and Non-Planting images is similar for both datasets, accounting for 82.5% of the total. However, there are notable differences between them. For high-quality Planting images, Mapbox has a higher proportion at 50.7%, compared to Google's 41.3%. Conversely, for medium-

350 quality Non-Planting images, Mapbox accounts for 32.2%, while Google has a higher proportion at 40.8%.

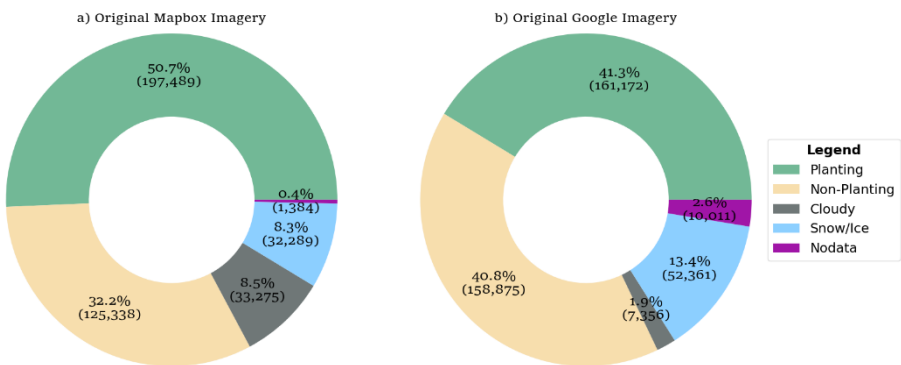



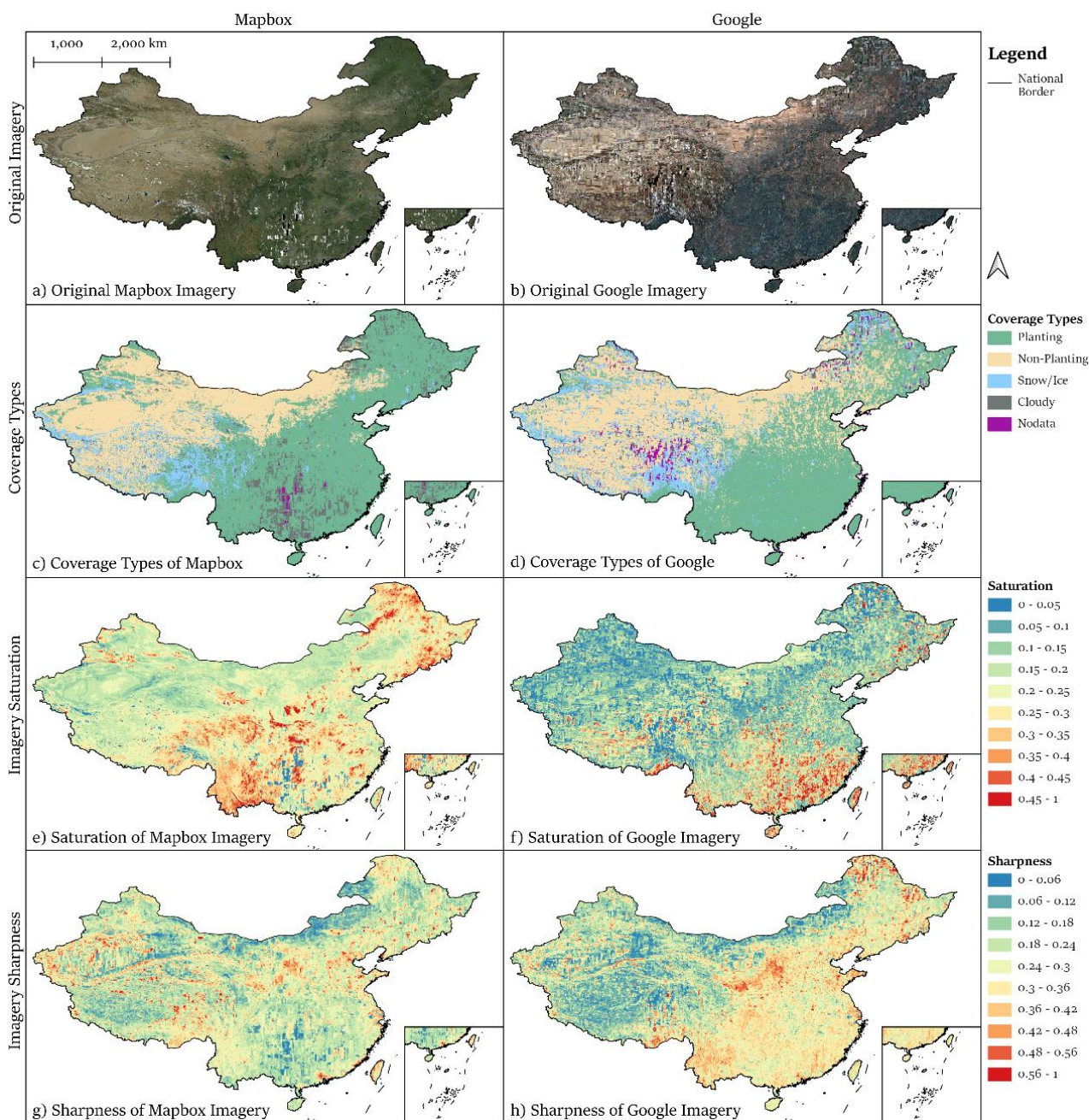
Figure 5: Statistics of Coverage Types for Original Mapbox and Google Imagery.

The combined percentage of the three lower-quality categories (Cloudy, Snow/Ice, and Nodata) is roughly 17.5% for both datasets. Specifically, the Cloudy category makes up 8.5% of Mapbox's data, significantly higher than Google's 1.9%. On

355 the other hand, Google shows a higher proportion of Snow/Ice (13.4%) and Nodata (2.6%) issues compared to Mapbox (8.3% and 0.4%, respectively).

The geographic distribution of image quality, shown in Figure 6, further underscores the differences between the two datasets:

For Mapbox, image quality is generally poorer in southern China, with a higher prevalence of Cloudy and Nodata blocks.
360 In contrast, northern China exhibits better image quality, though Cloudy and Snow/Ice imagery is concentrated in the northeastern region.

For Google, image quality is comparatively better in southern China, with more Nodata and Snow/Ice blocks observed in the northeastern and western regions. Additionally, northern China has a higher number of Non-Planting images.

365   **Figure 6:** Original Mapbox/ Google Imagery and Image Quality Assessment (IQA). a) Original imagery from Mapbox; b) Original imagery from Google; c) Coverage types of Mapbox imagery; d) Coverage types of Google imagery; e) Saturation of Mapbox imagery; f) Saturation of Google imagery; g) Sharpness of Mapbox imagery; h) Sharpness of Google imagery. The HR remote-sensing images in the figure are from © Mapbox 2023 and © Google Maps 2023.

4.2 Sample distribution

370   The active learning sampling process was iterated approximately 50 times. To assess the distribution of samples across provincial units, we calculated the sampling density, defined as the ratio of the number of sampled blocks (both positive and negative) to the total number of blocks within each provincial unit (Figure 7). Among China's 32 provincial units, only six— Beijing, Ningxia, Tianjin, Guangdong, Qinghai, and Shanghai—exceed the national average sampling density of 0.0033. Notably, Shanghai exhibits the highest sampling density at 0.0135, significantly surpassing the next-ranked province,

375   Taiwan, with a sampling density of 0.0025.

Conversely, provinces with the lowest sampling densities include Shaanxi, Shanxi, Hebei, Hainan, Henan, Hubei, Anhui, Zhejiang, and Jiangxi. Except for Hainan, these provinces are concentrated within China's major grain-producing regions: the Huang-Huai-Hai Plain and the middle and lower reaches of the Yangtze River Plain. The underlying reasons for these patterns will be further examined in conjunction with the extraction results in Section 5.2.

380

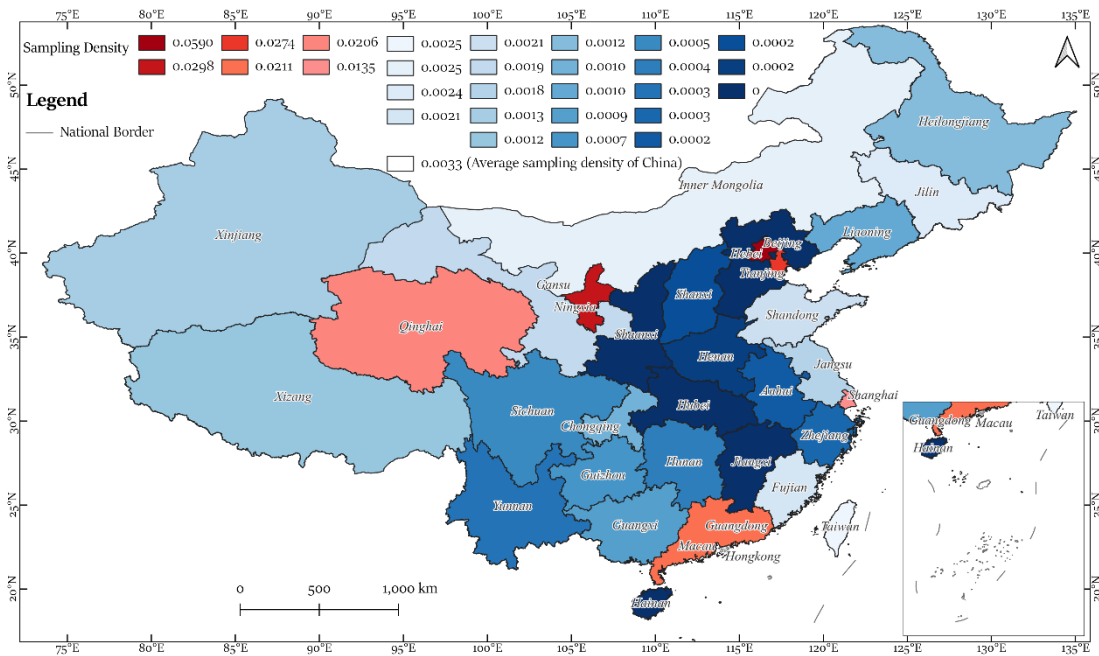



Figure 7: Sampling density for 32 provincial units.

4.3 Cropland extraction accuracy

A validation sample set was utilized to evaluate the performance of four segmentation models: PSPNet, PIDNet, Segformer, and Mask2Former. Among these, Mask2Former achieved the highest IoU score of 88.73% and was consequently selected for cropland extraction in this study.

385

Table 6: Model performance.

Model	Training timing	Test Accuracy in IoU
PSPNet	20h	69.13%
PIDNet	14h13m	82.37%
Segformer	5h41m	85.10%
Mask2Former	11h56m	88.73%

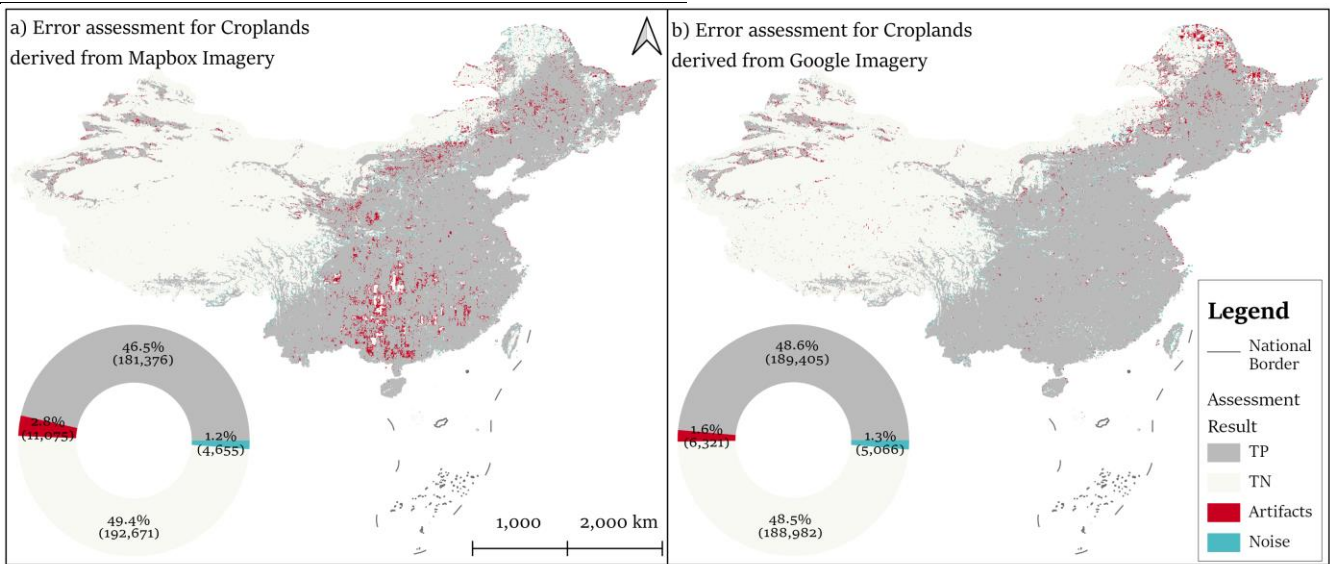



Figure 8: Distribution and statistics of error assessment for extracted croplands from a) Mapbox and b) © Google imagery.

The error assessment, conducted using the XGBoost model, achieved an overall accuracy of 94.3%. The combined TP and TN predictions reached accuracies of 95.9% for Mapbox and 97.1% for Google, demonstrating the strong performance of both sources in cropland identification (Figure 8). However, Mapbox imagery exhibited higher occurrences of Noise and Artifacts compared to Google imagery.

390

Spatially, Artifacts are more closely linked to low-quality imagery, as identified by the IQA analysis (Figure 6), rather than to topographical features such as Slope (Figure 1). In Mapbox imagery, Artifacts are primarily concentrated in southern China, particularly in mountainous and hilly regions frequently affected by Cloudy conditions (Figure 6). In contrast, Artifacts in Google imagery are more prevalent in northeastern China, largely due to Snow/Ice coverage, though they are predominantly observed in expansive plains.

395

4.4 Cropland results integration

The results integration was achieved through a combination of model predictions and manual modifications. The final results
400   integration are illustrated in Figure 9. The four integration strategies—Neither Source, Only Mapbox, Only Google, and
Merging Both—account for 46.7%, 15.7%, 19.7%, and 18.0% of the total blocks, respectively. Geographically, the
distribution of these options aligns closely with the Slope characteristics of the regions. The integration characteristics can be
broadly summarized as follows:

     1) Single Source Regions: These areas include major grain-producing regions such as the Northeast Plain, central and
405   eastern China, the North China Plain, and key agricultural zones in Xinjiang, the Hexi Corridor of Gansu, and Yunnan
Province. In these regions, one imagery source was typically preferred due to notably poor data quality from the alternative
source.

     2) Merged Source Regions: These regions primarily encompass southern China, mountainous and hilly areas of the Loess
Plateau, eastern Mongolia, terraced landscapes, and river valleys in Xizang.

410   3) Excluded Regions: Areas where neither imagery source was utilized are largely located in western China and Inner
Mongolia.

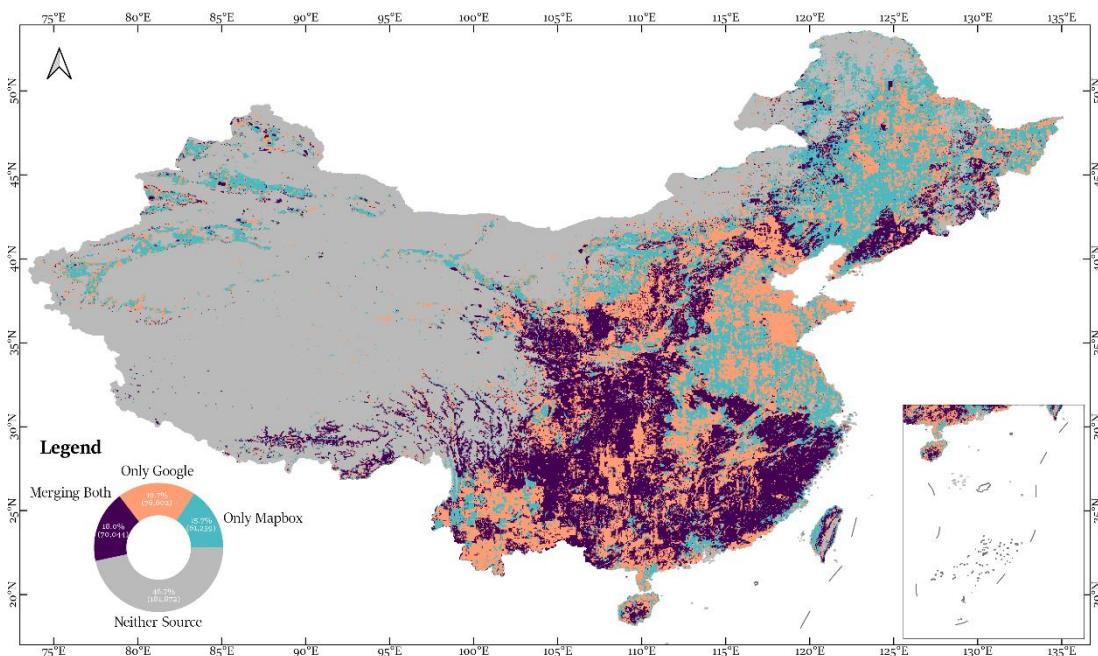



Figure 9: Cropland results integration

     Figure 10 presents the statistical cropland area values across the four scenarios. The x-axis represents the four scenarios
415   (Only Mapbox, Only Google, Complete Integration, and Selected Integration), while the y-axis lists 32 provincial units,
ordered roughly by latitude. These are divided into Northern provinces (Qinghai and above) and southern provinces (Xizang
and below).

In general, cropland estimates in Northern provinces are close to or slightly exceed statistical areas, while southern provinces exhibit widespread underestimation. The reasons behind these deviations will be discussed in detail in Section 5.2.

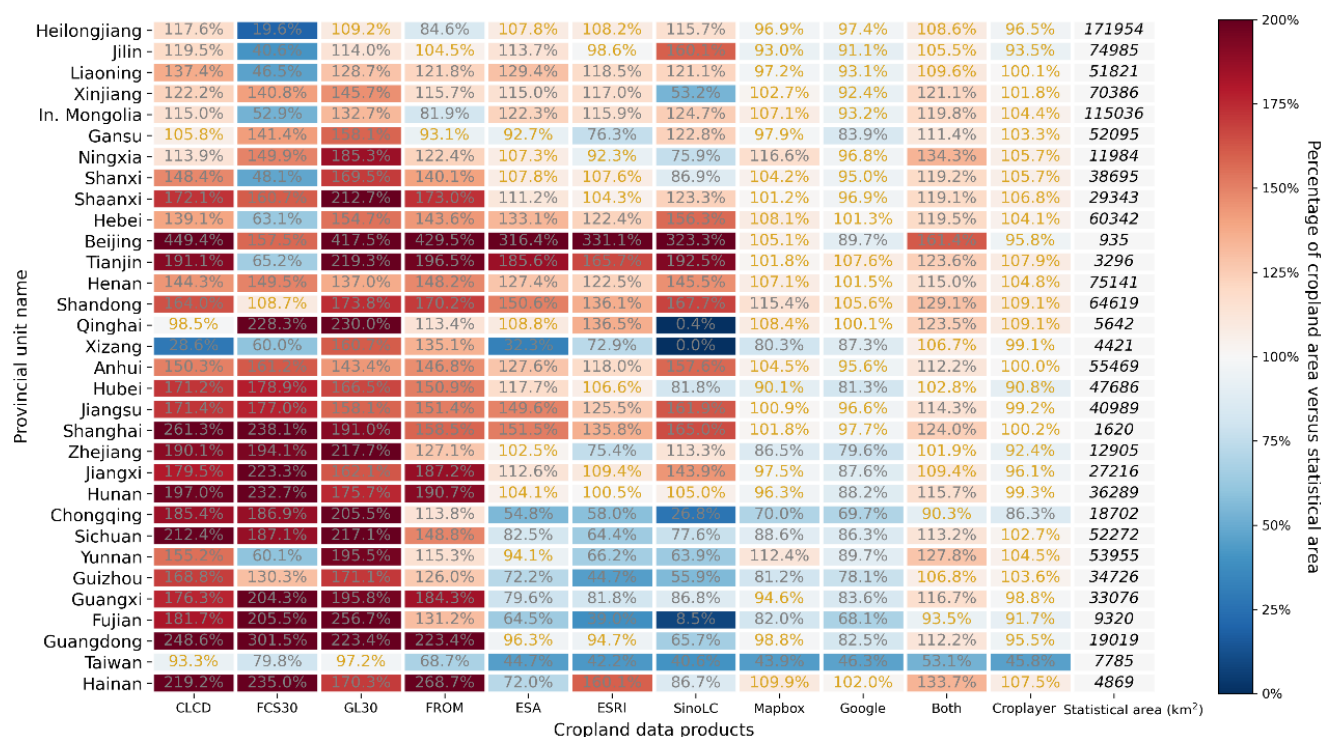



Figure 10: Percentage of cropland area versus statistical area (PCS) of 7 existing data, cropland extraction results derived by four scenarios (Only Mapbox, Only Google, Complete Integration, and Selected Integration) with statistical area. In. Mongolia represents Inner Mongolia.

4.5 Comparison with seven existing cropland data

We evaluated seven existing cropland datasets—CLCD, ESA, ESRI, FCS30, FROM, GL30, and SinoLC—alongside Croplayer (Figure 11 a-f, h, and i) using statistical area data. Additionally, we assessed the consistency of these datasets with Croplayer.

(1) Percentage of cropland area versus statistical area (PCS):

The cropland area for each dataset was calculated across 32 provincial administrative units and compared with statistical data. Differences in PCS within ±10% of the statistical area (90%–110%) are highlighted in gold in Figure 10.

Croplayer exhibited the highest alignment with statistical areas, with 30 provincial units showing PCS differences within ±10% (Figure 10). By contrast, ESA and ESRI met this standard for 9 units, CLCD and FROM for 2 units each, and FCS30, GL30, and SinoLC for only 1 unit each (Figure 11).

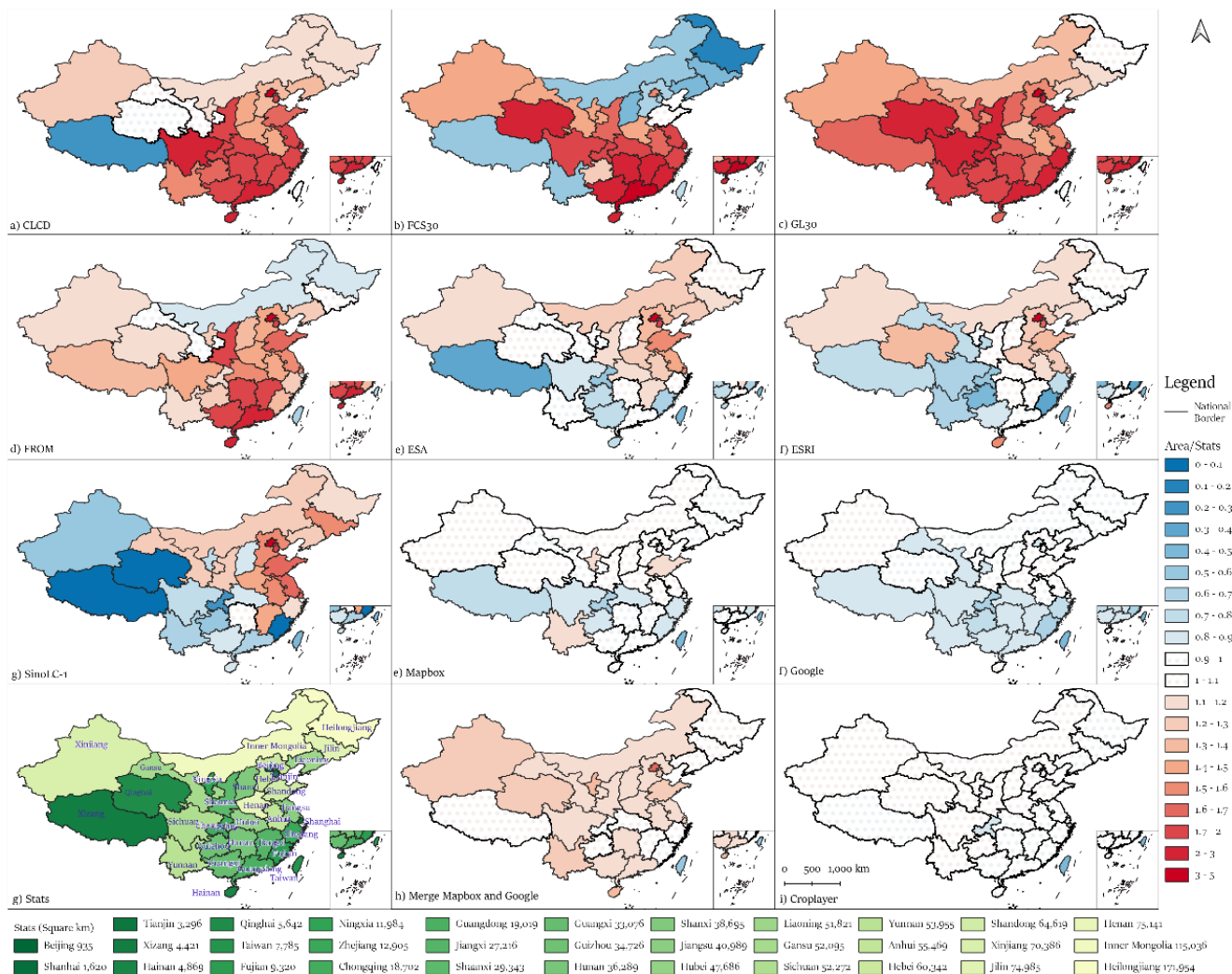



Figure 11. PCS distribution of 7 existing data, cropland extraction results derived by four scenarios (Only Mapbox, Only Google, Complete Integration, and Selected Integration) and statistical area.

(2) Consistency of 7 existing cropland data with Croplayer

Focusing on consistency between Croplayer and the seven existing datasets, two key observations emerge (Figure 12):

High-Consistency Datasets: A comparison of provincial units revealed that among the seven datasets, only ESA, ESRI, FROM, and CLCD achieved the highest provincial consistency with Croplayer. Among them, ESA had the most instances, aligning with Croplayer in 26 out of 32 provincial units. ESRI, CLCD, and FROM achieved the highest consistency in 4, 1, and 1 provincial units, respectively.

Regional Trends: High consistency between Croplayer and other datasets was predominantly observed in regions characterized by extensive plains, particularly in northern China and the southern plain provinces of Anhui, Hubei, and Jiangsu.

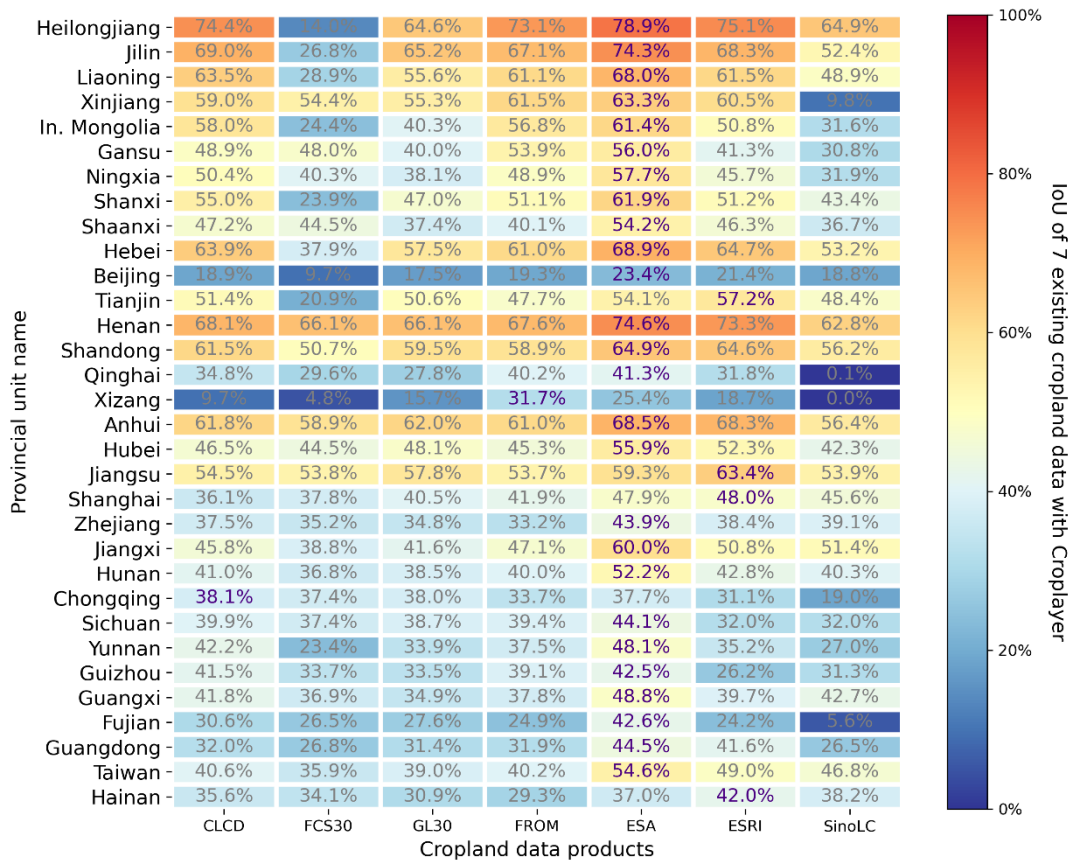



Figure 12. IoU of 7 existing cropland data with Croplayer. In. Mongolia represents Inner Mongolia.

(3) Visual Comparison

We selected ten representative sites for visual comparison across eight datasets, including Croplayer and seven publicly
450   available datasets (Figure 13). The analysis highlights the unique challenges in each region and underscores Croplayer's
superior performance in accurately identifying croplands while addressing the limitations of existing datasets:

a) Qiqihar, Heilongjiang: For large black soil croplands, most datasets successfully identified the main cropland areas,
except for FCS30, which exhibited significant omissions.

b) Changji, Xinjiang: In arid regions with extensive pivot irrigation croplands, most datasets performed well, although
455   GL30 lacked detail, and SinoLC showed overall omissions.

c) Pinggu, Beijing: In orchard-dominated areas, FCS30 slightly overestimated croplands. The other datasets consistently
misclassified orchards as croplands, leading to a 3–4 times overestimation of Beijing's cropland area.

d) Zhongwei, Ningxia: For regions dominated by sand-control grass grids, ESRI showed minor overestimations, while
other datasets contained substantial errors.

460

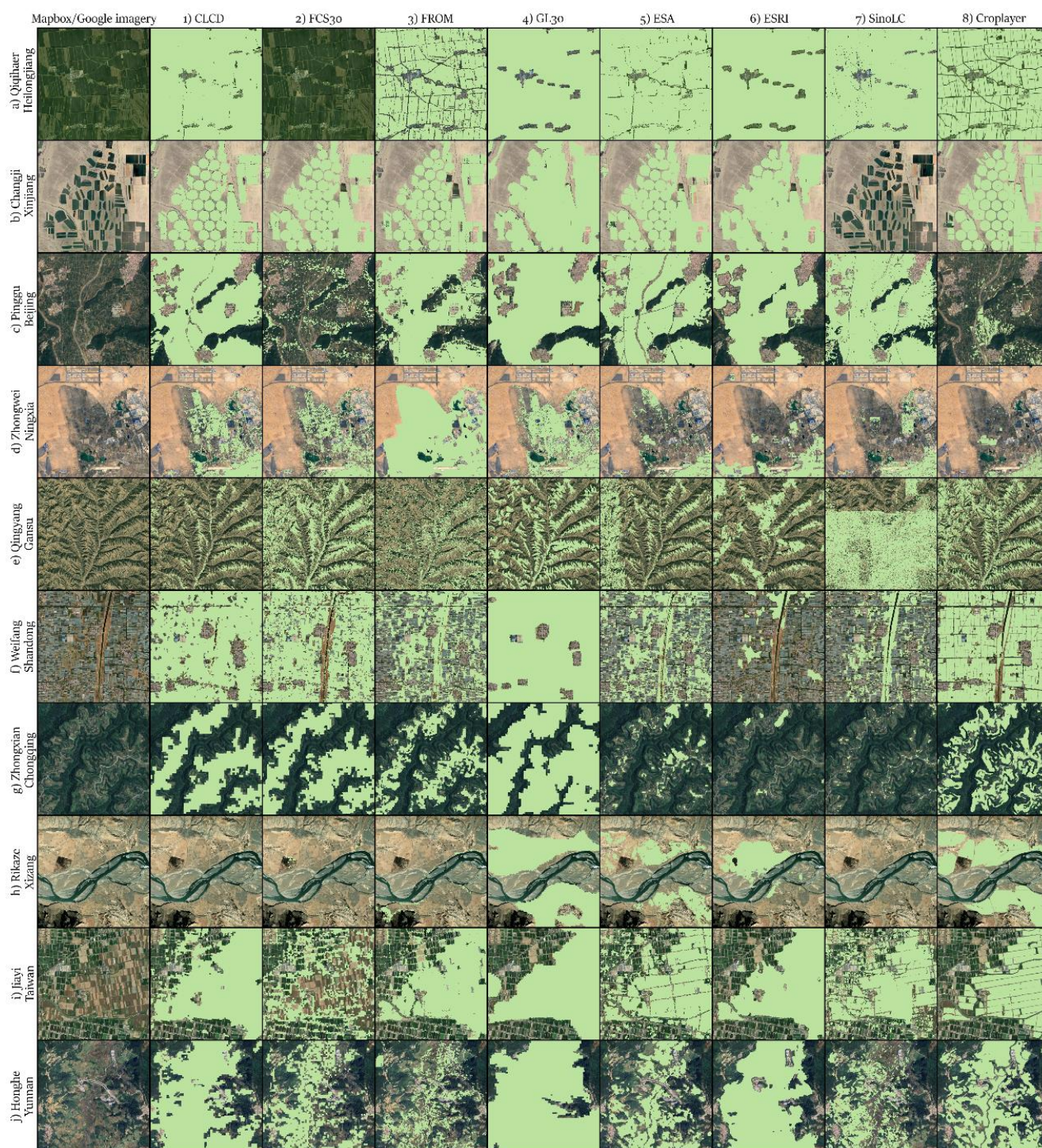



Figure 13. Example site of 7 existing cropland data and Croplayer. The HR remote-sensing images in the figure are from ©
Mapbox 2023 and © Google Maps 2023.

e) Qingyang, Gansu: On the Loess Plateau terraces, CLCD and GL30 performed best, while FROM and SinoLC produced
465 fragmented results. ESA and SinoLC exhibited artifacts, and ESRI underestimated cropland significantly.

f) Weifang, Shandong: In this vegetable-growing region with numerous greenhouses, only CLCD, FCS30, and GL30
accurately identified croplands. Other datasets, particularly ESA, significantly underestimated cropland coverage.

g) Zhongxian, Chongqing: In the complex southern folded hilly croplands, CLCD, FCS30, and FROM performed well.
GL30 overestimated extensively, while ESRI, ESA, and SinoLC largely missed croplands.

470 h) Rikaze, Xizang (Tibet): For high-altitude croplands, ESA performed best. GL30 and ESRI exhibited both
overestimations and underestimations, while other datasets largely omitted cropland areas.

i) Chiayi, Taiwan: In coastal areas with a mix of paddy fields and fish ponds, CLCD, FROM, GL30, and ESRI correctly
excluded fish ponds, while other datasets erroneously included some.

j) Honghe, Yunnan: For southern terrace areas, FCS30 and ESA accurately identified croplands. FROM and SinoLC
475 showed some omissions, while other datasets tended to overestimate cropland coverage.

These comparisons demonstrate Croplayer's robust performance in addressing the limitations of existing datasets across
diverse and complex agricultural landscapes.

5 Discussion

5.1 Impact of the image quality

480 In cropland remote sensing mapping, the quality of the data source sets the upper limit of mapping accuracy. Mapping
methodologies—including sample selection, extraction, and result integration—determine how closely this limit can be
approached. In remote sensing imagery, cropland parcels exhibit distinct features, such as shape and texture, which tend to
be more stable than spectral or temporal variations and less affected by weather conditions or sensor performance.
Consequently, cropland extraction that leverages the detailed information provided by HR data offers a more effective
485 approach than methods relying on temporal variation information derived from traditional medium-resolution data (e.g.,
Sentinel, Landsat).

However, obtaining large-scale HR imagery remains a significant challenge. Publicly available 2 m resolution data, such
as GF-1, GF-2, GF-6, and ZY-3, typically require extensive pre-processing, including registration, calibration, and color
rendering, all of which increase processing costs. By contrast, the Mapbox and Google datasets are pre-processed into RGB
490 images, making them more readily compatible with deep learning-based semantic segmentation algorithms. Nonetheless,
these datasets suffer from inconsistent quality and a lack of acquisition metadata (e.g., sensor type, acquisition time, cloud
cover), complicating the identification of low-quality imagery.

To address these limitations, we adopted a simplified image classification approach that relies solely on the imagery itself
to evaluate image quality and coverage types, despite the absence of acquisition metadata. As demonstrated in Section 4.1,
495 both the Mapbox and Google datasets exhibit considerable quality issues. The imagery was classified into five Coverage

types: Planting, Non-Planting, Cloudy, Snow/Ice, and Nodata, with the last three types significantly reducing extraction accuracy. Importantly, the distribution of low-quality data varies between the two datasets. Specifically, thick clouds and missing data in Mapbox imagery are concentrated in southern China and parts of northeastern China, whereas snow/ice and cloudy conditions are prominent in Google imagery, particularly over the Qinghai-Xizang Plateau and northeastern China.

500 This complementary distribution of low-quality data between the two sources enhances the potential accuracy of integrated results.

Additionally, sharpness and HSL features were employed to detect subtle cloud coverage that might otherwise be overlooked. Artifact errors were identified through error assessment and addressed by replacing or merging erroneous segments. Despite these measures, certain blocks contained low-quality imagery from both Mapbox and Google, which was

505 a major source of over- or under-segmentation errors in the Croplayer dataset.

This approach underscores the importance of integrating multiple data sources to achieve comprehensive and effective coverage, as relying on a single HR imagery source inherently limits spatial and data quality coverage. By leveraging complementary datasets, we can enhance spatial coverage while maintaining the necessary data quality for accurate cropland mapping.

510 **5.2 Sample Representativeness**

We evaluated the advantages of Croplayer using two objective metrics—Provincial Consistency Score (PCS) and Consistency—as well as subjective visual assessments (Figures 9 and 13). Using statistical cropland area data as a reference, 30 out of 32 provincial units exhibit PCS differences of less than ±10% with Croplayer. In contrast, the most accurate datasets among the other seven (ESA and ESRI) meet this standard in only eight provincial units. Furthermore, as

515 demonstrated in Figure 13, Croplayer provides finer and more stable details across 10 different cropland types compared to other datasets.

The superior performance of Croplayer is attributed to the rich detail provided by HR imagery and the active learning strategy employed. Statistical area data were integrated into the mapping process and used as termination criteria during sample collection. Artifacts were automatically identified through an error assessment process, enabling improvement in

520 poorly performing regions by supplementing new samples, which resulted in enhanced model predictions. However, the iterative process, including sample creation, model training, and result validation, remains time-consuming and labor-intensive.

Nevertheless, we contend that the representativeness of the samples significantly impacts cropland identification outcomes. By analyzing the 10 examples in Figure 12, we explain the substantial differences in area estimates observed among other

525 datasets (Figure 10):

Firstly, resolution is important but not the sole factor: ESA and ESRI datasets achieve the most accurate area estimates, with ESA exhibiting the highest consistency. Visually, ESA and FROM provide finer spatial details, often delineating field boundaries. Consequently, datasets with 10 m resolution (ESA, ESRI, and FROM) outperform those with 30 m resolution

530   CLCD performs comparably to ESA, ESRI, and FROM. At last, despite using 1 m resolution imagery, SinoLC does not achieve higher area accuracy than the other six datasets.

Secondly, sample representativeness is the primary factor for cropland area estimation: Existing systematic and rule-based sampling approaches have limitations. Based on our findings, we summarize three key insights:

1) Sampling quantity should reflect geospatial heterogeneity rather than area size:

535   Traditional stratified sampling allocates samples in proportion to area size, but this is less effective for semantic segmentation. Model generalization requires diverse samples capturing cropland heterogeneity and surrounding contexts rather than large quantities of similar samples. For example, the Huang-Huai-Hai Plain and the middle and lower reaches of the Yangtze River Plain, representing China's largest cropland areas, have the fewest samples. Provinces in these regions, including Hebei, Hainan, Henan, Shaanxi, Shanxi, Hubei, Anhui, Zhejiang, and Jiangxi, had almost no samples, while

540   Jiangsu and Shandong also had few.

2) Numerous special cases require dedicated sample creation

HR remote sensing segmentation encounters challenges in complex scenarios influenced by data quality, algorithm performance, and sampling expertise. Many errors only become apparent during the mapping process, necessitating the creation of additional samples to address specific issues. For example, tailored sampling was required for orchards in Pinggu

545   (c), greenhouses in Weifang (f), sand control grids in Zhongwei (d), and high-altitude croplands in Xizang (h).

Regions with the highest sample density (Figure 7), such as Beijing, Tianjin, and Shanghai, feature complex land cover types, including skyscrapers and their shadows, urban open spaces, orchards, and athletic fields. Similarly, provinces like Guangdong, Ningxia, and Qinghai, characterized by diverse landforms—encompassing orchards, fish ponds, sand control grids, salt pans, sand dunes, and photovoltaic panels—also demanded specialized sampling efforts.

550   3) Insufficient utilization of prior knowledge

Statistical data can guide sampling in two ways: Firstly, assisting in initial sample planning by aiding proportional allocation based on area size. Secondly, defining termination conditions for sampling, as seen in this study, where statistical data helped evaluate extraction performance.Using statistical constraints and iterative active learning in the extraction process effectively prevents extreme errors. While the seven other datasets showed errors where at least two provincial units

555   exhibited underestimations of 50% or overestimations exceeding 200%, our approach minimized such inaccuracies.

5.3 Potential use of the data

The Croplayer dataset provides a robust foundation for tasks such as identifying crop distribution, assessing growth, estimating yields, and evaluating crop damage. Its precise delineation of cropland boundaries ensures high consistency with provincial statistical values and enables its integration with broader research domains. The advantages of Croplayer can be

560   summarized as follows:

1) Enhanced data comparability: Croplayer closely aligns with provincial statistical values while utilizing only publicly accessible data sources, ensuring transparency, reproducibility, and facilitating data-sharing in agricultural and environmental research. Furthermore, advanced outputs like crop yield estimates derived from Croplayer remain consistent with other statistical datasets, enabling seamless integration across diverse analytical frameworks.

565 2) Advancing agricultural models: Accurate cropland delineation supports precise classification of rain-fed, irrigated, and paddy croplands, as well as more detailed subcategories (Salmon et al., 2015), while also advancing crop classification research (Dong et al., 2020). Croplayer's extensive coverage and high-resolution precision provide essential inputs for building agricultural foundation models (Li et al., 2024) and large-scale knowledge systems (Yang et al., 2024), driving advancements in predictive modeling for crop growth, yield estimation, and environmental impact assessment, thereby
570 fostering innovation in agricultural AI research.

3) Impacts on land cover studies: By improving cropland identification, Croplayer reduces classification errors in other land cover categories, such as forests, water bodies, and built-up areas. This enhancement indirectly increases the reliability of land cover datasets, supporting applications in ecological research, environmental management, and urban planning (Chen et al., 2022).

575 **5.4 Limitations and future directions**

This study primarily employs semantic segmentation for cropland identification, which, while effective for delineating cropland regions, falls short of providing instance segmentation to distinguish individual fields. This limitation constrains more granular analyses, such as precise field counts and area measurements, which are critical for agricultural statistics and field-level management. Incorporating instance segmentation in future iterations of the Croplayer dataset would significantly
580 enhance its utility for detailed agricultural applications.

Although the Croplayer dataset aligns closely with provincial statistical areas, the lack of comprehensive municipal- and county-level statistics poses challenges for fine-scale validation. While provincial-level estimates appear accurate, discrepancies at finer administrative scales, such as overestimations in urban-adjacent regions, highlight the need for more localized statistical data. Additionally, the rapid pace of urbanization further exacerbates these inaccuracies when imagery
585 updates lag behind land use changes, emphasizing the importance of more frequent and timely imagery acquisitions (Xian et al. 2009).

Finally, the study's reliance on 2020 data limits its capacity to capture temporal dynamics in cropland distribution. Multi-temporal datasets are crucial for understanding the impacts of agricultural policies, environmental changes, and land use transitions. As public imagery resources with higher spatial and temporal resolution continue to expand, future versions of
590 Croplayer can address these limitations. By integrating instance segmentation, multi-temporal analyses, and finer-scale validation, the dataset holds the potential to support more comprehensive agricultural and land use research.

6 Data availability and user guidelines

The Croplayer China 2020 dataset generated in this study is available on Zenodo: https://doi.org/10.5281/zenodo.14726428 (Jiang et al., 2025). The dataset is provided in Shapefile (.shp) format, with each file representing cropland within a county-
595 level administrative region in China. For clarity, file names follow the format cf_城市名_县名 (cf_city_county) and are organized by province, with each province packaged into a compressed file (.tar.gz). Additional datasets used in this study are introduced in Section 2, with download links provided.

7 Conclusion

This study addresses the discrepancies in existing public cropland datasets across China by proposing a comprehensive, data-
600 driven framework for high-resolution cropland mapping. Utilizing publicly available 2 m High Resolution (HR) imagery from Mapbox and Google, the framework integrates advanced deep learning and machine learning techniques in a three-stage process: (1) national imagery is partitioned into $0.05\degree \times 0.05\degree$ blocks for efficient parallel processing, accompanied by an Image Quality Assessment (IQA) to ensure data integrity despite the absence of metadata; (2) an active learning-based model is employed for cropland identification, incorporating segmentation and error assessment to refine the accuracy of the
605 predictions; and (3) an integration strategy merges four key feature groups—Geography, IQA, Region Property, and Consistency—resulting in a final cropland map referred to as Croplayer.

The proposed methodology achieves an overall mapping accuracy of 88.73%, with 30 out of 32 provincial units reporting area estimates within ±10% of official statistics. In contrast, only 1 to 9 provinces from seven other existing datasets meet the same accuracy standard. The Croplayer dataset demonstrates considerable potential for improving crop yield estimation,
610 facilitating agricultural structure analysis, and enhancing land-use research across China. Furthermore, the integration of open-access data, coupled with the high spatial resolution, enables future studies to refine agricultural monitoring models, making Croplayer a valuable tool for policymakers and researchers aiming to assess and manage China's agricultural landscape.

Author contributions

615 HJ designed the research, implement and wrote the paper. MK aggregated data. XZ performed analysis and provided financial support. QZ and YL provided technical support. JX, DL, and CW performed the analysis. JW revised the paper. JZ revised the graph. SC and JH give advices.

Competing interests

The authors declare that they have no conflict of interest

620 **Acknowledgements**

We would like to thank students who participated in the sample interpretation. During the preparation of this work the authors used ChatGPT in order to improve the language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Financial support

625 This research was funded by National Natural Science Foundation of China (No. 42071417), GDAS' Project of Science and Technology Development (2022GDASZH-2022010102), Guangdong Province Agricultural Science and Technology Innovation and Promotion Project (No. 2023KJ102), and Hunan Provincial Natural Science Foundation of China (2023JJ40025)

630

References

Bhosale, Y. H., Zanwar, S. R., Ali, S. S., Vaidya, N. S., Auti, R. A., and Patil, D. H.: Multi-plant and multi-crop leaf disease detection and classification using deep neural networks, machine learning, image processing with precision agriculture-A review, 2023 International Conference on Computer Communication and Informatics (ICCCI), 1-7,

635 Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Morin, D., Rabaute, T., and Savinaud, M.: "Sentinel-2 for agriculture": Supporting global agriculture monitoring, 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 4185-4188,

Burt, P. J., Hong, T.-H., and Rosenfeld, A.: Segmentation and estimation of image region properties through cooperative hierarchial computation, IEEE Transactions on Systems, Man, and Cybernetics, 11, 802-809, 1981.

640 Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794,

Chen, X., Yu, L., Du, Z., Liu, Z., Qi, Y., Liu, T., and Gong, P.: Toward sustainable land use in China: A perspective on China's national land surveys, Land Use Policy, 123, 106428, 2022.

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R.: Masked-attention mask transformer for universal image 645 segmentation, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1290-1299,

Cui, Y., Liu, R., Li, Z., Zhang, C., Song, X.-P., Yang, J., Yu, L., Chen, M., and Dong, J.: Decoding the inconsistency of six cropland maps in China, The Crop Journal, 12, 281-294, 2024a.

Cui, Y., Dong, J., Zhang, C., Yang, J., Chen, N., Guo, P., Di, Y., Chen, M., Li, A., and Liu, R.: Validation and refinement of cropland map in southwestern China by harnessing ten contemporary datasets, Scientific Data, 11, 671, 10.1038/s41597- 650 024-03508-5, 2024b.

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 System Science Data, 12, 3081-3095, 2020.

Gao, G., Li, L., Chen, H., Jiang, N., Li, S., Bian, Q., Bao, H., and Rao, C.: No-Reference Quality Assessment of Extended Target Adaptive Optics Images Using Deep Neural Network, Sensors, 24, 1, 2024.

655   Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., and Bai, Y.: Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Science Bulletin, 64, 370-373, 2019.

Hansen, M. C. and Loveland, T. R.: A review of large area monitoring of land cover change using Landsat data, Remote sensing of Environment, 122, 66-74, 2012.

660   He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778,

Jiang, H., Zhou, X., and Ku, M.: CropLayer: 2-meter resolution cropland mapping dataset for China in 2020 (v1.0) [dataset], https://doi.org/10.5281/zenodo.14726428, 2025.

Jiang, H., Xu, J., Zhang, X., Zhou, X., Liu, Y., Ku, M., Jia, K., Dai, X., Sun, Y., and Chen, S.: Cropland inundation mapping
665   in a mountain dominated region based on multi-resolution remotely sensed imagery and active learning for semantic segmentation, Journal of Hydrology, 639, 131609, https://doi.org/10.1016/j.jhydrol.2024.131609, 2024.

Jun, C., Ban, Y., and Li, S.: Open access to Earth land-cover map, Nature, 514, 434-434, 10.1038/514434c, 2014.

Kaijage, B., Belgiu, M., and Bijker, W.: Spatially Explicit Active Learning for Crop-Type Mapping from Satellite Image Time Series, Sensors, 24, 2108, 2024.

670   Kang, S. and Eltahir, E. A. B.: North China Plain threatened by deadly heatwaves due to climate change and irrigation, Nature Communications, 9, 2894, 10.1038/s41467-018-05252-y, 2018.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P.: Global land use/land cover with Sentinel 2 and deep learning, 2021 IEEE international geoscience and remote sensing symposium IGARSS, 4704-4707,

Kuang, W., Liu, J., Tian, H., Shi, H., Dong, J., Song, C., Li, X., Du, G., Hou, Y., and Lu, D.: Cropland redistribution to
675   marginal lands undermines environmental sustainability, National science review, 9, nwab091, 2022.

Li, J., Xu, M., Xiang, L., Chen, D., Zhuang, W., Yin, X., and Li, Z.: Foundation models in smart agriculture: Basics, opportunities, and challenges, Computers and Electronics in Agriculture, 222, 109032, 2024.

Li, Z., He, W., Cheng, M., Hu, J., Yang, G., and Zhang, H.: SinoLC-1: the first 1 m resolution national-scale land-cover map of China created with a deep learning framework and open-access data, Earth Syst. Sci. Data, 15, 4749-4780,
680   10.5194/essd-15-4749-2023, 2023.

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 Sensing of Environment, 239, 111624, 2020.

Liu, S., Wang, L., and Zhang, J.: The dataset of main grain land changes in China over 1985–2020, Scientific Data, 11, 1430, 2024.

685   Maryada, S. K., Devegowda, D., Curtis, M., and Rai, C.: Improved Data-Driven Method for the Prediction of Elastic Properties in Unconventional Shales from SEM Images, SPWLA Formation Evaluation Symposium of Japan, SPWLA-JFES-2024-N,

Mittal, S., Niemeijer, J., Schafer, J. P., and Brox, T.: Best Practices in Active Learning for Semantic Segmentation,

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., Friedlingstein, P., Liu, C., Tan,
690   K., Yu, Y., Zhang, T., and Fang, J.: The impacts of climate change on water resources and agriculture in China, Nature, 467, 43-51, 10.1038/nature09364, 2010.

Qiu, B., Lin, D., Chen, C., Yang, P., Tang, Z., Jin, Z., Ye, Z., Zhu, X., Duan, M., and Huang, H.: From cropland to cropped field: A robust algorithm for national-scale mapping by fusing time series of Sentinel-1 and Sentinel-2, International Journal of Applied Earth Observation and Geoinformation, 113, 103006, 2022.

695   Qiu, B., Liu, B., Tang, Z., Dong, J., Xu, W., Liang, J., Chen, N., Chen, J., Wang, L., Zhang, C., Li, Z., and Wu, F.: National-scale 10-m maps of cropland use intensity in China during 2018–2023, Scientific Data, 11, 691, 10.1038/s41597-024-03456-0, 2024.

Salmon, J. M., Friedl, M. A., Frolking, S., Wisser, D., and Douglas, E. M.: Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data, International Journal of Applied
700   Earth Observation and Geoinformation, 38, 321-334, https://doi.org/10.1016/j.jag.2015.01.014, 2015.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., Yan, N., Chang, S., Zhao, Y., and Dong, Q.: Challenges and opportunities in remote sensing-based crop monitoring: A review, National Science Review, 10, nwac290, 2023.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P.: SegFormer: Simple and efficient design for semantic segmentation with transformers, Advances in Neural Information Processing Systems, 34, 12077-12090, 2021.

705   Xinyu, L., Bin, F., Yi, L., Dongjie, L., and Shasha, H.: Spatio-temporal coupling evolution and zoning regulation of grain-to-arable value ratio and cropping structures in China, Acta Geographia Sinica, 77, 2721-2737, 2022.

Xu, J., Xiong, Z., and Bhattacharyya, S. P.: PIDNet: A real-time semantic segmentation network inspired by PID controllers, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 19529-19539,

Yang, J. and Huang, X.: The 30m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci.
710   Data, 13, 3907-3925, 10.5194/essd-13-3907-2021, 2021.

Yang, N., Liu, D., Feng, Q., Xiong, Q., Zhang, L., Ren, T., Zhao, Y., Zhu, D., and Huang, J.: Large-Scale Crop Mapping Based on Machine Learning and Parallel Computation with Grids, Remote Sensing, 11, 1500, 2019.

Yang, X., Gao, J., Xue, W., and Alexandersson, E.: Pllama: An open-source large language model for plant science, arXiv preprint arXiv:2401.01600, 2024.

715   Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., and Vergnaud, S.: ESA WorldCover 10 m 2020 v100 (Version v100)[Data set]. Zenodo, 2021.

Zhang, C., Dong, J., and Ge, Q.: Quantifying the accuracies of six 30-m cropland datasets over China: A comparison and evaluation analysis, Computers and Electronics in Agriculture, 197, 106946, 2022.

Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., and Chen, Q.: A generalized approach based on convolutional neural
720   networks for large area cropland mapping at very high resolution, Remote Sensing of Environment, 247, 111912, 2020.

Zhang, L., Xie, Y., Zhu, X., Ma, Q., and Brocca, L.: CIrrMap250: Annual maps of China's irrigated cropland from 2000 to 2020 developed through multisource data integration, Earth System Science Data Discussions, 2024, 1-31, 2024.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery, Earth System Science Data, 13, 2753-2776, 2021.

725   Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J.: Pyramid scene parsing network, Proceedings of the IEEE conference on computer vision and pattern recognition, 2881-2890,

Zhu, H., Li, L., Wu, J., Dong, W., and Shi, G.: MetaIQA: Deep meta-learning for no-reference image quality assessment, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 14143-14152,