remote sensing

MDPI

Article

Task-Oriented Evaluation of Eight Cropland Maps for Agricultural Monitoring in China

Yuan Xue [1,2], Guojin He [1,2,3,*], Ranyu Yin [1,2], Guizhou Wang [1,2,3], Zhaoming Zhang [1,2], Yan Peng [1,3], Tengfei Long [1,2,3] and Huichan Liu [1,2,3]

1 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 Lab of Big Earth Data and Sustainable Development Goal, Kashi Aerospace Information Research Institute, Kashi 844000, China
* Correspondence: hegj@aircas.ac.cn; Tel.: +86-136-1103-4123

Highlights

What are the main findings?

- A new task-oriented framework revealed different rankings of eight cropland maps for classification accuracy and area bias.
- Reference-label changes altered accuracy and area bias while map predictions remained fixed.

What are the implications of the main findings?

- Product selection should match the intended application, as no product performed best across all evaluated tasks.
- Reference-label definitions should be explicitly considered when comparing product accuracy and area bias.

Abstract

Differences in cropland definitions, spatial resolution, and mapping strategies complicate product selection. We harmonized and evaluated eight 2020 cropland and land-cover products across mainland China using hierarchically labeled probability samples. Fixed map predictions enabled comparisons across common task-specific and product-definition-matched reference labels. Evaluation covered spatial agreement, design-weighted classification accuracy, design-based area bias, cropland-type responses, and cross-scale performance. Only 26.8% of pixels in the cropland union were classified as cropland by all products. Within the consensus-cropland evaluation domain, CropLayer achieved the highest overall accuracy (0.961) and F1 score (0.844). CCropLand30 led 30 m products in F1, whereas the Esri land-cover product (ESRI-LC) had national area bias closest to zero (−2.26%). Most 30 m products showed commission-dominated errors and positive area bias. Product responses varied across cropland types, and reference labels altered performance estimates. Switching from unified-agriculture to product-definition-matched labels reduced ESRI-LC's F1 from 0.765 to 0.682 and shifted area bias from −8.08% to +14.88%. Spatial aggregation reduced F1 for some fine-resolution products without consistently changing local cropland-proportion error. These results show that classification accuracy and area bias yield different rankings, supporting product selection according to application, cropland definition, landscape context, evaluation metric, and spatial support.

Check for updates

Academic Editor: Guido D'Urso

Received: 27 September 2026
Revised: 29 October 2026
Accepted: 2 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: cropland mapping; land-cover products; accuracy assessment; area estimation; cropland definitions; China

1. Introduction

Accurate spatial information on cropland underpins food-security assessment, cropland protection, and agricultural-resource management [1–4]. Evaluating cropland products helps identify suitable data for agricultural monitoring and regional area estimation [5,6]. Cropland in China ranges from contiguous fields in major plains to fragmented fields in topographically complex agro-pastoral transition zones [7–10]. This heterogeneity means that product suitability may depend jointly on surface conditions, class definitions, and the scale of use.

Growing access to Earth-observation data, cloud computing, and machine learning has expanded cropland mapping beyond early medium- and coarse-resolution land-cover products [6,11,12]. Current products include annual 30 m maps, Sentinel-based 10 m products, and meter-scale thematic maps. Representative 30 m products include GlobeLand30 [13], GLC_FCS30 (Global 30 m Land-Cover Dynamic Monitoring Product with a Fine Classification System) [14,15], CLCD (China Land Cover Dataset) [16], CACD (China's Annual Cropland Dataset) [11], and CCropLand30 [17]. These products primarily use Landsat imagery or other medium-resolution data for annual or thematic mapping. Sentinel imagery has enabled 10 m products, including ESA-WC (ESA WorldCover 10 m 2020) [18], ESRI-LC (ESRI Land Cover 10 m 2020) [19], FROM-GLC10 [20], and Dynamic World [21]. Their finer grids provide greater spatial detail for representing cropland patches and boundaries. Recent studies have also combined very-high-resolution Google Earth or Mapbox imagery with deep learning to produce meter-scale products such as SinoLC-1 [22] and CropLayer [23]. These products provide finer delineation of small fields and field boundaries.

The growing diversity of cropland products provides valuable data for national and regional monitoring but complicates product selection. Before these products can support cropland monitoring, area statistics, or agricultural analysis, users need evidence on spatial agreement, classification accuracy, area performance, and conditions of applicability. Existing evaluations mainly use independent reference samples and error matrices to estimate classification accuracy, or examine spatial agreement and compare mapped areas with external statistics [5,24,25]. Harmonized reference-data initiatives provide additional resources for training and validation, while requiring attention to data quality and independence [26]. Zhang et al. [27] evaluated six 30 m cropland products at national, provincial, latitudinal, and longitudinal scales using 33,713 reference samples and statistics. Cui et al. [28] examined landscape fragmentation, climate, and agricultural management in a national comparison, whereas Cui et al. [29] evaluated and refined cropland maps in southwestern China. Peng et al. [30] and Jiang et al. [23] extended evaluations to higher-resolution products and multiple spatial levels. Global and regional comparisons likewise reveal variation across geographical settings, reference protocols, and terrain conditions [31–34]. These studies have established important evidence on product accuracy, spatial disagreement, and regional variation. Building on this evidence, a further question is whether product rankings and recommendations remain consistent when the evaluation target, reference-label protocol, or spatial support changes. Addressing this question requires comparisons that explicitly distinguish these evaluation conditions, particularly in the following three respects. First, existing evaluations do not always relate different performance dimensions explicitly to specific application needs. Spatial agreement, pixel-level classification accuracy, area

bias, and regional performance characterize different aspects of product quality and may therefore lead to different product rankings [5,35]. Aggregate metrics may conceal errors under heterogeneous surface conditions [32,34]. Commission and omission errors affect mapped-area bias and the efficiency of sample-based area estimation differently; a product with high overall accuracy is therefore not necessarily the most useful basis for estimating cropland area [36]. Recent work has explicitly linked map selection to stratified-sampling efficiency in African countries [37]. Consequently, the suitability of different products for cropland-location mapping, regional area estimation, and cropland monitoring in complex agricultural landscapes across China requires further investigation.

Second, class harmonization can confound semantic mismatch with mapping error. Evaluations commonly harmonize distinct classification systems into a shared class to improve comparability. Using a common FAO standard, Xue et al. [5] showed that classification-system differences still affected cropland area, spatial accuracy, and product rankings. Identical class names do not guarantee identical conceptual scopes, temporal attributes, or spatial decision rules [38–40]. Class names or crosswalk tables alone therefore cannot eliminate semantic heterogeneity [41,42]. Under a unified reference label, a pixel that conforms to a product's own definition may still be counted as an error [34,43,44]. This mismatch can alter estimated classification accuracy and area bias [45]. A key question is therefore how much estimated accuracy and area bias change when reference labels follow different cropland definitions while product predictions and the evaluation domain remain fixed. Such comparisons can quantify sensitivity to reference-label semantics without attributing these changes to differences in the underlying maps.

Third, the cross-scale behavior of high-resolution cropland products remains uncertain. Fine-resolution products can delineate field boundaries, small parcels, and fragmented cropland more precisely [22,23]. In practice, these products are often aggregated to coarser grids to match climatic, topographic, ecological, or statistical data. Previous studies have shown that changes in product resolution can alter class proportions and classification outcomes [46,47]. Most evaluations, however, assess products at their native resolutions or compare different products across resolutions. It remains unclear whether the spatial detail of fine-resolution products produces more reliable classifications and cropland-proportion estimates after aggregation to coarser spatial supports.

This study evaluated eight cropland and land-cover products for 2020 across mainland China at spatial resolutions of 2, 10, and 30 m. Products were aligned to a common 30 m spatial support for the main cross-product comparisons. Nationally stratified reference samples and reference plots centered on mixed cropland pixels supported four research questions:

1. Spatial agreement and mapped area: How do the eight products differ in spatial agreement, mapped area, and national and regional distribution patterns?
2. Classification accuracy, area bias, and robustness: Within the consensus-cropland evaluation domain, which products achieve higher pixel-level accuracy and lower design-based area bias? Which products maintain more stable classification performance as inter-product consistency decreases?
3. Semantic responses: Do documented cropland definitions match the observed secondary-class responses? How does reference-label recoding change accuracy and area bias when predictions and evaluation domains remain fixed within each product?
4. Cross-scale behavior: How do classification accuracy and local cropland-proportion error change when fine-resolution products are aggregated to coarser spatial supports?

Our task-oriented framework extends product comparison through a shared semantic benchmark and controlled comparisons of reference-label protocols and within-product aggregation. It clarifies the conditions shaping product assessments and supports selection for spatial identification, regional area estimation, and analyses of specific agricultural types.

2. Materials and Methods

2.1. Cropland and Land-Cover Products

This study selected eight representative cropland and land-cover products for 2020, with global or China-specific coverage and spatial resolutions of 2 m, 10 m, and 30 m (Table 1). Their mapping strategies encompass time-series classification, multi-source feature integration, deep learning, and multi-product fusion.

Table 1. Basic information of the eight cropland/land-cover products.

Product	Format	Data Source	Classification Method	Product-Reported Accuracy (%)	Reference
CLCD	Raster	Landsat	Random Forest	79.31	[16]
CACD	Raster	Landsat	Trajectory-based approach	93	[11]
CCropLand30	Raster	Multi-source fusion	Majority voting-fuzzy agreement score(MV-FAS)	88	[17]
GLC_FCS30	Raster	Landsat, Sentinel-1	Local adaptive random forest	80.88	[14]
GlobeLand30	Raster	Landsat, HJ-1, GF-1 WFV	Pixel-object-knowledge (POK)	85.72	[13]
ESA-WC	Raster	Sentinel-1 Sentinel-2	CatBoost gradient boosting	74.4	[18]
ESRI-LC	Raster	Sentinel-2	Deep learning (U-Net)	85	[19]
CropLayer	Vector	Mapbox, Google	Active learning with Mask2Former XGBoost	88.73	[23]

Accuracy values reported by the original product developers were obtained using different validation protocols and are therefore not directly comparable across products.

Among the 30 m products, CLCD and CACD both use Landsat time series but apply different temporal strategies. CLCD combines spectral and phenological features with random forest classification and spatiotemporal constraints, whereas CACD integrates automated sample generation, random forest probability estimation, and LandTrendr segmentation. GlobeLand30 applies the pixel–object–knowledge approach with object-based segmentation and expert rules, while GLC_FCS30 integrates Landsat, Sentinel-1 SAR, topography, and auxiliary data using locally adaptive random forests. Unlike these directly classified products, CCropLand30 fuses GlobeLand30, GLAD, CLUD, CLCD, and CACD and applies county-level TNLS constraints for area adjustment.

At 10 m, ESRI-LC combines Sentinel-2 imagery with U-Net semantic segmentation, whereas ESA-WC integrates Sentinel-1 and Sentinel-2 features using CatBoost and expert-rule post-processing.

CropLayer is a 2 m cropland vector dataset covering China, including Chinese Taiwan, but excluding Hong Kong and Macao. It was generated from high-resolution Mapbox and Google satellite imagery using a data-driven three-stage mapping framework. The dataset is organized by county-level administrative units and emphasizes field-boundary delineation and the mapping of small cropland parcels.

Cropland definitions vary systematically among products because of differences in classification schemes and mapping objectives (Table 2; Figure 1). All eight definitions include irrigated and rainfed cropland, and all except ESRI-LC explicitly include paddy fields. The main differences concern classes near the semantic boundary of cropland. CACD and ESA-WC restrict cropland to annual herbaceous crops harvestable within 12 months of sowing and explicitly exclude plantations. GlobeLand30 and GLC_FCS30 include orchards

and other plantations in cropland-related classes. CLCD, CCropLand30, and CropLayer adopt broader TNLS-aligned definitions that also include greenhouses, fallow land, and temporary cropland. ESRI-LC excludes paddy fields and plantations but includes fallow land. CACD explicitly specifies a minimum mapped area of 0.25 ha and a minimum width of 30 m.

Table 2. Cropland definitions of the eight cropland/land-cover products.

Product	Cropland Definition
CLCD	Cropland is defined according to the Third National Land Survey (TNLS) and includes paddy fields, irrigated land (including greenhouse cultivation), dryland cropland, and land temporarily used for medicinal plants, turf, flowers, or nursery stock. It also includes newly developed or reclaimed cropland, fallow land, and areas dominated by crop cultivation, even when sparsely interspersed with fruit trees or other vegetation
CACD	Annual cropland is defined as land with a minimum area of 0.25 ha (minimum width of 30 m) that is sown or planted and harvestable at least once within 12 months after the sowing or planting date. This excludes perennial crops, fruit, tea, and coffee plantations, greenhouse crops, and small plots that do not meet the minimum size criteria.
CCropLand30	Same as CLCD definition.
GLC_FCS30	Includes rainfed cropland, irrigated cropland, herbaceous cover, and tree or shrub cover (orchards).
GlobeLand30	Agricultural land includes horticulture, gardens, paddy fields, irrigated and dry farmland, vegetation gardens, fruit gardens.
ESA-WC	Annual crops are sown and harvested within 12 months, characterized by herbaceous cover possibly mixed with woody vegetation. Excludes perennial woody crops (mapped as tree/shrub cover) and greenhouses (mapped as built-up).
ESRI-LC	Human-planted cereals, grasses, and non-woody crops (e.g., corn, wheat, soy, fallow land). Rice paddies and other heavily irrigated and inundated agriculture are listed under flooded vegetation rather than Crops.
CropLayer	Same as CLCD definition.

In the reference labels used here, irrigated cropland denotes non-paddy irrigated cropland; paddy fields, including rice–upland rotations, form a separate category.

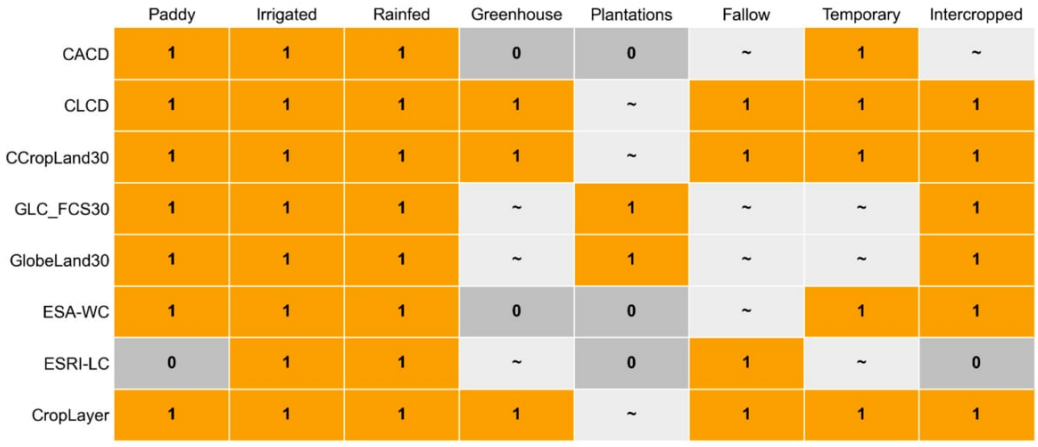


Figure 1. Cropland-class matrix derived from the documented definitions of eight cropland products. The symbols 1, 0 and ~ indicate that the corresponding class is explicitly included, explicitly excluded or not mentioned, respectively, in the documented product definition.

2.2. Ancillary Data

2.2.1. Agricultural Zoning Data

Regional variation was assessed to characterize product performance across agroecological settings. Data for China's nine agricultural zones were obtained from the Resource and Environmental Science Data Center, Chinese Academy of Sciences (Table 3; https://www.resdc.cn/data.aspx?DATAID=275 (accessed on 23 January 2026)).

Table 3. Agricultural zones used in this study and their corresponding spatial extents.

Agricultural Zone	Abbreviation	Spatial Extent
Northeast China Plain	NCP	Heilongjiang, Jilin, and Liaoning
Northern arid and semiarid region	NASR	Inner Mongolia, Ningxia, Xinjiang, and most of Gansu
Huang-Huai-Hai Plain	HHHP	Beijing, Tianjin, Hebei, Henan, and Shandong
Loess Plateau	LP	Shanxi, Shaanxi
Qinghai-Tibet Plateau	QTP	Qinghai, Tibet, western Sichuan, and southern Gansu
Sichuan Basin and surrounding regions	SBSR	Sichuan Basin, Chongqing
Middle-lower Yangtze Plain	MLYP	Hubei, Hunan, Jiangxi, Zhejiang, Anhui, and Jiangsu
Yunnan-Guizhou Plateau	YGP	Yunnan, Guizhou, and Guangxi
Southern China	SC	Guangdong, Fujian, and Hainan

Spatial extents follow the original agricultural zoning dataset, which includes Chinese Taiwan in the Southern China zone. Chinese Taiwan was excluded from all quantitative analyses in this study.

2.2.2. Digital Elevation Model

The NASA/USGS Shuttle Radar Topography Mission 30 m digital elevation model (SRTM DEM) was used to quantify how uncertainty in cropland identification varied with terrain [48].

2.3. Overall Workflow

Eight cropland and land-cover products were evaluated across mainland China to examine performance under different reference definitions, spatial supports, and evaluation tasks (Figure 2). Sampling was stratified by agricultural zone, inter-product agreement, slope, and cropland-edge status, without using any single product as the sampling frame. Neyman allocation guided the selection of 11,972 reference samples, which were visually assigned hierarchical labels for primary land status and secondary agricultural classes. These labels supported consensus-cropland evaluation, class-specific crop-assignment rates, and comparisons between unified-agriculture and product-definition-matched protocols while holding product maps constant. Maps were harmonized to 30 m for inter-product comparisons, while mixed-pixel plots supported controlled within-product aggregation from finer to coarser spatial supports. Assessments covered spatial agreement, design-weighted classification accuracy, design-based area bias, local cropland-proportion error, and robustness across inter-product consistency levels. Together, these comparisons related product performance to semantic definitions, spatial support, and classification or area-estimation tasks.

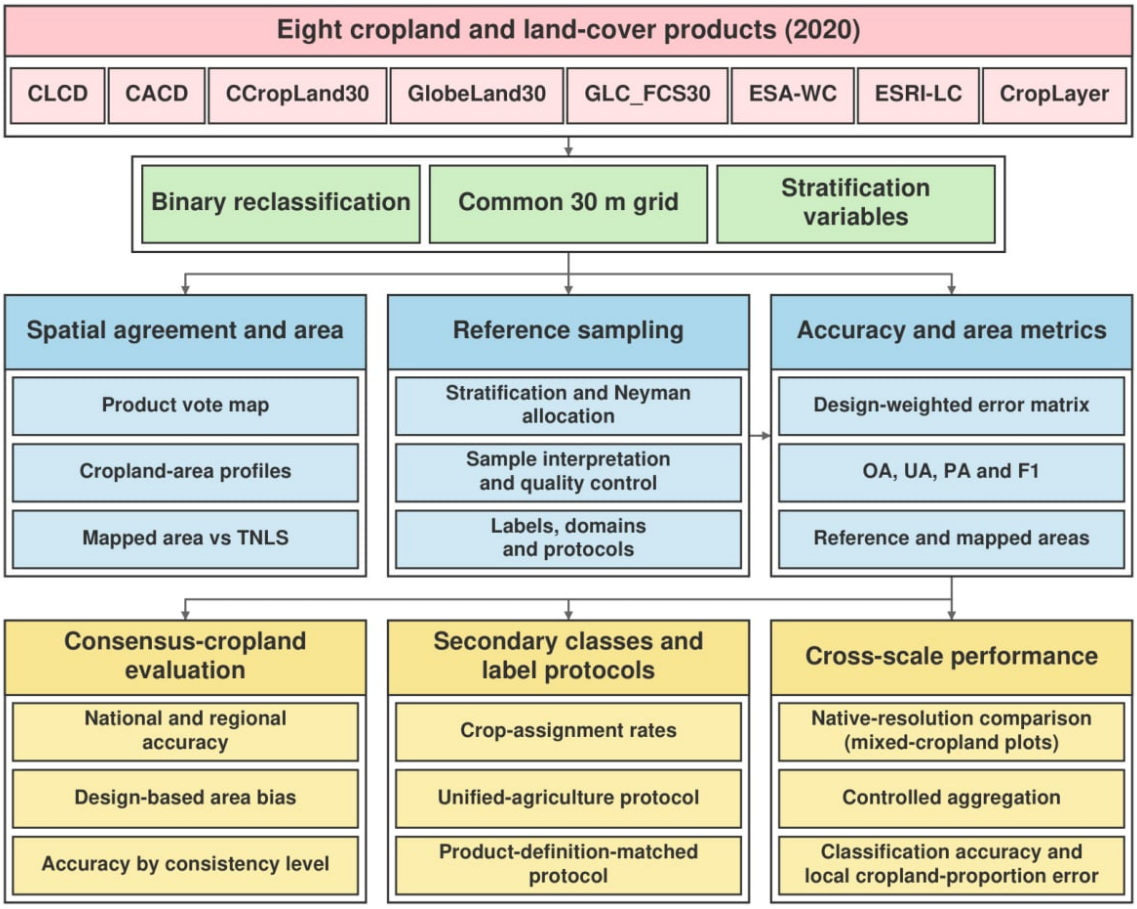


Figure 2. Overall study workflow.

2.4. Data Processing

The eight products were converted to comparable binary layers and aligned to a common 30 m grid for reference sampling and product evaluation.

2.4.1. Binary Reclassification and Spatial Harmonization

Following documented class definitions and original codes, the eight products were reclassified as cropland (1) or non-cropland (0). Table 4 presents the source classes and rules, and Figure 3 presents the resulting distributions. Product overlay, area calculation, stratified sampling, and accuracy assessment were restricted to mainland China, the shared coverage of CCropLand30 and TNLS. All valid classes not assigned to cropland were treated as non-cropland, whereas NoData cells were excluded. Binary harmonization standardized map representations but did not remove differences among product-specific cropland definitions; the effects of these definitional differences were evaluated through reference-label recoding under the product-definition-matched protocol. The 2 m and 10 m products were aggregated to the common 30 m grid by majority rule. A target cell was labeled cropland when cropland occupied at least 50% of its valid subpixels, and non-cropland otherwise. The five 30 m products retained their native resolution and were aligned to the same grid.

Table 4. Binary reclassification rules for the eight cropland and land-cover products.

Product	Original Cropland Class	Original Class Code	Reclassified Binary Code
CLCD	Cropland	1	1
CACD	Cropland	1	
CCropLand30	Cropland	1	
GLC_FCS30	Rainfed cropland	10	
	Herbaceous cover cropland	11	
	Tree or shrub cover (orchards) cropland	12	
	Irrigated cropland	20	
GlobeLand30	Cultivated land	10	
ESA-WC	Cropland	40	
ESRI-LC	Crops	5	
CropLayer	Cropland	1	

All classes listed above were reclassified as cropland (1). All other valid classes were reclassified as non-cropland (0), whereas NoData cells were excluded from subsequent analyses.

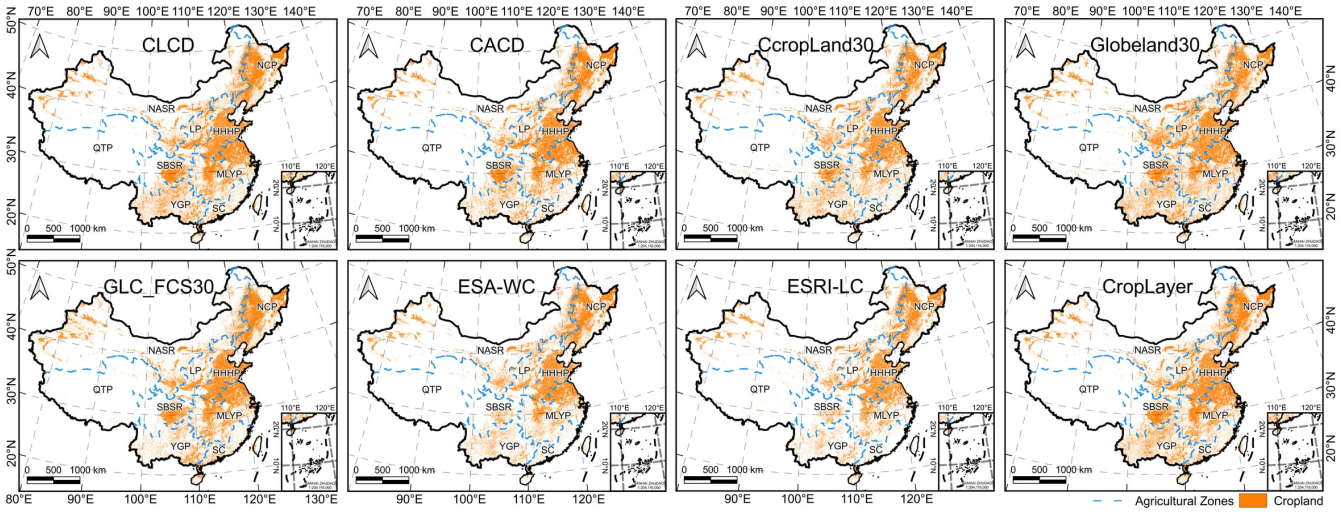


Figure 3. Spatial distribution of cropland mapped by eight cropland products.

2.4.2. Construction of Auxiliary Variables for Sampling

The eight binary products were overlaid on the common 30 m grid. The number of product votes for target cell (i,j) was defined as:

$$V_{ij} = \sum_{k=1}^{N} C_{ij}^{k}, \quad (1)$$

where $C_{ij}^{k} \in \{0, 1\}$ is the binary classification from product k, $N = 8$, and $V_{ij} \in [0,8]$. Pixels were grouped into high-consistency (7–8 votes), medium-consistency (4–6 votes), and low-consistency (1–3 votes) areas. Pixels with $V_{ij} > 0$ formed the union of the eight products. A 150 m buffer was constructed around the union boundary. Pixels with $V_{ij} = 0$ inside this buffer were classified as fringe non-cropland, and the remaining pixels with $V_{ij} = 0$ were classified as background non-cropland. The same eight-product vote map defined the consistency categories used for all product-specific F1 comparisons.

The nine-zone agricultural vector layer and SRTM 30 m DEM were aligned to this common grid before slope derivation. Product-agreement status, agricultural zone, slope class, and cropland-edge status jointly defined the stratification variables used in Section 2.5.1.

2.5. Reference Sampling Strategy and Semantic Labels

Reference samples were stratified and allocated across spatial conditions, then visually interpreted, quality-controlled, and assigned hierarchical semantic labels for design-weighted evaluation.

2.5.1. Stratified Reference Sampling

Reference samples supported national design-based accuracy and area estimation while increasing coverage of product-disagreement areas, cropland edges, complex terrain, and semantic-boundary classes. Sampling strata were defined by overlaying agricultural zones, product-agreement status, and slope classes. Table 5 lists the stratification variables and their definitions.

Table 5. Stratification variables for reference sampling and conditional accuracy assessment.

Stratification Variable	Class	Definition
Agricultural zone	Zone	Nine agricultural zones listed in Table 3
Product consistency level	High consistency	V = 7–8
	Medium consistency	V = 4–6
	Low consistency	V = 1–3
	Fringe consistency	V = 0, within the 150 m buffer around the cropland union of all products
	Background non-cropland	Remaining V = 0 area
Slope class	Low slope	<6°
	Medium slope	6–15°
	High slope	➢ 15°

Slope thresholds followed the cropland-slope classification used by the Third National Land Survey [49]. The key thresholds of 6° and 15° were retained to distinguish major terrain gradients while maintaining adequate sample sizes in multidimensional combined strata. Adjacent classes were merged into low (<6°), medium (6–15°), and high (>15°) slope. These rules produced 135 combined strata with nonzero area. The design sampled every agricultural zone, agreement level, terrain condition, and cropland-edge setting without using any single product as the sampling frame.

2.5.2. Sample Size Allocation and Selection

Sample-size allocation followed the stratified-sampling framework and example parameters summarized by Olofsson et al. [24]. Expected user's accuracy approximated within-stratum binary variance. Neyman allocation then assigned more samples to strata with larger areas or greater expected within-stratum variance [50].

For stratum h, let N_h denote the number of pixels and U_h the expected within-stratum accuracy during sampling design. The expected within-stratum standard deviation was approximated as:

$$S_h = \sqrt{U_h(1 - U_h)} \quad (2)$$

Given a target total sample size n, the initial sample allocation to stratum h under Neyman allocation was:

$$\tilde{n}_h = n \cdot \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h}, \quad (3)$$

where H is the number of valid strata. This study contained 135 combined strata with nonzero area. Following the sample-size range suggested by Olofsson et al. [24], a minimum of 50 samples per combined stratum was set to ensure basic coverage. The target total sample size was set to 12,000 to strengthen regional and conditional analyses while retaining capacity to allocate additional samples to high-variance strata.

Because the true reference-label distribution within each stratum was unknown before sampling, U_h was used to approximate within-stratum variance and guide allocation. These planning values were not used directly in subsequent design-based accuracy or area estimates. U_h was set to 0.95, 0.90, 0.60, 0.60, and 0.70 for background non-cropland, high agreement, medium agreement, low agreement, and fringe non-cropland, respectively. Consistent with the range of illustrative planning values in Olofsson et al. [24], these settings assigned lower anticipated variance to background and high-agreement strata, near-maximum variance to medium- and low-agreement strata, and intermediate variance to fringe non-cropland. After applying the minimum sample size, an upper limit for background non-cropland, integer rounding, and redistribution constraints, 11,972 samples were allocated across the 135 non-empty strata. The sample allocation across all 135 strata, together with the corresponding planning values of U_h, is provided in Supplementary Table S1.

2.5.3. Reference-Sample Interpretation and Quality Control

Reference samples were visually interpreted in Collect Earth Online [51] using April–October 2020 Sentinel-2 and Landsat imagery, supplemented by 2018–2019 and 2021–2022 observations where needed. Imagery and monthly Landsat NDVI and NDWI composites from 2017–2023 supported identification of fallow land, temporary cropland, and intercropping. Google Earth and Mapbox high-resolution imagery provided supplementary spatial evidence; acquisition dates were unavailable in the interface used. Two interpreters assigned primary and secondary labels by cross-checking multi-source imagery and temporal indices, with uncertain cases reviewed by a third interpreter.

2.5.4. Hierarchical Semantic Labels, Evaluation Domains and Label Protocols

A hierarchical reference-label system linked sample interpretation to evaluations of shared cropland semantics, secondary-class responses, and sensitivity to reference-label protocols. At the primary level, samples were labeled as cropland, non-cropland, or mixed cropland pixels. Cropland samples received secondary labels for rainfed cropland, irrigated cropland, paddy fields, fallow land, temporary cropland, greenhouses, intercropped cropland, or perennial crops (Figure 1). Drawing on the TNLS classification, rainfed cropland mainly depends on precipitation, whereas irrigated cropland denotes non-paddy land with water supplies and irrigation facilities for growing non-aquatic crops. Paddy fields form a separate category, including rice fields and rice–upland rotations such as rice–wheat and rice–rapeseed systems. These classes were interpreted using high-resolution imagery and multitemporal vegetation- and water-index observations. Four evaluation scenarios were constructed from these labels.

1. Consensus-cropland evaluation. Rainfed and irrigated cropland, explicitly included in all eight product definitions, were recoded as cropland, while non-cropland labels were

retained. Other secondary classes were excluded, establishing a common semantic domain for product comparison. Paddy fields were excluded because the documented ESRI-LC definition did not include them.

2. Secondary-class response evaluation. For each secondary cropland class, the crop-assignment rate was calculated as the weighted proportion of reference samples mapped as cropland by each product. Rates were interpreted in relation to each product's documented cropland definition.
3. Label-protocol evaluation. Two reference-label protocols were compared within the same extended evaluation domain. The unified-agriculture protocol recoded all secondary agricultural classes as cropland; the product-definition-matched protocol assigned reference labels according to each product's documented definition. Product predictions and domain membership were held constant, so changes in accuracy and area bias estimates reflected reference-label recoding.
4. Cross-scale application evaluation. Area-based reference plots supported comparisons of classification accuracy and local cropland-proportion error, together with controlled within-product aggregation from finer to coarser spatial supports.

2.5.5. Area-Based Reference Plots for Cross-Scale Analysis

After reference interpretation and hierarchical labeling, 674 samples comprising mixed cropland pixels and additional samples labeled as pure cropland or non-cropland were selected from the national stratified sample. A square 120 m × 120 m reference plot was centered on each sample. Cropland polygons were delineated throughout each plot using the imagery described in Section 2.5.3, and the interpreted background was assigned to non-cropland (0). Reference cropland proportions were calculated as cropland area divided by valid evaluation area, excluding NoData locations. Reference and mapped proportions used the same valid support within each evaluation; the denominator was 14,400 m^2 where the entire plot was valid. A geometric check found no overlapping plot footprints.

2.6. Evaluation Methods

Evaluation covered spatial agreement, design-weighted classification accuracy, cropland-area performance, and cross-scale performance. Spatial agreement was quantified using product votes, whereas area performance included distribution profiles, raw mapped area, design-based area bias, and local cropland-proportion error. Metric definitions were held constant across tasks, while evaluation domains, reference labels, and spatial supports varied with the analytical objective.

2.6.1. Spatial-Consistency Analysis

Spatial agreement was evaluated through overall multi-product agreement and pairwise agreement.

For overall agreement, the vote map generated by the spatial overlay in Section 2.4.2 was used to compare products nationally and across the nine agricultural zones. A larger vote count indicated that more products classified a pixel as cropland and therefore represented stronger agreement. A smaller count indicated greater product disagreement. We quantified the spatial distributions and area proportions of each agreement level. Product agreement measures concurrence among products, not classification correctness. Actual accuracy therefore required independent reference samples. Accuracy was further compared across product-agreement categories to examine variation in classification performance [52]. The category-specific F1 scores provide a conditional description of performance within the strata jointly defined by the eight products. Because each evaluated product also contributed

to the definition of these strata, the comparisons should not be interpreted as an independent assessment of classification robustness.

2.6.2. Design-Weighted Accuracy Assessment Across Evaluation Domains

A common design-weighted estimator supported accuracy assessment across evaluation domains and conditions, following the stratified estimation framework for cases where sampling strata differ from map classes [53]. Domain membership and reference-label assignments varied by evaluation task, whereas the original stratum weights were retained.

For stratum h, the area weight was:

$$W_h = A_h / A_{\text{tot}}, \tag{4}$$

where A_h and A_{tot} denote the stratum area and total study-area extent, respectively. Let n_h be the total number of reference samples in stratum h, before restricting the samples to an evaluation domain. Let $n_{hij,k}^{(D)}$ be the number of samples with map class i and reference class j in the evaluation domain, where $i,j \in \{0,1\}1$ denotes cropland, and 0 denotes non-cropland. The study-area proportion associated with each domain-specific error-matrix cell was estimated as:

$$P_{ij,k}^{(D)} = \sum_{h=1}^{H} W_h n_{hij,k}^{(D)} / n_h, \tag{5}$$

Retaining n_h in the denominator preserved the estimated area contribution of domain D within each stratum. The conditional error matrix was then obtained by normalizing these proportions by the estimated domain area share, $P_D = \sum_{i=0}^{1} \sum_{j=0}^{1} P_{ij,k}^{(D)}$:

$$p_{ij,k}^{(D)} = P_{ij,k}^{(D)} / P_D \tag{6}$$

Thus, $P_{ij,k}^{(D)}$ describes a cell's share of the entire study area, whereas $p_{ij,k}^{(D)}$ describes its share within domain D. For a common domain with valid predictions from all products P_D is shared across products.

Classification metrics were calculated from the conditional matrix:

$$OA_k^{(D)} = p_{00,k}^{(D)} + p_{11,k}^{(D)}, \tag{7}$$

$$UA_k^{(D)} = \frac{p_{11,k}^{(D)}}{p_{1\cdot,k}^{(D)}}, \tag{8}$$

$$PA_k^{(D)} = p_{11,k}^{(D)} / p_{\cdot 1,k}^{(D)}, \tag{9}$$

$$F1_k^{(D)} = \frac{2UA_k^{(D)} PA_k^{(D)}}{UA_k^{(D)} + PA_k^{(D)}} \tag{10}$$

Here, $p_{1\cdot,k}^{(D)} = p_{10,k}^{(D)} + p_{11,k}^{(D)}$ and $p_{\cdot 1,k}^{(D)} = p_{01,k}^{(D)} + p_{11,k}^{(D)}$. OA measures overall correctness within the domain; UA and PA quantify cropland reliability and completeness, respectively. Commission and omission-error rates are $1 - \text{UA}$ and $1 - \text{PA}$, respectively; F1 is the harmonic mean of UA and PA.

The weighted estimates supported the consensus-domain, secondary-class, and label-protocol evaluations defined in Section 2.5.4. For conditional accuracy assessment, weighted error-matrix cells were pooled across the three slope strata within each combination of agricultural zone and inter-product consistency level before calculating product-specific F1.

Uncertainty was quantified using a stratified percentile bootstrap [54]. Reference samples were resampled with replacement within each original sampling stratum, retaining the original stratum sample sizes and area weights. Domain-specific error matrices and OA, UA, PA, and F1 were recalculated for each of 400 replicates under the relevant evaluation domain and reference-label protocol. The 95% confidence limits were the 2.5th and 97.5th percentiles of the bootstrap estimates.

2.6.3. Analysis of Cropland Area Performance

Cropland-area performance was evaluated through longitudinal and latitudinal area profiles, raw mapped-area comparisons with TNLS statistics, and design-based bias estimation within the consensus-cropland domain.

For the area profiles, binary maps on the common 30 m grid were aggregated to a 1° × 1° WGS84 grid using average resampling. Cropland area in each cell was calculated as the aggregated cropland fraction multiplied by its geodesic area. Cell areas were then summed within 1° longitude and latitude bands to obtain the corresponding profiles.

Raw mapped cropland areas on the common 30 m grid were compared with TNLS statistics at national and agricultural-zone scales. TNLS-relative area deviation was calculated as the mapped area minus the TNLS area, divided by the TNLS area and multiplied by 100%. This comparison describes agreement with official area totals, whose thematic scope differs from the consensus-cropland domain, and does not directly assess classification accuracy.

To complement the TNLS-relative area comparison, design-based estimates of reference cropland area and area bias were calculated within the consensus-cropland domain. Using the design-weighted error matrix described in Section 2.6.2, reference cropland area within the evaluation domain D was estimated as:

$$\hat{A}_{\text{ref}} = A_{\text{tot}} P_{\cdot 1}^{(D)} \tag{11}$$

Here, $P_{\cdot 1}^{(D)}$ denotes the estimated proportion of the total study area occupied by reference cropland within domain D. This reference proportion is shared across products in the consensus-cropland evaluation.

The corresponding mapped cropland area for product k within the same evaluation domain D was estimated as:

$$\hat{A}_{\text{map},k} = A_{\text{tot}} P_{1\cdot,k}^{(D)} \tag{12}$$

Here, $P_{1\cdot,k}^{(D)} = P_{10,k}^{(D)} + P_{11,k}^{(D)}$ denotes the estimated proportion of the total study area mapped as cropland by product k within domain D.

Area bias was expressed relative to the estimated reference cropland area:

$$Bias_k = \frac{\hat{A}_{\text{map},k} - \hat{A}_{\text{ref}}}{\hat{A}_{\text{ref}}} \times 100\% \tag{13}$$

Positive and negative biases indicate overestimation and underestimation, respectively. Mapped-area estimates outside the reference area's 95% confidence interval were flagged for interpretation only; this comparison was not treated as a formal significance test. The same estimation framework was applied at national and agricultural-zone scales.

2.6.4. Cross-Scale Performance Analysis

Cross-scale analysis compared classification accuracy and local cropland-proportion error among the eight products at their native resolutions and within products before and after controlled aggregation. Native-resolution comparisons used reference polygons rasterized onto each product's native grid. Controlled aggregation comparisons used intersection areas between product pixels and reference polygons.

CropLayer was aggregated from 2 m to 10 m and 30 m, whereas ESA-WC and ESRI-LC were aggregated from 10 m to 30 m. The five 30 m products were evaluated at their native resolution. Aggregation followed the 50% majority-rule threshold used for harmonization to the common 30 m grid, assigning cropland when at least half of the valid subpixels were classified as cropland. This represents the predominant binary class within each target cell and allows the cross-scale analysis to examine the effects of the harmonization procedure

used in the main evaluation. Within-product comparisons included all three resolution pairs for CropLayer and the 10–30 m pair for ESA-WC and ESRI-LC.

For both comparisons, 95% confidence intervals were calculated from the 2.5th and 97.5th percentiles of 1000 bootstrap estimates, resampling whole plots with replacement and recalculating metrics from pooled confusion areas. Within-product changes were calculated as coarser minus finer support using the same resampled plots at both supports. This procedure retained within-plot structure but assumed independence between plots. Between-plot spatial dependence was not explicitly modeled and could affect confidence-interval coverage. The results apply to the evaluated reference plots and the specified aggregation rule.

3. Results

Results are presented in four parts: spatial agreement and raw mapped area, consensus-cropland classification accuracy and design-based area bias, secondary-class responses and label-protocol sensitivity, and cross-scale performance.

3.1. Agreement, Spatial Variation, and Mapped Area Differences Among the Products

Spatial overlay revealed pronounced regional variation in inter-product agreement (Figure 4a,b). Nationally, only 26.8% of pixels in the cropland union were mapped as cropland by all eight products. High-agreement pixels mapped as cropland by at least seven products accounted for 39.0% of the union and were concentrated in NCP, HHHP, and MLYP. Low-agreement areas occurred mainly in QTP and complex terrain across southwestern China. In QTP, 72.7% of union pixels were mapped as cropland by three or fewer products. Product disagreement was therefore spatially concentrated in complex terrain and agro-pastoral transition zones.

Longitudinal and latitudinal area profiles showed broad agreement in the locations of China's main cropland belts (Figure 4b,c), but differed markedly in peak width and rates of decline toward the margins. GlobeLand30, CLCD, and GLC_FCS30 generally showed higher profiles, broader peaks, and more gradual declines toward the margins. CropLayer showed lower profiles, narrower peaks, and more rapid declines, whereas ESA-WC and ESRI-LC were intermediate. Inter-product dispersion increased markedly across southwestern China.

Using TNLS statistics as an external reference, raw mapped cropland area varied substantially among products across mainland China (Figure 5b). CropLayer was closest to TNLS (−0.2%) and was the only product with a mapped area below the TNLS area. ESRI-LC and ESA-WC were 4.6% and 13.4% above TNLS, respectively, whereas CLCD and GlobeLand30 exceeded TNLS by 48.2% and 56.6%. Regional differences were larger (Figure 5a). Most products were comparatively close to TNLS in NCP, whereas discrepancies increased in QTP, SBSR, and YGP. In QTP, CACD and CLCD were 61.0% and 32.0% below TNLS, respectively, whereas GlobeLand30 was 101.0% above TNLS.

Overall, differences among products in spatial overlap, area profiles, and mapped area were more pronounced in regions with complex terrain and heterogeneous agricultural landscapes.

3.2. Classification Accuracy, Area Bias, and Performance Variation Within the Consensus-Cropland Evaluation Domain

The consensus-cropland evaluation domain comprised rainfed cropland, irrigated cropland, and unambiguous non-cropland. Independent reference samples supported assessments of national classification accuracy, national and regional area bias, spatial error patterns, and classification performance across inter-product consistency levels.

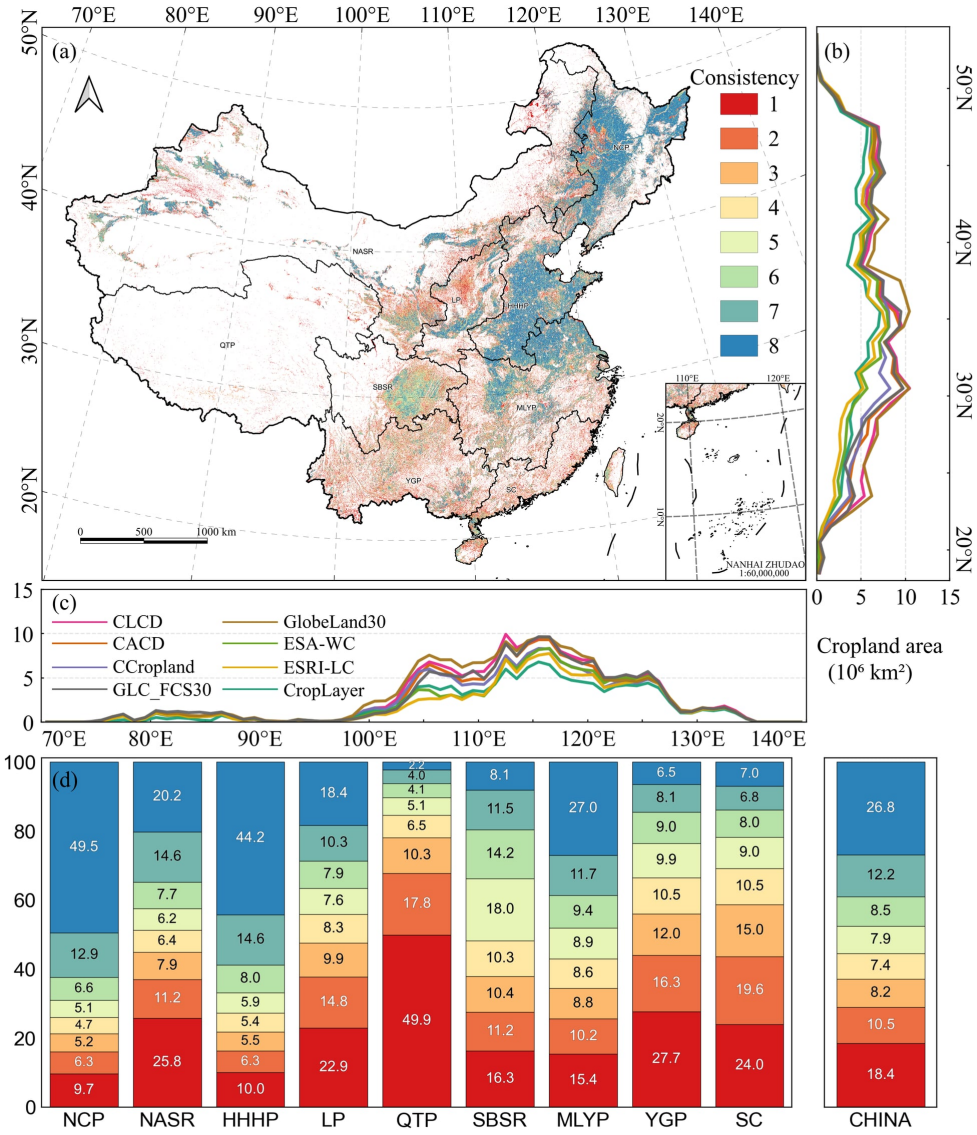


Figure 4. Spatial consistency and latitudinal and longitudinal cropland-area profiles across eight cropland products. (**a**) Spatial consistency map showing the number of products identifying each pixel as cropland within the union of all products, where legend values 1–8 represent the number of products voting for cropland. (**b**) Latitudinal cropland-area profiles aggregated using 1° latitude bands. (**c**) Longitudinal cropland-area profiles aggregated using 1° longitude bands. (**d**). Proportions of pixels at each consistency level within the nine agricultural zones and at the national scale.The colors in (**d**) correspond to the consistency levels (1–8) and their respective colors in the legend of (**a**).

3.2.1. Classification Accuracy and National and Regional Area Bias Across China

After excluding 938 samples representing cropland classes outside the consensus definition, 11,034 valid reference samples remained. National classification performance varied substantially among products. CropLayer achieved the highest OA and F1 (0.961 and 0.844), followed by ESA-WC (0.958 and 0.839; Table 6). CCropLand30 performed best among the five 30 m products (OA = 0.944; F1 = 0.797). GlobeLand30 and GLC_FCS30 had the lowest F1 (0.718 and 0.728, respectively), reflecting their low UA.

Regional biases were generally smaller for CropLayer, ESRI-LC, and ESA-WC than for the 30 m products, although substantial local deviations remained (Figure 6). CropLayer underestimated area in eight zones and slightly overestimated it in SC (+2.21%). ESRI-LC and ESA-WC had low to moderate biases in most zones but underestimated area in QTP, MLYP, and SC. Their MLYP biases reached −46.56% and −25.05%, respectively. CCropLand30,

GLC_FCS30, and GlobeLand30 overestimated area in all nine zones, with GlobeLand30 exceeding +100% in MLYP, YGP, and SC. CACD and CLCD underestimated area in QTP by 67.06% and 33.66%, respectively, but overestimated it elsewhere.

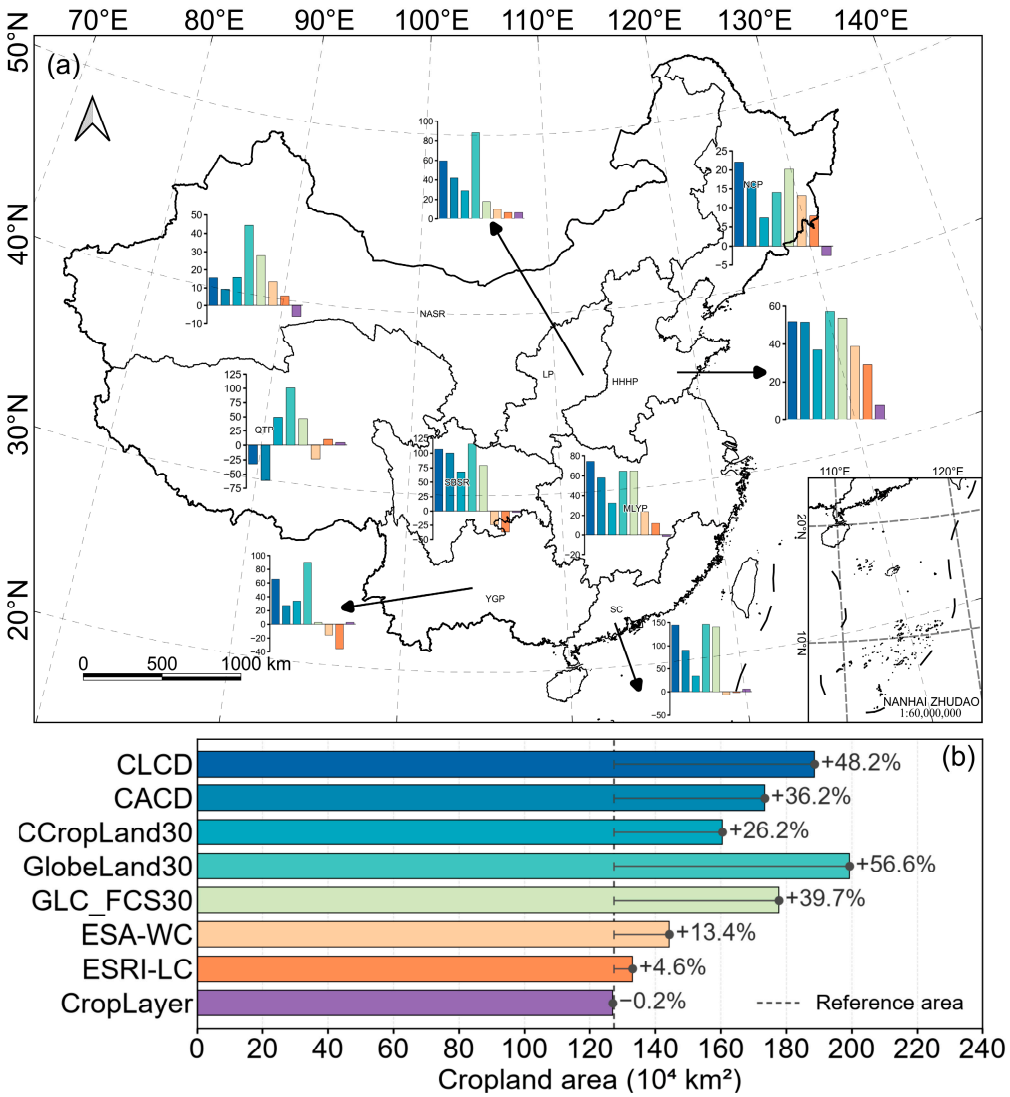


Figure 5. Comparison of cropland areas from eight products with statistics from the Third National Land Survey (TNLS). (**a**) Percentage area deviations across nine agricultural zones. (**b**) National mapped cropland areas and percentage deviations from TNLS.

Table 6. National design-weighted classification-accuracy metrics for the eight cropland products within the consensus-cropland evaluation domain.

Product	OA	UA	PA	F1
CACD	0.936	0.696	0.877	0.776
CLCD	0.928	0.657	0.908	0.762
CCropLand30	0.944	0.734	0.871	0.797
GlobeLand30	0.911	0.599	0.898	0.718
GLC_FCS30	0.920	0.642	0.840	0.728
ESA-WC	0.958	0.809	0.870	0.839
ESRI-LC	0.944	0.785	0.767	0.776
CropLayer	0.961	0.868	0.820	0.844

95% confidence intervals for OA, UA, PA, and F1 are provided in Table S2. The bootstrap procedure is described in Section 2.6.2.

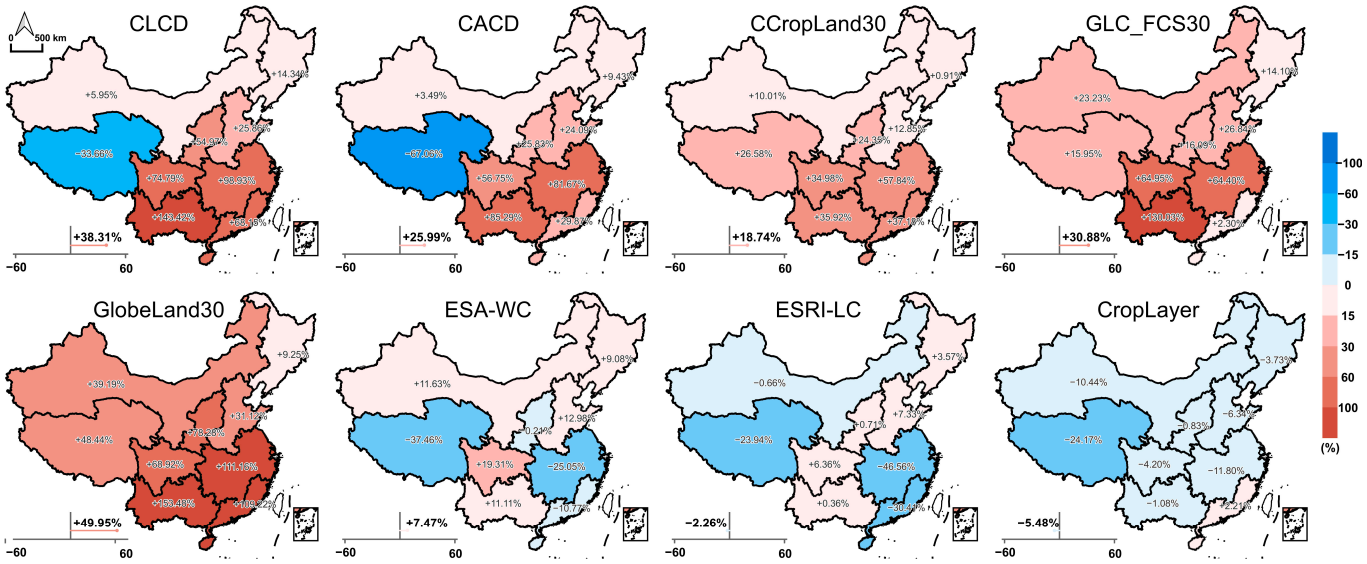


Figure 6. Design-based area bias rates of eight cropland products within the consensus-cropland evaluation domain. Maps show biases across nine agricultural zones; insets show national biases.

National area biases were calculated relative to a design-based consensus-cropland reference estimate of 1.154×10^6 km^2 (95% CI: 1.123–1.180×10^6 km^2). Bias directions agreed with the UA–PA relationship (Table 6; Figure 6). All five 30 m products overestimated area (+18.74% to +49.95%), whereas CropLayer and ESRI-LC slightly underestimated it (−5.48% and −2.26%, respectively). ESA-WC showed a positive bias of +7.47%.

Commission errors were widespread across eastern and southern China in the 30 m products, particularly GlobeLand30 and CLCD (Figure 7). Omission errors occurred along the eastern margin of QTP and across southwestern China, including in CropLayer and ESA-WC, which had fewer error samples overall. Omission samples outnumbered commission samples for CropLayer and ESRI-LC, whereas ESA-WC showed a more balanced distribution.

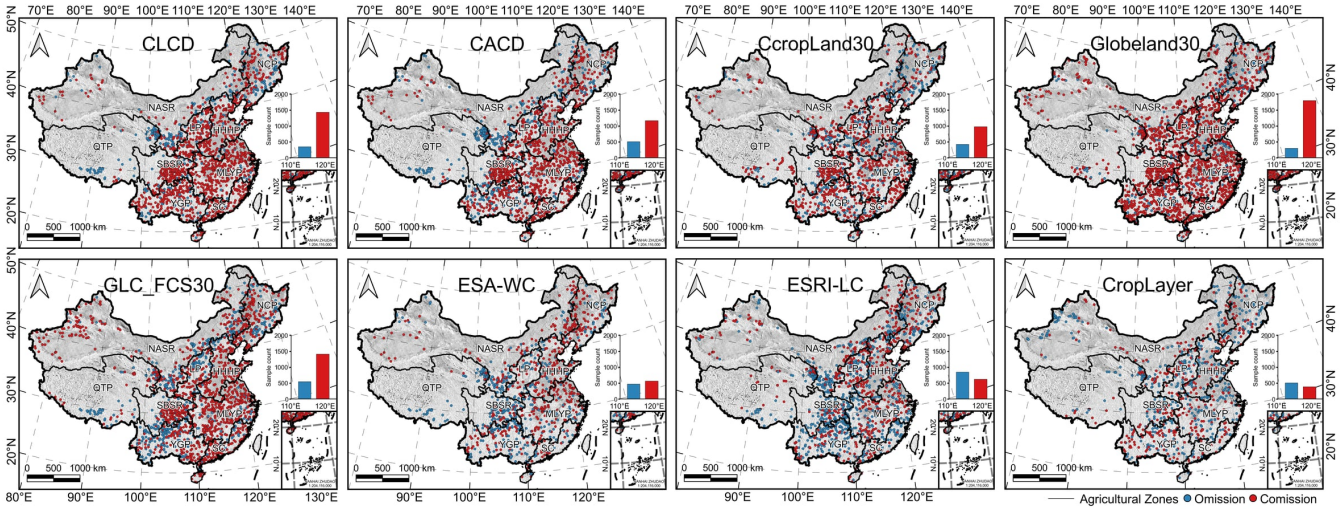


Figure 7. Spatial distribution of commission- and omission-error samples across eight cropland products. Inset bar charts show the unweighted number of samples in each error category.

3.2.2. Classification Performance Across Inter-Product Consistency Levels

Beyond area bias, F1 was examined across inter-product consistency levels (Figure 8).

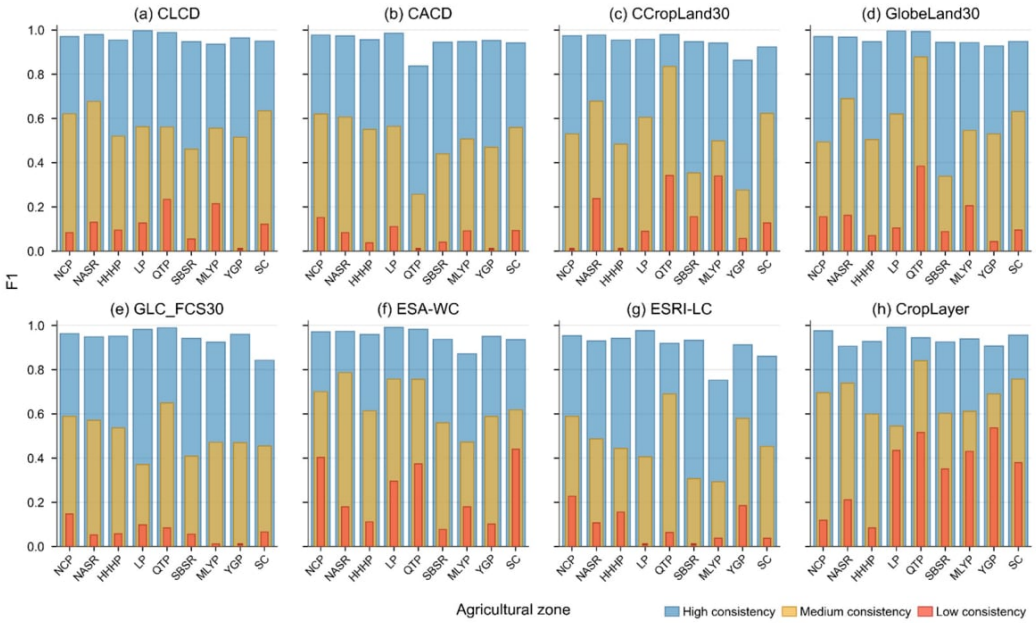


Figure 8. Product-specific classification performance across product-consistency levels within agricultural zones.

F1 values were generally similar under high inter-product consistency but diverged as consistency decreased. CropLayer and ESA-WC generally showed higher F1 under medium and low consistency. CCropLand30 and GlobeLand30 showed intermediate F1, whereas ESRI-LC, CLCD, CACD, and GLC_FCS30 exhibited lower F1 in the low-consistency subsets.

3.3. Secondary-Class Responses and Sensitivity to Reference-Label Protocols

Beyond the consensus-cropland evaluation, the extended semantic evaluation set was used to compare secondary-class crop-assignment rates and evaluate sensitivity to reference-label protocols. Product maps remained unchanged; only reference labels were recoded under the unified-agriculture and product-definition-matched protocols.

Weighted crop-assignment rates varied most among products for semantic-boundary classes, including greenhouses (29.5–91.3%), temporary cropland (19.2–87.2%), and intercropped cropland (24.6–91.7%; Figure 9). Variation also occurred within the consensus-cropland classes: rates ranged from 74.7% to 90.2% for rainfed cropland and from 61.6% to 100.0% for irrigated cropland.

Documented definitions only partly matched mapped responses. ESA-WC had low assignment rates for greenhouses (29.5%) and plantations (38.3%), consistent with their documented exclusion. GlobeLand30 and GLC_FCS30 include plantations and intercropped cropland and generally showed high assignment rates for both classes. However, several responses were inconsistent with the documented definitions. Although CropLayer includes intercropped cropland, only 24.6% of these samples were mapped as cropland. ESRI-LC excludes paddy fields and intercropped cropland, yet mapped 90.5% and 65.4% of the corresponding samples as cropland, respectively. Similarly, CACD excludes greenhouses but mapped 84.2% of greenhouse samples as cropland.

Table 7 compares classification accuracy and design-based area bias under the two label protocols. Switching from the unified-agriculture to the product-definition-matched protocol changed classification accuracy and design-based area bias only for CACD, ESA-WC, and ESRI-LC (Table 7). For ESRI-LC, F1 decreased from 0.765 to 0.682, and area bias reversed from −8.08% to +14.88%. For CACD, F1 decreased from 0.775 to 0.756, while area bias increased from +19.50% to +31.91%. For ESA-WC, F1 increased slightly from 0.819 to

0.823, but area bias increased from +0.52% to +10.92%. Results for the other five products remained unchanged because both protocols assigned identical reference labels.

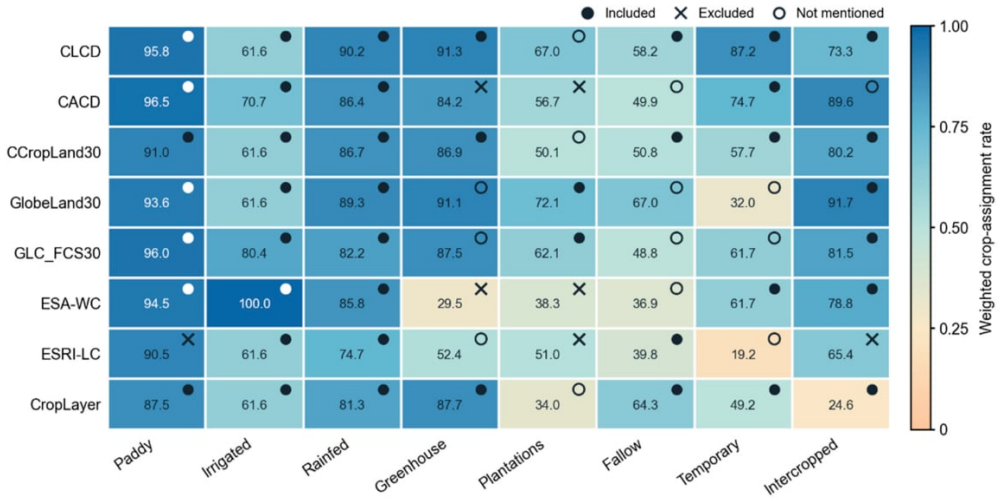


Figure 9. Design-weighted crop-assignment rates across secondary cropland classes and their correspondence with documented product definitions. Cell values are percentages; symbols indicate class inclusion status in the documented definitions.

Table 7. Accuracy and area bias estimates under unified-agriculture and product-definition-matched reference-label protocols.

Product	Reference-Label Protocol	OA	UA	PA	F1	Area Bias (%)
CACD	Unified-agriculture	0.931	0.711	0.850	0.775	19.50
	Product-definition-matched	0.929↓	0.664↓	0.876↑	0.756↓	31.91↑
ESA-WC	Unified-agriculture	0.950	0.817	0.822	0.819	0.52
	Product-definition-matched	0.953↑	0.783↓	0.868↑	0.823↑	10.92↑
ESRI-LC	Unified-agriculture	0.936	0.799	0.734	0.765	−8.08
	Product-definition-matched	0.923↓	0.638↓	0.732↓	0.682↓	14.88↑
CLCD	Same under both protocols	0.927	0.667	0.896	0.765	34.34
CCropLand30	Same under both protocols	0.941	0.742	0.857	0.795	15.44
CropLayer	Same under both protocols	0.960	0.873	0.811	0.841	−7.18
GlobeLand30	Same under both protocols	0.909	0.616	0.882	0.726	43.20
GLC_FCS30	Same under both protocols	0.917	0.658	0.822	0.731	24.93

Arrows indicate numerical changes under the product-definition-matched protocol relative to the unified-agriculture protocol; product maps were unchanged. The 95% confidence intervals for the accuracy and area-bias estimates are provided in Table S3. The bootstrap procedure is described in Section 2.6.2.

Overall, the two protocols distinguished performance under a unified agricultural definition from performance under each product's documented definition. Documented definitions only partly matched mapped responses, while reference-label recoding altered accuracy and area bias estimates without changing product maps.

3.4. Cross-Scale Product Performance

Within the evaluated mixed-cropland reference plots, classification performance and local cropland-proportion error varied among the eight products (Figure 10a,b). CropLayer had the highest F1 (0.793), followed by ESA-WC (0.707), whereas the other products ranged from 0.618

to 0.669. PA exceeded UA for all five 30 m products, indicating that commission errors predominated. These products also had positive local cropland-proportion errors of 13.8–25.0 percentage points, indicating local overestimation. By contrast, UA slightly exceeded PA for ESRI-LC and CropLayer, indicating omission-dominated error structures, although their local errors were close to zero (−0.8 and −0.9 percentage points, respectively). Thus, classification-error structure and local cropland-proportion error were not fully aligned near cropland boundaries.

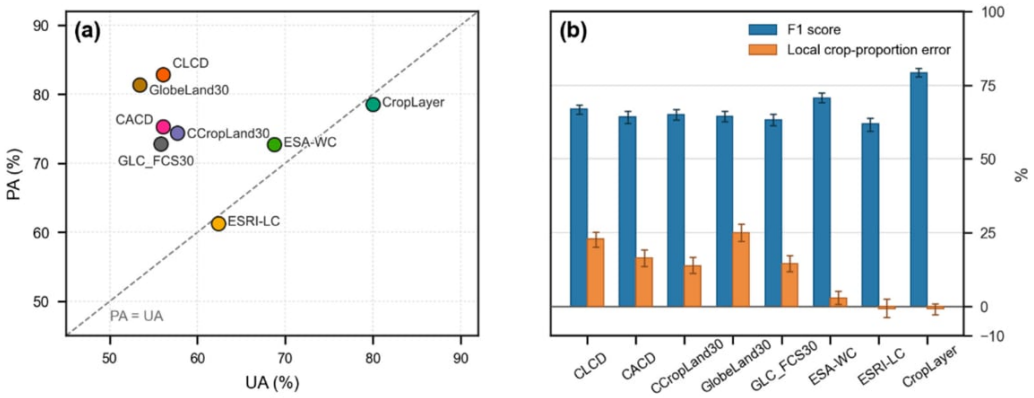


Figure 10. Classification performance and local cropland-proportion error within the evaluated mixed-cropland reference plots. (**a**) UA–PA relationships. (**b**) F1 scores (%) and local cropland-proportion errors (percentage points). Reference and mapped cropland proportions were derived from interpreted reference labels and product predictions, respectively. For each product, both proportions were calculated as cropland area divided by the same valid evaluation area, with areas pooled across plots. Local cropland-proportion error was calculated as mapped minus reference proportion. Error bars in (**b**) indicate 95% confidence intervals calculated using the plot-level bootstrap procedure described in Section 2.6.4.

Under the specified aggregation rule, CropLayer showed statistically significant F1 reductions across all three comparisons, ranging from 0.8 to 4.0 percentage points. The largest reduction occurred from 2 m to 30 m (change: −3.98 percentage points; 95% CI: −4.37 to −3.64). ESA-WC also showed a statistically significant F1 reduction of 0.8 percentage points after aggregation from 10 m to 30 m. For both products, the paired 95% confidence intervals were entirely below zero, supporting a decline in classification performance following aggregation (Figure 11a). In contrast, ESRI-LC's interval included zero, providing no clear statistical evidence of an F1 change. All paired intervals for changes in absolute local cropland-proportion error also included zero, providing no clear evidence of changes in this metric within the evaluated plots (Figure 11b).

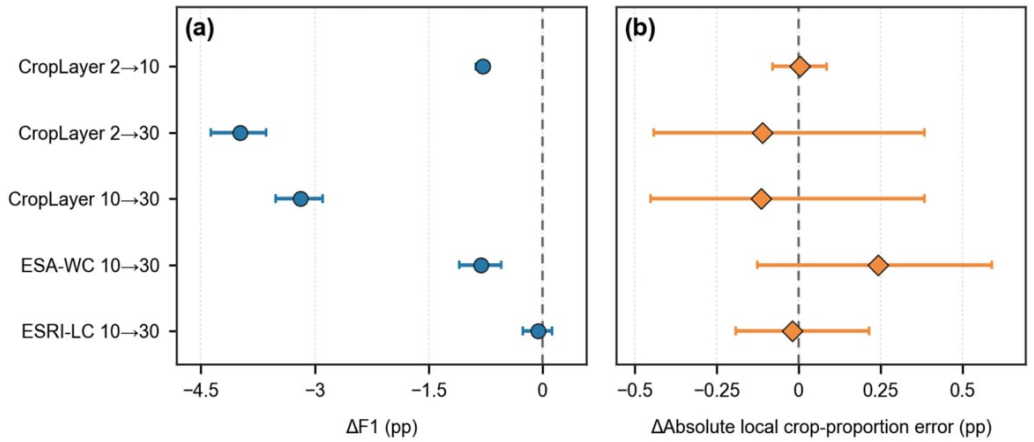


Figure 11. Paired changes following controlled spatial aggregation within the evaluated mixed-cropland reference plots: (**a**) F1 score and (**b**) absolute local cropland-proportion error. Changes

were calculated as coarser minus finer support and expressed in percentage points. Points indicate estimated changes; error bars show 95% confidence intervals from the paired plot-level bootstrap described in Section 2.6.4. Vertical dashed lines indicate zero change. Row labels denote aggregation from finer to coarser supports in meters.

4. Discussion

4.1. Stability and Variation Across Evaluation Frameworks

Broad performance patterns were relatively stable across evaluation frameworks, but exact rankings depended on the reference quantity and thematic scope (Figure 12a). Regional area deviations were more sensitive than national patterns, with some products changing direction across frameworks (Figure 12b). National agreement, therefore, does not necessarily imply suitability for regional area estimation.

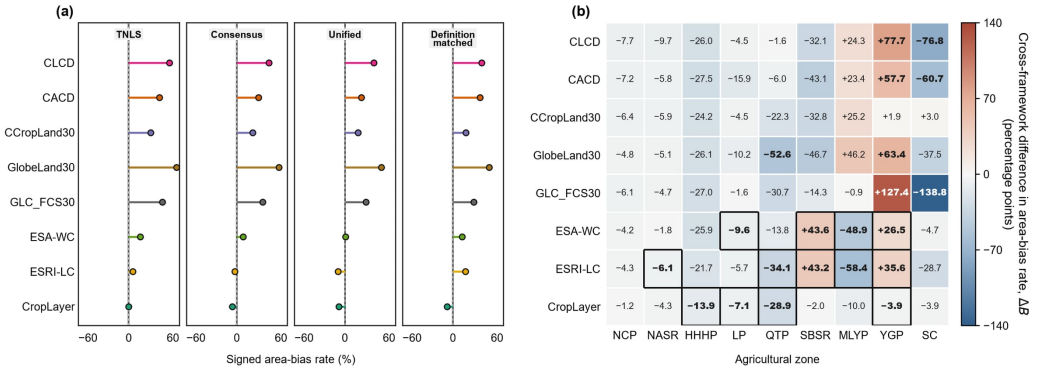


Figure 12. Relative cropland-area deviations under different evaluation frameworks. (**a**) National TNLS-relative area deviation and design-based area-bias estimates from the three sample-based frameworks. (**b**) Regional differences between consensus-domain design-based area bias and TNLS-relative area deviation, calculated as $\Delta B = B_{consensus} - B_{TNLS}$. Black outlines indicate sign reversals; bold values highlight sign reversals or absolute differences ≥50 percentage points. The contrasts reflect differences in thematic scope and reference quantities across frameworks.

Classification performance was comparatively stable between the consensus-cropland and unified-agriculture evaluations, whereas area-bias estimates were more sensitive (Tables 6 and 7). Product-definition matching introduced further product-specific changes, indicating that conclusions about performance depend partly on which agricultural categories the reference standard includes.

These patterns reflect differences in both error metrics and evaluation targets. F1 combines UA and PA, whereas area bias reflects the net balance of design-weighted commission and omission errors. A lower F1 can therefore coexist with a smaller absolute area bias, as illustrated by ESRI-LC relative to CropLayer within the consensus-cropland domain. Despite this ranking difference, the UA–PA imbalance remained consistent with the direction of area bias for both products. Frameworks also changed the evaluation target through the external TNLS area benchmark, the restricted consensus-cropland domain, and reference-class reassignment under the two label protocols. These distinctions help explain why ESA-WC and ESRI-LC responded differently across frameworks. Framework-dependent rankings should therefore be interpreted relative to the chosen reference basis and cropland definition, rather than as universal measures of product quality.

4.2. Sources of Product Differences and Implications for Evaluation

Class definitions determine which agricultural categories a product intends to include, whereas reference-label protocols determine how its predictions are evaluated. The protocol comparison changed reference labels while retaining the evaluation domain and predictions

for each product. Its metric changes therefore reflect a change in the evaluation criterion, not an improvement or deterioration of the underlying map. Documented definitions alone also cannot establish actual class coverage, because their implementation depends on training labels, observable image features, and mapping procedures.

Temporal information influences whether the image evidence distinguishes categories that are separate in the documented definitions. The transition from early-season inundation to subsequent canopy growth provides phenological evidence for paddy-rice identification, as illustrated by Ahmad et al. [55] using Sentinel-1/2 observations. When observations mainly capture developed crop canopies, their signatures may resemble other cropland more closely than flooded vegetation. This provides a possible explanation for ESRI-LC's paddy-field inclusion despite its documented class distinction (Figure 9). Observation gaps and temporal compositing may similarly weaken evidence of brief cultivation or harvest events, complicating the separation of fallow land and temporary cropland. Brief irrigation events may also be missed, introducing uncertainty into the distinction between rainfed and non-paddy irrigated cropland. Applications targeting these categories therefore require attention to temporal coverage as well as class definitions.

Training samples and mapping strategies influence both product differences and inter-product dependence. Pixel classification, temporal rules, and deep-learning segmentation use spectral, temporal, and spatial information differently, particularly at mixed boundaries and in heterogeneous backgrounds. CACD, for example, derives training samples from existing maps, screens them using vegetation time series, and applies spatial–temporal post-processing [11]. If training labels do not consistently represent greenhouse exclusions, learned class boundaries may extend beyond the intended cropland scope. Spatial filtering may additionally favor surrounding cropland labels at greenhouse margins. These processes offer possible explanations for the mismatch between documented exclusions and mapped greenhouse responses.

Product fusion may further propagate shared source errors and create dependence among maps. CCropLand30, for example, combines multiple candidate products and applies area constraints, so its agreement with source products may partly reflect their shared information. The unanimity rate also requires careful interpretation because its denominator is the union of all eight products, including disputed pixels. It therefore describes agreement within the combined mapped extent, rather than the proportion of actual cropland consistently identified. Product agreement can help locate stable and disputed areas and guide targeted verification. Independent reference samples remain necessary to establish classification accuracy, because agreement alone does not demonstrate correctness.

4.3. Spatial Support and Cross-Scale Performance of High-Resolution Products

Spatial aggregation can alter the representation of cropland boundaries and small patches within mixed-cropland settings. Under the 50% majority rule, each target cell receives a single binary label, replacing finer cropland–non-cropland patterns. Small cropland patches may disappear, while non-cropland gaps may be absorbed into surrounding cropland. These changes offer a possible explanation for the F1 reductions following aggregation (Figure 11a).

The implications differ between pixel-level classification and cropland-proportion estimation. F1 is affected by both commission and omission errors, whereas signed cropland-proportion error reflects their net difference in area. Opposing errors can therefore partly offset in pooled cropland proportions, even when classification agreement deteriorates. This provides a possible explanation for the less consistent response of local cropland-proportion error to aggregation (Figure 11b). Consequently, similar cropland proportions can coexist with differences in the spatial allocation of cropland.

Within the evaluated mixed-cropland plots, the classification advantage of finer spatial support did not consistently translate into more accurate local cropland-proportion estimates. Because these comparisons focused on mixed-cropland conditions under a single 50% majority rule, the findings cannot be directly extended to all agricultural landscapes. For applications in similar settings, retaining fine spatial detail may benefit pixel-level cropland identification, while local area estimation requires assessment of cropland-proportion error. When higher-resolution products are aggregated for comparison or joint use, both metrics should be evaluated at the intended spatial support.

4.4. Recommendations for Product Use

Cropland product selection should match the intended application. Food-security assessment, area estimation, and agricultural monitoring require different class definitions, spatial-identification capabilities, and target agricultural types. No single overall accuracy metric can therefore identify a universally best product.

The consensus-cropland results support product comparison within the shared scope of rainfed and irrigated cropland. CropLayer and ESA-WC are candidates for spatial cropland identification. For raster-oriented workflows, CropLayer's vector format requires additional rasterization and grid-alignment steps. Within this domain, ESRI-LC had the smallest design-based area bias and may be suitable for area estimation. Given its documented exclusion of paddy fields, applications in rice-dominated regions should also consider the secondary-class responses and extended-domain results.

For broader agricultural coverage, CLCD may provide a useful baseline because it retained relatively high proportions of several secondary agricultural classes. A multi-product union may likewise support candidate-area screening, but its extensive low-agreement component requires verification before area estimation. CLCD's inclusion of temporary cropland and fallow land may also help retain these areas in monitoring masks. However, distinguishing temporary inactivity from persistent abandonment requires multiyear observations of phenology and cultivation activity.

Secondary-class responses can further guide selection when inclusion criteria are explicit. ESA-WC's higher assignment rate for intercropped cropland than perennial crops may suit applications seeking this distinction. GlobeLand30 and CACD provided broader coverage of intercropped samples, although inclusion must be weighed against commission errors and area overestimation. These responses describe inclusion in a binary cropland class, not direct identification of cropping systems or perennial-crop types.

Regional agreement provides an additional basis for selection. Similar performance under high agreement supports considering CCropLand30 as a baseline for spatial assessment at 30 m. CropLayer and ESA-WC showed greater robustness under low agreement, supporting their consideration in transition zones and complex terrain. Such areas still require targeted verification and explicit uncertainty reporting.

4.5. Limitations and Future Directions

This study has five limitations. First, fallow land, temporary cropland, intercropped cropland, greenhouses, and perennial crops vary seasonally and with management practices. Although multi-source imagery and time series supported the interpretation of reference samples, labels for some semantic-boundary classes may remain uncertain. Second, the acquisition periods of input imagery, temporal compositing methods, and update cycles differed among products. Some observed differences may therefore partly reflect temporal mismatches among products. Third, the evaluation was limited to eight products nominally representing 2020 and to mainland China. The transferability of the findings and product-use recommendations to other years and regions remains to be tested. Fourth, reference-label

recoding addressed differences in cropland class inclusion, but reference interpretation did not apply product-specific minimum patch-size criteria. Disagreements involving small cropland patches may therefore partly reflect differences in these thresholds, whose contribution was not separately quantified. Fifth, operational considerations, including data volume, access permissions, and the availability of multiyear product archives, were not systematically evaluated.

Future evaluations should include additional independent reference samples in high-disagreement settings and for semantic-boundary classes. Multitemporal reference data could improve the reliability of interpreting seasonal agricultural types. Cross-scale evaluations should examine alternative aggregation rules, a broader range of target spatial supports, and continuous representations of cropland proportion. Multiyear and cross-regional evaluations are also needed to test the temporal stability and spatial transferability of the findings and recommendations.

5. Conclusion

A task-oriented framework evaluated eight cropland and land-cover products for 2020 across mainland China using stratified reference samples and design-based estimation. Hierarchical reference labels connected a shared consensus-cropland domain with evaluations of secondary-class responses and label-protocol sensitivity. Holding product maps constant during reference-label recoding, alongside controlled within-product aggregation, enabled explicit assessment of how semantic definitions and spatial support affected evaluation outcomes.

All products captured the broad national cropland pattern, yet only 26.8% of pixels in their cropland union were classified as cropland by all eight products. Disagreement concentrated along cropland edges, in complex terrain, and in agro-pastoral transition zones. Within the consensus domain, CropLayer and ESA-WC achieved the highest national F1 scores and greater robustness under medium and low inter-product consistency. CCrop-Land30 achieved the highest national F1 among the five 30 m products, whereas ESRI-LC had the smallest absolute national area bias. These contrasting strengths showed that classification performance and area-estimation performance require separate assessment.

The controlled comparisons further identified evaluation conditions that affected product performance. Reference-label recoding changed accuracy and area bias for three products, revealing product-specific sensitivity to semantic definitions. In mixed cropland settings, controlled aggregation reduced F1 for some products, but changes in absolute local cropland-proportion error were not statistically significant. Finer spatial support therefore benefited classification in these comparisons without clear evidence of improved local proportion estimation under the specified aggregation rule.

No single product performed best across all evaluated tasks. Product selection should jointly consider the target definition, spatial support, geographic setting, and whether the objective is spatial identification or area estimation. By making these conditions explicit, the framework supports interpreting product rankings and selecting cropland products for defined applications, with conclusions bounded by the evaluated domains and aggregation settings.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18213798/s1, **Table S1:** Reference-sample allocation across the 135 sampling strata; **Table S2:** National design-weighted accuracy estimates and 95% confidence intervals within the consensus-cropland evaluation domain; **Table S3:** Accuracy and area-bias estimates with 95% confidence intervals under unified-agriculture and product-definition-matched reference-label protocols.

Author Contributions: Conceptualization, Y.X., G.H. and R.Y.; methodology, Y.X. and G.H.; validation, Y.X., R.Y., Z.Z. and T.L.; formal analysis, Y.X.; data curation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, G.H., R.Y., G.W., Z.Z., Y.P., H.L. and T.L.; visualization, Y.X., Z.Z. and Y.P.; supervision, G.H.; project administration, G.H.; funding acquisition, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFF1304301), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19090300), the Second Tibetan Plateau Scientific Expedition and Research Program (Grant No. 2019QZKK030701), and the "Future Star" Young Talent Program of the Aerospace Information Research Institute, Chinese Academy of Sciences, and in part by the Science and Technology Program of Kashi Prefecture (Grant No. KS2024013).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors acknowledge the organizations and research teams that made the cropland, land-cover, and ancillary datasets used in this study publicly available. During the preparation of this manuscript, the authors used ChatGPT (OpenAI; GPT-5.6 and GPT-6.1)for language editing and polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Becker-Reshef, I.; Barker, B.; Whitcraft, A.; Oliva, P.; Mobley, K.; Justice, C.; Sahajpal, R. Crop Type Maps for Operational Global Agricultural Monitoring. *Sci. Data* **2023**, *10*, 172. [CrossRef] [PubMed]
2. Duan, C.; Li, J.; Wu, S.; Yu, L.; Feng, H.; Siddique, K.H.M. Monitoring Abandoned Cropland in the Hilly and Gully Regions of the Loess Plateau Using Landsat Time Series Images. *J. Integr. Agric.* **2025**, *24*, 4430–4450. [CrossRef]
3. Potapov, P.; Turubanova, S.; Hansen, M.C.; Tyukavina, A.; Zalles, V.; Khan, A.; Song, X.-P.; Pickens, A.; Shen, Q.; Cortez, J. Global Maps of Cropland Extent and Change Show Accelerated Cropland Expansion in the Twenty-First Century. *Nat. Food* **2022**, *3*, 19–28. [CrossRef] [PubMed]
4. Yu, Q.; Wu, W.; Wu, H.; Hu, Q.; Wang, X.Y.; Shi, Z.; Chen, S.; Du, Y.; Yang, J.Y.; Gao, B.B.; et al. Development of a Remote Sensing-Based System for Monitoring Well-Facilitated Farmland. *Natl. Remote Sens. Bull.* **2025**, *29*, 1053–1073. [CrossRef]
5. Xue, J.; Zhang, X.; Chen, S.; Hu, B.; Wang, N.; Shi, Z. Quantifying the Agreement and Accuracy Characteristics of Four Satellite-Based LULC Products for Cropland Classification in China. *J. Integr. Agric.* **2024**, *23*, 283–297. [CrossRef]
6. Zhang, C.; Kerner, H.; Wang, S.; Hao, P.; Li, Z.; Hunt, K.A.; Abernethy, J.; Zhao, H.; Gao, F.; Di, L.; et al. Remote Sensing for Crop Mapping: A Perspective on Current and Future Crop-Specific Land Cover Data Products. *Remote Sens. Environ.* **2025**, *330*, 114995. [CrossRef]
7. Deng, O.; Ran, J.; Huang, S.; Duan, J.; Reis, S.; Zhang, J.; Zhu, Y.-G.; Xu, J.; Gu, B. Managing Fragmented Croplands for Environmental and Economic Benefits in China. *Nat. Food* **2024**, *5*, 230–240. [CrossRef] [PubMed]
8. Liu, Y.; Yu, Q.; Zhou, Q.; Wang, C.; Bellingrath-Kimura, S.D.; Wu, W. Mapping the Complex Crop Rotation Systems in Southern China Considering Cropping Intensity, Crop Diversity, and Their Seasonal Dynamics. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9584–9598. [CrossRef]
9. Qiu, B.; Yu, L.; Yang, P.; Wu, W.; Chen, J.; Zhu, X.; Duan, M. Mapping Upland Crop-Rice Cropping Systems for Targeted Sustainable Intensification in South China. *Crop J.* **2024**, *12*, 614–629. [CrossRef]
10. Zhao, N.; Chen, K.; Wu, X.; Zhang, L.; Wang, W. Cropland Fragmentation Change across China over the Last Two Decades. *Agric. Syst.* **2024**, *218*, 104010. [CrossRef]
11. Tu, Y.; Wu, S.; Chen, B.; Weng, Q.; Bai, Y.; Yang, J.; Yu, L.; Xu, B. A 30 m Annual Cropland Dataset of China from 1986 to 2021. *Earth Syst. Sci. Data* **2024**, *16*, 2297–2316. [CrossRef]
12. Van Tricht, K.; Degerickx, J.; Gilliams, S.; Zanaga, D.; Battude, M.; Grosu, A.; Brombacher, J.; Lesiv, M.; Bayas, J.C.L.; Karanam, S.; et al. WorldCereal: A Dynamic Open-Source System for Global-Scale, Seasonal, and Reproducible Crop and Irrigation Mapping. *Earth Syst. Sci. Data* **2023**, *15*, 5491–5515. [CrossRef]
13. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global Land Cover Mapping at 30 m Resolution: A POK-Based Operational Approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]

14. Zhang, X.; Zhao, T.; Xu, H.; Liu, W.; Wang, J.; Chen, X.; Liu, L. GLC_FCS30D: The First Global 30 m Land-Cover Dynamics Monitoring Product with a Fine Classification System for the Period from 1985 to 2022 Generated Using Dense-Time-Series Landsat Imagery and the Continuous Change-Detection Method. *Earth Syst. Sci. Data* **2024**, *16*, 1353–1381. [CrossRef]
15. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global Land-Cover Product with Fine Classification System at 30 m Using Time-Series Landsat Imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [CrossRef]
16. Yang, J.; Huang, X. The 30 m Annual Land Cover Dataset and Its Dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]
17. Zhang, L.; Wang, W.; Ma, Q.; Hu, Y.; Ma, H.; Zhao, Y. CCropLand30: High-Resolution Hybrid Cropland Maps of China Created through the Synergy of State-of-the-Art Remote Sensing Products and the Latest National Land Survey. *Comput. Electron. Agric.* **2024**, *218*, 108672. [CrossRef]
18. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. *ESA WorldCover 10 m 2020 V100*; International Institute for Applied Systems Analysis (IIASA): Laxenburg, Austria, 2021.
19. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global Land Use/Land Cover with Sentinel 2 and Deep Learning. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021*; IEEE: New York, NY, USA, 2021; pp. 4704–4707.
20. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. 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. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef] [PubMed]
21. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
22. Li, Z.; He, W.; Cheng, M.; Hu, J.; Yang, G.; 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* **2023**, *15*, 4749–4780. [CrossRef]
23. Jiang, H.; Ku, M.; Zhou, X.; Zheng, Q.; Liu, Y.; Xu, J.; Li, D.; Wang, C.; Wei, J.; Zhang, J.; et al. CropLayer: A 2 m Resolution Cropland Map of China for 2020 from Mapbox and Google Satellite Imagery. *Earth Syst. Sci. Data* **2025**, *17*, 6703–6729. [CrossRef]
24. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
25. Stehman, S.V.; Foody, G.M. Key Issues in Rigorous Accuracy Assessment of Land Cover Products. *Remote Sens. Environ.* **2019**, *231*, 111199. [CrossRef]
26. Boogaard, H.; Pratihast, A.K.; Laso Bayas, J.C.; Karanam, S.; Fritz, S.; Van Tricht, K.; Degerickx, J.; Gilliams, S. Building a Community-Based Open Harmonised Reference Data Repository for Global Crop Mapping. *PLoS ONE* **2023**, *18*, e0287731. [CrossRef] [PubMed]
27. Zhang, C.; Dong, J.; Ge, Q. Quantifying the Accuracies of Six 30-m Cropland Datasets over China: A Comparison and Evaluation Analysis. *Comput. Electron. Agric.* **2022**, *197*, 106946. [CrossRef]
28. Cui, Y.; Liu, R.; Li, Z.; Zhang, C.; Song, X.-P.; Yang, J.; Yu, L.; Chen, M.; Dong, J. Decoding the Inconsistency of Six Cropland Maps in China. *Crop J.* **2024**, *12*, 281–294. [CrossRef]
29. Cui, Y.; Dong, J.; Zhang, C.; Yang, J.; Chen, N.; Guo, P.; Di, Y.; Chen, M.; Li, A.; Liu, R. Validation and Refinement of Cropland Map in Southwestern China by Harnessing Ten Contemporary Datasets. *Sci. Data* **2024**, *11*, 671. [CrossRef] [PubMed]
30. Peng, X.; Li, L.; Cao, X.; Li, F.; Ding, M.; Liu, L.; Fu, S.; Sun, Y.; Zhang, C.; Liu, W.; et al. Accuracy Comparison and Synergistic Strategies of Seven High-Resolution Cropland Maps (1–10 m) in China. *Remote Sens.* **2025**, *17*, 3121. [CrossRef]
31. Kerner, H.; Nakalembe, C.; Yang, A.; Zvonkov, I.; McWeeny, R.; Tseng, G.; Becker-Reshef, I. How Accurate Are Existing Land Cover Maps for Agriculture in Sub-Saharan Africa? *Sci. Data* **2024**, *11*, 486. [CrossRef] [PubMed]
32. Naboureh, A.; Li, A.; Bian, J.; Moharrami, M.; Ebrahimy, H.; Lei, G.; Nan, X.; Zhang, Z.; Feizizadeh, B.; Dabove, P.; et al. Accuracies, Discrepancies, and Challenges of the 10 m Global Land Cover Products in Mountains. *GIScience Remote Sens.* **2025**, *62*, 2556064. [CrossRef]
33. Wang, Y.; Xu, Y.; Xu, X.; Jiang, X.; Mo, Y.; Cui, H.; Zhu, S.; Wu, H. Evaluation of Six Global High-Resolution Global Land Cover Products over China. *Int. J. Digit. Earth* **2024**, *17*, 2301673. [CrossRef]
34. Xu, P.; Tsendbazar, N.-E.; Herold, M.; de Bruin, S.; Koopmans, M.; Birch, T.; Carter, S.; Fritz, S.; Lesiv, M.; Mazur, E.; et al. Comparative Validation of Recent 10 M-Resolution Global Land Cover Maps. *Remote Sens. Environ.* **2024**, *311*, 114316. [CrossRef]
35. Nelson, M.D.; Garner, J.D.; Tavernia, B.G.; Stehman, S.V.; Riemann, R.I.; Lister, A.J.; Perry, C.H. Assessing Map Accuracy from a Suite of Site-Specific, Non-Site Specific, and Spatial Distribution Approaches. *Remote Sens. Environ.* **2021**, *260*, 112442. [CrossRef]
36. Skakun, S. The Impact of Map Accuracy on Area Estimation with Remotely Sensed Data within the Stratified Random Sampling Design. *Remote Sens. Environ.* **2025**, *326*, 114805. [CrossRef]

37. Adebayo, A.D.; Nakalembe, C.; Skakun, S.; Frimpong, D.B.; Ginsburg, A.; Gadhiya, Y.H.; Adjei, M.; Kerner, H.; Becker-Reshef, I. Optimizing Satellite-Based Cropland Area Estimation through Integrated Map Accuracy Assessment and Stratified Sampling Design across Six African Countries. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *148*, 105223. [CrossRef]
38. Jin, G.; Wei, Y.; Yu, Q.; Wu, W. How Do Scientists Use Terminology Related to Cropland? Examining the Disparity across Disciplines and Regions. *J. Integr. Agric.* **2025**, *24*, 4776–4786. [CrossRef]
39. Tubiello, F.N.; Conchedda, G.; Casse, L.; Hao, P.; De Santis, G.; Chen, Z. A New Cropland Area Database by Country circa 2020. *Earth Syst. Sci. Data* **2023**, *15*, 4997–5015. [CrossRef]
40. Tubiello, F.N.; Conchedda, G.; Casse, L.; Pengyu, H.; Zhongxin, C.; De Santis, G.; Fritz, S.; Muchoney, D. Measuring the World's Cropland Area. *Nat. Food* **2023**, *4*, 30–32. [CrossRef] [PubMed]
41. Niu, B.; Feng, Q.; Qiu, B.; Su, S.; Zhang, X.; Cui, R.; Zhang, X.; Sun, F.; Yan, W.; Zhao, S.; et al. Global-PCG-10: A 10 m Global Map of Plastic-Covered Greenhouses Derived from Sentinel-2 in 2020. *Earth Syst. Sci. Data* **2025**, *17*, 5065–5088. [CrossRef]
42. Qiu, B.; Liu, B.; Tang, Z.; Dong, J.; Xu, W.; Liang, J.; Chen, N.; Chen, J.; Wang, L.; Zhang, C.; et al. National-Scale 10-m Maps of Cropland Use Intensity in China during 2018–2023. *Sci. Data* **2024**, *11*, 691. [CrossRef] [PubMed]
43. Foody, G.M. Ground Truth in Classification Accuracy Assessment: Myth and Reality. *Geomatics* **2024**, *4*, 81–90. [CrossRef]
44. Wickham, J.; Stehman, S.V.; Sorenson, D.G.; Gass, L.; Dewitz, J.A.; Kilaru, J.V. Thematic Accuracy Assessment of the National Land Cover Database (NLCD2021) for the Conterminous United States. *GIScience Remote Sens.* **2026**, *63*, 2659490. [CrossRef]
45. Xing, D.; Stehman, S.V. Using Interpenetrating Subsampling to Incorporate Interpreter Variability into Estimation of the Total Variance of Land Cover Area Estimates. *Remote Sens. Environ.* **2024**, *311*, 114289. [CrossRef]
46. Gann, D.; Richards, J. Scaling of Classification Systems—Effects of Class Precision on Detection Accuracy from Medium Resolution Multispectral Data. *Landsc. Ecol.* **2023**, *38*, 659–687. [CrossRef]
47. Li, R.; Gao, X.; Shi, F.; Zhang, H. Scale Effect of Land Cover Classification from Multi-Resolution Satellite Remote Sensing Data. *Sensors* **2023**, *23*, 6136. [CrossRef] [PubMed]
48. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [CrossRef]
49. Ministry of Natural Resources of the People's Republic of China. *Technical Regulation of the Third Nationwide Land Survey*; Ministry of Natural Resources of the People's Republic of China: Beijing, China, 2019.
50. Neyman, J. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *J. R. Stat. Soc.* **1934**, *97*, 558. [CrossRef]
51. Saah, D.; Johnson, G.; Ashmall, B.; Tondapu, G.; Tenneson, K.; Patterson, M.; Poortinga, A.; Markert, K.; Quyen, N.H.; San Aung, K.; et al. Collect Earth: An Online Tool for Systematic Reference Data Collection in Land Cover and Use Applications. *Environ. Model. Softw.* **2019**, *118*, 166–171. [CrossRef]
52. Hu, Q.; Cai, Z.; You, L.; Fritz, S.; Zhang, X.; Yin, H.; Wei, H.; Yang, J.; Li, Z.; Yu, Q.; et al. Hcropland30: A Hybrid 30-m Global Cropland Map Generated by Leveraging Global Land Cover Products and Landsat Data via Deep Learning. *Remote Sens. Environ.* **2026**, *344*, 115552. [CrossRef]
53. Stehman, S.V. Estimating Area and Map Accuracy for Stratified Random Sampling When the Strata Are Different from the Map Classes. *Int. J. Remote Sens.* **2014**, *35*, 4923–4939. [CrossRef]
54. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [CrossRef]
55. Ahmad, M.N.; Gu, L.; Yang, S.; Yuan, X.; Bathrellos, G.D. A Training-Independent Sentinel-1/2 Phenological Approach for Interpretable Paddy Rice Mapping in Heilongjiang, China. *Remote Sens.* **2026**, *18*, 3511. [CrossRef]

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