remote sensing

MDPI

Article

Comparative Validation and Misclassification Diagnosis of 30-Meter Land Cover Datasets in China

Xiaolin Xu [1], Dan Li [1], Hongxi Liu [2], Guang Zhao [3], Baoshan Cui [1], Yujun Yi [1], Wei Yang [1] and Jizeng Du [1,*]

[1] State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Beijing Normal University, Beijing 100875, China; 202221180079@mail.bnu.edu.cn (X.X.); 202321180094@mail.bnu.edu.cn (D.L.); cuibs@bnu.edu.cn (B.C.); yiyujun@bnu.edu.cn (Y.Y.); yangwei@bnu.edu.cn (W.Y.)
[2] Advanced Institute of Natural Sciences, Beijing Normal University at Zhuhai, Zhuhai 519087, China; liuhongxi@bnu.edu.cn
[3] Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; zhaoguang@igsnrr.ac.cn
* Correspondence: dujz@bnu.edu.cn; Tel.: +86-10-58803143; Fax: +86-10-58800059

Abstract: Land cover maps with high accuracy are essential for environmental protection and climate change research. The 30-meter-resolution maps, with their better resolution and longer historical records, are extensively utilized to assess changes in land cover and their effects on carbon storage, land–atmosphere energy balance, and water cycle processes. However, current data products use different classification methods, resulting in significant classification inconsistency and triggering serious disagreements among related studies. Here, we compared four mainstream land cover products in China, namely GLC_FCS30, CLCD, Globeland30, and CNLUCC. The result shows that only 50.34% of the classification results were consistent across the four datasets. The differences between pairs of datasets ranged from 21.10% to 37.53%. Importantly, most inconsistency occurs in transitional zones among land cover types sensitive to climate change and human activities. Based on the accuracy evaluation, CLCD is the most accurate land cover product, with an overall accuracy reaching 86.98 ± 0.76%, followed by CNLUCC (81.38 ± 0.87%) and GLC_FCS30 (77.83 ± 0.80%). Globeland30 had the lowest accuracy (75.24 ± 0.91%), primarily due to misclassification between croplands and forests. Misclassification diagnoses revealed that vegetation-related spectral confusion among land cover types contributed significantly to misclassifications, followed by slope, cloud cover, and landscape fragmentation, which affected satellite observation angles, data availability, and mixed pixels. Automated classification methods using the random forest algorithm can perform better than those that depend on traditional human–machine interactive interpretation or object-based approaches. However, their classification accuracy depends more on selecting training samples and feature variables.

Keywords: land cover maps; inconsistency evaluation; accuracy assessment; misclassification diagnosis

Citation: Xu, X.; Li, D.; Liu, H.; Zhao, G.; Cui, B.; Yi, Y.; Yang, W.; Du, J. Comparative Validation and Misclassification Diagnosis of 30-Meter Land Cover Datasets in China. *Remote Sens.* **2024**, *16*, 4330. https://doi.org/10.3390/rs16224330

Academic Editor: Dino Ienco

Received: 15 October 2024
Revised: 18 November 2024
Accepted: 18 November 2024
Published: 20 November 2024

1. Introduction

Due to the influence of climate change and human activities, land cover has changed significantly in the past decades, e.g., deforestation, urbanization, desertification, and agricultural inversion [1,2]. Land cover changes significantly impact environmental processes, altering the water cycle through interception and evapotranspiration and affecting biodiversity conservation by modifying ecosystem landscape patterns [3,4]. In addition, land cover change determines land–atmosphere energy balance, controlling surface albedo, moisture regimes, and carbon fluxes; thus, it has an unignorable feedback to climate change [5–7]. For instance, deforestation reduces evapotranspiration and carbon storage, intensifying local and global warming [8,9]. Land cover products are vital for tracking these changes and predicting long-term trends [10,11]. Additionally, these products serve as crucial model

inputs in capturing the effects of land cover changes on climate prediction, hydrology, and ecosystems, supporting strategic planning for climate adaptation and mitigation [12–14]. However, observing large-scale and detailed land cover changes has always presented a formidable challenge, significantly limiting related research.

The development of satellite remote sensing technology has facilitated earth surface observation, allowing for high-resolution land cover mapping [15]. With the advancement in spatial resolution, satellite transits offer a narrower field of view, leading to revisiting periods to specific locations longer and reduced temporal resolution in remote sensing observations [16,17]. Among the range of resolutions available, the 30-m (30 m) remote sensing image is the ideal balance between temporal and spatial resolution [18]. The Landsat satellite series has consistently provided accessible 30 m-resolution images since the 1980s, enabling historical comparisons and long-term trend analysis of land cover changes [19–21]. Although the advent of the Sentinel-2 satellite has increased the resolution of remote sensing imagery to 10 m, the data released from its imagery for land cover products such as FROM_GLC10 and Esri series have been published since 2017 [22,23]. At the same time, the 2.5 m-resolution SPOT1-5 and 3 m-resolution Planet satellite have not been widely used in land cover mapping. Consequently, 30 m land cover maps have emerged as the predominant fine-resolution products worldwide.

Then, accurate classification based on remote sensing images is essential in producing reliable land cover maps. Initially, land cover classification relied on labor-intensive and man–machine interactive interpretation [24]. Technological advances shifted the focus to automated classification systems using statistical techniques such as maximum likelihood classifiers and decision trees, improving land cover classification efficiency and accuracy [25]. Recently, machine learning, mainly supervised learning methods like Support Vector Machines (SVMs) and random forest algorithms (RFs), has become a frequently used approach due to its robustness against overfitting and ability to handle high-dimensional data [26–28]. Deep learning, especially convolutional neural networks (CNNs), has significantly improved classification efficiency and accuracy by automating feature extraction and learning complex data patterns, reducing the need for manual feature design [29–31]. Using the classification methods mentioned above, several institutions have developed and released global or regional land cover products at a 30 m resolution, including GLC-FCS30, Globeland30, CNLUCC, etc. [32–34]. Notably, the Google Earth Engine (GEE) cloud-based platform offers a vast, multi-petabyte catalog of analysis-ready data and a high-performance, inherently parallel computation service, which has led to the development of many large-scale and high-resolution land cover products, such as China Land Cover Database (CLCD), FROM_GLC30 [35–37]. For most land cover products, the 2020 land cover data are the latest versions, using the most advanced classification technology and up-to-date remote sensing images as input, resulting in their higher classification accuracy than products from other periods [21].

However, the classification of land cover products differs significantly due to the differences in data sources, algorithms, and workflow designs. Previous studies have demonstrated that most land cover products achieve an overall accuracy of over 70%, with first-level classification accuracy reaching up to 90% [33,38,39]. However, less than 50% of areas show a consistent classification across existing land cover products, far larger than the area of land cover changes caused by climate change and human activities [40]. Consequently, these discrepancies limited their practical value and caused more significant uncertainty in related studies [41,42]. Before releasing land cover products, creators assess their consistency and reliability through comparative analyses with existing products [32,43]. Many studies have evaluated the classification accuracy of existing land cover products [44,45]. However, significant discrepancies in evaluation results among those studies pose a challenge for researchers when selecting suitable land cover products for their research [46,47].

One of the alternative reasons for disagreement among accuracy assessments is differences in the validation dataset used, which led to significant disputes in the accuracy

assessment of land cover products [48]. Currently, there are three primary sources of validation data: (1) Field surveys are the primary approach to building validation data early and the most reliable one. However, it is labor-intensive and thus has a limited sampling size, making it difficult to evaluate the accuracy of land cover products at a large scale. (2) Public validation datasets, such as Geo-Wiki or GLCVSS, are a widely used third-party data source for the validation of land cover products [46,49]. Nevertheless, they suffer from outdated and non-random samples, e.g., most samples were created before 2012 and less located in sparsely populated areas (see Figure S1). (3) Construct validation datasets based on existing land cover products are also used, rapidly generating numerous validation points through random sampling. Their reliability is constrained by the chosen land cover products, none of which obtained widely recognized classification accuracy. Apart from the limitation in the validation dataset, most studies focus primarily on consistency comparison and accuracy assessment without further exploring the underlying causes of inconsistencies and biases.

China is one of the regions with the most significant inconsistencies in land cover classification [40]. An insufficient validation dataset makes accuracy assessment and misclassification diagnosis more difficult in China [50]. China's complex climate and terrain contribute to various land cover types and extensive landscape fragmentation [51,52]. Meanwhile, rapid economic growth and population expansion have made China one of the regions with the world's largest and fastest land cover change in the last century [53,54]. Due to extensive development, China experienced much deforestation, overgrazing, and thus severe land desertification. However, since 2000, China has been actively focusing on ecological protection and implementing projects such as returning cropland to forests and afforestation in deserts [55]. As a result, China has been recognized as one of the fastest-growing regions in the world based on satellite observations by NASA [56]. These changes have significantly impacted hydrological processes and ecosystems, emphasizing the critical need for selecting reliable land cover products to explore their response and feedback to climate change and ecosystem balance [57–59].

To meet the above needs, we evaluated the consistency and classification accuracy of four popular land cover products in China in 2020 based on the validation database. Furthermore, we diagnosed key factors for inconsistencies among land cover products. In order to quantify the accuracy of each land cover product, we randomly selected about ten thousand sampling points in China. Based on Google Earth's high-resolution images, photos, and street view data, we accurately identified the land cover types of these points and thus constructed a more reliable and representative validation dataset. Through accuracy assessment and analysis, we identified vegetation coverage, terrain complexity, and other related variables as fundamental factors affecting products quality. Our study offers a guide for improving China's classification accuracy of land cover products and helps select suitable products for research requirements.

2. Dataset and Methods

2.1. Dataset

2.1.1. Land Cover Products

Here, we have chosen four main land cover products in China, namely GLC_FSC30, CLCD, Globeland30, and CNLUCC. All of these products have a spatial resolution of 30 m. They are mainly derived from Landsat images and utilize hyperspectral and higher-resolution images to improve land cover classification. These products have a long time series, with most having sequences from the 1980s to the present (except for Globeland30, which starts in 2000). These products are the most commonly used to explore land cover changes in China. They also serve as input data for regional climate simulations and common land models. Therefore, they play a crucial role in studies about land cover changes in China. More detailed information on the above land cover products is provided in Table 1.

Table 1. Basic information of four datasets.

Datasets	Research Institution	Remote Sensing Imagery	Classification Methods	Data Source
GLC_FCS30	Aerospace Information Research Institute, Chinese Academy of Sciences	Landsat TM/ETM+/OLI	Random forest	http://data.casearth.cn, accessed on 30 June 2023
CLCD	Wuhan University	Landsat	Random forest	https://zenodo.org/records/8176941, accessed on 30 June 2023
Globeland30	National Geomatics Center of China	Landsat TM/ETM+, HJ-1, GF-1	Pixel–Object–Knowledge	https://cloudcenter.tianditu.gov.cn/, accessed on 30 June 2023
CNLUCC	Institute of Geographic Sciences and Natural Resources Research	Landsat TM/ETM/8	Human–machine interactive interpretation	https://www.resdc.cn/, accessed on 30 June 2023

However, these products have notable differences, primarily observed in three points: (1) Spatial extent: Globeland30 and GLC_FCS30 offer global coverage, while CLCD and CNLUCC are limited to China. (2) Classification methods: GLC_FCS30 and CLCD utilize the random forest algorithm, while Globeland30 and CNLUCC use the integration of pixel- and object-based methods with knowledge (POK-based) method and human–computer interactive interpretation, respectively. Even with the same methods, differences in training samples and feature variables lead to significant classification differences. (3) Classification systems: CLCD and Globeland30 are relatively simple, classifying into 9 and 10 categories, whereas GLC_FCS30 classifies into 29 categories and CNLUCC into 6 primary categories and 25 secondary categories. Therefore, we must reconstruct a unified classification system based on the same standards to facilitate comparison (see Table S1).

According to the classification system adopted in each dataset, combined with the classification standard of land use status in China () and the classification standard proposed by the Food and Agriculture Organization of the United Nations (FAO), seven land cover types were determined, namely cropland, forest, grassland, wetland, impervious surface, bareland, and glacier, and their correspondence with the four datasets is shown in Table S1. Compared to the commonly used nine categories, we have merged shrubs into forests and water bodies into wetlands. The proportion of shrubland is small, and it is hard to distinguish between shrubland and forest using high-resolution imagery, which could reduce the reliability of our validation samples. Additionally, seasonal transitions between water bodies and other wetlands may lead to misclassification. By merging these categories, we can focus on the impact of approaches on the classification accuracy of land cover products.

2.1.2. Ancillary Datasets

Through preliminary analysis, we have identified the main factors affecting the classification accuracy of land cover products, which include vegetation conditions, topographic features, landscape fragmentation, and the accessibility of remote sensing imagery. These factors can be quantified using ancillary datasets on vegetation index, digital elevation, cloud cover, and their derived indices.

(1) Vegetation Index

The vegetation index is a critical feature variable in land cover classification, deployed to discern vegetation's spatial and temporal dynamics. However, certain land cover types, including croplands and grasslands, exhibit a higher similarity in vegetation index, increasing the risk of misclassifications. Here, we utilized Landsat 8 imagery data to compute the Normalized Difference Vegetation Index (NDVI) from 2019 to 2021. Subsequently, we extracted the maximum NDVI value for each year and calculated the three-year average as the vegetation index for validation points. This derived vegetation index was employed to assess the correlation between classification accuracy and vegetation indices [60].

(2) Digital Elevation

Variations in topography, such as slope and aspect, can affect the satellite image quality due to differences in illumination and viewing angles, which can introduce shadows or distortions. These factors may lead to inconsistencies in the spectral signatures for classifying different land cover types, potentially resulting in decreased classification accuracy, especially in rugged or hilly regions. We used SRTM 30m-resolution digital elevation data to represent the elevation and undulation of China's topography and calculated main topographic indexes based on the Google Earth Engine cloud computing platform, including topographic relief, slope, aspect, and hill shade. The calculation methods for slope and aspect have been explained in previous studies [61,62].

(3) Cloud Cover

Cloud cover can significantly impact land cover classification accuracy by masking the Earth's surface, resulting in incomplete or distorted images [63]. This degradation in data quality compromises the reliability of spectral signatures crucial for distinguishing various land cover types. Consequently, regions characterized by frequent or dense cloud cover are prone to higher misclassification rates in land cover products. We extracted the cloud mask from the Landsat 8 images from 2018 to 2022, including cloud cover and shadow [64]. Then, we calculated the annual mean cloud mask for each validation point.

(4) Land cover fragmentation and changes

Urbanization and agricultural expansion have resulted in severe fragmentation of land cover in regions subject to human activities [65]. Land cover fragmentation can significantly influence land cover classification accuracy from remote sensing images. Firstly, fragmentation increases the number and complexity of boundaries between land cover types, increasing mixed pixels. Secondly, with higher fragmentation, the likelihood of misclassification increases, as manual interpreters or automated classification algorithms struggle to correctly assign land cover types to tiny patches.

We first set up a 10 * 10 window around each validation point, and the landscape fragmentation is calculated as follows:

$$LCF = PN/TN \quad (1)$$

where *LCF* is the land cover fragmentation of the particular window; *PN* and *TN* are the number of patches and pixels within this window, respectively. Eight-cell connectivity was used to build patches. We calculated land cover fragmentation based on the four land cover products mentioned above and used their mean as the land cover fragmentation of validation points. The land cover change ratio was estimated using a similar approach, based on the average proportion of land area that has changed land cover types within the sliding window of four land cover products between 2000 and 2020. This process ensured that the calculation of these two factors did not rely on a specific product alone but rather on the average of the four products to obtain the most reasonable result.

2.2. Methodology

2.2.1. Sampling Designs

In this study, we have implemented two sampling designs to assess overall accuracy and class-specific accuracy, respectively (see Figure S2). The first is balanced random sampling, which can generate unbiased estimators of the population of interest. Another is stratified random sampling, which increases the sample size from rare land cover types and assesses classification accuracy on a class-specific scale [66].

(1) Balanced random sampling

The CEOS-LPV report on global land cover map validation established recommended standards for accuracy assessment [67]. Simple random sampling is a fundamental method for accurately assessing land cover products. It randomly chooses sample points from the

study area without systematic biases. Importantly, simple random sampling is cost-efficient, and the validation result is insensitive to changes in sample sizes [68]. However, compared to systemic sampling, simple random sampling may cause an uneven spatial distribution of samples, reducing their representativeness. Hence, we used a balanced random sampling approach by setting a threshold to keep a minimum distance between any two sample points [69]. Thus, the sample points can be distributed evenly within the study area and have a better spatial balance [70].

As shown in Figure S3a, with the distance increases, the density of the sample in grids has better spatial balance. Therefore, we set 5 km as the minimum distance between samples to compromise the spatial balance and probability of the sampling design. We used ArcGIS to select 10,000 random sample points. Figure S3b shows the number of sample points in the $1^\circ \times 1^\circ$ grids. We find that the number of samples in each grid is consistent except for those in edges being smaller due to incomplete area, indicating that the density of samples has an even spatial consistency with a normal distribution (see Figure S3c). We examined the correlation between site density and spatial variables based on grid points. Our analysis found no significant correlation, indicating no spatial autocorrelation in our sample point selection (see Figure S3d). In total, 2049 croplands, 2566 forests, 2718 grasslands, 225 wetlands, 339 impervious surfaces, 2016 barelands, and 123 glaciers were selected to validate the accuracy of four land cover products. The proportion of land cover types matches well with the area proportion of land cover products, indicating reliable representativeness in spatial patterns and land cover types for the validation dataset (see Figure 1b,c).

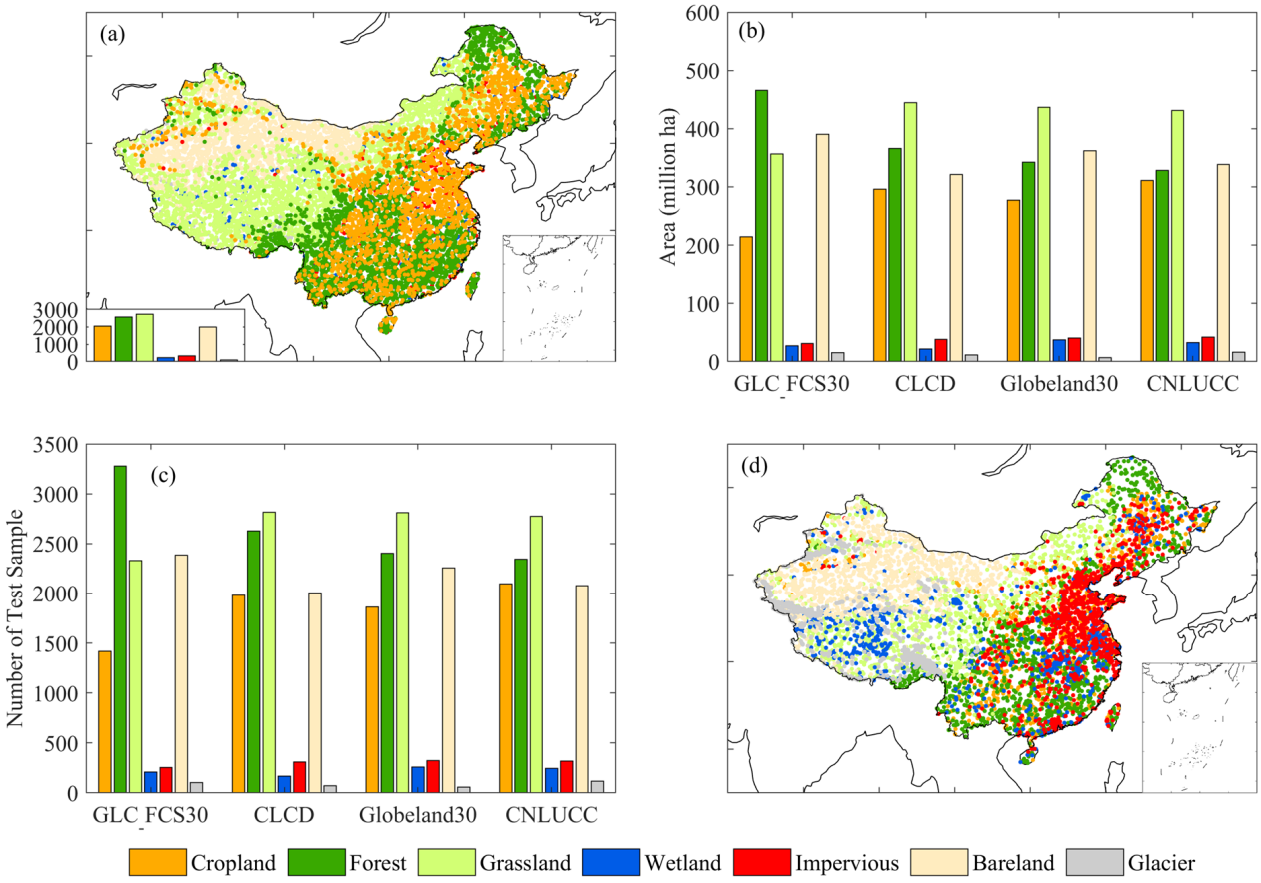

Figure 1. The composition of land cover types and spatial distribution of validation samples. (**a**) Validation samples used in assessing the classification accuracy of land cover products; (**b**) the total area of land cover types for products; (**c**) the number of validation samples for land cover types; (**d**) validation samples are generated by stratified random sampling based on land cover types.

(2) Stratified random sampling

Stratified sampling designs are used to increase the sample size of rare land cover types [42]. For instance, glaciers, impervious surfaces, and wetlands have a smaller area proportion but play a vital role in regional climates and ecosystems and are more sensitive to climate change. By increasing their sample size, stratified random sampling can reduce the standard errors of accuracy assessments for rare land cover types. Due to the irregular shape of many land cover patches, designing sample points using systematic sampling in specific classes is challenging. Hence, we selected sample units within each stratum using a simple random protocol, which is the preferred method because it allows for easy adjustments to the sample size within strata.

We cannot use existing land cover products to determine the spatial extent of each land cover type to ensure the selection of sampling sites is independent of them. As a result, acquiring samples for rare and dispersed land cover types is challenging. We initially pinpointed areas where specific land cover types could be present based on factors such as latitude, altitude, and vegetation index to boost sampling efficiency. We then gathered random samples from these areas and employed visual interpretation to recognize sample points of specific classes. We determine sample size by a common-use formula based on the user's accuracy for stratified random sampling as follows [66]:

$$n = z^2 p(1-p)/d^2 \quad (2)$$

where z = 1.96 for a 95% confidence interval and d = 0.05 for the desired half-width of the 95% confidence interval. We set the p to 0.54, the minimum user's accuracy of every land cover type calculated from balance random sampling. Finally, the sample size formula yields n = 382. In the subsequent uncertainty analysis, we need to randomly extract 80% of samples for accuracy assessment. Therefore, we set the number of samples for each land cover type to 500 (see Figure 1d), which can meet the sample size requirements of the accuracy assessment.

(3) Response design

The response design is crucial to our methodology, ensuring the agreement between land cover products and reference validation datasets. It controls accuracy assessment by defining the spatial unit, choosing reference data sources, and establishing labeling protocols [48]. For reproducibility, we provide a detailed description of the response design, which includes the spatial assessment unit, classes, sources of reference data, specific information collected, rules for assigning reference class labels, and how to define the agreement between the map and reference classifications [66]. This thorough approach, including using a 30 m × 30 m grid as the spatial unit and alignment of reference data labeling with the classification system, ensures the reliability of our research.

In our study, three experts and six interpreters were divided into three groups, each consisting of one expert and two interpreters. The experts have extensive experience in satellite-based land cover analysis and remote sensing image interpretation. They are responsible for training interpreters and ensuring quality control of classification results. Each group independently classified the land cover type of sample pixels. When the classification results of the three groups differed, the three experts discussed and analyzed the pixels with more auxiliary data, significantly reducing subjective errors and effectively improving identification accuracy.

Classifying sample pixels is as follows: All sample pixels are categorized into pure or mixed pixels. Pure pixels contain a single land cover type, which makes them relatively easy to identify. We created a KML file based on the boundary of pixels and imported them into Google Earth Pro and Gaode Maps (Version 14.02), a popular navigation software in China that has a lot of street view data (see Figure 2). High-resolution images, field photos, and street views from 2020 can help interpret the land cover type of sample pixels. To increase the reliability of the identification, we also used a time series of NDVI derived from MODIS, Landsat, and Proba-V data for plant phenology, which helps distinguish

cropland from forest. Moreover, when NDVI profiles show no significant change in pixels, they indicate stability in land cover type. Therefore, historical high-resolution images or field photos of nearby years can be used in interpretation.

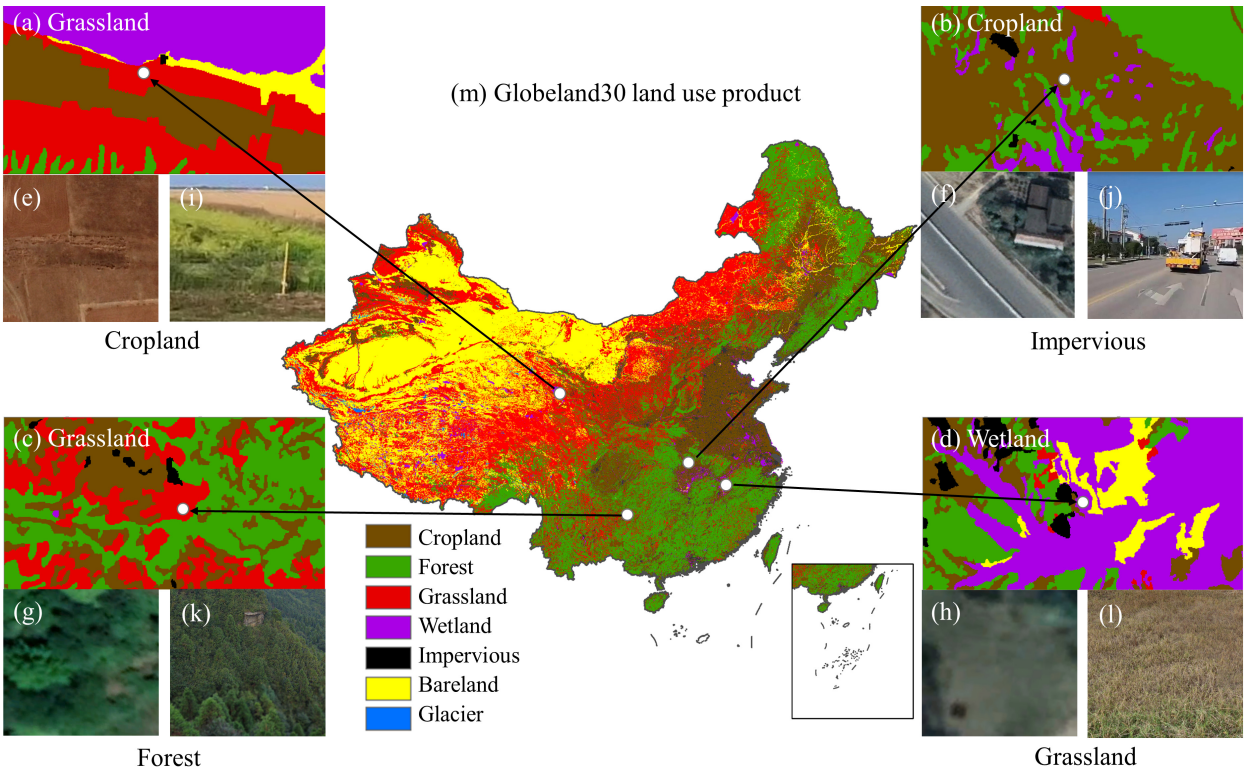

Figure 2. The difference in four typical points between the identification in the Globeland30 and the high-resolution images and the high-definition real picture from Google Earth Pro. (**a**–**d**) The land cover map of Globeland30; (**e**–**h**) the high-resolution images from Google Earth Pro; (**i**–**l**) the high-definition real pictures from Google Earth Pro and street view from Gaode Maps (Version 14.02).

For mixed pixels, based on the spectral features of land cover types provided by pure pixels above, we trained an automated image segmentation model to divide the mixed pixels into polygons of several land cover types (see Figure 3). Then, subpixels with land cover type with the most significant proportion of area were labeled accordingly and counted for all land cover types. The land cover type with the highest number of subpixels was used as the label for the pixel. If there were two equally abundant land cover types, we prioritized wetland, city, glacier, forest, farmland, grassland, and bareland in that order. Following these rules, we assigned a classification to each sample pixel.

(4) Uncertain estimation of sampling

Although we have employed various methods to enhance the randomness and spatial balance of sampling, estimators still have biases due to the uneven distribution of land cover types. In statistics, mathematics, and various fields, Monte Carlo simulation deals with uncertainty and variability by performing repeated random sampling to obtain numerical results. It can be a valuable tool in assessing the accuracy of land cover products with uncertainty in the sampling process. Here, we implemented a robust random sampling method with replacement, which involved selecting 80% of the total sample for accuracy assessment and repeating the process 1000 times. The 95% confidence level of accuracy assessment is reliably estimated based on the distribution characteristics of the results from all repetitions (Figure S2).

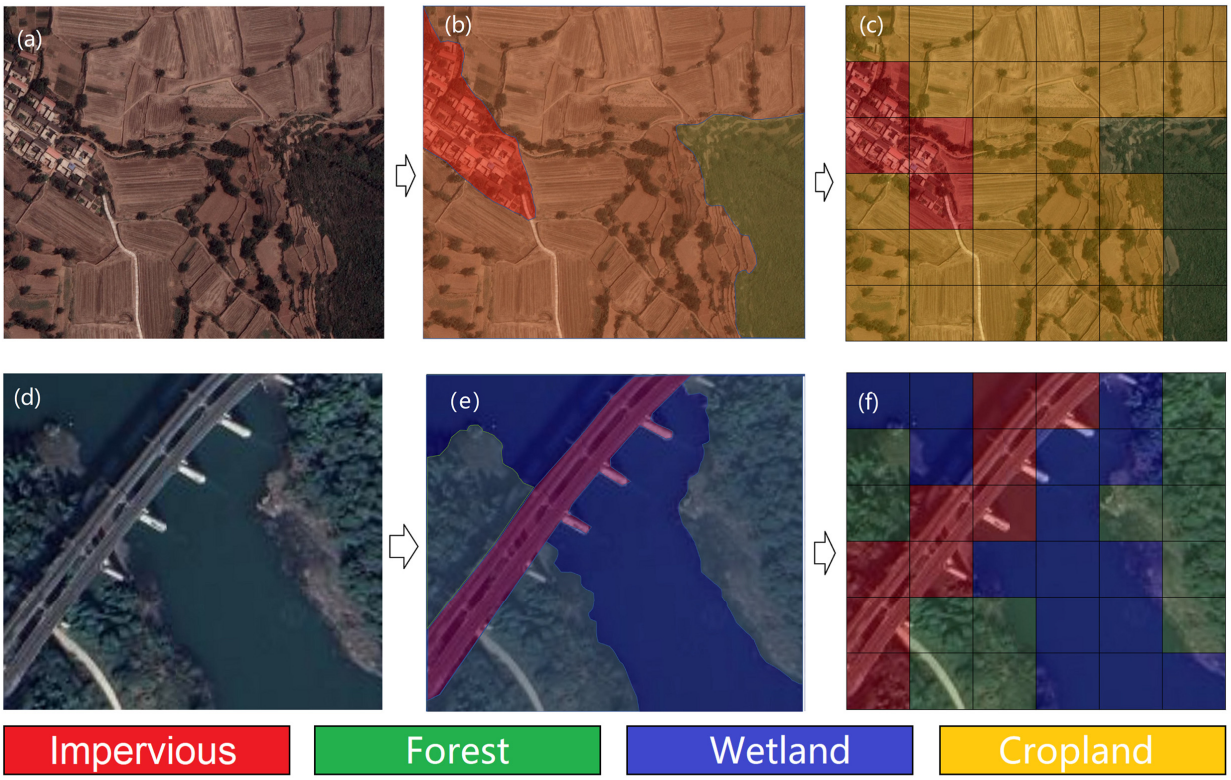

Figure 3. Examples of manual visual interpretation are shown here. (**a**,**d**) High-resolution images from Google Earth Pro; (**b**,**e**) land use types regardless of pixel; (**c**,**f**) the box represents the footprint of an interpretation unit and includes 36 grid cells. We divided the box into polygons based on land cover classification, identified grid cell type by area of polygons, and assigned the land cover type with the most grid cells as the box's land cover type. When the two types of land cover have the largest and the same number of grid cells, we set the land cover type of the box according to the priority order of glacier, wetland, impervious surface, cropland, forest, grassland, and bareland.

2.2.2. Inconsistency Analysis

We compared each pixel across the products to check for classification consistency among different products. Pixels classified in the same land cover types in all products were labeled "consistent", while those with differing classification results were categorized as "inconsistent". Then, we analyzed spatial patterns to see if inconsistencies are clustered in certain areas or randomly distributed. Using confusion matrices, we identified the leading causes and critical regions of inconsistent classification among land cover products.

We use balanced random sampling and stratified [71] random sampling approaches to create confusion matrices between land cover products. Considerable research has been conducted on land cover changes, predominantly centered on alterations in area, such as the estimation of land carbon storage. Therefore, the balanced random sampling method is the preferred approach for evaluating the influence of inconsistencies on the divergence of research outcomes derived from various land cover products. However, this method overlooks the contribution of land cover types with the smaller areas but essential ecological or hydrological functions, such as wetlands. The utilization of the confusion matrix derived from stratified random sampling allows for a more precise representation of the influence of various classification techniques on the disparity between land cover products. This method meticulously considers the proportion of pixels with inconsistent classification in the total number of pixels for every land cover type, which are expressed as follows:

$$P_{i,j} = X_i Y_j / X_i * 100\% \quad (3)$$

where $P_{i,j}$ is the pixel ratio of inconsistency between land cover products; X_iY_j is the number of pixels classified as class i in product X and class j in product Y; and X_i is the total number of pixels classified as class I in dataset X.

2.2.3. Accuracy Assessment

The accuracy of land cover products was evaluated using confusion matrices of land cover products with validation dataset, which included the producer's accuracy (*PA*), user's accuracy (*UA*), overall accuracy (*OA*), and the *F1* score [48,72]. The F1 score represents the balance between *PA* and *UA* and is calculated as follows:

$$F1 = 2(PA \times UA)/(PA + UA) \times 100\% \quad (4)$$

These metrics comprehensively assess land cover products' classification accuracy and reliability.

2.2.4. Inconsistency Diagnosis

Numerous studies have concentrated on assessing the consistency and accuracy but have not delved into the reasons for the differences among land cover products. This oversight hinders the improvement of land cover classification methods. To address this gap, our study investigated the causes of differences in classification accuracy by examining the spatiotemporal patterns of land cover products. We specifically looked into how various factors, such as vegetation indices, topographic variables, and landscape fragmentation, affected the variations in classification accuracy.

(1) Gradient analysis

To investigate the impact of related factors on classification accuracy, we encountered two main notable challenges. Firstly, validation points must be more evenly distributed across variable intervals. For example, when studying slope's impact on accuracy, most samples are in lower slope ranges, reducing the representation of higher slopes and thus underestimating the effect of slope on classification accuracy. To tackle these issues, we used gradient analysis, which divided environmental factors into intervals and calculated classification accuracy within each. It then examines accuracy changes with increasing gradients of environmental factors, counteracting uneven sample distribution. Additionally, it smooths data by adjusting interval sizes, reducing background noise and abnormal values interference, and aiding in identifying accurate responses to variable changes.

Gradient analysis can be divided into two steps: firstly, set a sliding window of target factor with a width that allows for no less than 100 samples in each window and calculate the classification accuracy of samples in each window. Consequently, each value of the target factor is assigned with a classification accuracy. Then, we set gradients to compromise the sample size and the noise distribution from other factors and calculated the mean and percentage of classification accuracy in each gradient interval. The effect of the target factor on classification accuracy was analyzed based on differences in classification accuracy with respect to interval value of a specific target factor. For instance, the slope range is approximately 0–40 degrees. We divided this interval into gradients every 4 degrees and computed each gradient's mean and percentiles of classification accuracy. Based on the variation in the classification accuracy's mean and percentile values with the gradients, we assessed the influence of slope on classification accuracy and the stability of this influence.

(2) Attribution analysis

Using the gradient analysis mentioned earlier, we can determine the classification accuracy within different impact factor intervals. Further analysis through correlation and regression helps us thoroughly evaluate the relationship and sensitivity of classification accuracy to these factors. When the classification accuracy exhibits a linear change with the gradient of the impact factor, we used the linear regression method to assess the

correlation and sensitivity of classification accuracy to impact factors. The regression of spatial variation of classification accuracy with impact factors can be expressed as follows:

$$CA = Sen * IF + \varepsilon \quad (5)$$

where CA is the classification accuracy, and IF is the related impact factor. Sen is the sensitivity of classification accuracy to IF, and ε is the contant.

We conducted a Pearson correlation analysis to assess the relationship between land cover classification accuracy and the impact factors. We identified the influential factors significantly associated with classification accuracy based on correlation analysis. Further, we investigated the synergistic effects among these factors and the difference in sensitivity of classification accuracy to these factors across land cover types. The variation of classification may be related to several impact factors, and there are likely significant correlations among impact factors. Hence, partial correlation analysis was taken into account only to analyze the correlation between classification accuracy and one influence factor, excluding the impact of the other [51]. The partial correlation coefficient was calculated as follows:

$$PR_{xy,z}(t) = (R_{xy}(t) - R_{xz}(t)R_{zy}(t)) / \sqrt{(1 - R_{xz}^2(t))\left(1 - R_{yz}^2(t)\right)} \quad (6)$$

where x is land cover classification accuracy, y is one influence factor, and z represents all influence factors apart from y. $PR_{xy,z}(t)$ is the partial correlation coefficient between variables x and y after controlling the effect from z.

3. Results

3.1. Consistency Analysis

Figure 4 shows the spatial pattern of inconsistent pixels among four 30m-resolution land cover products in China for 2020. The consistency index across the four products is 50.34%. Notably, regions with high consistency often feature a singular landscape, such as barren lands in northwest deserts, grasslands in Inner Mongolia, agricultural zones in the North China Plain, and forests in the Southwest and Northeast (see Figure 4g). Figure 4a–f illustrate spatial inconsistency between any two datasets, ranging from a minimum consistency index of 21.10% between CNLUCC and CLCD datasets (Figure 4d) to a maximum of 37.53% between GLC_FCS30 and Globeland30 (Figure 4b). Inconsistent classification primarily occurs in (1) land cover transitional zones, like agriculture–pasture transition belts in Northwest China and forest–grassland transition zones in Northeast China; (2) regions with significant topographical variations, primarily in mountainous areas of Southwest China; (3) landscapes with severe fragmentation, including terraced fields in mountainous areas interspersed with forests and grasslands and suburban areas around cities with intermingled cropland and buildings; and (4) areas experiencing rapid land cover changes, both interannual and intra-annual, such as some bareland prone to be misclassified as glacier due to seasonal snow cover interference. Fluctuating water levels can cause misclassifications of wetlands, grasslands, and barelands.

As shown in Figure S4, the land cover types with the highest consistent areas are forest land, bareland, grassland, and cropland, China's main land cover types, accounting for 93.7% of the total area. Meanwhile, the misclassification among these land cover types is also the main reason for the divergence of land cover products. The top eight types of misclassifications in terms of area are related to them, such as grassland and bareland, forest land and cropland, grassland and forest, etc. Additionally, compared to other land cover products, Globeland30 is prone to classifying bareland as grassland or forest as cropland (see Figure S4a–c); GLC-FCS30 is prone to classifying grasslands as forest or bareland (see Figure S4b,e,f). These inconsistencies often occur between land cover types with similar vegetation coverage, such as cropland and forest.

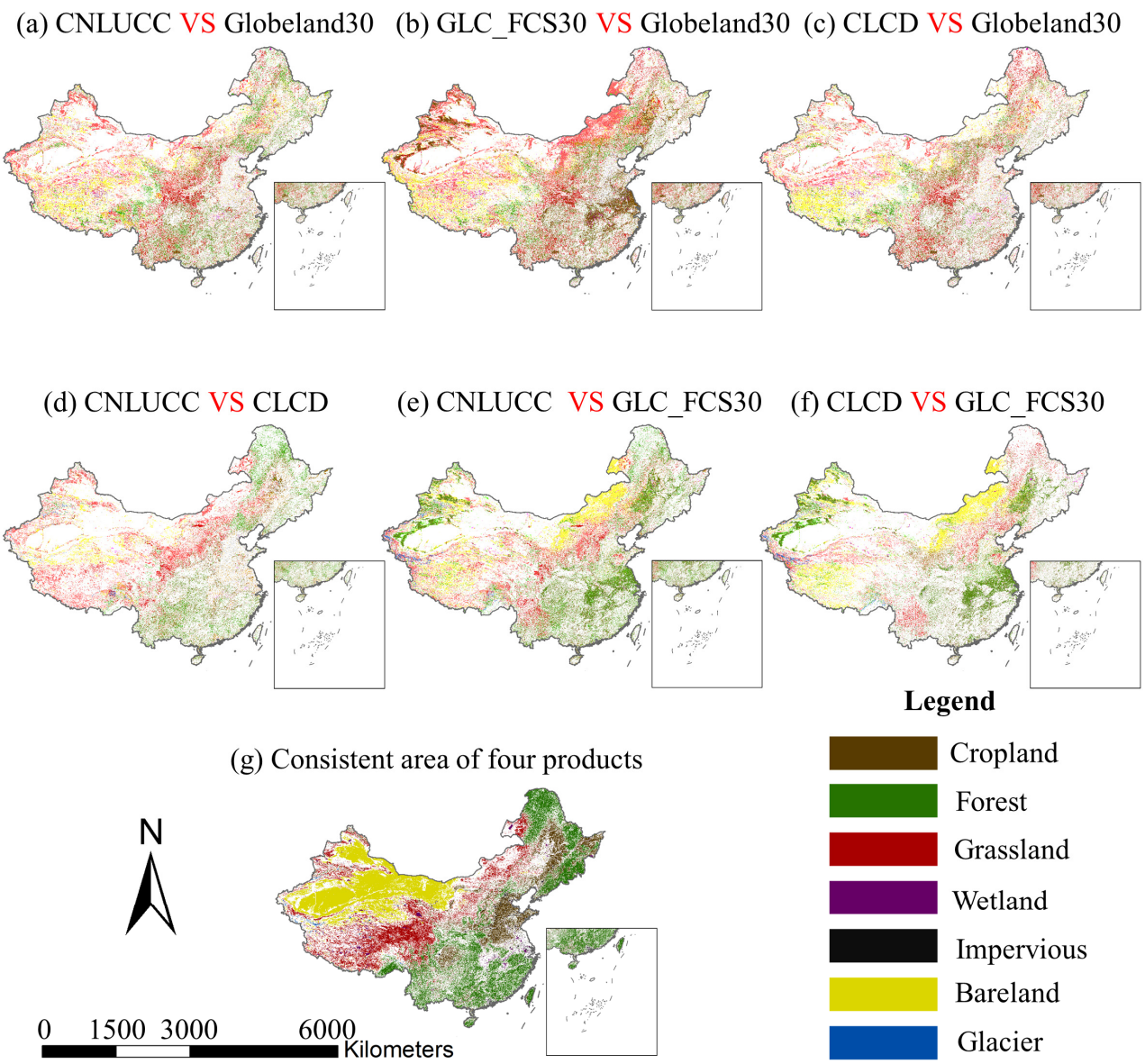

Figure 4. The spatial pattern among four land cover products at 30m spatial resolution in 2020. (**a**–**f**) The inconsistent pixels between each two land cover products; (**g**) the consistent pixels of four land cover products.

The pixel ratio-based consistency analysis eliminates the impact of variations in land cover proportions, demonstrating classification consistency across different land cover types. Figure 5g indicates that forest, wetland, and bareland exhibit high classification consistency among land cover products. These land cover types possess distinct spectral characteristics and minimal spectral confusion with other land cover types. The inconsistency types with higher pixel ratios include cropland and forest, grassland and glaciers, and cropland and impervious surfaces (see Figure 5a–f). Apart from similar vegetation indices, there are two other apparent features for inconsistent regions: firstly, areas with intensive human activity, like cropland and impervious areas, exhibit severe landscape fragmentation and rapid land cover changes; secondly, areas with complex terrains, such as the mountainous regions of the Tibetan Plateau, pose challenges in identifying glaciers or bareland.

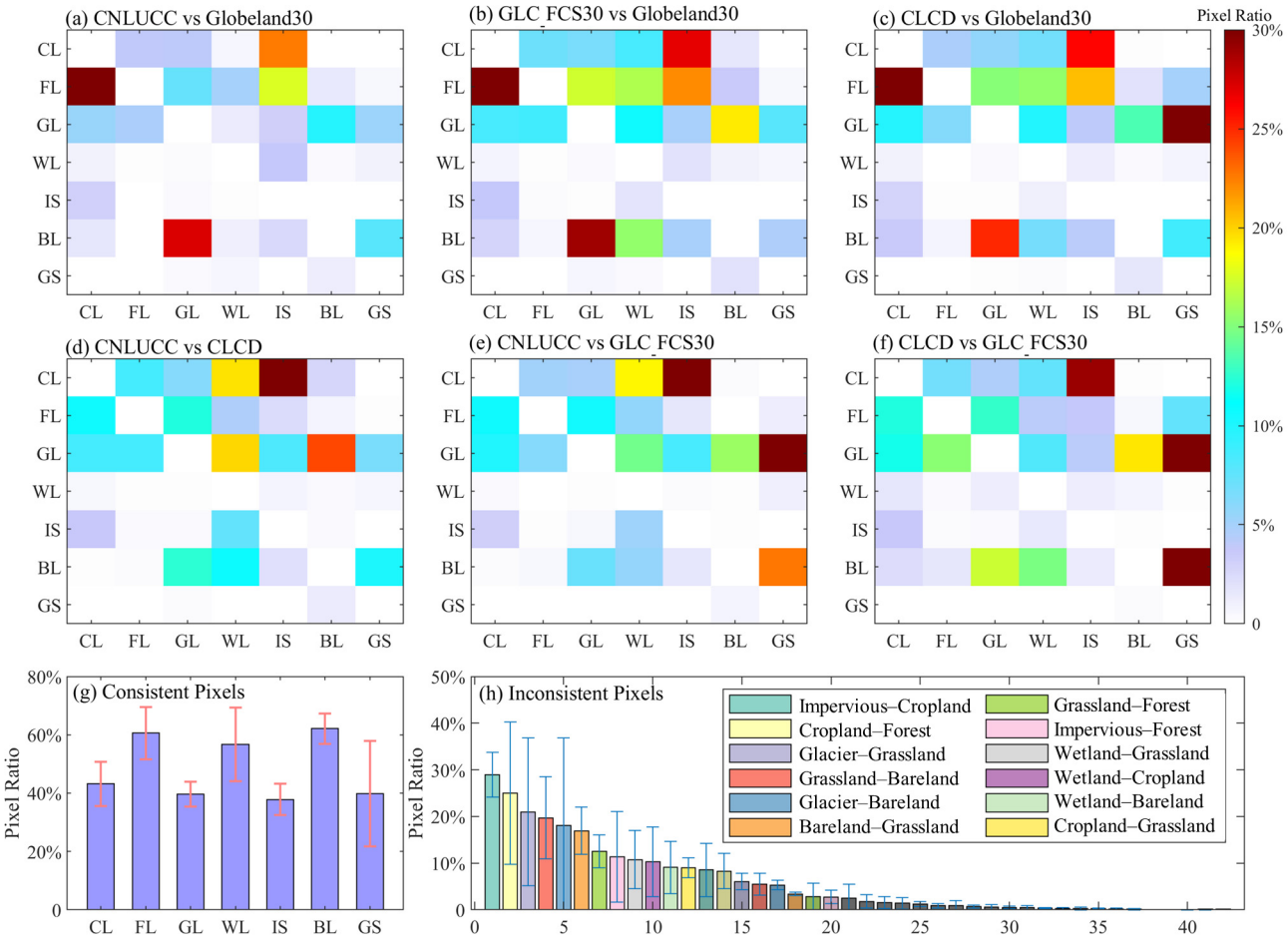

Figure 5. Pixel ratio-based confusion matrix of land cover types between land cover products (**a**–**f**). (**g**) Proportion of the consistent land cover for four land cover products; (**h**) proportion of land cover inconsistencies between four land cover products. Proportions in (**g**,**h**) are calculated based on average area of land cover type or land cover combination area in four products. CL: cropland, FL: forest, GL: grassland, WL: wetland, IS: impervious, BL: bareland, GS: glaciers.

Our comparative analysis revealed significant classification inconsistency in the existing 30m-resolution land cover products, consistent with previous studies [45,73]. Notably, the proportion of areas with inconsistent classifications far exceeds the estimated land cover changes caused by climate change or human activities based on these products. Additionally, these inconsistent pixels were mainly concentrated in transitional zones among land cover types and regions with fast urbanization or deforestation/afforestation, which are most sensitive to climate change and human activities. Consequently, these disparities will inevitably lead to significant variations in land cover change and attribution analysis based on different product evaluations, significantly undermining the trustworthiness of relevant research. Therefore, establishing an accurate validation dataset, assessing its precision, and delving into the underlying causes of classification disparities lay the groundwork for comprehending the credibility of land cover products and the uncertainty of related research outcomes.

3.2. Accuracy Assessment

According to the validation dataset from balanced random sampling, CLCD achieved the highest overall classification accuracy at 86.98 ± 0.76%, followed by CNLUCC at 81.38 ± 0.87%. GLC_FCS30 and Globeland30 showed lower accuracy at 77.83 ± 0.80% and 75.24 ± 0.91%, respectively (see Tables 2 and S2). None of the options consistently outperforms the others across all land cover types. Therefore, it is crucial to carefully

evaluate the strengths and weaknesses of each land cover dataset across various land cover types. As shown in Figure 6, the misclassification between grassland and bareland has the highest proportion, primarily concentrated in the northern regions of the Inner Mongolia Plateau, the Tibetan Plateau hinterland, and the Tarim Basin periphery within China. Subsequently, the forest was misclassified as cropland, particularly in GLC_FCS30, consistent with the significant divergence of GLC_FCS30 with other land cover products in cropland and forest.

Table 2. Comparison of mapping accuracy based on the balanced random sampling for GLC_FCS30, CLCD, Globeland30, and CNLUCC.

		Validation Dataset							
		Cropland	**Forest**	**Grassland**	**Wetland**	**Impervious**	**Bareland**	**Glacier**	**OA (%)**
GLC_FCS30	PA (%)	82.72	72.37	84.49	77.40	83.07	75.97	64.71	77.83
	UA (%)	57.25	92.48	72.37	73.52	62.24	90.74	76.74	
	F1 (%)	67.67	81.20	77.96	75.41	71.16	82.70	70.21	
CLCD	PA (%)	87.63	91.02	82.97	78.31	79.15	87.53	88.73	86.98
	UA (%)	84.72	93.22	85.87	69.41	71.68	87.88	73.26	
	F1 (%)	86.15	92.11	84.40	66.74	75.23	87.71	80.25	
Globeland30	PA (%)	79.44	82.47	69.56	63.98	69.35	74.18	83.05	75.24
	UA (%)	72.23	77.16	71.89	76.26	66.08	83.88	56.98	
	F1 (%)	75.66	79.73	70.71	69.58	67.67	78.73	67.59	
CNLUCC	PA (%)	82.17	87.97	76.50	70.20	76.73	83.25	53.45	81.38
	UA (%)	84.14	80.36	78.11	78.54	71.98	86.63	72.09	
	F1 (%)	83.14	83.99	77.30	74.14	74.28	84.91	61.39	

Based on stratified random sampling, there are substantial differences in identification accuracy across different products (see Table 3). CLCD holds a distinct advantage in identifying cropland, forests, grasslands, and impervious surfaces, though its performance in wetlands is much weaker than other products (see Figure 7b,e). CNLUCC shows good accuracy in recognizing impervious layers and cropland, and it outperforms other land cover products in wetland identification. However, its accuracy in forests is lower than that of other products, and it is prone to misclassifying forests as grassland or cropland. The GLC_FCS30 accurately identifies bareland and glaciers, outperforming other land cover products. However, its accuracy in identifying cropland and impervious surfaces could be improved, as many forests are misclassified as cropland or grassland. Although Globeland30 excels in wetland identification, it exhibits lower accuracy in other land cover types, particularly glaciers, where its accuracy is significantly below that of different products (see Figure 7c,e).

By comparison, misclassification samples have mainly distributed in inconsistent regions among land cover products (see Figures 4 and 6). Hence, it is reasonable to expect that the causes of misclassification are consistent with those of inconsistencies among land cover products. More misclassification in vegetation transitional zones means that the vegetation index plays a crucial role in classification accuracy. The lower classification accuracy in mountainous areas and humid regions indicated it may be sensitive to terrain relief and cloud cover, owing to their effect on the spectral reflection and available images.

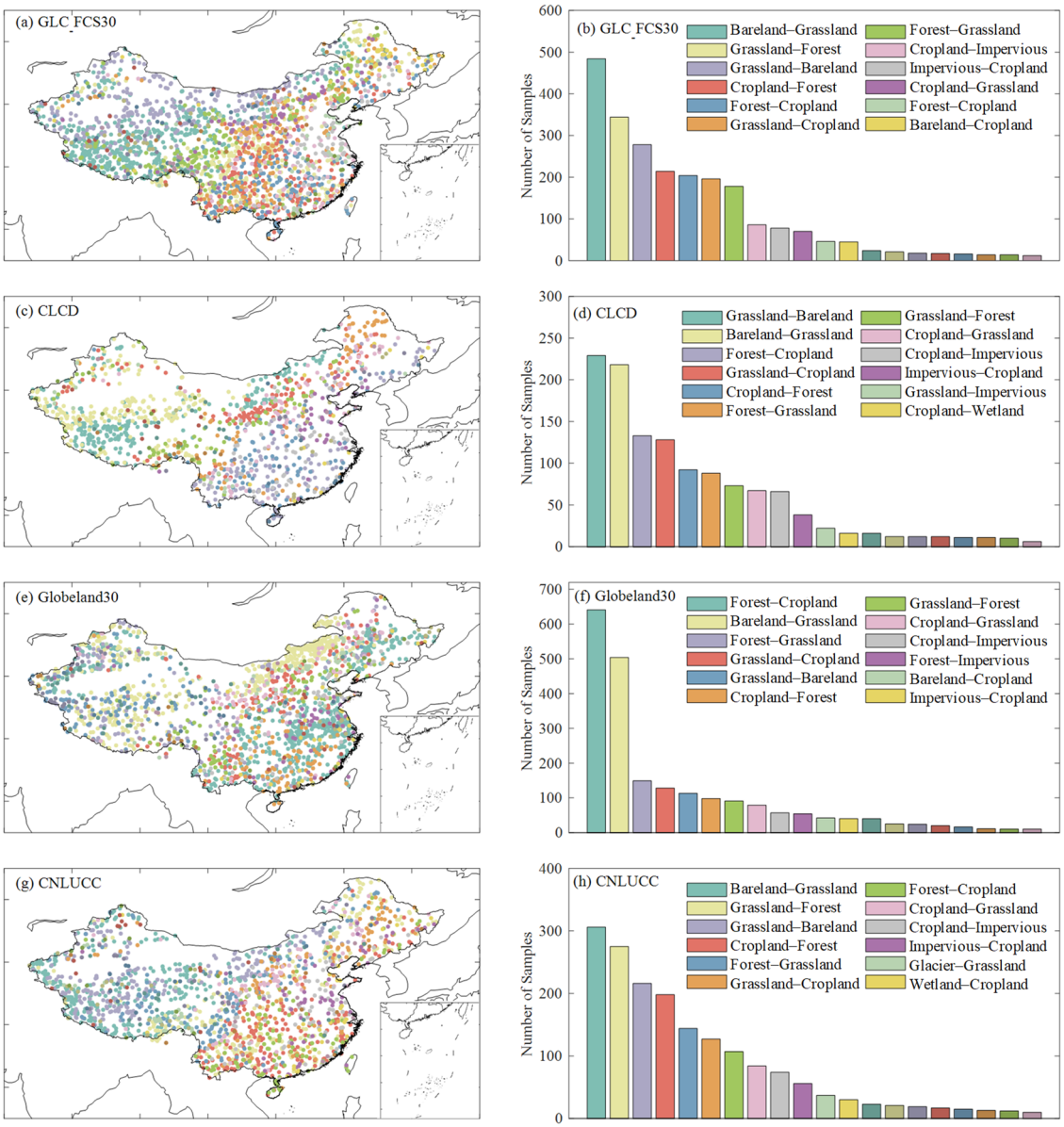

Figure 6. (**a**,**c**,**e**,**g**) The spatial pattern of the misclassification sample of land cover products; (**b**,**d**,**f**,**h**) the number of samples ranked in descending order of quantity. The legends only show the top 12 types of misclassification in terms of sample size.

Table 3. Comparison of mapping accuracy based on the stratified random sampling for GLC_FCS30, CLCD, Globeland30, and CNLUCC.

		Validation Dataset							
		Cropland	**Forest**	**Grassland**	**Wetland**	**Impervious**	**Bareland**	**Glacier**	**OA (%)**
GLC_FCS30	PA (%)	73.89	63.17	77.23	90.32	94.77	78.02	88.69	75.70
	UA (%)	57.30	92.45	72.31	73.52	62.24	90.65	76.74	
	F1 (%)	88.10	88.38	90.73	97.15	94.72	93.40	98.93	
CLCD	PA (%)	78.85	87.93	73.83	96.42	90.94	86.69	97.39	83.79
	UA (%)	84.62	93.16	85.83	69.41	71.68	87.88	73.26	
	F1 (%)	92.80	96.29	91.57	97.25	95.45	95.16	99.07	

Table 3. *Cont.*

		Validation Dataset							
		Cropland	**Forest**	**Grassland**	**Wetland**	**Impervious**	**Bareland**	**Glacier**	**OA (%)**
Globeland30	PA (%)	70.55	77.81	61.86	87.63	90.03	74.18	95.59	74.33
	UA (%)	72.23	76.98	71.91	76.26	66.08	83.92	56.98	
	F1 (%)	89.05	91.49	86.31	97.14	94.71	91.44	98.52	
CNLUCC	PA (%)	73.89	63.17	77.23	90.32	94.77	78.02	88.69	80.34
	UA (%)	57.30	92.45	72.31	73.52	62.24	90.65	76.74	
	F1 (%)	88.10	88.38	90.73	97.15	94.72	93.40	98.93	

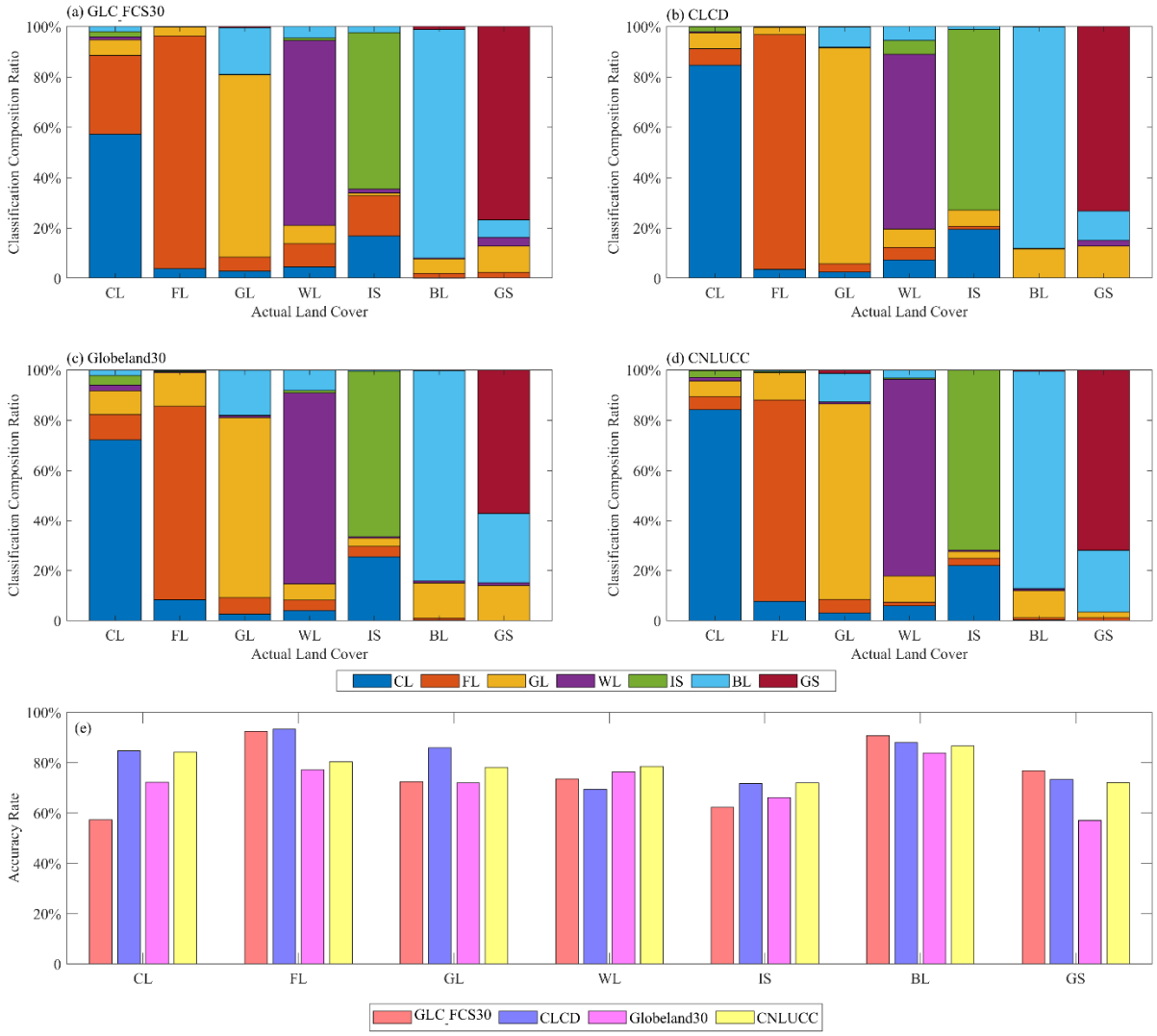

Figure 7. (**a**–**d**) The confusion proportions, a key factor in understanding the classification accuracy, for each of the land cover types in the validation dataset; (**e**) classification accuracy of land cover types in products. CL: cropland, FL: forest, GL: grassland, WL: wetland, IS: impervious, BL: bareland, GS: glaciers.

3.3. Misclassification Diagnosis

It is worth nothing that China's main land cover types are all related to vegetation. Vegetation index plays a crucial role in accurately distinguishing them. In Figure 8a–d, we observe a significant non-linear relationship between the classification accuracy and

the Normalized Difference Vegetation Index (NDVI), especially in GLC_FCS30. As NDVI increases, the classification accuracy initially decreases before rising again. The point of particular interest is the lowest accuracy around an NDVI value of 0.5. According to the probability density distribution (PDF), the overlap ratio of NDVI for cropland, grassland, forest, and impervious surfaces reaches the highest values around NDVI = 0.5. This overlap makes it challenging to accurately differentiate between these land cover types based on vegetation indicators, which may explain the lower classification accuracy in this NDVI interval.

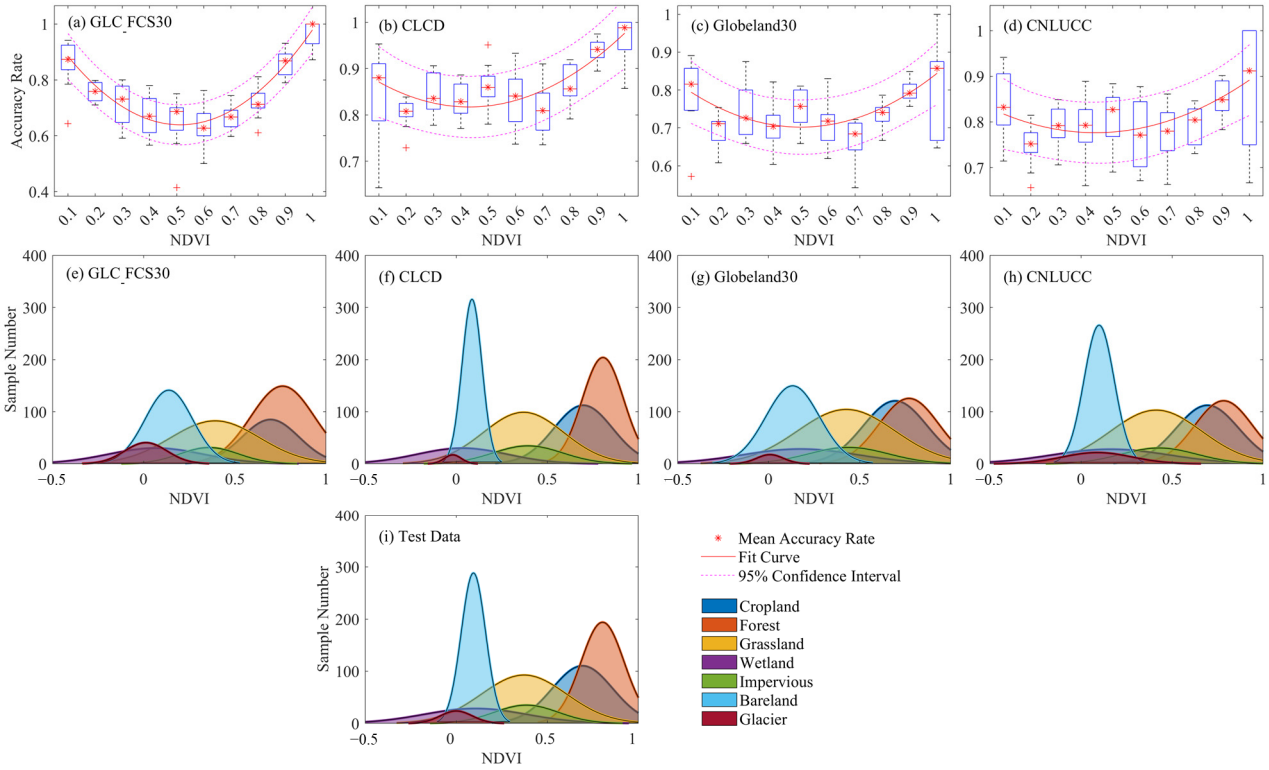

Figure 8. (**a**–**d**) The relationship between classification accuracy and NDVI; (**e**–**i**) the probability density distribution of NDVI for land cover types in each land cover product and validation dataset.

The land cover types in Globeland30 and GLC-FCS30 show significant differences in the PDF of NDVI compared to the validation dataset (see Figure 8e–i). In areas with high NDVI values, there is a broader overlap in PDF between cropland and forest. On the other hand, in regions with low NDVI values, the overlap interval of NDVI between bareland and grassland is significantly broader than that in the validation dataset. The higher overlap rate between land cover types indicates more significant spectral confusion, leading to more misclassification among related land cover types. This phenomenon elucidates the lower classification accuracy of vegetation-related land cover types such as grassland and forest in Globeland30 and GLC-FCS30.

As depicted in Figure 9, there is a clear negative correlation between classification accuracy and terrain relief. Among the terrain factors, the slope has the most significant impact on classification accuracy, with all products showing a noticeable decrease in accuracy as the slope increases. Specifically, classification accuracy decreases by approximately 1.1% to 2.6% for each degree increase in slope. Hill shade has a negative relationship with the classification accuracy of CLCD but no apparent effect on other products. Additionally, there is no significant difference in classification accuracy across eight aspects.

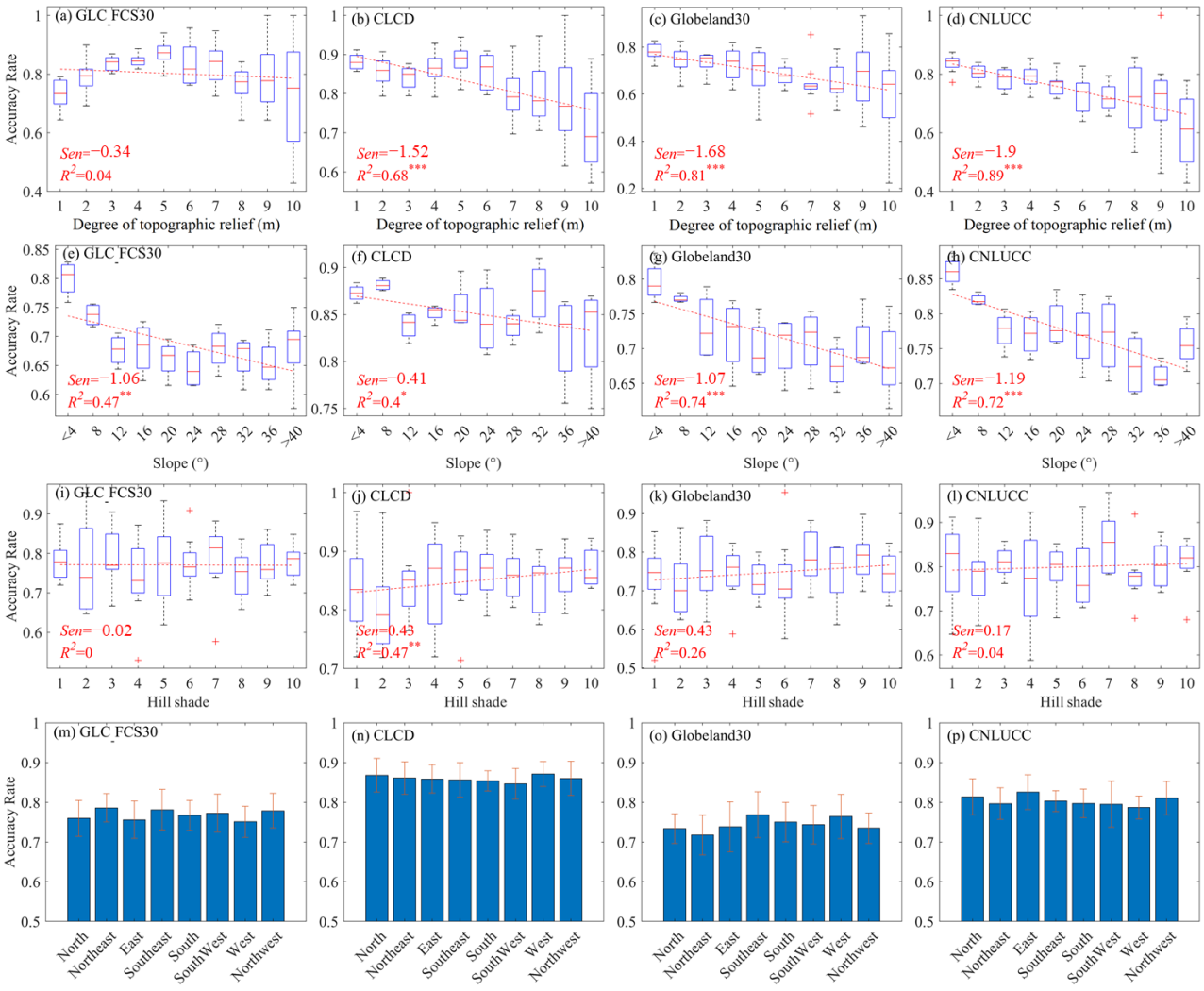

Figure 9. The relationship of classification accuracy with degree of topographic relief (**a**–**d**), slope (**e**–**h**), hill shade (**i**–**l**), and aspect (**m**–**p**). * means that the significance level is less than 0.1 ($\alpha < 0.1$), ** ($\alpha < 0.05$), *** ($\alpha < 0.01$).

To comprehensively assess the impact of mixed pixels and image availability, we carefully analyzed three key indices: landscape fragmentation, land cover change, and cloud mask (incorporating cloud cover and shadow). The presence of mixed pixels due to landscape fragmentation poses a challenge in accurately identifying land cover types. However, it is essential to note that the change in landscape fragmentation predominantly affects GLC_FCS30 (see Figure 10a), which may explain the higher misclassification of cropland and the impervious surface in GLC_FCS30 (see Figure 7e). Notably, the classification accuracy of all products is significantly affected by the land cover change, mainly showing a substantial decrease with an increased proportion of the land cover change in adjacent years. In particular, the sensitivity of CLCD and CNLUCC to land cover change surpasses that of GLC_FCS30 and Globeland30. Furthermore, cloud coverage significantly impacts Globeland30 and CNLUCC, with an increase in cloud cover notably decreasing the classification accuracy of these categories. However, cloud cover less affects the classification accuracy of GLC_FCS30 and CLCD.

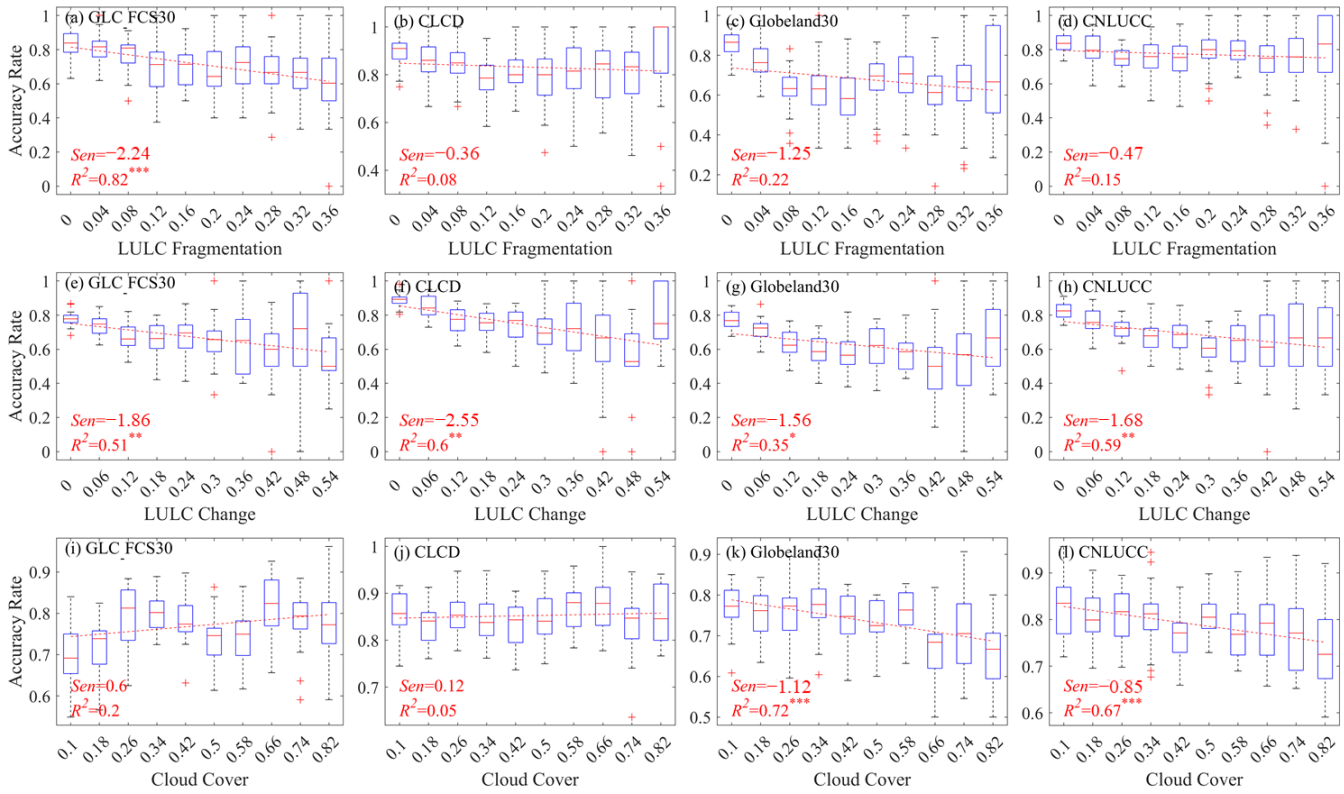

Figure 10. The relationship of classification accuracy with land cover fragmentation (LULC Fragmentation, **a**–**d**), land cover change (LULC Change, **e**–**h**), and cloud cover (**i**–**l**). * means that the significance level is less than 0.1 ($\alpha < 0.1$), ** ($\alpha < 0.05$), *** ($\alpha < 0.01$).

The partial correlation analysis examined the relationship between each variable and the classification accuracy of different land cover types. As shown in Figure 11, increasing NDVI increases the classification accuracy of forests but reduces those of bareland and impervious surfaces, resulting in the nonlinear relationship between NDVI and classification accuracy in Figure 8a–d. The slope reduces classification accuracy for all products, especially for land cover types distributed in mountainous areas. The wetland in CLCD is most sensitive to the effect of hill shade, which explains why the classification accuracy of wetlands in CLCD is considerably lower than that of other land cover products. Cropland and impervious surfaces, greatly affected by human activities, are more sensitive to land fragmentation in GLC_FCS30. Glaciers, cropland, and impervious layers are more sensitive to recent land cover changes and cloud impacts than other types. These types can change quickly due to human activities or climate change, so the data availability and timeliness requirements for classification are higher than other types.

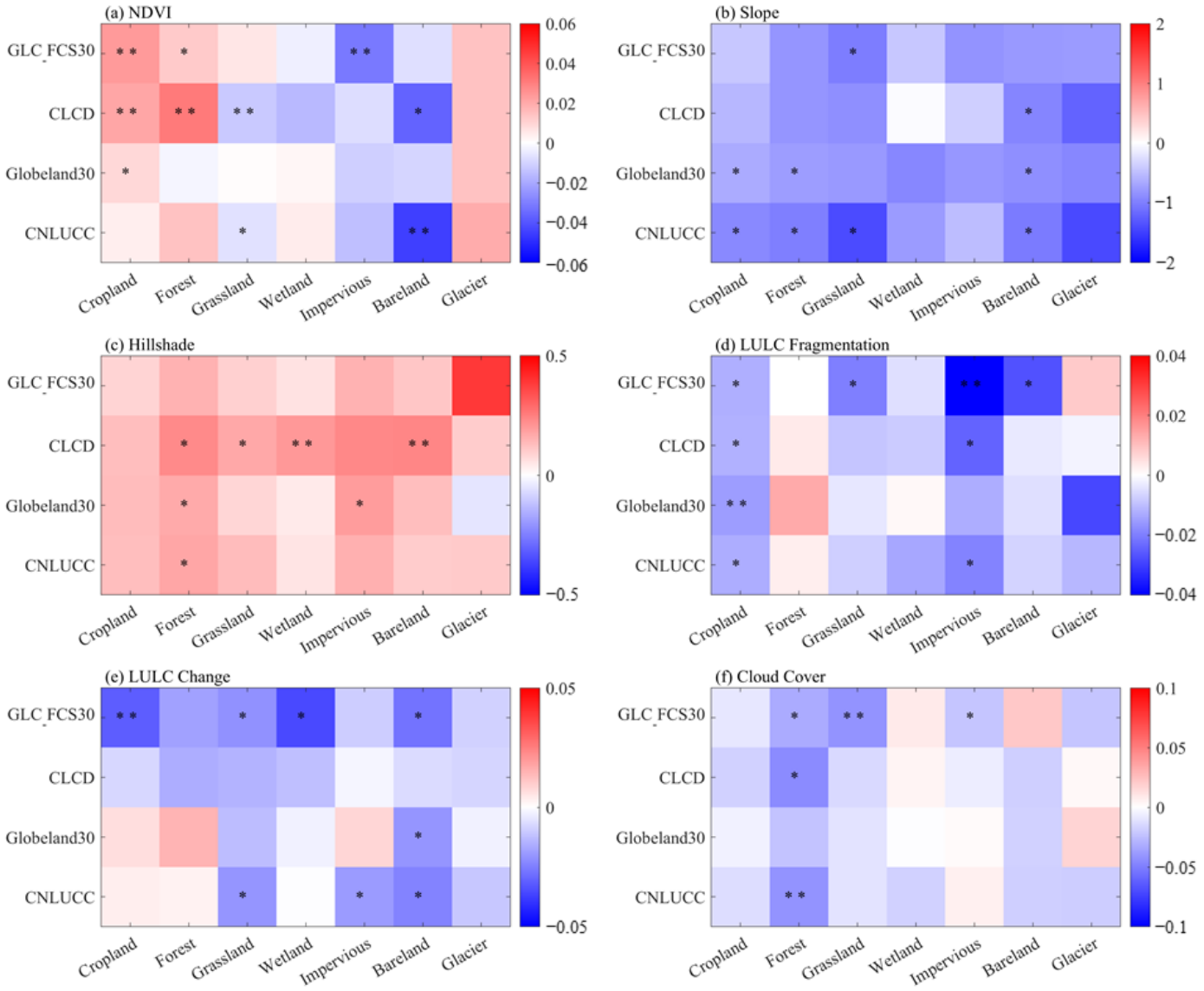

Figure 11. The partial correlation coefficients between land cover types and impact factors in different land cover products. * means that the significance level is less than 0.05 ($\alpha < 0.05$), ** ($\alpha < 0.01$).

4. Discussion

4.1. Accuracy Assessment Difference with Existing Studies

Compared to existing studies, our classification accuracy assessment generally aligns with previous assessments (see Table S3). However, there are still notable distinctions. The overall accuracy estimated in our study is slightly higher than that reported in previous studies. The classification accuracy of specific land cover types, such as cropland and impervious surfaces, is somewhat lower than those evaluated by existing studies. Those differences can be attributed to the following three main reasons.

Firstly, as mentioned earlier, to reduce the impact of possible biases in validation samples, we merge shrubs (except for sparse shrubland in GLC_FCS30) in forests and water bodies in wetlands, thereby reducing the effect of the misclassification among them on overall accuracy, which may explain why our evaluation results were better than previous studies (see Tables 2 and S3). Secondly, we evaluated the products from 2020, which have undergone more optimization in their classification methods compared to products from other periods, resulting in better performance. The variety and number of remote sensing images and auxiliary data available for land cover classification are also more abundant, making the classification accuracy in 2020 significantly better than in other

periods, consistent with previous results [21]. Additionally, with quasi-real-time high-resolution images, field photos, and street view data, we can more accurately determine the land cover type of the validation samples. We set more validation samples than previous studies in remote mountainous areas, such as the Tibetan Plateau, to give our validation dataset a better spatial representation.

4.2. Underlying Mechanism for Inconsistency and Misclassification

4.2.1. Spatial Extent

As mentioned earlier, Globeland30 and GLC_FCS30 are global classification products, whereas CLCD and CNLUCC are specific to China. Generally, global products have a larger spectral confusion among land cover types than those on a regional scale [71]. The optimal classification schemes for global datasets may not be as effective when applied to specific regions like China. In contrast, land cover products tailored for China can account for the region's notable heterogeneity and produce more accurate classifications. Secondly, regional products are specifically designed for China, leveraging local expertise and knowledge about the region's characteristics. This localized focus ensures that the classification models are fine-tuned to the specific conditions and variations within China's diverse landscapes. Land cover products in China often integrate various data sources, including satellite imagery, aerial photography, and ground surveys [33]. This multi-source integration provides a more comprehensive and accurate representation of land cover, leading to better classification performance.

Compared to Globeland30, GLC_FCS30 used local adaptive classification modeling to conduct the local classifier. They divided the globe into approximately 948 5° × 5° geographical tiles (equivalent to about 3 × 3 Landsat scenes) [32]. For every tile, they conducted a local classifier and employed it to create a corresponding regional land cover map and finally merge them into a global map [36]. This method alleviates the spectral confusion between land cover types caused by the increase in spatial extent, which may partly explain the better classification accuracy of GLC_FCS30 than Globeland30.

4.2.2. Classification Method

CNLUCC employs a human–machine interactive interpretation for its land cover classification. Visual interpretation is more effective for land cover types with complex spectral features and high internal heterogeneity, such as wetlands [74]. This method performs better in identifying land cover types with regular shapes, consistent with its better performance in cropland and impervious surfaces [75]. Additionally, this approach can classify mixed pixels depending on the importance of land cover types rather than their area proportion. For example, many rivers are discontinuous in other products because narrow parts of rivers may be identified as other land cover types. CNLUCC can maintain the continuity of rivers well, thereby improving their effectiveness and availability in hydrological process analysis and model simulation (see Figure S5). However, visual interpretation also has significant shortcomings, such as using remote sensing images at a certain time rather than fusing multi-source data and limited spectral information available. Therefore, it is susceptible to data availability, possibly due to its sensitivity to cloud cover and land cover change. In addition, manual visual interpretation requires a lot of labor and subjective judgment from interpreters, making it challenging to apply to large-scale land cover classification.

Globeland30 employs the POK (Pixel–Object–Knowledge) classification method to achieve broader coverage and higher efficiency [76]. This approach effectively integrates the advantages of pixel scale, object scale, and prior knowledge, significantly enhancing the accuracy of automated classification and thereby considerably improving the efficiency and accuracy [34]. Our results indicate that Globeland30 accurately identifies land cover types with relatively regular shapes, such as cropland. Additionally, the object-based classification method used by Globeland30 effectively reduces the salt-and-pepper phenomenon commonly seen in pixel-based methods [36,77]. However, this method also has certain

limitations. For instance, in many regions of China, artificial forests with regular shapes are often misclassified as cropland (see Figure S6). Additionally, Globeland30's classification of vegetation-related types relies solely on vegetation indices, resulting in lower classification accuracy. The accuracy of Globeland30 is also heavily dependent on prior knowledge. For objects with greater internal heterogeneity, such as wetlands, the classification accuracy of Globeland30 is noticeably lower.

Both CLCD and GLC_FCS30 employ the random forest algorithm in classification, which is highly efficient and can reflect more spatial details of land cover types. By effectively utilizing available remote sensing images, it significantly reduces sensitivity to the availability of a single data source. As a result, it is less affected by factors such as cloud cover and can produce products with higher temporal resolution. For instance, CLCD publishes annual land cover products. Moreover, the random forest algorithm is less influenced by prior knowledge and subjective errors, providing a more uniform time series of land cover products. This consistency supports the trend analysis in land cover changes. However, the random forest algorithm highly depends on training samples and feature variables. Consequently, there is a significant difference in classification accuracy between CLCD and GLC_FCS30. The higher classification accuracy of CLCD can be attributed to the greater number of training samples in China and the use of more vegetation indices as feature variables, which significantly improves the accuracy of vegetation type identification.

GLC_FCS30 primarily used a pixel-based random forest classification, which may result in numerous misclassified pixels, known as the "salt-and-pepper effect" (see Figure S7). CLCD used a spatial–temporal post-processing method to enhance classification accuracy. This approach uses a filter and logical reasoning to refine mapping results by suppressing illogical land cover conversions caused by misclassifications [21]. Hence, it can effectively utilize classification results from adjacent years and prior knowledge to correct misclassifications, improving accuracy and reducing the "salt-and-pepper phenomenon" in pixel-based classifiers [78].

4.2.3. Training Samples

For supervised large-scale land cover (LC) mapping, the accuracy and adequacy of training and validation samples are crucial [79]. Generally, the strategies for collecting training samples for large-scale mapping tasks include (1) visually interpreted samples and (2) samples automatically derived from existing land cover products [80–82]. The visual interpretation method yields high-quality samples but requires significant human labor. Conversely, automatic sample extraction using existing land cover products can generate many randomly distributed samples, although the quality depends on the products used.

The training samples of Globeland30 and CNLUCC involve the participation of multiple experts based on the spectral features of satellite images. With field surveys, creators refer to relevant geographical maps and analyze land cover types' shape, color, texture, and spatial distribution, establishing a harmonious interpretation marker library [76]. Owing to the above object-based features, the land cover types with regular shapes or unique textures, such as cropland and impervious surfaces, are more accurately classified in CNLUCC and Globeland30 [83]. However, in some exceptional cases, object-based features may lead to misclassification of land cover types, such as the large number of artificial forests distributed in China, which have regular shapes and are easily misclassified as cropland. Due to terrain limitations, many croplands present irregular shapes and are easily misclassified as forests in mountainous areas. Additionally, land cover types are classified by qualitative features rather than quantitatively defined thresholds; some transitional areas among land cover types, such as sparse grassland, are prone to misclassification due to the subjective differences between workers (see Figures 4 and 6).

Both CLCD and GLC_FCS30 employ the random forest algorithm, but the overall accuracy of the former is notably higher than the latter. This difference underscores the crucial role of training samples in enhancing classification accuracy. CLCD and GLC_FCS30

are built on training samples derived from CNLUCC and CCI_LC, respectively, with the reliability of these samples verified using MODIS products like MODIS NDVI and EVI [21,32]. In contrast to CNLUCC and Globeland30, the spatial distribution of training samples in CLCD and GLC_FCS30 is more random, allowing for stratified sampling based on different land cover types. This approach ensures the representativeness of land cover types in the training samples, leading to superior classification accuracy in remote areas, such as the Tibetan Plateau.

Additionally, CLCD's training samples are derived from CNLUCC with a resolution of 30 m, significantly higher than the 300m resolution of CCI_LC [84]. Furthermore, CLCD has refined and optimized its training samples further using high-resolution imagery and field photos in Google Earth, which is likely one of the reasons for its superior classification accuracy compared to GLC_FCS30. Moreover, CLCD employs more vegetation indices as feature vectors, significantly enhancing its accuracy in identifying vegetation types, particularly in distinguishing between cropland and forest, and in the classification accuracy of bareland and grassland, which is markedly superior to that of GLC_FCS30. The extensive use of spectral classification can lead to significant inaccuracies in CLCD due to factors such as mountain shadows affecting surface reflection characteristics. Figure S5 shows that the bareland and glaciers under mountain shadows are misclassified as wetlands in CLCD, which may explain the poor accuracy of wetlands.

5. Conclusions

The information on land surfaces is crucial for understanding environmental changes, ensuring food security, and coordinating global change mitigation efforts. Accurate and reliable 30m-resolution land cover products are essential for studying land cover change. However, there are significant discrepancies among available products. According to our comparison, the consistent area across the products accounts for 50.34% of the total area. The inconsistent regions between each pair of products range from 21.10% to 37.53%. This substantial inconsistency has introduced significant uncertainties in related studies about land cover changes.

Overall, we analyzed the main types of inconsistencies and spatial patterns by comparing four existing 30m-resolution land cover products in China in 2020. We evaluated their classification accuracy using more accurate validation samples derived from multi-source datasets. CLCD had the highest overall accuracy (86.98 $\pm$ 0.76%), followed by CNLUCC and GLC_FCS30, with overall accuracies of 81.38 $\pm$ 0.87% and 77.83 $\pm$ 0.80%, respectively. However, no product showed superior detection accuracy across all land cover types. CLCD had higher classification accuracy for vegetation types, but its recognition accuracy for wetlands was lower than that of other products. Globeland30 had the lowest overall accuracy, but its classification accuracy for glaciers, barelands, and forests was higher than other products. Therefore, selecting land cover products should be based on the research subject.

Importantly, we analyzed the relationship between classification accuracy and related factors and explored the underlying causes of inconsistency and misclassification among land cover products. Spectral confusion due to similar vegetation cover significantly affects classification accuracy. Slope, landscape fragmentation, and cloud cover also decrease classification accuracy by altering spectral features and reducing available images. Land fragmentation, land cover change, and cloud cover can explain the low classification accuracy of most land use types in GLC_FCS30. CLCD accuracy is more sensitive to variations in vegetation cover and topography. In Globeland30 and CNLUCC, the response of land cover accuracy to each influencing factor shows no clear pattern.

Digging deeper into the mechanism behind the results, human–machine interactive interpretation performs well in land cover types with unique textures, while automatic classification algorithms become more efficient and accurate, especially for vegetation-related land cover types. Object-based automatic classification and expert knowledge verification significantly improve classification accuracy based on unique shape or texture

features while increasing the risk of misclassification caused by texture confusion. Random forest algorithms improve land classification efficiency and provide detailed land cover patterns. Additionally, their accuracy is also highly sensitive to training samples.

These research results can provide theoretical support for selecting appropriate land cover products for associated studies and optimizing existing classification algorithms in the future.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16224330/s1, refs. [85–92].

Author Contributions: Methodology, X.X. and D.L.; Validation, X.X.; Formal analysis, W.Y.; Investigation, X.X., D.L., G.Z. and W.Y.; Resources, H.L.; Data curation, D.L.; Writing—original draft, X.X.; Writing—review & editing, J.D.; Visualization, H.L.; Supervision, H.L., B.C. and Y.Y.; Project administration, B.C.; Funding acquisition, X.X., G.Z. and B.C. 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 (2023YFC3209003, 2022YFC3202001, and 2022YFF1301801); the National Science Fund for Distinguished Young Scholars (52025092), and the National Natural Science Foundation of China (41930970, 42475198, and 52109074).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Winkler, K.; Fuchs, R.; Rounsevell, M.; Herold, M. Global land use changes are four times greater than previously estimated. *Nat. Commun.* **2021**, *12*, 2501. [CrossRef] [PubMed]
2. Maeda, E.E.; Aragão, L.E.O.C.; Baker, J.C.A.; Balbino, L.C.; de Moura, Y.M.; Nobre, A.D.; Nunes, M.H.; Silva Junior, C.H.L.; dos Reis, J.C. Land use still matters after deforestation. *Commun. Earth Environ.* **2023**, *4*, 29. [CrossRef]
3. Hu, J.; Wu, Y.; Wang, L.; Sun, P.; Zhao, F.; Jin, Z.; Wang, Y.; Qiu, L.; Lian, Y. Impacts of land-use conversions on the water cycle in a typical watershed in the southern Chinese Loess Plateau. *J. Hydrol.* **2021**, *593*, 125741. [CrossRef]
4. Liang, Y.; Liu, L. Simulating land-use change and its effect on biodiversity conservation in a watershed in northwest China. *Ecosyst. Health Sustain.* **2017**, *3*, 1335933. [CrossRef]
5. Sterling, S.M.; Ducharne, A.; Polcher, J. The impact of global land-cover change on the terrestrial water cycle. *Nat. Clim. Change* **2013**, *3*, 385–390. [CrossRef]
6. Qin, Z.; Zhu, Y.; Canadell, J.G.; Chen, M.; Li, T.; Mishra, U.; Yuan, W. Global spatially explicit carbon emissions from land-use change over the past six decades (1961–2020). *One Earth* **2024**, *7*, 835–847. [CrossRef]
7. Feddema, J.J.; Oleson, K.W.; Bonan, G.B.; Mearns, L.O.; Buja, L.E.; Meehl, G.A.; Washington, W.M. The Importance of Land-Cover Change in Simulating Future Climates. *Science* **2005**, *310*, 1674–1678. [CrossRef]
8. Xu, H.; Yue, C.; Zhang, Y.; Liu, D.; Piao, S. Forestation at the right time with the right species can generate persistent carbon benefits in China. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2304988120. [CrossRef]
9. Ge, J.; Liu, Q.; Zan, B.; Lin, Z.; Lu, S.; Qiu, B.; Guo, W. Deforestation intensifies daily temperature variability in the northern extratropics. *Nat. Commun.* **2022**, *13*, 5955. [CrossRef]
10. Gomes, L.C.; Bianchi, F.J.J.A.; Cardoso, I.M.; Schulte, R.P.O.; Arts, B.J.M.; Fernandes Filho, E.I. Land use and land cover scenarios: An interdisciplinary approach integrating local conditions and the global shared socioeconomic pathways. *Land Use Policy* **2020**, *97*, 104723. [CrossRef]
11. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [CrossRef]
12. Feng, H.; Wang, S.; Zou, B.; Nie, Y.; Ye, S.; Ding, Y.; Zhu, S. Land use and cover change (LUCC) impacts on Earth's eco-environments: Research progress and prospects. *Adv. Space Res.* **2023**, *71*, 1418–1435. [CrossRef]
13. Schewe, J.; Gosling, S.N.; Reyer, C.; Zhao, F.; Ciais, P.; Elliott, J.; Francois, L.; Huber, V.; Lotze, H.K.; Seneviratne, S.I.; et al. State-of-the-art global models underestimate impacts from climate extremes. *Nat. Commun.* **2019**, *10*, 1005. [CrossRef] [PubMed]
14. Venter, Z.S.; Barton, D.N.; Chakraborty, T.; Simensen, T.; Singh, G. Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sens.* **2022**, *14*, 4101. [CrossRef]
15. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [CrossRef]

16. Zhu, Z.; Qiu, S.; Ye, S. Remote sensing of land change: A multifaceted perspective. *Remote Sens. Environ.* **2022**, *282*, 113266. [CrossRef]
17. Toure, S.I.; Stow, D.A.; Shih, H.-C.; Weeks, J.; Lopez-Carr, D. Land cover and land use change analysis using multi-spatial resolution data and object-based image analysis. *Remote Sens. Environ.* **2018**, *210*, 259–268. [CrossRef]
18. Pelorosso, R.; Apollonio, C.; Rocchini, D.; Petroselli, A. Effects of Land Use-Land Cover Thematic Resolution on Environmental Evaluations. *Remote Sens.* **2021**, *13*, 1232. [CrossRef]
19. Li, X.; Gong, P.; Liang, L. A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. *Remote Sens. Environ.* **2015**, *166*, 78–90. [CrossRef]
20. Xu, Y.; Yu, L.; Peng, D.; Zhao, J.; Cheng, Y.; Liu, X.; Li, W.; Meng, R.; Xu, X.; Gong, P. Annual 30-m land use/land cover maps of China for 1980–2015 from the integration of AVHRR, MODIS and Landsat data using the BFAST algorithm. *Sci. China Earth Sci.* **2020**, *63*, 1390–1407. [CrossRef]
21. 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]
22. Gong, P.; Liu, H.; Zhang, M.N.; Li, C.C.; Wang, J.; Huang, H.B.; Clinton, N.; Ji, L.Y.; Li, W.Y.; Bai, Y.Q.; 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]
23. 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 IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4704–4707.
24. Gong, P. Remote sensing of environmental change over China: A review. *Chin. Sci. Bull.* **2012**, *57*, 2793–2801. [CrossRef]
25. Otukei, J.R.; Blaschke, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S27–S31. [CrossRef]
26. Adugna, T.; Xu, W.; Fan, J. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sens.* **2022**, *14*, 574. [CrossRef]
27. Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* **2012**, *121*, 93–107. [CrossRef]
28. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
29. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
30. Scott, G.J.; England, M.R.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [CrossRef]
31. Yang, C.; Rottensteiner, F.; Heipke, C. Classification of land cover and land use based on convolutional neural networks. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *IV-3*, 251–258. [CrossRef]
32. 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]
33. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W.; Zhang, S.; Li, R.; Yan, C. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geogr. Sci.* **2014**, *24*, 195–210. [CrossRef]
34. 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]
35. Li, W.; Dong, R.; Fu, H.; Wang, J.; Yu, L.; Gong, P. Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping. *Remote Sens. Environ.* **2020**, *237*, 111563. [CrossRef]
36. Zhang, X.; Liu, L.; Wu, C.; Chen, X.; Gao, Y.; Xie, S.; Zhang, B. Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform. *Earth Syst. Sci. Data* **2020**, *12*, 1625–1648. [CrossRef]
37. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
38. Zhao, T.; Zhang, X.; Gao, Y.; Mi, J.; Liu, W.; Wang, J.; Jiang, M.; Liu, L. Assessing the Accuracy and Consistency of Six Fine-Resolution Global Land Cover Products Using a Novel Stratified Random Sampling Validation Dataset. *Remote Sens.* **2023**, *15*, 2285. [CrossRef]
39. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [CrossRef]
40. Bai, Y.; Feng, M.; Jiang, H.; Wang, J.; Zhu, Y.; Liu, Y. Assessing Consistency of Five Global Land Cover Data Sets in China. *Remote Sens.* **2014**, *6*, 8739–8759. [CrossRef]
41. Hsiao, L.-H.; Cheng, K.-S. Assessing Uncertainty in LULC Classification Accuracy by Using Bootstrap Resampling. *Remote Sens.* **2016**, *8*, 705. [CrossRef]

42. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131. [CrossRef]
43. Wickham, J.D.; Stehman, S.V.; Gass, L.; Dewitz, J.; Fry, J.A.; Wade, T.G. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* **2013**, *130*, 294–304. [CrossRef]
44. Cai, S.; Liu, D.; Sulla-Menashe, D.; Friedl, M.A. Enhancing MODIS land cover product with a spatial–temporal modeling algorithm. *Remote Sens. Environ.* **2014**, *147*, 243–255. [CrossRef]
45. Gao, Y.; Liu, L.; Zhang, X.; Chen, X.; Mi, J.; Xie, S. Consistency Analysis and Accuracy Assessment of Three Global 30-m Land-Cover Products over the European Union using the LUCAS Dataset. *Remote Sens.* **2020**, *12*, 3479. [CrossRef]
46. Zhao, Y.; Gong, P.; Yu, L.; Hu, L.; Li, X.; Li, C.; Zhang, H.; Zheng, Y.; Wang, J.; Zhao, Y.; et al. Towards a common validation sample set for global land-cover mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814. [CrossRef]
47. Gong, P. Some essential questions in remote sensing science and technology. *J. Remote Sens.* **2009**, *13*, 1–23.
48. 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]
49. Fritz, S.; See, L.; Perger, C.; McCallum, I.; Schill, C.; Schepaschenko, D.; Duerauer, M.; Karner, M.; Dresel, C.; Laso-Bayas, J.-C.; et al. A global dataset of crowdsourced land cover and land use reference data. *Sci. Data* **2017**, *4*, 170075. [CrossRef] [PubMed]
50. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy assessment of seven global land cover datasets over China. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 156–173. [CrossRef]
51. Ye, S.; Ren, S.; Song, C.; Du, Z.; Wang, K.; Du, B.; Cheng, F.; Zhu, D. Spatial pattern of cultivated land fragmentation in mainland China: Characteristics, dominant factors, and countermeasures. *Land Use Policy* **2024**, *139*, 107070. [CrossRef]
52. Qian, F.; Chi, Y.; Lal, R.; Lorenz, K. Spatio-temporal characteristics of cultivated land fragmentation in different landform areas with a case study in Northeast China. *Ecosyst. Health Sustain.* **2020**, *6*, 1800415. [CrossRef]
53. Yao, S.; Zhang, Z. Regional Growth in China Under Economic Reforms. *J. Dev. Stud.* **2001**, *38*, 167–186. [CrossRef]
54. Kuang, W.; Liu, J.; Dong, J.; Chi, W.; Zhang, C. The rapid and massive urban and industrial land expansions in China between 1990 and 2010: A CLUD-based analysis of their trajectories, patterns, and drivers. *Landsc. Urban Plan.* **2016**, *145*, 21–33. [CrossRef]
55. Yin, R.; Yin, G. China's primary programs of terrestrial ecosystem restoration: Initiation, implementation, and challenges. *Environ. Manag.* **2010**, *45*, 429–441. [CrossRef]
56. Chen, C.; Park, T.; Wang, X.; Piao, S.; Xu, B.; Chaturvedi, R.K.; Fuchs, R.; Brovkin, V.; Ciais, P.; Fensholt, R.; et al. China and India lead in greening of the world through land-use management. *Nat. Sustain.* **2019**, *2*, 122–129. [CrossRef] [PubMed]
57. Zhang, Y.; Wang, X.; Lian, X.; Li, S.; Li, Y.; Chen, C.; Piao, S. Asymmetric impacts of forest gain and loss on tropical land surface temperature. *Nat. Geosci.* **2024**, *17*, 426–432. [CrossRef]
58. Findell, K.L.; Berg, A.; Gentine, P.; Krasting, J.P.; Lintner, B.R.; Malyshev, S.; Santanello, J.A.; Shevliakova, E. The impact of anthropogenic land use and land cover change on regional climate extremes. *Nat. Commun.* **2017**, *8*, 989. [CrossRef]
59. Butt, E.W.; Baker, J.C.A.; Bezerra, F.G.S.; von Randow, C.; Aguiar, A.P.D.; Spracklen, D.V. Amazon deforestation causes strong regional warming. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2309123120. [CrossRef]
60. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef]
61. Purinton, B.; Bookhagen, B. Validation of digital elevation models (DEMs) and comparison of geomorphic metrics on the southern Central Andean Plateau. *Earth Surf. Dynam.* **2017**, *5*, 211–237. [CrossRef]
62. 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]
63. Shen, S.; Mo, X.; Zhang, Q. Land Use/Cover Classification of Cloud-Contaminated Area by Multitemporal Remote Sensing Images. In Proceedings of the 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, Washington, DC, USA, 26–27 August 2014; pp. 156–159.
64. Skakun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Frantz, D.; Gascon, F.; Gómez-Chova, L.; Hagolle, O.; et al. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990. [CrossRef]
65. Irwin, E.G.; Bockstael, N.E. The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20672–20677. [CrossRef] [PubMed]
66. Stehman, S.V.; Foody, G.M. Key issues in rigorous accuracy assessment of land cover products. *Remote Sens. Environ.* **2019**, *231*, 111199. [CrossRef]
67. Strahler, A.H.; Boschetti, L.; Foody, G.M.; Friedl, M.A.; Hansen, M.C.; Herold, M.; Mayaux, P.; Morisette, J.T.; Stehman, S.V.; Woodcock, C.E. Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *Eur. Communities Luxemb.* **2006**, *51*, 1–60.
68. Stehman, S.V. Sampling designs for accuracy assessment of land cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272. [CrossRef]
69. Hedayat, A.S.; Stufken, J. Sampling designs to control selection probabilities of contiguous units. *J. Stat. Plan. Inference* **1998**, *72*, 333–345. [CrossRef]
70. Benedetti, R.; Piersimoni, F.; Postiglione, P. Spatially Balanced Sampling: A Review and A Reappraisal. *Int. Stat. Rev.* **2017**, *85*, 439–454. [CrossRef]

71. Small, C.; Sousa, D. Spectral Characteristics of the Dynamic World Land Cover Classification. *Remote Sens.* **2023**, *15*, 575. [CrossRef]
72. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]
73. Wang, J.; Zhao, Y.; Li, C.; Yu, L.; Liu, D.; Gong, P. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250m resolution. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 38–47. [CrossRef]
74. Tarko, A.; Tsendbazar, N.-E.; de Bruin, S.; Bregt, A.K. Producing consistent visually interpreted land cover reference data: Learning from feedback. *Int. J. Digit. Earth* **2021**, *14*, 52–70. [CrossRef]
75. Mas, J.-F.; Lemoine-Rodríguez, R.; González-López, R.; López-Sánchez, J.; Piña-Garduño, A.; Herrera-Flores, E. Land use/land cover change detection combining automatic processing and visual interpretation. *Eur. J. Remote Sens.* **2017**, *50*, 626–635. [CrossRef]
76. Chen, J.; Cao, X.; Peng, S.; Ren, H. Analysis and Applications of GlobeLand30: A Review. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 230. [CrossRef]
77. Behnia, N.; Zare, M.; Moosavi, V.; Khajeddin, S.I. Evaluation of a Hierarchical Classification Method and Statistical Comparison with Pixel-Based and Object-Oriented Approaches. *Ecopersia* **2020**, *8*, 209–219.
78. Wehmann, A.; Liu, D. A spatial–temporal contextual Markovian kernel method for multi-temporal land cover mapping. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 77–89. [CrossRef]
79. Foody, G.M.; Arora, M.K. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *Int. J. Remote Sens.* **1997**, *18*, 799–810. [CrossRef]
80. Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tsend-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaaideh, B.; et al. Production of global land cover data—GLCNMO. *Int. J. Digit. Earth* **2011**, *4*, 22–49. [CrossRef]
81. Xie, S.; Liu, L.; Zhang, X.; Chen, X. Annual land-cover mapping based on multi-temporal cloud-contaminated landsat images. *Int. J. Remote Sens.* **2019**, *40*, 3855–3877. [CrossRef]
82. Zhu, Z.; Gallant, A.L.; Woodcock, C.E.; Pengra, B.; Olofsson, P.; Loveland, T.R.; Jin, S.; Dahal, D.; Yang, L.; Auch, R.F. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 206–221. [CrossRef]
83. Ma, J.; Sun, Q.; Xu, L.; Wen, B.; Li, Y. Comparison Analysis of GlobeLand 30 and Volunteered Geographic Information. *J. Geo-Inf. Sci.* **2018**, *20*, 1225–1234. [CrossRef]
84. Lamarche, C.; Santoro, M.; Bontemps, S.; D'Andrimont, R.; Radoux, J.; Giustarini, L.; Brockmann, C.; Wevers, J.; Defourny, P.; Arino, O. Compilation and Validation of SAR and Optical Data Products for a Complete and Global Map of Inland/Ocean Water Tailored to the Climate Modeling Community. *Remote Sens.* **2017**, *9*, 36. [CrossRef]
85. *GB/T 21010-2017*; Current Land Use Condition Classification. MLRC (Ministry of Land and Resource of China): Beijing, China, 2017.
86. Bie, Q.; Shi, Y.; Li, X.; Wang, Y. Contrastive Analysis and Accuracy Assessment of Three Global 30 m Land Cover Maps Circa 2020 in Arid Land. *Sustainability* **2023**, *15*, 741. [CrossRef]
87. Cui, P.; Chen, T.; Li, Y.; Liu, K.; Zhang, D.; Song, C. Comparison and Assessment of Different Land Cover Datasets on the Cropland in Northeast China. *RemoteSens.* **2023**, *15*, 5134. [CrossRef]
88. Hao, X.; Qiu, Y.; Jia, G.; Menenti, M.; Ma, J.; Jiang, Z. Evaluation of Global Land Use-Land Cover Data Products in Guangxi, China. *Remote Sens.* **2023**, *15*, 1291. [CrossRef]
89. Hou, M.; Ge, J.; Xiu, Y.; Meng, B.; Liu, J.; Feng, Q.; Liang, T. The urgent need to develop a new grassland map in China: Based on the consistency and accuracy of ten land cover products. *Sci. China* **2023**, *66*, 385–405. [CrossRef]
90. Liu, J.; Ren, Y.; Chen, X. Regional Accuracy Assessment of 30-Meter GLC_FCS30, GlobeLand30, and CLCD Products: A Case Study in Xinjiang Area. *Remote Sens.* **2024**, *16*, 82. [CrossRef]
91. Sun, B.; Chen, X.; Zhou, Q. Uncertainty assessment of globeland30 land cover data set over central asia. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 1313–1317. [CrossRef]
92. Xu, H.; Jiang, L.; Liu, Y. Assessing the Accuracy and Consistency of Cropland Products in the Middle Yangtze Plain. *Land* **2024**, *13*, 301. [CrossRef]