

Article

Consistency and Accuracy of Four High-Resolution LULC Datasets—Indochina Peninsula Case Study

Hao Wang ^{1,2} , Huimin Yan ^{1,2}, Yunfeng Hu ^{1,2,*} , Yue Xi ^{2,3} and Yichen Yang ^{1,2} 

¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; wangh.20s@igsrr.ac.cn (H.W.); yanhm@igsrr.ac.cn (H.Y.); yangyc.20s@igsrr.ac.cn (Y.Y.)

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; xiy.19s@igsrr.ac.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

* Correspondence: huyf@reis.ac.cn

Abstract: Open and high-temporal- and spatial-resolution global land use/land cover (LULC) mapping data form the foundation of global change research and cross-scale land management planning. However, the consistency and reliability of the use of multisource LULC datasets in specific regions need to be quantitatively assessed. In this study, we selected the Indochina Peninsula as the research area; considered four datasets: LSV10, GLC_FCS30, ESRI10, and Globeland30; and analyzed them from four dimensions: the similarity of composition type, the degree of category confusion, spatial consistency, and data accuracy. The results show that: (1) the land composition descriptions of the different datasets are consistent. The study area is dominated by forest and cropland, supplemented by grassland, shrubland, and other land types. (2) The correlation coefficient between datasets is between 0.905 and 0.972; the spatial consistency of datasets is good; and the high-consistency area accounts for 77.87% of the total. (3) The overall accuracy of LSV10 is the highest (83.25%), and that of GLC_FCS30 is the lowest (72.27%). The accuracy of cropland, forest, water area, and built-up land is generally high (above 85%); the accuracy of grassland, shrubland, and bare land is low (below 60%). Therefore, researchers must conduct validation for specific regions and specific land types before using the above datasets. Our findings provide a basis for selecting LULC datasets in related research on the Indochina Peninsula and a reference method for assessing the reliability of multisource LULC datasets in other regions.

Keywords: land mapping; dataset quality; consistency; data accuracy; Southeast Asia



Citation: Wang, H.; Yan, H.; Hu, Y.; Xi, Y.; Yang, Y. Consistency and Accuracy of Four High-Resolution LULC Datasets—Indochina Peninsula Case Study. *Land* **2022**, *11*, 758. <https://doi.org/10.3390/land11050758>

Academic Editors: Christine Fürst, Chuanrong Zhang, Hossein Azadi, Thomas Panagopoulos, Le Yu and Saskia Keesstra

Received: 1 May 2022

Accepted: 19 May 2022

Published: 22 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land use/land cover (LULC) is the result of a combination of natural and artificial forces, reflecting the natural attributes of land and the impact of human activities [1]. The spatial distribution pattern of and dynamic changes in LULC not only affect the regional economic-social development but also regional environmental and climate change [2]. Satellite remote sensing technology provides strong technical support for the large-scale and rapid acquisition of LULC information [3]. In 1995, the International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions Programme on Global Environmental Change (IHDP) jointly proposed the Land Use and Land Cover Change project to evaluate the ecological and environmental effects of LULC change by studying the mechanisms through which human society, the ecological environment, and LULC change interact. Researchers are now paying more attention to LULC changes, which have become a frontier and hot topic in current global change research [4–6].

Several medium- and high-resolution global-scale LULC datasets are available for users to browse and download for free. Some well-known datasets include the IGBP

DISCover product of the United States Geological Survey (<https://daac.ornl.gov/>, accessed on 5 January 2022) [7], the UMD product of the University of Maryland (<https://idn.ceos.org/>, accessed on 21 January 2022) [8], the GLC product of European Commission Joint Research Centre (<https://forobs.jrc.ec.europa.eu/>, accessed on 21 January 2022) [9], the MODIS LC product of Boston University (<https://lpdaac.usgs.gov/>, accessed on 21 January 2022) [10], and the GLOBCOVER product of the European Space Agency (ESA, <https://www.esa.int/>, accessed on 22 January 2022). Most of them have a spatial resolution of 300 m, 500 m, or 1 km, and the update frequency is 5 to 10 years. Since the early 2010s, with the rapid increase in remote sensing observation platforms, the enhancement in remote sensing cloud computing capabilities, and the improvement in LULC mapping technology, increasing numbers of LULC datasets with a high spatial resolution (10 to 30 m) have been constructed [11]. These LULC datasets include the Globeland30 dataset developed by the National Geomatics Center of China (NGCC) [12], which provides global 30 m resolution data in three issues of 2000, 2010, and 2020, which can be freely obtained by users (<http://www.globallandcover.com/>, accessed on 31 January 2022); the GLC_FCS30 dataset developed by the Aerospace Information Research Institute, Chinese Academy of Sciences (AIR, CAS) [13–15], with an update cycle of 5 years, containing eight 30 m resolution global data from 1985 to 2020, which can be freely obtained (<https://data.casearth.cn/>, accessed on 31 January 2022); the LSV10 dataset developed by the European Space Agency (ESA) [16], with a resolution of 10 m in 2020, which can be freely obtained (<https://esa-worldcover.org/>, accessed on 31 January 2022); and the ESRI 2020 Land Cover 10 m (ESRI10) dataset developed by the Environmental Systems Research Institute (ESRI) [17], with an update cycle of one year and containing five phases of global 10 m resolution data from 2017 to 2021, which can be freely accessed online (<https://www.arcgis.com/>, accessed on 31 January 2022).

The above LULC datasets provide basic data support for scholars worldwide in conducting research in the fields of nature, ecology, environment, and resources [18,19]. However, different LULC datasets are different in terms of the classification system, product accuracy, etc. Therefore, their basic characteristics must be understood before conducting research in a specific study area. Scholars have evaluated the authenticity of different LULC datasets in different regions. Using field survey data as a reference, Heiskanen et al. tested and evaluated the vegetation type data of GLC, MODIS LC, and MODIS VCF datasets in northern Finland. The results showed that the overall accuracy of the first-level class of the three products was high, but their fine-type accuracy was substantially lower [20]. Xu et al. evaluated the accuracy and consistency of CGLS-LC100, ESA-S2-LC20, and FROM-GLC-Africa30 datasets in Africa using measured sample data and statistics from the Food and Agriculture Organization of the United Nations (FAO). The results show that the overall accuracy of the three products was greater than 60%, and the CGLS-LC100 results were most consistent with the FAO statistics, but they found significant differences in the spatial details among the different products [21]. David et al. used the stratified sampling method to evaluate the accuracy of the NLCD dataset in Alaska; the results showed that the overall accuracy of the first- and second-level classes was 83.9% and 76.2%, respectively [22].

Analyzing the consistency of multisource LULC datasets is also important to explore the characteristics and application potential of LULC datasets, which is basic work required to improve and optimize existing datasets [23,24]. In this direction, Hu et al. proposed a basic process of analyzing the consistency of multisource LULC datasets, summarizing specific models of the similarity of the composition type, degree of category confusion, spatial consistency, and reference accuracy [25,26]. Yang et al. applied index methods such as statistical area, distribution pattern, and spatial analysis to compare the application potential of the IGBP DISCover, UMD, GLC, MODIS LC, GLCNMO, CCI-LC, and GlobeLand30 datasets in China [27]. McCallum et al. set up test sites in South America, North America, Europe, Africa, Australia, Russia, and Asia, and compared the accuracy and consistency of GLC, UMD, MODIS LC, and IGBP DISCover datasets in these different regions using classification merging and grid-by-kilometer statistical area methods [28]. Although different

LULC datasets credibly describe the overall pattern of a region, their spatial consistency may widely vary in different study areas, spatial scales, and land types [23–28]. In addition, previous researchers mostly focused on the comparative analysis of multisource LULC datasets with coarse resolution. No comparative study of high-precision LULC datasets with 10 and 30 m resolution emerging since 2017 has yet been conducted.

The Indochina Peninsula is a key region connecting East Asia, South Asia, and Europe, and its geopolitical and economic statuses are important. In the Indochina Peninsula, except for Vietnam and Thailand, the remainder of the countries (Laos, Cambodia, and Myanmar) are the world's least developed countries, as determined by the United Nations. Due to the underdevelopment of the regional economy and society, the Indochina Peninsula generally lacks national-scale LULC datasets developed for each country, and enough human and material resources to conduct ground truth tests on the global LULC datasets. However, the Indochina Peninsula serves as a bridge linking developed regions such as East Asia and Western Europe. Especially in the context of China's Belt and Road initiative, the Indochina Peninsula has become an important area for China's ambitious "six corridors and six channels serving multiple countries and ports" infrastructure planning. Therefore, the latest LULC status in the Indochina Peninsula must be understood, the dynamic history and future changes of LULC in this region need to be analyzed, and the economic, environmental, and climate change effects that may be caused by infrastructure construction must be explored. This has become an important issue of concern to scholars both in and outside the region. In this context, fully using the current internationally renowned LULC datasets to conduct analysis provides convenient source for researchers in the field.

To this end, in this study, we selected the LSV10, GLC_FCS30, ESRI10, and Globeland30 data products, and conducted accuracy and spatial consistency analyses in Myanmar, Vietnam, Thailand, Cambodia, and Laos, which are located on the Indochina Peninsula. To evaluate the degree of agreement and reliability between different datasets, we aimed to achieve three objectives:

- (1) To reveal the law of land use/land cover composition on the Indochina Peninsula;
- (2) To test the spatial consistency of the four LULC datasets on the Indochina Peninsula;
- (3) To evaluate the overall and classification accuracies of the four LULC datasets to provide a basis for data selection in subsequent studies.

2. Study Area and Data

2.1. Study Area

The Indochina Peninsula is located between China and the South Asian subcontinent; it borders China's Guangxi, Yunnan, and Tibet regions to the north; the Bay of Bengal and the Andaman Sea on the northern edge of the Indian Ocean to the west; Malaysia to the south; and the South China Sea on the western edge of the Pacific Ocean to the east. The area ranges from approximately 92.0 to 109.5° E and 5.5 to 28.5° N. The study area included all the territories of Vietnam, Thailand, Laos, Cambodia, and Myanmar on the Indochina Peninsula, having a total area of $2.065 \times 10^6 \text{ km}^2$ (Figure 1).

The Indochina Peninsula has a tropical monsoon climate, experiencing high temperatures throughout the year, with the year divided into two seasons: dry and rainy. June–October is the rainy season, when the southwest monsoon prevails, and precipitation is abundant; November–May is the dry season, with the northeast monsoon prevailing and little rainfall. On the Indochina Peninsula, many mountains, rivers, and valleys are located in the north and south. The main mountain ranges are the Arakan Yoma in the west, a series of mountain ranges extending southward from the Shan Plateau in the middle, and the Truong Son Ra in the east. The main rivers are the Irrawaddy, Salween, Red, Mekong, and Chao Phraya Rivers. The overall terrain of the Indochina Peninsula is high in the north and low in the south. The northern mountains are high and the valleys are deep, being mostly plateau canyons and hills. The southern river valleys are open and the terrain is flat, with many estuary deltas and alluvial plains.

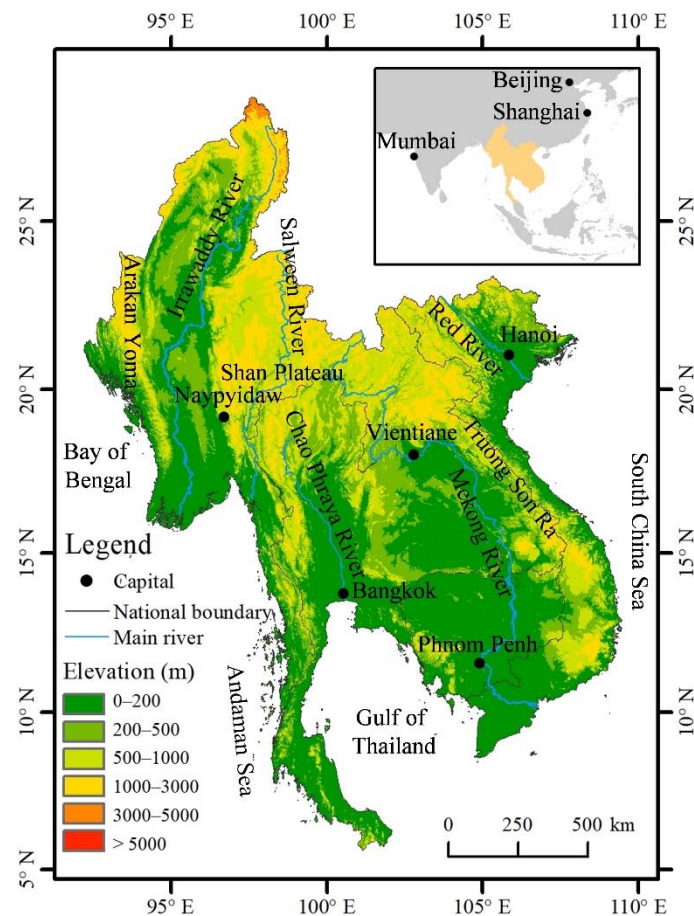


Figure 1. The location and topography of the study area.

In 2018, the total population of the five countries was 0.242×10^9 , accounting for 3.10% of the world's total population; the regional GDP was USD 0.864×10^{12} , accounting for 1.01% of the global economic total. The contribution of the added value of the primary, secondary, and tertiary industries to the regional GDP was roughly 12.1%, 52.0%, and 35.9%, respectively. Compared with the GDP proportion per population of the world, that of the Indochina Peninsula lags considerably behind. Compared with the industrial structure of developed countries, Indochina Peninsula is still in the initial stage of industrialization, which is reflected in the high contributions of primary and secondary industries to the total GDP. With economic globalization, especially the industrial restructuring and industrial transfer from neighboring China, the economy of the Indochina Peninsula grew rapidly from 2000 to 2020, with an average annual GDP growth rate of 9.32%, which has led to considerable improvements in the economic and social quality of the countries.

2.2. LULC Datasets

International academic institutions have released several global LULC datasets based on satellite remote sensing images. Due to the differences in satellite platforms and sensors, land classification systems, LULC mapping methods, and LULC dataset release times, the LULC datasets differ in terms of the current situation, spatial resolution, and product accuracy. In this study, we selected 4 LULC datasets that were the most current (2020), had the highest resolution (10–30 m), and had high precision (overall accuracy between 74% and 86%) for consistency evaluation: LSV10 [16], GLC_FCS30 [13–15], ESRI10 [17], and Globeland30 [12]. A brief summary of the datasets is provided in Table 1.

Table 1. Brief summary of the 4 LULC datasets.

Dataset	Institution	Remote Sensing Image	Date Range	Classification System	Classification Quantity	Classification Method	Spatial Resolution (m)	Overall Accuracy (%)	Source Website
LSV10	ESA	Sentinel-2 multispectral image, Sentinel-1 SAR image	January–December 2020	LCCS	11	Decision tree	10	74.4	https://esa-worldcover.org/ , accessed on 31 January 2022
GLC_FCS30	AIR, CAS	Landsat image, Sentinel-1 SAR image	2019–2020	GLC_FCS30-2020	30	Random forest	30	81.4	https://data.casearth.cn/ , accessed on 31 January 2022
ESRI10	ESRI	Sentinel-2 image	January–December 2020	/	10	Deep learning	10	85.96	https://www.arcgis.com/ , accessed on 31 January 2022
Globeland30	NGCC	Landsat8-OLI image, GF-1 multispectral image	Vegetation growth season within two years before and after 2020	/	10	POK	30	85.72	http://www.globallandcover.com/ , accessed on 31 January 2022

2.3. Data Preprocessing

Before consistency analysis, we preprocessed the 4 LULC datasets, which included image stitching and cropping, projection transformation, upscaling transformation, and classification system merging.

First, we stitched together 38 LSV10, 16 GLC_FCS30, 10 ESRI10, and 16 Globeland30 datasets. Then, we converted the coordinate system of the stitched data to the WGS_1984_UTM_Zone_47N coordinate system (the central meridian was 99° E and the longitude range was 96–102° E). Then, we used the unified study area boundary data to cut them down and finally obtain 4 study-area LULC datasets.

As these 4 LULC datasets have different resolutions (10 or 30 m), we used the maximum area aggregation method to convert the 10 m resolution LSV10 and ESRI10 data to 30 m to be consistent with GLC_FCS30 and Globeland30.

Because the 4 LULC datasets use different LULC classification systems (Table 2), we needed to reclassify them to the unified LULC classification system. Drawing on previous research experience [25–28], we identified 9 common LULC types (Table 3).

Table 2. Classification system of 4 LULC datasets.

LSV10		GLC_FCS30		ESRI10		Globeland30	
Code	Definition	Code	Definition	Code	Definition	Code	Definition
10	Tree cover	10	Rainfed cropland	82	Closed deciduous needle-leaved forest (fc > 0.4)	180	Wetlands
20	Shrubland	11	Herbaceous cover	91	Open mixed leaf forest (broad-leaved and needle-leaved)	190	Impervious surfaces
30	Grassland	12	Tree or shrub cover (Orchard)	92	Closed mixed leaf forest (broad-leaved and needle-leaved)	200	Bare areas
40	Cropland	20	Irrigated cropland	120	Shrubland	201	Consolidated bare areas
50	Built-up	51	Open evergreen broadleaved forest	121	Evergreen shrubland	202	Unconsolidated bare areas
60	Bare/sparse vegetation	52	Closed evergreen broadleaved forest	122	Deciduous shrubland	210	Water body
70	Snow and ice	61	Open deciduous broadleaved forest (0.15 < fc < 0.4)	130	Grassland	220	Permanent ice and snow
80	Permanent water bodies	62	Closed deciduous broadleaved forest (fc > 0.4)	140	Lichens and mosses	250	Filled value
90	Herbaceous wetland	71	Open evergreen needle-leaved forest (0.15 < fc < 0.4)	150	Sparse vegetation (fc < 0.15)	9	Snow/I\ice
95	Mangroves	72	Closed evergreen needle-leaved forest (fc > 0.4)	152	Sparse shrubland (fc < 0.15)	10	Clouds
100	Moss and lichen	81	Open deciduous needle-leaved forest (0.15 < fc < 0.4)	153	Sparse herbaceous (fc < 0.15)	100	Permanent snow and ice

Note: fc = forest cover.

Table 3. Correspondence between old and new LULC systems.

New System	LSV10	GLC_FCS30	ESRI10	Globeland30
1 Cropland	40	10, 20	5	10
2 Forest	10	51–92	2	20
3 Grassland	30, 100	11, 130–150, 153	3	30
4 Shrubland	20, 95	12, 120–122, 152	6	40
5 Wetland	90	180	4	50
6 Water area	80	210	1	60
7 Built-up land	50	190	7	80
8 Bare land	60	200–202	8	90
9 Snow and ice	70	220	9	100

After the above operations, the LULC data obtained in the four study areas had a small number of missing pixel values. We removed these pixels without including them in the analysis. This avoided introducing new uncertainties.

3. Study Method

We analyzed the multisource LULC dataset consistency evaluation process and index model proposed by Hu et al. [26]. We focused on evaluating the LULC datasets from four aspects: the similarity of composition type, the degree of category confusion, spatial consistency, and accuracy.

3.1. Similarity of Composition Type

For different LULC datasets, we counted the areas corresponding to LULC types to form multiple area sequences. From this, we calculated the correlation coefficient between the area series values of different LULC datasets to evaluate the similarity of the regional land composition of different LULC datasets [26]. The formula for calculating the correlation coefficient is:

$$R_{XY} = \frac{\sum_{k=1}^9 (X_k - \bar{X})(Y_k - \bar{Y})}{\sqrt{\sum_{k=1}^9 (X_k - \bar{X})^2 \sum_{k=1}^9 (Y_k - \bar{Y})^2}} \quad (1)$$

where R_{XY} is the correlation coefficient of the land cover area series of the two LULC datasets (X and Y); k is the land cover type; X_k and Y_k are the areas of k in LULC datasets X and Y , respectively, in km^2 ; and $\bar{X}$ and $\bar{Y}$ are the average of the area of 9 land cover types in LULC datasets X and Y , respectively, in km^2 .

3.2. Degree of Category Confusion

From the correlation analysis results obtained based on the area series, we evaluated the similarity of the regional land composition of different LULC datasets, but using this method, we could not describe the spatial confusion of each land type. As such, we further adopted spatial overlapping, judging the pixel confusion state, and counting the number of confused pixels to analyze the degree of category confusion.

Our specific method was as follows: first, we obtained the corresponding relationship of any two LULC datasets at the pixel scale by the spatial stacking method; second, according to the land type one-by-one, we counted the same area and the changed area per the land type attribute. We compared them with the total study area to obtain the degree of confusion for a particular LULC type. Before and after stacking, the pixels with the same land type were pure pixels; the pixels with changed land types were confused pixels. For each land type, the more pure pixels there were and the larger the area ratio, the easier it was to accurately extract this type of land; otherwise, this type of land was easily confused with other land types. The formulas for calculating the purity degree and confusion degree are as follows:

$$DP_{AB}(k) = \frac{S(kk)}{S(k)} \quad (2)$$

$$DC_{AB}(kp) = \frac{S(kp)}{S(k)} \quad (3)$$

where k and p are the LULC types; $DP_{AB}(k)$ and $DC_{AB}(kp)$ are the purity degree of k and the confusion degree of k and p in the combination of the A/B LULC datasets, respectively; $S(k)$ is the area of k in A , in km^2 ; $S(kk)$ is the area of k identified by A and B , in km^2 ; and $S(kp)$ is the area identified as k in A and as p in B .

3.3. Spatial Consistency

Although the category confusion analysis can quantitatively describe the degree of confusion of each category between different datasets, the results are statistical, not

intuitive, and direct. To visualize the degree of agreement between different LULC datasets, a spatially consistent mapping method is required.

Our specific method was as follows: first, we used the spatial stacking method to obtain the spatial correspondence of the LULC datasets at the pixel scale. Second, we determined whether the land cover types that they indicated were the same pixel-by-pixel. Third, according to the number of LULC datasets judged to be the same type, we sorted them to form a thematic map. In this study, we divided the spatial consistency into 4 levels: full agreement (spatial consistency of 100%, that is, the pixel was recognized as the same land type by the 4 datasets), high agreement (spatial consistency of 75%, that is, the pixel was recognized as the same land type by the 4 datasets), low agreement (spatial consistency of 50%, that is, only two datasets recognized the pixel as the same land type), and no agreement (spatial consistency of 0% to 25%, that is, the recognized types of the pixel of the 4 datasets were completely different). The formula for calculating spatial consistency is:

$$N_c(k) = \frac{\sum_{L=1}^4 (c_L == k)}{4} \quad (4)$$

where $N_c(k)$ is the spatial consistency of LULC type k on pixel c ; c_L is the LULC type of c recognized by LULC dataset L .

3.4. Accuracy Analysis

By checking the consistency of all the pixels of the LULC datasets through analysis, we revealed their differences in terms of number and spatial distribution. However, the findings did not indicate which LULC dataset had higher overall accuracy, not which land types in the dataset had higher accuracy [29–31]. Therefore, we first referred to the high-resolution images on Google Earth and, based on constructing a $0.5^\circ \times 0.5^\circ$ regular grid and central sampling points (556 in total, as shown in Figure 2), we established the LULC validation sample dataset for 2020 in the study area by means of visual interpretation. Then, we calculated the accuracy of the four sets of LULC data with the above validation sample data as the reference.

To evaluate the accuracy of the LULC datasets, we relied on constructing an error matrix. According to the error matrix, we further calculated the user, producer, and overall accuracies of the LULC datasets [25,26]. The formulas are as follows:

$$UA(k) = \frac{S(kk)}{\sum_{p=1}^9 S(kp)} \quad (5)$$

$$PA(k) = \frac{S(kk)}{\sum_{p=1}^9 S(pk)} \quad (6)$$

$$OA = \frac{\sum_{p=1}^9 S(pp)}{\sum_{k=1}^9 \sum_{p=1}^9 S(kp)} \quad (7)$$

where $UA(k)$ and $PA(k)$ are the user and producer accuracies of LULC k in the LULC dataset, respectively; OA is the overall accuracy of the LULC dataset; $S(kk)$ is the area of correctly classified k , in km^2 ; and $S(kp)$ is the area of k that is wrongly classified into the land cover type p , in km^2 .

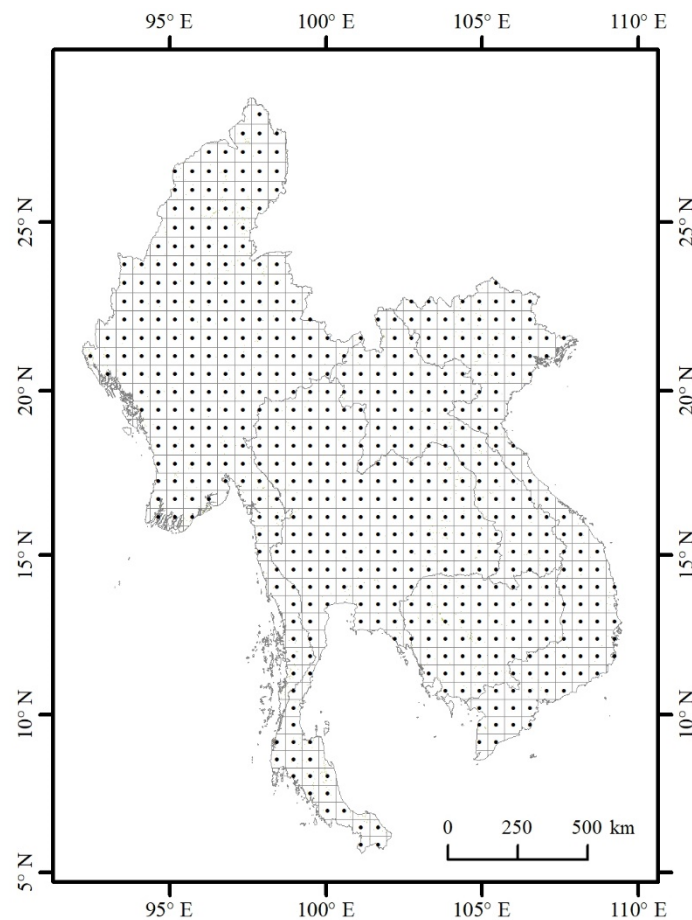


Figure 2. Distribution of LULC samples.

4. Result Analysis

4.1. Regional Land Composition

From the four LULC datasets, the LULC composition of the Indochina Peninsula is shown in Figure 3. In general, different LULC products provided consistent descriptions of the land composition of the Indochina Peninsula: mainly dominated by forest, followed by cropland, shrubland and grassland, then built-up land and water area; the areas of wetland, bare land, and snow and ice were small.

In the four LULC datasets, forest is the land type with the largest area (44.3–63.6%), followed by cropland. The proportion of cropland in the LSV10 and ESRI10 datasets is smaller (21.7–22.4%); the proportion of cropland in the GLC_FCS30 and Globeland30 datasets is larger (29.2–33.6%). In the GLC_FCS30 and Globeland30 datasets, the proportions of grassland are similar (about 5.3%); the proportion of grassland in the LSV10 dataset exceeds 9.3%; and the proportion of grassland in the ESRI10 dataset is only 0.3%. The area of shrubland differed the most among the four LULC datasets (coefficient of variation is 104.4%). The shrubland areas in the GLC_FCS30 and ESRI10 datasets account for a similar proportion (about 17%), while they only account for 1% in the LSV10 dataset, and the least in the Globeland30 dataset (0.7%).

By conducting correlation analysis on the land area series of the four LULC products (Table 4), we found that the correlation coefficients between different LULC products are all above 0.9, which is a high correlation. The correlation between GLC_FCS30 and ESRI10 is the highest, at 0.972; the correlation between LSV10 and Globeland30 is next, at 0.969; the correlations between GLC_FCS30 and Globeland30, LSV10, and ESRI10, and ESRI10 and Globeland30 are in the middle, ranging from 0.929 to 0.943; the correlation between LSV10 and GLC_FCS30 is the lowest, but also above 0.9.

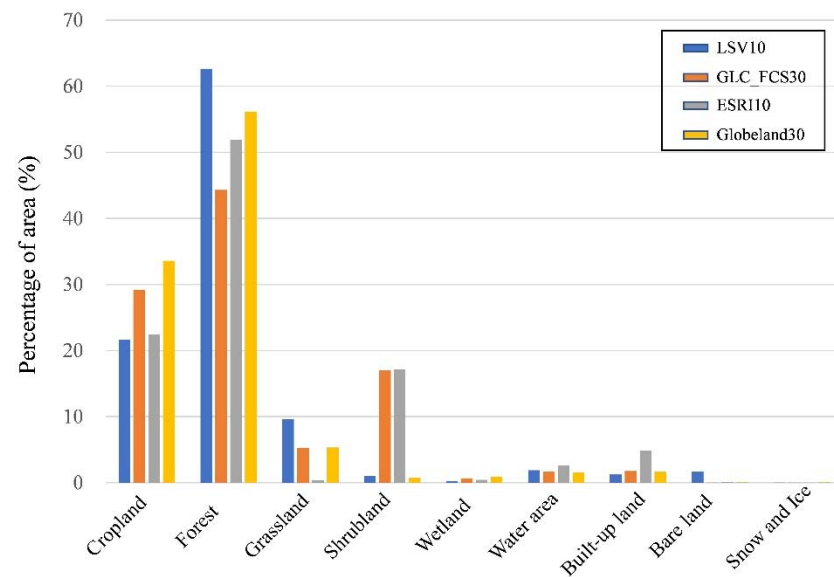


Figure 3. LULC composition of the study area.

Table 4. Correlation between different LULC datasets.

Dataset	LSV10	GLC_FCS30	ESRI10	Globeland30
LSV10	1.000	0.905	0.931	0.969
GLC_FCS30	0.905	1.000	0.972	0.943
ESRI10	0.931	0.972	1.000	0.929
Globeland30	0.969	0.943	0.929	1.000

4.2. Confusion of Land Type

We analyzed the degree of land type confusion for the different LULC datasets, and the results are shown in Figure 4a. When the LULC type on the abscissa is the same as the land type on the ordinate, the pixels are pure; otherwise, the pixel type is confused.

The correct pixel identification of cropland, forest, water area, and built-up land is high; the degree of confusion with other land types is low. The proportion of these land types that were consistent in the pairwise comparison of the multisource LULC datasets is mostly above 70%. Among them, the LSV10/GLC_FCS30 (Figure 4a) and the ESRI10/Globeland30 combinations (Figure 4f) have the highest accuracy of cropland, reaching over 89%. The GLC_FCS30/Globeland30 (Figure 4e) and the ESRI10/Globeland30 combinations (Figure 4f) have the highest forest identification accuracy, reaching over 86%. The LSV10/ESRI10 combination (Figure 4b) and the GLC_FCS30/ESRI10 combination (Figure 4d) have the highest purity degree of water area, reaching over 95%. The LSV10/ESRI10 (Figure 4b) and the GLC_FCS30/ESRI10 combinations (Figure 4d) have the highest identification degree of built-up land, reaching over 97%. Conversely, the pixel identification of grassland, shrubland, and wetland is low; the degree of confusion with other land types is high. The proportion of these land types that were consistent in the pairwise comparison of the multisource LULC datasets is mostly less than 30% and even less than 10%. Among them, the LSV10/Globeland30 combination (Figure 4c) has the highest purity degree of grassland, at 16%; the LSV10/ESRI10 (Figure 4b) and the GLC_FCS30/ESRI10 combinations (Figure 4d) have the lowest purity degree of grassland, at only 1%. The LSV10/ESRI10 combination (Figure 4b) has the highest identification degree of shrubland, at 22%; the GLC_FCS30/Globeland30 combination (Figure 4e) has the lowest, less than 1%. The GLC_FCS30/Globeland30 combination (Figure 4e) has the highest-accuracy identification of wetland, at 66%; the LSV10/GLC_FCS30 combination (Figure 4a) has the lowest, at 13%.

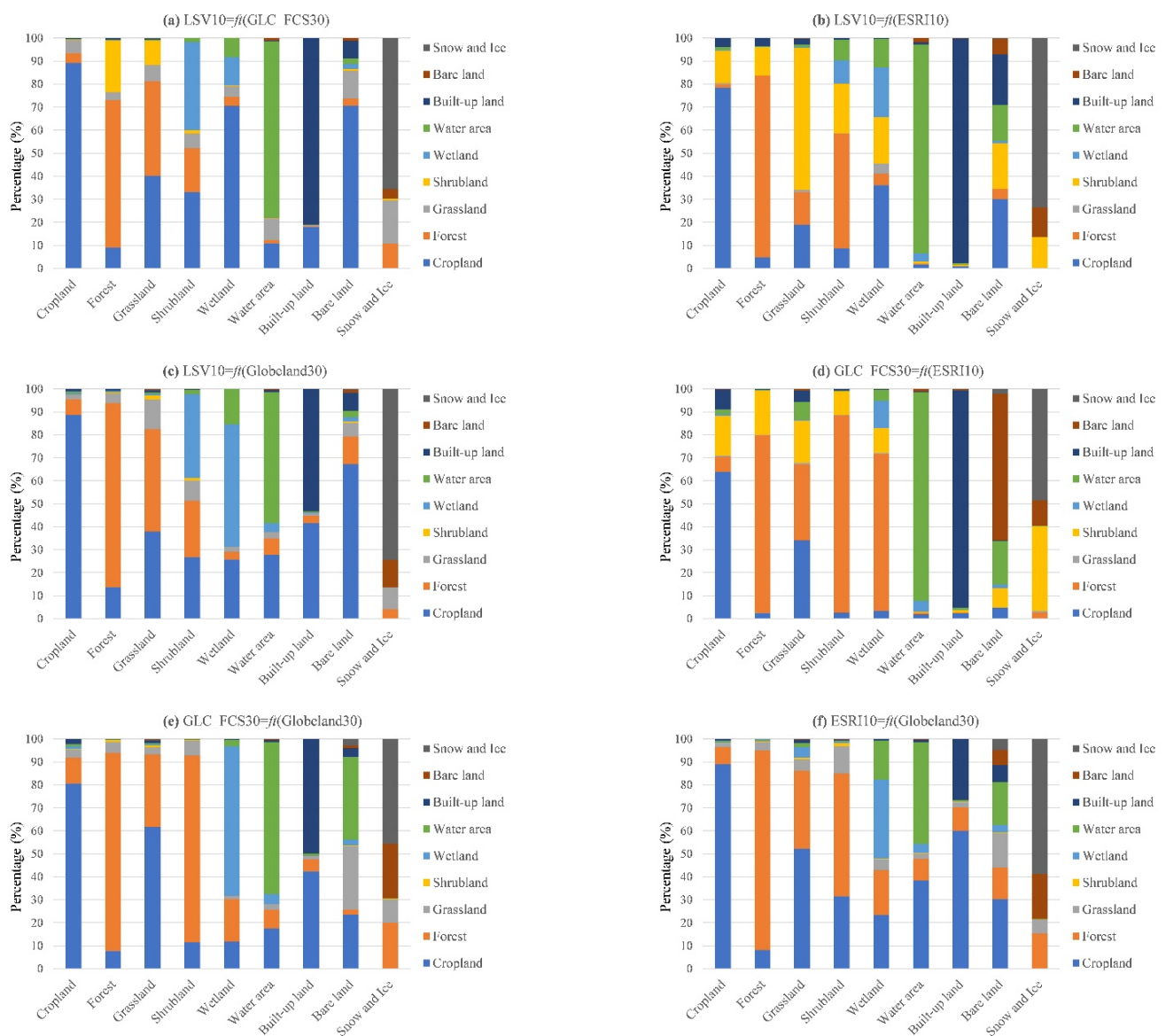


Figure 4. Degree of land type category confusion of different LULC datasets.

In conclusion, from the analysis of the confusion of land types, we found that the four LULC datasets have a high pixel identification accuracy and good consistency of cropland, forest, water area, and built-up land. This partly reflects the high accuracy of different LULC datasets for these land types. The four LULC datasets have low identification accuracy and poor consistency for grassland, shrubland, and wetland, reflecting that the accuracy of different LULC datasets in these land types is inconsistent, so further verification is needed.

4.3. Spatial Consistency

To analyze the spatial differentiation characteristics of the consistency of land cover type identification, we selected four land types (cropland, forest, grassland, and shrubland), which account for the largest proportion of the study area, and conducted a pixel-by-pixel comparative analysis. The results (Figure 5) show that the four LULC datasets have the highest spatial consistency for forest and cropland, and most of the pixels are identified as the same land type by three to four LULC datasets. The spatial consistency of grassland identification is lower, and that of shrubland is lowest: only a few pixels are identified as the same land type by two LULC datasets, and very few pixels are identified as the same land type by three to four LULC datasets.

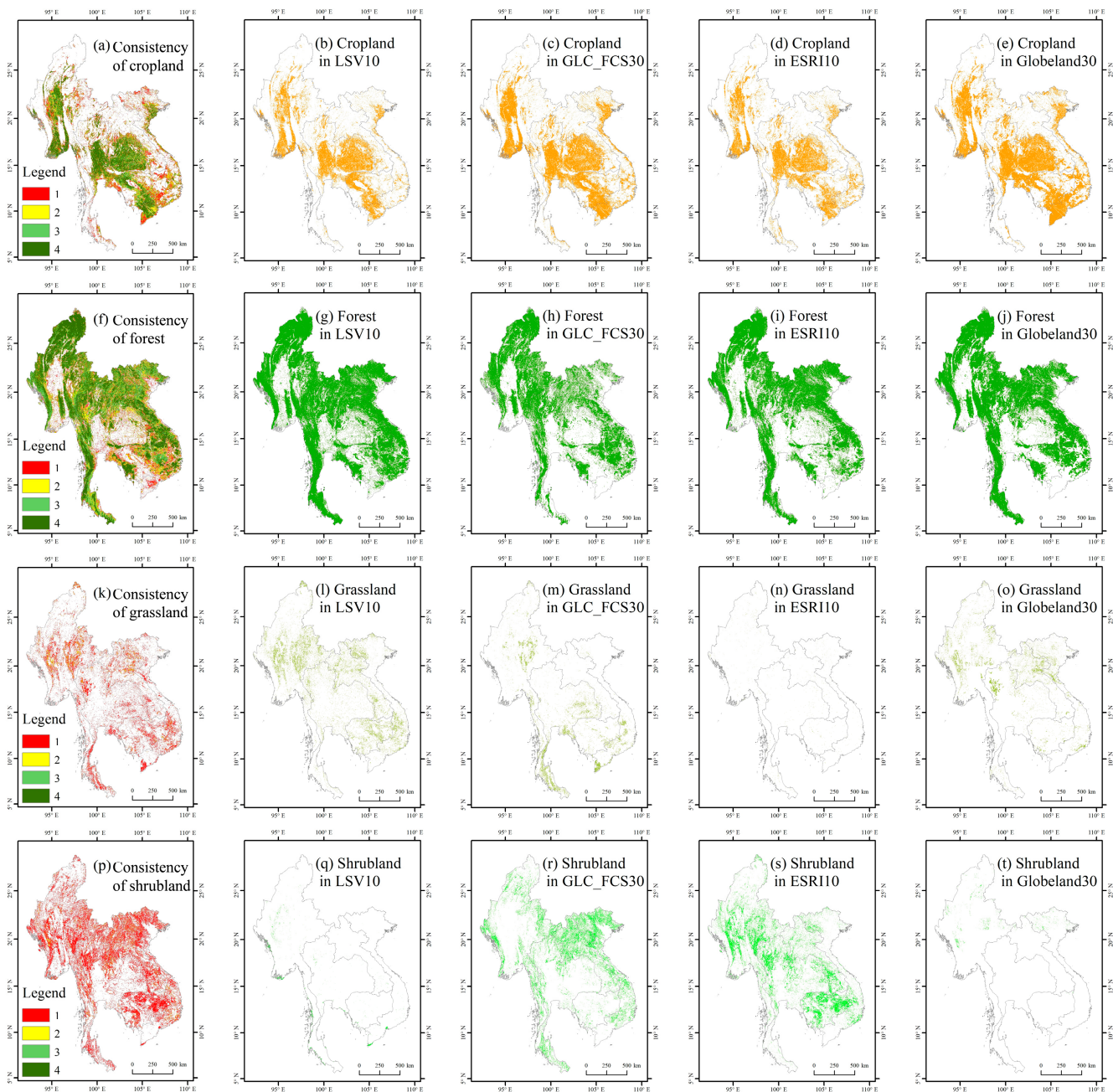


Figure 5. Spatial consistency of different LULC types.

Cropland (Figure 5a–e) is an important land type on the Indochina Peninsula and is widely distributed throughout the region. For the four LULC datasets, the cropland spatial distribution pattern is the same. In the GLC_FCS30 and Globeland30 datasets, the cropland pixels are denser and the cropland area is larger. Conversely, in the LSV10 and ESR10 datasets, the cropland area is smaller. To analyze the consistency of cropland identification, the proportion of cropland identified by three to four LULC datasets is 56.27%; that identified by one to two LULC datasets is 43.73%. The spatial distribution of cropland is the most concentrated in the northwest (the middle and lower reaches of the Irrawaddy River), the middle (the middle and lower reaches of the Chao Phraya and Mekong Rivers), and the eastern and southern coastal areas. In these areas, we found that the spatial consistency of cropland is the highest, and three to four LULC datasets generally identify these pixels as cropland. The distribution of cropland is fragmented in the southeast (south of Truong Son Ra), the northeast (the area of the Red River basin,

excluding the delta), and the south (the southernmost peninsula of Vietnam). In these areas, the spatial consistency of multisource LULC datasets is low, and one to two LULC datasets generally identify these pixels as cropland.

Forest (Figure 5f,g) is the most important land type on the Indochina Peninsula, having the largest area and the widest spatial distribution. In the LSV10, ESRI10, and Globeland30 datasets, the forest spatial distribution pattern and quantity are similar. In the GLC_FCS30 dataset (Figure 5j), the forest pixels are sparse, and their number is low. To analyze the consistency of forest identification, we found that the proportion of forest identified by three to four LULC datasets is 71.60% and that identified by one to two LULC datasets is 28.40%. From the perspective of spatial distribution, forest is most concentrated in the northwest (the upper reaches of the Irrawaddy River, the Arakan Yoma, and the Shan Plateau), the vast area in the north-central region (the upper reaches of the Chao Phraya River and the Mekong River), the east (Truong Son Ra), the southwest (Mountains in southern Myanmar). In these areas, the spatial consistency of multisource LULC datasets is the highest, and three to four LULC datasets generally identify these pixels as forest. In the southeast (Mekong Delta and southern Truong Son Ra), the forest distribution is more fragmented. In these areas, the spatial consistency of the four LULC datasets is low, and one to two LULC datasets generally identify these pixels as forest.

The spatial consistency of the identification of grassland (Figure 5k–o) is poor in the four LULC datasets. The spatial distribution and quantity of grassland identified by the datasets widely vary. The grassland area in the LSV10 dataset is the largest, and most of the pixels are located in the northeastern half of the study area. The grassland areas in the GLC_FCS30 and Globeland30 datasets are similar, but most of the pixels in the GLC_FCS30 dataset are located in the southwest half of the region and most of the pixels in the Globeland30 dataset are located in the northern half. The grassland area in the LSV10 dataset is the smallest, being much less than in the other three. As shown in Figure 5l, the grassland area is barely visible. From analyzing the consistency of grassland identification, we found that 88.81% of the pixels are identified as grassland by only one LULC dataset, 10.69% of the pixels are identified as grassland by two LULC datasets, and only 0.49% of the pixels are identified as grassland by three to four LULC datasets.

The spatial consistency of shrubland in the Indochina Peninsula (Figure 5p–t) was the worst in the multisource LULC datasets. The amount of shrubland identified in the datasets widely varies, as do the spatial distribution patterns. The GLC_FCS30 and ESRI10 datasets identify the largest shrubland area, reporting similar values. The pixels in the GLC_FCS30 dataset are the densest in the northeast area, followed by the southwest area, and the middle area is sparse. Conversely, the shrubland pixels in the ESRI10 dataset are less distributed in the northeast and southwest regions, and more distributed in the middle. The shrubland area in the LSV10 and Globeland30 datasets is small: Figure 5q,t show that the shrubland distribution is sporadic. From our analysis of the consistency of shrubland identification, we found that 92.85% of the pixels are correctly identified as shrubland by only one LULC dataset, 6.42% of the pixels are identified as shrubland by two LULC datasets, and only 0.06% of the pixels are identified as shrubland by three to four LULC datasets. From a spatial point of view, the shrubland and forest distribution patterns are highly consistent across the datasets, and the grassland and shrubland patterns are generally consistent. Shrubby and grassland are mainly distributed in the low mountain and hilly areas, and distribution relationship with forest and cropland shows a mosaic pattern. Due to the complexity of the spatial distribution, the fragmentation of the spatial form, and the spectral characteristics of the same object with different spectra, we expected the spatial consistency of the above two land types to be low.

We analyzed the overall consistency of the identification of various types of land among the datasets, and the results are shown in Figure 6. In the Shan Plateau, Arakan Yoma, Truong Son Ra, the upper reaches of the three rivers (Mekong, Chao Phraya, and Irrawaddy Rivers), where forest is concentrated, and in the middle and lower reaches of the three rivers, and the eastern and southern coastal areas, where cropland is concentrated,

the spatial consistency of the multisource LULC datasets is the highest. Most datasets reach full agreement on all pixels. In southern Truong Son Ra, the lower Mekong River basin, and the Mekong Delta, where cropland, forest, grassland, and shrubland are mixed, the spatial consistency of the multisource LULC datasets is the least accurate, and with low or no agreement on many pixels.

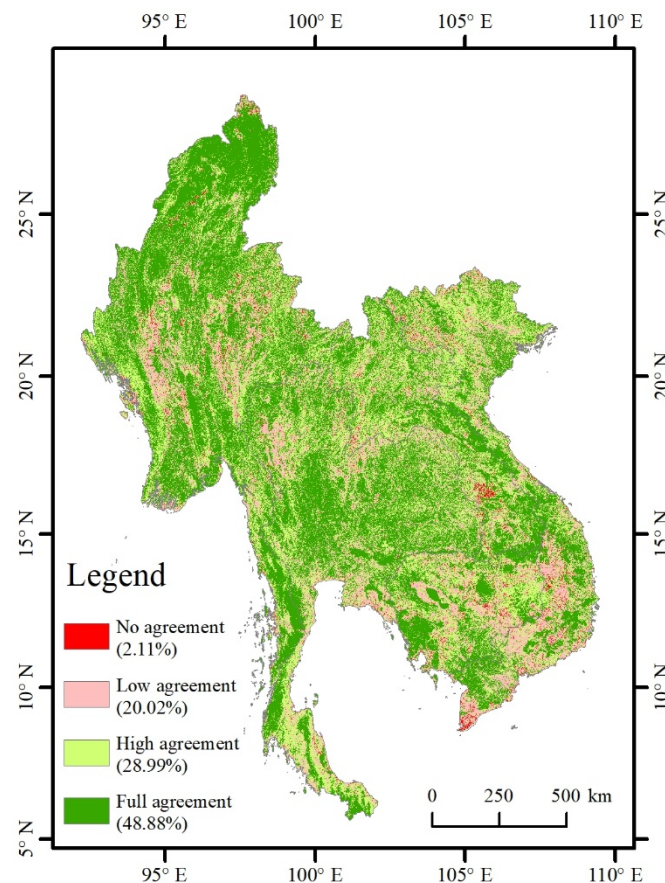


Figure 6. Overall pixel identification spatial consistency of multisource LULC datasets.

We also calculated additional statistics on various types of land. The results show that the area of full agreement on land types of the four LULC datasets accounts for 48.88% of the total area of the Indochina Peninsula; the area with high agreement accounts for 28.99% of the region; the area with low agreement accounts for 20.02% of the region; and the area with no agreement accounts for 2.11% of the region. Considering a reliability of 75% (that is, of the four LULC datasets, at least three identify pixels as the same land type), 77.87% of the land type information in the Indochina Peninsula is reliable; the other 22.13% of the land type information is uncertain.

4.4. Data Accuracy

On the Indochina Peninsula (Table 5), the overall accuracy of the four LULC datasets is between 72% and 83%, as follows: LSV10 (83.25%) > ESRI10 (77.27%) > Globeland30 (73.88%) > GLC_FCS30 (72.27%).

For the LSV10 dataset, the accuracy of cropland, forest, water area, and built-up land identification is highest. Their user and producer accuracies are all above 86%. The accuracy of grassland identification is at a medium level, but the dataset considerably overestimated the grassland area. The identification accuracy is the worst for shrubland, wetland, and bare land.

Table 5. Validation accuracy of the multisource LULC dataset (%).

Land Type	LSV10		GLC_FCS30		ESRI10		Globeland30	
	UA	PA	UA	PA	UA	PA	UA	PA
Cropland	92.72	89.41	88.26	65.01	82.30	86.74	90.14	72.56
Forest	96.76	86.81	88.06	82.83	97.84	84.99	90.67	77.80
Grassland	90.36	53.17	17.01	23.18	26.78	65.12	52.66	42.39
Shrubland	36.25	96.75	41.62	43.20	29.80	31.68	14.90	94.98
Wetland	23.89	65.81	19.67	86.60	31.62	94.41	26.23	41.79
Water area	90.75	90.03	92.63	92.88	98.39	70.92	81.23	88.34
Built-up land	90.24	94.69	79.02	91.42	70.34	94.65	77.76	94.92
Bare land	47.10	57.96	11.78	80.26	56.76	61.64	44.02	57.72
OA	83.25		72.27		77.27		73.88	

For the GLC_FCS30 dataset, the identification accuracy of forest, water area, and built-up land was highest. Their user and producer accuracies are all above 79%. The identification accuracy of cropland is in the middle, but the dataset substantially overestimated the cropland area. Wetland, grassland, shrubland, and bare land identification accuracies are the lowest.

For the ESRI10 dataset, the cropland and forest identification accuracies are highest, and their user and producer accuracies are above 82%. The identification accuracy of water area, built-up land, and bare land is in the middle, but water area is considerably overestimated, built-up land is substantially underestimated, and bare land is underestimated. The identification accuracies of grassland, shrubland, and wetland are the lowest.

For the Globeland30 dataset, the cropland, forest, built-up land, and water area identification accuracies are the highest. The user and producer accuracies are above 72% in general. The dataset remarkably overestimates cropland and forest and notably underestimates built-up land. The identification accuracy of bare land is in the middle, whereas those of grassland, shrubland, and wetland are the lowest.

In conclusion, the LSV10 dataset has the highest overall accuracy. In four LULC datasets, the accuracy of cropland, forest, water area, and built-up land is generally high. For grassland, shrubland, and wetland, the accuracy of the four datasets is not ideal.

5. Discussion

In this study, we found that the overall accuracy of the four LULC datasets in the Indochina Peninsula is ranked from high to low as LSV10 (83.25%) > ESRI10 (77.27%) > Globeland30 (73.88%) > GLC_FCS30 (72.27%). These are $\pm 10\%$ deviations from the accuracy claimed by the original authors (74.4%, 85.96%, 85.72%, and 81.4%, respectively) [12–17]. As the assessment of the accuracy by LULC dataset makers is based on the global validation sample base, our accuracy assessment was limited to the Indochina Peninsula, which is a region with specific LULC types. Moreover, the Indochina Peninsula is in the tropics, with lush vegetation and more complex surface cover. Therefore, their results would differ from those we obtained in this study. This also shows that quantitatively analyzing the actual accuracy of each LULC dataset is necessary before conducting land science research in a specific area.

Our findings show that the four LULC datasets have high identification accuracy of cropland, forest, water area, and built-up land. LSV10 has the highest cropland identification accuracy, whereas the difference between user accuracy and producer accuracy is less than 3%, indicating that the dataset is suitable for research and applications based on cropland, such as grain production potential and cultivated land protection policy. LSV10 and ESRI10 have the highest forest identification accuracies, both exceeding 95%, and so are suitable for forest-based research and applications, such as wildlife conservation, forest resource assessment, and forest ecological services research. GLC_FCS30 has the highest water area identification accuracy, and so is suitable for natural hydrology-related studies, such as fishery production evaluation, hydropower generation evaluation, and water area

ecological protection. LSV10 has the highest built-up land identification accuracy, and so is suitable for research and applications based on built-up land, such as urban heat island, urban dynamic expansion, and urban and rural development planning research.

Our findings show that the identification accuracy of the four datasets for grassland, shrubland, and wetland is low, so they are unsuitable for scientific research and land management planning of these land types. The classification error of these land types is an important factor that reduces the overall accuracy of LULC datasets. On the one hand, the spectral characteristics and texture characteristics of these land types are similar, and their spatial distributions are fragmented. On the other hand, we found differences in the definition of grassland-shrubland and wetland-grassland in the different datasets, which could have led to confusion in the construction of the training samples and final mapping results of the LULC datasets [32]. Therefore, grassland-shrubland, wetland-grassland, and wetland-water area must be more scientifically defined, and more accurate training samples must be built to provide an appropriate foundation for improving the classification accuracy of grassland, shrubland, and wetland in the future. In addition, the classification accuracies of grassland, shrubland, and wetland must be improved to enable the development of process monitoring methods based on time series image data by integrating laser tree height data and radar water retrieval data [33].

We evaluated the multisource LULC datasets from two dimensions: consistency and accuracy. Consistency analysis does not introduce validation samples: LULC datasets were used as a reference to analyze the similarity of land composition, degree of category confusion, and spatial consistency. The advantage of this method is that all pixels of the datasets can be included in the analysis to reveal the difference. The disadvantage is that the results are relative and cannot be absolutely evaluated [34]. The accuracy test method introduces verification samples as close to reality as possible, and calculates the accuracy parameters of each dataset. The advantage is that it can evaluate the overall accuracy of each LULC dataset and the classification accuracy of each land class. The disadvantages are that the representativeness of sample data is limited [35], and that the validation of samples produces new uncertainties [36].

In this study, scaling up and reclassification provided the basis for follow-up evaluation, and were important sources of uncertainty [37]. Scaling has little effect on large areas of forest, large areas of cropland, and water that have higher spatial continuity. However, with mixed land types, scaling leads to changes in fine spatial information and quantity [38,39]. The classification systems of LSV10, ESRI10, and Globeland30 show high correspondence with the merging system, and the subjective factors have little influence on the reclassification process. GLC_FCS30 has 30 fine classes, so the reclassification is easily affected by the knowledge level of the researchers [40,41]. In addition, due to their similar morphology (forest, shrubland, grassland, wetland, etc.), mixed distribution and limited image resolution, sampling methods based on Google Earth and visual interpretation are uncertain, and thereby affect the accuracy verification results [42,43].

6. Conclusions

In this study, we converted four well-known LULC datasets (LSV10, GLC_FCS30, ESRI10, and Globeland30, with a spatial resolution of 10 or 30 m) into a unified and comparable benchmark through scaling up and classification merging. We analyzed the similarity of composition type, degree of LULC confusion, spatial consistency, and validation accuracy in detail. In this study, we are the first to evaluate the spatial consistency and accuracy of LULC datasets in the Indochina Peninsula. Our findings provide a quantitative basis for people in various countries and fields in selecting LULC data and provide a reference method for assessing the accuracy of multisource LULC datasets in other regions.

Our findings show that forest is the main land type on the Indochina Peninsula, followed by cropland, shrubland, grassland, built-up land, and water. The overall accuracy of different LULC datasets is between 72% and 83%. LSV10 has the highest overall accuracy.

The accuracy and consistency of each LULC dataset are higher for cropland, forest, water, and built-up land and lower for grassland, shrubland, and wetland.

Based on the above analysis results, we provided clear suggestions for countries in the Indochina Peninsula countries for relevant researchers when selecting LULC datasets for agriculture, fisheries, forest, and ecological protection, and urban and rural development. We provided some recommendations to improve some land mapping methods that have a low spatial consistency and accuracy. In future research, strengthening field investigation, scientifically defining classification systems, and integrating multisource data such as laser detection and radar inversion will be important for improving the accuracy and availability of LULC datasets.

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

Funding: This research was supported by the National Natural Science Foundation of China (42130508), the Network Security and Information Program of the Chinese Academy of Sciences (CAS-WX2021SF-0106), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA20010202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We express our sincere thanks to the anonymous reviewers for their comments and suggestions that considerably helped to improve the quality of this paper.

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

References

1. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global consequences of land use. *Science* **2005**, *309*, 570–574. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ye, J.; Hu, Y.; Zhen, L.; Wang, H.; Zhang, Y. Analysis on Land-Use Change and Its Driving Mechanism in Xilingol, China, during 2000–2020 Using the Google Earth Engine. *Remote Sens.* **2021**, *13*, 5134. [\[CrossRef\]](#)
3. Liu, J.; Tian, H.; Liu, M.; Zhuang, D.; Melillo, J.M.; Zhang, Z. China's changing landscape during the 1990s: Large-scale land transformations estimated with satellite data. *Geophys. Res. Lett.* **2005**, *32*, 2405. [\[CrossRef\]](#)
4. Zhang, Y.; Luo, Y.; Liu, J.; Zhuang, D. Land use and landscape pattern change in Hetao irrigation district, Inner Mongolia Autonomous Region. *Nongye Gongcheng Xuebao* **2005**, *21*, 61–65.
5. Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tseng-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaadeh, B. Production of global land cover data—GLCNMO. *Int. J. Digit. Earth* **2011**, *4*, 22–49. [\[CrossRef\]](#)
6. Wang, H.; Hu, Y.; Yan, H.; Liang, Y.; Guo, X.; Ye, J. Trade-off among grain production, animal husbandry production, and habitat quality based on future scenario simulations in Xilinhot. *Sci. Total Environ.* **2022**, *817*, 153015. [\[CrossRef\]](#)
7. Loveland, T.; Brown, J.; Ohlen, D.; Reed, B.; Zhu, Z.; Yang, L.; Howard, S.; Hall, F.; Collatz, G.; Meeson, B. ISLSCP II IGBP DISCover and SiB land cover, 1992–1993. *ORNL DAAC* **2009**, *5*, 2257.
8. Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331–1364. [\[CrossRef\]](#)
9. Bartholome, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [\[CrossRef\]](#)
10. Friedl, M.A.; Sulla-Menasse, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [\[CrossRef\]](#)
11. Defourny, P.; Schouten, L.; Bartalev, S.; Bontemps, S.; Cacetia, P.; De Wit, A.; Di Bella, C.; Gérard, B.; Giri, C.; Gond, V. *Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience*; International Center for Remote Sensing of Environment: Amsterdam, The Netherlands, 2009.
12. Defourny, P.; Vancutsem, C.; Bicheron, P.; Brockmann, C.; Nino, F.; Schouten, L.; Leroy, M. GLOBCOVER: A 300 m Global Land Cover Product for 2005 Using ENVISAT MERIS Time Series. In Proceedings of the ISPRS Commission VII Mid-Term Symposium: Remote Sensing: From Pixels to Processes, Enschede, The Netherlands, 8–11 May 2006; ISPRS: Hannover, Germany, 2006; pp. 8–11.

13. Oliphant, A.J.; Thenkabail, P.S.; Teluguntla, P.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *81*, 110–124. [\[CrossRef\]](#)
14. Chen, J.; Ban, Y.; Li, S. Open access to Earth land-cover map. *Nature* **2014**, *514*, 434.
15. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [\[CrossRef\]](#)
16. 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. [\[CrossRef\]](#)
17. Liu, L.; Zhang, X.; Gao, Y.; Chen, X.; Shuai, X.; Mi, J. Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects. *J. Remote Sens.* **2021**, *2021*, 5289697. [\[CrossRef\]](#)
18. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S. ESA WorldCover 10 m 2020 v100. 2021. Available online: https://developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v100 (accessed on 31 January 2022).
19. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global Land Use/Land Cover with Sentinel 2 and Deep Learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 4704–4707.
20. Hu, Q.; Xiang, M.; Chen, D.; Zhou, J.; Wu, W.; Song, Q. Global cropland intensification surpassed expansion between 2000 and 2010: A spatio-temporal analysis based on GlobeLand30. *Sci. Total Environ.* **2020**, *746*, 141035. [\[CrossRef\]](#)
21. Liang, D.; Zuo, Y.; Huang, L.; Zhao, J.; Teng, L.; Yang, F. Evaluation of the consistency of MODIS Land Cover Product (MCD12Q1) based on Chinese 30 m GlobeLand30 datasets: A case study in Anhui Province, China. *ISPRS Int. Geo-Inf.* **2015**, *4*, 2519–2541. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#)
23. Heiskanen, J. Evaluation of global land cover data sets over the tundra-taiga transition zone in northernmost Finland. *Int. J. Remote Sens.* **2008**, *29*, 3727–3751. [\[CrossRef\]](#)
24. Xu, Y.; Yu, L.; Feng, D.; Peng, D.; Li, C.; Huang, X.; Lu, H.; Gong, P. Comparisons of three recent moderate resolution African land cover datasets: CGLS-LC100, ESA-S2-LC20, and FROM-GLC-Africa30. *Int. J. Remote Sens.* **2019**, *40*, 6185–6202. [\[CrossRef\]](#)
25. Selkowitz, D.J.; Stehman, S.V. Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska. *Remote Sens. Environ.* **2011**, *115*, 1401–1407. [\[CrossRef\]](#)
26. Bai, Y.; Feng, M. Data fusion and accuracy evaluation of multi-source global land cover datasets. *Acta Geogr. Sin* **2018**, *73*, 2223–2235.
27. Chen, Y.; Shao, H.; Li, Y. Consistency analysis and accuracy assessment of multi-source land cover products in the Yangtze River Delta. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 142–150.
28. Dai, Z.; Hu, Y.F.; Zhang, Q. Agreement analysis of multi-source land cover products derived from remote sensing in South America. *Remote Sens. Inf.* **2017**, *32*, 137–148.
29. Hu, Y.; Zhang, Q.; Dai, Z.; Huang, M.; Yan, H. Agreement analysis of multi-sensor satellite remote sensing derived land cover products in the Europe Continent. *Geogr. Res.* **2015**, *34*, 1839–1852.
30. 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. [\[CrossRef\]](#)
31. McCallum, I.; Obersteiner, M.; Nilsson, S.; Shvidenko, A. A spatial comparison of four satellite derived 1 km global land cover datasets. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 246–255. [\[CrossRef\]](#)
32. Baig, M.F.; Mustafa, M.R.U.; Baig, I.; Takaijudin, H.B.; Zeshan, M.T. Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia. *Water* **2022**, *14*, 402. [\[CrossRef\]](#)
33. Batar, A.K.; Watanabe, T.; Kumar, A. Assessment of land-use/land-cover change and forest fragmentation in the Garhwal Himalayan Region of India. *Environments* **2017**, *4*, 34. [\[CrossRef\]](#)
34. Nedd, R.; Light, K.; Owens, M.; James, N.; Johnson, E.; Anandhi, A. A synthesis of land use/land cover studies: Definitions, classification systems, meta-studies, challenges and knowledge gaps on a global landscape. *Land* **2021**, *10*, 994. [\[CrossRef\]](#)
35. Szatmári, D.; Kopecká, M.; Feranec, J. Accuracy Assessment of the Building Height Copernicus Data Layer: A Case Study of Bratislava, Slovakia. *Land* **2022**, *11*, 590. [\[CrossRef\]](#)
36. Giuliani, G.; Rodila, D.; Külling, N.; Maggini, R.; Lehmann, A. Downscaling Switzerland Land Use/Land Cover Data Using Nearest Neighbors and an Expert System. *Land* **2022**, *11*, 615. [\[CrossRef\]](#)
37. Zhang, E.; Chen, X.; Wang, L. Consistent discriminant correlation analysis. *Neural Processing Lett.* **2020**, *52*, 891–904. [\[CrossRef\]](#)
38. Wang, Y.; Zhang, J.; Liu, D.; Yang, W.; Zhang, W. Accuracy assessment of GlobeLand30 2010 land cover over China based on geographically and categorically stratified validation sample data. *Remote Sens.* **2018**, *10*, 1213. [\[CrossRef\]](#)
39. Stehman, S.V.; Olofsson, P.; Woodcock, C.E.; Herold, M.; Friedl, M.A. A global land-cover validation data set, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *Int. J. Remote Sens.* **2012**, *33*, 6975–6993. [\[CrossRef\]](#)
40. van der Kwast, J.; Van de Voorde, T.; Canters, F.; Uljee, I.; Van Looy, S.; Engelen, G. Inferring urban land use using the optimised spatial reclassification kernel. *Environ. Model. Softw.* **2011**, *26*, 1279–1288. [\[CrossRef\]](#)

41. Su, B.; Noguchi, N. Discrimination of Land Use Patterns in Remote Sensing Image Data using Minimum Distance Algorithm and Watershed Algorithm. *Eng. Agric. Environ. Food* **2013**, *6*, 48–53. [[CrossRef](#)]
42. Wu, F.; Zhan, J.; Yan, H.; Shi, C.; Huang, J. Land Cover Mapping Based on Multisource Spatial Data Mining Approach for Climate Simulation: A Case Study in the Farming-Pastoral Ecotone of North China. *Adv. Meteorol.* **2013**, *2013*, 520803. [[CrossRef](#)]
43. Jepsen, M.R.; Levin, G. Semantically based reclassification of Danish land-use and land-cover information. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2375–2390. [[CrossRef](#)]