

Accuracy Assessment of the FROM-GLC30 Land Cover Dataset Based on Watershed Sampling Units: A Continental-Scale Study

Zitian Guo ^{1,2}, Chunmei Wang ^{1,2,3,*}, Xin Liu ^{1,2}, Guowei Pang ^{1,2,3}, Mengyang Zhu ^{1,2} and Lihua Yang ⁴

¹ Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, Northwest University, Xi'an 710127, China; jyguozitian@163.com (Z.G.); liuxin11081108@163.com (X.L.); gwpang@nwu.edu.cn (G.P.); myangyang@stumail.nwu.edu.cn (M.Z.)

² College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China

³ Key Laboratory of National Forestry Administration on Ecological Hydrology and Disaster Prevention in Arid Regions, Xi'an University of Technology, Xi'an 710048, China

⁴ Institute of Soil and Water Conservation Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, China; yanglihua17@mailsucas.ac.cn

* Correspondence: cmwang@nwu.edu.cn

Received: 8 September 2020; Accepted: 9 October 2020; Published: 13 October 2020

Abstract: Land cover information plays an essential role in the study of global surface change. Multiple land cover datasets have been produced to meet various application needs. The FROM-GLC30 (Finer Resolution Observation and Monitoring of Global Land Cover) dataset is one of the latest land cover products with a resolution of 30 m, which is a relatively high resolution among global public datasets, and the accuracy of this dataset is of great concern in many related researches. The objective of this study was to calculate the accuracy of the FROM-GLC30 2017 dataset at the continental scale and to explore the spatial variation differences of each land type accuracy in different regions. In this study, the visual interpretation land cover results at 20,936 small watershed sampling units based on high-resolution remote sensing images were used as the reference data covering 65 countries in Asia, Europe, and Africa. The reference data were verified by field survey in typical watersheds. Based on that, the accuracy assessment of the FROM-GLC30 2017 dataset was carried out. The results showed (1) the area proportion of different land cover types in the FROM-GLC30 2017 dataset was generally consistent with that of the reference data. (2) The overall accuracy of the FROM-GLC30 2017 dataset was 72.78%, and was highest in West Asia–Northeast Africa, and lowest in South Asia. (3) Among all the seven land cover types, the accuracy of bareland and forest was relatively higher than that of others, and the accuracy of shrubland was the lowest. The accuracy for each land cover type differed among regions. The results of this work can provide useful information for land cover accuracy assessment researches at a large scale and promote the further practical applications of the open-source land cover datasets.

Keywords: FROM-GLC30; high-resolution remote sensing images; land cover accuracy; continental scale

1. Introduction

Land cover information plays an important role in the study of land surface processes and global environment changes [1–3]. It is widely used in many fields, such as soil erosion [4], urban

change [5], and disasters [6]. With the development of remote sensing technology, the global land cover and land use data at resolutions of 1 km, 500 m, 250 m, and 30 m have been released [7–10]. Due to the different data sources and spatial resolutions, the accuracy of those land cover data products is different. The accuracy of the land cover dataset is of great concern since it can directly affect the modeling results in many surface processes [11,12].

There have been many studies on the accuracy assessment of different land cover datasets, at multiple scales such as continental scale [13], national scale [14], and regional scale [15]. Mainly two types of methods have been utilized. The first one compares different land cover datasets products without identifying which dataset is reference data; the second, which is more commonly used, compares a specific dataset with a more precise reference dataset. It is hard to know the accuracy by using the first methodology since no reference dataset is utilized. In the second method, a reference dataset is usually obtained at sampling units, which can be pixels or small watersheds. The sampling method based on pixels is often used for the data accuracy assessment when the density of sampling units is large, which makes the use of a reference dataset more convenient and efficient [16]. In geography, a watershed is a defined unit, having the similar regional characteristics of climate, hydrology, soil, vegetation, and so on with other watersheds around. In theory, it could be more representative to take the small watershed as the accuracy assessment unit.

Among various land cover data products, the FROM-GLC data (Finer Resolution Observation and Monitoring of Global Land Cover) published in 2013 is a 30 m resolution global land cover dataset based on Landsat TM and ETM+ data [17]. To further optimize the data, solve problems related to the impacts of different seasons, and improve the accuracy of data products, a series of products such as FROM-GLC-seg, FROM-GLC-agg, and FROM-GLC-Hierarchy were released in 2013 and 2014 [18–20]. In 2018, FROM-GLC30 2017, the latest product of the FROM-GLC30 data series, was released. This data product is based on Landsat8 images and uses all-season samples for land cover classification to reduce the impact of seasonal problems. The series products of the FROM-GLC30 dataset have been widely used in many fields, and their accuracy has also been of great concern. Lu [21] evaluated the accuracy of cultivated land of FROM-GLC30 and the other four commonly used land cover datasets and concluded that the overall accuracy of cultivated land of FROM-GLC30 in China reached 76.23%. Chen [22] calculated the cultivated land accuracy of FROM-GLC30 and the other three land cover datasets and concluded that the overall accuracy of the FROM-GLC30 dataset was 77.67 in Shaanxi Province, China. Due to the short time since release of the FROM-GLC30 2017 dataset, the accuracy assessment of this dataset is still insufficient. There is an urgent need to conduct a comprehensive accuracy assessment of the latest FROM-GLC series dataset to improve the understanding of the dataset quality and to serve the further application in surface process researches, especially at a large scale.

Accuracy assessments of high-resolution land cover data at the continental scale are relatively rare. The several existing continental-scale land cover data accuracy researches were mainly carried out based on a reference dataset using pixels as the sampling units [13,16] or by comparing different dataset products without reference data [23,24]. The method based on small watershed sampling unit is rarely used since it is difficult to obtain such a reference dataset. In addition, due to the short time since release of the FROM-GLC30 2017 dataset, fewer researches have reported on the accuracy assessment of this latest 30 m global land cover open-source data product based on a unified reference dataset and sampling method at the continental scale. This is a current limitation for its application.

The aim of this research was to clarify the spatial variation of the FROM-GLC30 2017 dataset accuracy at the continental scale using visually interpreted results from 20,936 small watershed sampling units as reference dataset based on sub-meter Google Earth high-resolution images and field survey as well as other information about accuracy of land cover both in different regions and for each cover type. The results of this study should be helpful for applications related to land cover in different locations of the Pan-Third Pole Area since land cover accuracy varies with location as well as land covers of interest. They could also be important to help improve other large-scale land

cover accuracy assessments and produce research that improves the effective application of land cover datasets in various fields.

2. Data and Methods

2.1. Study Area

The study area was the Pan-Third Pole Area, including the world's third pole Tibetan Plateau, Pamir Plateau, Iranian Plateau, and the Carpathian Mountains, and other mountains, covering 65 countries (Figure 1) [25]. The Pan-Third Pole Area spans parts of Asia, Europe, and Africa, with a total area of about 51.46 million square kilometers. It is one of the most ecologically vulnerable areas worldwide and is sensitive to human activities. The impact of changes in land cover on global climate change and ecological sustainability in the Pan-Third Pole Area has attracted worldwide attention [26,27]. The study area is divided into eight regions, including Central and Eastern Europe, Central Asia, China, Mongolia, Russia, South Asia, Southeast Asia, and West Asia–Northeast Africa (Figure 1).

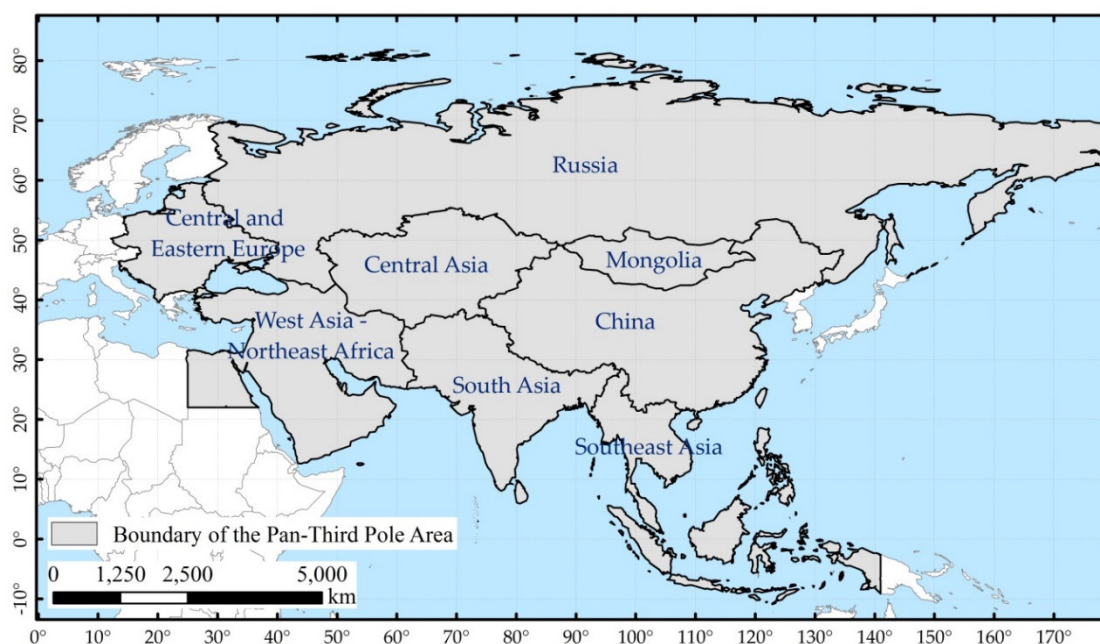


Figure 1. The boundary of the Pan-Third Pole Area.

2.2. Base Data

In this study, two types of land cover data were utilized, the FROM-GLC30 2017 dataset and the reference dataset. The FROM-GLC30 2017 dataset was the data to be assessed. The reference data were acquired by the manual vision interpretation of high-resolution images for 20,936 sampling units. The accuracy assessment of this study was based on small watershed units, and regional statistics were generated in the above eight regions. Since the FROM-GLC30 2017 dataset had no projection information when released, we applied the projection method of the GlobeLand30 from the same series of 30 m land cover data, and unified the projection of both the reference data and the FROM-GLC30 2017 dataset into the UTM projection.

2.2.1. Sampling Units (SUs) and Source of Reference Data

The sampling strategy is similar to that used for the general survey of soil erosion in China [28,29]. Firstly, the study area was divided into zones, with each zone occupying a specified width and length. This width and length were set to be 0.5° latitude and 1° longitude between 60° N and 70° N; 0.5° latitude and 0.75° longitude between 40° N and 60° N; 0.5° latitude and 0.5° longitude

below 40° N. In this way, the ground size of each zone did not differ too much. In each zone, we identified the central 5 km by 5 km extent as a control area, and SUs were selected randomly inside the control area. Small watersheds with an area of 0.2–3 km² were used as SUs in the mountainous area, and the SUs were square in shape with size of 1 km by 1 km in plain areas where watersheds are less clearly defined. A total number of 20,936 SUs were set up in this study.

Research has shown that Google Earth images with high resolution can be an important data source to evaluate the accuracy of land cover products [30,31]. The reference datasets in this study were the results of manual visual interpretation of 20,936 sampling units based on high-resolution remote sensing images from Google Earth. Figure 2 displays the specific processing flow of the reference data in this study. There were four main steps, interpretation of remote sensing images, conversion of projection and data format, scale transformation of raster data, and data quality inspection. In the study area, more than 78% of the sampling units had remote sensing images with a sub-meter resolution. In the other 22% of the sampling units, the spatial resolution of the remote sensing images could also reach the meter level. To maintain the consistency of image time with the FROM-GLC30 2017 dataset, most interpretation of the reference data was based on Google Earth images around 2015. Google Earth images with different years and seasons were also utilized to improve the accuracy of the interpretation results of the time-sensitive land cover types (such as water bodies, glaciers, and permanent snow). The visual interpretation accuracy met requirements of 1:10,000 scale, and the reference data were unified into grid data of 1 m resolution after format conversion. The reference data and the data to be assessed need to be consistent in scale for data accuracy assessment. For this study, the grid size of the reference data was transformed from 1 m to 30 m by scale transformation. The resampling method was Majority. That means the land cover type of each 30 m grid cell after resampling is consistent with that accounting for the largest proportion of the corresponding 900 grid cells with 1 m resolution. After resampling, the attribute values of each 30 m by 30 m grid unit were consistent with those occurring most in the 900 grid units with a size of 1 m by 1 m (Figure 2).

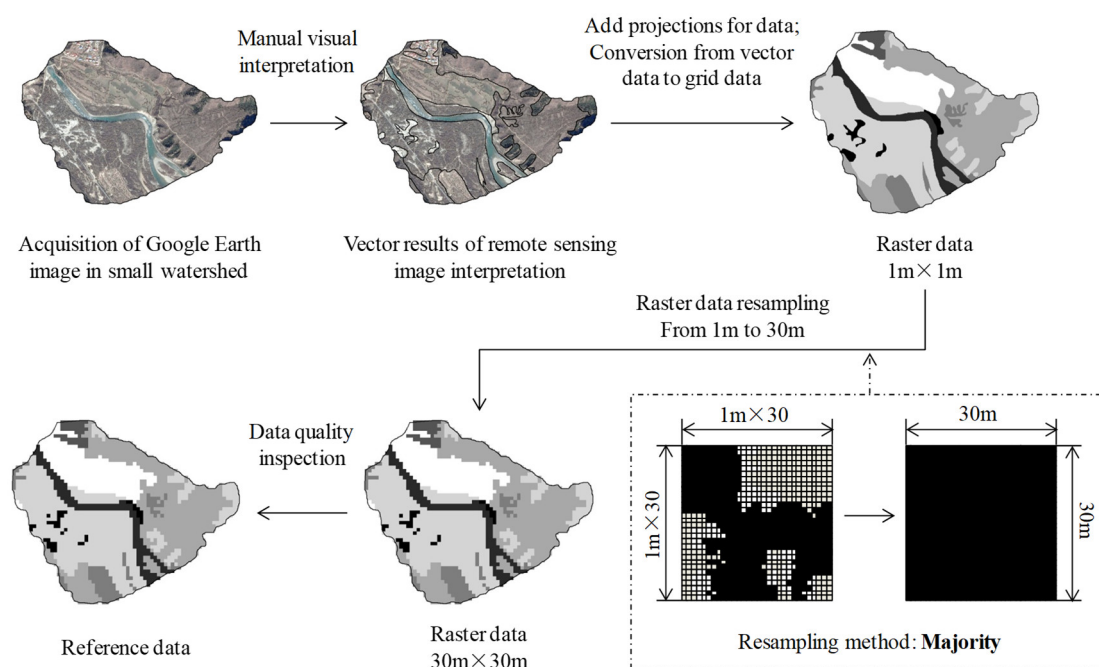


Figure 2. The flow diagram of the reference data processing.

In order to improve the quality of the reference data, four field surveys were organized in Thailand, Pakistan, Tibet, and Xinjiang, China in 2018 and 2019. Based on the field survey of 53 small watershed sampling units (Figure 3), the land cover interpretation results were verified. According to the results of field surveys, some common errors in the interpretation were identified.

For example, in the Tibet Plateau, the most common mistake was to interpret grassland as bareland. In images, many objects look like bareland in color but are actually low cover of grassland at high elevation. After the field survey, the reference data were revised according to the common errors not only at the surveyed SUs but also at SUs with similar conditions in the same regions. That helped improve the reference data.

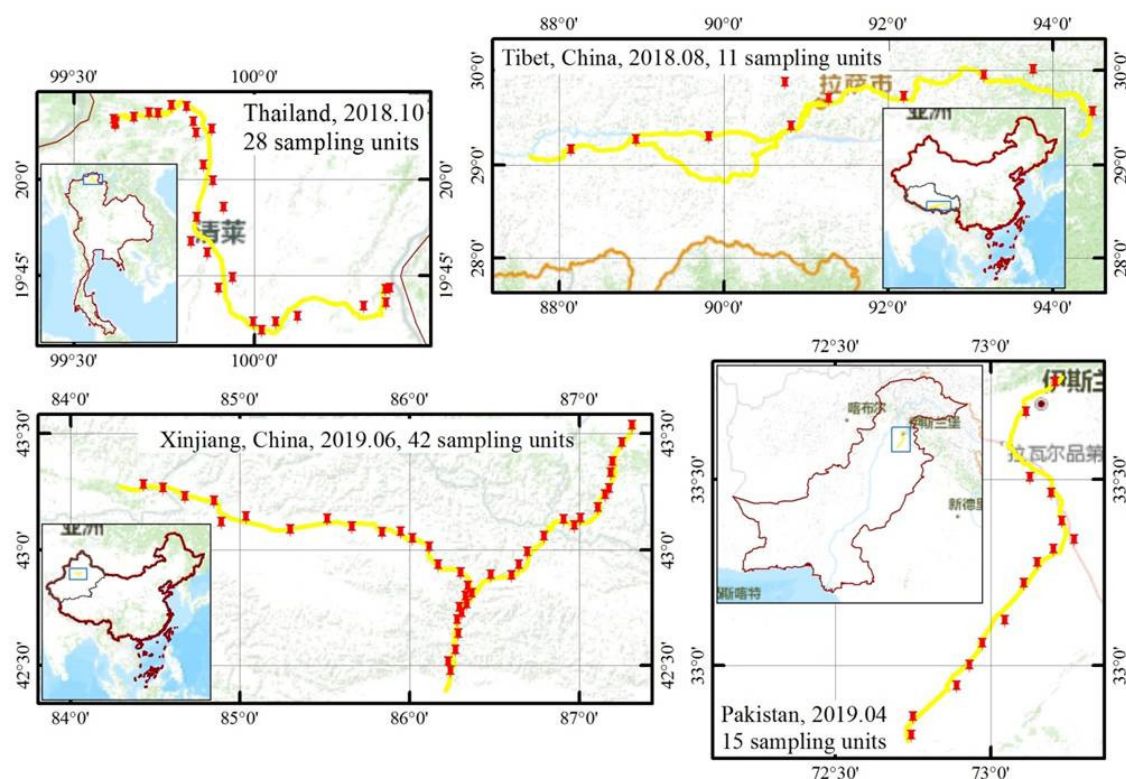


Figure 3. Field survey routes for the validation of reference data.

2.2.2. FROM-GLC30 2017 Dataset

Global Land Cover (GLC) maps can provide important information for agriculture, forestry, and other industries, and are of great significance. Different applications and research requirements have spawned a variety of GLC maps with various resolutions and from different data sources. The FROM-GLC30 dataset, the data to be assessed in this study, is one of the Chinese GLC maps generated by Tsinghua University with a resolution of 30 m. The FROM-GLC30 dataset is one of the commonly used open data land cover datasets at present. It provides a new data source for land cover and land cover researches at different research scales. It has been widely used in the research of climate change, regional development, regional soil erosion, and so forth. At the end of 2018, the FROM-GLC30 2017 dataset was released as the latest data product of the FROM-GLC30 series [32]. The FROM-GLC30 2017 dataset was generated by using a supervised classification method, taking Landsat8 images (mainly in the year 2015) as the primary data source, combining the high-resolution Chinese satellite data, high-resolution SRTM DEM and ASTER DEM elevation data, MOD13Q1 (NDVI), and the global night light data of 500 m spatial resolution published by NASA in 2016.

The reference land cover data was originally derived for the land cover interpretation in a Pan Third-Pole erosion project, rather than within the specific research study presented here. One change was made, as described earlier, to the resolution of the dataset from 1 m to 30 m in order to make sure the accuracy was calculated at the same scale as the FROM-GLC30 2017 dataset. In addition, based on the original project requirements, the land cover classification system of the reference data was more focused on regional soil erosion and its applications. As it differed in details from the land cover classification system of the FROM-GLC30 2017 dataset, the classification system of the two datasets needed to be partially consolidated and some categories mapped. The unified

classification system included cropland, forest, shrubland, grassland, impervious surface, water, and bareland. Table 1 shows the correspondence of the classification system between the FROM-GLC30 2017 dataset and the reference data and describes the definition of the unified seven land cover types. Figure 4a, b displays the comparison of Google Earth images, interpretation results, the reference data, and the FROM-GLC30 2017 dataset of two sampling units under the new classification system.

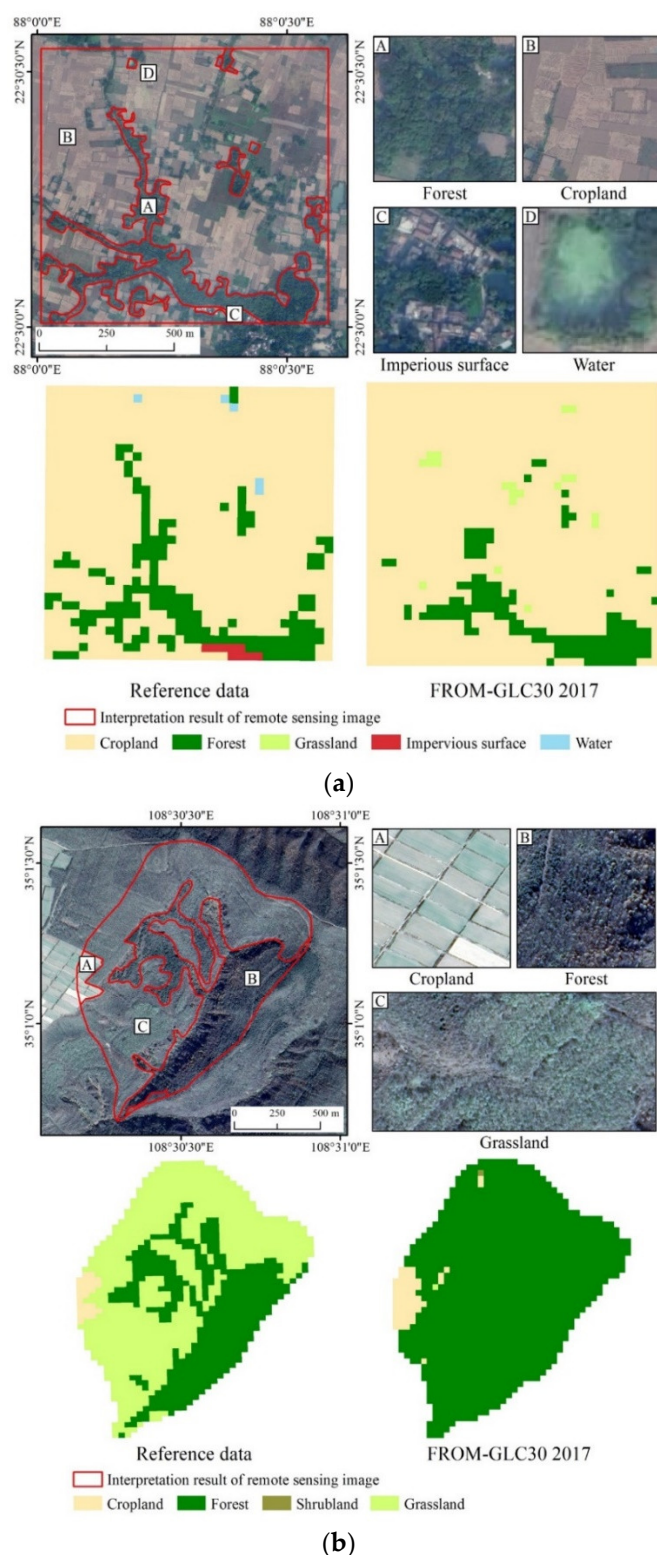


Figure 4. Google Earth images, interpretation results, reference data, and the Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC30) 2017 dataset of two sampling

units. (a) Google Earth image, interpretation result, reference data, and the FROM-GLC30 2017 dataset of sampling unit 1. (b) Google Earth image, interpretation result, reference data, and the FROM-GLC30 2017 dataset of sampling unit 2.

Table 1. Classification schemes employed in the reference data and the FROM-GLC30 2017 dataset. The two schemes are converted to match to seven classes.

Unified Classification	Reference Data	FROM-GLC30 2017	Definition
1 Cropland	1 Cropland	1 Cropland	Land used for growing crops and economic crops.
2 Forest	2 Forest	2 Forest	The land covered by trees with coverage over 30% and the sparse forest land with crown coverage of 10–30%.
3 Shrubland	3 Shrubland	4 Shrubland	The land with shrub coverage higher than 30%, and the desert shrub with desert area coverage higher than 10%.
4 Grassland	4 Grassland	3 Grassland, 7 Tundra	The land mainly covered with herbaceous vegetation, and the vegetation coverage is more than 10%, including the land covered by bryophytes, lichens, and cold-resistant herbaceous and shrub vegetation in the alpine area.
5 Impervious surface	5 Impervious surface	8 Impervious surface	Including urban land, industrial and mining land, commercial land, storage land, transportation facilities land, etc.
6 Water	6 Water	5 Wetland, 6 Water, 10 Snow/Ice	Including the areas covered by liquid water, glaciers, and permanent snow within the land area, as well as the land with shallow water or soil moisture at the water/land interface.
7 Bareland	7 Bareland	9 Bareland	Including bare soil, bare rock, desert, sandy land, gravel land, saline alkali land, and other natural land with vegetation coverage less than 10%.

2.3. Accuracy Assessment Methods

2.3.1. Area Proportion Analysis

The analysis of area proportion shows the composition of land cover types of the reference data and the FROM-GLC30 2017 dataset. It can be used to obtain the abundance of the land cover types in each dataset and compare the differences of land cover proportion between the reference data and the FROM-GLC30 2017 dataset. The area proportion of each land cover type is obtained by calculating the percentage of the total area of the specific land cover type in the total area of all the sampling units.

2.3.2. Accuracy Assessment Index

Overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) are the common indexes for the accuracy assessment of land cover data [33,34]. OA is a macro description of data accuracy, which shows the area proportion of the correct type in all land cover types. UA and PA represent data accuracy from the perspective of different land cover types. OA, UA, and PA can be found by aggregating the SU error matrices. For UA and PA, this was done as the variation at SU level for these is large. For OA, because the variation in SU size is not large, we chose to look at general statistics of OA, such as histogram, box plot, mean, SD, median, and so forth. We found OA from aggregating close to the mean.

In this study, the overall accuracy was calculated at each sampling unit. Then, the average values of sampling units within each region were calculated as the overall accuracy values of that region, which means each sampling unit has equal importance in the regional overall accuracy calculation. User's accuracy (UA) and producer's accuracy (PA) were calculated for each land cover type by summarizing all the SUs pixels in regions or the whole study area. By doing this, we could have equal importance for each pixel in calculating UA and PA within a certain domain, which also fits our expectations most, because in some SUs there is only quite a small number of pixels or no pixel with a certain land cover type.

The F_β statistic was used for a combined description of UA and PA [35]. In this manuscript, β was set to be 1, so F_β was then F_1 , which means the same importance of UA and PA.

The calculations used for OA, UA, PA and F_β were as follows:

$$OA = \frac{X}{N} \times 100\% \quad (1)$$

$$UA_i = \frac{X_i}{N_i} \times 100\% \quad (2)$$

$$PA_i = \frac{X_i}{N'_i} \times 100\% \quad (3)$$

$$F_{\beta i} = \frac{(1 + \beta^2)}{\frac{PA_i \times UA_i}{(\beta^2 \times PA_i + UA_i)}} \quad (4)$$

$$F_{1i} = \frac{2}{\frac{1}{PA_i} + \frac{1}{UA_i}} \quad (5)$$

where OA represents the overall accuracy; UA_i and PA_i represent the user's accuracy and the producer's accuracy of the land cover type i ; N is the total grid number of pixels; X is the number of pixels with the same land cover types in reference data and the FROM-GLC30 2017 dataset; X_i is the number of pixels with consistent attributes (the i th land cover type) in the FROM-GLC30 2017 dataset and the reference dataset. $F_{\beta i}$ and F_{1i} refer to the F_β and F_1 values for land cover type i .

3. Results

3.1. The Area Proportion Analysis of Each Land Cover Type

Figure 5 displays the area proportion of the land cover types in the FROM-GLC30 2017 dataset and the reference data. The total area in each case is then the total area of sampling units. The results showed that bareland, grassland, and forest were the three most common land cover types in both datasets, followed by cropland and water. There were also some differences in the two datasets. Compared with the reference data, the area proportion of bareland and grassland was higher in the FROM-GLC30 dataset, and the area proportion of forest was lower in the FROM-GLC30 dataset. However, the findings indicated that the area proportion of each land cover type of the FROM-GLC30 2017 generally conformed to the field condition in the Pan-Third Pole Area.

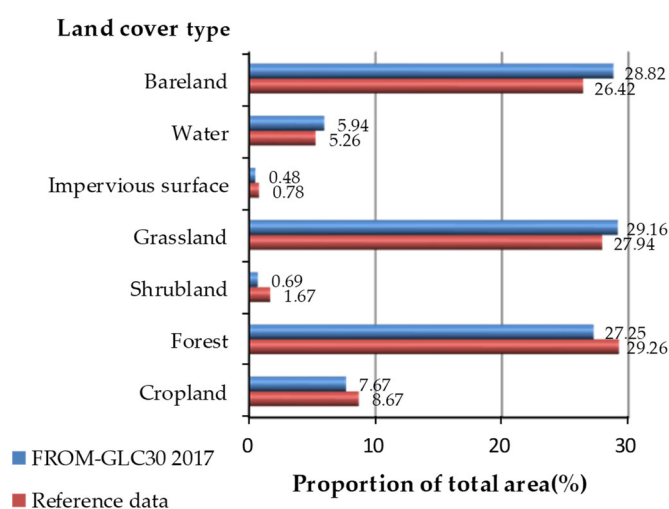


Figure 5. Area comparison of different land types in the FROM-GLC30 2017 dataset and the reference data.

3.2. Overall Accuracy in Different Regions

In the Pan-Third Pole Area, the overall accuracy of the FROM-GLC30 2017 dataset calculated based on the reference data was 72.78%. Statistical results of the overall accuracy of each sampling unit are given in Figure 6. It shows that 78.66% of the sampling units had an overall accuracy of more than 50%, and 54.35% had an overall accuracy between 80% and 100%. The result shows that the accuracy of the FROM-GLC30 2017 dataset is quite high in most of the sampling units.

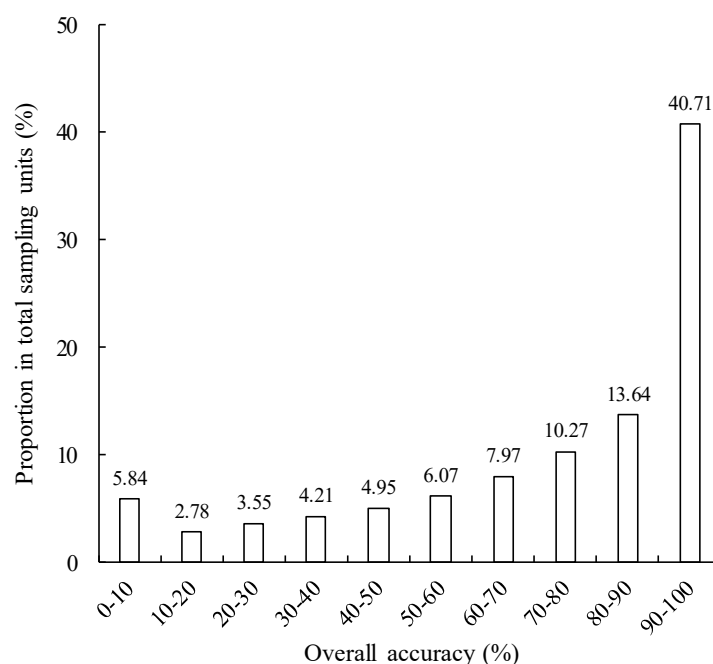


Figure 6. Overall accuracy distribution of sampling units in the Pan-Third Pole Area.

Figure 7 shows the box-plot of overall accuracy in eight regions. The box-plot presents five standard statistics in a plot. The five lines from top to bottom represent the maximum, upper quartile (75%), median, lower quartile (25%), and minimum. The boxes for Central and Eastern Europe and West Asia–Northeast Africa were relatively short, and the median of West Asia–Northeast Africa was the highest of all regions. It indicates that the distribution of the overall accuracy of sampling units in these two regions is relatively centralized, and the overall accuracy in West Asia–Northeast Africa is the highest. The boxes of Central Asia, Mongolia, and South Asia were long, and the median of South Asia was the lowest in all regions. It indicates that there is a large difference between the overall accuracy values of each sampling unit in these three regions, and the overall accuracy in South Asia is the lowest. The medians of all regions were all close to the upper quartile. It indicates that the overall accuracy of the FROM-GLC30 2017 dataset is relatively high.

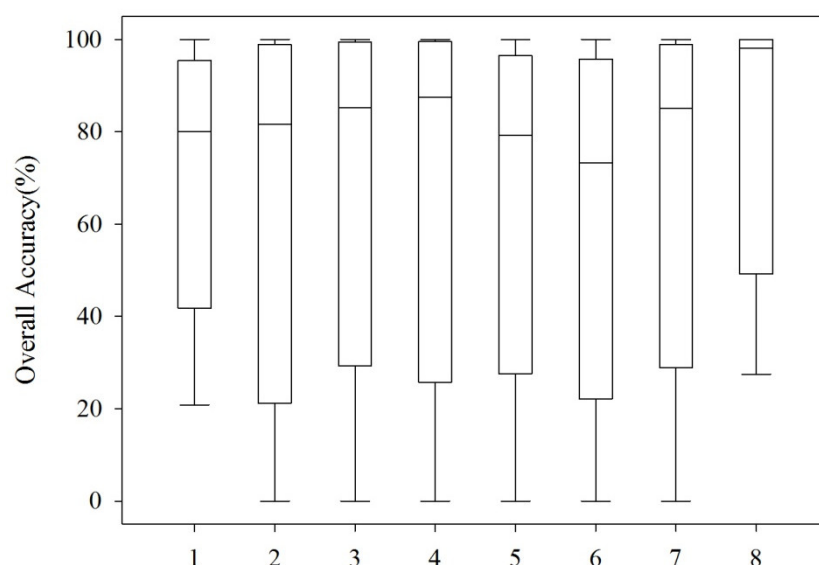


Figure 7. Box-plot of overall accuracy in eight regions. Regions: 1: Central and Eastern Europe; 2: Central Asia; 3: China; 4: Mongolia; 5: Russia; 6: South Asia; 7: Southeast Asia; 8: West Asia–Northeast Africa.

Table 2 shows the basic statistics of the overall accuracy of the sampling units in eight regions and in the whole Pan-Third Pole Area. Mean OA values were high in West Asia–Northeast Africa (80.11%), and low in South Asia (65.36%) and Central Asia (67.71%). For the other five regions, mean OA values were medium or between 70% and 75%. The median of the OA value was also high in West Asia–Northeast Africa (98.04%) and low in South Asia (73.37%). The median values of OA were larger than the mean value, since there were much longer quartiles and minimum values in the box-plot of OA (Figure 7).

Table 2. Mean, Median, and Standard Deviation of OA values.

Regions	Mean of OA (%)	Median of OA (%)	Standard Deviation of OA (%)
Central and Eastern Europe	74.42	80.13	21.01
Central Asia	67.71	81.64	34.58
China	74.38	85.31	28.88
Mongolia	71.53	87.54	33.91
Russia	70.81	79.29	27.78
South Asia	65.36	73.37	30.86
Southeast Asia	74.86	85.05	26.50
West Asia—Northeast Africa	80.11	98.04	31.22
Pan-Third Pole Area	72.78	83.70	29.38

OA: Overall Accuracy, calculated by an average of OA values over all sampling units.

Figure 8 displays the spatial distribution of the overall accuracy or geometric center of each sampling unit. The sampling units with higher overall accuracy (85–100%) were mainly distributed in northwestern China, central and northwest Russia, southern West Asia–Northeast Africa, and parts of the border between Central Asia and West Asia–Northeast Africa. The sampling units with lower overall accuracy (0–30%) were mainly concentrated in central and northern China, the central and southern regions of Central Asia, and the southwestern regions of South Asia. The findings show that the overall accuracy of the sampling unit has regional differences, which may be related to the complexity of the land cover composition in different regions.

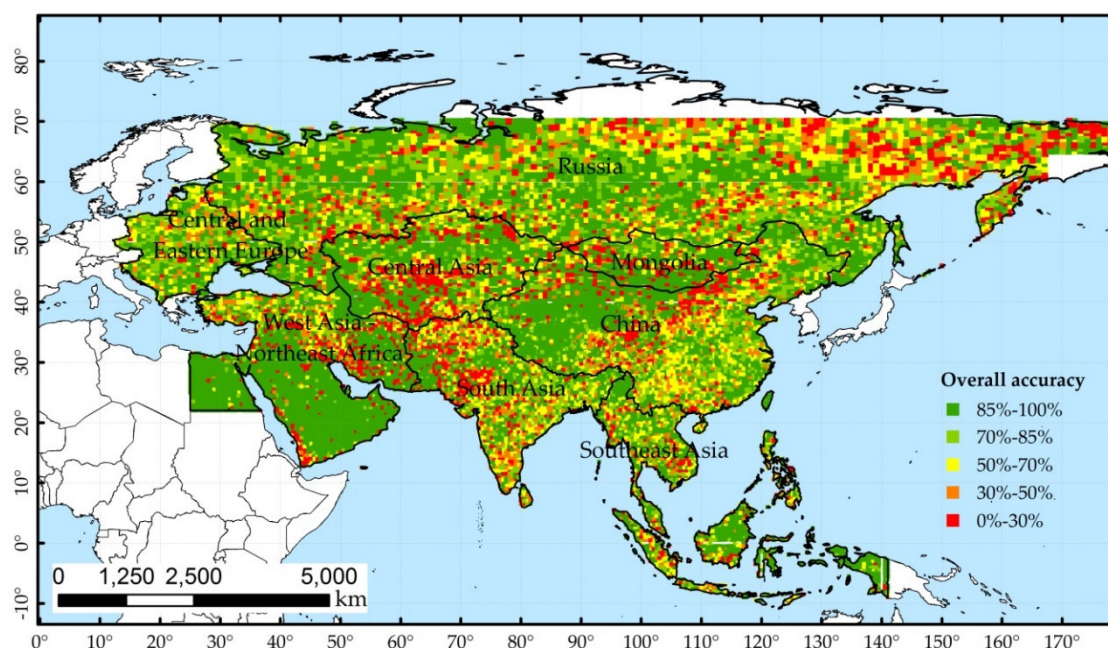


Figure 8. Spatial distribution of overall accuracy in sampling units.

3.3. Regional Accuracy of Different Land Cover Types

Table 3 records the user's accuracy, the producer's accuracy, and F_1 of the seven land cover types, which were calculated using formula (2), (3), and (5) for the whole study area as domain. F_1 values indicate the combined accuracy value of user's accuracy and the producer's accuracy of each land cover type. The accuracies for bareland and forest were highest with F_1 values slightly greater than 80%. The accuracies for water, cropland, and grassland were medium, with F_1 values of 75.69%, 70.16%, and 63.81%. The accuracies for shrubland and impervious surface were lowest, with F_1 values of only 4.67% and 34.26%. The user's accuracy of the forest was the highest, with a value of 83.95%, followed by bareland (77.65%) and cropland (74.44%). The user's accuracy of shrubland was the lowest. The producer's accuracy for bareland was the highest, with a value of 85.31%, followed by water (80.87%) and forest (78.73%). The producer's accuracy of shrubland was the lowest. Comparing the results of user's accuracy and producer's accuracy of the same land cover type, it can be found that the absolute value of the difference between the user's accuracy of impervious surface and the producer's accuracy of impervious surface was the largest, which was 16.94%, followed by water, which was 9.73%.

Table 3. User's accuracy (UA), producer's accuracy (PA), and F_1 of different land cover types.

Code	Land Cover Type	UA (%)	PA (%)	F_1 (%)	Difference between UA and PA (%)
1	Cropland	74.44	66.34	70.16	8.1
2	Forest	83.95	78.73	81.26	5.22
3	Shrubland	7.95	3.31	4.67	4.64
4	Grassland	62.26	65.44	63.81	−3.18
5	Impervious surface	44.71	27.77	34.26	16.94
6	Water	71.14	80.87	75.69	−9.73
7	Bareland	77.65	85.31	81.30	−7.66

The F_1 values, which are a combined value of user's accuracy and producer's accuracy in Table 4, showed the accuracy for each land cover type differs in regions. The accuracy for cropland was high in Central and Eastern Europe (79.6%) and South Asia (77.21%), low in Central Asia (8.51%) and Mongolia (22.9%). The accuracy for forest was high in Southeast Asia (89.17%) and Central and Eastern Europe (85.96%) and low in Central Asia (47.87%). The accuracy for shrubland

was under 20% in all the eight regions, and for grassland, it was high in Central Asia (76.61) and Mongolia (74.02), low in Russia (13.44%) and Southeast Asia (20.3%). The accuracy for impervious surface was under 45% in all the eight regions. The accuracy for water was high in West Asia–Northeast Africa (95.87%), Central and Eastern Europe (89.75%), and Central Asia (87.01%), and low in South Asia (15.67%). The accuracy for bareland was high in West Asia–Northeast Africa (92.32%), low in Central and Eastern Europe (0.82%), Russia (7.14%), and Southeast Asia (11.78%).

Table 4. F_1 , user's accuracy, and producer's accuracy of different land cover types in eight regions.

Regions	Cropland (%)	Forest (%)	Shrubland (%)	Grassland (%)	Impervious Surface (%)	Water (%)	Bareland (%)
1	79.60 (92.24, 70.01)	85.96 (86.27, 85.66)	0.23 (3.50, 0.12)	38.43 (26.59, 69.29)	41.89 (66.80, 30.51)	89.75 (92.72, 86.97)	0.82 (0.63, 1.18)
2	8.51 (72.11, 34.52)	47.87 (70.92, 36.13)	1.84 (28.26, 0.95)	76.61 (77.15, 76.07)	14.46 (16.89, 12.64)	87.01 (84.56, 89.60)	65.42 (55.71, 79.23)
3	71.18 (74.43, 68.21)	82.90 (80.17, 85.82)	3.14 (9.60, 1.88)	52.27 (78.81, 39.10)	44.89 (48.90, 41.49)	63.87 (58.22, 70.74)	57.42 (42.18, 89.90)
4	22.90 (15.94, 40.67)	71.93 (89.47, 60.14)	Null (Null, Null)	74.02 (82.72, 66.98)	Null (Null, Null)	72.18 (66.63, 78.75)	76.75 (65.53, 92.60)
5	61.94 (85.02, 48.72)	79.50 (82.24, 76.94)	1.35 (2.44, 0.93)	13.44 (64.84, 67.50)	27.06 (44.79, 19.39)	70.03 (63.59, 77.93)	7.14 (4.06, 29.55)
6	77.21 (75.35, 79.17)	76.94 (86.50, 69.29)	9.2 (8.92, 9.49)	37.69 (50.40, 30.10)	26.16 (51.02, 17.59)	15.67 (50.35, 9.28)	66.58 (53.62, 87.80)
7	57.26 (46.81, 73.72)	89.17 (91.85, 86.64)	4.21 (3.65, 4.96)	20.30 (24.16, 17.51)	29.98 (56.08, 20.46)	74.79 (76.44, 73.21)	11.78 (14.95, 9.72)
8	66.76 (75.61, 59.77)	68.16 (86.08, 56.42)	17.01 (17.96, 16.16)	44.59 (60.99, 35.14)	19.31 (19.90, 18.75)	95.87 (96.43, 95.32)	92.32 (87.01, 98.32)

Regions: 1: Central and Eastern Europe; 2: Central Asia; 3: China; 4: Mongolia; 5: Russia; 6: South Asia; 7: Southeast Asia; 8: West Asia–Northeast Africa. Values in this table refer to F_1 (user's accuracy, producer's accuracy). The value NULL indicates no such land cover pixels were found in all the sampling units (SUs) in a certain region.

4. Discussion

4.1. The Influence of Sampling Units on Land Cover Accuracy: Small Watershed vs. Pixel

In previous research on land cover data accuracy assessment, sampling units were mostly based on pixels [23,36], while in this study, small watersheds were used. A small watershed is the basic unit with specified geographical characteristics, such as climate, soil, terrain, vegetation, and also land cover, which is similar as in the other nearby watersheds. That is why it is more likely to be able to represent the land cover characteristics and its accuracy in regions by using watersheds as sampling units.

Figure 9 shows the spatial interpolation results of the overall accuracy, based respectively on pixel units and small watershed units in Yunnan and Guizhou Provinces, China. The value of the points used for interpolation in Figure 9a was the accuracy of each pixel, and the value was 0 (incorrect) or 100% (correct). In Figure 9b, the point values were derived from the average accuracy of the small watershed sampling unit. The value was calculated according to Formula 1 and could be any values between 0 and 100%. Using the same sampling scheme, it is found that the averaged overall accuracy results of the FROM-GLC30 2017 dataset based on pixel and small watershed were 71.49% and 68.74%, respectively, which were close to each other. In the Pan-Third Pole Area, the overall accuracy of the FROM-GLC30 2017 dataset calculated in this research was 72.78%, also similar to the accuracy of 72.35% published by the data producer [32]. However, the spatial interpolation results were quite different. The interpolation result based on small watershed had more spatial continuity and could better express the overall characteristics of the region, while the interpolation result based on pixel was relatively broken in pattern, and the overall accuracy results in each sampling unit had a greater contingency.

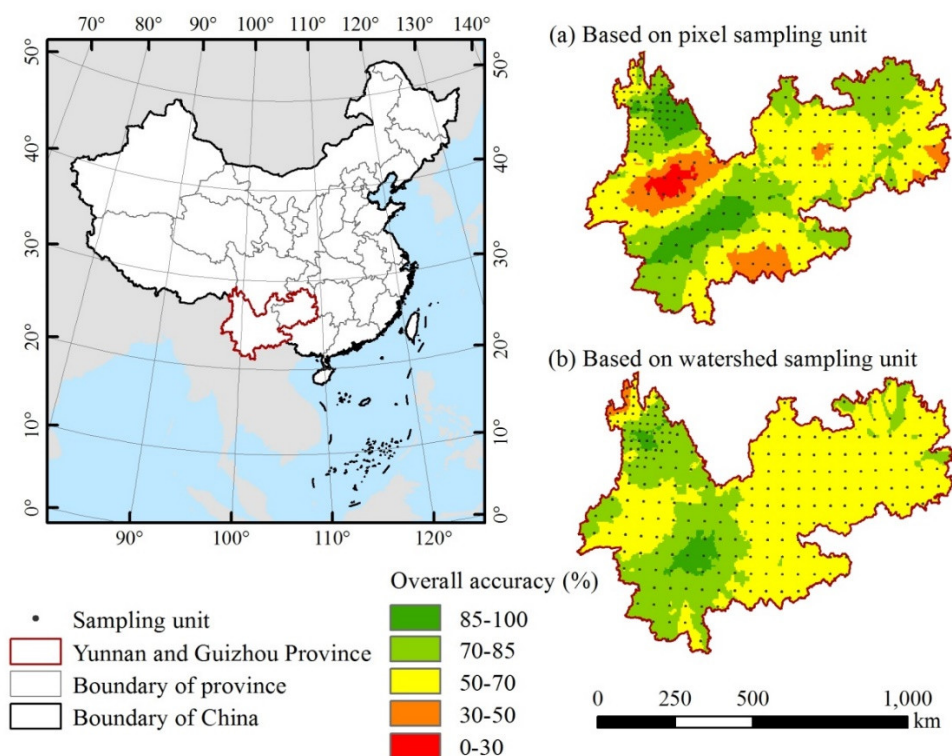


Figure 9. Spatial interpolation results of the overall accuracy of the FROM-GLC30 2017 dataset in Yunnan and Guizhou Province, China. (a) Based on the sampling units of pixels; (b) based on the sampling units of the small watersheds.

To further illustrate the ability of the methods to express the spatial differences of regional land cover accuracy based on small watersheds sampling units and pixels, this study carried out validation experiments in Yunnan and Guizhou Province. In the experimental area, 85% of the sampling units (212) were randomly selected as sample sets for the overall accuracy interpolation, and the remaining 15% of the sampling units (37) were used as verification samples for the comparison with the accuracy interpolation results. Root Mean Square Error (RMSE) between the verification sampling units accuracy and interpolation results was calculated based on small watersheds and pixels respectively. The smaller the RMSE was, the better the sampling method was for accurate expression of the spatial accuracy difference. The results showed that the RMSE based on the small watershed units was 21.2%, and the RMSE based on pixel units was 42.7%. In summary, there was not much difference between the overall accuracy values based on small watershed unit and pixel unit. However, taking small watersheds as the sampling units could better reflect the difference of classification accuracy spatially.

4.2. Geographical Interpretation of the Spatial Differentiation of the FROM-GLC30 2017 Dataset Accuracy

According to the spatial distribution results of the overall accuracy in each sampling unit, the accuracy of the FROM-GLC30 2017 dataset in different regions was quite different. In Section 3.2, Figure 8 shows that the overall accuracy of sampling units in northwestern China, central and northwest Russia, south of West Asia–Northeast Africa, and parts of the border between Central Asia and West Asia–Northeast Africa was high, being more than 85%. The possible reason is that these areas generally have wide area and sparse population, and the composition of land cover type is often single with a large coverage area. In these places, different land cover types are usually distributed in concentrated and contiguous units, mostly consisting of desert, bareland, forest, grassland, or cropland. These land cover types have unique colors, shapes, and texture features in remote sensing images, which are easy to distinguish, so the overall accuracy of these areas is relatively higher (Figure 10, H1–H3).

The areas with lower overall accuracy were mainly located in central and northern China, the central and southern regions of Central Asia, and the southwestern regions of South Asia. The possible reasons are that in these areas, the level of urbanization is not very high and the land cover types are complex; the degree of aggregation of the same land type is not high, the plots of different land cover types are scattered, and the area of each plot is not large (Figure 10, L1–L4). Thus, it is reasonable to assume the fragmentation of land cover in these regions will influence and lower their accuracy. The relationships between fragmentation and accuracy will need more detailed exploration in a future study.

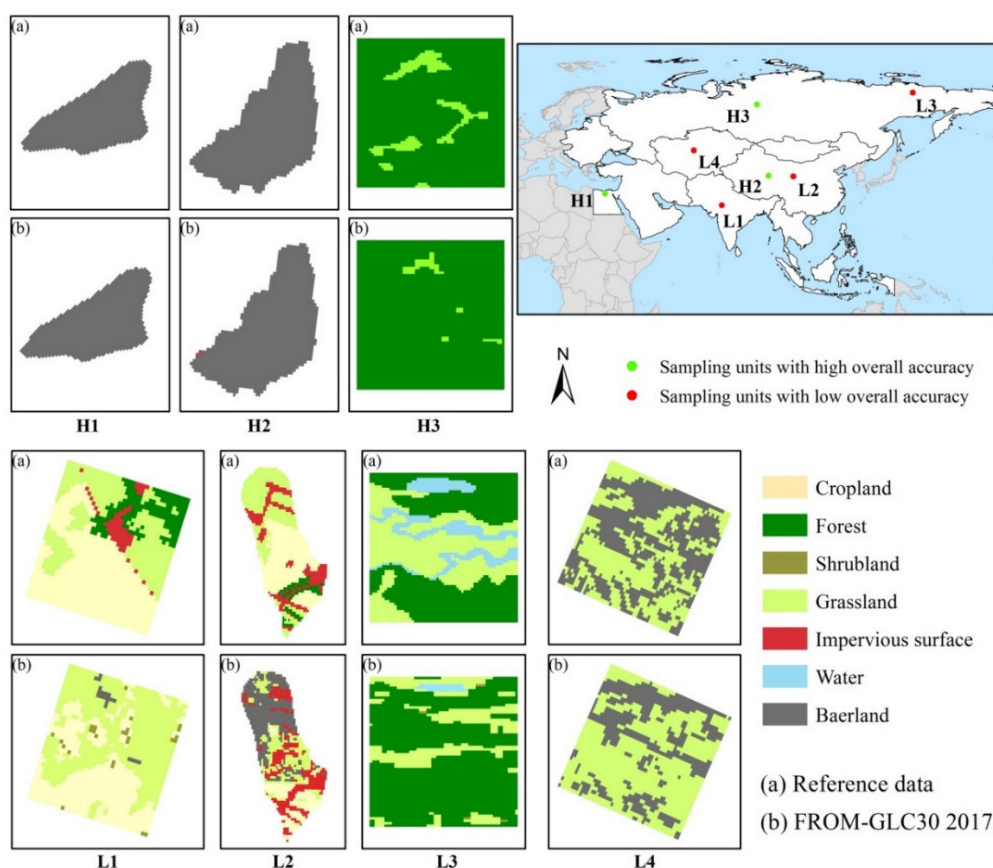


Figure 10. Land cover composition in typical sampling units.

4.3. Explanation of the Accuracy Differences between Different Land Cover Types

The accuracies of forest, bareland, cropland, and water were relatively higher among the seven land cover types. The possible reason is that these four land cover types usually have wider and more concentrated distributions. They also have relatively unique textures and colors in remote sensing images, which makes them easier for identification. The accuracy of the impervious surface was relatively lower. This may be because the distribution of this land cover type is not centralized, especially in non-urban areas. Scattered houses or buildings with areas only of hundreds of square meters or roads with a width of several meters failed to be displayed in the FROM-GLC30 2017 dataset. This might be a reason for the low accuracy of impervious surface. The area proportion of shrubland in the whole study area was the smallest, at less than 1.7% of the total area. The distribution of shrubland is relatively scattered, and it often exists with grassland and forest in the same place. Because of the different growth forms, forest and grassland should be able to be separated, but shrubland is hard to separate, so it is challenging to distinguish shrubland accurately, whether in the interpretation of reference data or in the interpretation of the FROM-GLC30 2017 dataset. This is most likely the reason for the low accuracy of shrubland.

The accuracy for a particular land cover also differed between regions. For example, the accuracy of bareland in West Asia–Northeast Africa was 92.32%, and was highest among all regions.

In this region, there are some deserts with a large area, such as the Rub' al Khali Desert and the Neford Desert. It is possible that if a particular type of land cover is distributed widely and is concentrated, then the land cover product accuracy will be high, otherwise it is likely low. Other geographical and environmental factors, such as slope, aspect, terrain relief, and so forth, may also have some effects on the classification accuracy of land cover types, which could also lead to the spatial difference of the accuracy for the same land cover types. These spatial factors deserve further study.

4.4. Scale Effect on Reference Data

In this study, to ensure the consistency between the reference data and the FROM-GLC30 2017 dataset on the spatial scale, the grid size of the reference data was synthesized from 1 m to 30 m. This process has specific impacts on the reference data in reflecting the real land cover condition. Figure 11 compares the reference data at the scale of 1 m with those at 30 m in two watershed sampling units. In the local enlargements, the reference data at the scale of 1 m can show more detailed and continuous information. With the scale transformation of the reference data, part of the ground information was lost at the scale of 30 m. In particular, the scattered patch of dissimilar land cover types and irregular boundaries between land cover types are greatly affected.

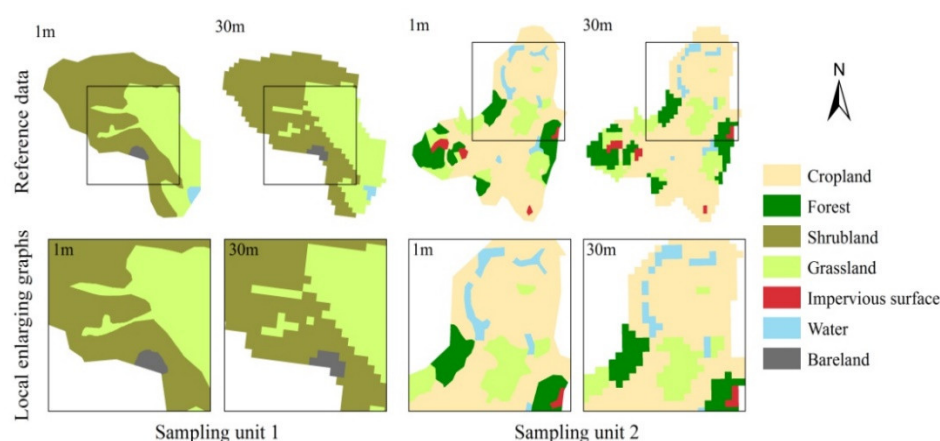


Figure 11. The influence of scale transformation on the reference data.

According to statistics, there is little difference in the area proportion of each land type before and after the scale transformation. However, in both sampling units, some patches become scattered, and the location of some patches changes. These changes are caused by the scale transformation from 1 m to 30 m and may certainly have an effect on the accuracy results for sampling units. In further research, the mechanism of the scale effect and its influence on the accuracy of land cover data should be further explored.

5. Conclusions

In this paper, the accuracy of the FROM-GLC30 2017 dataset in the Pan-Third Polar Area was studied using reference data based on interpretations of high-resolution remote sensing images and common sampling units. The results for overall accuracy, user's accuracy, and producer's accuracy vary with location as well as land covers of interest. These may be helpful for applications in different locations of the Pan-Third Pole Area. The main conclusions were as follows:

- (1) In the study area, the proportion of land cover types in the FROM-GLC30 dataset was similar to that in the reference data. The difference between the two datasets for all the land types was small.
- (2) The overall accuracy of the FROM-GLC30 2017 dataset in the Pan-Third Polar Area was 72.78%. The sample units with an overall accuracy of more than 50% accounted for 78.66% of the total sample units, and the sample units with an overall accuracy of 80–100% accounted for 54.35%.

of the total sample units. The regions with the highest and lowest overall accuracy were located in West Asia–Northeast Africa and South Asia, respectively.

- (3) The accuracy for different land cover types differed. Generally, the accuracy for bareland and forest was high, which was higher than 80%, the accuracy for water, cropland, and grassland was medium, and the accuracy was low for shrubland and impervious surface, which was only 4.67% and 34.26%. The accuracy for each land cover type differed in different regions.

The summary information from this study will support applications in the Pan-Third Polar Area using FROM-GLC30 2017 data where regional differences and land cover differences in accuracy may be more important than overall accuracy.

Author Contributions: Conceptualization, C.W. and G.P.; Methodology, all authors; Software, X.L.; Data Curation, Z.G. and M.Z.; Writing—Original Draft Preparation, Z.G.; Writing—Review and Editing, all authors; Visualization, Z.G.; Supervision, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences, Pan-Third Pole Environment Study for a Green Silk Road (Pan-TPE) (XDA20000000), National Natural Science Foundation of China, (41977062, 41601290), Program for Key Science and Technology Innovation Team in Shaanxi Province, (2014KCT-27).

Acknowledgments: Thanks to Liu Baoyuan and Zhang Wenbo from Beijing Normal University, researcher Zhang Xiaoping from Institute of soil and water conservation Chinese Academy of Sciences, Chang Qingrui from Northwest A&F University, Shi Yun from Ningxia University, and all of their team members who participated in the land cover interpretation of the reference data. We also acknowledge the data support from “the National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (<http://www.geodata.cn>)”.

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

References

1. Hailu, B.T.; Fekadu, M.; Nauss, T. Availability of global and national scale land cover products and their accuracy in mountainous areas of Ethiopia: A review. *J. Appl. Remote Sens.* **2018**, *12*, 041502.
2. Liang, D.; Zuo, Y.; Huang, L. 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.
3. Son, S.; Kim, J. Accuracy assessment of global land cover datasets in South Korea. *Kor. J. Remote Sens.* **2018**, *34*, 601–610.
4. Senanayake, S.; Pradhan, B.; Huete, A. Assessing Soil Erosion Hazards Using Land-Use Change and Landslide Frequency Ratio Method: A Case Study of Sabaragamuwa Province, Sri Lanka. *Remote Sens.* **2020**, *12*, 1483.
5. Minaei, M.; Shafizadeh-Moghadam, H.; Tayyebi, A. Spatiotemporal nexus between the pattern of land degradation and land cover dynamics in Iran. *Land Degrad. Dev.* **2018**, *29*, 2854–2863.
6. Li, S.; Cui, Y.; Liu, M. Integrating Global Open Geo-Information for Major Disaster Assessment: A Case Study of the Myanmar Flood. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 201–219.
7. 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.
8. Congalton, R.G.; Gu, J.Y.; Yadav, K. Global land cover mapping: A review and uncertainty analysis. *Remote Sens.* **2014**, *6*, 12070–12093.
9. Hansen, M.C.; Defries, R.S.; Townshend, J.R.G. Global land cover classification at 1km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331–1364.
10. Loveland, T.R.; Reed, B.C.; Brown, J.F. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330.
11. Liu, Q.H.; Zhang, Y.L.; Liu, L.S. The spatial local accuracy of land cover datasets over the Qiangtang Plateau, High Asia. *J. Geogr. Sci.* **2019**, *29*, 1841–1858.
12. Wang, Y.; Zhang, J.; Liu, D. Accuracy assessment of GlobeLand30 2010 land cover over China based on geographically and categorically stratified validation sample data. *Remote Sens.* **2018**, *10*, 1213.

13. Tsensdbazar, N.E.; Herold, M.; Bruin, S.D. Developing and applying a multi-purpose land cover validation dataset for Africa. *Remote Sens. Environ.* **2018**, *219*, 298–309.
14. Arsanjani, J.J.; See, L.; Tayyebi, A. Assessing the suitability of GlobeLand30 for mapping land cover in Germany. *Int. J. Digit. Earth* **2016**, *9*, 873–891.
15. Yang, Z.Q.; Dong, J.W.; Liu, J.Y. Accuracy Assessment and Inter-Comparison of Eight Medium Resolution Forest Products on the Loess Plateau, China. *ISPRS Int. Geo Inf.* **2017**, *6*, 152–167.
16. Samasse, K.; Hanan, N.P.; Tappan, G. Assessing cropland area in West Africa for agricultural yield analysis. *Remote Sens.* **2018**, *10*, 1785–1803.
17. Gong, P.; Wang, J.; Yu, L. 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.
18. Ji, L.Y.; Gong, P.; Geng, X.R. Improving the accuracy of the water surface cover type in the 30 m FROM-GLC product. *Remote Sens.* **2015**, *7*, 13507–13527.
19. Yu, L.; Wang, J.; Gong, P. Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets. *Int. J. Remote Sens.* **2013**, *34*, 5851–5867.
20. Yu, L.; Wang, J.; Cao, X.L. A multi-resolution global land cover dataset through multisource data aggregation. *Sci. China Earth Sci.* **2014**, *57*, 2317–2329.
21. Lu, M.; Wu, W.; Zhang, L. A comparative analysis of five global cropland datasets in China. *Sci. China Earth Sci.* **2016**, *59*, 2307–2317.
22. Chen, X.Y.; Lin, Y.; Zhang, M. Assessment of the cropland classifications in four global land cover datasets: A case study of Shaanxi Province, China. *J. Integr. Agric.* **2017**, *16*, 298–311.
23. Kang, J.M.; Wang, Z.H.; Sui, L.C. Consistency Analysis of Remote Sensing Land Cover Products in the Tropical Rainforest Climate Region: A Case Study of Indonesia. *Remote Sens.* **2020**, *12*, 1410.
24. Islam, S.; Zhang, M.; Yang, H. Assessing inconsistency in global land cover products and synthesis of studies on land use and land cover dynamics during 2001 to 2017 in the southeastern region of Bangladesh. *J. Appl. Remote Sens.* **2019**, *13*, 048501.
25. Yao, T.D.; Chen, F.H.; Cui, P. From Tibetan Plateau to Third Pole and Pan-Third Pole. *Bull. Chin. Acad. Sci.* **2017**, *32*, 924–931. (In Chinese).
26. Ma, W.Q.; Zhong, L. Monitoring and Modeling the Tibetan Plateau's climate system and its impact on East Asia. *Sci. Rep.* **2017**, *7*, 44574–44579.
27. Berkowitz, R. Move a plateau, change a climate. *Phys. Today* **2018**, *71*, 21–23.
28. Yin, S.Q.; Zhu, Z.Y.; Wang, L. Regional soil erosion assessment based on a sample survey and geostatistics. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 1695–1712.
29. Liu, B.Y.; Guo, S.Y.; Li, Z.G. China water erosion survey based on sampling strategy. *Soil Water Conserv. China (Soil Water Conserv. China)* **2013**, *34*, 30–38. (In Chinese)
30. Clark, M.L.; Aide, T.M.; Grau, H.R. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the dry chaco ecoregion of South America. *Remote Sens. Environ.* **2010**, *114*, 2816–2832.
31. Friedl, M.A.; Sulla-Menashe, D.; Tan, B. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182.
32. National Earth System Science Data Center, National Science & Technology Infrastructure of China. Available online: <http://www.geodata.cn> (accessed on 21 September 2018).
33. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.
34. Herold, M.; Mayaux, P.; Woodcock, C.E. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* **2008**, *112*, 2538–2556.
35. Rijsbergen, V.; Joost, C.K. *Information Retrieval*, 2nd ed.; Butterworths: Waltham, MA, USA, 1979.
36. Manakos, I.; Karakizi, C.; Gkinis, I. Validation and Inter-Comparison of Spaceborne Derived Global and Continental Land Cover Products for the Mediterranean Region: The Case of Thessaly. *Land Degrad. Dev.* **2017**, *6*, 34–48.

