Research Article

Lichun Sui, Junmei Kang, Xiaomei Yang, Zhihua Wang*, and Jun Wang

Inconsistency distribution patterns of different remote sensing land-cover data from the perspective of ecological zoning

https://doi.org/10.1515/geo-2020-0014
received November 12, 2019; accepted January 21, 2020

Abstract: Analyzing consistency of different land-cover data is significant to reasonably select land-cover data for regional development and resource survey. Existing consistency analysis of different datasets mainly focused on the phenomena of spatial consistency regional distribution or accuracy comparison to provide guidelines for choosing the land-cover data. However, few studies focused on the hidden inconsistency distribution rules of different datasets, which can provide guidelines not only for users to properly choose them but also for producers to improve their mapping strategies. Here, we zoned the Sindh province of Pakistan by the Terrestrial Ecoregions of the World as a case to analyze the inconsistency patterns of the following three datasets: GlobeLand30, FROM-GLC, and regional land cover (RLC). We found that the inconsistency of the three datasets was relatively low in areas having a dominant type and also showing homogeneity characteristics in remote sensing images. For example, cropland of the three datasets in the ecological zoning of Northwestern thorn scrub forests showed high consistency. In contrast, the inconsistency was high in areas with strong heterogeneity. For example, in the southeast of the Thar desert ecological zone where cropland, grassland, shrubland, and bareland were interleaved and the surface cover complexity was relatively high, the inconsistency of the three datasets was relatively high. We also found that definitions of some types in different classification systems are different, which also increased the inconsistency. For example, the definitions of grassland and bareland in GlobeLand30 and RLC were different, which seriously affects the consistency of these datasets. Hence, producers can use the existing land-cover products as reference in ecological zones with dominant types and strong homogeneity. It is necessary to pay more attention on ecological zoning with complex land types and strong heterogeneity. An effective way is standardizing the definitions of complex land types, such as forest, shrubland, and grassland in these areas.

Keywords: remote sensing, land-cover data, terrestrial ecoregions, spatial consistency

1 Introduction

Land cover is one of the key factors affecting the energy balance of surface solar radiation [1–4]. Changes in land cover and its spatial distribution have important impacts on global environmental change, climate change, the energy cycle of earth ecosystems, regional resources, and the sustainable development of economy and society [5–8]. At present, surface land-cover information and dynamic changes form the basis of scientific research in many fields, such as environmental modeling and soil erosion [9,10]. Traditional land-cover information is acquired through field surveys, which are time-consuming, costly, and mostly inaccurate. Satellite remote sensing is the only method for obtaining surface information quickly, in large scale, and periodically with the development of remote sensing and geographic information system technology [11–16]. Therefore, the global and regional land-cover data produced by remote

* **Corresponding author: Zhihua Wang,** State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; University of Chinese Academy of Sciences, Beijing 100049, China, e-mail: zhwang@lreis.ac.cn, tel: +86-10-6488-8955
Lichun Sui, Junmei Kang, Jun Wang: College of Geological Engineering and Geomatics, Chang'an University, Xi'an Shaanxi 710054, China
Xiaomei Yang: State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; University of Chinese Academy of Sciences, Beijing 100049, China

sensing have been successively launched. The most commonly used datasets are Global Land Cover 2000 (GLC2000), produced by the European Union Joint Research Center [17]; Moderate Resolution Imaging Spectroradiometer (MODIS), produced by Boston University [18]; Fine Resolution Observation and Monitoring of Global Land Cover (FROM-GLC), produced by Tsinghua University [19]; and GlobeLand30, produced by the National Geomatics Center of China [20]. Due to the differences in remote sensing platforms, classification methods, and the spatial resolution, it is particularly important to carry out consistency analysis of multi-source remote sensing land-cover data in specific applications.

Researchers have conducted the consistency analyses on different remote sensing land-cover data. Dai et al. [21] analyzed the consistency of five land-cover datasets (GLC2000, GlobCover2005, GlobCover2009, MODIS2000, and GlobeLand30) in South America by area similarity, confusion matrix, and spatial consistency methods. They found that the consistency of the five datasets ranged between 42.27% and 87.59%, and the consistency of the forest land type was the highest. Hongli and Xiaonan [22] evaluated the category accuracy of four types of data, including FROM, MODIS, European Space Agency Climate Change Initiative, and GLOBCOVER, using indicators such as overall accuracy, producer accuracy, and user accuracy. The results showed that the FROM and MODIS products possessed the highest overall consistency with the reference land-cover data, obtaining 0.69 and 0.67, respectively, but the overall consistency between the GLOBCOVER products and the reference data was relatively low, at only 0.55. Giri et al. [23] assessed the consistency of the MODIS and GLC2000 datasets by confusion matrix and area comparison analysis methods. The results indicated that the land-cover types were generally consistent, except for savanna/bush and wetland, but upon comparing finer land-cover types, the inconsistency area increased. Yang et al. [24] analyzed the consistency of seven different land-cover data in China and found that these land-cover data differed significantly in the area type and spatial distribution. Using GLCD-2005 data, Bai et al. [25] evaluated the consistency and inconsistency of the five popular land-cover data (GLCC, UMD, GLC2000, MODIS LC, and GlobCover) in China. The results showed that the consistency among five different land-cover data of grassland and cropland types was higher, while the consistency of shrubland and wetland types was lower.

However, these previous studies typically focused on the consistency and inconsistency between different land-cover data, paying less attention to the regularity behind the distribution of the inconsistencies. Hua et al. [26] studied the consistency and distribution rules of five kinds of land-cover data, including GLC2000, CCILC, MCD12, GLOBCOVER, and GLCNMO, on the global and continental scales by area similarity, confusion matrix, and spatial consistency methods. The results showed that the overall consistency of the five land-cover data ranged between 49.2% and 67.63% at the global scale, and the value for the European zone was relatively high at 66.57%. More importantly, the overall consistency of the datasets on the six continents fluctuated greatly with the change in elevation. When the elevation was greater than 3,500 m, the consistency of six continents showed a downward trend, and the consistency was lower than 1,000–3,500 m. When the elevation was less than 3,500 m, the consistency variation trend of the datasets in six continents differed greatly. For example, the overall consistency increased rapidly in North America, while the overall consistency declined below 1,000 m but increased significantly at 1,000–3,500 m in South America. In addition, this difference was also reflected in the climate gradient. For example, the overall consistency of the EF (icecap) climate zone was significantly higher than that of the other climate zones, 95%, while the overall consistency of the DS (subarctic) and ET (tundra) climate zones was poor, and the average values were about 45.44% and 41.92%. Hua's study not only provides data consistency conclusions on a global and continental scale but also compares and analyzes the two perspectives of elevation and climate division. The research results can be used to classify global land cover, especially the region in complex areas, and the selection of representative classification samples is of great significance and has a very high reference value. However, it is still insufficient to reflect the consistency between different land-cover data and the distribution rule from the perspective of elevation and climate. The incorporation of additional factors, such as slope and precipitation, will help to gain deeper insight into this. In order to close the gaps in this field, we explored the consistency and distribution rules between different land-cover data from the perspective of the Terrestrial Ecoregions of the World (TEOW).

TEOW was established by the World Wildlife Fund for conservation purposes. The ecological zones were divided into 8 biogeographic realms and 14 biomes. Based on these two basic layers, the world was divided into a total of 867 ecological zones [27]. Each ecological zone in TEOW with similar biomes, which are relatively stable and remain static in a certain period, is divided based on certain geographical attributes (such as elevation, slope, and precipitation). Therefore, TEOW can accurately reflect the distribution of species and communities.

Therefore, this study used the GlobeLand30, FROM-GLC, and regional land cover (RLC) data to study the inconsistency patterns of the products in different ecological zones by spatial superposition methods. In addition, in order to highlight the consistent distribution regularities of different land cover products in different ecological zones, we chose the Sindh province with insignificant differences in elevation and climate as the research area. The research results can provide guidance for reasonable sample layout and improve mapping strategies in the process of land-cover classification to improve mapping accuracy.

2 Study area and data source

2.1 Study area

The Islamic Republic of Pakistan is located in south Asia, south of the Arabian Sea, and borders India, China, Afghanistan, and Iran on the east, north, and west. The Sindh, whose area ranks third in the four provinces of Pakistan, is located in the southwest of the south Asian subcontinent, the southeast of Pakistan, and downstream of the Indus River (Figure 1). Sindh has a subtropical climate with hot temperatures of up to 46°C in summer and cold temperatures as low as 2°C in winter. It is affected by the southwest monsoon from February to September each year. It has a dry climate with an annual rainfall of 180 mm. The regional economy is dominated by agriculture, and it is a major production area of cotton, wheat, and rice, with cotton output accounting for one-third of the country.

2.2 Data source and preprocessing

This aim of this study is to analyze the consistency of multi-source remote sensing land-cover data with higher resolution. Currently, available products with higher resolution land cover include GlobeLand30, FROM-GLC, and RLC, but these products were made at different times. Therefore, to reduce the consistency error caused by time difference, we selected GlobeLand30-2010, FROM-GLC2017, and RLC2016 products from these available products for analysis. Although values may differ across individual years, the natural ecosystem usually changes significantly within 10 years or longer [28]. Furthermore, we use Google Earth high-resolution images as a reference, and the 330 random sample points (Figure 2) in the research area were identified for their type values in 2010, 2016, and 2017. Then, the land cover change matrix was calculated. The experimental results (Tables 1–3) indicate that the ecological status of

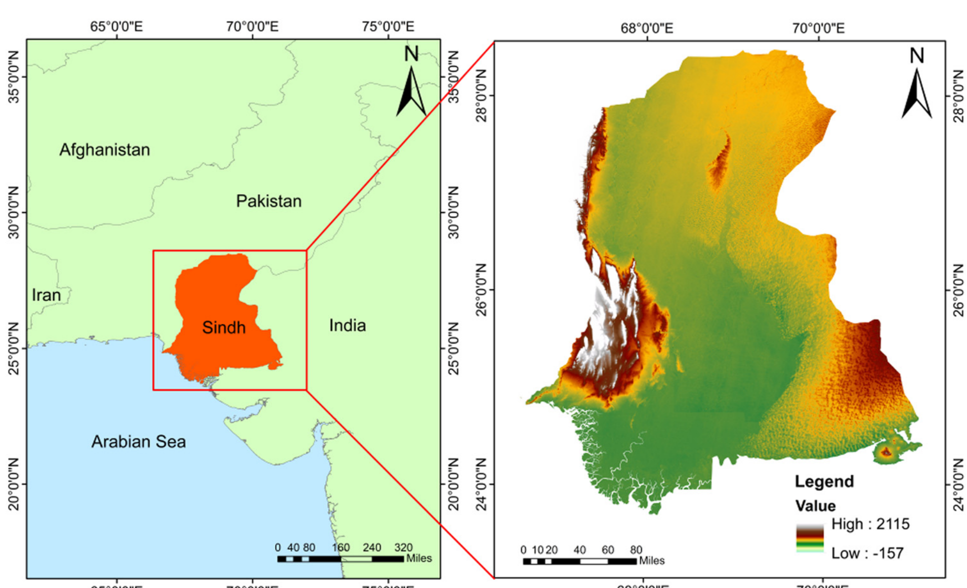

Figure 1: Geographical location and topographic map of the study area.

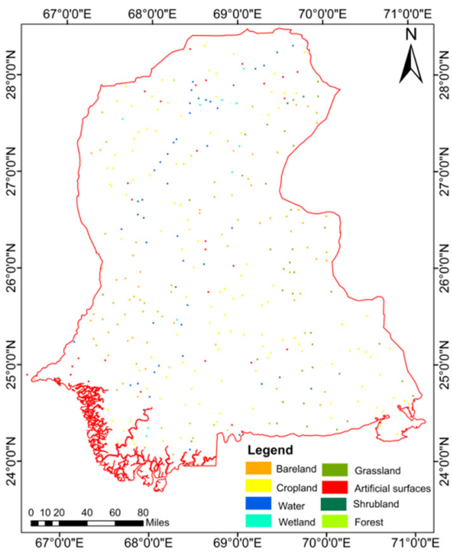

Figure 2: Spatial distribution of sample points in 2010.

the study area is relatively stable from 2010 to 2017. The variation error caused by dynamic change of land cover is much smaller than that caused by different producers, different data sources, and different interpretation methods. In addition, the RLC data were produced by the Institute of Geographic Sciences and Natural Resources Research. The production of the RLC product, atmospheric correction, orthographical correction, and region clipping are pre-processed on the GF-1 remote sensing image, and then a multi-feature adaptive classification model based on geological understanding is used for automatic classification. The main parameters of the three datasets are shown in Table 4.

Table 1: Land cover change matrix from 2010 to 2016

Year	2010									
	Type	**BL**	**CP**	**WT**	**WL**	**GL**	**AS**	**SL**	**FR**	**Row total**
2016	BL	84	0	1	0	1	0	0	0	86
	CP	0	114	1	0	0	0	0	0	115
	WT	0	0	33	0	0	0	0	0	33
	WL	0	0	1	10	0	0	0	0	11
	GL	1	0	0	0	36	0	0	0	37
	AS	0	0	0	0	0	27	0	0	27
	SL	0	1	0	0	0	0	13	0	14
	FR	0	0	0	0	0	0	0	7	7
	Column total	85	115	36	10	37	27	13	7	330
	Type changes	1	1	3	0	1	0	0	0	6

BL: bareland; CP: cropland; WT: water; WL: wetland; GL: grassland; AS: artificial surfaces; SL: shrubland; FR: forest.

Table 2: Land cover change matrix from 2010 to 2017

Year	2010									
	Type	**BL**	**CP**	**WT**	**WL**	**GL**	**AS**	**SL**	**FR**	**Row total**
2017	BL	84	0	2	0	2	0	0	0	88
	CP	0	112	0	0	0	0	0	0	112
	WT	0	0	33	0	0	0	0	0	33
	WL	0	0	1	10	0	0	0	0	11
	GL	1	0	0	0	35	0	0	0	36
	AS	0	0	0	0	0	27	0	0	27
	SL	0	1	0	0	0	0	13	0	14
	FR	0	0	0	0	0	0	0	9	9
	Column total	85	113	36	10	37	27	13	9	330
	Type changes	1	1	3	0	2	0	0	0	7

BL: bareland; CP: cropland; WT: water; WL: wetland; GL: grassland; AS: artificial surfaces; SL: shrubland; FR: forest.

Table 3: Land cover change matrix from 2016 to 2017

Year	2016									
	Type	**BL**	**CP**	**WT**	**WL**	**GL**	**AS**	**SL**	**FR**	**Row total**
2017	BL	86	0	0	0	0	0	0	0	86
	CP	0	114	0	0	0	0	1	0	115
	WT	0	0	33	0	0	0	0	0	33
	WL	0	0	0	11	0	0	0	0	11
	GL	0	0	0	0	37	0	0	0	37
	AS	0	0	0	0	0	27	0	0	27
	SL	0	1	0	0	0	0	13	0	14
	FR	0	0	0	0	0	0	0	7	7
	Column total	86	115	33	11	37	27	14	7	330
	Type changes	0	1	0	0	0	0	1	0	2

BL: bareland; CP: cropland; WT: water; WL: wetland; GL: grassland; AS: artificial surfaces; SL: shrubland; FR: forest.

Land-cover data preprocessing is required prior to consistency analysis and mainly includes data clipping, projection conversion, and unifying the resolution and classification systems among the different products. Using ArcGIS software [29], the land-cover dataset of the research area was cropped by the Sindh region vector boundary data. The coordinates of all land-cover data were unified into Universal Transverse Mercator (UTM) projection, WGS84 coordinate system. The three land cover datasets selected in this study have different spatial resolutions, the resampling method is used to unify their spatial resolution. In the sampling process of this method, the pixel value with the highest frequency in the window

Table 4: Characteristics of three datasets

Dataset	Resolution (m)	Base year	Classification number	Classification method	Publication organization	Sensor
GlobeLand30	30	2010	10 classes	POK (based on pixels, objects, and knowledge rules)	National Geomatics Center of China	Landsat TM, ETM+, HJ-1A/B
FROM-GLC	30	2017	10 classes	Random forest	Tsinghua University	Landsat TM/ETM+
RLC	16	2016	10 classes	Multi-feature adaptive classification model based on geological understanding	Institute of Geographic Sciences and Natural Resources Research, CAS	GF-1

range is assigned to the output grid. This method can not only maintain the original raster data-type value but also has a fast processing speed and is easy to be applied to those discrete data representing a classification or a thematic, such as land cover data and vegetation type data.

Various land classification systems, many of which use remote sensing to obtain the characteristics of surface features, currently exist [30–33]. These classification systems are well suited for different application needs. But, the classification system of different datasets is not uniform [34,35]. Therefore, this study standardized the codes of the same land-cover types in different products to establish a unified classification system. The spatial distribution map of the cover types of the three land-cover data after preprocessing is shown in Figure 3. The original classification and standardized classification systems for the selected data are shown in Tables 5 and 6, respectively.

TEOW (https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world) is a biogeographic division based on terrestrial biodiversity. Each ecological zone has a unique code, encoded as six digits, that can be interpreted as follows: the first two digits represent the geographical zone, the middle two are the biome types, and the last two are the natural property identifiers. Figure 4 and Table 7 show the locations, codes, and names of the five terrestrial ecological zones included in the study area. The PA1307 ecological zone belongs to the palearctic biogeographic realm, deserts, and xeric shrubland biomes, with the main type being dry shrubbery. The other four ecological zones belong to the Indo-Malay Biogeographic realm. IM1303 and IM1304 are deserts and xeric shrubland biomes, which primarily include brambles, shrubs, and sparse arid grasslands. The IM1403 ecological zone is located in the Indus delta, which runs through the dry Thar Desert of Pakistan and into the Arabian sea. The IM1403 ecoregion belongs to the Mangrove Biome. The IM0901 ecological zone is located at the end of the Luni River and eventually dissipates in the arid salt flats represented by this ecological zone. Grassland and dry spiny shrubs were the main land cover types in this ecological zone. The spatial distribution map of the three land-cover data corresponding to the five terrestrial ecological zones is shown in Figure 5.

3 Methods

We used the overlay analysis method to analyze the types and overall consistency between the different data

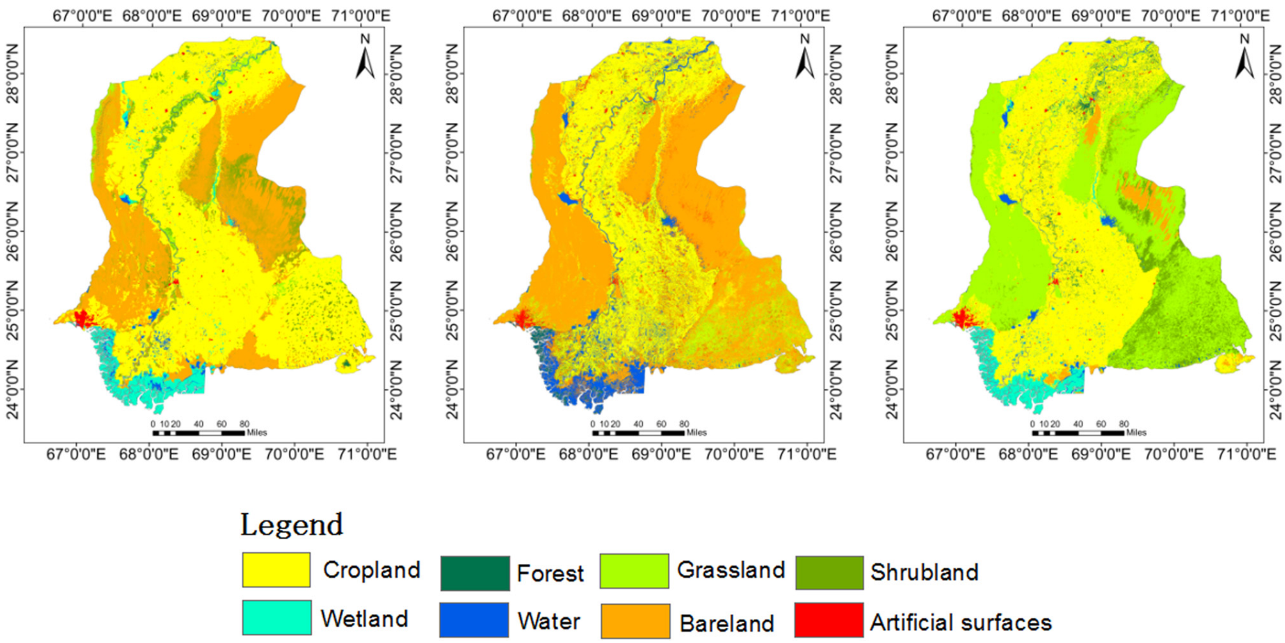

Figure 3: The spatial distribution map of three land-cover data, GlobeLand30, FROM-GLC, and RLC data from left to right, of the study area.

Table 5: Original classification systems of three datasets

Type ID	GlobeLand30	Type ID	FROM-GLC	Type ID	RLC
10	Cropland	1	Cropland	10	Cropland
20	Forest	2	Forest	20	Forest
30	Grassland	3	Grassland	30	Shrubland
40	Shrubland	4	Shrubland	40	Grassland
50	Wetland	5	Wetland	50	Wetland
60	Water bodies	6	Water	60	Water
70	Tundra	7	Tundra	70	Artificial surfaces
80	Artificial surfaces	8	Impervious surfaces	80	Bareland
90	Bareland	9	Bareland	90	Permanent snow
100	Permanent snow and ice	10	Snow/ice	100	ice

Table 6: Standardized classification system for three land-cover data included in the study area

Type ID	Name	GlobeLand30	FROM-GLC	RLC
1	Cropland	10	1	12
2	Forest	20	2	20
3	Grassland	30	3	40
4	Shrubland	40	4	30
5	Wetland	50	5	50
6	Water	60	6	60
7	Artificial surfaces	80	8	70
8	Bareland	90	9	80

products. The process is shown in Figure 6. First, in order to quantitatively compare the consistency of the three land-cover datasets, the area consistency in the different data types was analyzed throughout the entire study area. Second, since area consistency analysis cannot depict the spatial distribution of the same land-cover type among different datasets, the spatial overlay method was adopted to analyze the different types and the overall spatial consistency distribution among the different data in the entire study area. Then, in order to research the distribution rules of the consistent and inconsistent areas throughout the study area, the same method was adopted for the consistency analysis of the three land-cover datasets in different ecological zones. Ultimately, its combination with high-resolution remote sensing image analysis leads to major factors of inconsistency. Moreover, based on the raster calculator of Spatial Analyst under the ArcGIS platform, we analyzed the spatial consistency of the same types in three

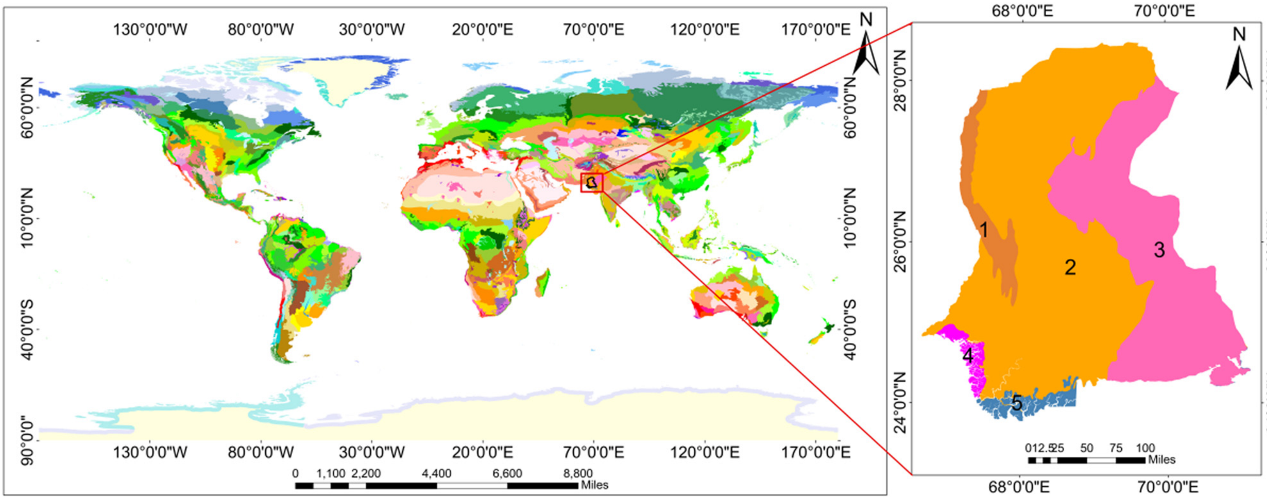

Figure 4: TEOW data in the study area.

Table 7: Code and name of TEOW data included in the study area

Zone ID	Type ID	Name
1	PA1307	Baluchistan xeric woodlands
2	IM1303	Northwestern thorn scrub forests
3	IM1304	Thar desert
4	IM1403	Indus River Delta-Arabian Sea mangroves
5	IM0901	Rann of Kutch seasonal salt marsh

land-cover datasets throughout the entire study area and different ecological zones.

The specific process was as follows. First, the same land-cover types in different land cover products were superimposed. Then, the 30 m × 30 m pixels spatial resolution of the land cover product as the minimum unit and the pixel-by-pixel statistics were used to determine whether the different product definitions were the same at each grid point by pixel. For example, when analyzing the spatial map of forest type, the forest type values in the three land-cover products were set to 1 and the other types were set to 0 by the reclassification method. Then, the forest land types of three kinds of

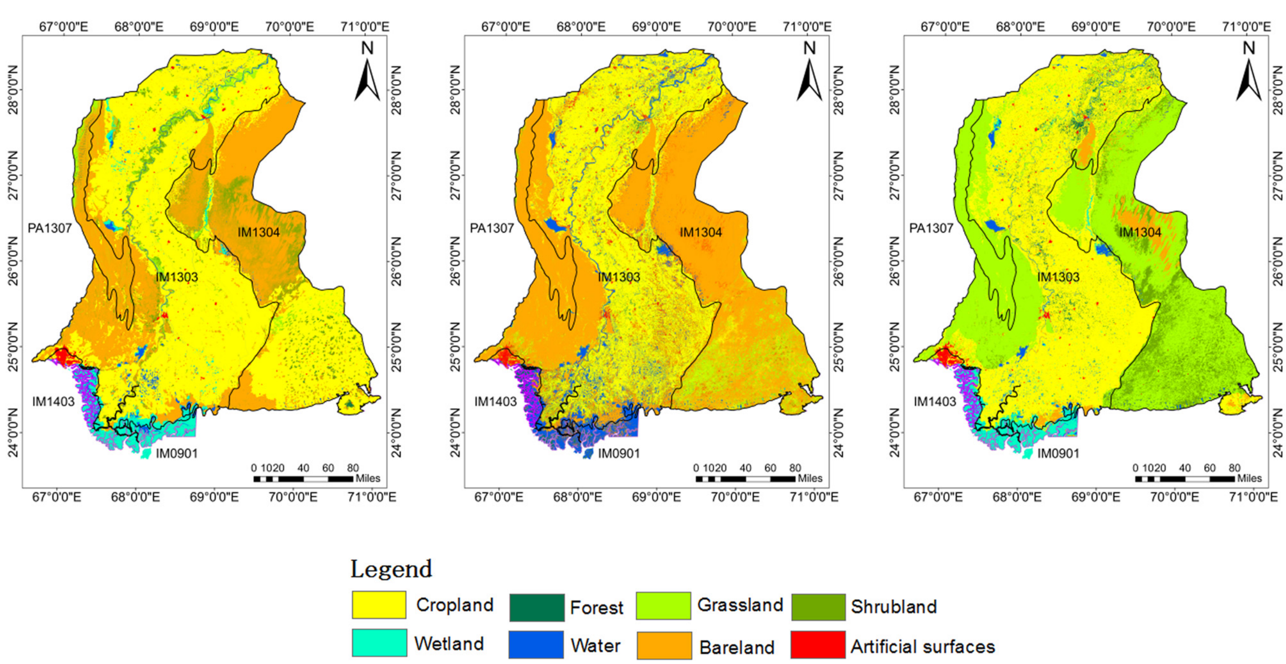

Figure 5: The spatial distribution map of three land-cover data, GlobeLand30, FROM-GLC, and RLC data from left to right, in different ecological zones.

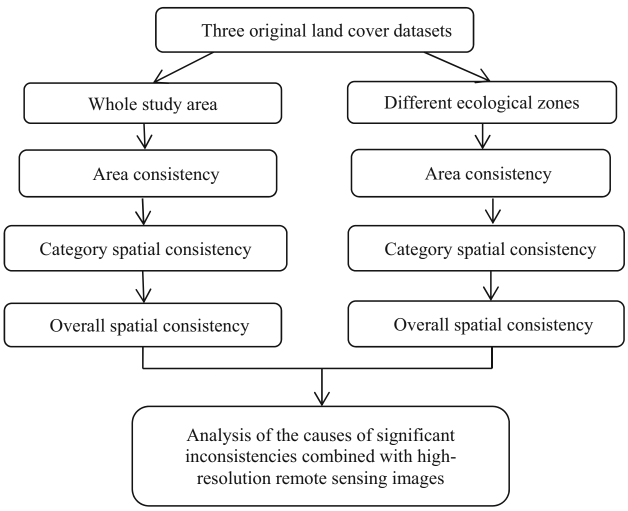

Figure 6: Overall analysis flowchart.

land-cover data are calculated based on pixel superposition by a grid calculator. In the final output, if the pixel value is 3, it indicates that the three land-cover products on this grid point are uniformly defined as forest. If the pixel value is 2, it indicates that any two land-cover products on this grid point are defined as forest. If the pixel value is 1, it indicates that the three land-cover products on this grid point are not forest. The calculation process of other types of spatial consistency distribution is consistent with the above-mentioned forest type.

Based on the obtained consistency maps of each category, according to the same number of land cover types, the degree of consistency between the three products was divided into three levels from low to high: completely inconsistent (the three product indication types are different), basically consistent (the two product indication types are the same), and completely consistent (the three product indication types are all the same). In addition, the main research purpose is to try to study the overall inconsistency of land cover products from the perspective of differences in different ecological zones, so as to provide a basis for the introduction of ecological zones to improve the interpretation accuracy of land cover. Therefore, the error caused by the mixed pixels is not considered in the spatial consistency analysis.

4 Results

4.1 Spatial consistency throughout the entire study area

4.1.1 Consistency analysis of the land-cover area

Area is an important attribute in land-cover products and is of more practical significance to compare the area of three land-cover data. Figure 7 shows the type area composition of the three land-cover data in Sindh, Pakistan.

According to Figure 7, GlobeLand30, FROM-GLC, and RLC all reflect that cropland is the dominant land-cover type in study area, and its area proportion in three datasets is 54.37%, 34.77%, and 42.71%, respectively.

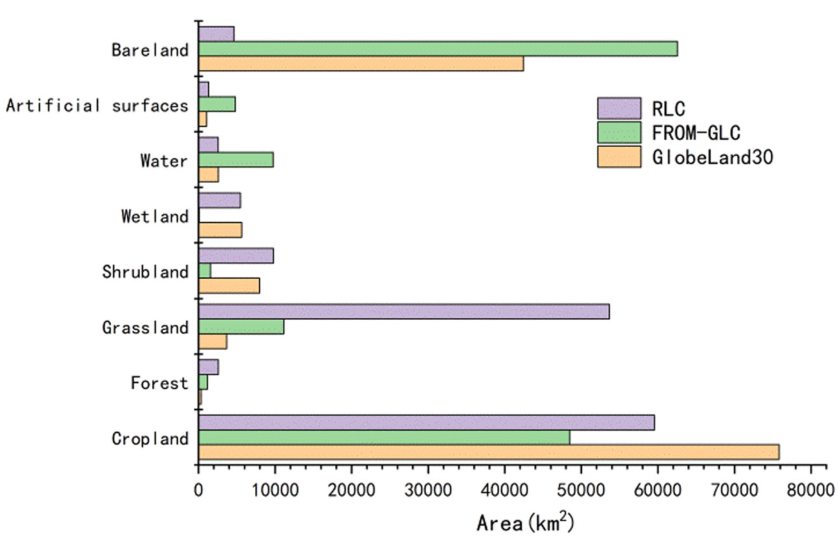

Figure 7: The comparison of the land-cover type area.

The area consistency of each cover type between the three products is low. In the GlobeLand30 and FROM-GLC data, there is a high area consistency between bareland at 30.41% and 44.85% and grassland at 2.62% and 7.97%, respectively. The other types of GlobeLand30 and FROM-GLC data are less consistent. In the GlobeLand30 and RLC data, the consistency of artificial surfaces, water, wetland, and shrubland area is relatively high, while the consistency of grassland and bareland is relatively low. The difference between grassland and bareland is about 35% and 24%, respectively. In the FROM-GLC and RLC data, the consistency of cropland and forest area is good, while the consistency of the other types is poor, particularly that of bareland, grassland, and wetland. The proportion of the bareland area in the FROM-GLC data is about 11 times that of the RLC data, while the proportion of grassland area in the RLC data is about five times that of the FROM-GLC data.

4.1.2 Category consistency and overall consistency analysis

For analyzing the spatial distribution patterns of the consistency degree of the land-cover types, a raster-by-raster comparative analysis was carried out for the land-cover types in Sindh. Cropland (Figure 8a) is the dominant type in Sindh, and the differences among remote sensing land-cover products exhibited high consistency in identifying cropland. The completely consistent region accounted for 27.7% of the entire study area, mainly distributed in the Indus plain region in central Sindh province, which is the main plain agricultural area and the main food growing region. The completely inconsistent region for cropland type in the three datasets was mainly distributed in Tharparkar, the southeast of Sindh province, accounting for 18.2% of the entire study area.

The forest-type consistency in the three land-cover data was relatively low, with the completely consistent area only accounting for 0.00015% of the entire study area, while the basically consistent area accounted for 0.17% of the total study area and was mainly distributed in the central area of Sindh province (Figure 8b). The completely inconsistent area of forest types was predominantly distributed near the Arabian Sea in southwest Sindh province, around the Indus River in northern Sindh province, and in central Sindh province in three datasets. The different land-cover data differed greatly in the consistency of forest cover.

Different satellite remote sensing land-cover products have a low consistency in grassland identification (Figure 8c). Only a few areas in northern Sindh province in the three products were indicated as grassland, accounting for 0.067% of the total study area. The basically consistent and completely inconsistent grassland types of the three land-cover data were mainly distributed in the southeast of the study area and the east and west of the Indus River, among which the completely inconsistent area accounted for a larger proportion. On the whole, there were great differences in the identification of grassland in three datasets.

Figure 8d presents shrubland spatial consistency of the three land-cover data. The completely inconsistent type of the three datasets accounted for 11.89%, mainly distributed in the vicinity of the Indus River and the eastern and southeastern parts of Sindh Province, with less distribution in other areas. Completely consistent and basically consistent areas occupied less area and were mainly distributed in the eastern and southeastern Sindh province.

In Figure 8e, the basically consistent area of wetland types in the three land-cover data accounted for the largest proportion (3.02%) and was mainly distributed around the Arabian Sea in the south of Sindh province. Completely consistent and completely inconsistent areas of wetland of the three datasets accounted for less area, mainly distributed in the Indus plain area in the south and north of the study area, and the overall consistency was low.

As indicated in Figure 8f, the consistency of the water types of the three datasets was relatively low. The completely consistent area accounted for 0.6% of the total area of the study area, distributed in the Indus River and some lakes. The completely inconsistent water types of the three data products were mainly concentrated near the Arabian Sea in the southern of the study area and were less concentrated in the Indus River and surrounding plains.

In Figure 8g, the completely consistent areas of artificial surface of the three data products were mainly distributed in the southwest of the study area, accounting for 0.33% of the whole area. The completely inconsistent areas were mainly scattered over the Indus plain.

The bareland spatial consistency of the three datasets was relatively low (Figure 8h). The completely consistent area was mainly located in the eastern Thar Desert, accounting for 2.56% of the total area. The basic consistency and complete inconsistency of the three land-cover datasets of bareland types have significant regional distribution. The basic consistent areas were mainly distributed in the east and west of Sindh

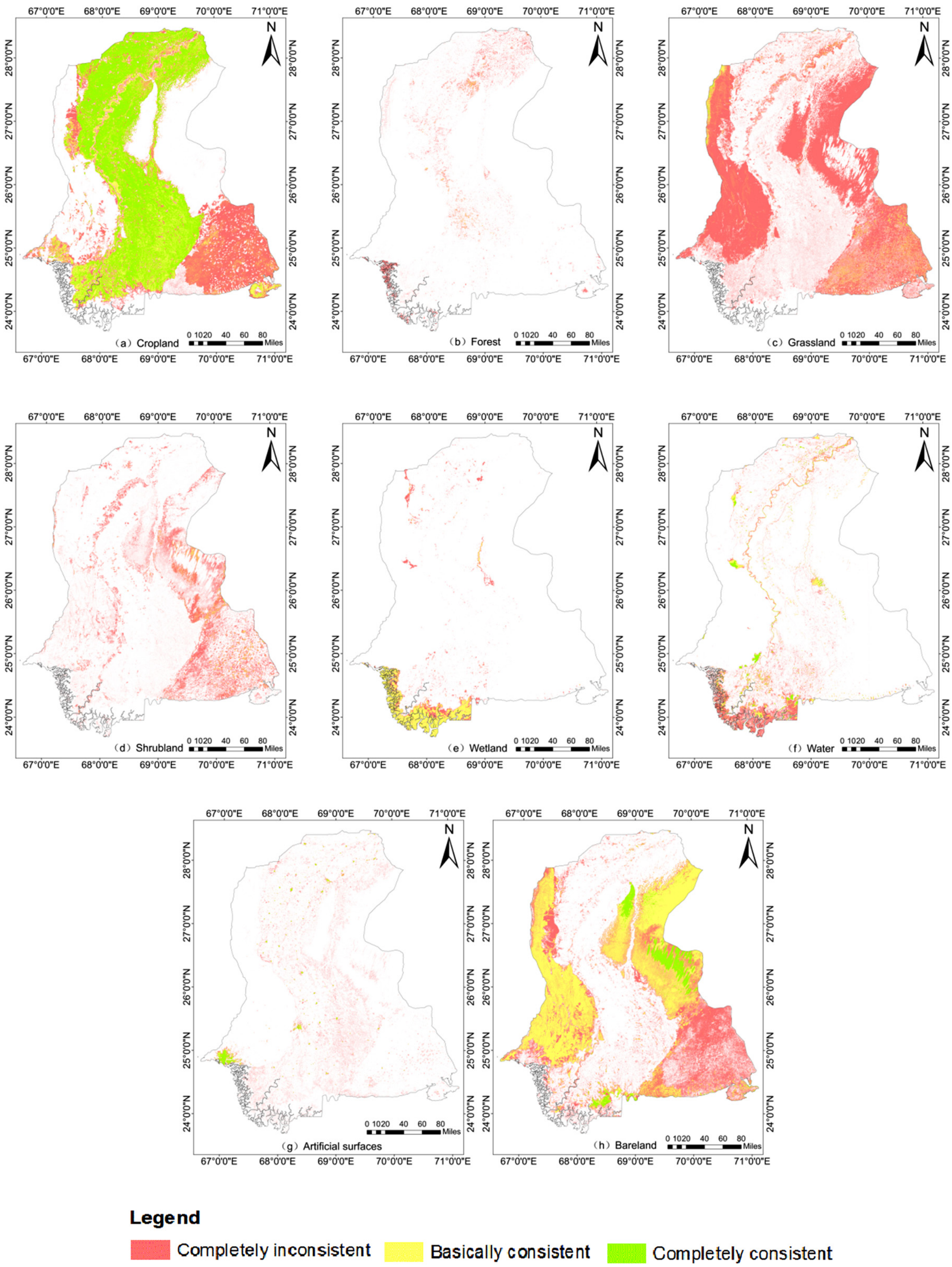

Figure 8: Spatial consistency distribution map of different types in the study area. (a) Cropland, (b) Forest, (c) Grassland, (d) Shrubland, (e) Wetland, (f) Water, (g) Artificial surfaces and (h) Bareland.

Province, accounting for 24.98% of the total area. The completely inconsistent areas were mainly distributed in the southeastern part of Sindh Province, accounting for 20.93% of the total area.

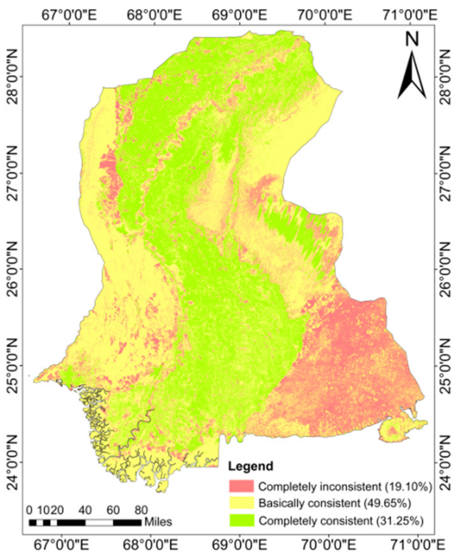

Figure 9: Spatial consistency distribution map of all land covers in the study area.

Based on the spatial consistency analysis of a single land-cover type, the overall consistency analysis of the three datasets in the study area was assessed (Figure 9). The completely consistent areas of the three datasets were in the Indus Plain, accounting for 31.25% of the entire study area and mainly constituted cropland. The basically consistent areas in the three datasets were mainly distributed in the eastern Thar Desert and the western plain area of the study area, accounting for 49.65% of the total study area, being mainly composed of bareland, shrubland, and grassland. The completely inconsistent areas of the three datasets were mainly distributed in the Thar Desert area in the southeast of the study area, accounting for 19.10% of the entire study area and mainly including cropland, bareland, shrubland, and grassland. In general, the overall consistency of the three datasets in the study area was low.

4.2 Spatial consistency of TEOW

4.2.1 Area consistency analysis in different ecological zones

Figure 10 shows the area composition of the following five terrestrial ecological zones: PA1307, IM1303, IM1304, IM1403, and IM0901. For the PA1307 ecological zone, grassland and bareland constituted the main types, and their consistency among the three datasets was low. For the PA1307 ecological zone, the proportion of grassland area in the total area in GlobeLand30, FROM-GLC, and RLC was 12.52%, 2.47%, and 97.1%, respectively, and the proportion of bareland area was 80.3%, 95.1%, and 0.08%, respectively. The consistency for the grassland and bareland between the GlobeLand30 and FROM-GLC data was higher. For the IM1303 ecological zone, cropland was the main land-cover type. The cropland area of the GlobeLand30, FROM-GLC, and RLC data accounted for 69.38%, 53.45%, and 68.52% of the total area of the IM1303 ecological zone. The grassland and bareland types of the three datasets differed greatly. The bareland area of the GlobeLand30 and FROM-GLC data was higher than that of the grassland, while the RLC data demonstrated the opposite. For the IM1304 ecological zone, the consistency of the three datasets, including forests, wetlands, water, and artificial surface, was high, while the consistency of the other types was low. The bareland type exhibited the greatest difference. The largest area of bareland in the total area of the IM1304 ecological zone was from the FROM-GLC data, which was 74.89%, followed by the GlobeLand30 data at 46.76% and RLC data at 6.26%. For the IM1403 and IM0901 ecological zones, the consistency of the wetland types of the three datasets was relatively low. The wetland area of the GlobeLand30 and RLC data was much greater than the water area, which indicates a better consistency between them. However, the wetland area in the FROM-GLC data was much smaller than that of water, and the consistency with other two land-cover data products was low.

4.2.2 Category consistency in different ecological zones

In order to obtain the spatial distribution consistency characteristics for each type in different terrestrial ecological zones, a raster-by-raster comparative analysis was conducted on the land-cover types of the five different ecological zones. Figure 11 shows the spatial consistency map of typical categories in five ecological zones. For cropland types, the consistency of the three data products was relatively high, being mainly distributed in the Indus plain in the IM1303 ecological zone. However, in the southeast of ecological zone IM1304, the consistency of the three data cropland types was low. For grassland types, the consistency of the three data products was relatively low, and the completely inconsistent areas were mainly located in the north of the IM1304 ecological zone, the west of the IM1303 ecological zone, and most of the PA1307 ecological zone. For shrubland types, the three data inconsistencies were mainly distributed throughout the IM1304 ecological

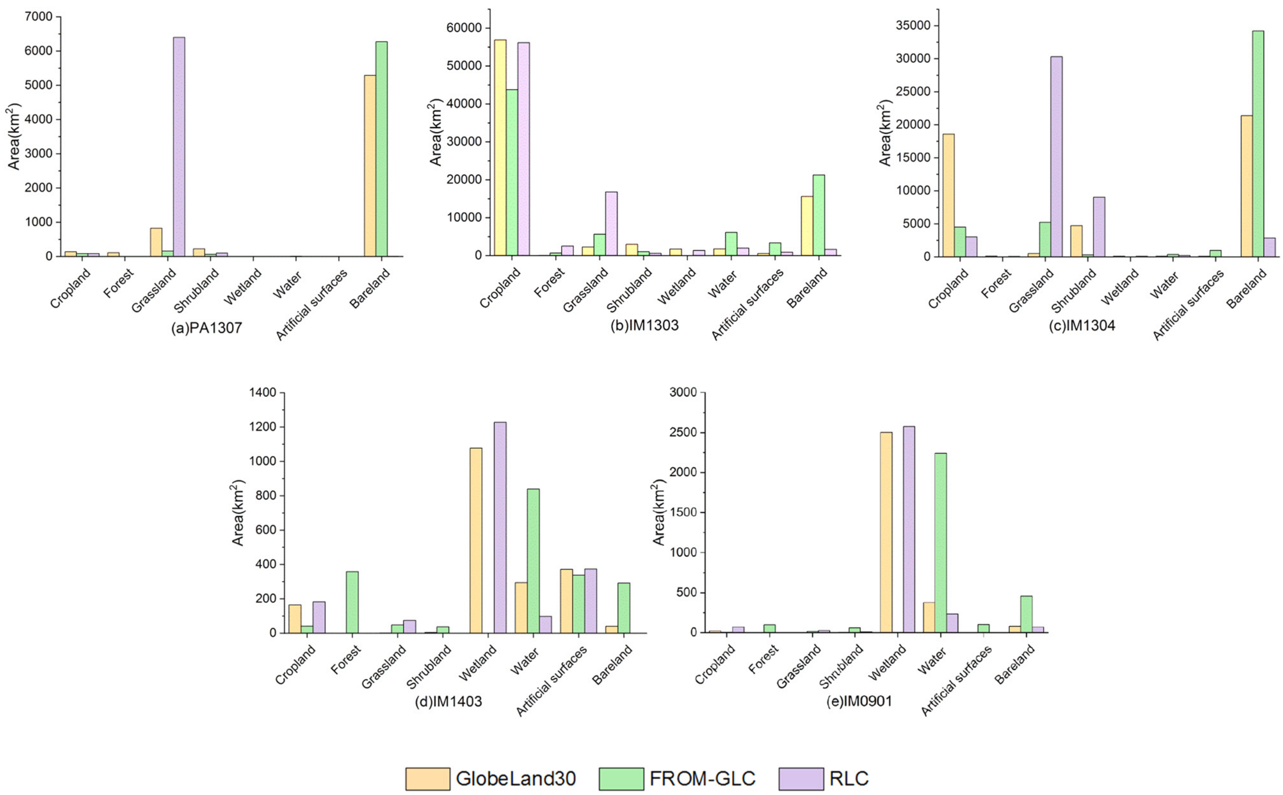

Figure 10: The land-cover type area of different ecological zones. (a) PA1307, (b) IM1303, (c) IM1304, (d) IIM1403 and (e) IM0901.

zone and near the Indus River in the IM1303 ecological zone. For bareland types, the three land-cover data inconsistencies were mainly distributed in the south-eastern part of the IM1304 ecological zone. In the northern part of the IM1304 ecological zone, the western part of the IM1303 ecological zone, and most of the PA1307 ecological zone, there were two types of bare land types. For wetland and water types, the consistency among the three datasets was low, and the inconsistent area was mainly distributed in IM1403 and IM0901 ecological zones.

Figure 12 shows the consistent area statistics for the selected datasets in different ecological zones. The results show that the consistency of the grassland and bareland types in the PA1307 and IM1304 ecological zones was low. The completely inconsistent area of the grassland type accounted for 85.79% and 66.8% of the entire area of the respective zones, while the completely inconsistent area of the bareland type accounted for 17.67% and 38.12% of the total area of the respective zones. The consistency of wetland and water type in ecological zones IM1403 and IM0901 was low. The completely inconsistent area of the wetland type accounted for 16.73% and 16.16% of the total area of the respective zones. In the IM1303 ecological zone, the area of grassland, cropland, and bareland in the completely inconsistent areas accounted for the most. The confusion phenomenon between these types was severe.

4.2.3 Overall consistency in different ecological zones

Based on the category spatial consistency analysis, the overall consistency of the three land cover products in five different terrestrial ecological zones was analyzed. According to Figure 13, the consistency of the different ecological zones in Sindh province differed. The highest overall consistency for the selected datasets was mainly in the IM1303 ecological zone in the Indus Plain area, with cropland constituting the main cover type. The consistency of the three land-cover data products in the IM1304 ecological zone was the lowest, mainly being distributed in the southeast of the IM1304 with grassland, bareland, and cropland as the main cover types. For the PA1307, IM1403, and IM0901 ecological zones, the consistency of the three products was relatively low. In most areas in these ecological zones, there were two land-cover products with the same results, with the main types being bareland, wetland, and water.

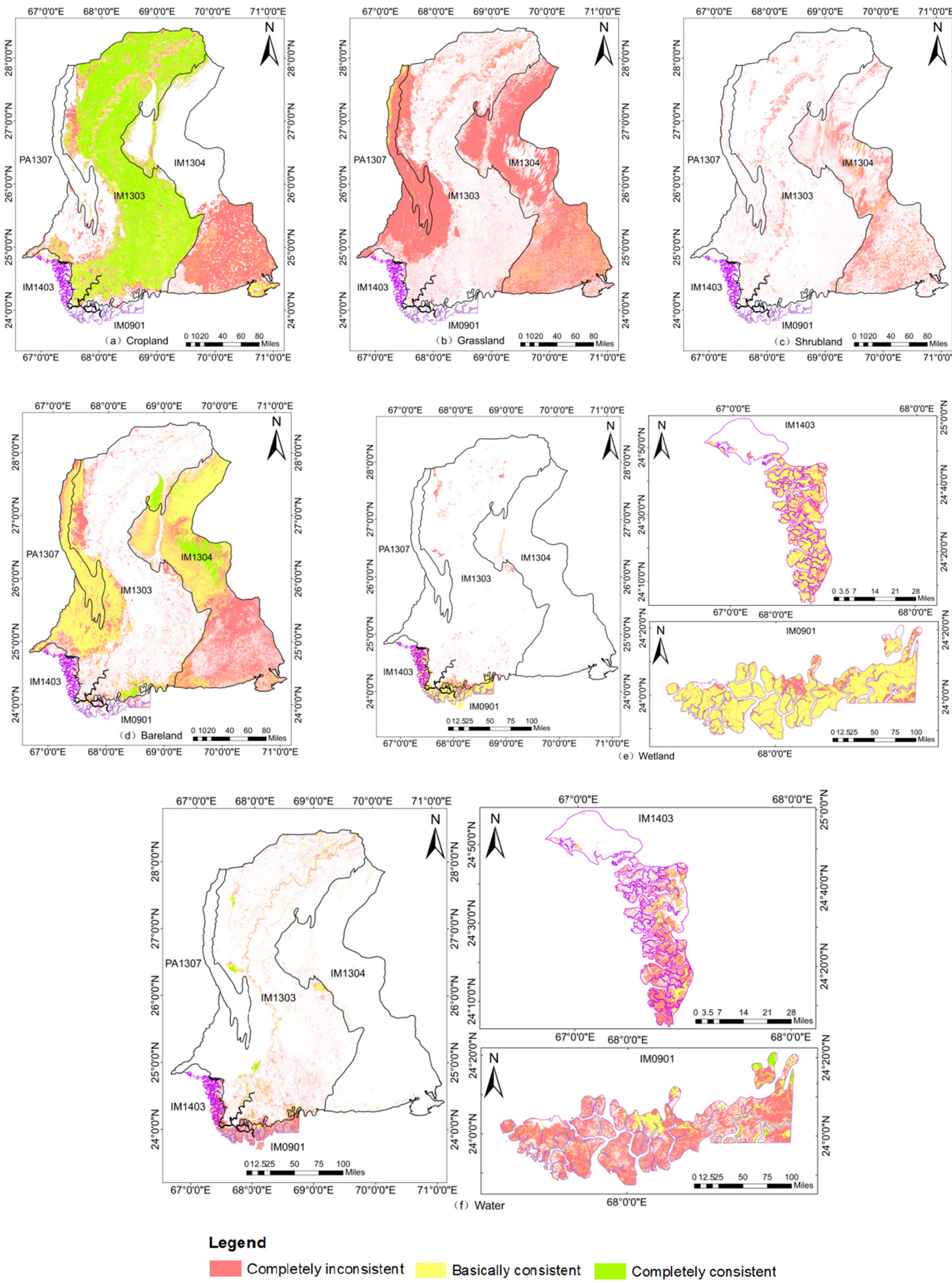

Figure 11: Category consistency distribution map in different ecological zones. (a) Cropland, (b) Grassland, (c) Shrubland, (d) Bareland, (e) Wetland and (f) Water.

The statistical results of the overall consistent area of the five terrestrial ecological zones (Figure 14) show that the completely consistent area of the IM1303 zone accounted for 47.64% of the entire area, which was higher than the basically consistent and completely inconsistent areas, while the other ecological zones indicated the opposite.

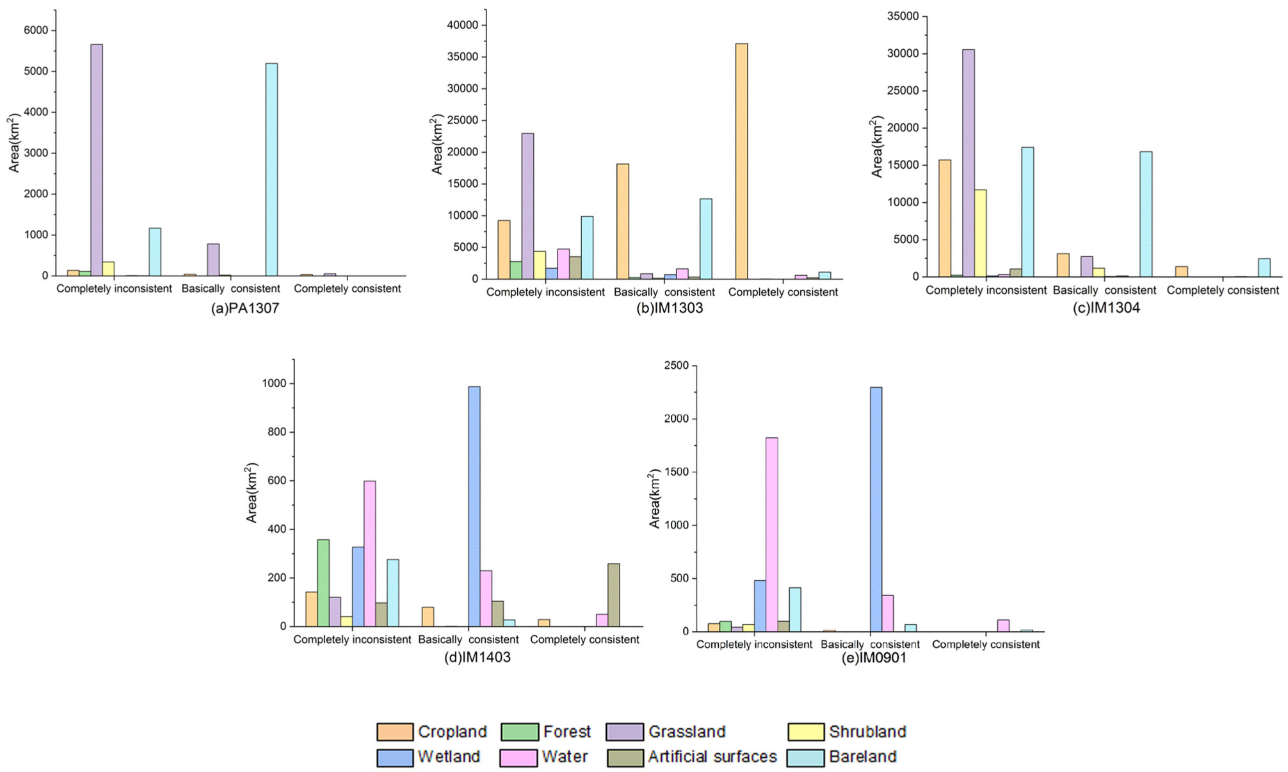

Figure 12: Category consistency area in different ecological zones. (a) PA1307, (b) IM1303, (c) IM1304, (d) IIM1403 and (e) IM0901.

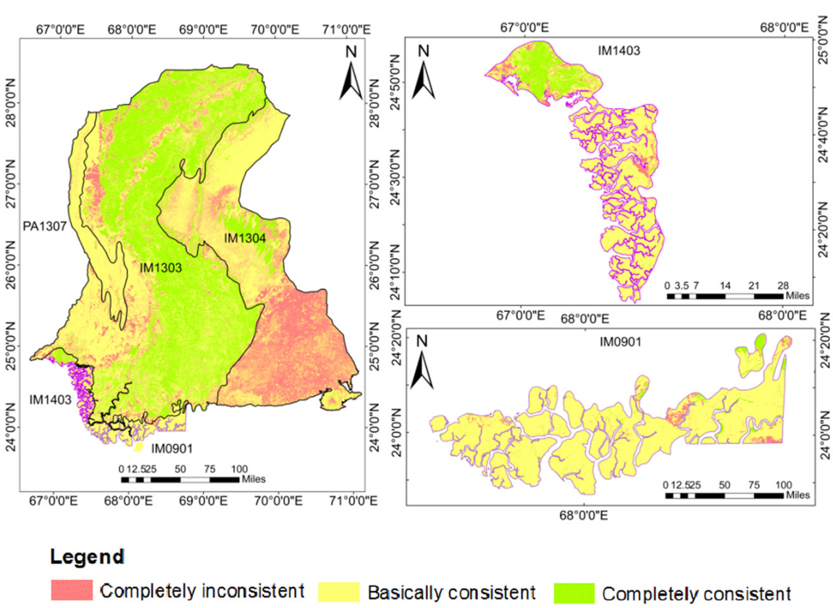

Figure 13: Overall consistency distribution maps in different ecological zones.

The completely consistent areas of the five terrestrial ecological zones from high to low were in the following order: IM1303 > IM1304 > IM1403 > IM0901 > PA1307; the relatively consistent areas were IM1303 > IM1304 > PA1307 > IM0901 > IM1403, and the completely inconsistent areas were IM1304 > IM1303 > PA1307 > IM1403 > IM0901. It can thus be concluded that differences exist in the consistency of the different ecological zones.

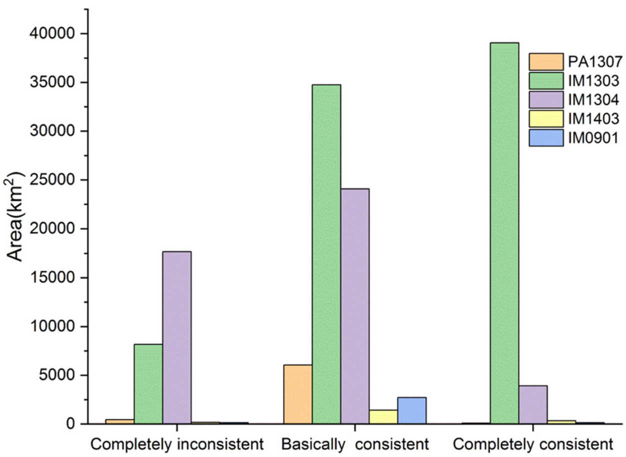

Figure 14: Overall consistency area in different ecological zones.

5 Discussion

Through the above consistency analysis, we determined that the spatial consistency among these land-cover datasets is relatively low. Even though the date of data acquisition has a certain impact on the results of land cover change comparison [36,37], it is not a major factor, because natural ecosystems usually change over 10 years or more. In the land-cover mapping classification, the classification system differences and the subordinate definition of the land-use types between datasets are the main factors affecting the classification accuracy. The inconsistency of the classification system will make the assessment of multi-source remote sensing land-cover data difficult [38,39]. The selected three land-cover products utilize their respective defined classification systems and different definitions for individual land-cover types. For example, in the GlobeLand30 data, grassland types are defined as all grasslands with vegetation coverage greater than 10%, while bareland types are defined as natural land with vegetation coverage lower than 10%. In contrast, the definition of grassland type by the RLC data includes not only grasslands with vegetation coverage greater than 10% but also shrub grasslands predominantly composed of grasslands and sparse forest grasslands with forest coverage below 10%. The data are limited to less than 10% for bareland-type vegetation coverage. However, the FROM-GLC data do not provide a detailed vegetation coverage range for grassland and bareland types. Therefore, the spatial consistency for three datasets is relatively low in the southeast of the ecological zone IM1304 and the northwest of IM1303 (Figure 13). The IM1303 and IM1304 ecological zones constitute arid shrub and desert biomes in the 14 biomes of the global ecological zone. In the process of land-cover interpretation, the subordinate definition of grassland and bareland types in this ecological zone is inaccurate, which will lead to major uncertainty between grassland and bareland types. Therefore, the classification principles for these datasets and the designation for the land-cover types need a deeper study to reduce the inconsistency between the different land-cover datasets.

The spectral information characteristics among different land-cover types fluctuate with seasonal changes [40,41]. The original remote sensing images obtained by the different land-cover data production institutions may differ between months. For example, if the selected image was obtained in the dry season, it will have a certain impact on the identification of water type. If the selected images were acquired during the non-growing vegetative period, the recognition of vegetation types will be reduced. The study area in this research is arid. In the southeast of ecological zone IM1304, the images of the different seasons will affect the interpretation of cropland, bareland, and grassland types, which is also a factor affecting the low spatial consistency for three selected datasets in this ecological zone. Furthermore, the ecological zone IM1303 is traversed by the flooding of the Indus River, resulting in Sindh experiencing extremely dangerous flood hazard levels [42,43]. This directly affects the interpretation accuracy of the water body and its surrounding vegetation types in the IM1303 ecological division, which in turn affects the consistency in different land-cover datasets.

The three datasets adopt different classification methods, which is also an important factor leading to large spatial differences between the different land-cover datasets [44,45]. The GlobeLand30 data adopt the pixel-object-knowledge method, which mainly includes three steps of pixel method classification, object filtering, and human–computer interaction checking. The purpose is to highlight the advantages of various classification algorithms and fully utilize the knowledge and human experience to improve classification quality. Besides, the classified images, a large amount of auxiliary data and reference materials are used, including the existing surface coverage products (global, regional), MODIS NDVI data, global basic geographic information data, global DEM data, and various thematic data and online high-resolution images (such as Google Earth high-resolution images, Bing images, and OpenStreet high-resolution images). The FROM-GLC data divided the world into 16 regions and trained one classifier for each region and applied it to classify the images. The main process of classification is digital elevation model (DEM) data as auxiliary data. However, in the process of interpretation, the impervious type cannot be well differentiated from bareland or other vegetation classes due to the nature of the surface cover complexity in human settlement areas and the high proportion of spectrally mixed pixels. Therefore, nighttime data are introduced as auxiliary data. The RLC data adopt a multi-feature adaptive classification model based on geological knowledge, and DEM, MODIS, and NDVI data are used as auxiliary data. The differences in classification methods and auxiliary data among these datasets led to the low consistency in the southeast of the ecological zone IM1304. Meanwhile, due to the differences in classification methods, the FROM-GLC data mainly interpreted the IM1403 and IM0901 ecological zones as water type, while the GlobeLand30 and RLC products mainly interpreted these as wetland type, resulting in the low consistency of the three datasets in the IM1403 and IM0901 ecological zones. In addition, when the different datasets were assessed by the human–computer interaction, the difference in people's earth knowledge was also a factor affecting the final product quality.

In addition to the aforementioned internal factors, there are also some external factors that reduce the consistency among these datasets, including map projection, uniform resolution, and classification system merge. These operations will lead to the loss or misinterpretation of the original land-cover information, affecting the comprehensiveness and consistency of the classification results [46,47]. The subsequent classification of product developers should pay attention to the regions with high land-cover complexity to address the issue of the spectral mixing of forest, grassland, and shrubland types, thus further improving the classification accuracy.

6 Conclusions

Using the Sindh province in Pakistan as an example in TEOW, we analyzed the inconsistency of the GlobeLand30, FROM-GLC, and RLC land-cover data. The results showed that (1) the consistency among the three land-cover data across the entire Sindh province was low, with the completely consistent area only accounting for 31.25% of the entire study area. The consistent areas were mainly distributed in IM1303 ecological zones with a single structure of land-cover type and obvious homogeneity of the surface category. (2) The differences exist in the spatial consistency of the different terrestrial ecological zones throughout the study area. The proportion of completely consistent areas in the total study area from high to low was in the following order: IM1303 > IM1304 > IM1403 > IM0901 > PA1307. (3) The definitions of some land-cover types in the classification system adopted by the three land-cover data were different due to different research purposes and scales. For example, the definitions of grassland and bareland types in GlobeLand30 and RLC were different, which leads to the high spatial inconsistency of these land-cover data in the southeast of IM1304 ecological zone with complex surface types (interleaved distribution of cropland, grassland, shrubland, and bareland types) and strong heterogeneity.

The above conclusions inspire us that the existing land-cover data can be adopted with high confidence in ecological zoning with low heterogeneity and obvious dominant land types when using land-cover data. However, the consistency of the existing available datasets is low, which should be used with caution, especially the definition consistency of these land types in ecological zoning with high heterogeneity and diverse land-cover types.

These conclusions also have important enlightenment for the production of land-cover data. For example, we can use a heterogeneity index of remote sensing images to partition remote sensing images, and then keep the results of the existing land-cover products in areas with low heterogeneity, while take targeted measures to improve the interpretation accuracy in areas with high heterogeneity, such as selecting more sample points.

Acknowledgments: This study was supported by the National Key Research and Development Program of China, Grant No. 2016YFB0501404; the CAS Earth Big Data Science Project, Grant No. XDA19060303; the National Science Foundation of China, Grant No. 41671436; and the Innovation Project of LREIS, Grant No. O88RAA01YA.

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

References

[1] Masoudi M, Tan PY. Multi-year comparison of the effects of spatial pattern of urban green spaces on urban land surface

temperature. Landsc Urban Plan. 2019;184:44–58. doi: 10.1016/j.landurbplan.2018.10.023

[2] Middel A, Brazel AJ, Gober P, Myint SW, Duh JD. Land cover, climate, and the summer surface energy balance in Phoenix, AZ, and Portland, OR. Int J Climatol. 2012;32:2020–32.

[3] Schulz JP, Vogel G, Becker C, Kothe S, Rummel U, Ahrens B. Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil. In Proceedings of Egu General Assembly Conference; p. 607–20.

[4] Zhang X, Liu L, Henebry GM. Impacts of land cover and land use change on long-term trend of land surface phenology: a case study in agricultural ecosystems. Environ Res Lett. 2019;14:4. doi: 10.1088/1748-9326/ab04d2.

[5] Hereher ME. Effects of land use/cover change on regional land surface temperatures: severe warming from drying Toshka lakes, the Western Desert of Egypt. Nat Hazards. 2017;88:1789–803. doi: 10.1007/s11069-017-2946-8.

[6] Jolliet O, Anton A, Boulay A-M, Cherubini F, Fantke P, Levasseur A, et al. Global guidance on environmental life cycle impact assessment indicators: impacts of climate change, fine particulate matter formation, water consumption and land use. Int J Life Cycle Assess. 2018;23:2189–207. doi: 10.1007/s11367-018-1443-y.

[7] Liu G, Wang J, Li S, Li J, Duan P. Dynamic evaluation of ecological vulnerability in a lake watershed based on RS and GIS technology. Pol J Environ Stud. 2019;28:1785–98. doi: 10.15244/pjoes/89981.

[8] M'Mboroki KG, Wandiga S, Oriaso SO. Climate change impacts detection in dry forested ecosystem as indicated by vegetation cover change in -Laikipia, of Kenya. Environ Monit Assess. 2018;190:4. doi: 10.1007/s10661-018-6630-6.

[9] Esteban Lucas-Borja M, Zema DA, Antonio Plaza-Alvarez P, Zupanc V, Baartman J, Sagra J, et al. Effects of different land uses (abandoned farmland, intensive agriculture and forest) on soil hydrological properties in Southern Spain. Water. 2019;11:3. doi: 10.3390/w11030503.

[10] Holmberg M, Aalto T, Akujarvi A, Arslan AN, Bergstrom I, Bottcher K, et al. Ecosystem services related to carbon cycling – modeling present and future impacts in Boreal Forests. Front Plant Sci. 2019;10:343. doi: 10.3389/fpls.2019.00343.

[11] Chen D, Liu Y, Niu T, Chen Z, Li J. Land resources information acquisition based on multi-temporal TM images. In Proceedings of IEEE International Conference on Computer Science & Automation Engineering.

[12] Zhang J, Qisheng HE, Cui T, Shi P, Yang T. Information extraction and dynamic changes of wetland in Jiangsu coastal area based on RS and GIS. Journal of Yangtze River Scientific Research Institute; 2017;34:4.

[13] Zhang J, Zhang J, Du X, Kang H, Qiao M. An overview of ecological monitoring based on geographic information system (GIS) and remote sensing (RS) technology in China. IOP Conf Series: Earth Environ Sci. 2017;94:012056.

[14] Wang Z, Lu C, Yang X. Exponentially sampling scale parameters for the efficient segmentation of remote-sensing images. Int J Remote Sens. 2018;39:1628–54. doi: 10.1080/01431161.2017.1410297.

[15] Wang Z, Yang X, Lu C, Yang F. A scale self-adapting segmentation approach and knowledge transfer for automatically updating land use/cover change databases using high spatial resolution images. Int J Appl Earth Observation Geoinf. 2018;69:88–98. doi: 10.1016/j.jag.2018.03.001.

[16] Chen Z, Yu B, Zhou Y, Liu H, Yang C, Shi K, et al. Mapping global urban areas from 2000 to 2012 using time-series nighttime light data and MODIS products. IEEE J Sel Top Appl Earth Observations Remote Sens. 2019;12:1143–53. doi: 10.1109/jstars.2019.2900457.

[17] Bartholome E, Belward AS. GLC2000: a new approach to global land cover mapping from Earth observation data. Int J Remote Sens. 2005;26:1959–77. doi: 10.1080/01431160412331291297.

[18] Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, et al. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens Environ. 2010;114:168–82. doi: 10.1016/j.rse.2009.08.016.

[19] Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int J Remote Sens. 2013;34:2607–54. doi: 10.1080/01431161.2012.748992.

[20] Chen J, Chen J, Liao A, Cao X, Chen L, Chen X, et al. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J Photogrammetry Remote Sens. 2015;103:7–27. doi: 10.1016/j.isprsjprs.2014.09.002.

[21] Dai Z, Yunfeng HU, Zhang Q. Agreement analysis of multi-source land cover products derived from remote sensing in South America. Remote Sens Inf. 2017;32:137–48.

[22] Hongli S, Xiaonan Z. Precision validation of multi-sources land cover products derived from remote sensing. Remote Sens Land & Resour. 2018;30:26–32.

[23] Giri C, Zhu ZL, Reed B. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. Remote Sens Environ. 2005;94:123–32. doi: 10.1016/j.rse.2004.09.005.

[24] Yang Y, Xiao P, Feng X, Li H. Accuracy assessment of seven global land cover datasets over China. Isprs J Photogrammetry Remote Sens. 2017;125:156–73. doi: 10.1016/j.isprsjprs.2017.01.016.

[25] Bai Y, Feng M, Jiang H, Wang J, Zhu Y, Liu Y. Assessing consistency of five global land cover data sets in China. Remote Sens. 2014;6:8739–59. doi: 10.3390/rs6098739.

[26] Hua T, Zhao W, Liu Y, Wang S, Yang S. Spatial consistency assessments for global land-cover datasets: a comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. Remote Sens. 2018;10:11. doi: 10.3390/rs10111846.

[27] Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, et al. Terrestrial ecoregions of the World: a new map of life on earth. Bioscience. 2001;51:933–8.

[28] Zhang J, Qian Z, Wanggu XU, Zhang H, Wang Z. Ecosystem pattern variation from 2000 to 2010 in national nature reserves of China. Acta Ecologica Sin. 2017;37:8067–76.

[29] Maguire DJ. ArcGIS: General-Purpose GIS Software. 2008.

[30] Hansen MC, Defries RS, Townshend JRG, Sohlberg R. Global land cover classification at 1km spatial resolution using a classification tree approach. Int J Remote Sens. 2000;21:1331–64. doi: 10.1080/014311600210209.

[31] Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, et al. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogramm Eng Remote Sens. 2007;73:337–41.

[32] Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, et al. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int J Remote Sens. 2000;21:1303–30. doi: 10.1080/014311600210191.
[33] Yang H, Li S, Chen J, Zhang X, Xu S. The standardization and harmonization of land cover classification systems towards harmonized datasets: a review. ISPRS Int J Geo-Information. 2017;6:154. doi: 10.3390/ijgi6050154.
[34] Latifovic R, Olthof I. Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. Remote Sens Environ. 2004;90:153–65. doi: 10.1016/j.rse.2003.11.016.
[35] Kang J, Sui L, Yang X, Wang Z, Huang C, Wang J. Spatial pattern consistency among different remote-sensing land cover datasets: a case study in Northern Laos. ISPRS Int J Geo-Information. 2019;8:201. doi: 10.3390/ijgi8050201.
[36] Jung M, Henkel K, Herold M, Churkina G. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens Environ. 2006;101:534–53. doi: 10.1016/j.rse.2006.01.020.
[37] Kaptue Tchuente AT, Roujean J-L, De Jong SM. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. Int J Appl Earth Observation Geoinf. 2011;13:207–19. doi: 10.1016/j.jag.2010.11.005.
[38] Congalton RG, Gu J, Yadav K, Thenkabail P, Ozdogan M. Global land cover mapping: a review and uncertainty analysis. Remote Sens. 2014;6:12070–93. doi: 10.3390/rs61212070.
[39] Herold M, Mayaux P, Woodcock CE, Baccini A, Schmullius C. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1km datasets. Remote Sens Environ. 2008;112:2538–56.
[40] Chen Y. Seasonal patterns of spatial differentiation of landcover change in China. Chin Sci Bull. 1999;44:362–4.
[41] Verbesselt J, Hyndman R, Newnham G, Culvenor D. Detecting trend and seasonal changes in satellite image time series. Remote Sens Environ. 2010;114:106–15.
[42] Gaurav K, Sinha R. The Indus flood of 2010 in Pakistan: a perspective analysis using remote sensing data. Nat Hazards. 2011;59:1815–26.
[43] Kazi A. A review of the assessment and mitigation of floods in Sindh, Pakistan. Nat Hazards. 2014;70:839–64.
[44] Rodriguez-Galiano VF, Chica-Rivas M. Evaluation of different machine learning methods for land cover mapping of a mediterranean area using multi-seasonal landsat images and digital terrain models. Int J Digital Earth. 2014;7:492–509.
[45] Singh AN, Deepthi GV, Singhal A. Land use/land cover information from various classification methods of Shekhawati region of Rajasthan. In Proceedings of International Conference on Electrical.
[46] Feng CC, Flewelling DM. Assessment of semantic similarity between land use/land cover classification systems. Comp Environ & Urban Syst. 2004;28:229–46.
[47] Hui Y, Li S, Chen J, Zhang X, Xu S. The standardization and harmonization of land cover classification systems towards harmonized datasets: a review. ISPRS Int J Geo-Information. 2017;6:154.