Identifying and Quantifying Pixel-Level Uncertainty among Major Satellite Derived Global Land Cover Products

Hao GAO, Gensuo JIA, and Yu FU

1 National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081
2 Key Laboratory of Regional Climate–Environment for Temperate East Asia (RCE-TEA), Chinese Academy of Sciences, Beijing 100029
3 Climate Change Research Center, Chinese Academy of Sciences, Beijing 100029

(Received November 21, 2019; in final form June 4, 2020)

ABSTRACT

Accurate global land cover (GLC), as a key input for scientific communities, is important for a wide variety of applications. In order to understand the current suitability and limitation of GLC products, the discrepancy and pixel-level uncertainty in major GLC products in three epochs are assessed in this study by using an integrated uncertainty index (IUI) that combines the thematic uncertainty and local classification accuracy uncertainty. The results show that the overall spatial agreements (Ao values) between GLC products are lower than 58%, and the total areas of forests are very consistent in major GLC products, but significant differences are found in different forest classes. The misclassification among different forest classes and mosaic types can account for about 20% of the total disagreements. The mean IUI almost reaches 0.5, and high uncertainty mostly occurs in transition zones and heterogeneous areas across the world. Further efforts are needed to make in the land cover classifications in areas with high uncertainty. Designing a classification scheme for climate models, with explicit definitions of land cover classes in the threshold of common attributes, is urgently needed. Information of the pixel-level uncertainty in major GLC products not only give important implications for the specific application, but also provide a quite important basis for land cover fusion.

Key words: global land cover (GLC), discrepancies, thematic similarity, local accuracy, pixel-level uncertainty

Citation: Gao, H., G. S. Jia, and Y. Fu, 2020: Identifying and quantifying pixel-level uncertainty among major satellite derived global land cover products. J. Meteor. Res., 34(4), 806–821, doi: 10.1007/s13351-020-9183-x.

1. Introduction

Land cover plays an important role in the biogeochemistry of earth system, which influences the carbon, moisture, energy, and momentum exchanges between the land and atmosphere (Findell et al., 2007; Sterling et al., 2013). As an essential variable considered by the Global Climate Observing System (GCOS), land cover is used in science and policy applications, such as the climate change, food security, biodiversity conservation, ecosystem assessment, and hydrological and dynamic vegetation modeling.

A number of global land cover (GLC) products have been produced based on the remote sensing data in response to the need for land cover information and dynamics in the past decades (Hansen et al., 2000; Bartholomé and Belward, 2005; Friedl et al., 2010; Bontemps et al., 2011; Gong et al., 2013; Tateishi et al., 2014; Chen et al., 2015; Defourny et al., 2016), which have been widely used in various scientific studies. Many applications have benefited from using satellite derived land cover products as an input variable in the model for improving the model performance. However, the quality of these products has a strong impact on the final model output, which has significant effects on these applications (Ge et al., 2007; Sertel et al., 2010; Santos-Alamillos et al., 2015; Madhusoodhanan et al., 2017). Thus, accurate GLC products are vital for global applications, and associated information of the accuracy and bias is required for better employing GLC products in scientific studies.
Currently, there are several prevalent methods for accuracy and uncertainty assessments of land cover products, including the validation using high-quality reference data with the confusion matrix (Iwao et al., 2006; Zhao et al., 2014; Tsendbazar et al., 2016; Wickham et al., 2017; Yu et al., 2018), and comparison with existing land cover products (McCallum et al., 2006; Fritz and See, 2008; Kaptué et al., 2011; Pérez-Hoyos et al., 2012; Bai et al., 2014) or statistical land cover information (Nakaegawa, 2011; Comber, 2013; Tsendbazar et al., 2015; Tsutsumida and Comber, 2015; Hua et al., 2018). It is shown that despite efforts in advancing land cover mapping, the overall accuracy of the existing GLC products typically varies between 67% and 81%, while the class-specific accuracy varies significantly (Mayaux et al., 2006; Friedl et al., 2010; Gong et al., 2013; Tateishi et al., 2014; Chen et al., 2015; Defourny et al., 2016). Meanwhile, due to the differences in data sources, classification methodologies, and spatial resolution, inevitable discrepancies among these products have been found at both the global and regional scales through inter-comparisons (McCallum et al., 2006; Herold et al., 2008; Gao and Jia, 2012; Pérez-Hoyos et al., 2012; Yang et al., 2017).

Although many efforts have been made to provide valuable quality information of GLC products, it is still unable to fulfill requirements of different users in providing a guide for data usage. For instance, the cross validation with reference data provides only overall and specific accuracy estimations for the entire products, which could not represent the spatial variations (Foody, 2002). In terms of the spatial agreement, the inter comparison with the existing land cover products does not provide information on the comparative accuracy of these products. Recently, several studies have estimated the accuracy for different applications (Fritz and See, 2008; Gao and Jia, 2013; Tsendbazar et al., 2016), analyzed the thematic similarity (Ahlqvist, 2005; Neumann et al., 2007; Pérez-Hoyos et al., 2012), and evaluated the spatial variations in classification accuracy (Foody, 2005; See et al., 2015; Tsendbazar et al., 2015; Comber et al., 2017; Yang et al., 2017). The thematic analysis indicates the discrepancies in the definition of land cover class and positional errors among products, which are often ignored when the classification scheme is translated. The spatial variation assessment informs the detailed spatial variability in the classification accuracy, which is crucial for scientific applications. However, no previous studies, to our knowledge, have evaluated the uncertainty of current GLC products with the combination of both the thematic similarity and spatial variation.

In this study, we develop a methodology to identify the spatial distribution of uncertainty among GLC products that combines both the thematic uncertainty and local classification accuracy uncertainty. First, the thematic uncertainty index (TUI) is calculated based on the thematic similarity analysis, which accommodates both the partial definition overlaps between classes and positional errors by the per-pixel comparison. Second, the local classification accuracy uncertainty index (AUI) is calculated by using the reference dataset via a geostatistical model. Finally, the integrated uncertainty index (IUI) is calculated based on the combination of the thematic uncertainty and local classification accuracy uncertainty to quantify the pixel-level uncertainty of GLC products, offering new insights to understand the spatial distribution of disagreement and implication for both the users and producers of GLC products.

2. Materials and methods

2.1 GLC products

The major GLC products derived from satellite data for three epochs (2000–2005, 2008–2010, and 2013–2015) are used in this study, including the GLC2000 dataset from the European Commission’s Joint Research Center, land cover dataset based on the Moderate Resolution Imaging Spectroradiometer (MODIS) data, GLC map (GlobCover) from European Space Agency (ESA), GLC by National Mapping Organizations (GLCNMO) dataset from the International Steering Committee for Global Mapping, and Climate Change Initiative Land Cover dataset (CCI-LC) from ESA. These products are developed by different national or international initiatives for different scientific purposes, with diverse classification characteristics, which are widely used as a baseline for GLC conditions and dynamics over the past decades. Characteristics of these different products are summarized in Table 1.

We clipped all the GLC products to the overlap extent of the globe, projected, and co-registered them to the same projection and spatial resolution to reduce errors in geo-registration. Since the resolution of the products ranges from 300 m to 1 km, the resampling strategies for the three epochs are not the same, e.g., the products are resampled to a resolution of 1 km for 2000–2005 and 500 m for periods of both 2008–2010 and 2013–2015 by using the nearest neighbor rule.

These GLC products have different classification schemes (Table 1). In order to quantify the uncertainty among these products for climate models, legends associated with each GLC product are translated into a common scheme with 14 land cover classes (Table 2), by us-
ing the Land Cover Classification System (LCCS) legend translation protocols developed by Herold et al. (2008).

### 2.2 Reference datasets

Several existing GLC reference datasets are used in this paper, including the consolidated GLC2000 reference dataset (Schultz et al., 2015), GLCNMO2008 training dataset (Tateishi et al., 2011), consolidated GlobCover2005 reference dataset (Bicheron et al., 2008), System for Terrestrial Ecosystem Parameterization (STEP) reference dataset (Friedl et al., 2010), Visible Infrared Imaging Radiometer Suite (VIIRS) Surface Type validation database (Olofsson et al., 2012), GLC Ground Truth (GLCGT) database, Global Forest Resources Assessment (Global FRA) reference (Potapov et al., 2011), Geo-Wiki validations (Fritz et al., 2009), and a common validation sample set for the Finer Resolution Observation and Monitoring (FROM) of GLC project (Zhao et al., 2014) and Global Flux sites, which are generated by different international organizations through enormous interpretation efforts. All these reference datasets are publicly accessible through the Global Observation of Forest Cover and Land Dynamics (GOFC-GOLD) reference data portal, Geo-Wiki, and Internet, and can be freely used by any researcher. It should be noted that time periods for these reference datasets are discrepant, but errors in the processes of land cover changing over

| GLC product | GLC2000 | MODIS2001/2010/2015 | GlobCover2005/2009 | GLCNMO2003/2008/2013 | CCI-LC2001/2010/2015 |
|-------------|---------|---------------------|--------------------|----------------------|----------------------|
| Satellite sensor | SPOT VGT | MODIS Terra and Aqua | Envisat MERIS | MODIS Terra and Aqua | Envisat MERIS |
| Input data | Daily mosaic of 4 spectral bands | 16-day nadir BRDF adjusted reflectance (1–7), EVI, LST | MERIS: Bi-monthly reflectance from 10-day composites | 16-day nadir BRDF adjusted reflectance (1–7), NDVI, DMSP-OLS, Landsat ETM+ | MERIS: 7-day composite reflectance, SPOT VGT time series data |
| Spatial resolution | 1 km | 500 m | 300 m | 1 km/500 m | 300 m |
| Time of data collection | 1999.11–2000.12 | Calendar year 2001/2010/2015 | 2004.12–2006.6 Calendar year 2009 | 2003/2008/2013 | Calendar year 2001/2010/2015 |
| Classification scheme | LCCS (22 classes) | Multi-scheme including the IGBP (17 classes) | LCCS (22 classes) | LCCS (20 classes) | LCCS (22 classes) |
| Classification method | General unsupervised, depending on the partner | Supervised decision tree boosting | Unsupervised/supervised spatio–temporal clustering expert-based labeling | Supervised decision tree | Combination of supervised and unsupervised; machine learning classification |
| Overall accuracy (%) | Globally 68.6 | Globally 74.8 | Globally 73.1/67.5 | Globally 81.2/82.6 | Globally 73.2 |
| Reference | Bartholomé and Belward (2005), Mayaux et al. (2006) | Friedli et al. (2010) | Bontemps et al. (2011) and Defourny et al. (2011) | Tateishi et al. (2011, 2014) | Defourny et al. (2016) |

Notes: SPOT VGT—Systeme Probatoire d’Observation de la Terrestre Vegetation, BRDF—Bidirectional Reflectance Distribution Function, EVI—Enhanced Vegetation Index, LST—Land Surface Temperature, IGBP—International Geosphere–Biosphere Programme, MERIS—Medium Resolution Imaging Spectrometer, NDVI—Normalized Difference Vegetation Index, DMSP-OLS—Defense Meteorological Satellite Program Operational Linescan System, ETM+—Enhanced Thematic Mapper.
time can be neglected compared to misclassification errors as most of the reference points and scientific sites are located in homogeneous and stable areas. The class of land cover at the centroid of sample sites is considered as the reference land cover class. Legends of all the reference datasets mentioned above are translated into a 14-class common scheme listed in Table 2. Totally, 17,361 reference sites within the globe are adopted in this study. The spatial distributions and corresponding size of each reference dataset as well as reference land cover classes are shown in Fig. 1.

2.3 Comparison analysis

In this study, the area and spatial comparisons are applied. The area comparison shows the total area differences of individual land cover class in the common scheme among different GLC products. The spatial comparison is based on the per-pixel comparison among GLC products, and the overall spatial agreement ($A_o$) as well as class-specific spatial agreement ($A_s$) values are...
calculated for quantifying the differences of spatial patterns for each class by using the following equations.

$$A_0 = \frac{\sum N XY_{ii}}{M} \times 100\%,$$

(1)

$$A_i = \frac{XY_{ii}}{(X_i+Y_i)/2} \times 100\%,$$

(2)

where $X_i$ is the number of pixels of class $i$ in GLC product $X$, and $Y_i$ is the number of pixels of class $i$ in GLC product $Y$. $XY_{ii}$ is the number of pixels of class $i$ in both GLC products $X$ and $Y$. $N$ is the number of land cover classes (14 in this study), which is used to calculate $A_0$, and $M$ is the total number of pixels in the entire global area.

### 2.4 Thematic uncertainty

The thematic similarity is defined as the affinity between different land cover classes based on a set of common land cover attributes from the LCCS translation in previous studies (Ahlqvist, 2005; Neumann et al., 2007; Fritz and See, 2008; Pérez-Hoyos et al., 2012). It allows us to examine the agreement between different land cover products with different classification schemes. The overlap property is firstly computed for each separate categorical or continuous attribute, and then those values for all the common considered attributes are combined together to yield an overlap metric, ranging from 0 (unrelated classes) to 1 (coincidence of all attributes). The thematic similarity indicates the degree of agreement between two land cover classes, and the thematic uncertainty represents the discrepancies in land cover definitions and positional errors. The thematic uncertainty of GLC products can be obtained through the following steps.

Step 1: Following the method described by Pérez-Hoyos et al. (2012), land cover classes of GLC products are translated into LCCS and a set of land cover attributes are defined, and the overlap property is calculated for each separate common attribute. Then, the overlap property values for all common attributes are aggregating into a single quantity by using the following equations to indicate the thematic similarity for any pair of GLC products.

$$O_{ij} = \sum_{k=1}^{n} W_k \times O_k(C_i, C_j).$$

(3)

The overlap metric is a similarity score of the attributes that two land cover classes have in common, ranging from 0 (total disagreement between the two land cover classes) to 1 (the two land cover classes are considered to be identical). $O_{ij}$ is the overlap metric, $W_k$ represents the weight of each common attribute (each attribute is assumed to have the same contribution to the overlap metric), and $O_k(C_i, X_j)$ is the overlap property for each common attribute between classes $i$ and $j$.

Step 2: TUI is defined as below.

$$TUI = \sqrt{a_{ts}^2 + p_{ts}^2},$$

(4)

$$a_{ts} = \text{True}_{ts} - \text{Mean}_{ts},$$

(5)

$$p_{ts} = \text{STD}_{ts},$$

(6)

where $a_{ts}$ is the accuracy, which is the difference value between the true and mean thematic similarities, and $p_{ts}$ is the precision, which is the standard deviation (STD) of thematic similarities. The True$_{ts}$ is set as 1. For a given pixel, if all attributes of the two land cover classes are the same, the thematic similarity is 1, otherwise, it is computed for any pair of GLC products, so there are several thematic similarities of pair-combinations for a given pixel. For example, there are totally six pair-combinations for the four GLC products, and thus, six thematic similarities are obtained for a given pixel. Mean$_{ts}$ and STD$_{ts}$ are the mean value and STD of the six thematic similarities for a given pixel in the same epoch, respectively.

### 2.5 Local classification accuracy uncertainty

The accuracy of GLC products is often expressed in terms of the global accuracy, without informing about the spatial variation in the classification accuracy. The local classification accuracy denotes the local probability that a specific GLC product is correct, and it can be modeled with substantial GLC reference datasets to assess the local classification accuracy of global-scale products in previous studies (See et al., 2015; Tsendbazar et al., 2015; Comber et al., 2017). We followed the method elaborated by Tsendbazar et al. (2015) for calculating the local classification accuracy in this work. The local classification accuracy uncertainty of GLC products is calculated through the following two steps.

Step 1: Correspondences between GLC products and reference data are indicator-coded. If the land cover class of a reference site matched with that of a GLC product, an indicator code of 1 is assigned to the reference site, otherwise, an indicator code of 0 is given to that site. Then, the spatial autocorrelation of the indicator-coded data is analyzed by using indicator semivariograms. More detailed information can be found in (Tsendbazar et al., 2015). Semivariogram examples of the spatial correspondence for GLC in 2008–2010, and the used fitted models for the indicator kriging, are demonstrated in Fig. 2.
Finally, the local classification accuracy is generated by indicator kriging for each GLC product. The value of local classification accuracy ranges from 0 to 1, which denotes the local probability that a specific GLC product is correct. Here, a correct land cover class has a value of 1, while a probably wrong land cover class has a value of less than 1.

Step 2: AUI is defined as follows.

\[
A_{UI} = \sqrt{a_{la}^2 + p_{la}^2}, \tag{7}
\]

\[
a_{la} = \text{True}_{la} - \text{Mean}_{la}, \tag{8}
\]

\[
p_{la} = \text{STD}_{la}, \tag{9}
\]

where \( \text{True}_{la} \) is set as 1, and \( \text{Mean}_{la} \) and \( \text{STD}_{la} \) are the mean value and STD of the local classification accuracy among all GLC products in the same epoch, respectively.

### 2.6 Integrated uncertainty of GLC products

In this study, an IUI, which combines both the thematic uncertainty and local classification accuracy uncertainty, is defined as below.

\[
IUI = W_i \times TUI + W_j \times AUI, \tag{10}
\]

where IUI is the GLC integrated uncertainty index, TUI is the thematic uncertainty index, and AUI is the local classification accuracy uncertainty index. \( W_i \) and \( W_j \) denote the weights of TUI and AUI, respectively. Although in this study identical weights are assigned to TUI and AUI (\( W_i \) and \( W_j \) are equal to 0.5), different weights can also be applied to reveal the relative importance of the two uncertainty indexes in terms of land cover characters.

### 3. Results

#### 3.1 Areal comparisons

The comparison of total percent area for 14 aggregated classes of major GLC products in three epochs is shown in Fig. 3. The analysis excluding Antarctica shows that the grassland and barren or sparsely vegetated are dominant cover types over the globe. Differences among the GLC products are significant. Major discrepancies occur in all the aggregated land cover classes except for the evergreen broadleaf forest, croplands, snow and ice, and inland water. Surprisingly, the urban and built-up with distinct spectral–temporal signals has a relative higher discrepancy among the products. The coefficient of variance (CV) values of the total percent area for each aggregated class in major GLC products in the same epoch (red line in Fig. 3) range from 2.0% to 88.4%, 2.8% to 81.9%, and 8.7% to 77% for the three epochs, respectively. The relative small val-
ues are found in the evergreen broadleaf forest (2.0%, 2.8%, and 8.7%), snow and ice (10.8%, 11.9%, and 14.2%), inland water (22.7%, 19.2%, and 27.6%), and croplands (26.7%, 23.8%, and 19.8%) for all the three epochs, representing the high classification agreement in these land cover classes. While CV values for the deciduous needleleaf forest (88.4%, 81.9%, and 77.0%), evergreen needleleaf forest (62.3%, 61.0%, and 59.2%), de-

Fig. 3. The total area comparison of land cover classes among the major GLC products with CV in (a) 2000–2005, (b) 2008–2010, and (c) 2013–2015 (1 is evergreen needleleaf forest, 2 is evergreen broadleaf forest, 3 is deciduous needleleaf forest, 4 is deciduous broadleaf forest, 5 is mixed forest, 6 is shrublands, 7 is grassland, 8 is croplands, 9 is permanent wetlands, 10 is urban and built-up, 11 is cropland/natural vegetation mosaic, 12 is snow and ice, 13 is barren or sparsely vegetated, and 14 is inland water).
The relative higher classification quality in GLC products. Meanwhile, the mean spatial agreement of the cropland/natural vegetation mosaic (13.3%, 14.0%, and 12.8%), permanent wetlands (20.2%, 20.7%, and 20.0%), and deciduous needleleaf forest (27.9%, 26.3%, and 16.9%) are extremely low in the three epochs, which reveals the relatively lower classification quality in GLC products. Mean spatial agreements of the other land cover classes are all lower than 50%.

When the five types of forest classes and cropland/natural vegetation mosaic in all GLC products are aggregated into forests and croplands, \( A_0 \) values between GLC products are 50.8%–66.0%, 51.8%–66.1%, and 60.6%–66.9% for the three epochs, which have respectively increased by 8.2%–14.1%, 14.5%–15.7%, and 13.8%–18.1% than the results in a detailed scheme for the three epochs. The misclassification among the forests of different plant functional types and mosaic types can account for approximately 20% of disagreements between the GLC products. Naturally, the spatial agreements will decrease sharply when they are estimated based on schemes that contain detailed classes. The analysis indicates that improving the classification accuracy of different plant functional type forests is the key point for GLC products, and mosaic classes are the main factors causing discrepancies among these GLC products.

3.3 Comprehensive analysis between area and per-pixel comparisons

A comprehensive analysis on the results between area and per-pixel comparisons is performed. The results show that the aggregated land cover classes with high consistency in the total area among the major GLC products also have high \( A_0 \) values (Figs. 3, 4), such as the evergreen broadleaf forest, croplands, urban and built-up, snow and ice, barren or sparsely vegetated, and inland water. Similarly, \( A_0 \) of aggregated land cover classes with poor consistency in the total area among different products is found to be poor (Figs. 3, 4), such as the evergreen needleleaf forest, deciduous needleleaf forest, mixed forest, shrublands, permanent wetlands, and cropland/natural vegetation mosaic. No aggregated land cover classes with high consistency in the total area among different products have low \( A_0 \). In other word, central/core land cover classes were correctly identified in all major GLC products, while land cover classes distributing along edges and transition zones were misclassified. It is non-trivial to discriminate clear boundaries between the major ecologically similar land cover classes.

Furthermore, area consistencies and \( A_0 \) values of specific land cover classes vary greatly among the major
GLC products, such as the deciduous needleleaf forest, deciduous broadleaf forest, shrublands, etc. (Figs. 3, 4). For the deciduous needleleaf forest, the total areas among GLC2000, GlobCover, and CCI-LC are relatively consistent, meanwhile the corresponding $A_o$ is also higher than that in other products. Similar cases also occur in specific land cover classes, such as the deciduous broadleaf forest, mixed forest, shrublands, permanent wetlands, and barren or sparsely vegetated class. It could be largely explained with the LCCS used in GLC2000, GlobCover, and CCI-LC. In addition, the same agency mainly responsible for generating these three land cover

![Diagram](image_url)

Fig. 4. The overall spatial agreement ($A_o$, %) and class-specific spatial agreement ($A_s$, %) among the different GLC products in (a) 2000–2005, (b) 2008–2010, and (c) 2013–2015 ($A_1$ is evergreen needleleaf forest, $A_2$ is evergreen broadleaf forest, $A_3$ is deciduous needleleaf forest, $A_4$ is deciduous broadleaf forest, $A_5$ is mixed forest, $A_6$ is shrublands, $A_7$ is grassland, $A_8$ is croplands, $A_9$ is permanent wetlands, $A_{10}$ is urban and built-up, $A_{11}$ is cropland/natural vegetation mosaic, $A_{12}$ is snow and ice, $A_{13}$ is barren or sparsely vegetated, and $A_{14}$ is inland).
products is also likely to have contributed to this distinguishing characteristic. It means that different land cover definitions used in classification schemes for the major GLC products are likely to make great contributions to the discrepancy.

### 3.4 Thematic uncertainty analysis

The spatial thematic uncertainty of the major GLC products based on their original schemes for the three epochs is depicted in Fig. 5, with values typically ranging from 0 (low uncertainty) to 1 (high uncertainty). The mean thematic similarity and thematic similarity STD are also shown in Fig. 5 for better understanding the quality of GLC products. Mean thematic uncertainties of GLC products in the three epochs are 0.49, 0.46, and 0.43, respectively. The mean thematic similarity with STD of GLC products in the three epochs are 0.58 ± 0.24, 0.60 ± 0.21, and 0.61 ± 0.16, respectively. In terms of the thematic uncertainty, mean thematic similarity, and their STD in classification schemes, their spatial patterns in the three epochs are very similar over the entire globe. Generally, areas with low thematic uncertainty also have high mean thematic similarity and low STD of thematic similarity. Low thematic uncertainties are found in Greenland, Sahara Desert, Arabian Peninsula, Taklimakan Desert, tropical rainforest regions (the Amazon, Congo River, and Southeast Asia), and main arable land (North China Plain, India, southern Brazil, Southeast Argentina, and northern central United States), whereas the Arctic Circle, central Asia, Australia, southern Africa, Sahel, western North America, and eastern South America generally present high thematic uncertainties.

Different GLC products are produced with different classification schemes. As presented in Table 1, the LCCS scheme has more classes than IGBP. Definitions among the classification schemes do not match well. The definition of common attributes varies significantly, which are recognized as the main cause of the thematic uncertainty, leading to confusions in the spatial distribution and therefore to inaccurate classifications. Many inconsistencies are evident among the thresholds for common attributes. For example, the definition of the density cover attribute in forest classes varies markedly in different classification schemes. The tree height attribute also has different definitions depending on the classification schemes for different GLC products. Moreover, the leaf type and leaf phenology attributes among the forest types are also very important factors for the thematic uncertainty.

Another example is the density cover attribute in barren and sparsely vegetated areas. Thresholds for the density of cover separating the barren and sparsely veget-
ated areas (i.e., < 10% for MODIS, 1%–20% for GLC2000 and GLCNMO, < 15% for GlobCover and CCI-LC, and 0–4% for the barren) vary among different products, thus resulting in significant disagreements. Therefore, this is one reason why high thematic uncertainty is generally observed in transition zones of the major ecosystems across the world. The sparse vegetation, bare area, open shrubland, and grassland are always confused with each other among different land cover products. One important reason is the overlap among their density cover attributes.

3.5 **Analysis on the local classification accuracy uncertainty**

Figure 6 shows the local classification accuracy uncertainty in the existing GLC products during the three epochs. Local classification accuracy uncertainties among different GLC products are 0.47, 0.44, and 0.39 in the three epochs, respectively. The mean local classification accuracy with STD of GLC products in the three epochs are 0.57 ± 0.17, 0.60 ± 0.16, and 0.64 ± 0.13, respectively. Patterns of the local classification accuracy uncertainty, mean local classification accuracy, and their STD in the three epochs are generally similar to those over the entire globe, and areas with low local classification accuracy uncertainty also have high local classification accuracy and low STD of the local classification accuracy. As shown in Fig. 6, areas with high uncertainty, analyzed based on the aggregated scheme, are found in the central Asia, Australia, northern Canada, northern Russia, eastern and southern Africa, Sahel, eastern and southern parts of South America, and southern part of China. On the other hand, low uncertainty occurs in Greenland, Sahara Desert, Arabian Peninsula, Taklimakan Desert, tropical rainforest regions, and main arable land. It is very similar to the thematic uncertainty pattern. For example, the major land cover of reference data is the shrubland in eastern Africa, shrubland, and barren or sparsely vegetated class in southern Africa during 2000–2005. While the major land cover for these areas is the shrubland or cropland in CCI-LC, grassland in MODIS, grassland, barren or sparsely vegetated class in GLC2000, shrubland or cropland in GLCNMO, and grassland, barren or sparsely vegetated class in GlobCover. For other regions in northern Canada, the major land cover class is barren or sparsely vegetated, but it is classified as the grassland in CCI-LC, shrubland in MODIS, barren or sparsely vegetated in GLC2000, and grassland or shrubland in GLCNMO. Additionally, reference data show the shrubland or cropland/natural vegetation mos-

![Fig. 6. Spatial patterns of the (a–c) local classification accuracy uncertainty, (d–f) mean local classification accuracy, and (g–i) STD of local classification accuracy among the different GLC products in 2000–2005, 2008–2010, and 2013–2015, respectively.](image-url)
aic in eastern Brazil which has high uncertainty, as it is
classified as the grassland in MODIS, cropland in
GLC2000, cropland/natural vegetation mosaic in
GLCNMO. GLC mapping with the 14-class common
scheme during 2000–2005 are shown in supplementary
figures. In general, confusions among the shrubland,
grassland, and cropland/natural vegetation mosaic are al-
ways serious, and their local accuracies are relatively low
due to similar temporal and spectral features. Hence,
those areas should be paid more attention for GLC
producers to generate high-quality products.

3.6 Integrated uncertainty analysis

The integrated uncertainty among GLC products to
quantify the total classification uncertainty, which com-
bines both the thematic uncertainty and local classifica-
tion accuracy uncertainty, is shown in Fig. 7. Mean val-
ues of the spatial integrated uncertainty among GLC
products are 0.48, 0.45, and 0.41 in the three epochs, re-
spectively. The spatial pattern of the integrated uncer-
tainty among GLC products is very similar to that of the
thematic uncertainty and local classification accuracy un-
certainty. High uncertainty occurs in northern Russia,
northern part of North America, western and central
America, eastern Brazil and Argentina, Australia, central
Asia, southern part of China, eastern and southern parts
of Africa, and Sahel. While low uncertainty is observed
in Greenland, Sahara Desert, Arabian Peninsula, Takli-
makan Desert, tropical rainforest regions, and the main
arable land over the world in the three epochs. Obvi-
ously, transition zones of the major ecosystems across
the world and areas with high landscape complexity, e.g.,
mountainous areas or hilly regions, have relatively high
integrated uncertainty. It is not surprising that the trans-
ition zones and landscape heterogeneity have been re-
garded as major factors influencing the GLC accuracy.

---

**Fig. 7.** Spatial patterns of the (a–c) integrated uncertainty and (d–f) identified area with high thematic similarity but low local classification ac-
curacy among the different GLC products in 2000–2005, 2008–2010, and 2013–2015, respectively.
On the one hand, similar temporal and spectral features of land cover classes in transition zones largely contribute to disagreements in this area. On the other hand, due to the complicated spectral characteristics, it is difficult for classification on the land cover in areas with high landscape heterogeneity.

Moreover, the mean thematic similarity and local classification accuracy are used to identify specific areas for validating the integrated uncertainty. Pixels can be considered as the uncertain area, if a pixel shows high thematic similarity among GLC products but low local classification accuracy. STD of the mean thematic similarity and mean local classification accuracy are selected as the threshold for identifying areas with high thematic similarity but low local classification accuracy (Fig. 7). It can be seen that the identified areas with high thematic similarity but low local classification accuracy are all captured in the integrated uncertainty among GLC products. These areas should be paid more attention for both GLC product producers and users in the future.

The integrated uncertainty contains the thematic uncertainty, which presents discrepancies in land cover definitions and positional errors. Despite of the high local classification accuracy uncertainty in some areas, the wrong classification is not vital, which is just a confusion between similar land cover classes. However, compared with the local classification accuracy uncertainty, the integrated uncertainty would decrease by combining the thematic uncertainty together. For example, the main land cover is found to be the deciduous needleleaf forest with evergreen needleleaf forest in northern Russia over 2008–2010, but the area is classified as the mixed forest in MODIS, deciduous needleleaf forest in GlobCover, mixed forest or deciduous needleleaf forest in GLCNMO, and evergreen needleleaf forest or mixed forest in CCI-LC, respectively. In terms of the classification accuracy, it is wrong, but for the thematic uncertainty, these land cover classes have some similarity. The integrated uncertainty has well captured these characteristics among the GLC products. In fact, these types of the uncertainty can be found in many areas in the integrated uncertainty figure. Therefore, the integrated uncertainty in this study provides more implications for understanding the uncertainty among the existing GLC products.

4. Conclusions and discussion

Study on the pixel-level uncertainty is essential for understanding the suitability and limitation of GLC products, as well as the analysis on the propagation of this kind of uncertainty in various models, particularly for areas and classes with high uncertainties. In this study, both area and per-pixel comparisons are conducted among the existing GLC products for identifying the bias of land cover products. This paper has also outlined a method that provides the pixel-level uncertainty combining both the thematic uncertainty (indicating discrepancies in land cover definitions and positional errors) and local classification accuracy uncertainty (depicting the spatial variability of classification accuracy). Based on the major existing GLC products, this methodology is performed in three epochs.

Comparisons show that, for all the three epochs, the total percent area of each land cover class among different GLC products varies significantly. Major discrepancies are found in the deciduous needleleaf forest, evergreen needleleaf forest, deciduous broadleaf forest, and cropland/natural vegetation mosaic, whereas the evergreen broadleaf forest, croplands, snow and ice, and inland water are with high consistency. This may be the result of the good spectral discrimination of the evergreen broadleaf forest, snow and ice, inland water, great pixel homogeneity, and relative stable distribution of croplands. Total areas of the forest are very consistent for the existing GLC products, but significant differences are observed in different forest classes. $A_o$ values between GLC products are 35.9%–57.8%, 36.1%–51.6%, and 46.8%–48.8% for the three epochs, respectively. The highest $A_o$ of 57.8% is found between CCI-LC 2001 and GLC2000, followed by that between CCI-LC 2010 and GlobCover2009 (51.9%), and the lowest value (35.9%) occurs between GlobCover2005 and MODIS2001. Meanwhile, $A_o$ values between the other GLC products are all lower than 50%. $A_o$ exhibits significantly different characteristics. Classes with high spatial agreement are the snow and ice, evergreen needleleaf forest, barren or sparsely vegetated, inland water, and cropland, whereas the evergreen broadleaf forest, deciduous needleleaf forest, and permanent wetlands have extremely low spatial agreements. This may be partly associated to the mixed class pixels with the high intra-class variability and fragmented land cover, which has a negative influence on the classification accuracy of land cover. Mean spatial agreements of other land cover classes are lower than 50%. When the five classes of forests are aggregated into forests, $A_o$ values of the products increase by 8.2%–18.1% for the three epochs. The misclassification among different forest classes and mosaic types can account for about 20% of the total disagreement. A key point of improving the quality of GLC products is to en-
hance the accuracy of forest types and mosaic classes.

The spatial uncertainty analysis has further shown that the mean thematic uncertainty, local classification accuracy uncertainty, and integrated uncertainty all nearly reach 0.5 for the three epochs. The results show that different definitions of common attributes among classification schemes with GLC products are recognized as the main factor of the thematic uncertainty, leading to a confusion and inaccurate classification. The local classification accuracy uncertainty demonstrates where the relatively high confidence on classification accuracy exists, and to what degree it is. Generally, areas with the ecologically similar land cover classes tend to have low values of local accuracy due to the low separability of spectral–temporal signal features. Most of the high uncertainty occurs in transition zones of major ecosystems across the world and areas with high complexity in the landscape. Compared with the previous studies, the integrated uncertainty in this work, which contains both the thematic uncertainty and local classification accuracy uncertainty, provides more implications for understanding the uncertainty among the existing GLC products. The distribution of the integrated uncertainty is very similar to that of Yu et al. (2018). The spatial distributions of the difficult to map regions (DMRs) identified by Yu et al. (2018) are the most consistent with high slope variance areas attached with shrubland-related or grassland-related land cover classes or biomes. Therefore, DMRs or areas with high integrated uncertainty in our study should be taken as focus areas to improve the accuracy of GLC products.

Significantly, the difference in classification schemes is one of the major factors driving disagreements among the GLC products (Herold et al., 2006, 2008). Scientists have highlighted the harmonization and adoption of common classification schemes. In our study, a common classification scheme with 14 aggregated land cover classes are used for area and per-pixel comparisons. Classes of the evergreen broadleaf forest, cropland, urban and built-up, snow and ice, and inland water have the relative high consistency of total areas and $A_o$ values. Interestingly, it is found that these aggregated land cover classes have explicit definitions in different classification systems. For instance, the evergreen broadleaf forest, urban and built-up, snow and ice, and inland water with high $A_o$ values are the only class in the classification system of GLC products. The classification system cross-working process for the GLC product comparison may induce some uncertainty. It is shown that designing a common classification scheme for different application goals, which gives explicit definitions of land cover classes in the thresholds of common attributes, is still necessary and exigent. Moreover, according to the previous studies, enhancing the ability to discriminate the ecologically and spectrally similar land cover classes and improving the classification accuracy for transition zones and heterogeneous landscapes are key approaches in a short term. Recently, the emerging novel data such as hyperspectral data and crucial attribute data, and methods such as artificial intelligence as well as deep learning, have provided a new opportunity for improving the classification accuracy of land cover.

There are also some limitations with our method. Firstly, it is considered that the thematic uncertainty and local classification accuracy uncertainty are of equal importance for GLC products in this study. However, the integrated uncertainty information has more implications for specific applications. Different weights can be applied to evaluate the relative importance of the two properties, regarding to the suitability of this method for specific applications. For example, if the local classification accuracy uncertainty is regarded as the extremely crucial factor affecting the results in an application, and the harmonization could solve the difference between land cover classes, the influence of thematic uncertainty can be ignored. Then the weight for the thematic uncertainty can be assigned with a low value, while a relatively high value is assigned to the local classification accuracy uncertainty. Secondly, the thematic uncertainty is implemented based on different classification schemes with common attributes from the viewpoint of product producers. The result is universal for common applications, but is not fit for specific applications, such as climate modeling. In order to obtain the optimal uncertainty sources for specific applications, the thematic uncertainty could be calculated based on special attributes that affect the application outcomes, as done in the previous studies (Fritz and See, 2008; Gao and Jia, 2013; Tsendbazar et al., 2016). In addition, the local classification accuracy uncertainty is achieved with both the GLC products and reference datasets can induce new uncertainties. Using enough credible reference data for the local classification accuracy uncertainty analysis with the classification scheme of products can provide more accurate results. The integrated uncertainty could not only enhance the understanding of the quality of GLC products, but also provide a very important basis for the land cover fusion. Thus, our next step is to generate the integrated GLC products with high quality based on information of the integrated uncertainty.
Acknowledgments. The authors thank all the international organizations for providing the satellite derived GLC products or reference datasets used in this study, and appreciate all anonymous reviewers for their constructive comments on this paper.

REFERENCES

Ahlqvist, O., 2005: Using uncertain conceptual spaces to translate between land cover categories. Int. J. Geogr. Inf. Sci., 19, 831–857, doi: 10.1080/13658810500106729.

Bai, Y., M. Feng, H. Jiang, et al., 2014: Assessing consistency of five global land cover data sets in China. Remote Sens., 6, 8739–8759, doi: 10.3390/rs6098739.

Bartholomé, E., and A. S. Belward, 2005: GLC2000: A new approach to global land cover mapping from earth observation data. Int. J. Remote Sens., 26, 1959–1977, doi: 10.1080/014311600210209.

Bicheron, P., P. Defourny, C. Brockmann, et al., 2008: GlobCover: Products description and validation report. MEDIAS-France, Toulouse, 1–46.

Bontemps, S., P. Defourny, E. Van Bogaert, et al., 2011: GLOBCOVER 2009 products description and validation report. European Space Agency, Frascati, Italy, 1–51.

Chen, J., J. Chen, A. P. Liao, et al., 2015: Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J. Photogramm. Remote Sens., 103, 7–27, doi: 10.1016/j.isprsjprs.2014.09.002.

Comber, A. J., 2013: Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. Remote Sens. Lett., 4, 373–380, doi: 10.1080/2150704X.2012.736694.

Comber, A., C. Brunsdon, M. Charlton, et al., 2017: Geographically weighted correspondence matrices for local error reporting and change analyses: Mapping the spatial distribution of errors and change. Remote Sens. Lett., 8, 234–243, doi: 10.1080/2150704X.2016.1258126.

Defourny, P., S. Bontemps, L. Schouten, et al., 2011: GLOBCOV-ER 2005 and GLOBCOVER 2009 Validation: Learnt lessons. Proceedings of GOFC-GOLD Global Land Cover & Change Validation Workshop, Laxenburg, Austria.

Defourny, P., G. Kirches, C. Brockmann, et al., 2016: Land cover CCI: Product user guide version 2. Available online at http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-PUG-v2.5.pdf. Accessed October 5, 2019.

Fendell, K. L., E. Shevliakova, P. C. D. Milly, et al., 2007: Modeled impact of anthropogenic land cover change on climate. J. Climate, 20, 3621–3634, doi: 10.1175/JCLI4185.1.

Foody, G. M., 2002: Status of land cover classification accuracy assessment. Remote Sens. Environ., 80, 185–201, doi: 10.1016/S0034-4257(01)00295-4.

Foody, G. M., 2005: Local characterization of thematic classification accuracy through spatially constrained confusion matrices. Int. J. Remote Sens., 26, 1217–1228, doi: 10.1080/01431640500133126521.

Friedl, M. A., D. Sulla-Menashe, B. Tan, et al., 2010: Modis Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sens. Environ., 114, 168–182, doi: 10.1016/j.rse.2009.08.016.

Fritz, S., and L. See, 2008: Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. Global Change Biol., 14, 1057–1075, doi: 10.1111/j.1365-2486.2007.01519.x.

Fritz, S., I. McCallum, C. Schill, et al., 2009: Geo-Wiki.Org: The use of crowdsourcing to improve global land cover. Remote Sens., 1, 345–354, doi: 10.3390/rs1030345.

Gao, H., and G.-S. Jia, 2012: Spatial and quantitative comparison of satellite-derived land cover products over China. Atmos. Oceanic Sci. Lett., 5, 426–434, doi: 10.1080/16742834.2012.11447026.

Gao, H., and G. S. Jia, 2013: Assessing disagreement and tolerance of misclassification of satellite-derived land cover products used in WRF model applications. Adv. Atmos. Sci., 30, 125–141, doi: 10.1007/s00376-012-2037-4.

Ge, J. J., J. G. Qi, B. M. Lofgren, et al., 2007: Impacts of land use/cover classification accuracy on regional climate simulations. J. Geophys. Res. Atmos., 112, D05107, doi: 10.1029/2006JD00704.

Gong, P., J. Wang, L. Yu, et al., 2013: Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens., 34, 2607–2654, doi: 10.1080/0143116012.748992.

Hansen, M. C., R. S. Defries, J. R. G. Townshend, et al., 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens., 21, 1331–1364, doi: 10.1080/014311600210209209.

Herold, M., C. E. Woodcock, A. Di Gregorio, et al., 2006: A joint initiative for harmonization and validation of land cover datasets. IEEE Trans. Geosci. Remote Sens., 44, 1719–1727, doi: 10.1109/TGRS.2006.871219.

Herold, M., P. Mayaux, C. E. Woodcock, et al., 2008: Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ., 112, 2538–2556, doi: 10.1016/j.rse.2007.11.013.

Hua, T., W. W. Zhao, Y. X. Liu, et al., 2018: Spatial consistency assessments for global land-cover datasets: A comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. Remote Sens., 10, 1846, doi: 10.3390/rs10111846.

Iwao, K., K. Nishida, T. Kinoshita, et al., 2006: Validating land cover maps with Degree Confluence Project information. Geophys. Res. Lett., 33, L23404, doi: 10.1029/2006GL027768.

Kaptue Tchuente, A. T., J.-L. Roujean, and S. M. De Jong, 2011: Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover datasets at the African continental scale. Int. J. Appl. Earth Obs. Geoinf., 13, 207–219, doi: 10.1016/j.jag.2010.11.005.

Madhusoodhanan, C. G., K. G. Sreeja, and T. I. Eldho, 2017: Assessing uncertainties in global land cover products for hydro-climate modeling in India. Water Resour. Res., 53, 1713–1734, doi: 10.1002/2016WR020193.

Mayaux, P., H. Eva, J. Gallego, et al., 2006: Validation of the global land cover 2000 map. IEEE Trans. Geosci. Remote Sens., 44, 1728–1739, doi: 10.1109/TGRS.2006.864370.

McCallum, I., M. Obersteiner, S. Nilsson, et al., 2006: A spatial comparison of four satellite derived 1 km global land cover datasets. Int. J. Appl. Earth Obs. Geoinf., 8, 246–255, doi: 10.1016/j.jag.2005.12.002.
Nakaegawa, T., 2011: Uncertainty in land cover datasets for global land-surface models derived from 1-km global land cover datasets. *Hydrol. Process.*, 25, 2703–2714, doi: 10.1002/hyp.8011.

Neumann, K., M. Herold, A. Hartley, et al., 2007: Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *Int. J. Appl. Earth Obs. Geoinf.*, 9, 425–437, doi: 10.1016/j.jag.2007.02.004.

Olafsson, P., S. V. Stehman, C. E. Woodcock, et al., 2012: A global land-cover validation data set, part I: Fundamental design principles. *Int. J. Remote Sens.*, 33, 5768–5788, doi: 10.1080/01431161.2012.674230.

Pérez-Hoyos, A., F. J. García-Haro, and J. San-Miguel-Ayanz, 2012: Conventional and fuzzy comparisons of large scale land cover products: Application to CORINE, GLC2000, MODIS and GlobCover in Europe. *ISPRS J. Photogramm. Remote Sens.*, 74, 185–201, doi: 10.1016/j.isprsjprs.2012.09.006.

Potapov, P., M. C. Hansen, A. M. Gerrand, et al., 2011: The global Landsat imagery database for the FAO FRA remote sensing survey. *Int. J. Digit. Earth*, 4, 2–21, doi: 10.1080/17538941003777521.

Santos-Alamillos, F. J., D. Pozo-Vázquez, J. A. Ruiz-Arias, et al., 2015: Influence of land-use misrepresentation on the accuracy of WRF wind estimates: Evaluation of GLCC and CORINE land-use maps in southern Spain. *Atmos. Res.*, 157, 17–28, doi: 10.1016/j.atmosres.2015.01.006.

Schultz, M., N. E. Tsendlbazar, M. Herold, et al., 2015: Utilizing the global land cover 2000 reference dataset for a comparative accuracy assessment of 1 km global land cover maps. Proceedings of the 36th International Symposium on Remote Sensing of Environment, ISPRS—The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Berlin, 503–510.

See, L., D. Schepaschenko, M. Lesiv, et al., 2015: Building a hybrid land cover map with crowdsourcing and geographically weighted regression. *ISPRS J. Photogramm. Remote Sens.*, 103, 48–56, doi: 10.1016/j.isprsjprs.2014.06.016.

Sertel, E., A. Robock, and C. Ormeci, 2010: Impacts of land cover data quality on regional climate simulations. *Int. J. Climatol.*, 30, 1942–1953, doi: 10.1002/joc.2036.

Sterling, S. M., A. Ducharme, and J. Polcher, 2013: The impact of global land-cover change on the terrestrial water cycle. *Nat. Clim. Change*, 3, 385–390, doi: 10.1038/nclimate1690.

Tateishi, R., B. Uriyangqai, H. Al-Bilbisi, et al., 2011: Production of global land cover data—GLCNMO. *Int. J. Digit. Earth*, 4, 22–49, doi: 10.1080/17538941003777521.

Tateishi, R., N. T. Hoan, T. Kobayashi, et al., 2014: Production of global land cover data—GLCNMO2008. *J. Geogr. Geol.*, 6, 99–123, doi: 10.5539/jgg.v6n3p99.

Tsendbazar, N.-E., S. de Bruin, S. Fritz, et al., 2015: Spatial accuracy assessment and integration of global land cover datasets. *Remote Sens.*, 7, 15,804–15,821, doi: 10.3390/rs71215804.

Tsendbazar, N. E., S. de Bruin, B. Mora, et al., 2016: Comparative assessment of thematic accuracy of GLC maps for specific applications using existing reference data. *Int. J. Appl. Earth Obs. Geoinf.*, 44, 124–135, doi: 10.1016/j.jag.2015.08.009.

Tsutsumida, N., and A. J. Comber, 2015: Measures of spatio–temporal accuracy for time series land cover data. *Int. J. Appl. Earth Obs. Geoinf.*, 41, 46–55, doi: 10.1016/j.jag.2015.04.018.

Wickham, J., S. V. Stehman, L. Gass, et al., 2017: Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). *Remote Sens. Environ.*, 191, 328–341, doi: 10.1016/j.rse.2016.12.026.

Yang, Y. K., P. F. Xiao, X. Z. Feng, et al., 2017: Accuracy assessment of seven global land cover datasets over China. *ISPRS J. Photogramm. Remote Sens.*, 125, 156–173, doi: 10.1016/j.isprsjprs.2017.01.016.

Yu, L., X. X. Liu, Y. Y. Zhao, et al., 2018: Difficult to map regions in 30 m global land cover mapping determined with a common validation dataset. *Int. J. Remote Sens.*, 39, 4077–4087, doi: 10.1080/01431161.2018.1455238.

Zhao, Y. Y., P. Gong, L. Yu, et al., 2014: Towards a common validation sample set for global land-cover mapping. *Int. J. Remote Sens.*, 35, 4795–4814, doi: 10.1080/01431161.2014.930202.

Tech & Copy Editor: Meili ZHANG