

A comparative analysis of five global cropland datasets in China

LU Miao¹, WU WenBin^{1*}, ZHANG Li¹, LIAO AnPing², PENG Shu² & TANG HuaJun^{1†}

¹Key Laboratory of Agri-informatics, Ministry of Agriculture/Institute of Agricultural Resources and Regional Planning,
Chinese Academy of Agricultural Sciences, Beijing 100081, China;

²National Geomatics Centre of China, Beijing 100830, China

Received September 18, 2015; accepted April 25, 2016

Abstract Accurate information of cropland area and spatial location is critical for studies of national food security, global environmental change, terrestrial ecosystem geophysics and the geochemical cycle. In this paper, we compared five global cropland datasets in circa 2010 of China from in terms of cropland area and spatial location, including GlobalLand30, FROM-GLC, GlobCover, MODIS Collection 5, and MODIS Cropland. The results showed that the accuracies of cropland area and spatial location of GlobeLand30 were higher than the other four products. The cropland areas of the five products varied in most of the provinces. Compared with the statistical data, the best goodness of fit was obtained using GlobeLand30, followed by MODIS Collection 5 and FROM-GLC, with MODIS Cropland and GlobCover having the poorer accuracies. Regarding the spatial location of cropland, GlobeLand30 achieved the best accuracy, followed by FROM-GLC and MODIS Collection 5, with GlobCover and MODIS Cropland having the poorer accuracies. In addition, the spatial agreement of the five datasets was reduced from agricultural production area to pastoral area and significantly affected by elevation and slope factors. Although the spatial resolution of MODIS Collection 5 was the lowest, accuracies of the cropland area and spatial location were better than those of GlobCover and MODIS Cropland. Therefore, high spatial resolution remote sensing images can help to improve the accuracy of the dataset during land cover mapping, while it is also important to select a suitable classification method. Furthermore, in northwestern and southeastern China, spectral mixing pixels are universal because of the complicated landscape and fragmented topography and result in uncertainty and poor consistency when using the five products. Therefore, these regions require additional attention in future cropland mapping studies.

Keywords Land cover, Cropland area, Spatial location, Accuracy, Spatial agreement

Citation: Lu M, Wu W B, Zhang L, Liao A P, Peng S, Tang H J. 2016. A comparative analysis of five global cropland datasets in China. *Science China Earth Sciences*, doi: 10.1007/s11430-016-5327-3

1. Introduction

Land cover plays an important role in studies of climate change and the biogeochemistry of earth systems (Chen et al., 2011). With the development of satellite sensors and computer cartography, remote sensing has become an increasingly important technology for obtaining land cover

information from the regional to global scales (Giri et al., 2013; Ban et al. 2015; Chen et al., 2013). Early global land cover datasets were produced with 1 km spatial resolution satellite imagery, such as the UMD (University of Maryland) product (Hansen et al., 2000), IGBP-DISCover (International Geosphere-Biosphere Programme Data and Information Systems Cover) product (Hansen et al., 2000; Loveland et al., 2000), MODIS land cover product (Friedl et al., 2002), and GLC2000 (Global Land Cover 2000) (Bartholome and Belward, 2005) dataset. Although these datasets have been widely used, their accuracies and qualities

*Corresponding author (email: wuwenbin@caas.cn)

†Corresponding author (email: tanghua jun@caas.cn)

are not satisfactory for many applications (Chen et al., 2015). Therefore, land cover products with higher resolution and accuracy are urgently needed. Recently, Boston University produced a new global land cover dataset called the MODIS Collection 5 using the MODIS data from 2000–2012 with a spatial resolution of 500 m (Friedl et al., 2010). In addition, based on MERIS (Medium Resolution Imaging Spectrometer) reflection data, the European Space Agency (ESA) produced the GlobCover land cover dataset for 2005 and 2009 with a spatial resolution of 300 m (Bicheron et al., 2008; Bontemps et al., 2011). With the support of national science and technology planning projects of China, the first global land cover datasets with 30 m spatial resolution were produced by Chinese scientists; these products include the FROM-GLC (Finer Resolution Observation and Monitoring-Global Land Cover) dataset of Tsinghua University and the GlobeLand30 dataset from the National Geomatic Center of China (Gong et al., 2013; Cao et al., 2014; Liao et al., 2014; Chen et al., 2015). Although several land cover datasets exist, they were developed by using different mapping standards and classification methodologies (Herold et al., 2008), resulting in inevitable discrepancies of the available information. It is important to compare land cover products to understand their strengths and weaknesses. Product users need the accuracy of land cover products to understand which map is most useful for their application, and map producers need to understand the quality of the products to improve their accuracy (Congalton et al., 2014; Wang and Liu, 2014). Several studies have compared the consistency of global land cover datasets at the global, regional, and national scales (Hansen and Reed, 2000; Giri et al., 2005; Jung et al., 2006; Herold et al., 2008; Wu et al., 2009; Song and Zhang, 2012; Li et al., 2010; Yang et al., 2014). For example, Herold et al. (2008) harmonized the classification systems of four land cover products, including IGBP DISCover, UMD, MODIS 1 km, and GLC2000, to compare and assess their advantages and disadvantages. Tchuenté et al. (2011) studied the agreement and quality of GLC2000, GlobCover, MODIS and ECOCLIMAP data over the Africa. Congalton et al. (2014) analyzed the uncertainty of IGBP DISCover, UMD, GLC2000 and GlobCover using an error budget approach. However, these studies focused on global land cover products of coarse spatial resolution in circa 2000. Few studies have compared the global land cover datasets for 2010, especially for the GlobeLand30 and FROM-GLC datasets with a spatial resolution of 30 m.

As one of the most important types of land cover, cropland is a key component of the subsistence and development of human beings. Cropland area and spatial location are highly related to the security of human food sources, global environmental change, terrestrial ecosystem geophysics and the geochemical cycle (Fritz et al., 2013; Tang et al., 2015). Hence, focusing on cropland in China, this study compares cropland datasets derived from five global

land cover products, GlobeLand30, FROM-GLC, GlobCover, MODIS Collection 5, and MODIS Cropland in circa 2010. The five cropland products are compared for cropland area accuracy and spatial location accuracy. To compare the accuracy of the cropland area, the cropland area is calculated at the province and regional scales using the five cropland products listed above and then compared with the statistical data. The spatial location accuracies of the five products are assessed using validation samples by error matrix, and the spatial agreements are analyzed using elevation and slope data. The comparison of the five cropland products can help product users select appropriate cropland data for application, and help map producers improve their datasets in the future.

2. Data sources and preprocessing

2.1 Data sources

Table 1 lists the basic information of the five global datasets from 2010. The GlobeLand30 dataset, including 10 land cover types, was produced using the POK (pixel-object-knowledge) classification methodology based on 30 m resolution images from the Landsat satellite and China's environmental satellite (HJ) (Chen et al., 2015). Compared with GlobeLand30, the FROM-GLC dataset utilizes the same classification system and images but a different classification method. FROM-GLC employed the automatic classification algorithm, i.e., SVM (support vector machine) and MLC (maximum likelihood classifier), and improved the classification results using multi-source data, including MODIS time series image, DEM (Digital Elevation Model) and soil water data (Yu et al., 2013). According to the LCCS (Land Cover Classification System) established by the Food and Agriculture Organization (FAO), GlobCover2009 combined supervised and unsupervised classification using the temporal and spectral features of MERIS images to divide the land cover into 22 types (Bontemps et al., 2011). Based on the 500 m MODIS time series data, the dataset of MODIS Collection 5 followed the IGBP land cover classification system and used the decision tree classification to obtain 17 land cover types at the global scale (Friedl et al., 2010). South Dakota State University used 250 m MODIS time series data from 2000 to 2008 to produce global cropland extent data using decision tree classification and statistical data, yielding the so-called MODIS Cropland dataset (Pittman et al., 2010).

The statistical data of cropland area are sourced from the 2010 "Regional Economical Statistical Yearbook of China", which was produced by the Ministry of Land and Resources. The data include cropland area and the corresponding proportion in each province (excluding the regions of Hong Kong, Macao, and Taiwan) (Sheng and Yan, 2010). In addition, the DEM data of SRTM (Shuttle Radar Topography Mission), with a 90 m resolution, are applied to calculate the slope data.

Table 1 Overview of the five cropland datasets used in this study

| Product | Satellite sensor | Spatial resolution (m) | Cropland class |
|--------------------|--------------------|------------------------|--|
| GlobeLand30 | Landsat/TM, HJ/CCD | 30 | Cropland |
| FROM-GLC | Landsat/TM | 30 | Cropland, un-mulched farmland |
| GlobCover | ENVISAT/MERIS | 300 | Irrigated soil, dry farmland, mixed cropland (50–70%)/vegetation (20–50%), mixed vegetation (50–70%)/cropland (20–50%) |
| MODIS Collection 5 | TERRA/MODIS | 500 | Cropland, mixed cropland/natural vegetation farmland |
| MODIS Cropland | TERRA/MODIS | 250 | Cropland |

2.2 Data preprocessing

Data preprocessing, mainly including coordinate transformation, standardization of spatial resolution and harmonization of cropland definition, is important before comparing the five cropland maps.

First, the five datasets for China were obtained by clipping the original global land cover datasets using vector data of China's boundary. All five datasets are under the strict control of geometrical accuracy during production, for example, the accuracy of GlobeLand30 based on Landsat imagery is within 25 m (Rengarajan et al., 2015), while that of GlobCover is within 150 m (Bontemps et al., 2011). However, the diversities of the geometrical accuracy are obvious. To reduce the effects from the differences among the geometrical accuracies of the datasets, the WGS-84 (World Geodetic System-1984) of GlobeLand30 dataset was used as the reference ellipsoid, and the geographical coordinate systems of the other four datasets were normalized to the WGS-84. Then, all input data were standardized to the spatial resolution of 500 m using the nearest neighbor resampling method. The classification systems of the five land cover datasets are different; thus, it is necessary to harmonize the cropland class. Referring to the definition of croplands given by the FAO, the types of cropland in the five datasets are shown in Table 1. Both GlobeLand30 and MODIS Cropland contain only one type of cropland, while FROM-GLC contains both cropland and un-mulched farmland. The GlobCover dataset contains four types of cropland, including irrigated soil, dry farmland, mixed cropland (50–70%)/vegetation (20–50%), and mixed vegetation (50–70%)/cropland (20–50%). The MODIS Collection 5 dataset contains two types of cropland, including cropland and mixed cropland/natural vegetation farmland. These types of cropland were extracted from the products, and the five cropland datasets in China were obtained as shown in Figure 1.

3. Comparison method

In this study, the five cropland datasets in China were compared according to cropland area and spatial location. For cropland area, we analyzed the differences among the five products in terms of the proportion of cropland area and the discrepancies relative to the statistical data at the province

and regional scales. For the spatial distribution, we assessed the accuracy based on validation samples and analyzed the spatial consistency of the five datasets.

3.1 Comparison of the cropland area

The cropland area in each province was calculated from the five datasets and then used to summarize the total national cropland area, followed by calculations for the proportions of cropland area in each province relative to the total national area. According to its geographical location, China was divided into six areas: northeast, north, southeast, central, southwest and northwest. Depending on the calculated proportion of cropland area in each province, the cropland proportions of each product in the six regions were estimated. Then, based on the proportion of cropland area referenced in the 2010 “Regional Economical Statistical Yearbook of China”, the deviation can be calculated between the proportion of statistical data and the proportions of cropland area in each province and region. Furthermore, the root mean square error (RMSE) between the proportion of statistical data and the proportion of cropland area estimated by products can be calculated to reflect the dispersion between the reported cropland and statistical data products. The deviation Δx_i and RMSE can be calculated as follows:

$$\Delta x_i = x_i - y_i, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}, \quad (2)$$

where x_i is the proportion of cropland area in province i calculated using the cropland dataset, y_i is the deviation of the cropland area from the statistical data in province i , and n is the total number of provinces.

Finally, a correlation analysis was performed with the following equation to reflect the goodness of fit between the proportion of cropland area in the cropland datasets and the statistics.

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3)$$

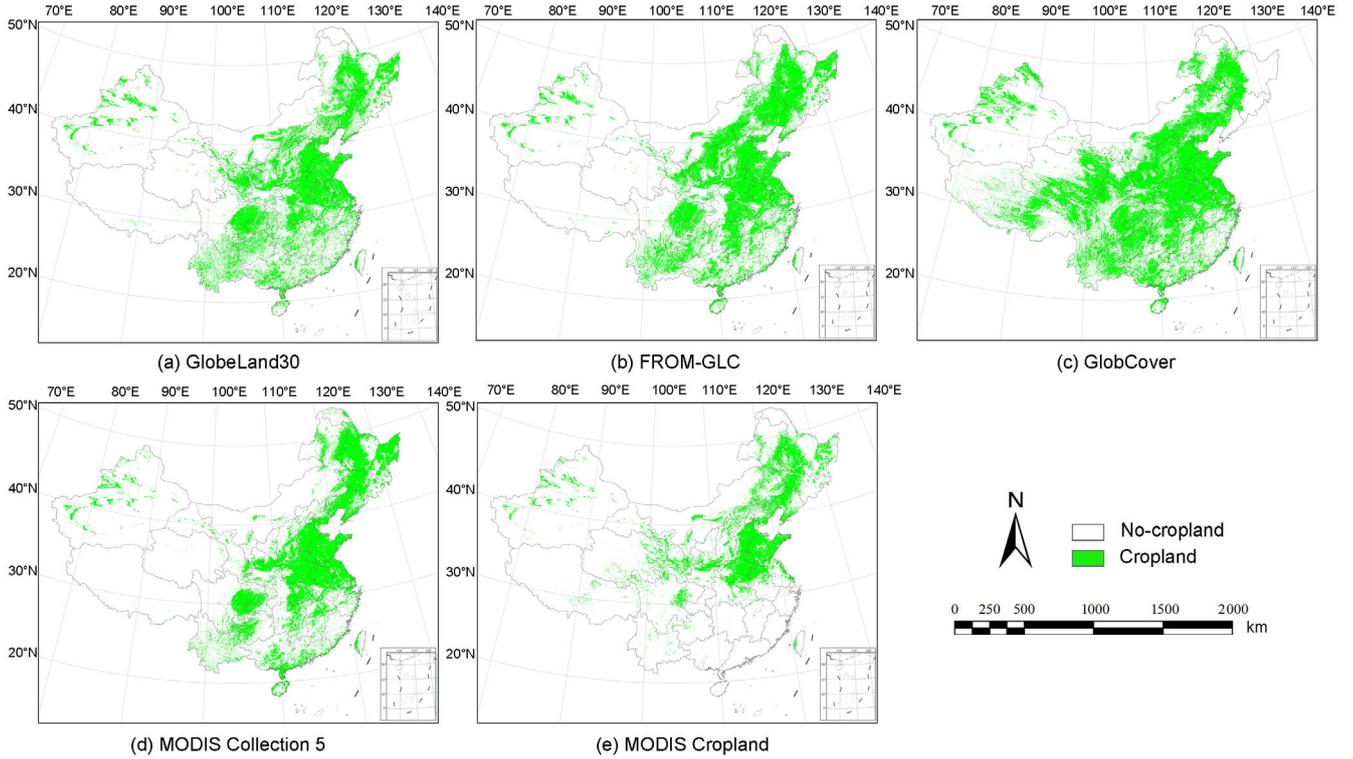


Figure 1 Five cropland datasets of China after preprocessing.

where \bar{x} is the average proportion of the cropland area that was estimated from the cropland dataset and \bar{y} is the average proportion of the statistical area. Higher correlation coefficients correspond with better goodness of fit between the cropland dataset and the statistical data.

3.2 Comparison of the spatial location

In general, land cover mapping can easily describe the area characteristics of land cover types, but it is difficult to indicate the spatial distribution accurately (Wu et al., 2008). Traditionally, spatial locations are compared by analyzing the consistency between the target product and the reference product (Li et al., 2010; Pflugmacher et al., 2011). This method relies on the accuracy of the reference data; however, the reference data maybe inaccurate and imprecise, especially in large scale mapping efforts (Herold et al., 2008; Yang et al., 2014). In this paper, the accuracy of spatial location was analyzed by error matrix based on validation samples and the spatial consistency of the five datasets. The error matrix method is the most popular method for accuracy analysis in remote sensing. The error matrix conducted by validation samples is a cross-tabulation of the mapped land cover class against the reference land cover class, and the parameters represent the errors include the overall accuracy, Kappa coefficient, producer accuracy and user accuracy. The amount and distribution of validation samples significantly affect the results of the error matrix (Foody,

2010). In this paper, the validation samples were derived from two different sources to ensure the quality. The first source was developed for validating the FROM-GLC dataset (Gong et al., 2013; Yu et al., 2014). Equal-area stratified random sampling was employed for the sample design. The global land area was partitioned into approximately 7000 equal-area hexagons, and 5 samples were randomly selected from each hexagon. Then, the land cover types of the samples were identified using visual interpretation and field surveys. There are 443 cropland samples and 1687 no-cropland samples in China, which hardly satisfies the requirement of precise cropland assessment. Hence, 0.001% pixels were randomly selected at different consistency levels of agreement map to develop supplementary samples. Next, samples were labeled as cropland or no-cropland by visual interpretation via high-resolution Google Earth images, including 2716 cropland samples and 858 no-cropland samples. The samples from the two sources were combined to obtain the validation samples for the five cropland datasets, as shown in Figure 2, with 3159 cropland samples and 2545 no-cropland samples. Based on these samples, the error matrix was used to estimate the accuracies of spatial location for the five cropland products at the country and region scales.

Spatial agreement analysis describes the differences and similarities at a given location by counting the number of occurrences of agreement for the five cropland maps (Yang et al., 2014; Tchuenté et al., 2011). In the spatial agreement

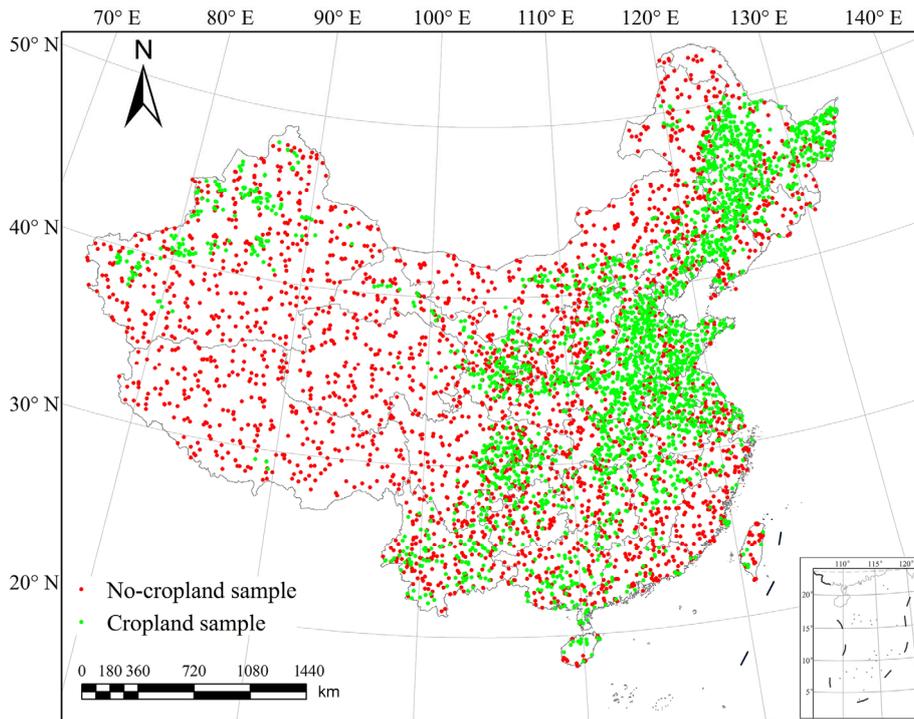


Figure 2 Spatial distribution of the validation samples.

map, the pixel value describes the numerical agreement that occurs for the five cropland datasets, and greater pixel values imply higher agreement among the datasets. For example, a pixel value of 5 indicates that all five data products are cropland in the pixel, while a value of 4 indicates that four of the cropland products are the same in the pixel. According to the pixel value, spatial agreement can be divided into five levels: complete agreement, high agreement, medium agreement, low agreement, and disagreement, which correspond to pixel values of 5, 4, 3, 2 and 1, respectively. Based on the agreement map, we analyzed the spatial consistencies of the five cropland datasets in the six regions of China.

In addition, although the cropland area is significantly affected by human activity, its spatial distribution is restricted by environmental factors, including topography and precipitation, which poses a challenge to cropland mapping. Therefore, in this study, we analyzed the variations of the spatial agreement of the five datasets based on elevation and slope. According to the elevation, the topography is divided into plain (<20 m), upland (20–200 m), low mountain (200–500 m), medium mountain (500–1500 m), and high mountain (>1500 m) (Chai, 1989). We analyzed the diversities of the spatial agreement of different topographies for all five cropland datasets. According to the “Technical Rules of Land Use Survey”, slopes were divided into five levels, including $\leq 2^\circ$, $2^\circ\text{--}6^\circ$, $6^\circ\text{--}15^\circ$, $15^\circ\text{--}25^\circ$, and $>25^\circ$, to analyze the distribution characteristics of the spatial agreement of the five cropland datasets for each slope interval.

4. Results and discussion

4.1 Comparison of the cropland area accuracy

Table 2 shows the proportions of the cropland area in each province and region relative to the national total area from all five datasets and their deviation compared with the statistical data. Overall, the differences in the proportions of cropland area between the cropland datasets and statistics are obvious in most of the provinces. The provinces with greater deviation include Qinghai, Gansu, and Yunnan. In comparison, the provinces with smaller cropland area deviation include Anhui, Shanxi, and Hebei. Compared with the statistical data, the proportion obtained through GlobeLand30 is relatively consistent with the statistical data, with respect to different levels of either overestimation or underestimation in the other four products. For FROM-GLC, the greatest deviation relative to the statistical data is found in the northeast region, which had an overestimation of 5.42% cropland area. However, cropland is underestimated in the north, southeast, and southwest regions, with deviations of -1.72% , -1.35% , and -1.99% , respectively, and the results obtained for the central and northwest regions are similar to the statistical data. For GlobCover, the cropland area is significantly underestimated in the northeast, with a deviation of -10.21% relative to the statistical data, whereas the deviation in terms of overestimated area in the northwest reaches 11.62%. In addition, MODIS Collection 5 overestimates the cropland area in the northeast by 6.42% and underestimates the cropland area in the south-west and northwest

Table 2 Proportions of cropland area (%) for different provinces and regions calculated by five datasets and their deviations relative to the statistical data

| Region | Province | Globeland30 | | FROM-GLC | | GlobCover | | MODIS Collection 5 | | MODIS Cropland | | Statistics |
|-----------|----------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------------|--------------|----------------|---------------|--------------|
| | | Proportion | Deviation | Proportion | Deviation | Proportion | Deviation | Proportion | Deviation | Proportion | Deviation | |
| Northeast | Inner Mongolia | 6.71 | 0.84 | 12.31 | 6.44 | 6.34 | 0.47 | 5.72 | -0.15 | 11.93 | 6.05 | 5.87 |
| | Heilongjiang | 9.15 | -0.57 | 8.80 | -0.92 | 3.30 | -6.42 | 13.82 | 4.10 | 12.31 | 2.59 | 9.72 |
| | Jilin | 4.35 | -0.20 | 4.37 | -0.18 | 1.47 | -3.08 | 5.18 | 0.64 | 5.67 | 1.12 | 4.55 |
| | Liaoning | 3.50 | 0.15 | 3.44 | 0.09 | 2.17 | -1.19 | 5.19 | 1.84 | 6.48 | 3.12 | 3.36 |
| | Total | 23.72 | 0.23 | 28.92 | 5.42 | 13.28 | -10.21 | 29.92 | 6.42 | 36.38 | 12.89 | 23.49 |
| North | Beijing | 0.27 | 0.08 | 0.26 | 0.07 | 0.29 | 0.10 | 0.28 | 0.09 | 0.33 | 0.14 | 0.19 |
| | Tianjin | 0.38 | 0.01 | 0.34 | -0.02 | 0.30 | -0.06 | 0.44 | 0.08 | 0.51 | 0.15 | 0.36 |
| | Hebei | 4.92 | -0.27 | 5.24 | 0.05 | 5.46 | 0.27 | 5.01 | -0.18 | 8.68 | 3.49 | 5.19 |
| | Shanxi | 3.42 | 0.09 | 3.79 | 0.46 | 4.51 | 1.18 | 2.89 | -0.44 | 3.13 | -0.20 | 3.33 |
| | Shandong | 6.10 | -0.08 | 4.94 | -1.23 | 5.38 | -0.79 | 7.26 | 1.08 | 10.93 | 4.76 | 6.17 |
| | Henan | 5.36 | -1.15 | 5.46 | -1.06 | 5.78 | -0.73 | 7.14 | 0.62 | 9.76 | 3.25 | 6.51 |
| | Total | 20.44 | -1.32 | 20.04 | -1.72 | 21.72 | -0.04 | 23.02 | 1.26 | 33.34 | 11.58 | 21.76 |
| Southeast | Shanghai | 0.17 | -0.03 | 0.17 | -0.03 | 0.20 | 0.00 | 0.22 | 0.02 | 0.05 | -0.15 | 0.20 |
| | Jiangsu | 3.60 | -0.32 | 3.22 | -0.70 | 3.31 | -0.60 | 4.33 | 0.42 | 2.40 | -1.52 | 3.91 |
| | Zhejiang | 1.61 | 0.04 | 1.46 | -0.12 | 1.39 | -0.19 | 1.14 | -0.44 | 0.07 | -1.50 | 1.58 |
| | Fujian | 1.21 | 0.11 | 1.01 | -0.08 | 0.52 | -0.58 | 0.65 | -0.44 | 0.02 | -1.07 | 1.09 |
| | Guangdong | 2.36 | 0.03 | 2.01 | -0.32 | 1.48 | -0.84 | 2.79 | 0.46 | 0.04 | -2.29 | 2.33 |
| | Hainan | 0.39 | -0.21 | 0.50 | -0.10 | 0.33 | -0.27 | 0.72 | 0.12 | 0.03 | -0.57 | 0.60 |
| | Total | 9.34 | -0.37 | 8.36 | -1.35 | 7.23 | -2.48 | 9.84 | 0.13 | 2.60 | -7.11 | 9.71 |
| Central | Anhui | 4.12 | -0.59 | 3.99 | -0.72 | 3.94 | -0.77 | 4.97 | 0.26 | 3.73 | -0.98 | 4.71 |
| | Jiangxi | 2.32 | -0.01 | 2.44 | 0.12 | 2.16 | -0.17 | 2.22 | -0.10 | 0.16 | -2.16 | 2.32 |
| | Hubei | 4.04 | 0.21 | 3.74 | -0.09 | 4.39 | 0.55 | 4.16 | 0.32 | 1.81 | -2.02 | 3.83 |
| | Hunan | 3.24 | 0.13 | 4.26 | 1.15 | 3.87 | 0.76 | 2.65 | -0.46 | 0.06 | -3.05 | 3.11 |
| | Total | 13.72 | -0.26 | 14.43 | 0.45 | 14.35 | 0.38 | 14.00 | 0.02 | 5.76 | -8.21 | 13.98 |
| Southwest | Guangxi | 3.31 | -0.16 | 2.73 | -0.74 | 2.48 | -0.98 | 2.37 | -1.10 | 0.14 | -3.33 | 3.47 |
| | Chongqing | 2.01 | 0.18 | 1.44 | -0.39 | 1.64 | -0.20 | 1.68 | -0.16 | 0.05 | -1.79 | 1.84 |
| | Sichuan | 5.82 | 0.93 | 5.41 | 0.52 | 7.65 | 2.77 | 5.23 | 0.35 | 3.56 | -1.33 | 4.89 |
| | Guizhou | 2.95 | -0.74 | 2.72 | -0.96 | 3.06 | -0.62 | 2.41 | -1.27 | 0.32 | -3.37 | 3.69 |
| | Yunnan | 5.23 | 0.24 | 4.57 | -0.42 | 4.76 | -0.23 | 3.13 | -1.86 | 0.68 | -4.31 | 4.99 |
| | Total | 19.31 | 0.45 | 16.88 | -1.99 | 19.60 | 0.73 | 14.81 | -4.05 | 4.74 | -14.12 | 18.86 |
| Northwest | Tibet | 0.22 | -0.08 | 0.22 | -0.08 | 3.88 | 3.58 | 0.11 | -0.19 | 1.26 | 0.96 | 0.30 |
| | Shaanxi | 3.10 | -0.22 | 3.97 | 0.64 | 4.54 | 1.21 | 2.79 | -0.54 | 2.75 | -0.58 | 3.33 |
| | Gansu | 4.16 | 0.34 | 2.74 | -1.09 | 4.46 | 0.64 | 1.75 | -2.08 | 4.49 | 0.66 | 3.83 |
| | Qinghai | 0.57 | 0.13 | 0.19 | -0.26 | 5.54 | 5.09 | 0.11 | -0.34 | 3.27 | 2.82 | 0.45 |
| | Ningxia | 1.09 | 0.18 | 0.84 | -0.07 | 0.40 | -0.51 | 0.26 | -0.65 | 0.70 | -0.22 | 0.91 |
| | Xinjiang | 4.33 | 0.94 | 3.41 | 0.03 | 5.00 | 1.61 | 3.40 | 0.01 | 4.71 | 1.32 | 3.39 |
| | Total | 13.47 | 1.28 | 11.38 | -0.82 | 23.82 | 11.62 | 8.42 | -3.78 | 17.18 | 4.98 | 12.20 |

regions by 4.05% and 3.78%, respectively. Particularly, the results obtained through MODIS Cropland show significant discrepancies with the statistical data. The areas in the northeast, north, and northwest are overestimated by 12.89%, 11.58%, and 4.98%, respectively, while areas in the south, central, and southwest regions are underestimated by -7.11%, -8.21%, and -14.12%.

The RMSEs and correlation coefficients were calculated to assess the dispersion and goodness of fit between the cropland areas estimated from the five datasets and the statistical data, respectively, as shown in Figure 3. The dashed line is the 1:1 line that reflects the dispersion qualitatively. The RMSE values of GlobeLand30, FROM-GLC, GlobCover, MODIS Collection 5, and MODIS Cropland compared with the statistical data are 0.43%, 1.29%, 1.89%, 1.07%, and 2.46%, respectively. Higher RMSE values rep-

resent more significant dispersion from the statistical data. Hence, GlobeLand30 has the lowest dispersion because it results in the smallest RMSE value, and the data points for GlobeLand30 are mainly distributed on or near the 1:1 line, as shown in Figure 3a. The dispersion of MODIS Collection 5 is also relatively low, as shown in Figure 3d, with a distribution on or near the 1:1 line; however, a large dispersion can be observed for a few of the data points. By comparison, MODIS Cropland has the greatest dispersion because it has its highest RMSE, and few data points are on or near the 1:1 line, as shown in Figure 3e. The solid line is fitted according to the cropland proportion for each dataset and the statistical data. The correlation coefficients for the GlobeLand30, FROM-GLC, GlobCover, MODIS Collection 5, and MODIS Cropland compared with the statistical data are 0.96, 0.70, 0.29, 0.82, and 0.25, respectively. In

general, greater correlation coefficients represent increasingly preferable goodness of fits with the statistical data. The best goodness of fit was obtained for the proportion of cropland area data from GlobeLand30, followed by MODIS Collection 5. The poorest fit occurred when using MODIS Cropland. From Figure 3, it is easy to observe that GlobeLand30 has the lowest dispersion and greatest goodness of fit, which implies an accurate result for the cropland area. In addition, the correlation coefficient calculated by the MODIS Collection 5 dataset, with a spatial resolution of 500m, is higher than the correlation coefficients calculated by the 30 m resolution FROM-GLC dataset, 300 m resolution GlobCover dataset, and 250 m resolution MODIS Cropland dataset, indicating that excellent classification results can also be achieved when using the proper technique, even with the use of low-resolution data.

4.2 Comparison of the spatial location accuracy

The spatial location accuracies of the five cropland datasets are obtained by error matrix using the samples shown in Figure 2. Table 3 lists the results including overall accuracy, Kappa coefficient, commission error, and omission error. According to this table, the best accuracy is obtained by GlobeLand30, with an overall accuracy of 79.61% and a Kappa coefficient of 0.58. In addition, the overall accuracy

and Kappa coefficient of FROM-GLC reach 76.23% and 0.52, and the overall accuracy and Kappa coefficient of MODIS Collection 5 are 73.49% and 0.47, respectively. Comparatively, the accuracies obtained with GlobCover and MODIS Cropland are relatively lower. The overall accuracy and Kappa coefficient for GlobCover are 70.13% and 0.39, respectively, and the values for MODIS Cropland are 67.99% and 0.38, respectively. In addition, the largest commission error of 28.97% is found for GlobCover because this dataset significantly overestimated the cropland area in the northwest region, as shown in Table 2. Moreover, the highest omission error of 49.86% occurs for MODIS Cropland because of the significant underestimation of cropland area in the southeast, central, and southwest regions.

Figure 4 provides the overall accuracies of the five datasets in the northeast, north, southeast, central, southwest, and northwest regions. In the northeast, the best overall accuracy, 79.40%, is obtained through FROM-GLC, and the overall accuracy of GlobCover is the lowest because of the high commission error. In north China, GlobCover provides the highest accuracy of 78.90%. In southeast China, the best overall accuracy is achieved through MODIS Collection 5 (68.72%). GlobeLand30 provides the best overall accuracies in the central, southwest, and northwest regions, corresponding to accuracies of 78.73%, 66.41%, and 86.62%,

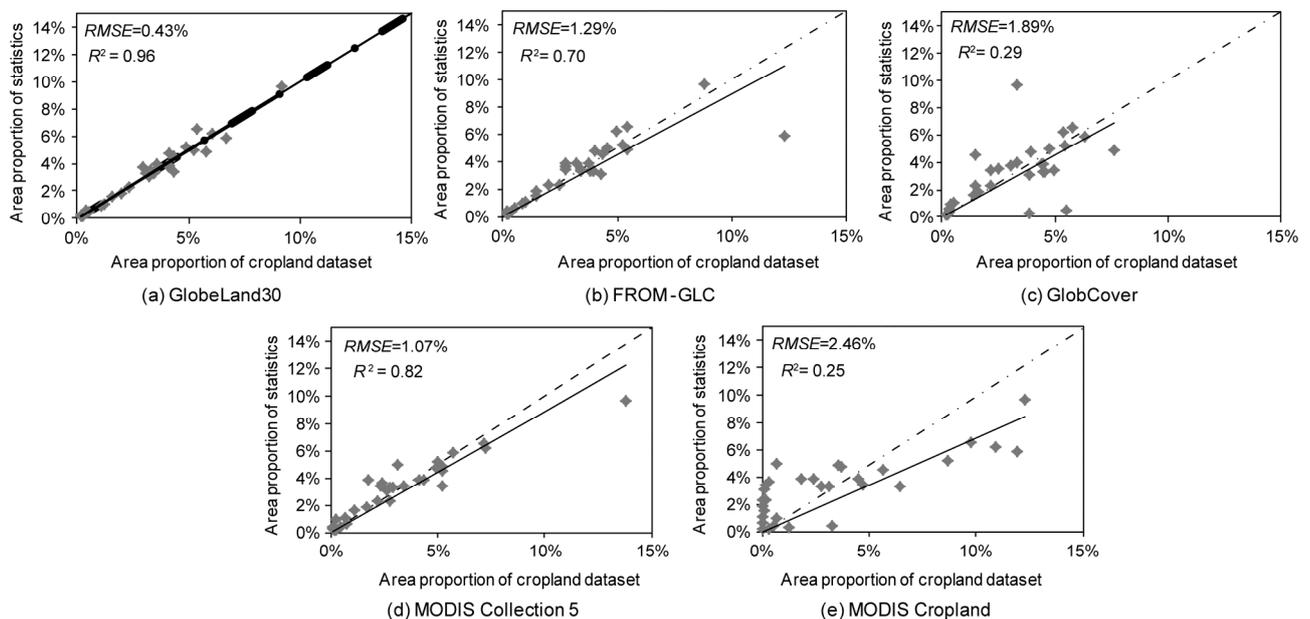


Figure 3 Analyses of dispersion and correlation of the area proportion between the cropland dataset and statistical data.

Table 3 The spatial location accuracies of the five cropland datasets

| | GlobeLand 30 | FROM-GLC | GlobCover | MODIS Collection 5 | MODIS Cropland |
|----------------------|--------------|----------|-----------|--------------------|----------------|
| Overall accuracy (%) | 79.61 | 76.23 | 70.13 | 73.49 | 67.99 |
| Kappa coefficient | 0.58 | 0.52 | 0.39 | 0.47 | 0.38 |
| Commission error (%) | 19.70 | 21.66 | 28.97 | 20.79 | 13.68 |
| Omission error (%) | 21.33 | 21.11 | 22.22 | 29.31 | 49.86 |

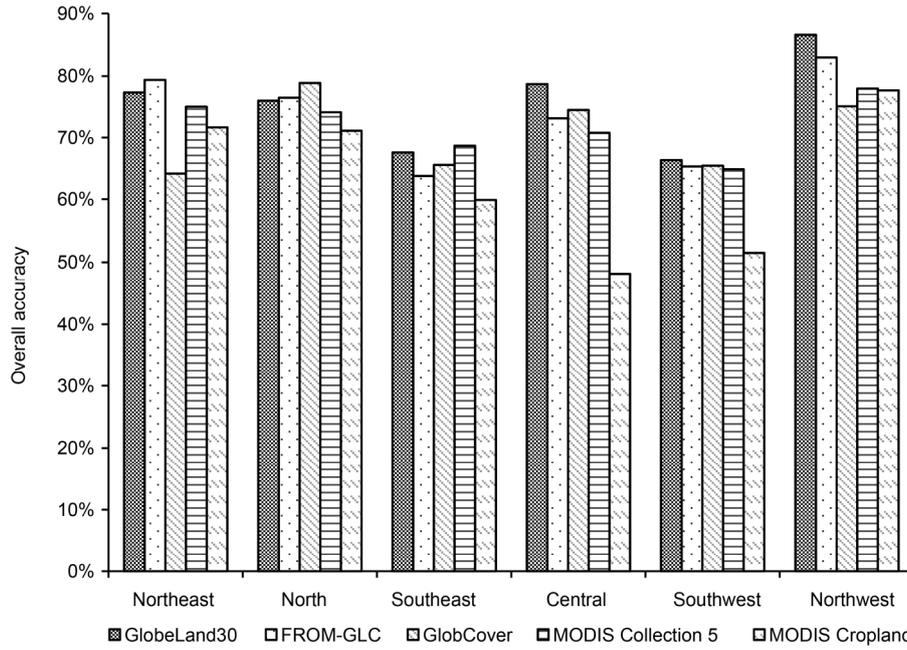


Figure 4 Overall accuracies of the five cropland datasets in the northeast, north, southeast, central, southwest, and northwest regions.

respectively. In the six regions, the overall accuracies obtained through all five products are higher in the northwest and central regions, while those in the southwest region are relatively lower. Among the five products, GlobeLand30 exhibits the higher overall accuracy in each region, while the lower overall accuracy is achieved with MODIS Cropland.

4.3 Comparison of the spatial agreement

Figure 5 is the spatial agreement map of the five datasets. The higher level of consistency is obtained in the major grain producing areas, including the Northeast Plain, the Huang-Huai-Hai Plain, the Yangtze River Basin, and the Weihe Plain, in which the cropland is distributed intensively and continuously with a simple geographical landscape, which leads to accurate results from all five products for the extraction of cropland data. The spatial consistency become lower from the major grain producing area to the farming-pastoral zone, and the pastoral zone has the poorest consistency. For example, the spatial consistency of the five datasets is good in the major grain producing area of the Northeast Plain, including Heilongjiang, Jilin, and Liaoning provinces, and is lower in the north farming-pastoral zone located in the south rim of the Inner Mongolia Plateau and the region along the Great Wall. However, the worst consistency of the five products is obtained in the pastoral zone of the Inner Mongolia Plateau.

Based on Figure 5, the distribution of the spatial agreement of the five products is quantitatively analyzed at the regional scale (Figure 6). In the six regions, the proportion of the cropland area in the north region is increased, with an

increasing agreement value. In this region, the proportion of complete agreement reaches up to 39.36%, implying that the best consistency of the five products is obtained in the north region. In contrast, this proportion is reduced with the increasing agreement value in the northwest and southwest regions, yielding poor consistency. The highest disagreement is found in the northwest region (61.79%), while the percentage of the area with complete agreement is only 5.75%. In the southwest, the ratio of the disagreement cropland area is 41.27%, with only 2.22% in complete agreement. In the northwest and southwest regions, the topography is complicated, and the landscape is fragmented, leading to universal mixed pixels in remote sensing imagery. Hence, the classification of cropland is difficult in these regions, resulting in poor consistency among the five products.

Figure 7 shows the variations in the spatial agreements of the five datasets with elevation. Plains with an elevation lower than 20 m and uplands with an elevation range of 20–200 m have large proportions of high and complete agreement, implying significant consistency among the five datasets in these areas. The plains and upland areas are mainly distributed in most provinces of North and Central China, and some provinces of the Northeast (including Heilongjiang, Liaoning and Jilin) and South China (such as Jiangsu, Shanghai, and Zhejiang) which are the main grain producing areas. Because the landscape is homogeneous and cropland is the dominant land cover type in these areas, cropland can be accurately classified. For the low mountain area with an elevation range of 200–500 m, the agreement is reduced with increasing elevation. The medium mountain area, with an elevation range of 500–1500 m, is distributed

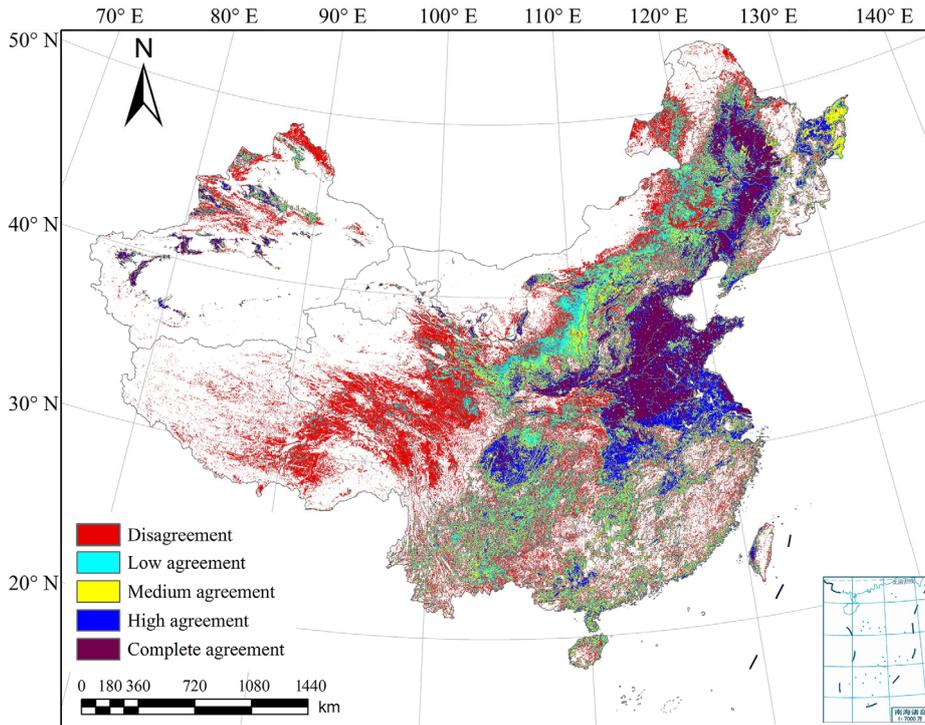


Figure 5 Spatial agreement map of the five cropland products.

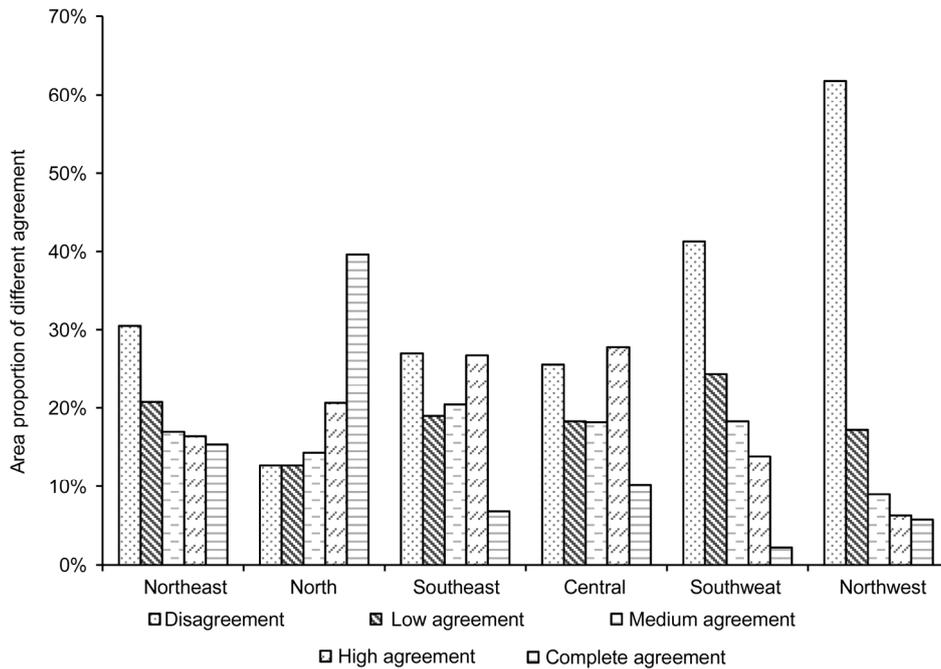


Figure 6 Analysis of spatial agreement in the six regions.

within the Inner Mongolia Plateau, the Tarim Basin in Xinjiang, the Loess Plateau, and the Yunnan-Guizhou Plateau. Except for the Inner Mongolia Plateau and the Tarim Basin, it is difficult to extract cropland in other areas because of the multi-mountain topography and fragmented geography, resulting in disagreement of up to 40.16% in this region.

The high mountain region, with an elevation greater than 1500 m, is mainly distributed in the northwest Tibet Plateau, with a disagreement of 70.76%.

Slope is an important factor affecting cropland utilization. Figure 8 shows the spatial agreement distribution with different slope intervals. In general, the consistency of the

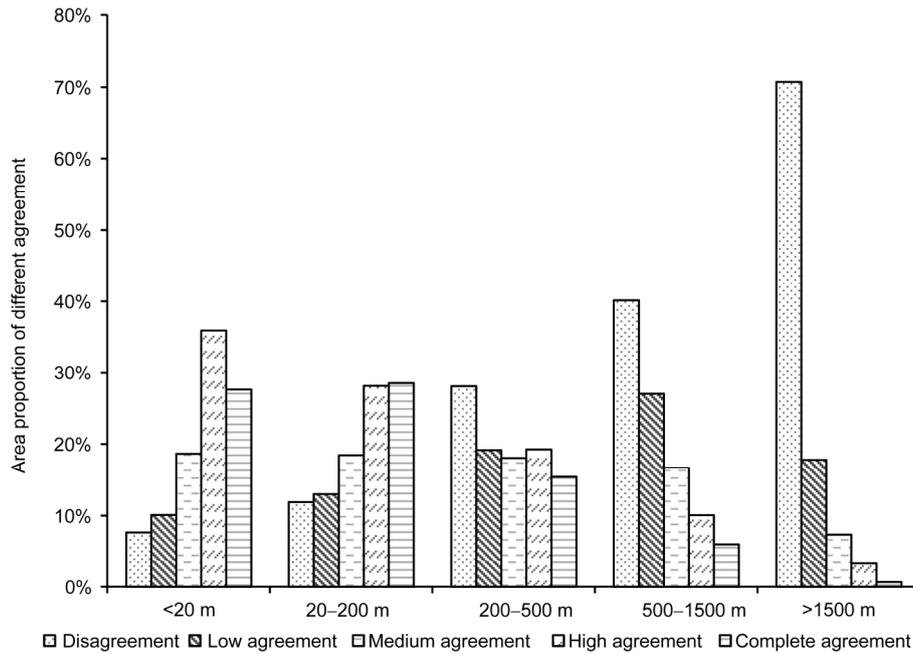


Figure 7 Analysis of spatial agreement in different elevation intervals.

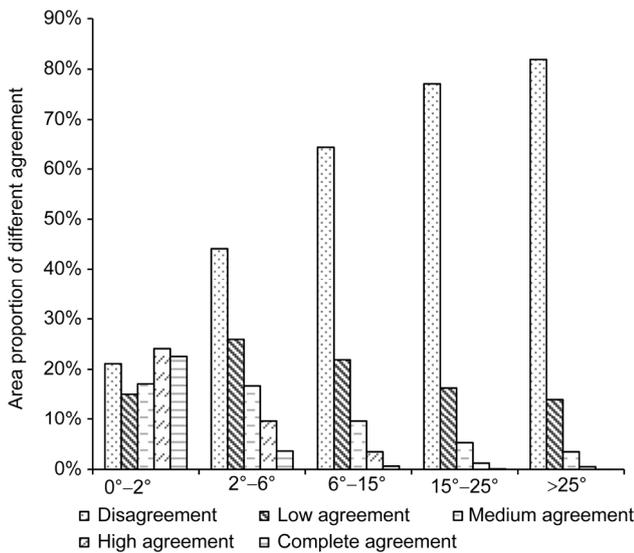


Figure 8 Analysis of spatial agreement with different slope intervals.

five datasets decreases with increasing slope. The area with slope values below 2° is mainly distributed in the plains and basins in China. Because of the gentle terrain and homogeneous geographical landscape, the consistency of the five cropland datasets is high, and the complete agreement and high agreement are 24.09% and 22.55%, respectively. In the second slope interval, which is $2^\circ-6^\circ$, the disagreement increases to 41.11%, and the complete agreement decreases to 3.66%. Furthermore, the areas with slope between $15^\circ-25^\circ$ and over 25° are mainly distributed in the mountainous area of the Tibet Plateau, where the proportions of disagreement of the five datasets are high, corresponding to 77.09% and

81.92%.

5. Conclusion

In this paper, we compared five cropland datasets in circa 2010 for China, GlobeLand30, FROM-GLC, GlobCover, MODIS Collection 5, and MODIS Cropland. The results indicated that GlobeLand30 performed better than the other four datasets, regardless of the cropland area and the spatial location. Regarding the comparison of the cropland area accuracy, the best goodness of fit with the statistical data was obtained through GlobeLand30, followed by MODIS Collection 5 and FROM-GLC, with the poorest fits achieved by MODIS Cropland and GlobCover. Regarding the spatial location accuracy, the highest accuracy was found using the GlobeLand30 dataset, followed by FROM-GLC and MODIS Collection 5, while the poorest accuracy was obtained using GlobCover and MODIS Cropland. The spatial agreement of the five datasets was reduced from the major grain producing areas to pastoral zones and was significantly affected by the elevation and slope.

The spatial resolution of the imagery is believed to significantly affect the cartographical accuracy of the land cover remote sensing product (Hansen and Loveland, 2012; Ban et al., 2015). Although GlobeLand30 and FROM-GLC use similar images for land cover mapping, the diversity of the classification method results in significant differences in cropland area and spatial location between the two datasets. Moreover, regarding cropland area, the MODIS Collection 5 dataset with the lower spatial resolution had a better

goodness of fit with the statistical data compared with that from the FROM-GLC, GlobCover, and MODIS Cropland datasets. For the accuracy of spatial location, the overall accuracy and Kappa coefficient of MODIS Collection 5 were better than those obtained with GlobCover and MODIS Cropland. Therefore, it is reasonable to consider that improving the classification method is important for enhancing the accuracy of cropland mapping and that images with lower spatial resolution can deliver better results if the classification method is advanced. In north and northeast China, characterized by low elevations, gentle slopes and homogeneous landscapes, the cropland is distributed continuously and intensively with high separability. Therefore, the spatial agreements of these five products in the two regions are better than the spatial agreements in other regions. By comparison, the northwest and southwest regions are characterized as having high elevations, large slopes, and complicated topographies with fragmented geographical landscapes, and the cropland is heavily mixed with other land cover types with lower separability of the remote sensing spectra, resulting in low accuracies of cropland area and spatial location. For example, FROM-GLC largely misclassifies pasture as cropland in Inner Mongolia, MODIS Cropland has a high omission error of cropland in southeast and southwest China, and GlobCover confuses cropland and other vegetation types in southwest and northwest China. Therefore, regions with fragmented topography and heterogeneous landscapes require additional attention in future remote sensing cropland mapping studies.

Acknowledgements *This study was supported by the National Natural Science Foundation of China (Grant No. 41501483) and Research Foundation for Mapping Geographic Information Public Welfare of China (Grant No. 201512028).*

References

- Ban Y F, Gong P, Giri C. 2015. Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS-J Photogramm Remote Sens*, 103: 1–6
- Bartholome E, Belward A S. 2005. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int J Remote Sens*, 26: 1959–1977
- Bicheron P, Defourny P, Brockmann C, Schouten L, Vancutsem C, Huc M, Bontemps S, Leroy M, Achard F, Herold M, Ranera F, Arino O. 2008. *GlobCover Products Description and Validation Report*
- Bontemps S, Defourny P, Bogaert E V, Arino O, Kalogirou V, Perez J R. 2011. *GLOBCOVER 2009 Products Description and Validation Report*
- Cao X, Chen J, Chen L J, Liao A P, Sun F D, Li Y, Li L, Lin Z, Pang Z, Chen J, He C Y, Peng S. 2014. Preliminary analysis of spatiotemporal pattern of global land surface water. *Sci China Earth Sci*, 57: 2330–2339
- Chai Z X. 1989. *A Proposal to Divide the Basic Forms of Landscape by Relative Height* (in Chinese). Beijing: Surveying and Mapping Press
- Chen J, Chen J, Liao A P, Cao X, Chen L J, Chen X H, He C Y, Han G, Peng S, Lu M, Zhang W, Tong X H, Mills J. 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS-J Photogramm Remote Sens*, 103: 7–27
- Chen J, Lu M, Chen X H, Chen J, Chen L J. 2013. A spectral gradient difference based approach for land cover change detection. *ISPRS J Photogramm Remote Sens*, 85: 1–12
- Chen J, Chen J, Gong P, Liao A P, He C Y. 2011. Higher resolution global land cover mapping (in Chinese). *Geomatics World*, 2: 12–14
- Congalton R G, Gu J, Yadav K, Thenkabail P, Ozdogan M. 2014. Global land cover mapping: A review and uncertainty analysis. *Remote Sens-Basel*, 6: 12070–12093
- Foody G M. 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sens Environ*, 114: 2271–2285
- Friedl M A, Mciver D K, Hodges J C F, Zhang X Y, Muchoney D, Strahler A H, Woodcock C E, Gopal S, Schneider A, Cooper A, Baccini A, Gao F, Schaaf C. 2002. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens Environ*, 83: 287–302
- Friedl M A, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ*, 114: 168–182
- Fritz S, See L, You L, Justice C, Becker-Reshef I, Bydekerke L, Cumani R, Defourny P, Erb K, Foley J, Gilliams S, Gong P, Hansen M, Hertel T, Herold M, Herrero M, Kayitakire F, Latham J, Leo O, McCallum I, Obersteiner M, RamanKutty N, Rocha J, Tang H, Thornton P, Vancutsem C, Velde M, Wood S, Woodcock C. 2013. The need for improved maps of global cropland. *Eos Trans Am Geophys Union*, 94: 31–32
- Giri C, Pengra B, Long J, Loveland T R. 2013. Next generation of global land cover characterization, mapping, and monitoring. *Int J Appl Earth Obs*, 25: 30–37
- Giri C, Zhu Z, Reed B. 2005. A comparative analysis of the global land cover 2000 and MODIS land cover datasets. *Remote Sens Environ*, 94: 123–132
- Gong P, Wang J, Yu L, Zhao Y C, Zhao Y Y, Liang L, Niu Z G, Huang X M, Fu H H, Liu S, Li C C, Li X Y, Fu W, Liu C X, Yue X, Wang X Y, Cheng Q, Hu L Y, Yao W B, Zhang H, Zhu P, Zhao Z Y, Zhang H Y, Zheng Y M, Ji L Y, Zhang Y W, Chen H, Yan A, Guo J H, Liang Y, Wang L, Liu X J, Shi T T, Zhu M H, Chen Y L, Yang G W, Tang P, Xu B, Giri C, Clinton N, Zhu Z L, Chen J, Chen J. 2013. Finer resolution observation and monitoring of GLC: First mapping results with Landsat TM and ETM+ data. *Int J Remote Sens*, 34: 2607–2654
- Hansen M C, Loveland T R. 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sens Environ*, 122: 66–74
- Hansen M C, Reed B. 2000. A comparison of the IGBP DISCover and university of Maryland 1 km global land cover products. *Int J Remote Sens*, 21: 1365–1373
- Hansen M C, Defries R S, Townshend J R G, Sohlberg R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int J Remote Sens*, 21: 1331–1364
- Herold M, Mayaux P, Woodcock C E, Baccini A, Schmullius C. 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
- Jung M, Henkel K, Herold M, Churkina G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sens Environ*, 101: 534–553
- Li B B, Fang X Q, Ye Y, Zhang X Z. 2010. Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data—A case study in northeast china. *Sci China Earth Sci*, 53: 1689–1699
- Liao A P, Chen L J, Chen J, He C Y, Cao X, Chen J, Peng S, Sun F D, Gong P. 2014. High-resolution remote sensing mapping of global land water. *Sci China Earth Sci*, 57: 2305–2316
- Loveland T R, Reed B C, Brown J F, Ohlen D O, Zhu Z, Yang L, Merchant J W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int J Remote Sens*, 21: 1303–1330
- Pflugmacher D, Krankina O N, Cohen W B, Friedl M A, Sulla-Menashe D, Kennedy R E, Nelson P, Loboda T V, Kueemmerle T, Dyukarev E, Elsakov V, Kharuk V I. 2011. Comparison and assessment of coarse resolution land cover maps for northern Eurasia. *Remote Sens Environ*, 115: 3539–3553

- Pittman K, Hansen M C, Becker-Reshef I, Potapov P V, Justice C O. 2010. Estimating global cropland extent with multi-year MODIS data. *Remote Sens-Basel*, 2: 1844–1863
- Rengarajan R, Sampath A, Storey J, Choate M. 2015. Validation of geometric accuracy of global land survey (GLS) 2000 data. *Photogramm Eng Remote Sens*, 81: 131–141
- Sheng L, Yan J. 2010. *Regional Economical Statistical Yearbook of China 2010* (in Chinese). Beijing: China Statistics Press. 86
- Song H, Zhang X. 2012. Precision analysis and validation of multi-sources land cover products derived from remote sensing in China (in Chinese). *Trans Chin Soc Agric Eng*, 28: 207–214
- Tang H J, Wu W B, Yu Q Y, Xia T, Yang P, Li Z G. 2015. *Key Research Priorities for Agricultural Land System Studies* (in Chinese). *Sci Agr Sin*, 48: 900–910
- Tchuente A T K, Roujean J L, Jong S M D. 2011. 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 Obs*, 13: 207–219
- Wang Z, Liu L. 2014. Assessment of coarse-resolution land cover products using CASI hyperspectral data in an arid zone in northwestern China. *Remote Sens-Basel*, 6: 2864–2883
- Wu W B, Shibasaki R, Yang P, Ongaro L, Zhou Q B, Tang H. 2008. Validation and comparison of 1 km global land cover products in China. *Int J Remote Sens*, 29: 3769–3785
- Wu W B, Yang P, Zhang L, Tang H J, Zhou Q B, Ryosuke S. 2009. Accuracy assessment of four global land cover datasets in China (in Chinese). *Trans Chin Soc Agric Eng*, 25: 167–173
- Yang Y, Xiao P, Feng X, Li H X, Chang X, Feng W. 2014. Comparison and assessment of large-scale land cover datasets in China and adjacent regions. *J Remote Sens*, 18: 453–475
- Yu L, Wang J, Gong P. 2013. Improving 30 m global land cover map FROM-GLC with time series MODIS and auxiliary datasets: A segmentation based approach. *Int J Remote Sens*, 34: 5851–5867
- Yu L, Wang J, Li X, Li C, Zhao Y, Gong P. 2014. A multi-resolution global land cover dataset through multisource data aggregation. *Sci China Earth Sci*, 57: 2317–2329