

Dear Reviewer #4:

We would like to express our sincere appreciations for your efforts on reviewing our submission. As shown in the revision, we carefully considered the concerning points and made efforts to improve the rigor, logic, and clarity of this manuscript, and carefully responded to all of the comments and concerns.

Thanks again to the careful inspection and professional comments, the normalization problems of figures and table are corrected accordingly. We made changes to keep it clear, organized, and easy to understand.

According to your comments/ suggestions, we made clarifications in the manuscript and attach the detailed item-by-item response to all comments and suggestions for the evaluation.

Yours sincerely,

Chunqiao Song and co-authors

This study undertakes a comprehensive and in-depth assessment of cultivated land classification accuracy within Northeast China, utilizing key metrics including Overall Accuracy (OA), Producer Accuracy, and User Accuracy across a diverse array of 13 global and national-scale land cover datasets. The precision of cultivated land delineation assumes paramount significance in the realms of food security, agricultural strategizing, and resource management, thus conferring a substantive research imperative upon this study.

However, as a study centered on the evaluation of product accuracy, it is imperative for the authors to comprehensively address potential factors that could influence the evaluation results during the experimental phase itself, rather than reserving their impact for the discussion section. For instance, challenges related to divergent field definitions, classifications, and mismatches between validation sample intervals and product data can be effectively resolved through an expansion of the experimental design.

Furthermore, although this study successfully identified datasets that exhibited high accuracy, there is a pressing need for a more extensive comparative analysis. This entails offering a more comprehensive description and analysis of the strengths and weaknesses inherent in each dataset, accompanied by a robust discussion elucidating the underlying reasons for the disparities in dataset performance.

Reply: Sincere thanks for the concise summary and positive comment on this work. The concerning points raised by the reviewer are very helpful for us to improve the manuscript. We carefully addressed these points listed below and made changes accordingly.

We have previously taken into account this matter, specifically prior to comparing the data, identifying the factors that contribute to the accuracy or inaccuracy of the product within a specific location or category. In this study, we attribute the possible influencing factors mainly to four aspects. Hence, we separate and extend this part of analyses on assessing the product accuracy of these ten types of products while analyzing their influencing factors as a single section (Section 4.1 & 4.2) in the discussion. In this part, we incorporate overall accuracy comparison based on different factors, including the spatial resolution, satellite/sensor type, product time, and mapping algorithm.

Minor:

Line 135: Figure 1.

Figure 1: The quality of Figure 1 needs to be improved, such as the use of map symbols, and the proper use of legends.

Reply: Thanks for your suggestion, the figure was updated accordingly.

Section 2.2.1: In addition to the temporal coverage summary provided in Table 1, it is essential to include a corresponding explanation within the main body of the description.

Reply: Thanks for your suggestion, we have added corresponding explanation in the body of the description.

Line 225: 2.2.2 Other auxiliary dataset

Reply: It was corrected accordingly.

Line 245: What is the rationale behind selecting 2,000 samples for each year? Please provide an explanation for the appropriateness of this sample size. Additionally, could you discuss whether the distribution of samples across different crops has an impact on verification accuracy?

Reply: Sincere thanks for the reviewer's concern on the validation samples. It is a very key point. We understand the assessment of classification accuracy of land cover types based on larger sample size may produce more robust results. We selected about 2,000 samples for each period by compromising the assessment robustness and time cost. Each period of validation points derived from high-resolution imagery over the study area took the researcher approximately 60 hours. That is to say, we spent more than four weeks of full-time work to generate the four periods of validation points. The determination of assessment sample size is a balance of the time cost and spatial representativeness. We generated the points by spatially random sampling over the study area boundary and then manually determine the types of sampling points over high-resolution imagery one by one. Regretfully, only based on single-date optical images, we cannot derive the precise results of crop land types (e.g., corn, maize, rice, etc.), thus we did not assess the accuracy of land cover products by crop types. However, the reviewer indeed raises an important issue. In the future work, we may choose several sets products with high accuracy and further assess their performances by crop types.

Section 2.4.1 and 2.4.3: The formula does not specify the parameters used (such as X_{+i} , X_{i+} , x_i and y_i), and it is possible that the author has misreported or incorrectly described these parameters.

Reply: Thanks for your suggestion, we have checked and made corresponding modifications.

Figure 3: It is advisable to present the distribution of sample points for each period separately. If concerns regarding text length arise, this information can alternatively be included in an appendix.

Reply: The suggestion is appreciated and followed. We update Figure 3 accordingly in the revision.



Article

Comparison and assessment of different land cover datasets on the cropland in Northeast China

Peipei Cui ^{1,2}, Tan Chen ¹, Yingjie Li ², Kai Liu ¹, Dapeng Zhang ¹ and Chunqiao Song ^{1,3,4*}

¹ Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

² School of Geography, Geomatics and Planning, Jiangsu Normal University, 101 Shanghai Road, Tongshan District, Xuzhou, Jiangsu 221116, China

³ [University of Chinese Academy of Sciences, Nanjing \(UCASNJ\), Nanjing 211135, China](#)

⁴ [University of Chinese Academy of Sciences, Beijing 100049, China](#)

* Correspondence: Correspondence: cqsong@niglas.ac.cn; Tel.: +86-25-86882293

Abstract: ~~Accurate and reliable cropland information on extent and distribution is essential for food security evaluation, agricultural planning, resource management.~~ The provision of precise and dependable information regarding the extent and distribution of cropland is imperative for the evaluation of food security, agricultural planning, and resource management. Cropland is one of the most critical elements of land cover type and can be available from many existing global/regional land cover products. Cropland is an important component of land cover type and is offered in several multiple existing global/regional land cover products. However, global-scale accuracy evaluation may not be representative for class-specific or local-area accuracy, such as Northeast China, which is an important grain-producing region of China and has various types of cultivated land (e.g., wheat, rice) and diverse terrains. It poses a great challenge ~~on-in~~ generating precise cropland classification by automated mapping. Thus, it is indispensable to evaluate the accuracy and reliability in these various land cover datasets before using them. In this study, we collected thirteen sets of global or national-scale land cover datasets. Through visual interpretation of high-resolution ~~image images~~, ground “truth” samples were collected to evaluate the data accuracy across Northeast China. The overall accuracy (OA) evaluation results in Phase-2020 shows that CLCD has the highest value with 0.914, followed by GlobeLand30 (0.906), GLC_FCS30 (0.902), and Esri (0.896) for cropland classification in Northeast China. CGLS-LC100 has the lowest OA (0.710). For the commission and omission errors of six datasets of Phase-2020, CGLS-LC100 ~~have-has~~ an obvious overestimation (larger commission error), while the two national-scale datasets (CLCD and CLUDs) suffers from slight underestimation of cropland area and CLCD performs best relatively better. In the aspect of spatial consistency, high spatial agreement among the nine Phase-2015 datasets or in the six Phase-2020 datasets could be discovered in traditional agricultural regions like the Sanjiang-Songnen-Liaohe Plain, and low agreement is found in the transition areas of mountains (hills) and plains with the ~~mix-mixed~~ landscape of forest (grassland) and farmland. In the aspect of comparison pairwise data, CLCD is in good agreement with GLC_FCS30, GlobeLand30, and Esri, while the CGLS-LC100 is in the poorest agreement with any other dataset. The comparison and evaluation results are expected to provide the reference on which aspects and what extent these land cover products may be consistent and guiding guide the cropland data product selection for Northeast China.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date



Copyright: © 2023 by the authors

Submitted for possible open access

publication under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

Keywords: land cover; Northeast China; cropland classification; accuracy assessment; spatial agreement; regional comparison

Abbreviations

GLC2000, Global Land Cover 2000; **FAO-GLCshare**, The Global Land Cover-share created by FAO (Food and Agriculture Organization, FAO); **FROM-GLC**, Finer Resolution Observation and Monitoring of Global Land Cover; **Esri**, a global map of land use/land cover (LULC) derived from ESA Sentinel-2 imagery at 10m resolution; **GLASS-GLC**, Global Land Surface Satellite-Global Land Cover; **GLCNMO**, Global Land Cover by National Mapping Organizations; **CCI-LC**, Climate Change Initiative-Land Cover; **GlobCover**, Global Land Cover Map; **CGLS-LC100**, Copernicus Global Land Service-Land Cover 100m; **GlobeLand30**, World's First Global Land Cover Datasets at a 30m; **GLC_FCS30**, Global Land Cover product with Fine Classification System at 30m; **CLUDs**, China's Land-use/cover datasets; **CLCD**, China Land Cover Dataset

1. Introduction

Agricultural land meets global demands such as human food, stock feed, and biofuel. These demands are growing rapidly due to the increasing ~~of~~ population and consumption [1,2]. Therefore, understanding the range, distribution and dynamics change of cropland ~~are-is~~ required for food demands and environmental sustainability, especially for agricultural planning, cropland monitoring, and food security assessment, among others. An increasing number of global and regional land use/cover (LULC) datasets [3], including ~~the~~ farmland category, have been produced by the researchers and ~~are~~ freely available to the public with the rapid development of remote sensing and new technologies (i.e., cloud computing, big data and machine learning). Coarse resolution images (i.e., pixels of 5 km or 1 km) were generally used in the early period [4-6]; however, ~~its-their~~ application at national or regional levels is limited by the low precision and image quality. Then, the spatial resolution of products was enhanced to 500 m or 300 m [7,8]. Nowadays, land cover mapping using high-resolution images, such as 30 m or 10 m resolution satellite data [9,10], has received increasing attention from the scientific community. All these datasets offer precious ~~croplands-cropland~~ information for multi-discipline ~~researches-research~~ and applications. However, there are considerable inconsistencies and uncertainties among these LULC datasets, especially in the areas with complex transition landscapes, since they have been generated using different ~~satellites-satellites~~/sensors, classification features, classifiers and methods, etc. [11,12]. A few previous studies have also shown that discrepancies and uncertainties of cropland also arise due to cropland definitions or different classification schemes [13-15]. Therefore, the accuracy assessment and comparative studies of these LULC datasets are necessary for relevant application research before using them [16,17].

In the existing studies, many researchers have performed accuracy evaluations for various LULC datasets. For example, multiple prior studies have compared and evaluated the spatial (in)consistency of global land cover products from various perspectives, e.g., at the same spatial resolution (i.e., 1 km or 30 m) [18-21], for certain type of satellite sensors [22], on a global scale [20,21,23], local regions (like China or one basin) [24-28], or the global and continental scales by elevation and climate partitions [29]. Others have compared and summarized an uncertainty analysis of global land cover datasets using an error budget approach [14]. In addition, a few researchers compared a certain element or land cover theme (such as grassland, wetland, etc.) in some land cover datasets to investigate their comparative advantages [30-32]. These ~~prior~~ evaluation ~~research-work~~ provides important references for choosing land cover products and understanding the spatial characteristics of LULC in global and national scales. However, findings on the global/national evaluations of the product accuracy for all LULC elements (classes) may not be replaceable for local-area or class-specific accuracy. For example, a total number of 159,874 validation samples were selected by the producers for the assessment of GlobeLand30-2010. The product overall accuracy was estimated to be 80.3% [9]. Regional or country-wide accuracy evaluations of GlobeLand30 have also been evaluated by many researchers ~~and reported different assessment results~~. For ~~exampleinstance~~, a number of 1467 sample points were used by Meng et al. [33] to verify the ~~accuracy-performance~~ of GlobeLand30-2010 in Shaanxi Province, which indicates an overall accuracy of 80.0%. The

overall accuracy of GlobeLand30-2010 performed by Ma et al. is 81.5% for the study area in Henan province [34]. ~~The number of~~ Another assessment based on 8400 verification sample pixels ~~in China were used~~ by Wang et al. [35] indicates that the overall accuracy ~~for of the~~ GlobeLand30-2010 data in China is 84.2% ~~[35]~~. Regional accuracy ~~evaluation evaluations~~ of GlobeLand30 were also being carried out in other countries and their evaluation results are ~~also~~ different. Overall accuracies were evaluated by Kussul et al. [36] ~~at~~ about 89.7% for Ukraine, ~~while accuracy~~ assessments in Iran by Jokar Arsanjani et al. [37] and in Thessaly by Manakos et al. [38] indicated ~~the~~ overall accuracies of 77.9% and 86.0%, respectively. A summary of the comparison of the regionally achieved accuracies from different land cover datasets vs the claimed global accuracies typically reported by global land cover datasets can be found in [Supplementary Table S1](#).

Northeast China is an important production base of grain and other agricultural products, and is one of the three main grain-producing regions in China, which provides the highest grain food [production](#) and the highest grain commodity rate to the country. Thus, it plays a critical supporting role to ~~protect~~ [ensure](#) the national food security. Moreover, in Northeast China, the different types of croplands and the diversity of terrains and landscapes have led to an obvious transitional zone of cropland distribution. Under these complex conditions, the remote sensing extraction of cropland is very challenging. At present, cropland information can be obtained from various remote sensing data. However, due to the different purposes, study areas, evaluation methods, and other reasons, results on the accuracy evaluation of global/national land cover datasets are inconsistent and incongruous. Thus, it is ~~essential inevitable~~ to investigate the accuracy of different land cover datasets ~~over Northeast China and to for~~ better depicting ~~the its~~ cropland distribution ~~of Northeast China~~.

The purpose of this study is to ~~examine and contrast the farmland category of thirteen distinct land cover datasets with varying resolutions and satellite data sources over Northeast China, compare and evaluate the differences and similarities of the cropland category of totally thirteen land cover datasets with different resolutions and satellite data sources over Northeast China~~. The evaluation proceeds from ~~the~~ following three aspects: (1) approximately 2,000 “ground-truth” samples for each epoch (the years around 2000, 2010, 2015, and 2020) were manually interpreted across Northeast China to verify the dataset accuracies. (2) The spatial agreement among these datasets was compared and analyzed at different scales, including provincial, longitudinal, and latitudinal dimensions. (3) The possible causes for the uncertainties and discrepancies of the accuracies among these evaluated datasets were discussed eventually. Motivated by these targets, this paper conducts a relatively comprehensive comparison analysis and accuracy evaluation of various datasets collected across Northeast China, which is hoped to further understand the accuracies and abilities of these datasets to represent the cropland area in Northeast China and to provide [an](#) important reference for data users and to improve the cropland mapping and advanced applications in the future.

2. Materials and Methods

2.1. Study area

Northeast China is a major agricultural region (115°52'–135°09' E, 38°72'–53°55' N), including the provinces Heilongjiang, Jilin, and Liaoning, and several prefecture-level cities in the east of Inner Mongolia Autonomous Region (Hulunbuir, Xing'an League, Tongliao, and Chifeng). With the altitude ranging from 0 to 2665 m, Northeast China has various types of landforms, including the mountains of ~~Daxing'an~~ [the Greater Khingan Mountains](#), ~~Xiaoxing'an~~ [the lesser Khingan Mountains](#) and ~~the~~ Changbai [Mountains](#), ~~Liaodong~~ [the eastern Liaoning](#) and ~~Liaoxi Hills~~ [the western Liaoning](#), Horqin Sandy Land, Sanjiang-Songnen-Liaohu Plains, Songhua [River](#) and ~~Non-Rivers~~ [Nenjiang River](#), wetlands

of various sizes, and transition zone of these mega-landforms, which leads to its complicated natural landscape conditions. The Figure 1 shows its geographic location and topographic distributions.

From east to west, the territory is divided into temperate, warm temperate, and cold temperate zones, with monsoon and continental climate types. From east to west, the region is characterized by temperate, warm and cold temperate zones successively, with monsoon and continental climate types. The elevation over the study area increases significantly from the hinterland to the outer, and the terrain changes are mainly shaped by the surrounding mountains and drainage network. All these factors contribute to the complex and diversified landforms of the region. Cropland is concentrated in the Sanjiang-Songnen-Liaohe Plain and mosaic with forest, grassland, wetland and other land cover types. Therefore, Northeast China is a typical representative of the cropland classification analysis evaluation.

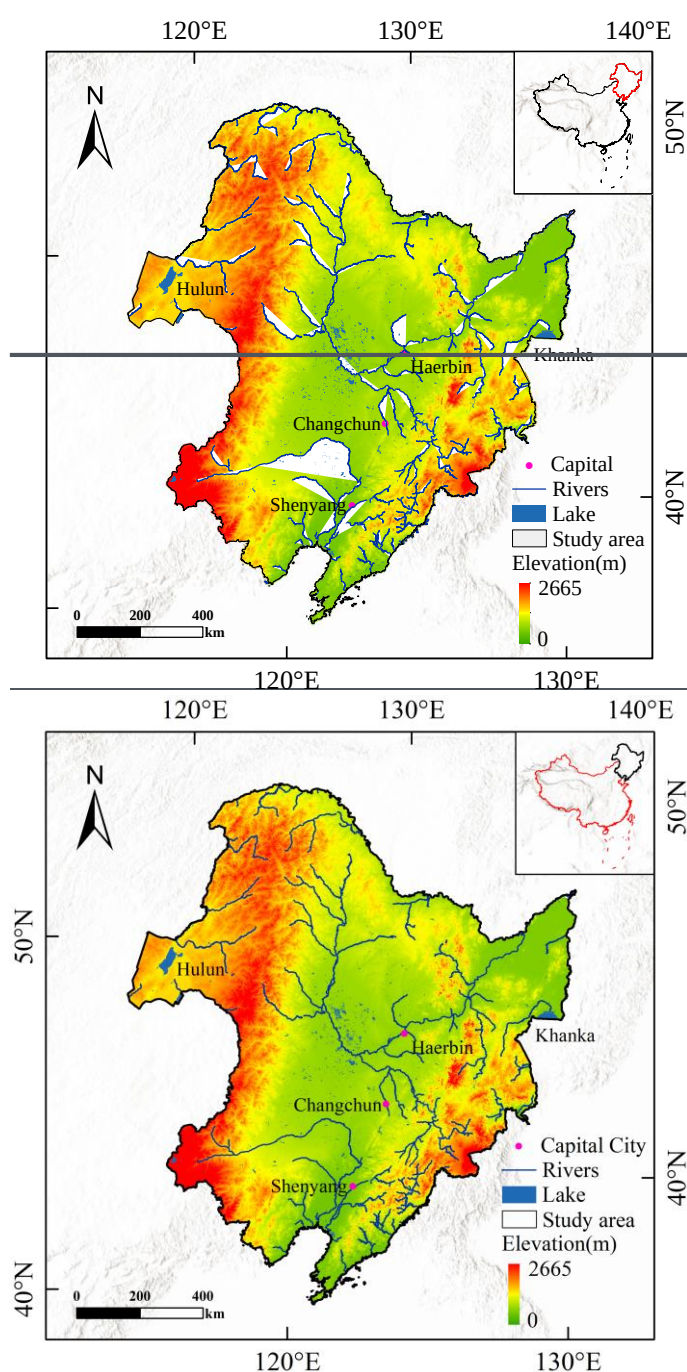


Figure 1. Geographical location and topography map of Northeast China.

2.2. Datasets

2.2.1 Global/regional Land cover datasets

Global/regional land cover datasets are important information sources for understanding the complex interactions between human activities and global change. In this study, we focus on comparing and assessing the cropland classification performance of land cover data products at a global scale. Additionally, in order to confirm whether the classification with more input samples of land cover types at local scales can enhance the data accuracy, we add two another national-scale datasets in China. This study totally collected eleven global-scale datasets, among which the GLC2000, FAO-GLCshare-2014, FROM-GLC-2017 and Esri-2020 products only contain one year data layer, the GLASS-GLC, GLCNMO, CCI-LC, GlobCover, CGLS-LC100, GlobeLand30, and GLC_FCS30 contain multi-year data layers. ~~In addition, we included The two recently released national-scale data products at a national scale, are~~ CLUDs and CLCD. CLUDs has a 5-year interval from 1990 to 2020, while CLCD has annual data layers from 1990 to 2019. The investigated land cover datasets are briefly described below. In addition, we have sorted out and summarized the satellites used in data production, the spatial resolution, and the classification techniques and algorithms used in land cover types. ~~The~~those key information about the thirteen datasets is ~~summarized~~shown in Table 1.

(1) The Global Land Cover 2000 (GLC2000) dataset was developed by the European Commission's Joint Research Center with a spatial resolution of 1 km. The GLC2000 legend is classified into 22 classes using unsupervised clustering and its overall accuracy is 68.6% [39].

(2) FAO-GLCshare (the Global Land Cover-share created by FAO (Food and Agriculture Organization)), produced by the United Nations' (UN) FAO in 2014 [40]. It uses the data fusion method to integrate the available national, regional, and global datasets with a resolution of 1 km. The product encompasses 11 classes and has an accuracy of 80%.

(3) The Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) was produced by using the data of Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) [10]. This data represents the land cover around 2017 with a spatial resolution of 30 m and encompasses ten land-cover classes.

(4) Another global land cover map, which we named Esri in this paper, is a 10 m resolution imagery derived from ESA Sentinel-2. It was generated using a deep learning model that used more than 5 billion Sentinel-2 pixels and sampled from more than 20,000 sites. These sites are found in all major biomes in the world [41,42].

(5) The Global Land Surface Satellite-Global Land Cover (GLASS-GLC) is an annual dynamic record of global land cover product from 1982 to 2015. It was generated on the Google Earth Engine (GEE) platform with the latest version of GLASS (the Global Land Surface Satellite) CDRs (Climate Data Records) [43]. It has a resolution of 5 km and an average overall accuracy is 82.81%.

(6) The National Mapping Organizations (GLCNMO) Global Land Cover was generated by the International Steering Committee for Global Mapping. ~~The version~~Version I (2003) has a spatial resolution of 1 km and ~~the~~version II (2008) and version III (2013) datasets [44,45] have an improved spatial resolution at 500 m [46].

(7) The Land Cover (LC) project of the Climate Change Initiative (CCI) by the European Space Agency (ESA) provides a series of annual datasets with 300m resolution from 1992 to 2015, termed as CCI-LC dataset [47]. The product uses the unsupervised spatio-temporal clustering and machine learning classification method, with a total of 22 land cover classes.

(8) The GlobCover global land cover map is developed by the ESA in collaboration with international network of partners [8,48,49]. This dataset contains 22 classes defined by the UN LCCS.

(9) The Copernicus Global Land Service Land Cover at the 100 m resolution (CGLS-LC100)-collection 3 was released by the Copernicus Global Land Service. It was derived from high quality land cover training sites based on PROBA-V satellite observations and multiple auxiliary datasets [50]. This global LULC map contains 23 classes.

(10) The GlobeLand30 was a ~~30-30-m~~ global land cover product ~~produced-generated~~ by the National Geomatics Center of China using ~~the~~ Pixel-Object-Knowledge operational method [9,51]. It has an overall classification accuracy ~~of~~ more than 80% worldwide [52]. It comprises 10 land cover classes.

(11) The global 30 m land cover dataset with a fine classification system (GLC_FCS30) in 2015 was produced using ~~a~~ time series image of Landsat and high-quality training data of the Global Spatial Temporal Spectra Library (GSPECLib) on the GEE platform. Then the GLC_FCS30-2020 was generated with the prior knowledge of experts and the multi-source auxiliary datasets. Both ~~of~~ them contain 30 land-cover classes [21].

(12) The China's land-use/cover datasets (CLUDs) were provided by the Resource and Environment Science and Data Center. Its resolution is 1 km and documented in detail China's land cover in the 1980s, 1990, 1995, 2000, 2005, 2010, 2015 and 2020. It includes 6 level-1 classes, which are cropland, grassland, forest, built-up area, water and barren, and 25 level-2 classes [53,54].

(13) The annual China Land Cover Dataset (CLCD) was derived from Landsat imagery on the GEE platform by Yang et al. [55], which contains annual land cover data layers at the 30 m spatial resolution in China from 1990 to 2019. Including 9 land cover types: cropland, forests, shrub, grasslands, water, snow/ ice, barren, impervious, and wetlands, its overall accuracy is reported at 79.31 %.

Table 1. The summary of main information of all collected datasets in this study. They are sorted in order of the spatial resolution from lowest to highest.

Datasets	Satellites or sensor	Time	Spatial resolution	Classification technique	Classification scheme
GLASS-GLC	AVHRR GLASS CDR	1982-2015	5 km	Random forest & LandTrendr	7 classes
GLC2000	SPOT VGT	2000	1 km	Generally unsupervised classification	22 classes
FAO-GLCshare	----	2014	1 km	Data fusion	11 classes
CLUDs	Landsat	1980 1990 1995 2000 2005 2010 2015 2020	1 km	Extraction of remote sensing information	6 classes*
GLCNMO	Terra MODIS	2003	1 km	Supervised classification	20 classes
	Terra and Aqua MODIS	2008 2013	500 m	Supervised classification	20 classes
CCI-LC	ENVISAT MERIS SPOT VGT	1992-2015	300 m	Unsupervised spatio-temporal clustering & Machine learning classification	22 classes
GlobCover	MERIS	2005 2009	300 m	Generally unsupervised classification	22 classes
CGLS-LC100	PROBA-V	2015-2019	100 m	Supervised classification & Random forest	23 classes
GlobeLand30	Landsat TM/ETM+ HJ-1	2000 2010 2020	30 m	Pixel-Object-Knowledge classification approach	10 classes
CLCD	Landsat	1990-2019	30 m	Random forest	9 classes
GLC_FCS30	Landsat TM/ETM+/OLI	2015 2020	30 m	Operational SPECLib-based approach & Random forest	30 classes

FROM-GLC	Landsat TM/ETM+/OLI Sentinel-2	2017	30 m	Random forest	10 classes
Esri	Sentinel-2	2020	10 m	Deep learning model	10 classes

Note: 6 classes* means 6 classes in the first level and 25 classes in the second level.

According to the time characteristics of the mapping data sources, we classified all data into four phases: 2000, 2010, 2015 and 2020. The Phase-2000 comprises eight datasets, Phase-2010 also includes eight datasets, Phase-2015 predominantly consists of nine datasets, and Phase-2020 comprises six datasets. Table 2 displays the data contained in these four periods.

Table 2. The datasets included in four ~~periods~~phases.

Phase-2000	Phase-2010	Phase-2015	Phase-2020
GLASS-GLC	GLASS-GLC	GLASS-GLC	
GLC2000		FAO-GLCshare-2014	
CLUDs	CLUDs	CLUDs	CLUDs
GLCNMO-2003	GLCNMO-2008	GLCNMO-2013	
CCI-LC	CCI-LC	CCI-LC	
		CGLS-LC100	CGLS-LC100-2019
GlobeLand30	GlobeLand30		GlobeLand30
CLCD	CLCD	CLCD	CLCD-2019
	GlobCover-2009	GLC_FCS30	GLC_FCS30
GlobCover-2005	GlobCover-2005	FROM-GLC-2017	Esri

2.2.24 Other auxiliary dataset

We also collected the auxiliary data of cropland distribution in Northeast China reconstructed by Yu et al. [56] as statistical data. This dataset was reconstructed by a multi-source data fusion method. By coordinating the data accuracy, time and spatial resolution of diverse data sources, the continuous annual cropland distribution dataset of China was reconstructed ~~by using~~ the spatialized cropland distribution approach. ~~Four statistical datasets were used in this process, including the Land and Resources Statistical Yearbook, the National Land and Resources Bulletin, the Chinese Agriculture Yearbook, and the national crop production from the Chinese Statistical Yearbook, as well as multiple remote sensing products, including global cropland data provided by the Global Food Security Analysis-Support Data, GlobeLand30, the China Land Use and Cover Change, FROM-GLC, and so on.~~ ~~In this process, four statistical datasets including the Land and Resources Statistical Yearbook, the National Land and Resources Bulletin, the Chinese Agriculture Yearbook, and the national crop production from the Chinese Statistical Yearbook, and multiple sources of remote sensing products including the global cropland data provided by the Global Food Security Analysis-Support Data, GlobeLand30, the China Land Use and Cover Change, FROM-GLC and so on were used.~~ Overall, Yu et al. [56] assimilated the trend information and spatial pattern of various satellite datasets to ~~produced~~produce series of cropland percentage maps, which are spatially explicit and ~~spans~~span the time period from 1900 to 2016. Through intensive comparisons, the authors are confident that this data provides reliable sources of cropland maps [56].

2.3. Data processing procedures

The flowchart of this study is shown in Figure 2. ~~The following sections describe the key data processing steps. Three key steps are involved in the accuracy assessment and product comparison. The following describes the key processing steps of the data. Accuracy assessment and product comparison include three key steps.~~ First, a random sampling approach was employed to obtain validation sample data. ~~This process resulted~~

in four-phase validation samples (about 2000 samples for each) in 2000, 2010, 2015, and 2020, respectively. ~~his process generated four phase validation samples (about 2000 samples for each) in 2000, 2010, 2015, and 2020, respectively.~~ These validation sample points were put into the Google Earth and World Imagery Wayback platform, and interpreted by referring to the high-resolution images. Here the images of ~~phrase~~Phase-2000 and ~~Phasephrase~~Phase-2010 are provided by Google Earth Pro, and the images of ~~Phasephrase~~Phase-2015 and ~~Phasephrase~~Phase-2020 are from September 20th, 2015 and 2020 of ~~the~~ World Imagery Wayback platform. Those sample points were marked according to the actual situation of the ground (marked 1 represents cropland, 0 represents non-cropland). Second, the function of “Extract Multi Values to Points” was used in ArcGIS to assign the cropland and non-cropland of each data here to the attributes of the sample point. Third, we constructed the confusion matrices of all datasets in the four years, that is, the number of products and interpretive sample points simultaneously identified cropland/cropland, cropland/non-cropland, non-cropland/cropland and non-cropland/non-cropland were counted (The confusion matrix is shown in supplementary data). The accuracy evaluation indexes were calculated according to the formula, and the comparison chart of the accuracy evaluation indexes of each data was drawn.

About spatial agreement comparison, in order to facilitate the comparison of these different land cover products (different coordinate systems and resolution), we ~~firstly-first~~ processed the global/national land cover datasets for the study area by using the boundary of Northeast China. Secondly, we carried out reprojection, resampling and reclassification processing successively in ArcGIS. ~~The datasets were reprojected using the Albers Equal-Area projection. The reprojection was that datasets transferred into the Albers Equal Area projection.~~ Using the nearest neighbor method, all data were resampled to 30 m. We combined all cropland subclasses of each dataset into one class to integrate cropland information. Here, the mosaic cropland (>50%) with other elements in some data (e.g., GLCNMO, CCI-LC, and GlobCover) are classified into cropland and the classification of cropland subcategory is shown in [Supplementary Table S5](#). After this processing, the datasets were only reclassified into cropland and non-cropland. Finally, the “Raster Calculator” was used to obtain the agreement levels of consistency and discrepancy of cropland in the same spatial position of different land cover datasets, and the area proportion of different agreement levels was calculated.

In the comparison of cropland area in Section 3.4, we used pixel counting to calculate the cropland area of Northeast China under the projection of Albers, which is the pixel number of all cropland areas multiplied by the single pixel area.

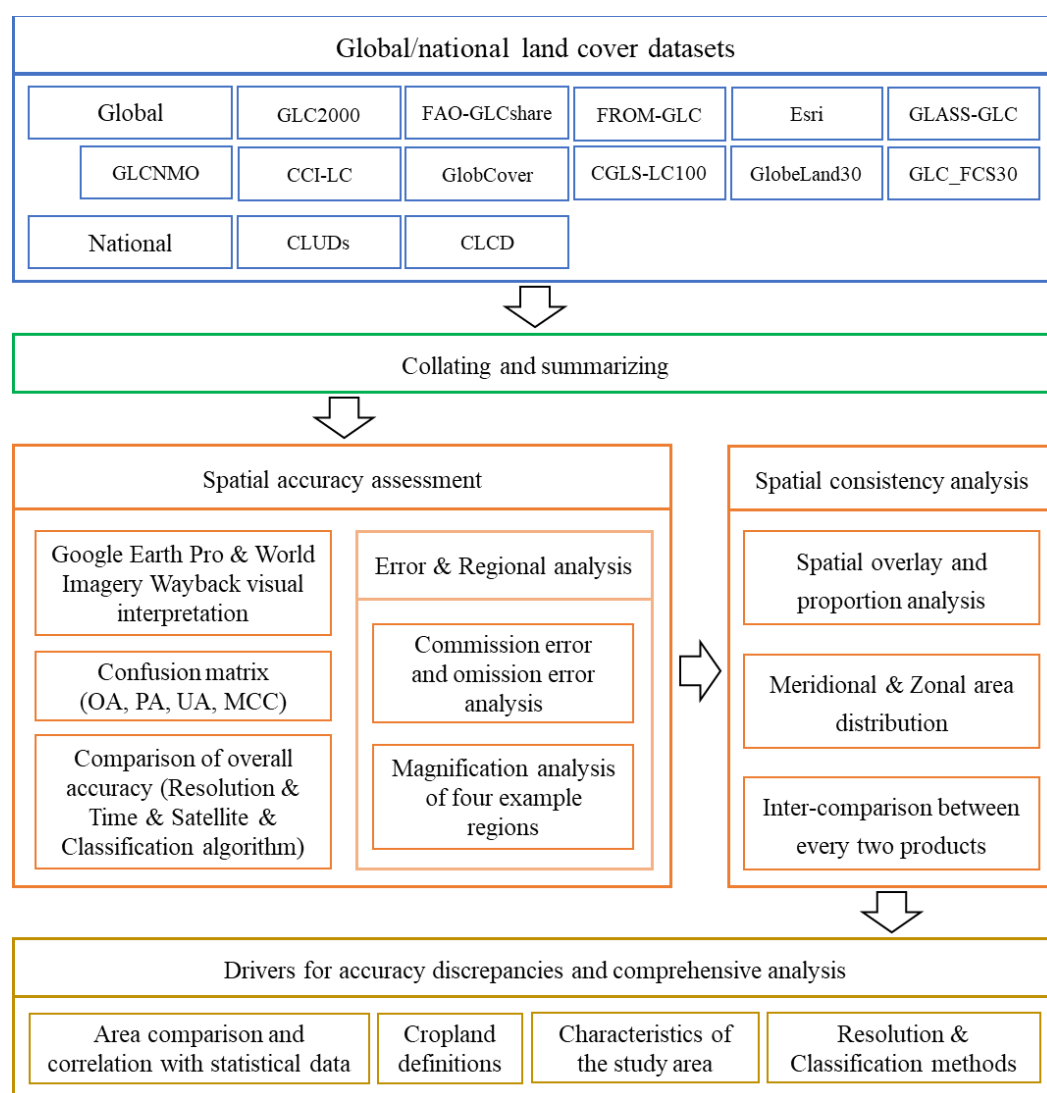


Figure 2. The workflow of this research. OA = overall accuracy, PA = producer's accuracy, UA = User's accuracy, MCC = the Matthews correlation coefficient.

2.4. Methodology

2.4.1 Accuracy assessment metrics

First, the accuracy assessment is performed by the error matrix or confusion, which is a crosstab of various land cover products and the ground truth sample point. It directly reflects the overall accuracy of the mapping [57-59]. Using the confusion matrix, four classical evaluation parameters are calculated, which are the user's accuracy (UA) (commission error), producer's accuracy (PA) (omission error), overall accuracy (OA), and the Matthews correlation coefficient (MCC), respectively. The UA, PA, and OA respectively refer to the percentages of the products classified correctly, the validation points classified correctly, and the correctly classified number. The Matthews correlation coefficient ~~generate~~ ~~generates~~ reliable results independently the ratios of positive and negative elements on the binary classification predictions [60].

The four evaluation metrics are the most commonly used methods in the accuracy assessment of land cover datasets [11,61-63]. They are defined as

$$\text{User's accuracy} = \frac{X_{ii}}{X_{i+}} \times 100\% \quad (1)$$

$$\text{Producer's accuracy} = \frac{X_{ii}}{X_{+i}} \times 100\% \quad (2)$$

$$\text{Overall accuracy} = \frac{\sum_i^2 X_{ii}}{N^2} \times 100\% \quad (3)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

where X_{ii} is the amount of class i pixels that were correctly classified, in this study it refers to the amount of cropland and non-cropland are correctly classified. N refers to the total number of the examined pixels, X_{+i} represents the number of class i pixels of the classification result, which is the total number of pixels of cropland in various land cover datasets; X_{ii} on behalf of the class i pixel number for the reference data, it is the total pixel number of cropland in the validation sample. The indices TP/TN (true positive/negative) respectively refers to the reference positive/negative is classified correctly as positive/negative; the FP/FN (false positive/negative) means that the reference negative/positive is classified wrongly as positive/negative. TP (true positive) indicates the reference positive is correctly classified as positive; TN (true negative) indicates the reference negative is correctly classified as negative; FP (false positive) represents the reference negative is wrongly classified as positive; FN (false negative) represents the reference positive is wrongly classified as negative.

2.4.2 Inter-comparison method

The comparative assessments of various land cover datasets include the location and the quantity of each class [64]. In this study, 13 sets of LULC datasets were compared and analyzed in Northeast China from the two aspects of spatial location and cropland quantity. The comparison of location is usually a matter of spatial (dis)agreement pixel by pixel. The purpose of spatial agreement analysis is to explore the characteristics of the spatial distribution of cropland in various datasets. Spatial agreements include visual comparison and pixel-by-pixel comparison. Visual comparison can show the spatial variation of the datasets among regions subjectively. Per-pixel comparison is to calculate the proportion of pixels with the same attributes in the total pixels of various datasets in the same spatial location. The comparison of quantity mainly refers to areal comparison of cropland. We calculate the cropland area of Northeast China before reprojection and resampling.

2.4.3 Pairwise data and prefecture-level city cropland area validation

Correlation analysis can verify the relationship of the pairwise data and the ability of represent the prefecture-level cropland area. The following formulas (5), (6) and (7) are used for correlation analysis to detect the goodness of fit between datasets:

$$\Delta x_i = x_i - y_i \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \Delta x_i^2}{n}} \quad (6)$$

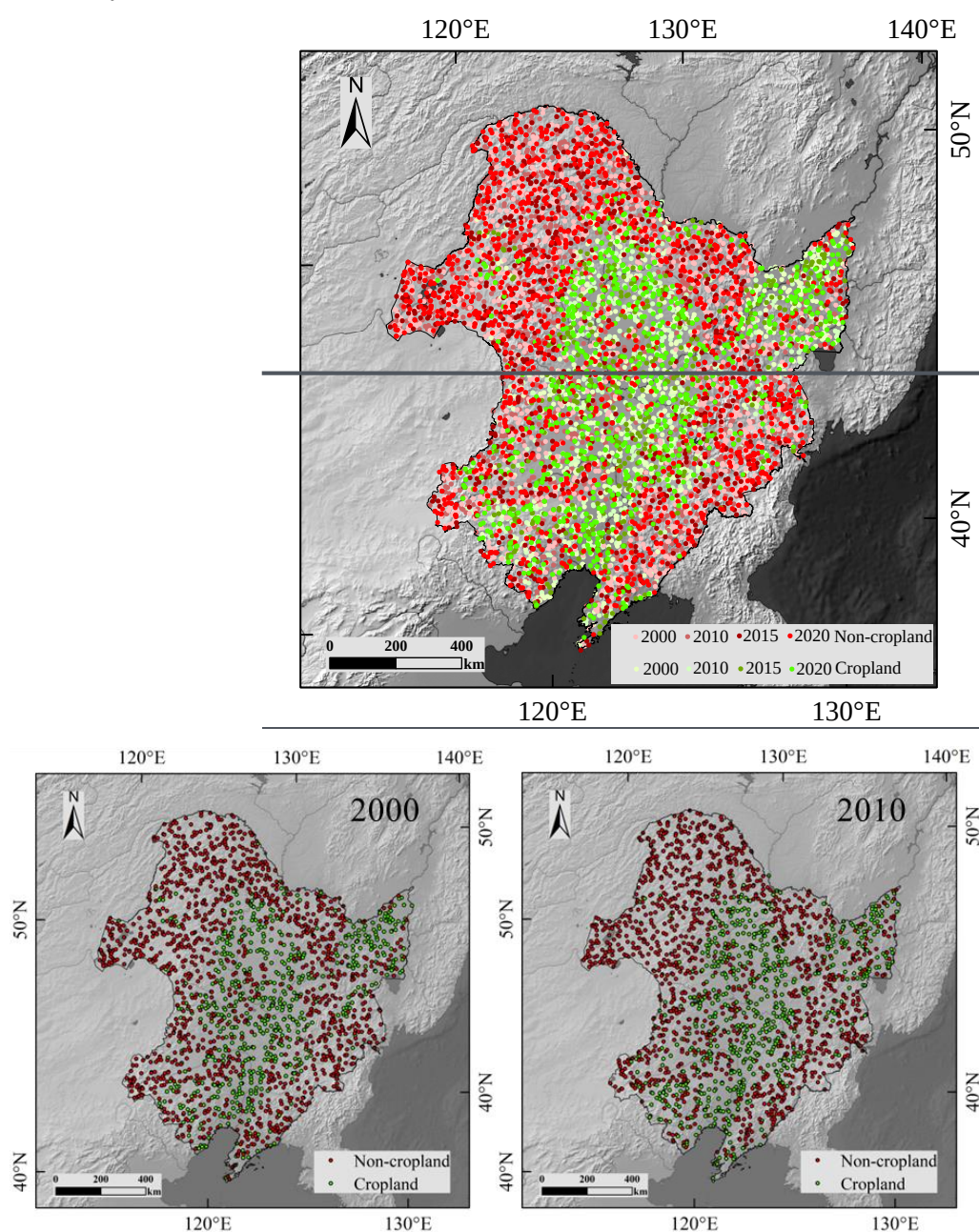
$$r^2 = \frac{(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

where Δx_i is the difference between the cropland area (x_i) of a data and another data (y_i), n is the total number of regions of aggregation. $\bar{x}$ and $\bar{y}$ are the average of cropland area.

3. Results

3.1. Accuracy evaluation of the thirteen datasets in the four phases

Figure 3 shows the distribution of sample points across Northeast China for four years. The distribution of ground-truth points in four periods is roughly similar in Figure 3, and the majority of cropland is distributed in the Sanjiang-Songnen-Liaohe Plain, with the main types of non-cropland being mountains, hills, lakes and rivers, grasslands, sandy and bare land, such as the Greater Khingan Mountains, the Changbai Mountains, and so on. Scattered cropland distribution is found at the transition areas of plains and hills, beside lakes and rivers, or in areas such as grassland or bare, the difference between cropland and non-cropland in the four phases were mainly concentrated in these areas.



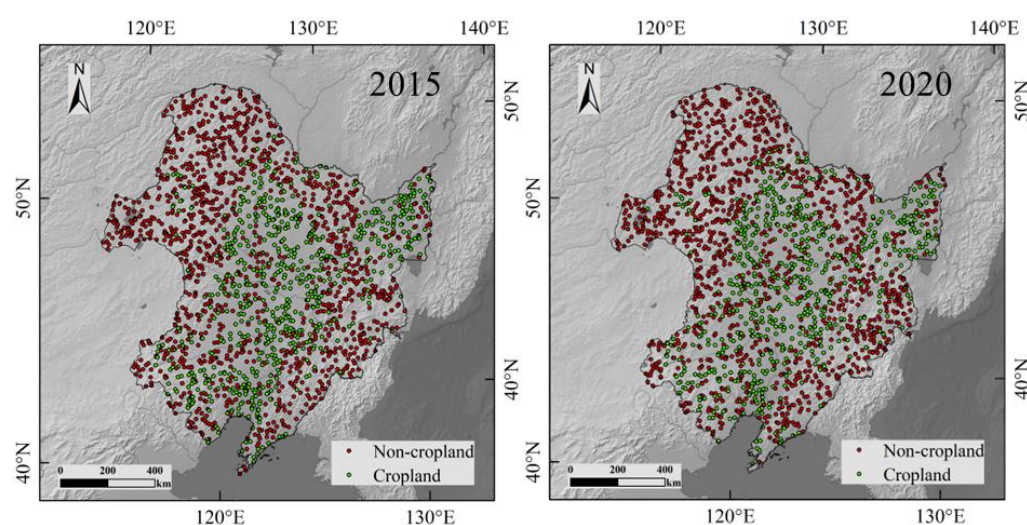


Figure 3. Distribution of validation samples across Northeast in Phases 2000, 2010, 2015 and 2020. The graduated red points represent true non-cropland, graduated green indicate real cropland.

Based on the ground truth validation samples, we constructed the confusion matrices of all datasets in four phases and obtained the spatial accuracy measures (i.e., producer's accuracy, user's accuracy, overall accuracy, and the Matthews correlation coefficient-), they were computed from confusion matrices (Figure 4). In Phase-2020, the average overall accuracy of six datasets is the highest at 0.86. On the whole, the average of overall accuracy of six datasets in Phase 2020 is the highest of 0.86, followed by Phase-2000, Phase-2010, and Phase-2015, all around 0.83 (Figure 5 ¹³). CLUDs and CLCD data are available in all the four phases, while GLASS and CCI-LC data are available in three phases, 2000, 2010 and 2015. The following is a comparison of the data evaluation metrics by phase. The following is a comparative analysis of the evaluation metrics of the data by phase.

The overall accuracies of the eight datasets in Phase-2000 all exceed 0.73. GlobeLand30 has the highest OA of 0.90, followed by CLCD (0.89), CCI-LC (0.87). From the perspective of PA, ClobCover-2005 ranks the lowest with the accuracy of 0.57, followed by GLC2000 (0.66). As for UA, ClobCover-2005 also has the lowest value of 0.62, followed by GLCNMO-2003 (0.63). Overall, GlobeLand30 has the balanced performance, with higher OA, UA and PA in 2000, followed by CLCD, and their MCC are all over 0.75. In Phase-2010, GlobCover-2005 has the lowest UA, PA and OA, followed by GlobCover-2009. CLCD has the highest UA, PA and OA, followed by ClobLand30. Other datasets have medium accuracy. The overall accuracies in the nine datasets of Phase-2015 are over 0.76, CLCD has the highest OA value of 0.90, followed by CCI-LC (0.87) and GLC_FCS30 (0.85). From the perspective of PA, GLASS ranks the lowest with 0.68. As for UA, GLCNMO-2013 has the lowest value of 0.63, followed by CGLS-LC100 (0.64) and FAO-GLCshare (0.68). In general, CLCD has the higher OA, UA and PA in Phase-2015, its MCC is over 0.77. Following it, CCI-LC and GLC_FCS30 also have good performances with MCC greater than 0.70. The overall accuracies in the six datasets of Phase-2020 are more than 0.70, the OA of CLCD is 0.914, followed by GlobeLand30 (0.906), GLC_FCS30 (0.902), and Esri (0.896). In terms of PA, CLUDs ranks the lowest with 0.69. This corresponds to a higher omission error, indicating that the estimation of cropland area across Northeast China is low (Figure 5 and Figure 6). As for UA, CGLS-LC100 has the lowest value of 0.57, and it is clear to observe that the cropland area in CGLS-LC100 data is overestimated than other data (Figure 5, Figure 6, Figure 8, and Figure 9). CLCD has the best performance in Phase-2020, followed by GlobeLand30, and all of their MCC are greater than 0.79. In a word, CLCD has the best performance in Phase 2020, followed by GlobeLand30, and all their MCC are over 0.79 (Figure 4 and, Figure 5 and Figure 6). Following them are GLC_FCS30 and Esri, both of which have MCCs greater than 0.77. Following them,

GLC_FCS30 and Esri also have good performances with MCC of over 0.77. For detailed numerical information, please see [Supplementary Table S2](#).

In summary, CLCD has the highest accuracy in four phases, followed by GlobeLand30, GLC_FCS30, and Esri. The low PAs values of GlobCover, GLC2000, CLUDs, GLASS and GLCNMO reflect a low estimation of the area in different regions of Northeast China. The UAs of CGLS-LC100, GlobCover, GLCNMO and FAO-GLCshare are all relatively low, all of which have different degrees of commission errors of cropland in different regions of the Northeast, and the cropland area is relatively overestimated than other datasets.

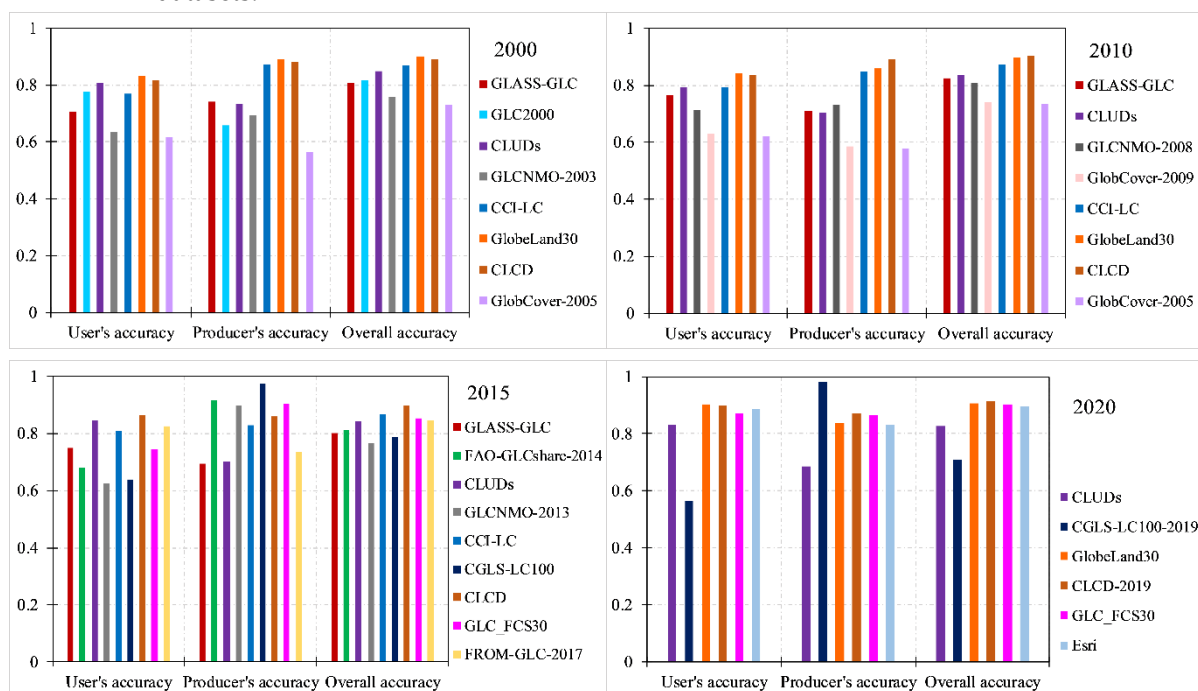
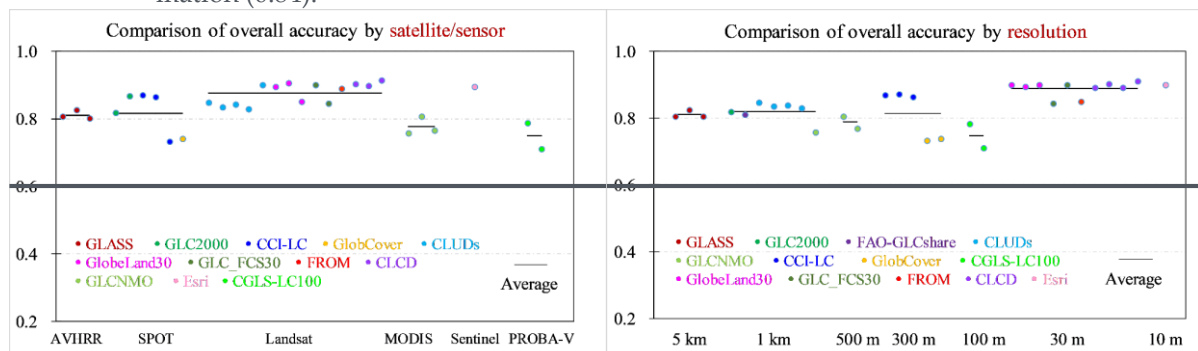


Figure 4. Comparisons of the spatial accuracy indexes of the datasets in 2000, 2010, 2015 and 2020.

We further grouped and compared the overall accuracy of these datasets by resolution, time, satellite sensors, and classification algorithms, and the results are illustrated in Fig. 5. It can be found the overall accuracy rises with the improvement of resolution and the development of mapping time in general. The average of the overall accuracy of the products derived from Sentinel and Landsat observations are overall better than the data produced by other satellite sensors, while the data produced by the PROBA satellite is the worst. The average overall accuracy of the data produced by random forest, Pixel Object Knowledge classification approaches and deep learning are all over 0.89, followed by the operational SPECLib based approach & random forest (0.88), the unsupervised spatio-temporal clustering & machine learning (0.87), and extraction of remote sensing information (0.84).



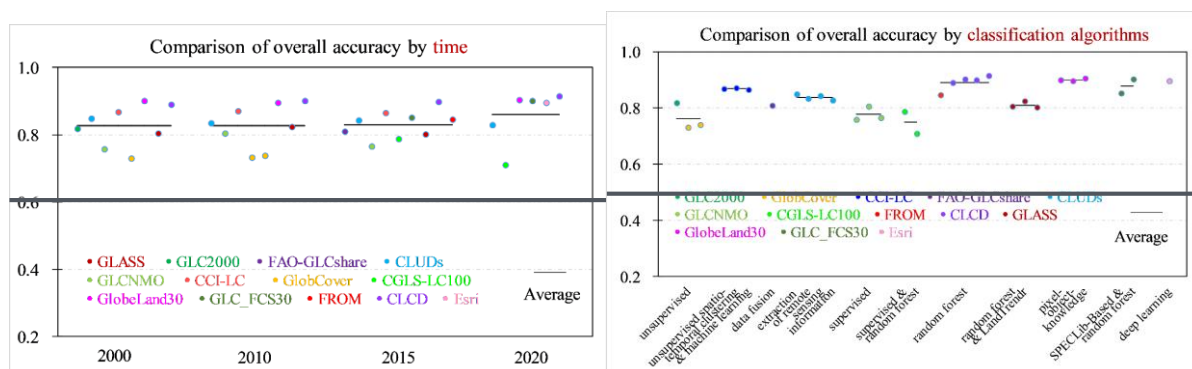


Figure 5. Scatter plots of overall accuracy compared by resolution, time, sensor, and classification algorithm, with different individual colors markers for different land cover datasets.

3.2. Commission and omission error analysis on cropland in Northeast China

The Phase-2015 and Phase-2020 contain higher precisions and more products, covering almost all data products except for GLC2000 and GlobCover, thus we choose the two phases to further analyze the omission and commission error of these datasets. Based on ground verification points, the omission and commission error distribution of datasets in 2015 and 2020 across Northeast China is illustrates in [Figure 5](#) and [Figure 6](#). The ginger-pink points indicate the higher commission errors and the spruce-green points represent the higher omission issues.–

Lots of ginger-pink points exist in CGLS-LC100 in [Figure 5](#) and [Figure 6](#), indicating serious commission ~~issue-issues~~ and illustrating the overestimation of cropland area in both meridional and zonal dimensions ([Figure 8](#) and [Figure 9](#)). Followed by GLCNMO and FAO-GLCshare, they are suffered from commission issues in different regions. GLCNMO-2013 has serious commission errors in ~~the~~ western Inner Mongolia, especially in the western part of Hulunbuir City, where the overestimation is more serious, followed by Heihe City in Heilongjiang Province. Chifeng City in Inner Mongolia has a lower omission phenomenon. FAO-GLCshare has high misclassification in the north of Heihe City in Heilongjiang, the central part of Hulunbuir City, as well as the junction of mountains and plains (~~Da Xing'an~~ [the Greater Khingan Mountains](#), ~~the~~ Changbai Mountains, etc.), while it has low leakage in Greater Khingan Mountains. GLASS has many spruce-green points and it has less cropland area than other datasets ([Figure 8 aA](#), [Figure 5](#), [Figure 6](#)). CLUDs have omission issues in the Sanjiang-Songnen-Liaohe Plain and Chifeng of Inner Mongolia with many spruce-green points. It is clear that CGLS-LC100 and GLCNMO-2013 have lower omission errors with fewer spruce-green points, but they suffer from a little more commission issue, CLUDs is higher omission errors with more spruce-green points ([Figure 5](#) and [Figure 6](#)). Furthermore, CLCD has the fewest spruce-green and ginger-pink points, indicating that it has few omission and commission errors, which corresponds to the highest accuracy evaluation in [Figure 4](#). In addition, CLCD has the fewest spruce-green and ginger-pink points, which means that it suffers from a little omission and commission errors, which also corresponds to the highest accuracy evaluation in [Figure 4](#). The numbers of the commission and omission error in each product in phrase 2015 and phrase 2020 is shown in [Supplementary Table S3](#).

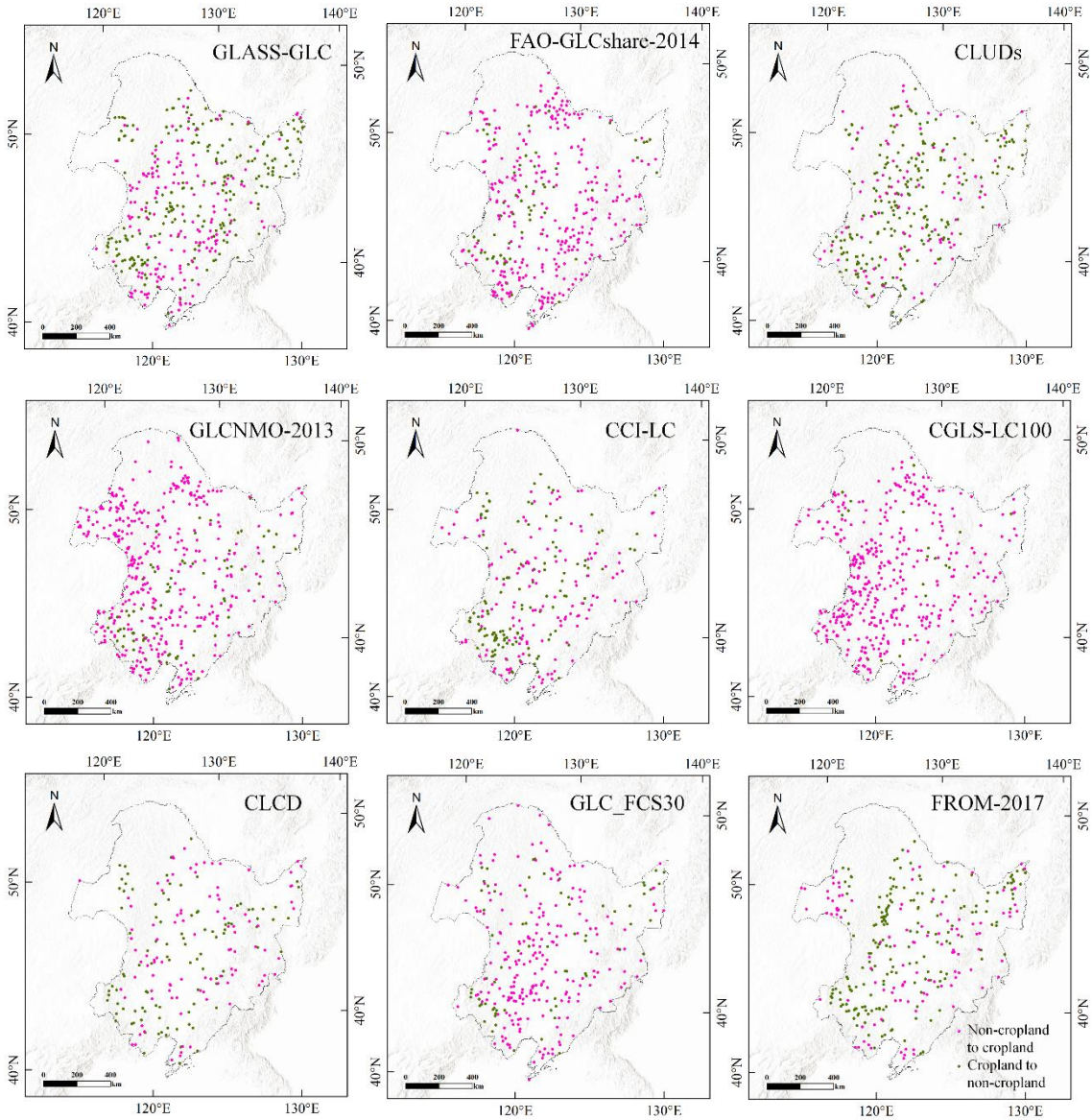
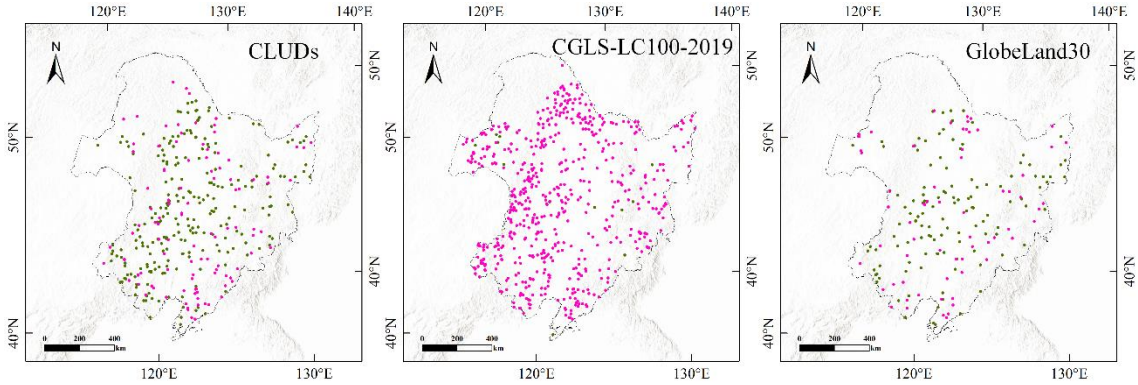


Figure 5(4) The commission and omission error distribution of nine data in 2015.



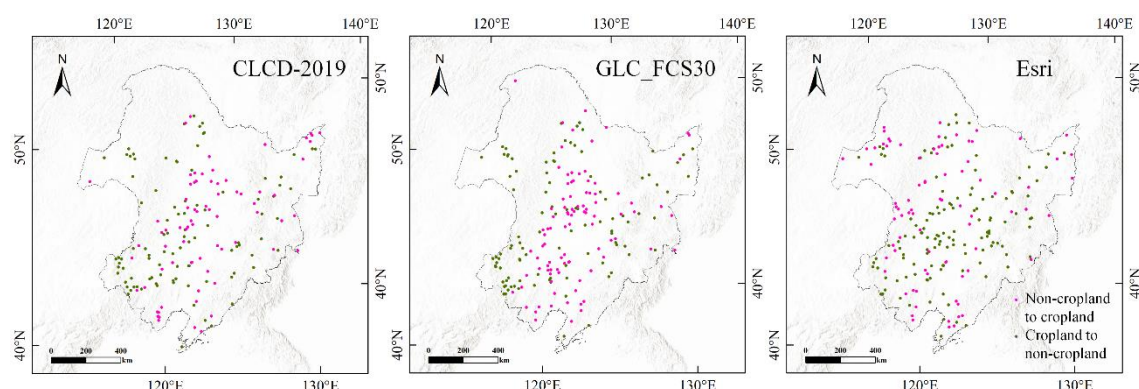
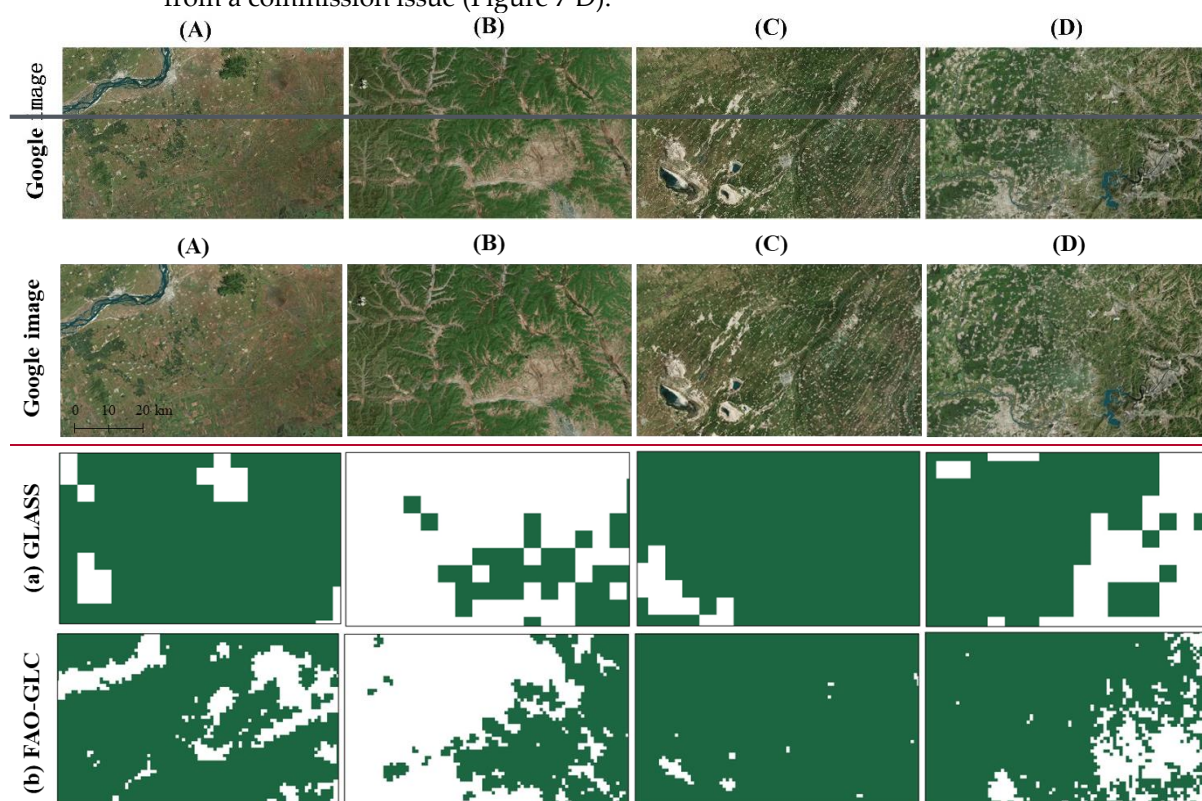


Figure 6. The commission and omission error distribution of six data in 2020.

Figure 6. The disagreements distribution of verified points and the corresponding position in each dataset. In Figure 5 and Figure 6, The the ginger-pink dots represent that true pixels of non-cropland are classified as cropland (representing commission error). The spruce-green dots indicate that true pixels of cropland are classified as non-cropland (representing omission error).

To better understand the spatial accuracies of these datasets in more detail, we selected four regions for Phase-2015 (as it contains the most data products) for zoom-in exemplification to better observe the varied accuracies, as shown in Figure 7. Figure 7 shows their positions. It can be seen that except GLASS, the other eight datasets could capture the distribution of cropland in the Sanjiang and Songnen Plain. GLASS not only has low resolution, but also suffers from omission and commission issue. In the southeast of Hulunbuir city of Inner Mongolia (Figure 7 B), FROM-GLC and GLASS have a large number of missing, and fail to detect the cropland, followed by FAO-GLCshare and CLUDs. CLCD and GLC_FCS30 have a good description of mountains and farmlands, and CGLS-LC100 has a significantly overestimation of cropland area. Five datasets perform well in the Liaohe Plain except FAO, GLASS, GLCNMO, and CLUDs which suffer from a commission issue (Figure 7 D).



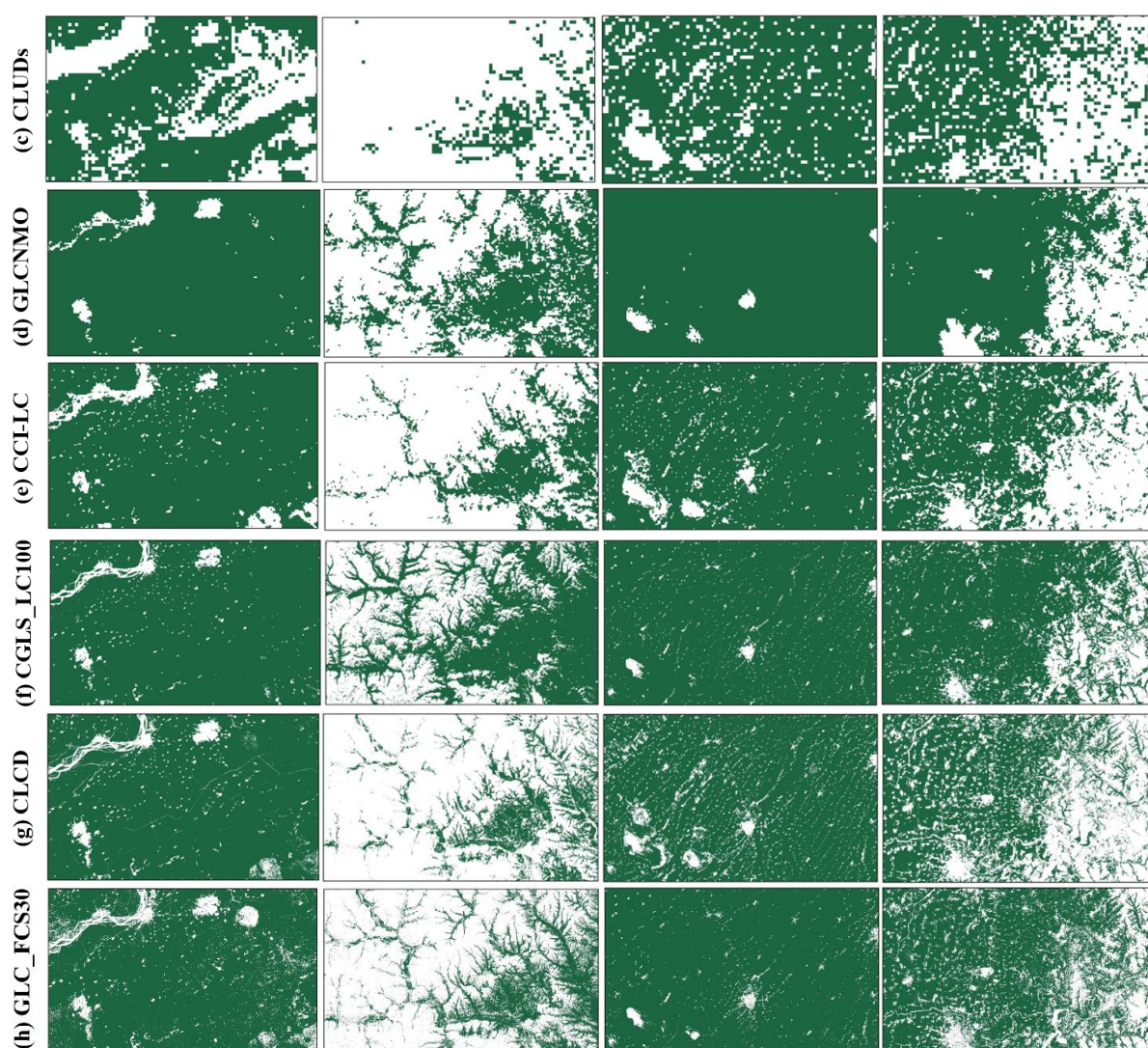


Figure 7. Demonstration of the cropland magnified performances of four regions in different datasets. Dark-green indicates cropland and white indicates non-cropland. The positions of A-D are shown in Fig. 8 A-b.

3.3. Spatial agreement and discrepancies

The spatial agreement is shown in Figure 8 and Figure 9, which illustrates spatial similarities and differences of cropland. In terms of longitudinal cropland distribution curve, most croplands of all datasets are clustered between 125°E and 130°E (Figure 8). In summary, the region with agreement level 9 or 6 (highest spatial consistency) is mainly distributed in traditional agricultural regions of Northeast China, such as the Sanjiang-Songnen-Liaohe Plain, which include Harbin, Hegang, Jixi, Jiamusi, Qiqihar, and Suihua in Heilongjiang. Liaoyuan City, Siping City, Song yuan City, Changchun City in Liaoning Province, and Fuxin City, Jinzhou City, Shenyang City, and Tieling City in Liaoning Province. All of these zones have homogeneous and clustered cropland and can be easily captured by remote sensing imagery. The regions of lower spatial consistency (agreement level ≤ 2) are distributed in mountainous, hilly zones and transition areas between them in the east of Inner Mongolia, Jilin, and Liaoning, and the north of Heilongjiang.

The overlapping result of nine datasets in Phase-2015 is shown in Figure 8 (a-c), the area identified as cropland by the nine datasets account-accounts for 19.7% of the total cropland area. From the meridional area curve, we can see that CGLS-LC100 has the highest area curve between 121°E and 128°E, FAO-GLCshare has the highest area curve between 128°E and 135°E compared with other datasets. FROM-2017 at around 121-125°E,

CLUDs at about 126–129°E, and GLASS at 130–132°E are the lowest. From the perspective of zonal area distribution, FAO-GLCshare, CGLS-LC100 and GLCNMO-2013 have the highest area curves in the range of about from 40°N to 42°N, from 42°N to 48°N, and from 48°N to 51°N.

Figure 8–9 (a–c) shows the overlapping result of six datasets in Phase-2020. The stacked bar shows that 28.5% are simultaneously labeled as cropland in the total identified cropland regions, and 42.2% of the data are identified as cropland in agreement level five or higher. According to the distribution of meridional cropland area, the area curve of CGLS-LC100 is the highest in all longitudes, this indicates that CGLS-LC100 has a high estimation value of cropland area. CLUDs ranks lowest between 123°E and 129°E, revealing low area estimation within this longitude range. In the zonal distribution curve of cropland area, CGLS-LC100 also showed the highest regional distribution among the six curves. Esri has the lowest area around 45°N, corresponding to the omission phenomenon in Songyuan and Changchun city of Jilin Province. [Supplementary Table S4](#) of supplementary data is the area proportion of different consistency levels.

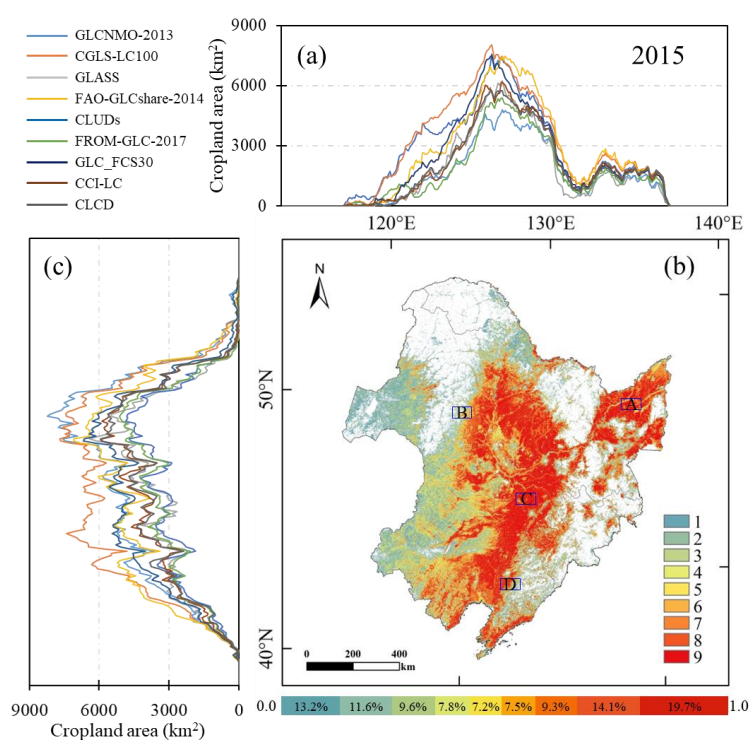


Figure 8. Illustration of the spatial agreement level, cropland area curves of meridional and zonal of nine datasets in 2015. Charts (a) and (c) are the area curves of the meridional and zonal of datasets, respectively. The overlapping result of datasets is shown in subplot (b), representing the resampled evaluated data at a resolution of 30 m, and the digital numbers indicate different consistency levels. The stacked bar chart is the cropland area proportion (%) in the agreement levels.

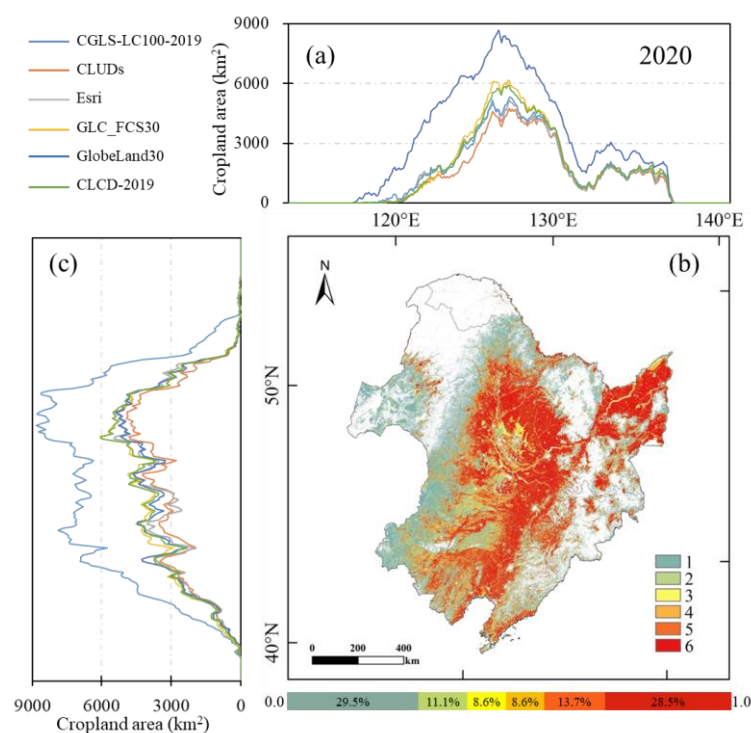
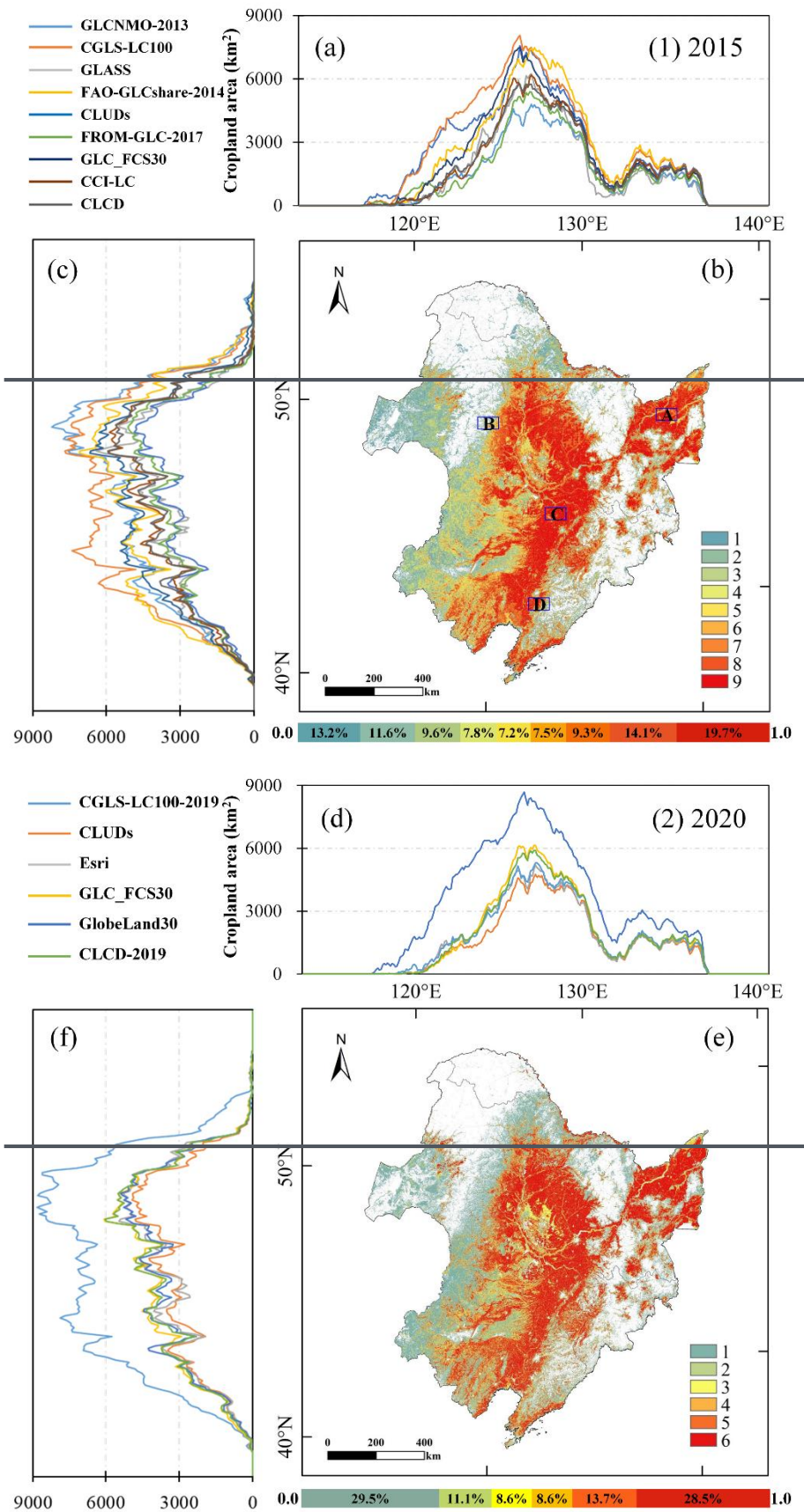


Figure 89. Illustration of the spatial agreement level, cropland area curves of meridional and zonal of six datasets in 2020. In Figure 8 and Figure 9, Charts (a), (d) and (c), (f) are the area curves of the meridional and zonal of datasets, respectively. The overlapping result of datasets is shown in subplot (b) and (e), subfigure (b) shows nine data in 2015—representing the resampled evaluated data at a resolution of 30 m to 30 m, and subfigure (e) shows six data in 2020 resampled to 30 m, respectively, and the digital numbers indicate different consistency levels. The stacked bar chart is the cropland area proportion (%) in the agreement levels.



Because the overall accuracy of six data in Phase-2020 is higher, we further explore their agreement relationships with each other. We compared the pairwise data of six data

and their scatterplots is shown in Figure 9. CLCD has the highest agreement with GLC_FCS30 ($r^2 = 0.96$), followed by GlobeLand30 with CLUDs ($r^2 = 0.89$), GlobeLand30 and CLCD ($r^2 = 0.88$), all their RMSEs are less than 12 km². The CLUDs has the lowest agreement with CGLS_LC100, with the lowest r^2 of 0.38 and the highest RMSE of 36.15 km². CGLS-LC100 is in poor agreement with any data, and we can see that almost all points are above or below the 1:1 line. This is because the cropland area in CGLS-LC100 data is relatively overestimated than other data.

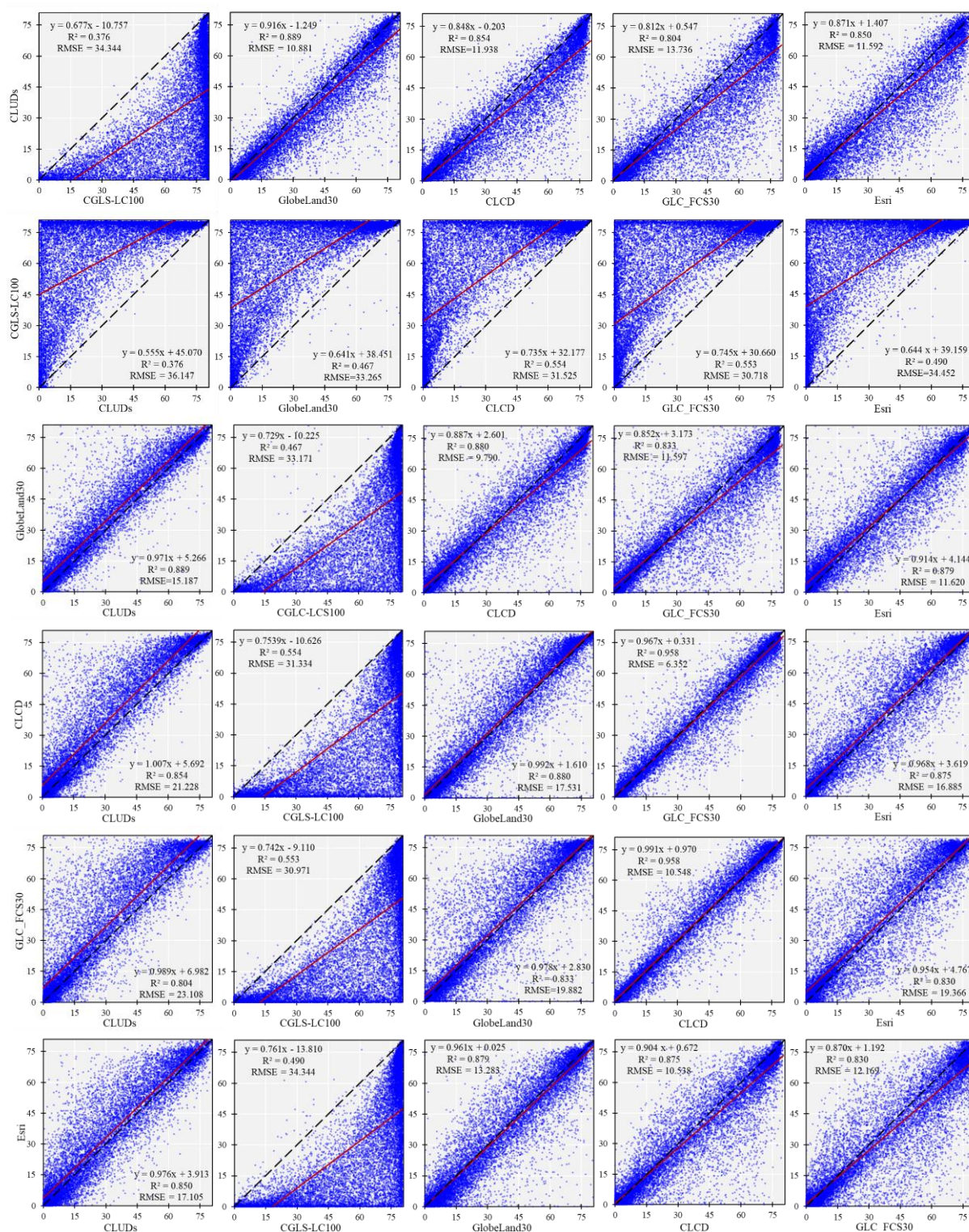


Figure 910. Scatterplots between the CLCD and the CGLS-LC100, CLUDs, Esri, GLC_FCS30, GlobeLand30. The axes represent the cropland area aggregation within the grid cell of 8.438 km×9.537 km across Northeast China of six datasets, which only includes the comparison of cropland area. The unit of RMSE is km².

3.4. Comparative analysis by referring to the statistical data

Figure 1011 compares the cropland area from collected datasets to statistical data in Northeast China. CLUDs-1980s slightly underestimates cropland area, whereas CGLS-LC100 and GLCNMO-2013 significantly overestimate cropland area in Northeast China. It can be seen that CLUDs-1980s slightly underestimates the cropland area, while CGLS-LC100 and GLCNMO-2013 seriously overestimate the cropland area in Northeast China. Compared with the reconstructed cropland data, the cropland area value of CLUDs, GlobCover, and GlobeLand30 products are relatively closer to the statistics. The area of GLASS, FAO-GLCshare, CCI-LC, CLCD, GLC_FCS30, FROM-GLC and Esri are concentrated around 600×10^3 km². The three-year area-areas of GLCNMO-2003, 2008 and 2013 are significantly different, and all products in the statistics of cropland area increase with time. On the whole, the cropland areas based on remote sensing are higher than the statistical data. The differences may be caused by inconsistent spatial resolutions of remote sensing products and official statistics. The remote sensing mapping probably tends to induce biases in the mixed pixels while the official yearbook statistical data was generated by field surveying, multi-level estimates, and census.

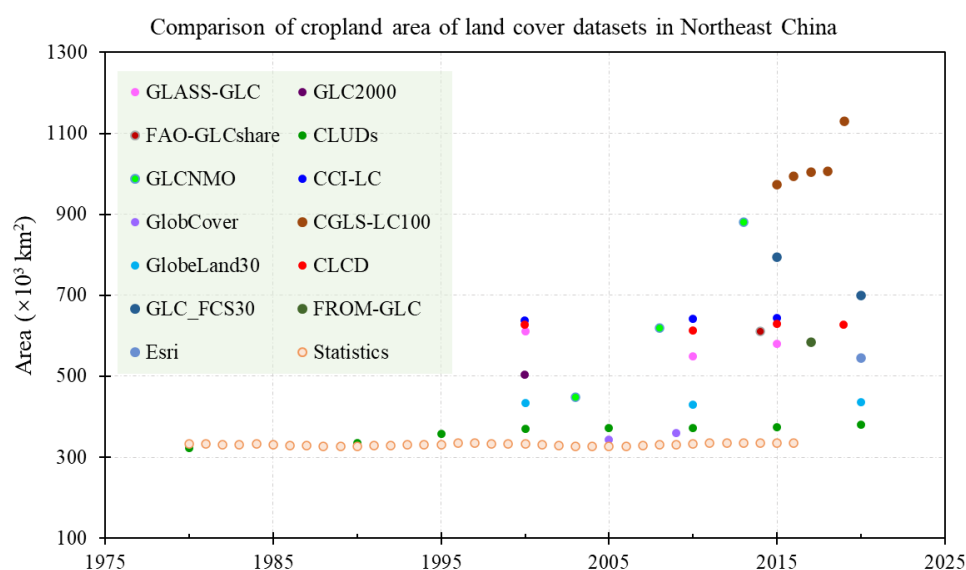


Figure 1011. A scatter plot comparing Comparison of the cropland area of all examined datasets with the statistic result by Yu et al. [56] in Northeast China.

Based on the Phase-2015 datasets, we explored the correlation between datasets and the reconstructed data on the cropland area of aggregated in prefecture-level (Figure 1112). We find that the r^2 of CLUDs is the highest (0.97) and the RMSE is the lowest (1.63 km²), followed by CLCD (0.95, 4.04 km²), and CCI-LC (0.96, 4.10 km²). CGLS-LC100 has the lowest r^2 of 0.53 and the highest RMSE of 15.96 km², besides, the fitting effect of FROM-GLC-2017 is also better. Compared with other datasets, CLUDs performs best in regional accuracy (prefecture-level). FROM-GLC-2017 and GLASS also have good manifestation with most the “plus” points centralized around the 1:1 line. On the contrary, CGLS-LC100 and GLCNMO perform poorly, and they suffer from obvious overestimation on prefecture-level cropland area.

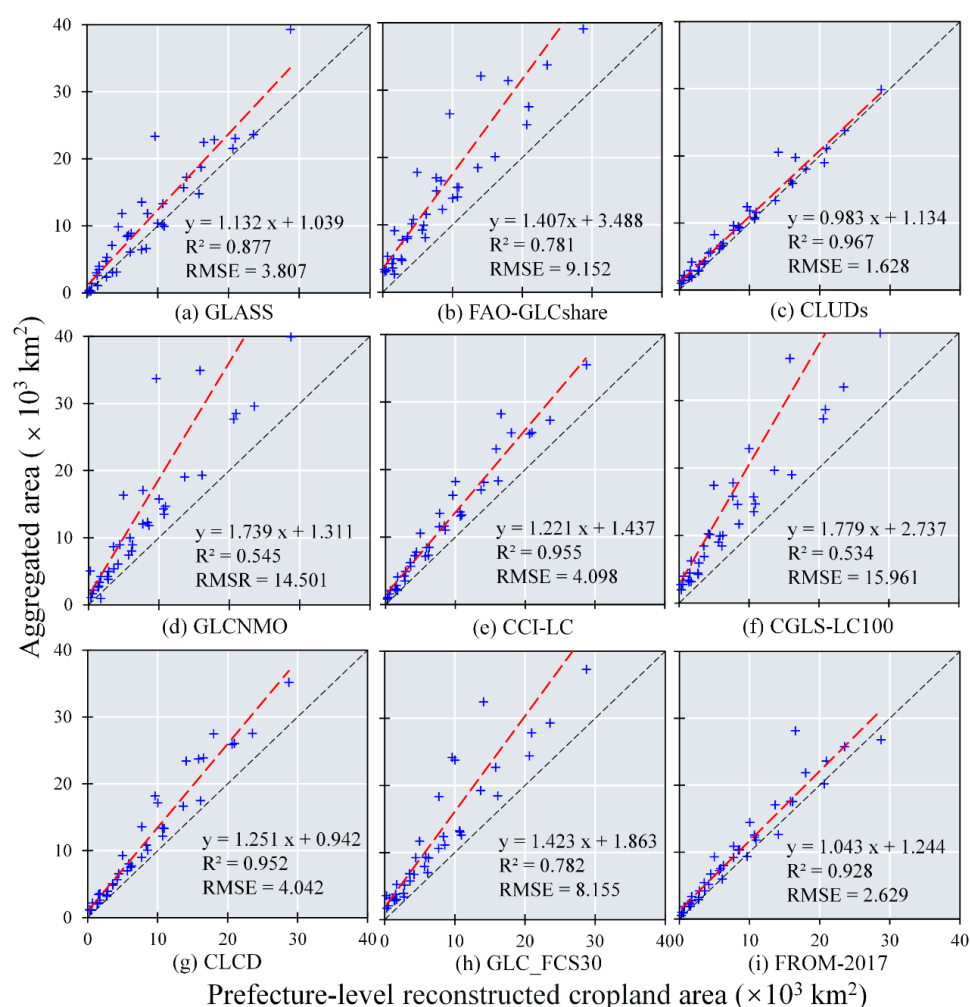


Figure 12. Scatterplots of the prefecture-level city reconstructed cropland area vs. the aggregated area of cropland from each dataset. The unit of RMSE is km^2 .

4. Discussion

4.1. Potential influencing factors of the data uncertainties accuracy

Multiple factors may have resulted in the difference and uncertainties in the cropland range among these datasets. We grouped and compared the overall accuracy of these datasets mainly from four dimensions, including by the data resolution, producing time, satellite sensors, and classification algorithms. and The results are illustrated in Figure 13. It can be found the overall accuracy rises with the improvement of resolution and the development of mapping time in general. For example, GLASS with 5 km resolution has the lowest overall accuracy, and Esri with 10 m resolution has the highest overall accuracy. The average of the overall accuracy of the products derived from Sentinel and Landsat observations are overall better than the data produced by other satellite sensors, while the data produced by the PROBA satellite is the worst. In addition, different classification strategies and mapping methods adopted by these datasets will may lead to difference in classification accuracies. The average overall accuracy of the data produced by random forest classification method, Pixel-Object-Knowledge classification approaches and deep learning are all over 0.89, followed by the operational SPECLib-based approach & random forest models (0.88), the unsupervised spatio-temporal clustering & machine learning (0.87), and extraction of remote sensing information (0.84).

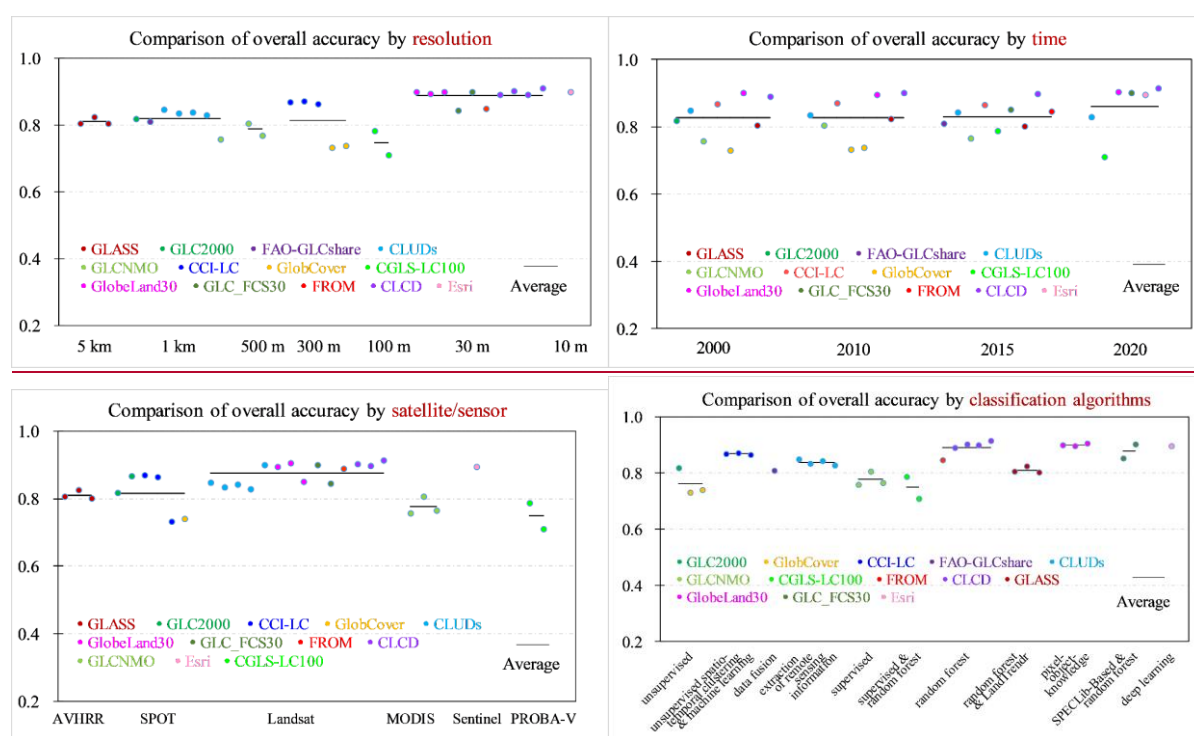


Figure 13. Scatter plots Comparison of the overall accuracy of compared different datasets grouping by data resolution, producing time, sensor, and classification algorithm, with different individual colors markers for different land cover datasets.

4.1. Potential influencing factors of the data uncertainties

Multiple factors may have resulted in the difference and uncertainties in the cropland range among these datasets. Firstly, the definitions and classification of cropland are diverse in these evaluated datasets. For example, in GLC2000, CGLS-LC100 and Esri, the croplands are “pure” cropland, including rainfed croplands and flooded crops, such as cereal, corn, wheat, rice. The croplands in GLASS, GlobeLand30, CLCD, GLC_FCS30 and FROM-GLC mainly refer to rainfed/irrigated cropland, orchard and temporally bare farmland, etc. However, in GLCNMO, the cropland includes herbaceous crop(s), the paddy field of graminoid crops and non-graminoid crops, and the mosaic of cropland and other vegetation. In CCI-LC, the cropland incorporates rainfed/irrigated cropland, herbaceous, tree or shrub cover, and the mosaic of cropland and natural vegetation (Table S5). In general, the mixture of herbaceous or shrubs and cropland and the narrow definition result in different overestimates and a little underestimation of cropland area respectively. The cropland definition of GLC2000, CGLS-LC100 and Esri are “pure” cropland, and their estimation of cropland area should be somewhat underestimated. However, we found that the cropland area of CGLS-LC100 is very high, while GLC2000 and Esri are relatively normal, and there must be some other reason for this (we suspect it may be because the cropland in CGLS-LC100 is an independent thematic cropland file rather than a single land cover layer). In addition to that, the cropland of GLASS, GLC2000, CLUDs, GlobeLand30, CLCD, and GLCFCS30 contains several different subcategories (as shown above), while FAO-GLCshare and CGLS-LC100 are thematic files of cropland, which only considered cropland and non-cropland. We found that both FAO and CGLS-LC100 have high cropland area.

Secondly, different classification strategies and mapping methods were adopted by these datasets. There is no doubt that they will lead to difference in classification accuracies. Some other factors such as the mapping extent and training samples may also affect the performance of accuracy. Through relevant reference, we find that CLCD data was produced by massive Landsat images, a large number of stable sufficient training samples and advanced algorithms. [55]. GLC_FCS30 was generated using Landsat imagery time

series, high quality training data of GSPECLib and random forest model, and 44,043 validation samples show that it has more spatial details and more diverse land cover types [21]. Esri was produced by Impact Observatory's deep learning artificial intelligence land classification model which uses a massive training dataset of billions of human labeled image pixels [41]. The resolution and precision of these data are very high. However, the accuracy of other data is not good in different aspects, such as rough resolution, old time, unadvanced satellite and classification algorithm, which lead to the differences in the identification of cropland in mixed regions. It should be stated that, our validation samples simply evaluated the Northeast China cropland and may be not representative for other regions accuracy assessments, more representative truth points of ground need to be collected for quantitative implementation in other regions in the future. Also, in this work, we merely pay attention to the cropland cover, and our conclusions on the accuracy evaluation of these datasets may not be representable for other land cover classes. Thirdly, there are mountains, hills, plains, rivers, wetlands and transition zones in Northeast China. This complicated landscape condition is also one of the reasons leading to the difference and uncertainties among these datasets.

4.2. Other ~~potential~~ possible influencing factors of the dataset ~~uncertainties~~ performance

The diverse definitions and classification of cropland in these evaluated datasets can also introduce ~~uncertainties~~ the difference of classification performance. For example, in GLC2000, CGLS-LC100 and Esri, the croplands are "pure" cropland, including rainfed croplands and flooded crops, such as cereal, corn, wheat, rice. The croplands in GLASS, GlobeLand30, CLCD, GLC FCS30 and FROM-GLC mainly refer to rainfed/irrigated cropland, orchard and temporally bare farmland, etc. However, in GLCNMO, the cropland includes herbaceous crop(s), the paddy field of graminoid crops and non-graminoid crops, and the mosaic of cropland and other vegetation. In CCI-LC, the cropland incorporates rainfed/irrigated cropland, herbaceous, tree or shrub cover, and the mosaic of cropland and natural vegetation (Supplementary Table S5). In general, the mixture of herbaceous or shrubs and cropland and the narrow definition inevitably result in ~~different~~ some overestimations and a little underestimation of cropland area, respectively. The cropland definition of GLC2000, CGLS-LC100 and Esri are "pure" cropland, and thus their estimation of cropland area should be somewhat underestimated. However, we found that the cropland area of CGLS-LC100 is very high, while GLC2000 and Esri are relatively normal, and there must be It could be attributed to some other reasons for this (we suspect it may be because the cropland in CGLS-LC100 is an independent thematic cropland file rather than a single land cover layer). In addition to that, the cropland of GLASS, GLC2000, CLUDs, GlobeLand30, CLCD, and GLCFCS30 contains several different subcategories (as shown above), while FAO-GLCshare and CGLS-LC100 are thematic files of cropland, which only considered cropland and non-cropland. We found that both FAO and CGLS-LC100 have high cropland area.

~~Also~~ Additionally, there are mountains, hills, plains, rivers, wetlands and transition zones in Northeast China. This complicated landscape condition is also one of the reasons leading to the difference and uncertainties among these datasets. It should be stated that, in this work, our validation samples simply evaluated the Northeast China cropland and our conclusions on the accuracy evaluation of these datasets may be not representative for other land cover classes or other regions accuracy assessments, more representative truth points of ground need to be collected for quantitative implementation in other regions in the future.

4.2.3. Some other factors that causes the errors ~~Uncertainties~~ existing in our evaluation work

One potential source of error in our accuracy assessment is the use of validation samples from the same phase year to evaluate datasets from different years, particularly when

land cover changes occur in specific regions. One possible reason for potential errors in accuracy assessment is the use of validation samples from the same phase year to evaluate datasets from different years, especially when land cover changes occur in certain regions. In this study, validation samples of four phrase years were generated through visual interpretation of high-resolution images. For example, samples collected in the same phase year of 2010 can be used for Phase-2010, GLCNMO-2008, GlobCover-2009, and GlobCover-2005. Using validation samples from the same phase year (2010) to evaluate datasets from different years (2008/2009/2005) may be problematic if land cover changes occur in some regions, reducing the reliability of commission/omission errors. For instance, for Phase-2010, GLCNMO-2008, GlobCover-2009 and GlobCover-2005 use the samples collected in the same phase year of 2010. There is no doubt that using validation samples from the same phase year (2010) to evaluate datasets from different years (2008/2009/2005) may be problematic if the land cover changes occur in some regions and further affecting the reliability of commission/omission errors. However, this situation occurred primarily in a few datasets, including CLCNMO-2003 and GlobCover-2005 in Phase-2000, CLCNMO-2008, GlobCover-2005 and GlobCover-2009 in Phase-2010, FAO-GLCshare-2014 and FROM-GLC-2017 in Phase-2015, and CLCD-2019 in Phase-2020. And, while the land cover type in some traditional cropland regions will not change in the short term, some changes will occur at the intersection of land cover types, such as cropland and grassland, shrub or bare land. And in some traditional region of cropland, the land cover type will not change in a short term, some changes tend to occur at the intersection of land cover types, such as cropland and grassland, shrub or bare land. The mismatch between the period of these verification samples and the data leads to the commission/omission errors, and the extent of this error cannot be quantifiable, which is a shortcoming point of our research. Another reason for this issue is limited to the accessibility of high resolution imagery at the annual timescale. We cannot obtain the high frequency reference imagery series in this study area.

Another factor issue that caused the error evaluation uncertainty is the cropland sub-categories mosaic. In terms of the mosaic cropland found in some land cover datasets, the cropland mosaic which accounts for more than 50% is directly divided into cropland in the data processing section. Such an issue is mainly related to three products GLCNMO, CCI-LC and GlobCover. We calculated the proportion of mosaic cropland pixel-pixels in the whole cropland pixel. Mosaic cropland accounted for 9.6% in GCLNMO-2003, 4.3% in GCLNMO-2008, and 14.5% in GCLNMO-2013. In CCI-LC, the proportion of these mosaics was 7.2% in CCI-LC-000, 7.0% in CCI-LC-2010, and 7.0% in CCI-LC-2015, respectively. However, in the GlobCover, this proportion is relatively large, 35.5% in GlobCover-2005 and 39.8% in GlobCover-2009. We looked up the definition of the mosaic class in GlobCover, which is "mosaic cropland (50-70%)/ vegetation (grassland/ shrubland/ forest) (20-50%)". Therefore, in these 35.5% or 39.8% of cropland mosaic pixels, the proportion of non-cropland is pretty small (e.g., 8.0%–19.9% in GlobCover-2009). However, there is no doubt that some omission and commission errors are also caused by these cropland mosaic pixels.

4.34. Comparison with the existing prior assessment articlework

In the recent study by Zhang et al.'s [65] article, a comparative analysis and accuracy assessment of six products with 30 m resolution of China in 2015 were conducted, where GLAD and GFSAD are thematic cropland data. Four accuracy indexes were used, including the overall accuracy, commission error, omission error, and Matthews correlation coefficient. We collected 13 sets of data for our study. In our study, we collected 13 sets of data and conducted a comparative analysis and assessment of the cropland in Northeast China for the four years-phases of around the years 2000, 2010, 2015, and 2020. During the evaluation process, we recognized that the years are not uniform, the data resolutions are different, and all data are land cover products and not thematic cropland products. The accuracy evaluation indexes we used include overall accuracy, producer's accuracy, user's

accuracy, and the Matthews correlation coefficient. Through our ~~rigorous~~ methodology, we aimed to provide valuable insights into the cropland changes in Northeast China and contribute to the scientific community's understanding of land use dynamics.

Both ~~papers~~ studies compare and analyze the spatial consistency of data at regional (China, Northeast China), provincial, meridional and zonal scales. However, Zhang et al' ~~paper~~ study goes into more detail on provincial agreements and differences, not only calculating the proportion of cropland area at different agreement levels in each province but also the relative area difference between each data and the statistics for each province. An important difference between the two papers is that in the discussion section, Zhang et al found the relationship between elevation and slope, and concluded that the area proportion with high agreement decreases with an increase in elevation and slope. When the elevation and slope increase, the data difference is obvious. Our ~~paper~~ study did not investigate either aspect, which should be strengthened in the next work.

5. Conclusions

As more and more global or national land cover datasets are published and available, it is becoming easier to obtain cropland information from these open-access products. However, the accuracy evaluation results of these datasets at a global/national scale for all LULC elements (classes) may not be replaceable for local-area or class-specific accuracy. Northeast China is a major agricultural region of China and ~~have~~ has complicated landscapes and varied terrains. This paper takes Northeast China to evaluate the performances of our collected thirteen sets of land cover datasets in the accuracy, spatial location, and area of croplands. The similarities and differences are shown in a multi-scale comparison between datasets.

Accuracy validation based on ground truth samples indicates that, in general, compared with other datasets, CLCD, GlobeLand30, GLC_FCS30, and Esri have the best performance in cropland classification with ~~the~~ overall accuracies of more than 0.89. On the contrary, the OAs of CGLS-LC100-2019 and GlobCover are lower than 0.75. For commission and omission error, CGLS-LC100 has a severe overestimation, while GLASS suffers from a significantly underestimation of cropland area, and the bias situation of CLCD is the least. In the aspect of spatial consistency and difference, the traditional agricultural regions such as Sanjiang-Songnen-Liaohe Plain in Northeast China have the highest spatial consistency. Lower consistency regions are distributed in mountainous, hilly regions and the transition areas between them. For the scatter plots between pairwise data, the agreement between CLCD and GLC_FCS30 is the highest ($r^2 = 0.96$), followed by GlobeLand30 with CLUDs (0.89), and GlobeLand30 with CLCD (0.88), all their RMSEs are less than 12 km². CGLS-LC100 is in poor agreement with any data, with r^2 lower than 0.56 and RMSE higher than 30 km².

In exploring the influence factors of the difference and uncertainties in the cropland range and location among all datasets, although the purpose of these datasets was to provide accurate cultivate information, factors such as the definitions of cropland, the landscape characteristics of Northeast China, classification scheme, and different satellite sensors, jointly produced different mapping results. It is arbitrary to say that this dataset is more suitable for this particular application, however, with the increasing demand for high-precision cropland datasets, users should consider the accuracy of cropland class when selecting a dataset. The accuracy evaluations and comprehensive comparison analysis of this work are expected to provide a meaningful and valuable reference for those (e.g., cropland ~~surveyors~~ surveyors) who use ~~using~~ these datasets in Northeast China. In addition, the detailed evaluation of this work could provide feedback to data generators, facilitate the improvement of data processing algorithms and classification techniques, better service with future cropland applications and mapping or cropland data fusion.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, C.S. and T.C.; methodology, P.C., T.C., and K.L.; software, P.C., and D.Z.; validation, P.C., T.C., D.Z. and K.L.; formal analysis, P.C., T.C., C.S. and K.L.; investigation, D.Z. and P.C.; resources, P.C. and D.Z.; data curation, P.C., T.C. and D.Z.; writing—original draft preparation, C.S., T.C., P.C. and D.Z.; writing—supervision, C.S., T.C., K.L.; visualization, P.C., T.C. and D.Z.; project administration, C.S. and T.C.; funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA28020503, XDA23100102), the Jiangsu Normal University Postgraduate Research & Practice Innovation Program (2021XKT0094), the National Key Research and Development Program of China (2018YFD0900804, 2019YFA0607101), and the National Natural Science Foundation of China (41971403, 41930102, 42171421).

Data Availability Statement: The GLC-2000 can be downloaded from the web <https://forobs.jrc.ec.europa.eu/products/glc2000/products.php>. The FAO-GLCshare is available at FAO Map Catalog. The FROM-GLC is accessible on the Tsinghua website at Finer Resolution Observation and Monitoring - Global Land Cover (tsinghua.edu.cn). The Esri can be downloaded from the web at <https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2>. The GLASS-GLC is available at <https://doi.org/10.1594/PANGAEA.913496>. The GLCNMO is available at https://github.com/globalmaps/gm_lc_v1. The CCI-LC is available at <http://maps.elie.ucl.ac.be/CCI/viewer/download.php>. The GlobCover can be downloaded from European space agency data user element at http://due.esrin.esa.int/page_globcover.php. The CGLS-LC100 can be downloaded from the web at <https://zenodo.org/record/3939038#.Y3TzSMdBxBM>. The GlobeLand30 are online accessible at <https://www.webmap.cn/commres.do?method=globeIndex>. The GLC_FCS30 data is available at <https://zenodo.org/record/3986872#.Y3Ty2sdBxBN>. The CLUDs is available on request from the corresponding author. The CLCD is freely available at <https://doi.org/10.5281/zenodo.4417810>. The auxiliary datasets of China cropland are available at <https://doi.org/10.6084/m9.figshare.13356680.v1>.

Acknowledgments: We would like to thank the editor and the anonymous reviewer, whose constructive comments will help to improve the presentation of this paper.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Pinstrup-Andersen, P. Perspectives in World Food and Agriculture 2004. 2008; pp. 87-97.
2. Gibbs, H.K.; Ruesch, A.S.; Achard, F.; Clayton, M.K.; Holmgren, P.; Ramankutty, N.; Foley, J.A. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc Natl Acad Sci U S A* 2010, 107, 16732-16737, doi:10.1073/pnas.0910275107.
3. Grekousis, G.; Mountrakis, G.; Kavouras, M. An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing* 2015, 36, 5309-5335, doi:10.1080/01431161.2015.1093195.
4. Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 2010, 21, 1303-1330, doi:10.1080/014311600210191.
5. Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* 2010, 21, 1331-1364, doi:10.1080/014311600210209.
6. Bartholomé, E.; Belward, A. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* 2005, 26, 1959-1977, doi:10.1080/01431160412331291297.
7. Friedl, M.A.; Sulla-Menashé, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 2010, 114, 168-182, doi:10.1016/j.rse.2009.08.016.
8. Bicheron, P.; Leroy, M.; Brockmann, C.; Krämer, U.; Miras, B.; Huc, M.; Niño, F.; Defourny, P.; Vancutsem, C.; Arino, O.; et al. Globcover: a 300 m global land cover product for 2005 using ENVISAT MERIS time series. *Proceeding of the Second International Symposium on Recent Advances in Quantitative Remote Sensing* 2006, 538-542.
9. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 2015, 103, 7-27, doi:10.1016/j.isprsjprs.2014.09.002.
10. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* 2012, 34, 2607-2654, doi:10.1080/01431161.2012.748992.
11. Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Schmullius, C. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment* 2008, 112, 2538-2556, doi:10.1016/j.rse.2007.11.013.
12. Kaptué Tchuenté, A.T.; Roujean, J.-L.; De Jong, S.M. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation* 2011, 13, 207-219, doi:10.1016/j.jag.2010.11.005.
13. Wu, W.; Shibasaki, R.; Yang, P.; Zhou, Q.; Tang, H. Remotely sensed estimation of cropland in China: a comparison of the maps derived from four global land cover datasets. *Canadian Journal of Remote Sensing* 2014, 34, 467-479, doi:10.5589/m08-059.
14. Congalton, R.; Gu, J.; Yadav, K.; Thenkabail, P.; Ozdogan, M. Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sensing* 2014, 6, 12070-12093, doi:10.3390/rs61212070.
15. Lu, M.; Wu, W.; Zhang, L.; Liao, A.; Peng, S.; Tang, H. A comparative analysis of five global cropland datasets in China. *Science China Earth Sciences* 2016, 59, 2307-2317, doi:10.1007/s11430-016-5327-3.
16. Deines, J.M.; Patel, R.; Liang, S.-Z.; Dado, W.; Lobell, D.B. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment* 2021, 253, doi:10.1016/j.rse.2020.112174.
17. You, N.; Dong, J. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing* 2020, 161, 109-123, doi:10.1016/j.isprsjprs.2020.01.001.
18. Giri, C.; Zhu, Z.; Reed, B. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment* 2005, 94, 123-132, doi:10.1016/j.rse.2004.09.005.
19. McCallum, I.; Obersteiner, M.; Nilsson, S.; Shvidenko, A. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation* 2006, 8, 246-255, doi:10.1016/j.jag.2005.12.002.
20. Liu, L.; Zhang, X.; Gao, Y.; Chen, X.; Shuai, X.; Mi, J. Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects. *Journal of Remote Sensing* 2021, 2021, 1-38, doi:10.34133/2021/5289697.
21. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data* 2021, 13, 2753-2776, doi:10.5194/essd-13-2753-2021.
22. Hansen, M.C.; Reed, B. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing* 2010, 21, 1365-1373, doi:10.1080/014311600210218.
23. Tsendbazar, N.-E.; de Bruin, S.; Fritz, S.; Herold, M. Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sensing* 2015, 7, 15804-15821, doi:10.3390/rs71215804.
24. Bai, Y.; Feng, M.; Jiang, H.; Wang, J.; Zhu, Y.; Liu, Y. Assessing Consistency of Five Global Land Cover Data Sets in China. *Remote Sensing* 2014, 6, 8739-8759, doi:10.3390/rs6098739.
25. Ning, J.; Zhang, S.; Cai, H.; Bu, K. A Comparative Analysis of the MODIS Land Cover Data Sets and Globcover Land Cover Data Sets in Heilongjiang Basin. *Journal of Geo-information Science* 2012, 14, 240-249.

26. Liu, Y.; Zhou, M. Comparative Analysis on Three Land Cover Datasets based on IGBP Classification System over Hanjiang River Basin. *Remote Sensing Technology and Application* 2017, 32, 575-584.
27. Yang, Y.; Xiao, P.; Feng, X.; Li, H.; Chang, X.; Feng, W. Comparison and assessment of large-scale land cover datasets in China and adjacent regions. *National Remote Sensing Bulletin* 2014, 18, 453-475.
28. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy assessment of seven global land cover datasets over China. *ISPRS Journal of Photogrammetry and Remote Sensing* 2017, 125, 156-173, doi:<https://doi.org/10.1016/j.isprsjprs.2017.01.016>.
29. Hua, T.; Zhao, W.; Liu, Y.; Wang, S.; Yang, S. Spatial Consistency Assessments for Global Land-Cover Datasets: A Comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. *Remote Sensing* 2018, 10, doi:10.3390/rs10111846.
30. Pérez-Hoyos, A.; Rembold, F.; Kerdiles, H.; Gallego, J. Comparison of Global Land Cover Datasets for Cropland Monitoring. *Remote Sensing* 2017, 9, doi:10.3390/rs9111118.
31. Gao, Y.; Guo, Y.; Wang, W.; Li, F.; Huang, P. Accuracy evaluation of different land use or land cover data in grassland of northern China. *Chinese Journal of Ecology* 2019, 38, 283-293, doi:10.13292/j.1000-4890.201901.003.
32. Niu, Z.; Guo, Y.; Shan, Y.; Zhang, H. Accuracy Assessment of Wetland Categories from the GlobCover2009 Data Over China. *Wetland Science* 2012, 10, 389-395, doi:10.13248/j.cnki.wetlandsci.2012.04.002.
33. Meng, W. Accuracy Assessment for Regional Land Cover Remote Sensing Mapping Product Based on Spatial Sampling: A Case Study of Shaanxi Province, China. *Journal of Geo-information Science* 2015.
34. Ma, J.; Qun, S.; Qiang, X.; Bowei, W. Accuracy Assessment and Comparative Analysis of GlobeLand30 Dataset in Henan Province. 2016.
35. Wang, Y.; Zhang, J.; Liu, D.; Yang, W.; Zhang, W. Accuracy Assessment of GlobeLand30 2010 Land Cover over China Based on Geographically and Categorically Stratified Validation Sample Data. *Remote Sensing* 2018, 10, 1213.
36. Kussul, N.; Shelestov, A.; Basarab, R.; Skakun, S.; Kussul, O.; Lavreniuk, M. Geospatial intelligence and data fusion techniques for sustainable development problems; 2015; Volume 1356.
37. Jokar Arsanjani, J.; Tayyebi, A.; Vaz, E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat International* 2016, 55, 25-31, doi:<https://doi.org/10.1016/j.habitatint.2016.02.003>.
38. Manakos, I.; Karakizi, C.; Gkinis, I.; Karantza, K. Validation and Inter-Comparison of Spaceborne Derived Global and Continental Land Cover Products for the Mediterranean Region: The Case of Thessaly. *Land* 2017, 6, 34.
39. Mayaux, P.; Eva, H.; Gallego, J.; Strahler, A.H.; Herold, M.; Agrawal, S.; Naumov, S.; De Miranda, E.E.; Di Bella, C.M.; Ordoyney, C.; et al. Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing* 2006, 44, 1728-1739, doi:10.1109/tgrs.2006.864370.
40. John, L.; Renato, C.; Ilaria, R.; Mario, B. Global Land Cover-share of year 2014- Beta-Release 1.0 FAO Global Land Cover Network(GLCN). Available online: <https://www.fao.org/uploads/media/glc-share-doc.pdf> (accessed on 2023-01-10).
41. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.; Mathis, M.; Brumby, S. Global land use / land cover with Sentinel 2 and deep learning; 2021; pp. 4704-4707.
42. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data* 2022, 9, doi:10.1038/s41597-022-01307-4.
43. Liu, H.; Gong, P.; Wang, J.; Clinton, N.; Bai, Y.; Liang, S. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth System Science Data* 2020, 12, 1217-1243, doi:10.5194/essd-12-1217-2020.
44. Tateishi, R.; Hoan, N.T.; Kobayashi, T.; Alsaadeh, B.; Tana, G.; Phong, D.X. Production of Global Land Cover Data – GLCNMO2008. *Journal of Geography and Geology* 2014, 6, doi:10.5539/jgg.v6n3p99.
45. Kobayashi, T.; Tateishi, R.; Alsaadeh, B.; Sharma, R.C.; Wakaizumi, T.; Miyamoto, D.; Bai, X.; Long, B.D.; Gegentana, G.; Maitiniyazi, A.; et al. Production of Global Land Cover Data – GLCNMO2013. *Journal of Geography and Geology* 2017, 9, doi:10.5539/jgg.v9n3p1.
46. Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tseng-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaadeh, B.; et al. Production of global land cover data – GLCNMO. *International Journal of Digital Earth* 2011, 4, 22-49, doi:10.1080/17538941003777521.
47. Pierre, D.; G, K.; C, B.; M, B.; M, P.; S, B.; C, L.; M, S.; M, S. Land Cover CCI: Product User Guide Version 2. Available online: <http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-PUG-v2.5.pdf> (accessed on 2023-01-10).
48. Bicheron, P.; Defourny, P.; Brockmann, C.; Schouten, L.; Vancutsem, C.; Huc, M.; Bontemps, S.; Leroy, M.; Frédéric, A.; Herold, M.; et al. GLOBCOVER: products description and validation report. 2008.
49. Defourny, P.; Bontemps, S.; Bogaert, E. GLOBCORINE 2009. Product Description Manual; 2010.
50. Buchhorn, M.; Smets, B.; Bertels, L.; Roo, B.D.; Lesiv, M.; Tsendbazar, N.-E.; Li, L.; Tarko, A. Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Product User Manual. 2020, doi:10.5281/zenodo.3938974.
51. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; Peng, S.; Han, G.; Zhang, H.; He, C.; et al. Concepts and Key Techniques for 30 m Global Land Cover Mapping. *Acta Geodaetica et Cartographica Sinica* 2014, 43, 551-557, doi:10.13485/j.cnki.11-2089.2014.0089.
52. Chen, J.; Chen, L.; Chen, F.; Ban, Y.; Li, S.; Han, G.; Tong, X.; Liu, C.; Stamenova, V.; Stamenov, S. Collaborative validation of GlobeLand30: Methodology and practices. *Geo-spatial Information Science* 2021, 24, 134-144, doi:10.1080/10095020.2021.1894906.

53. Liu, J.; Liu, M.; Zhuang, D.; Zhang, Z.; Deng, X. Study on Spatial Pattern of Land-use Change in China During 1995-2000. *Science in China Series D Earth Sciences* 2003, 46, 373-384, doi:10.1360/03yd9033.
54. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W. Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Journal of Geographical Sciences* 2014, 69, 3-14.
55. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth System Science Data* 2021, 13, 3907-3925, doi:10.5194/essd-13-3907-2021.
56. Yu, Z.; Jin, X.; Miao, L.; Yang, X. A historical reconstruction of cropland in China from 1900 to 2016. *Earth System Science Data* 2021, 13, 3203-3218, doi:10.5194/essd-13-3203-2021.
57. Olofsson, P.; Stehman, S.; Woodcock, C.; Sulla-Menashe, D.; Sibley, A.; Newell, J.; Friedl, M.; Herold, M. A global land-cover validation data set, part I: Fundamental design principles. *International Journal of Remote Sensing* 2012, 33, 5768-5788, doi:10.1080/01431161.2012.674230.
58. Fung, T.; LeDrew, E. The Determination of Optimal Threshold Levels for Change Detection Using Various Accuracy Indices. 2008.
59. Janssen, L.L.F.; Wel, F.v.d. Accuracy assessment of satellite derived land - cover data : a review. *Photogrammetric Engineering and Remote Sensing* 1994, 60, 419-426.
60. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020, 21, 6, doi:10.1186/s12864-019-6413-7.
61. Clark, M.L.; Aide, T.M.; Grau, H.R.; Riner, G. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment* 2010, 114, 2816-2832, doi:10.1016/j.rse.2010.07.001.
62. Ran, Y.; Li, X.; Lu, L. Accuracy Evaluation of the Four Remote Sensing Based Land Cover Products over China. *Journal of Glaciology and Geocryology* 2009, 31, 490-500.
63. Foody, G. Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment* 2002, 80, 185-201, doi:10.1016/S0034-4257(01)00295-4.
64. Lambin, E.; Geist, H. Land-Use and Land-Cover Change: Local Processes and Global Impacts. 2006; Volume 18.
65. Zhang, C.; Dong, J.; Ge, Q. Quantifying the accuracies of six 30-m cropland datasets over China: A comparison and evaluation analysis. *Computers and Electronics in Agriculture* 2022, 197, 106946, doi:https://doi.org/10.1016/j.compag.2022.106946.

Article

Comparison and assessment of different land cover datasets on the cropland in Northeast China

Peipei Cui ^{1,2}, Tan Chen ¹, Yingjie Li ², Kai Liu ¹, Dapeng Zhang ¹ and Chunqiao Song ^{1,3,4*}

¹ Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

² School of Geography, Geomatics and Planning, Jiangsu Normal University, 101 Shanghai Road, Tongshan District, Xuzhou, Jiangsu 221116, China

³ University of Chinese Academy of Sciences, Nanjing (UCASNJ), Nanjing 211135, China

⁴ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: Correspondence: cqsong@niglas.ac.cn; Tel.: +86-25-86882293

Abstract: The provision of precise and dependable information regarding the extent and distribution of cropland is imperative for the evaluation of food security, agricultural planning, and resource management. Cropland is an important component of land cover type and is offered in multiple existing global/regional land cover products. However, global-scale accuracy evaluation may not be representative for class-specific or local-area accuracy, such as Northeast China, which is an important grain-producing region of China and has various types of cultivated land (e.g., wheat, rice) and diverse terrains. It poses a great challenge in generating precise cropland classification by automated mapping. Thus, it is indispensable to evaluate the accuracy and reliability in these various land cover datasets before using them. In this study, we collected thirteen sets of global or national-scale land cover datasets. Through visual interpretation of high-resolution images, ground “truth” samples were collected to evaluate the data accuracy across Northeast China. The overall accuracy (OA) evaluation results in Phase-2020 shows that CLCD has the highest value with 0.914, followed by GlobeLand30 (0.906), GLC_FCS30 (0.902), and Esri (0.896) for cropland classification in Northeast China. CGLS-LC100 has the lowest OA (0.710). For the commission and omission errors of six datasets of Phase-2020, CGLS-LC100 has an obvious overestimation (larger commission error), while the two national-scale datasets (CLCD and CLUDs) performs relatively better. In the aspect of spatial consistency, high spatial agreement among the nine Phase-2015 datasets or in the six Phase-2020 datasets could be discovered in traditional agricultural regions like the Sanjiang-Songnen-Liaohe Plain, and low agreement is found in the transition areas of mountains (hills) and plains with the mixed landscape of forest (grassland) and farmland. In the aspect of comparison pairwise data, CLCD is in good agreement with GLC_FCS30, GlobeLand30, and Esri, while the CGLS-LC100 is in the poorest agreement with any other dataset. The comparison and evaluation results are expected to provide the reference on which aspects and what extent these land cover products may be consistent and guide the cropland data product selection for Northeast China.

Keywords: land cover; Northeast China; cropland classification; accuracy assessment; spatial agreement; regional comparison

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date



Copyright: © 2023 by the authors.

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abbreviations

GLC2000, Global Land Cover 2000; **FAO-GLCshare**, The Global Land Cover-share created by FAO (Food and Agriculture Organization, FAO); **FROM-GLC**, Finer Resolution Observation and Monitoring of Global Land Cover; **Esri**, a global map of land use/land cover (LULC) derived from ESA Sentinel-2 imagery at 10m resolution; **GLASS-GLC**, Global Land Surface Satellite-Global Land Cover; **GLCNMO**, Global Land Cover

by National Mapping Organizations; **CCI-LC**, Climate Change Initiative-Land Cover; **GlobCover**, Global Land Cover Map; **CGLS-LC100**, Copernicus Global Land Service-Land Cover 100m; **GlobeLand30**, World's First Global Land Cover Datasets at a 30m; **GLC_FCS30**, Global Land Cover product with Fine Classification System at 30m; **CLUDs**, China's Land-use/cover datasets; **CLCD**, China Land Cover Dataset

1. Introduction

Agricultural land meets global demands such as human food, stock feed, and biofuel. These demands are growing rapidly due to the increasing population and consumption [1,2]. Therefore, understanding the range, distribution and dynamics change of cropland is required for food demands and environmental sustainability, especially for agricultural planning, cropland monitoring, and food security assessment, among others. An increasing number of global and regional land use/cover (LULC) datasets [3], including the farmland category, have been produced by the researchers and are freely available to the public with the rapid development of remote sensing and new technologies (i.e., cloud computing, big data and machine learning). Coarse resolution images (i.e., pixels of 5 km or 1 km) were generally used in the early period [4-6]; however, their application at national or regional levels is limited by the low precision and image quality. Then, the spatial resolution of products was enhanced to 500 m or 300 m [7,8]. Nowadays, land cover mapping using high-resolution images, such as 30 m or 10 m resolution satellite data [9,10], has received increasing attention from the scientific community. All these datasets offer precious cropland information for multi-discipline research and applications. However, there are considerable inconsistencies and uncertainties among these LULC datasets, especially in the areas with complex transition landscapes, since they have been generated using different satellites/sensors, classification features, classifiers and methods, etc. [11,12]. A few previous studies have also shown that discrepancies and uncertainties of cropland also arise due to cropland definitions or different classification schemes [13-15]. Therefore, the accuracy assessment and comparative studies of these LULC datasets are necessary for relevant application research before using them [16,17].

In the existing studies, many researchers have performed accuracy evaluations for various LULC datasets. For example, multiple prior studies have compared and evaluated the spatial (in)consistency of global land cover products from various perspectives, e.g., at the same spatial resolution (i.e., 1 km or 30 m) [18-21], for certain type of satellite sensors [22], on a global scale [20,21,23], local regions (like China or one basin) [24-28], or the global and continental scales by elevation and climate partitions [29]. Others have compared and summarized an uncertainty analysis of global land cover datasets using an error budget approach [14]. In addition, a few researchers compared a certain element or land cover theme (such as grassland, wetland, etc.) in some land cover datasets to investigate their comparative advantages [30-32]. The prior evaluation work provides important references for choosing land cover products and understanding the spatial characteristics of LULC in global and national scales. However, findings on the global/national evaluations of the product accuracy for all LULC elements (classes) may not be replaceable for local-area or class-specific accuracy. For example, a total number of 159,874 validation samples were selected by the producers for the assessment of GlobeLand30-2010. The product overall accuracy was estimated to be 80.3% [9]. Regional or country-wide accuracy evaluations of GlobeLand30 have also been evaluated by many researchers and reported different assessment results. For instance, a number of 1467 sample points were used by Meng et al. [33] to verify the performance of GlobeLand30-2010 in Shaanxi Province, which indicates an overall accuracy of 80.0%. The overall accuracy of GlobeLand30-2010 performed by Ma et al. is 81.5% for the study area in Henan province [34]. Another assessment based on 8400 verification sample pixels by Wang et al. [35] indicates that the overall accuracy of GlobeLand30-2010 data in China is 84.2%. Regional accuracy evaluations of GlobeLand30 were also being carried out in other countries and their evaluation results are also different. Overall accuracies were evaluated by Kussul et al. [36] at about

89.7% for Ukraine, while assessments in Iran by Jokar Arsanjani et al. [37] and in Thessaly by Manakos et al. [38] indicated the overall accuracies of 77.9% and 86.0%, respectively. A summary of the comparison of the regionally achieved accuracies from different land cover datasets vs the claimed global accuracies typically reported by global land cover datasets can be found in Supplementary Table S1.

Northeast China is an important production base of grain and other agricultural products, and is one of the three main grain-producing regions in China, which provides the highest grain food production and the highest grain commodity rate to the country. Thus, it plays a critical supporting role to ensure the national food security. Moreover, in Northeast China, the different types of croplands and the diversity of terrains and landscapes have led to an obvious transitional zone of cropland distribution. Under these complex conditions, the remote sensing extraction of cropland is very challenging. At present, cropland information can be obtained from various remote sensing data. However, due to the different purposes, study areas, evaluation methods, and other reasons, results on the accuracy evaluation of global/national land cover datasets are inconsistent and incongruous. Thus, it is essential to investigate the accuracy of different land cover datasets over Northeast China for better depicting its cropland distribution.

The purpose of this study is to examine and contrast the farmland category of distinct land cover datasets with varying resolutions and satellite data sources over Northeast China. The evaluation proceeds from the following three aspects: (1) approximately 2,000 “ground-truth” samples for each epoch (the years around 2000, 2010, 2015, and 2020) were manually interpreted across Northeast China to verify the dataset accuracies. (2) The spatial agreement among these datasets was compared and analyzed at different scales, including provincial, longitudinal, and latitudinal dimensions. (3) The possible causes for the uncertainties and discrepancies of the accuracies among these evaluated datasets were discussed eventually. Motivated by these targets, this paper conducts a relatively comprehensive comparison analysis and accuracy evaluation of various datasets collected across Northeast China, which is hoped to further understand the accuracies and abilities of these datasets to represent the cropland area in Northeast China and to provide an important reference for data users and to improve the cropland mapping and advanced applications in the future.

2. Materials and Methods

2.1. Study area

Northeast China is a major agricultural region (115°52'–135°09' E, 38°72'–53°55' N), including the provinces Heilongjiang, Jilin, and Liaoning, and several prefecture-level cities in the east of Inner Mongolia Autonomous Region (Hulunbuir, Xing'an League, Tongliao, and Chifeng). With the altitude ranging from 0 to 2665 m, Northeast China has various types of landforms, including the mountains of the Greater Khingan Mountains, the lesser Khingan Mountains and the Changbai Mountains, the eastern Liaoning and the western Liaoning, Horqin Sandy Land, Sanjiang-Songnen-Liaohe Plains, Songhua River and Nenjiang River, wetlands of various sizes, and transition zone of these mega-landforms, which leads to its complicated natural landscape conditions. Figure 1 shows its geographic location and topographic distributions.

From east to west, the territory is divided into temperate, warm temperate, and cold temperate zones, with monsoon and continental climate types. The elevation over the study area increases significantly from the hinterland to the outer, and the terrain changes are mainly shaped by the surrounding mountains and drainage network. All these factors contribute to the complex and diversified landforms of the region. Cropland is concentrated in the Sanjiang-Songnen-Liaohe Plain and mosaic with forest, grassland, wetland and other land cover types. Therefore, Northeast China is a typical representative of the cropland classification evaluation.

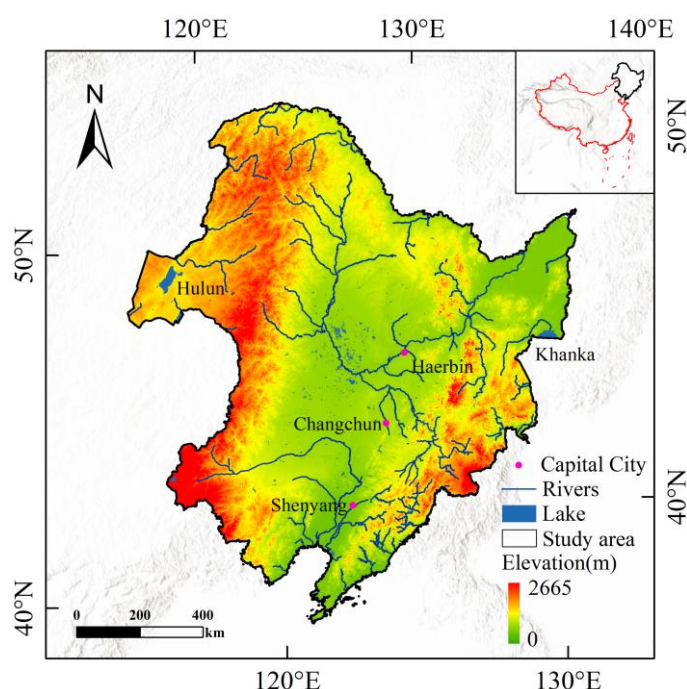


Figure 1. Geographical location and topography map of Northeast China.

2.2. Datasets

2.2.1 Global/regional land cover datasets

Global/regional land cover datasets are important information sources for understanding the complex interactions between human activities and global change. In this study, we focus on comparing and assessing the cropland classification performance of land cover data products at a global scale. Additionally, in order to confirm whether the classification with more input samples of land cover types at local scales can enhance the data accuracy, we add two another national-scale datasets in China. This study totally collected eleven global-scale datasets, among which the GLC2000, FAO-GLCshare-2014, FROM-GLC-2017 and Esri-2020 products only contain one year data layer, the GLASS-GLC, GLCNMO, CCI-LC, GlobCover, CGLS-LC100, GlobeLand30, and GLC_FCS30 contain multi-year data layers. The two national-scale data products are CLUDs and CLCD. CLUDs has a 5-year interval from 1990 to 2020, while CLCD has annual data layers from 1990 to 2019. The investigated land cover datasets are briefly described below. In addition, we have sorted out and summarized the satellites used in data production, the spatial resolution, and the classification techniques and algorithms used in land cover types, those key information about the thirteen datasets is shown in Table 1.

(1) The Global Land Cover 2000 (GLC2000) dataset was developed by the European Commission's Joint Research Center with a spatial resolution of 1 km. The GLC2000 legend is classified into 22 classes using unsupervised clustering and its overall accuracy is 68.6% [39].

(2) FAO-GLCshare (the Global Land Cover-share created by FAO (Food and Agriculture Organization)), produced by the United Nations' (UN) FAO in 2014 [40]. It uses the data fusion method to integrate the available national, regional, and global datasets with a resolution of 1 km. The product encompasses 11 classes and has an accuracy of 80%.

(3) The Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) was produced by using the data of Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) [10]. This data represents the land cover around 2017 with a spatial resolution of 30 m and encompasses ten land-cover classes.

(4) Another global land cover map, which we named Esri in this paper, is a 10 m resolution imagery derived from ESA Sentinel-2. It was generated using a deep learning

model that used more than 5 billion Sentinel-2 pixels and sampled from more than 20,000 sites. These sites are found in all major biomes in the world [41,42].

(5) The Global Land Surface Satellite-Global Land Cover (GLASS-GLC) is an annual dynamic record of global land cover product from 1982 to 2015. It was generated on the Google Earth Engine (GEE) platform with the latest version of GLASS (the Global Land Surface Satellite) CDRs (Climate Data Records) [43]. It has a resolution of 5 km and an average overall accuracy is 82.81%.

(6) The National Mapping Organizations (GLCNMO) Global Land Cover was generated by the International Steering Committee for Global Mapping. Version I (2003) has a spatial resolution of 1 km and version II (2008) and version III (2013) datasets [44,45] have an improved spatial resolution at 500 m [46].

(7) The Land Cover (LC) project of the Climate Change Initiative (CCI) by the European Space Agency (ESA) provides a series of annual datasets with 300m resolution from 1992 to 2015, termed as CCI-LC dataset [47]. The product uses the unsupervised spatio-temporal clustering and machine learning classification method, with a total of 22 land cover classes.

(8) The GlobCover global land cover map is developed by the ESA in collaboration with international network of partners [8,48,49]. This dataset contains 22 classes defined by the UN LCCS.

(9) The Copernicus Global Land Service Land Cover at the 100 m resolution (CGLS-LC100)-collection 3 was released by the Copernicus Global Land Service. It was derived from high quality land cover training sites based on PROBA-V satellite observations and multiple auxiliary datasets [50]. This global LULC map contains 23 classes.

(10) The GlobeLand30 was a 30-m global land cover product generated by the National Geomatics Center of China using the Pixel-Object-Knowledge operational method [9,51]. It has an overall classification accuracy of more than 80% worldwide [52]. It comprises 10 land cover classes.

(11) The global 30 m land cover dataset with a fine classification system (GLC_FCS30) in 2015 was produced using a time series image of Landsat and high-quality training data of the Global Spatial Temporal Spectra Library (GSPECLib) on the GEE platform. Then the GLC_FCS30-2020 was generated with the prior knowledge of experts and the multi-source auxiliary datasets. Both of them contain 30 land-cover classes [21].

(12) The China's land-use/cover datasets (CLUDs) were provided by the Resource and Environment Science and Data Center. Its resolution is 1 km and documented in detail China's land cover in the 1980s, 1990, 1995, 2000, 2005, 2010, 2015 and 2020. It includes 6 level-1 classes, which are cropland, grassland, forest, built-up area, water and barren, and 25 level-2 classes [53,54].

(13) The annual China Land Cover Dataset (CLCD) was derived from Landsat imagery on the GEE platform by Yang et al. [55], which contains annual land cover data layers at the 30 m spatial resolution in China from 1990 to 2019. Including 9 land cover types: cropland, forests, shrub, grasslands, water, snow/ ice, barren, impervious, and wetlands, its overall accuracy is reported at 79.31 %.

Table 1. The summary of main information of all collected datasets in this study. They are sorted in order of the spatial resolution from lowest to highest.

Datasets	Satellites or sensor	Time	Spatial resolution	Classification technique	Classification scheme
GLASS-GLC	AVHRR GLASS CDR	1982-2015	5 km	Random forest & LandTrendr	7 classes
GLC2000	SPOT VGT	2000	1 km	Generally unsupervised classification	22 classes
FAO-GLCshare	----	2014	1 km	Data fusion	11 classes
CLUDs	Landsat	1980 1990 1995	1 km	Extraction of remote	6 classes*

		2000 2005 2010 2015 2020		sensing information	
GLCNMO	Terra MODIS	2003	1 km	Supervised classification	20 classes
	Terra and Aqua MODIS	2008 2013	500 m	Supervised classification	20 classes
CCI-LC	ENVISAT MERIS SPOT VGT	1992-2015	300 m	Unsupervised spatio-temporal clustering & Machine learning classification	22 classes
GlobCover	MERIS	2005 2009	300 m	Generally unsupervised classification	22 classes
CGLS-LC100	PROBA-V	2015-2019	100 m	Supervised classification & Random forest	23 classes
GlobeLand30	Landsat TM/ETM+ HJ-1	2000 2010 2020	30 m	Pixel-Object-Knowledge classification approach	10 classes
CLCD	Landsat	1990-2019	30 m	Random forest	9 classes
GLC_FCS30	Landsat TM/ETM+/OLI	2015 2020	30 m	Operational SPECLib-based approach & Random forest	30 classes
FROM-GLC	Landsat TM/ETM+/OLI Sentinel-2	2017	30 m	Random forest	10 classes
Esri	Sentinel-2	2020	10 m	Deep learning model	10 classes

Note: 6 classes* means 6 classes in the first level and 25 classes in the second level.

According to the time characteristics of the mapping data sources, we classified all data into four phases: 2000, 2010, 2015 and 2020. The Phase-2000 comprises eight datasets, Phase-2010 also includes eight datasets, Phase-2015 predominantly consists of nine datasets, and Phase-2020 comprises six datasets. Table 2 displays the data contained in these four periods.

Table 2. The datasets included in four phases.

Phase-2000	Phase-2010	Phase-2015	Phase-2020
GLASS-GLC	GLASS-GLC	GLASS-GLC	
GLC2000		FAO-GLCshare-2014	
CLUDs	CLUDs	CLUDs	CLUDs
GLCNMO-2003	GLCNMO-2008	GLCNMO-2013	
CCI-LC	CCI-LC	CCI-LC	
		CGLS-LC100	CGLS-LC100-2019
GlobeLand30	GlobeLand30		GlobeLand30
CLCD	CLCD	CLCD	CLCD-2019
	GlobCover-2009	GLC_FCS30	GLC_FCS30
GlobCover-2005	GlobCover-2005	FROM-GLC-2017	Esri

2.2.2 Other auxiliary dataset

We also collected the auxiliary data of cropland distribution in Northeast China reconstructed by Yu et al. [56] as statistical data. This dataset was reconstructed by a multi-source data fusion method. By coordinating the data accuracy, time and spatial resolution of diverse data sources, the continuous annual cropland distribution dataset of China was reconstructed using the spatialized cropland distribution approach. Four statistical datasets were used in this process, including the Land and Resources Statistical Yearbook, the National Land and Resources Bulletin, the Chinese Agriculture Yearbook, and the national crop production from the Chinese Statistical Yearbook, as well as multiple remote

sensing products, including global cropland data provided by the Global Food Security Analysis-Support Data, GlobeLand30, the China Land Use and Cover Change, FROM-GLC, and so on. Overall, Yu et al. [56] assimilated the trend information and spatial pattern of various satellite datasets to produce series of cropland percentage maps, which are spatially explicit and span the time period from 1900 to 2016. Through intensive comparisons, the authors are confident that this data provides reliable sources of cropland maps [56].

2.3. Data processing procedures

The flowchart of this study is shown in Figure 2. The following sections describe the key data processing steps. Three key steps are involved in the accuracy assessment and product comparison. First, a random sampling approach was employed to obtain validation sample data. This process resulted in four-phase validation samples (about 2000 samples for each) in 2000, 2010, 2015, and 2020, respectively. These validation sample points were put into the Google Earth and World Imagery Wayback platform, and interpreted by referring to the high-resolution images. Here the images of Phase-2000 and Phase-2010 are provided by Google Earth Pro, and the images of Phase-2015 and Phase-2020 are from September 20th, 2015 and 2020 of the World Imagery Wayback platform. Those sample points were marked according to the actual situation of the ground (marked 1 represents cropland, 0 represents non-cropland). Second, the function of “Extract Multi Values to Points” was used in ArcGIS to assign the cropland and non-cropland of each data here to the attributes of the sample point. Third, we constructed the confusion matrices of all datasets in the four years, that is, the number of products and interpretive sample points simultaneously identified cropland/cropland, cropland/non-cropland, non-cropland/cropland and non-cropland/non-cropland were counted (The confusion matrix is shown in supplementary data). The accuracy evaluation indexes were calculated according to the formula, and the comparison chart of the accuracy evaluation indexes of each data was drawn.

About spatial agreement comparison, in order to facilitate the comparison of these different land cover products (different coordinate systems and resolution), we first processed the global/national land cover datasets for the study area by using the boundary of Northeast China. Secondly, we carried out reprojection, resampling and reclassification processing successively in ArcGIS. The datasets were reprojected using the Albers Equal-Area projection. Using the nearest neighbor method, all data were resampled to 30 m. We combined all cropland subclasses of each dataset into one class to integrate cropland information. Here, the mosaic cropland (>50%) with other elements in some data (e.g., GLCNMO, CCI-LC, and GlobCover) are classified into cropland and the classification of cropland subcategory is shown in Supplementary Table S5. After this processing, the datasets were only reclassified into cropland and non-cropland. Finally, the “Raster Calculator” was used to obtain the agreement levels of consistency and discrepancy of cropland in the same spatial position of different land cover datasets, and the area proportion of different agreement levels was calculated.

In the comparison of cropland area in Section 3.4, we used pixel counting to calculate the cropland area of Northeast China under the projection of Albers, which is the pixel number of all cropland areas multiplied by the single pixel area.

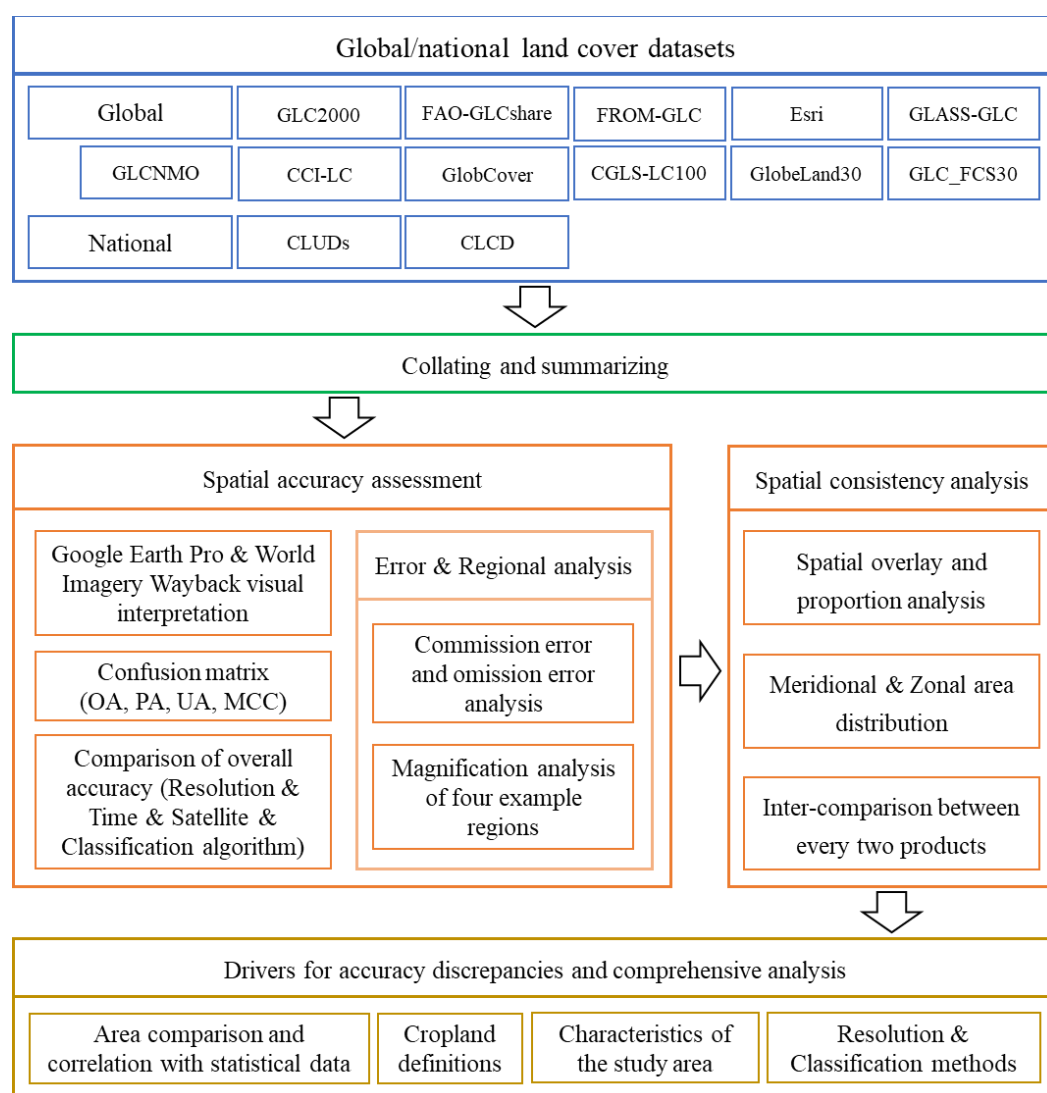


Figure 2. The workflow of this research. OA = overall accuracy, PA = producer's accuracy, UA = User's accuracy, MCC = the Matthews correlation coefficient.

2.4. Methodology

2.4.1 Accuracy assessment metrics

First, the accuracy assessment is performed by the error matrix or confusion, which is a crosstab of various land cover products and the ground truth sample point. It directly reflects the overall accuracy of the mapping [57-59]. Using the confusion matrix, four classical evaluation parameters are calculated, which are the user's accuracy (UA) (commission error), producer's accuracy (PA) (omission error), overall accuracy (OA), and the Matthews correlation coefficient (MCC), respectively. The UA, PA, and OA respectively refer to the percentages of the products classified correctly, the validation points classified correctly, and the correctly classified number. The Matthews correlation coefficient generates reliable results independently the ratios of positive and negative elements on the binary classification predictions [60].

The four evaluation metrics are the most commonly used methods in the accuracy assessment of land cover datasets [11,61-63]. They are defined as

$$\text{User's accuracy} = \frac{X_{ii}}{X_{i+}} \times 100\% \quad (1)$$

$$\text{Producer's accuracy} = \frac{X_{ii}}{X_{+i}} \times 100\% \quad (2)$$

$$\text{Overall accuracy} = \frac{\sum_i^2 X_{ii}}{N^2} \times 100\% \quad (3)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

where X_{ii} is the amount of class i pixels that were correctly classified, in this study it refers to the amount of cropland and non-cropland are correctly classified. N refers to the total number of the examined pixels, X_{+i} represents the number of class i pixels of the classification result, which is the total number of pixels of cropland in various land cover datasets; X_{+i} on behalf of the class i pixel number for the reference data, it is the total pixel number of cropland in the validation sample. TP (true positive) indicates the reference positive is correctly classified as positive; TN (true negative) indicates the reference negative is correctly classified as negative; FP (false positive) represents the reference negative is wrongly classified as positive; FN (false negative) represents the reference positive is wrongly classified as negative.

2.4.2 Inter-comparison method

The comparative assessments of various land cover datasets include the location and the quantity of each class [64]. In this study, 13 sets of LULC datasets were compared and analyzed in Northeast China from the two aspects of spatial location and cropland quantity. The comparison of location is usually a matter of spatial (dis)agreement pixel by pixel. The purpose of spatial agreement analysis is to explore the characteristics of the spatial distribution of cropland in various datasets. Spatial agreements include visual comparison and pixel-by-pixel comparison. Visual comparison can show the spatial variation of the datasets among regions subjectively. Per-pixel comparison is to calculate the proportion of pixels with the same attributes in the total pixels of various datasets in the same spatial location. The comparison of quantity mainly refers to areal comparison of cropland. We calculate the cropland area of Northeast China before reprojection and resampling.

2.4.3 Pairwise data and prefecture-level city cropland area validation

Correlation analysis can verify the relationship of the pairwise data and the ability of represent the prefecture-level cropland area. The following formulas (5), (6) and (7) are used for correlation analysis to detect the goodness of fit between datasets:

$$\Delta x_i = x_i - y_i \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \Delta x_i^2}{n}} \quad (6)$$

$$r^2 = \frac{(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

where Δx_i is the difference between the cropland area (x_i) of a data and another data (y_i), n is the total number of regions of aggregation. $\bar{x}$ and $\bar{y}$ are the average of cropland area.

3. Results

3.1. Accuracy evaluation of the thirteen datasets in the four phases

Figure 3 shows the distribution of sample points across Northeast China for four years. The distribution of ground-truth points in four periods is roughly similar in Figure 3, and the majority of cropland is distributed in the Sanjiang-Songnen-Liaohe Plain, with the main types of non-cropland being mountains, hills, lakes and rivers, grasslands, sandy and bare land, such as the Greater Khingan Mountains, the Changbai Mountains, and so on. Scattered cropland distribution is found at the transition areas of plains and hills, beside lakes and rivers, or in areas such as grassland or bare, the difference between cropland and non-cropland in the four phases were mainly concentrated in these areas.

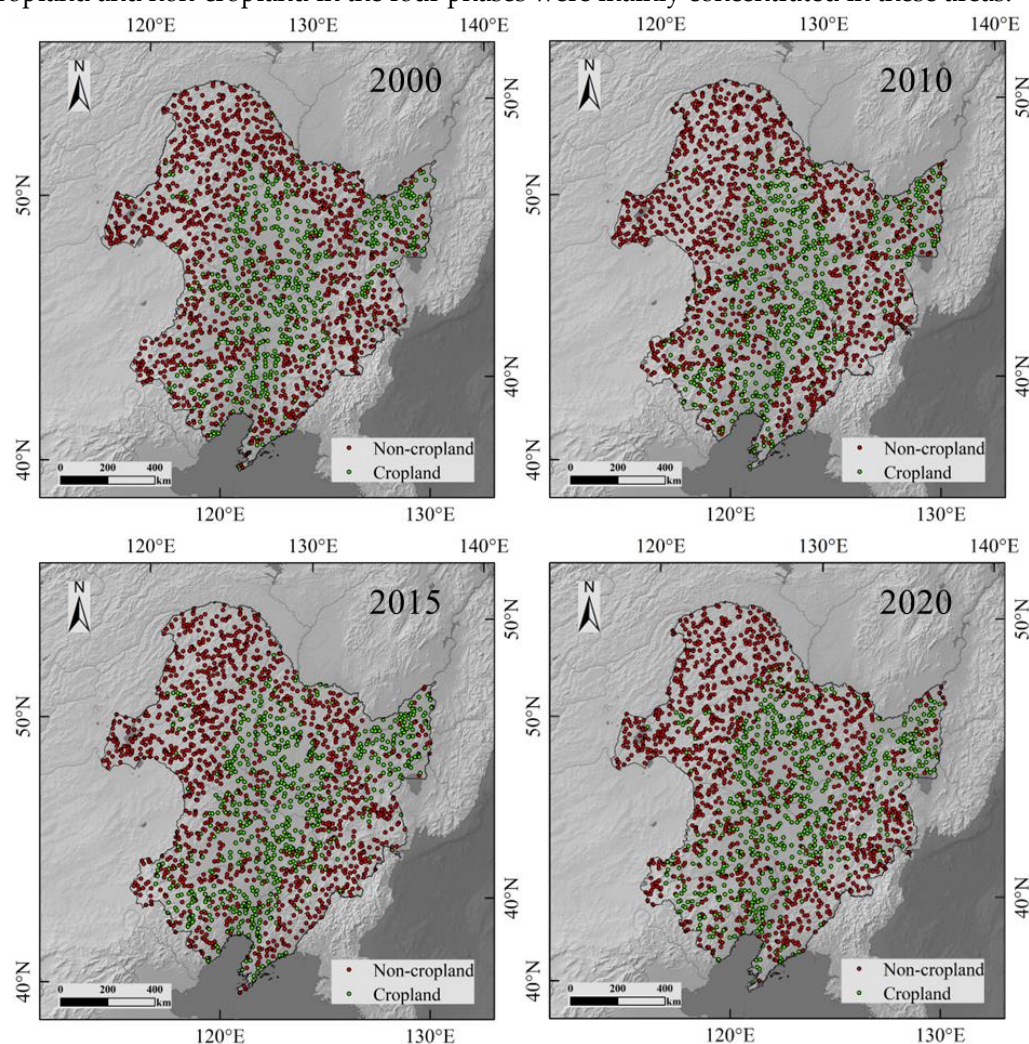


Figure 3. Distribution of validation samples across Northeast in Phases 2000, 2010, 2015 and 2020. The graduated red points represent true non-cropland, graduated green indicate real cropland.

Based on the ground truth validation samples, we constructed the confusion matrices of all datasets in four phases and obtained the spatial accuracy measures (i.e., producer's accuracy, user's accuracy, overall accuracy, and the Matthews correlation coefficient), they were computed from confusion matrices (Figure 4). In Phase-2020, the average overall accuracy of six datasets is the highest at 0.86, followed by Phase-2000, Phase-2010, and Phase-2015, all around 0.83 (Figure 13). CLUDs and CLCD data are available in all the four phases, while GLASS and CCI-LC data are available in three phases, 2000, 2010 and 2015. The following is a comparison of the data evaluation metrics by phase.

The overall accuracies of the eight datasets in Phase-2000 all exceed 0.73. GlobeLand30 has the highest OA of 0.90, followed by CLCD (0.89), CCI-LC (0.87). From the perspective of PA, ClobCover-2005 ranks the lowest with the accuracy of 0.57, followed by GLC2000 (0.66). As for UA, ClobCover-2005 also has the lowest value of 0.62,

followed by GLCNMO-2003 (0.63). Overall, GlobeLand30 has the balanced performance, with higher OA, UA and PA in 2000, followed by CLCD, and their MCC are all over 0.75. In Phase-2010, GlobCover-2005 has the lowest UA, PA and OA, followed by GlobCover-2009. CLCD has the highest UA, PA and OA, followed by Clobeland30. Other datasets have medium accuracy. The overall accuracies in the nine datasets of Phase-2015 are over 0.76, CLCD has the highest OA value of 0.90, followed by CCI-LC (0.87) and GLC_FCS30 (0.85). From the perspective of PA, GLASS ranks the lowest with 0.68. As for UA, GLCNMO-2013 has the lowest value of 0.63, followed by CGLS-LC100 (0.64) and FAO-GLCshare (0.68). In general, CLCD has the higher OA, UA and PA in Phase-2015, its MCC is over 0.77. Following it, CCI-LC and GLC_FCS30 also have good performances with MCC greater than 0.70. The overall accuracies in the six datasets of Phase-2020 are more than 0.70, the OA of CLCD is 0.914, followed by GlobeLand30 (0.906), GLC_FCS30 (0.902), and Esri (0.896). In terms of PA, CLUDs ranks the lowest with 0.69. This corresponds to a higher omission error, indicating that the estimation of cropland area across Northeast China is low (Figure 5 and Figure 6). As for UA, CGLS-LC100 has the lowest value of 0.57, and it is clear to observe that the cropland area in CGLS-LC100 data is overestimated than other data (Figure 5, Figure 6, Figure 8, and Figure 9). CLCD has the best performance in Phase-2020, followed by GlobeLand30, and all of their MCC are greater than 0.79. (Figure 4, Figure 5 and Figure 6). Following them are GLC_FCS30 and Esri, both of which have MCCs greater than 0.77. For detailed numerical information, please see Supplementary Table S2.

In summary, CLCD has the highest accuracy in four phases, followed by GlobeLand30, GLC_FCS30, and Esri. The low PAs values of GlobCover, GLC2000, CLUDs, GLASS and GLCNMO reflect a low estimation of the area in different regions of Northeast China. The UAs of CGLS-LC100, GlobCover, GLCNMO and FAO-GLCshare are all relatively low, all of which have different degrees of commission errors of cropland in different regions of the Northeast, and the cropland area is relatively overestimated than other datasets.

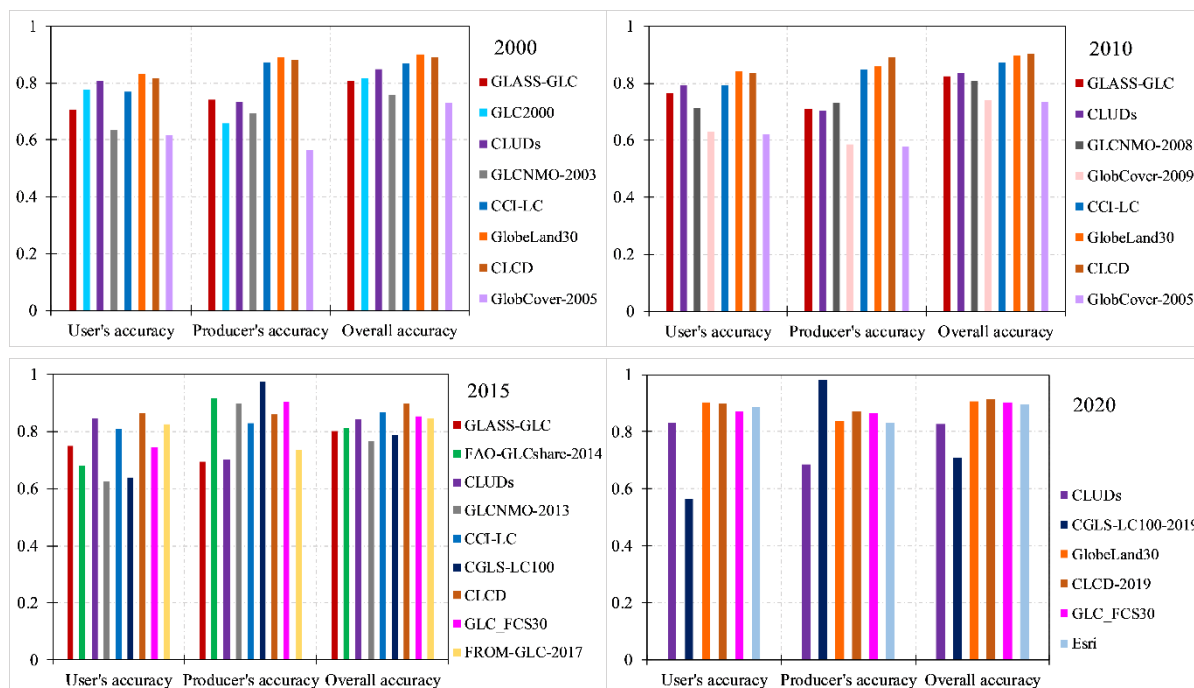
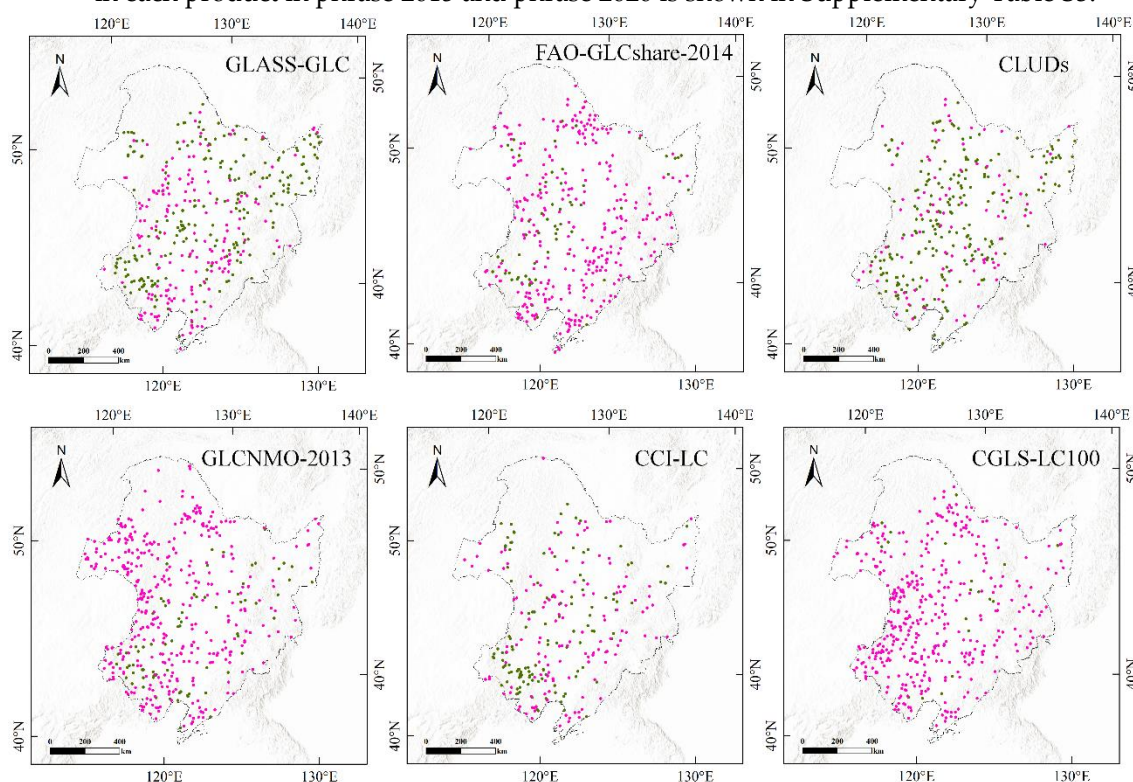


Figure 4. Comparisons of the spatial accuracy indexes of the datasets in 2000, 2010, 2015 and 2020.

3.2. Commission and omission error analysis on cropland in Northeast China

The Phase-2015 and Phase-2020 contain higher precisions and more products, covering almost all data products except for GLC2000 and GlobCover, thus we choose the two phases to further analyze the omission and commission error of these datasets. Based on ground verification points, the omission and commission error distribution of datasets in 2015 and 2020 across Northeast China is illustrates in Figure 5 and Figure 6. The ginger-pink points indicate the higher commission errors and the spruce-green points represent the higher omission issues.

Lots of ginger-pink points exist in CGLS-LC100 in Figure 5 and Figure 6, indicating serious commission issues and illustrating the overestimation of cropland area in both meridional and zonal dimensions (Figure 8 and Figure 9). Followed by GLCNMO and FAO-GLCshare, they are suffered from commission issues in different regions. GLCNMO-2013 has serious commission errors in western Inner Mongolia, especially in the western part of Hulunbuir City, where the overestimation is more serious, followed by Heihe City in Heilongjiang Province. Chifeng City in Inner Mongolia has a lower omission phenomenon. FAO-GLCshare has high misclassification in the north of Heihe City in Heilongjiang, the central part of Hulunbuir City, as well as the junction of mountains and plains (the Greater Khingan Mountains, the Changbai Mountains, etc.), while it has low leakage in Greater Khingan Mountains. GLASS has many spruce-green points and it has less cropland area than other datasets (Figure 8 a, Figure 5, Figure 6). CLUDs have omission issues in the Sanjiang-Songnen-Liaohe Plain and Chifeng of Inner Mongolia with many spruce-green points. It is clear that CGLS-LC100 and GLCNMO-2013 have lower omission errors with fewer spruce-green points, but they suffer from a little more commission issue, CLUDs is higher omission errors with more spruce-green points (Figure 5 and Figure 6). Furthermore, CLCD has the fewest spruce-green and ginger-pink points, indicating that it has few omission and commission errors, which corresponds to the highest accuracy evaluation in Figure 4. The numbers of the commission and omission error in each product in phrase 2015 and phrase 2020 is shown in Supplementary Table S3.



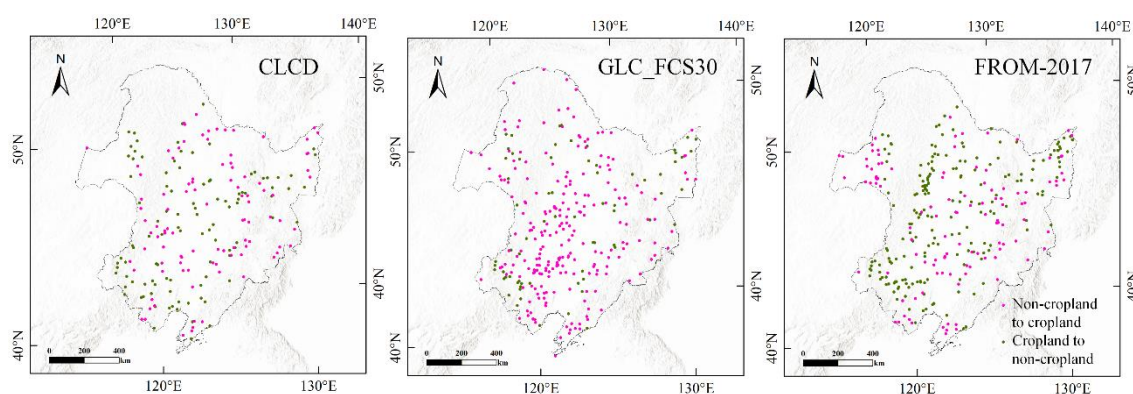


Figure 5 The commission and omission error distribution of nine data in 2015.

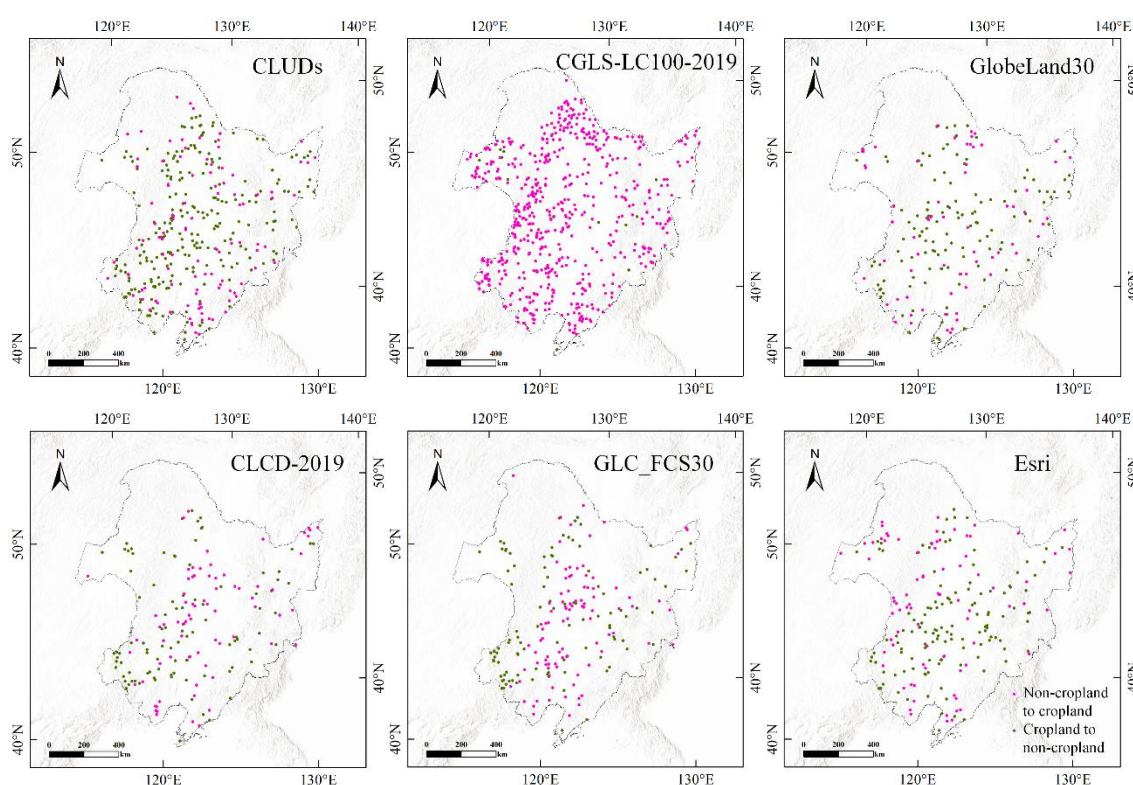


Figure 6 The commission and omission error distribution of six data in 2020. The disagreements distribution of verified points and the corresponding position in each dataset. In Figure 5 and Figure 6, the ginger-pink dots represent that true pixels of non-cropland are classified as cropland (representing commission error). The spruce-green dots indicate that true pixels of cropland are classified as non-cropland (representing omission error).

To better understand the spatial accuracies of these datasets in more detail, we select four regions for Phase-2015 (as it contains the most data products) for zoom-in exemplification to better observe the varied accuracies, as shown in Figure 7. It can be seen that except GLASS, the other eight datasets could capture the distribution of cropland in the Sanjiang and Songnen Plain. GLASS not only has low resolution, but also suffers from omission and commission issue. In the southeast of Hulunbuir city of Inner Mongolia (Figure 7 B), FROM-GLC and GLASS have a large number of missing, and fail to detect the cropland, followed by FAO-GLCshare and CLUDs. CLCD and GLC_FCS30 have a good description of mountains and farmlands, and CGLS-LC100 has a significantly over-estimation of cropland area. Five datasets perform well in the Liaohe Plain except FAO, GLASS, GLCNMO, and CLUDs which suffer from a commission issue (Figure 7 D).

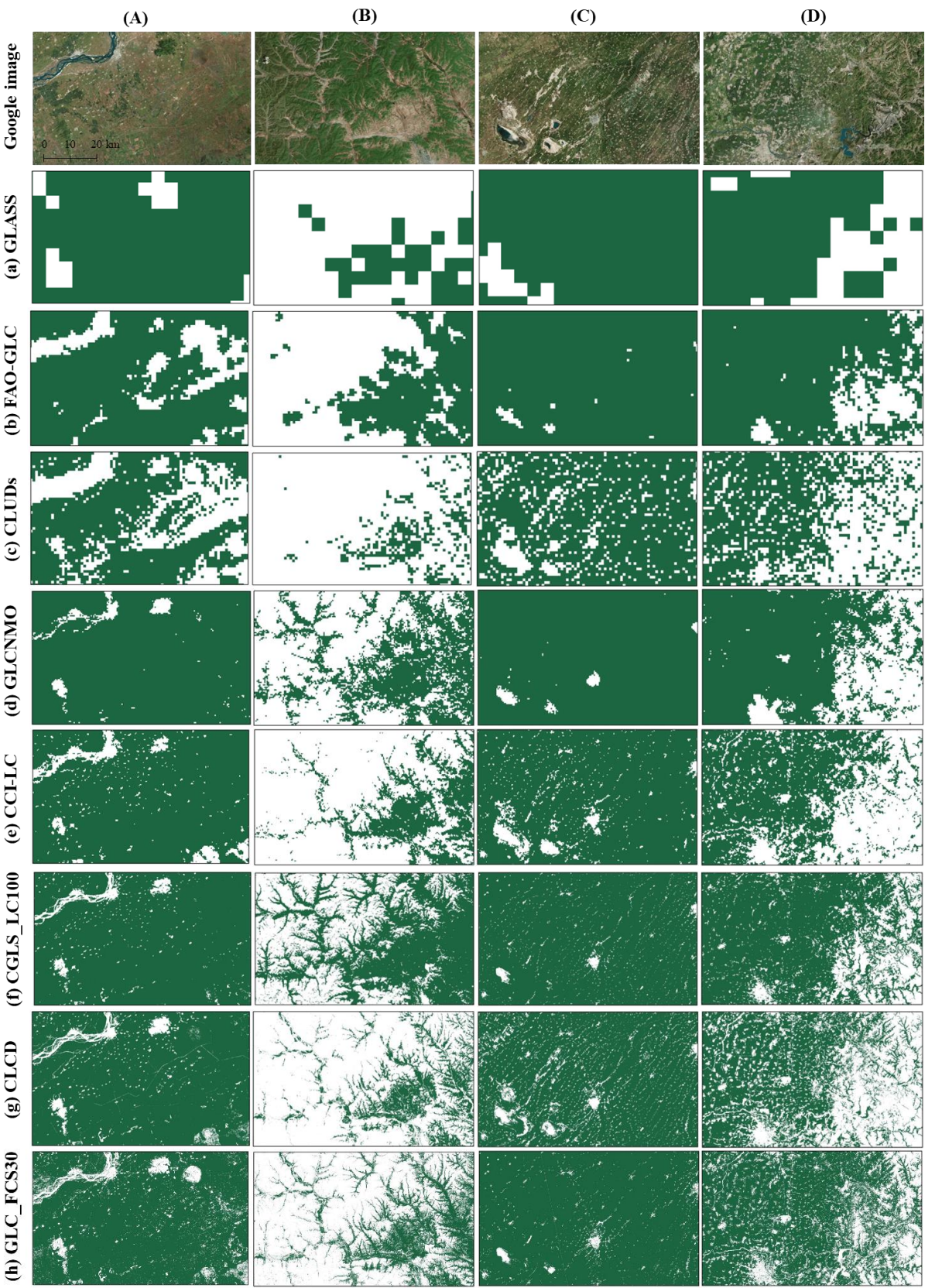


Figure 7. Demonstration of the cropland magnified performances of four regions in different datasets. Dark-green indicates cropland and white indicates non-cropland. The positions of A-D are shown in Fig. 8 b.

3.3. Spatial agreement and discrepancies

The spatial agreement is shown in Figure 8 and Figure 9, which illustrates spatial similarities and differences of cropland. In terms of longitudinal cropland distribution curve, most croplands of all datasets are clustered between 125°E and 130°E. In summary, the region with agreement level 9 or 6 (highest spatial consistency) is mainly distributed in traditional agricultural regions of Northeast China, such as the Sanjiang-Songnen-Liaohe Plain, which include Harbin, Hegang, Jixi, Jiamusi, Qiqihar, and Suihua in Heilongjiang. Liaoyuan City, Siping City, Song yuan City, Changchun City in Jilin Province, and Fuxin City, Jinzhou City, Shenyang City, and Tieling City in Liaoning Province. All of these zones have homogeneous and clustered cropland and can be easily captured by remote sensing imagery. The regions of lower spatial consistency (agreement level ≤ 2) are distributed in mountainous, hilly zones and transition areas between them in the east of Inner Mongolia, Jilin, and Liaoning, and the north of Heilongjiang.

The overlapping result of nine datasets in Phase-2015 is shown in Figure 8 (a-c), the area identified as cropland by the nine datasets accounts for 19.7% of the total cropland area. From the meridional area curve, we can see that CGLS-LC100 has the highest area curve between 121°E and 128°E, FAO-GLCshare has the highest area curve between 128°E and 135°E compared with other datasets. FROM-2017 at around 121-125°E, CLUDs at about 126-129°E, and GLASS at 130-132°E are the lowest. From the perspective of zonal area distribution, FAO-GLCshare, CGLS-LC100 and GLCNMO-2013 have the highest area curves in the range of about from 40°N to 42°N, from 42°N to 48°N, and from 48°N to 51°N.

Figure 9 (a-c) shows the overlapping result of six datasets in Phase-2020. The stacked bar shows that 28.5% are simultaneously labeled as cropland in the total identified cropland regions, and 42.2% of the data are identified as cropland in agreement level five or higher. According to the distribution of meridional cropland area, the area curve of CGLS-LC100 is the highest in all longitudes, this indicates that CGLS-LC100 has a high estimation value of cropland area. CLUDs ranks lowest between 123°E and 129°E, revealing low area estimation within this longitude range. In the zonal distribution curve of cropland area, CGLS-LC100 also showed the highest regional distribution among the six curves. Esri has the lowest area around 45°N, corresponding to the omission phenomenon in Songyuan and Changchun city of Jilin Province. Supplementary Table S4 of supplementary data is the area proportion of different consistency levels.

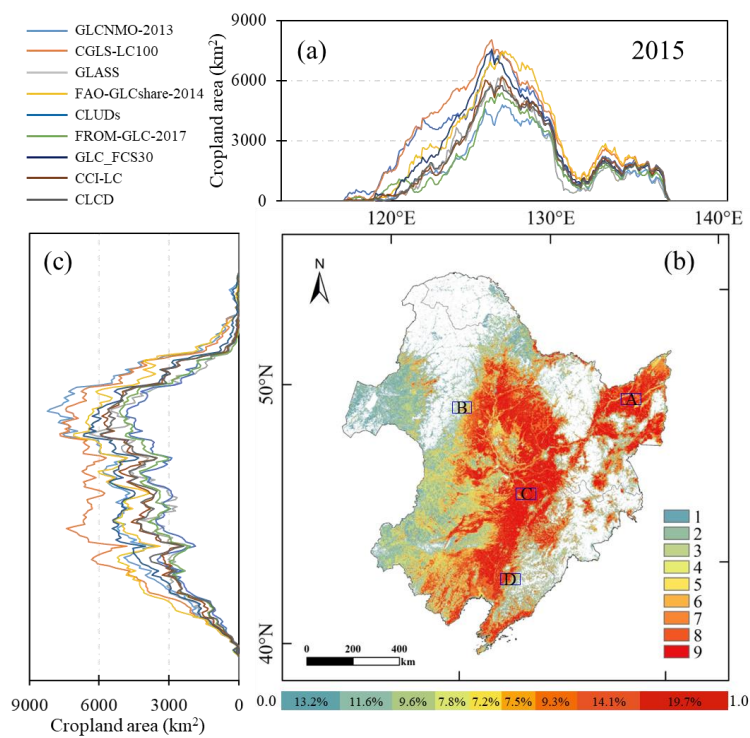


Figure 8. Illustration of the spatial agreement level, cropland area curves of meridional and zonal of nine datasets in 2015. Charts (a) and (c) are the area curves of the meridional and zonal of datasets, respectively. The overlapping result of datasets is shown in subplot (b), representing the resampled evaluated data at a resolution of 30 m, and the digital numbers indicate different consistency levels. The stacked bar chart is the cropland area proportion (%) in the agreement levels.

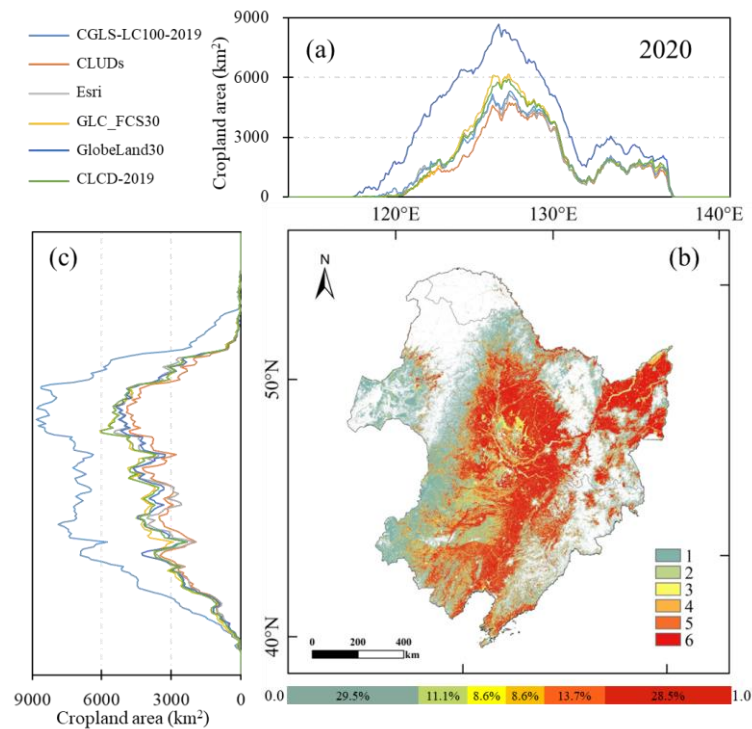
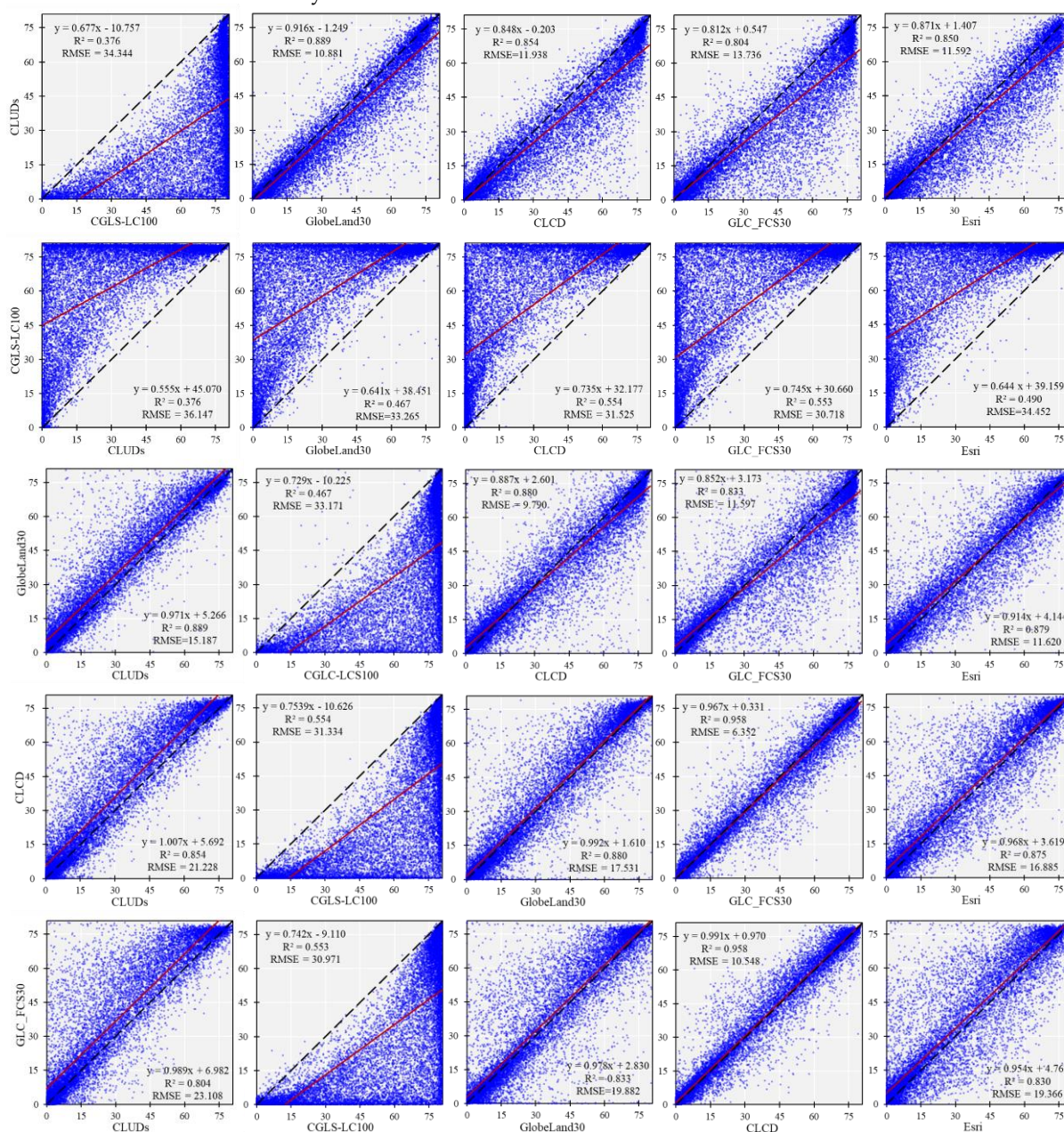


Figure 9. Illustration of the spatial agreement level, cropland area curves of meridional and zonal of six datasets in 2020. Charts (a) and (c) are the area curves of the meridional and zonal of datasets, respectively. The overlapping result of datasets is shown in subplot (b), representing the resampled

evaluated data at a resolution of 30 m, and the digital numbers indicate different consistency levels. The stacked bar chart is the cropland area proportion (%) in the agreement levels.

Because the overall accuracy of six data in Phase-2020 is higher, we further explore their agreement relationships with each other. We compared the pairwise data of six data and their scatterplots is shown in Figure 10. CLCD has the highest agreement with GLC_FCS30 ($r^2 = 0.96$), followed by GlobeLand30 with CLUDs ($r^2 = 0.89$), GlobeLand30 and CLCD ($r^2 = 0.88$), all their RMSEs are less than 12 km². The CLUDs have the lowest agreement with CGLS_LC100, with the lowest r^2 of 0.38 and the highest RMSE of 36.15 km². CGLS-LC100 is in poor agreement with any data, and we can see that almost all points are above or below the 1:1 line. This is because the cropland area in CGLS-LC100 data is relatively overestimated than other data.



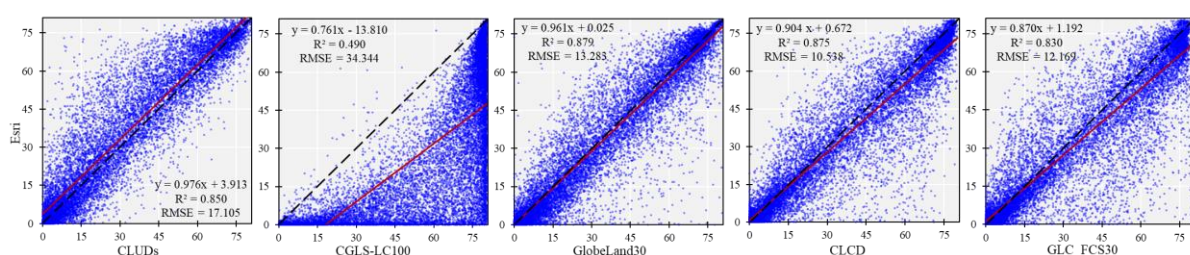


Figure 10. Scatterplots between the CLCD and the CGLS-LC100, CLUDs, Esri, GLC_FCS30, GlobeLand30. The axes represent the cropland area aggregation within the grid cell of 8.438 km×9.537 km across Northeast China of six datasets, which only includes the comparison of cropland area. The unit of RMSE is km².

3.4. Comparative analysis by referring to the statistical data

Figure 11 compares the cropland area from collected datasets to statistical data in Northeast China. CLUDs-1980s slightly underestimates cropland area, whereas CGLS-LC100 and GLCNMO-2013 significantly overestimate cropland area in Northeast China. Compared with the reconstructed cropland data, the cropland area value of CLUDs, GlobCover, and GlobeLand30 products are relatively closer to the statistics. The area of GLASS, FAO-GLCshare, CCI-LC, CLCD, GLC_FCS30, FROM-GLC and Esri are concentrated around 600×10³ km². The three-year areas of GLCNMO-2003, 2008 and 2013 are significantly different, and all products in the statistics of cropland area increase with time. On the whole, the cropland areas based on remote sensing are higher than the statistical data. The differences may be caused by inconsistent spatial resolutions of remote sensing products and official statistics. The remote sensing mapping probably tends to induce biases in the mixed pixels while the official yearbook statistical data was generated by field surveying, multi-level estimates, and census.

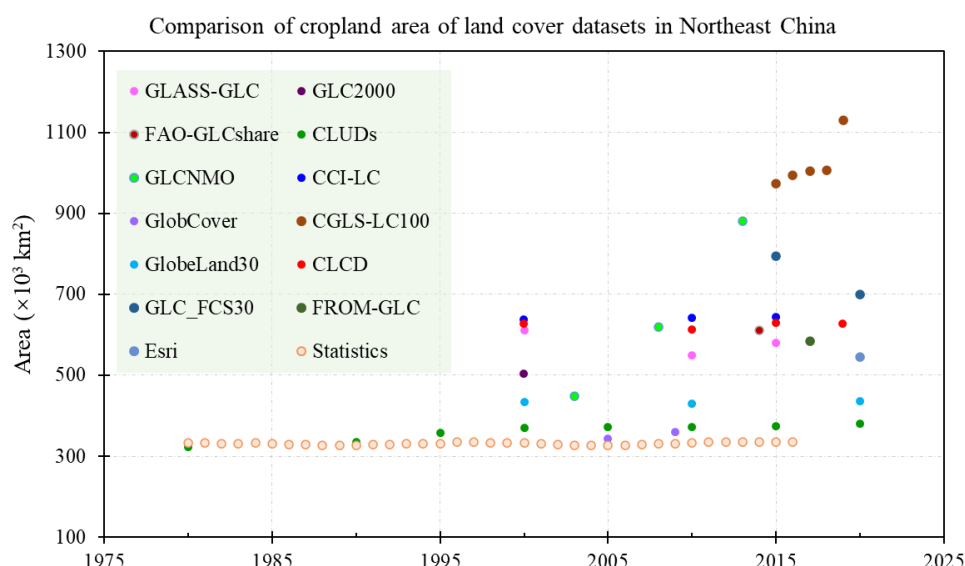


Figure 11. Comparison of the cropland area of all examined datasets with the statistic result by Yu et al. [56].

Based on the Phase-2015 datasets, we explored the correlation between datasets and the reconstructed data on the cropland area of aggregated in prefecture-level (Figure 12). We find that the r^2 of CLUDs is the highest (0.97) and the RMSE is the lowest (1.63 km²), followed by CLCD (0.95, 4.04 km²), and CCI-LC (0.96, 4.10 km²). CGLS-LC100 has the lowest r^2 of 0.53 and the highest RMSE of 15.96 km², besides, the fitting effect of FROM-GLC-2017 is also better. Compared with other datasets, CLUDs performs best in regional

accuracy (prefecture-level). FROM-GLC-2017 and GLASS also have good manifestation with most the “plus” points centralized around the 1:1 line. On the contrary, CGLS-LC100 and GLCNMO perform poorly, and they suffer from obvious overestimation on prefecture-level cropland area.

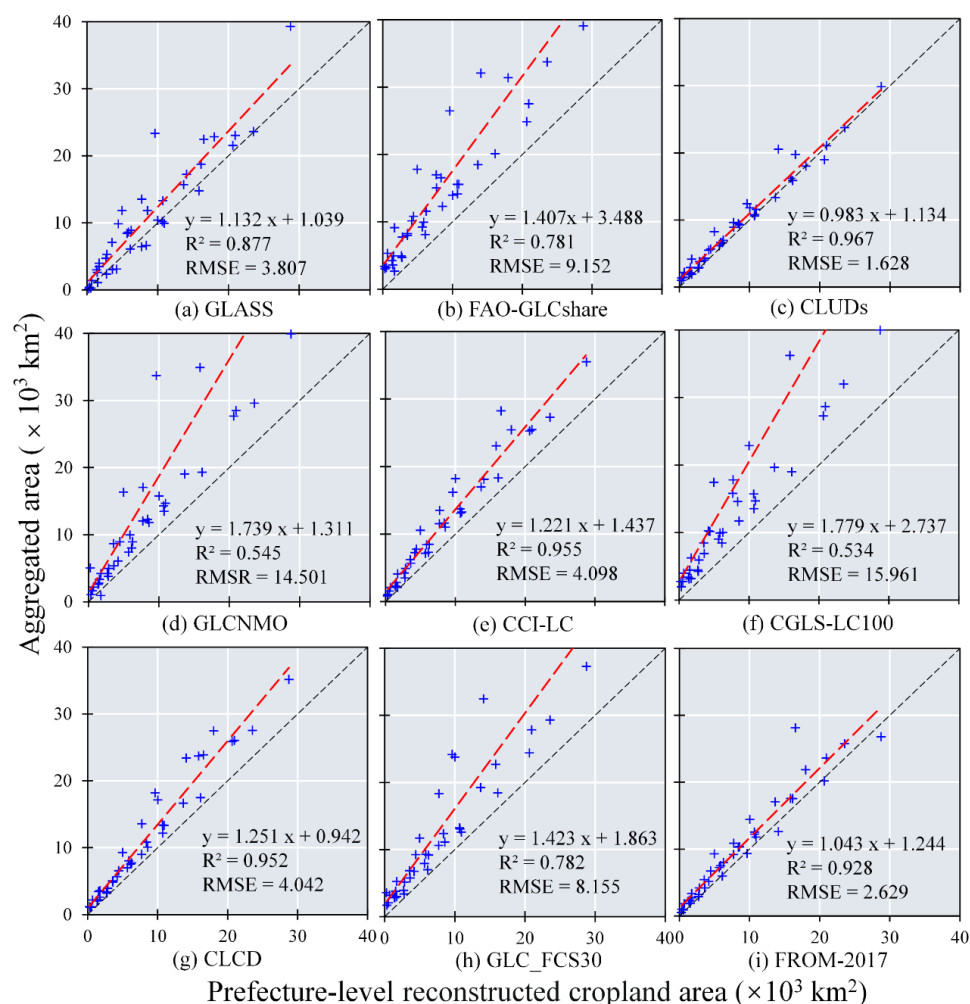


Figure 12. Scatterplots of the prefecture-level city reconstructed cropland area vs. the aggregated area of cropland from each dataset. The unit of RMSE is km^2 .

4. Discussion

4.1. Potential influencing factors of the data accuracy

Multiple factors might have resulted in the difference and uncertainties in the cropland range among these datasets. We grouped and compared the overall accuracy of these datasets mainly from four dimensions, including the data resolution, producing time, satellite sensors, and classification algorithms. The results are illustrated in Figure 13. It can be found the overall accuracy rises with the improvement of resolution and the development of mapping time in general. For example, GLASS with 5 km resolution has the lowest overall accuracy, and Esri with 10 m resolution has the highest overall accuracy. The average of the overall accuracy of the products derived from Sentinel and Landsat observations are overall better than the data produced by other satellite sensors, while the data produced by the PROBA satellite is the worst. In addition, different classification strategies and mapping methods adopted by these datasets may lead to difference in classification accuracies. The average overall accuracy of the data produced by random forest

classification method, Pixel-Object-Knowledge classification approaches and deep learning are all over 0.89, followed by the operational SPECLib-based approach & random forest models (0.88), the unsupervised spatio-temporal clustering & machine learning (0.87), and extraction of remote sensing information (0.84).

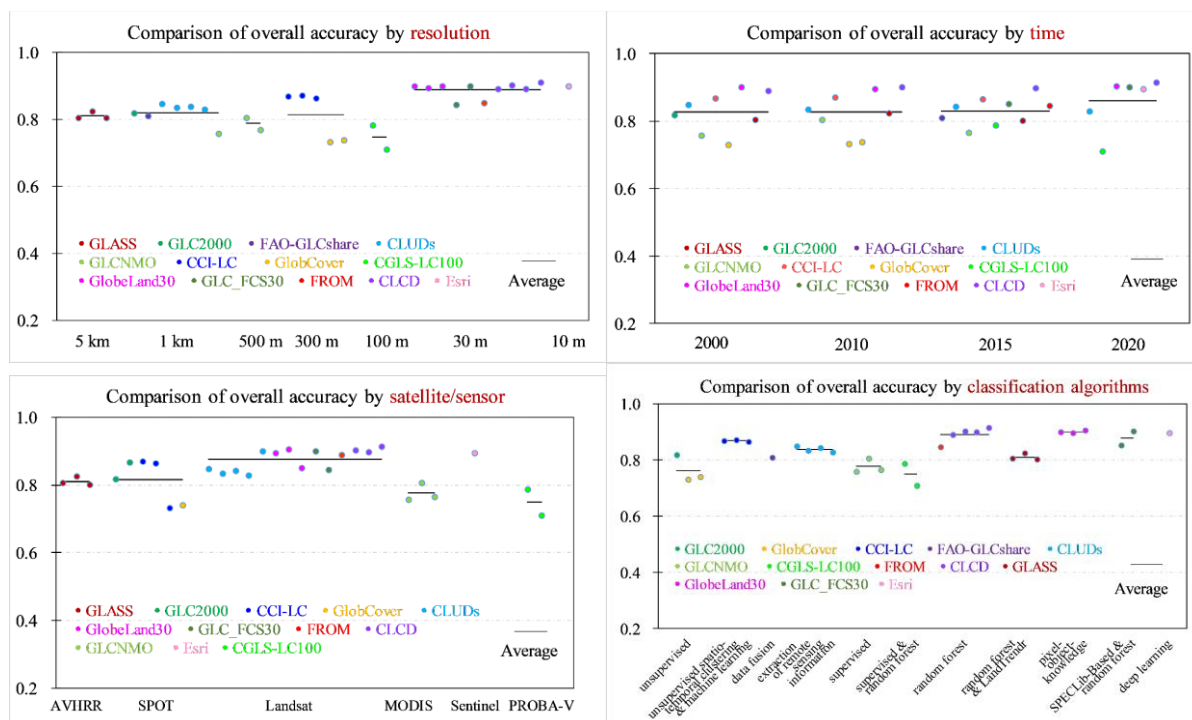


Figure 13. Comparison of the overall accuracy of different datasets grouping by data resolution, producing time, sensor, and classification algorithm, with different individual colors markers for different land cover datasets.

4.2. Other possible influencing factors of the dataset performance

The diverse definitions and classification of cropland in these evaluated datasets can also introduce the difference of classification performance. For example, in GLC2000, CGLS-LC100 and Esri, the croplands are “pure” cropland, including rainfed croplands and flooded crops, such as cereal, corn, wheat, rice. The croplands in GLASS, GlobeLand30, CLCD, GLC_FCS30 and FROM-GLC mainly refer to rainfed/irrigated cropland, orchard and temporally bare farmland, etc. However, in GLCNMO, the cropland includes herbaceous crop(s), the paddy field of graminoid crops and non-graminoid crops, and the mosaic of cropland and other vegetation. In CCI-LC, the cropland incorporates rainfed/irrigated cropland, herbaceous, tree or shrub cover, and the mosaic of cropland and natural vegetation (Supplementary Table S5). In general, the mixture of herbaceous or shrubs and cropland and the narrow definition inevitably result in some overestimations and a little underestimation of cropland area, respectively. The cropland definition of GLC2000, CGLS-LC100 and Esri are “pure” cropland, and thus their estimation of cropland area should be somewhat underestimated. However, we found that the cropland area of CGLS-LC100 is very high, while GLC2000 and Esri are relatively normal. It could be attributed to some other reasons (we suspect it may be because the cropland in CGLS-LC100 is an independent thematic cropland file rather than a single land cover layer). In addition to that, the cropland of GLASS, GLC2000, CLUDs, GlobeLand30, CLCD, and GLC_FCS30 contains several different subcategories, while FAO-GLCshare and CGLS-

LC100 are thematic files of cropland, which only considered cropland and non-cropland. We found that both FAO and CGLS-LC100 have high cropland area.

Additionally, there are mountains, hills, plains, rivers, wetlands and transition zones in Northeast China. This complicated landscape condition is also one of the reasons leading to the difference and uncertainties among these datasets. It should be stated that, in this work, our validation samples simply evaluated the Northeast China cropland and our conclusions on the accuracy evaluation of these datasets may be not representative for other land cover classes or other regions accuracy assessments, more representative truth points of ground need to be collected for quantitative implementation in other regions in the future.

4.3. Uncertainties existing in our evaluation work

One potential source of error in our accuracy assessment is the use of validation samples from the same phase year to evaluate datasets from different years, particularly when land cover changes occur in specific regions. In this study, validation samples of four phase years were generated through visual interpretation of high-resolution images. For example, samples collected in the same phase year of 2010 can be used for Phase-2010, GLCNMO-2008, GlobCover-2009, and GlobCover-2005. Using validation samples from the same phase year (2010) to evaluate datasets from different years (2008/2009/2005) may be problematic if land cover changes occur in some regions, reducing the reliability of commission/omission errors. However, this situation occurred primarily in a few datasets, including CLCNMO-2003 and GlobCover-2005 in Phase-2000, CLCNMO-2008, GlobCover-2005 and GlobCover-2009 in Phase-2010, FAO-GLCshare-2014 and FROM-GLC-2017 in Phase-2015, and CLCD-2019 in Phase-2020. While the land cover type in some traditional cropland regions will not change in the short term, some changes will occur at the intersection of land cover types, such as cropland and grassland, shrub or bare land. The mismatch between the period of these verification samples and the data leads to commission/omission errors, and the extent of this error cannot be quantifiable, which is a shortcoming point of our research.

Another issue that caused the evaluation uncertainty is the cropland subcategories mosaic. In terms of the mosaic cropland found in some land cover datasets, the cropland mosaic which accounts for more than 50% is directly divided into cropland in the data processing section. Such issue is mainly related to three products GLCNMO, CCI-LC and GlobCover. We calculated the proportion of mosaic cropland pixels in the whole cropland pixel. Mosaic cropland accounted for 9.6% in GCLNMO-2003, 4.3% in GCLNMO-2008, and 14.5% in GCLNMO-2013. In CCI-LC, the proportion of these mosaics was 7.2% in CCI-LC-000, 7.0% in CCI-LC-2010, and 7.0% in CCI-LC-2015, respectively. However, in the GlobCover, this proportion is relatively large, 35.5% in GlobCover-2005 and 39.8% in GlobCover-2009. We looked up the definition of the mosaic class in GlobCover, which is "mosaic cropland (50-70%)/ vegetation (grassland/ shrubland/ forest) (20-50%)". Therefore, in these 35.5% or 39.8% of cropland mosaic pixels, the proportion of non-cropland is pretty small (e.g., 8.0-19.9% in GlobCover-2009). However, there is no doubt that some omission and commission errors are also caused by these cropland mosaic pixels.

4.4. Comparison with the prior assessment work

In the recent study by Zhang et al. [65], a comparative analysis and accuracy assessment of six products with 30 m resolution of China in 2015 were conducted, where GLAD and GFSAD are thematic cropland data. Four accuracy indexes were used, including the overall accuracy, commission error, omission error, and Matthews correlation coefficient. We collected 13 sets of data for our study and conducted a comparative analysis and assessment of the cropland in Northeast China for the four phases around the years 2000, 2010, 2015, and 2020. During the evaluation process, we recognized that the years are not uniform, the data resolutions are different, and all data are land cover products and not

thematic cropland products. The accuracy evaluation indexes we used include overall accuracy, producer's accuracy, user's accuracy, and the Matthews correlation coefficient. Through our methodology, we aimed to provide valuable insights into the cropland changes in Northeast China and contribute to the scientific community's understanding of land use dynamics.

Both studies compare and analyze the spatial consistency of data at regional (China, Northeast China), provincial, meridional and zonal scales. However, Zhang et al' study goes into more detail on provincial agreements and differences, not only calculating the proportion of cropland area at different agreement levels in each province but also the relative area difference between each data and the statistics for each province. An important difference between the two papers is that in the discussion section, Zhang et al found the relationship between elevation and slope, and concluded that the area proportion with high agreement decreases with an increase in elevation and slope. When the elevation and slope increase, the data difference is obvious. Our study did not investigate either aspect, which should be strengthened in the next work.

5. Conclusions

As more and more global or national land cover datasets are published and available, it is becoming easier to obtain cropland information from these open-access products. However, the accuracy evaluation results of these datasets at a global/national scale for all LULC elements (classes) may not be replaceable for local-area or class-specific accuracy. Northeast China is a major agricultural region of China and has complicated landscapes and varied terrains. This paper takes Northeast China to evaluate the performances of our collected thirteen sets of land cover datasets in the accuracy, spatial location, and area of croplands. The similarities and differences are shown in a multi-scale comparison between datasets.

Accuracy validation based on ground truth samples indicates that, in general, compared with other datasets, CLCD, GlobeLand30, GLC_FCS30, and Esri have the best performance in cropland classification with overall accuracies of more than 0.89. On the contrary, the OAs of CGLS-LC100-2019 and GlobCover are lower than 0.75. For commission and omission error, CGLS-LC100 has a severe overestimation, while GLASS suffers from a significantly underestimation of cropland area, and the bias situation of CLCD is the least. In the aspect of spatial consistency and difference, the traditional agricultural regions such as Sanjiang-Songnen-Liaohe Plain in Northeast China have the highest spatial consistency. Lower consistency regions are distributed in mountainous, hilly regions and the transition areas between them. For the scatter plots between pairwise data, the agreement between CLCD and GLC_FCS30 is the highest ($r^2 = 0.96$), followed by GlobeLand30 with CLUDs (0.89), and GlobeLand30 with CLCD (0.88), all their RMSEs are less than 12 km². CGLS-LC100 is in poor agreement with any data, with r^2 lower than 0.56 and RMSE higher than 30 km².

In exploring the influence factors of the difference and uncertainties in the cropland range and location among all datasets, although the purpose of these datasets was to provide accurate cultivate information, factors such as the definitions of cropland, the landscape characteristics of Northeast China, classification scheme, and different satellite sensors, jointly produced different mapping results. It is arbitrary to say that this dataset is more suitable for this particular application, however, with the increasing demand for high-precision cropland datasets, users should consider the accuracy of cropland class when selecting a dataset. The accuracy evaluations and comprehensive comparison analysis of this work are expected to provide a meaningful and valuable reference for those (e.g., cropland surveyors) who use these datasets in Northeast China. In addition, the detailed evaluation of this work could provide feedback to data generators, facilitate the improvement of data processing algorithms and classification techniques, better service with future cropland applications and mapping or cropland data fusion.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, C.S. and T.C.; methodology, P.C., T.C., and K.L.; software, P.C., and D.Z.; validation, P.C., T.C., D.Z. and K.L.; formal analysis, P.C., T.C., C.S. and K.L.; investigation, D.Z. and P.C.; resources, P.C. and D.Z.; data curation, P.C., T.C. and D.Z.; writing—original draft preparation, C.S., T.C., P.C. and D.Z.; writing—supervision, C.S., T.C., K.L.; visualization, P.C., T.C. and D.Z.; project administration, C.S. and T.C.; funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA28020503, XDA23100102), the Jiangsu Normal University Postgraduate Research & Practice Innovation Program (2021XKT0094), the National Key Research and Development Program of China (2019YFA0607101), and the National Natural Science Foundation of China (41971403, 41930102, 42171421).

Data Availability Statement: The GLC-2000 can be downloaded from the web <https://forobs.jrc.ec.europa.eu/products/glc2000/products.php>. The FAO-GLCshare is available at FAO Map Catalog. The FROM-GLC is accessible on the Tsinghua website at Finer Resolution Observation and Monitoring - Global Land Cover (tsinghua.edu.cn). The Esri can be downloaded from the web at <https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2>. The GLASS-GLC is available at <https://doi.org/10.1594/PANGAEA.913496>. The GLCNMO is available at https://github.com/globalmaps/gm_lc_v1. The CCI-LC is available at <http://maps.elie.ucl.ac.be/CCI/viewer/download.php>. The GlobCover can be downloaded from European space agency data user element at http://due.esrin.esa.int/page_globcover.php. The CGLS-LC100 can be downloaded from the web at <https://zenodo.org/record/3939038#.Y3TzSMdBxBM>. The GlobeLand30 are online accessible at <https://www.webmap.cn/commres.do?method=globeIndex>. The GLC_FCS30 data is available at <https://zenodo.org/record/3986872#.Y3Ty2sdBxBN>. The CLUDs is available on request from the corresponding author. The CLCD is freely available at <https://doi.org/10.5281/zenodo.4417810>. The auxiliary datasets of China cropland are available at <https://doi.org/10.6084/m9.figshare.13356680.v1>.

Acknowledgments: We would like to thank the editor and the anonymous reviewer, whose constructive comments will help to improve the presentation of this paper.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Pinstrup-Andersen, P. Perspectives in World Food and Agriculture 2004. 2008; pp. 87-97.
- Gibbs, H.K.; Ruesch, A.S.; Achard, F.; Clayton, M.K.; Holmgren, P.; Ramankutty, N.; Foley, J.A. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc Natl Acad Sci U S A* 2010, 107, 16732-16737, doi:10.1073/pnas.0910275107.
- Grekousis, G.; Mountrakis, G.; Kavouras, M. An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing* 2015, 36, 5309-5335, doi:10.1080/01431161.2015.1093195.
- Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 2010, 21, 1303-1330, doi:10.1080/014311600210191.
- Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* 2010, 21, 1331-1364, doi:10.1080/014311600210209.
- Bartholomé, E.; Belward, A. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* 2005, 26, 1959-1977, doi:10.1080/01431160412331291297.
- Friedl, M.A.; Sulla-Menashé, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 2010, 114, 168-182, doi:10.1016/j.rse.2009.08.016.
- Bicheron, P.; Leroy, M.; Brockmann, C.; Krämer, U.; Miras, B.; Huc, M.; Niño, F.; Defourny, P.; Vancutsem, C.; Arino, O.; et al. Globcover: a 300 m global land cover product for 2005 using ENVISAT MERIS time series. *Proceeding of the Second International Symposium on Recent Advances in Quantitative Remote Sensing* 2006, 538-542.
- Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 2015, 103, 7-27, doi:10.1016/j.isprsjprs.2014.09.002.
- Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* 2012, 34, 2607-2654, doi:10.1080/01431161.2012.748992.
- Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Schmullius, C. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment* 2008, 112, 2538-2556, doi:10.1016/j.rse.2007.11.013.
- Kaptué Tchuenté, A.T.; Roujean, J.-L.; De Jong, S.M. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation* 2011, 13, 207-219, doi:10.1016/j.jag.2010.11.005.
- Wu, W.; Shibasaki, R.; Yang, P.; Zhou, Q.; Tang, H. Remotely sensed estimation of cropland in China: a comparison of the maps derived from four global land cover datasets. *Canadian Journal of Remote Sensing* 2014, 34, 467-479, doi:10.5589/m08-059.
- Congalton, R.; Gu, J.; Yadav, K.; Thenkabail, P.; Ozdogan, M. Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sensing* 2014, 6, 12070-12093, doi:10.3390/rs61212070.
- Lu, M.; Wu, W.; Zhang, L.; Liao, A.; Peng, S.; Tang, H. A comparative analysis of five global cropland datasets in China. *Science China Earth Sciences* 2016, 59, 2307-2317, doi:10.1007/s11430-016-5327-3.
- Deines, J.M.; Patel, R.; Liang, S.-Z.; Dado, W.; Lobell, D.B. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment* 2021, 253, doi:10.1016/j.rse.2020.112174.
- You, N.; Dong, J. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing* 2020, 161, 109-123, doi:10.1016/j.isprsjprs.2020.01.001.
- Giri, C.; Zhu, Z.; Reed, B. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment* 2005, 94, 123-132, doi:10.1016/j.rse.2004.09.005.
- McCallum, I.; Obersteiner, M.; Nilsson, S.; Shvidenko, A. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation* 2006, 8, 246-255, doi:10.1016/j.jag.2005.12.002.
- Liu, L.; Zhang, X.; Gao, Y.; Chen, X.; Shuai, X.; Mi, J. Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects. *Journal of Remote Sensing* 2021, 2021, 1-38, doi:10.34133/2021/5289697.
- Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data* 2021, 13, 2753-2776, doi:10.5194/essd-13-2753-2021.
- Hansen, M.C.; Reed, B. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing* 2010, 21, 1365-1373, doi:10.1080/014311600210218.
- Tsendbazar, N.-E.; de Bruin, S.; Fritz, S.; Herold, M. Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sensing* 2015, 7, 15804-15821, doi:10.3390/rs71215804.
- Bai, Y.; Feng, M.; Jiang, H.; Wang, J.; Zhu, Y.; Liu, Y. Assessing Consistency of Five Global Land Cover Data Sets in China. *Remote Sensing* 2014, 6, 8739-8759, doi:10.3390/rs6098739.
- Ning, J.; Zhang, S.; Cai, H.; Bu, K. A Comparative Analysis of the MODIS Land Cover Data Sets and Globcover Land Cover Data Sets in Heilongjiang Basin. *Journal of Geo-information Science* 2012, 14, 240-249.

26. Liu, Y.; Zhou, M. Comparative Analysis on Three Land Cover Datasets based on IGBP Classification System over Hanjiang River Basin. *Remote Sensing Technology and Application* 2017, 32, 575-584.
27. Yang, Y.; Xiao, P.; Feng, X.; Li, H.; Chang, X.; Feng, W. Comparison and assessment of large-scale land cover datasets in China and adjacent regions. *National Remote Sensing Bulletin* 2014, 18, 453-475.
28. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy assessment of seven global land cover datasets over China. *ISPRS Journal of Photogrammetry and Remote Sensing* 2017, 125, 156-173, doi:<https://doi.org/10.1016/j.isprsjprs.2017.01.016>.
29. Hua, T.; Zhao, W.; Liu, Y.; Wang, S.; Yang, S. Spatial Consistency Assessments for Global Land-Cover Datasets: A Comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. *Remote Sensing* 2018, 10, doi:10.3390/rs10111846.
30. Pérez-Hoyos, A.; Rembold, F.; Kerdiles, H.; Gallego, J. Comparison of Global Land Cover Datasets for Cropland Monitoring. *Remote Sensing* 2017, 9, doi:10.3390/rs9111118.
31. Gao, Y.; Guo, Y.; Wang, W.; Li, F.; Huang, P. Accuracy evaluation of different land use or land cover data in grassland of northern China. *Chinese Journal of Ecology* 2019, 38, 283-293, doi:10.13292/j.1000-4890.201901.003.
32. Niu, Z.; Guo, Y.; Shan, Y.; Zhang, H. Accuracy Assessment of Wetland Categories from the GlobCover2009 Data Over China. *Wetland Science* 2012, 10, 389-395, doi:10.13248/j.cnki.wetlandsci.2012.04.002.
33. Meng, W. Accuracy Assessment for Regional Land Cover Remote Sensing Mapping Product Based on Spatial Sampling: A Case Study of Shaanxi Province, China. *Journal of Geo-information Science* 2015.
34. Ma, J.; Qun, S.; Qiang, X.; Bowei, W. Accuracy Assessment and Comparative Analysis of GlobeLand30 Dataset in Henan Province. 2016.
35. Wang, Y.; Zhang, J.; Liu, D.; Yang, W.; Zhang, W. Accuracy Assessment of GlobeLand30 2010 Land Cover over China Based on Geographically and Categorically Stratified Validation Sample Data. *Remote Sensing* 2018, 10, 1213.
36. Kussul, N.; Shelestov, A.; Basarab, R.; Skakun, S.; Kussul, O.; Lavreniuk, M. Geospatial intelligence and data fusion techniques for sustainable development problems; 2015; Volume 1356.
37. Jokar Arsanjani, J.; Tayyebi, A.; Vaz, E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat International* 2016, 55, 25-31, doi:<https://doi.org/10.1016/j.habitatint.2016.02.003>.
38. Manakos, I.; Karakizi, C.; Gkinis, I.; Karantzalos, K. Validation and Inter-Comparison of Spaceborne Derived Global and Continental Land Cover Products for the Mediterranean Region: The Case of Thessaly. *Land* 2017, 6, 34.
39. Mayaux, P.; Eva, H.; Gallego, J.; Strahler, A.H.; Herold, M.; Agrawal, S.; Naumov, S.; De Miranda, E.E.; Di Bella, C.M.; Ordoyney, C.; et al. Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing* 2006, 44, 1728-1739, doi:10.1109/tgrs.2006.864370.
40. John, L.; Renato, C.; Ilaria, R.; Mario, B. Global Land Cover-share of year 2014- Beta-Release 1.0 FAO Global Land Cover Network(GLCN). Available online: <https://www.fao.org/uploads/media/glc-share-doc.pdf> (accessed on 2023-01-10).
41. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.; Mathis, M.; Brumby, S. Global land use / land cover with Sentinel 2 and deep learning; 2021; pp. 4704-4707.
42. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data* 2022, 9, doi:10.1038/s41597-022-01307-4.
43. Liu, H.; Gong, P.; Wang, J.; Clinton, N.; Bai, Y.; Liang, S. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth System Science Data* 2020, 12, 1217-1243, doi:10.5194/essd-12-1217-2020.
44. Tateishi, R.; Hoan, N.T.; Kobayashi, T.; Alsaadeh, B.; Tana, G.; Phong, D.X. Production of Global Land Cover Data – GLCNMO2008. *Journal of Geography and Geology* 2014, 6, doi:10.5539/jgg.v6n3p99.
45. Kobayashi, T.; Tateishi, R.; Alsaadeh, B.; Sharma, R.C.; Wakaizumi, T.; Miyamoto, D.; Bai, X.; Long, B.D.; Gegentana, G.; Maitiniyazi, A.; et al. Production of Global Land Cover Data – GLCNMO2013. *Journal of Geography and Geology* 2017, 9, doi:10.5539/jgg.v9n3p1.
46. Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tsend-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaadeh, B.; et al. Production of global land cover data – GLCNMO. *International Journal of Digital Earth* 2011, 4, 22-49, doi:10.1080/17538941003777521.
47. Pierre, D.; G, K.; C, B.; M, B.; M, P.; S, B.; C, L.; M, S.; M, S. Land Cover CCI: Product User Guide Version 2. Available online: <http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-PUG-v2.5.pdf> (accessed on 2023-01-10).
48. Bicheron, P.; Defourny, P.; Brockmann, C.; Schouten, L.; Vancutsem, C.; Huc, M.; Bontemps, S.; Leroy, M.; Frédéric, A.; Herold, M.; et al. GLOBCOVER: products description and validation report. 2008.
49. Defourny, P.; Bontemps, S.; Bogaert, E. GLOBCORINE 2009. Product Description Manual; 2010.
50. Buchhorn, M.; Smets, B.; Bertels, L.; Roo, B.D.; Lesiv, M.; Tsendbazar, N.-E.; Li, L.; Tarko, A. Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Product User Manual. 2020, doi:10.5281/zenodo.3938974.
51. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; Peng, S.; Han, G.; Zhang, H.; He, C.; et al. Concepts and Key Techniques for 30 m Global Land Cover Mapping. *Acta Geodaetica et Cartographica Sinica* 2014, 43, 551-557, doi:10.13485/j.cnki.11-2089.2014.0089.
52. Chen, J.; Chen, L.; Chen, F.; Ban, Y.; Li, S.; Han, G.; Tong, X.; Liu, C.; Stamenova, V.; Stamenov, S. Collaborative validation of GlobeLand30: Methodology and practices. *Geo-spatial Information Science* 2021, 24, 134-144, doi:10.1080/10095020.2021.1894906.

-
53. Liu, J.; Liu, M.; Zhuang, D.; Zhang, Z.; Deng, X. Study on Spatial Pattern of Land-use Change in China During 1995-2000. *Science in China Series D Earth Sciences* 2003, 46, 373-384, doi:10.1360/03yd9033.
54. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W. Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Journal of Geographical Sciences* 2014, 69, 3-14.
55. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth System Science Data* 2021, 13, 3907-3925, doi:10.5194/essd-13-3907-2021.
56. Yu, Z.; Jin, X.; Miao, L.; Yang, X. A historical reconstruction of cropland in China from 1900 to 2016. *Earth System Science Data* 2021, 13, 3203-3218, doi:10.5194/essd-13-3203-2021.
57. Olofsson, P.; Stehman, S.; Woodcock, C.; Sulla-Menashe, D.; Sibley, A.; Newell, J.; Friedl, M.; Herold, M. A global land-cover validation data set, part I: Fundamental design principles. *International Journal of Remote Sensing* 2012, 33, 5768-5788, doi:10.1080/01431161.2012.674230.
58. Fung, T.; LeDrew, E. The Determination of Optimal Threshold Levels for Change Detection Using Various Accuracy Indices. 2008.
59. Janssen, L.L.F.; Wel, F.v.d. Accuracy assessment of satellite derived land - cover data : a review. *Photogrammetric Engineering and Remote Sensing* 1994, 60, 419-426.
60. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020, 21, 6, doi:10.1186/s12864-019-6413-7.
61. Clark, M.L.; Aide, T.M.; Grau, H.R.; Riner, G. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment* 2010, 114, 2816-2832, doi:10.1016/j.rse.2010.07.001.
62. Ran, Y.; Li, X.; Lu, L. Accuracy Evaluation of the Four Remote Sensing Based Land Cover Products over China. *Journal of Glaciology and Geocryology* 2009, 31, 490-500.
63. Foody, G. Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment* 2002, 80, 185-201, doi:10.1016/S0034-4257(01)00295-4.
64. Lambin, E.; Geist, H. Land-Use and Land-Cover Change: Local Processes and Global Impacts. 2006; Volume 18.
65. Zhang, C.; Dong, J.; Ge, Q. Quantifying the accuracies of six 30-m cropland datasets over China: A comparison and evaluation analysis. *Computers and Electronics in Agriculture* 2022, 197, 106946, doi:https://doi.org/10.1016/j.compag.2022.106946.