Preprints.org

Article

Comprehensive Representations of Subpixel Land Use and Cover Shares by Fusing Multiple Geospatial Datasets and Statistical Data with Machine-Learning Methods

Yuxuan Chen , Rongping Li [*] , Yuwei Tu , Xiaochen Lu , Guangsheng Chen [*]

Article

Comprehensive Representations of Subpixel Land Use and Cover Shares by Fusing Multiple Geospatial Datasets and Statistical Data with Machine-Learning Methods

Yuxuan Chen [1,2], Rongping Li [3,*], Yuwei Tu [1,2], Xiaochen Lu[1,2] and Guangsheng Chen [1,2,*]

[1] State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, China; chenyuxuan@stu.zafu.edu.cn (Y.C.); tyuwei@stu.zafu.edu.cn (Y.T.); 2020602041069@stu.zafu.edu.cn (X.L.)

[2] College of Environmental and Resource Sciences, Zhejiang A&F University, Hangzhou 311300, China

[3] Institute of Atmospheric Environment, China Meteorological Administration, Shenyang 110166, China

* Correspondence: lirp@iae.sy.cn (R.L.); chengu1@zafu.edu.cn (G.C.)

Abstract: Land use and cover change (LUCC) is a key factor influencing global environmental and socio-economic systems. Many long-term geospatial LUCC datasets have been developed at various scales during the recent decades owing to the availability of long-term satellite data, statistical data and computational techniques. However, most existing LUCC products can not accurately reflect the spatiotemporal change patterns of LUCC at regional scale in China. Based on these geospatial LUCC products, Normalized Difference Vegetation Index (NDVI), socioeconomic data, and statistical data, we developed multiple procedures to represent both spatial and temporal changes of the major LUC types by applying machine-learning, regular decision tree and hierarchical assignment methods using the northeastern China (NEC) as a case study. In this approach, each individual LUC type was developed in sequence under different schemes and methods. The accuracy evaluation using sampling plots indicated that our approach can accurately reflect the actual spatiotemporal patterns of LUC shares in the NEC, with an overall accuracy of 0.82, Kappa coefficient of 0.77 and regression coefficient of 0.82. Further comparisons with existing LUCC datasets and statistical data also indicated our approach and dataset can more accurately and comprehensively represent the spatiotemporal patterns of all LUC types at subpixel level. Our approach unfolded the mixed pixel issue and integrated the strengths of all LUCC products through the fusion process. The analysis based on our developed dataset indicated that forest, cropland and built-up land area increased by 17.11×10^4 km^2, 15.19×10^4 km^2 and 2.85×10^4 km^2, respectively during 1980-2020, while grassland, wetland, shrubland and bareland decreased by 26.06×10^4 km^2, 4.24×10^4 km^2, 3.97×10^4 km^2, and 0.92×10^4 km^2, respectively. The temporal change patterns of all these LUC types were consistent with the provincial inventory data. Our developed approach can be widely applied in the entire China and worldwide, and our data products can provide accurate data supports for studying the LUCC consequences and making effective land use policies.

Keywords: Fractional land cover share; machine-learning method; the northeastern China; land use and cover change (LUCC); NDVI

1. Introduction

The land use/cover change (LUCC) is closely associated with human production and living, social and economic development, as well as ecological carrying capacity [1–3]. With the continuous development and releases of remote sensing images and advancements in image processing techniques, many LUCC products were developed during the recent decades at regional, national and global scales [4,5], such as the Global Land Cover map (DISCover) for 1992 [6], the Global Land Cover 2000 (GLC2000) [7], the MODIS series products [8], the 30-m global land cover data (Globeland30) [9], the 10-m Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) products for 2017 [10], the 30-m fine classification system global land cover product

(GLC_FCS30) [11], European Space Agency Climate Change Initiative (ESA-CCI) land cover product during 1992-2020 (300 m) [12], and the ESRI annual map of Earth's land surface for 2017-2023 [13]. Due to the requirements of higher temporal and spatial accuracy data, many LUCC products were also produced specifically for China, such as China's Land-Use/cover Dataset (CLUD) at 30-m resolution for the 1980s, 1995, 2000, 2005, 2010, 2015 and 2020 [14,15], annual China Land use/Land cover datasets (CLUD-A) [16] and China Land Cover Dataset (CLCD) [17]. Although these datasets have been validated with high accuracy, the intercomparisons indicated that there is a large discrepancy among these datasets, and none of the spatiotemporal patterns of these datasets match well with the China's statistical or inventory data at both regional and national scales [18]. Most datasets showed a slight increase or even decrease in forest area from the 1980s to present, and none of the datasets can match the temporal change trends of national statistical data. For example, Qin et al. [19] has compared several LUCC products and indicated that the forest area of five datasets ranged from 174×10^4 km² to 227×10^4 km² in 2010. Yang and Huang [17] reported that forest area in China has only increased by 4.34% during 1980-2019, significantly lower than the national forest inventory (NFI) released 77% increase from 1984-1988 (12.98% forest coverage) to 2014-2018 (22.96%). Yu et al. [20] indicated that most of the cropland data in the existing LUCC products are not consistent with the statistical data by comparing over 10 existing cropland datasets. Similarly, the wetland area in the CLUD, MODIS, CLCD and CLUD-A has changed less than ±5% during 1980-2020, while reports have indicated that China's wetland area has significantly reduced by 33% [21,22]. In addition, most of these existing datasets only targeted at a single LUC type, while few studies have comprehensively addressed the spatiotemporal patterns for all LUC types. Therefore, it is necessary to produce a more accurate and comprehensive long-term LUCC dataset for China.

Several attempts have been made to match areas based on statistical and field survey data. For instance, Xia et al. [23] reconstructed a new forest cover data set (CFCD) from 1980 to 2015 by combining several existing LUCC datasets and NFI; however, this approach only matched the temporal change patterns but sacrifices the spatial accuracy. To match the statistical cropland area and change trends, Yu et al. [20] developed a subpixel level cropland share dataset; however, this dataset only targeted at the cropland area and did not consider other LUC types. There are two major reasons for the misrepresentation of LUCC at spatiotemporal scales in China. The first reason is that most LUCC products were developed using the pixel-based classification methods [20,24]. In the pixel-based approach, each pixel is regarded binary value (either Boolean 0 or 1), i.e., each grid cell is completely occupied one land cover type [20]. This approach is more suitable for high-resolution images [25]. The small percentage of the pixels could be ignored based on this approach, resulting in an underestimation for the changes. Using forest as example, the forest area in China is defined as tree coverage greater than 10% within a minimum area of 0.5 ha. The pixels with tree coverage ranging from 10% to 100% are regarded as forest area, which will result in the failing reflection of the change of pixel-level tree coverage in the LUCC products and an underestimation of forest area increase is caused when the tree coverage increases from 10% to 100%. To develop a more accurate temporal change pattern of LUCC, it is necessary to produce a subpixel level LUCC dataset that can reflect the fractional shares of each LUC type within each pixel [20]. The second reason is that most LUCC products did not simultaneously match the spatiotemporal patterns of all LUC types with statistical data. For example, Xia et al. [18], Yu et al. [20], Gong et al. [26] and Niu et al. [21] only targeted at match the forest, cropland, urban and wetland area with inventory data, respectively. None of current long-term LUCC products in China can comprehensively match all LUC types with statistical data. It is a challenge to harmonize the area and its temporal changes for all LUC types within each pixel. Recently, several long-term assisting geospatial datasets such as the normalized difference vegetation index (NDVI) and leaf area index (LAI) datasets have been developed [27–29]. Based on the existing vegetation indices and LUCC products, it is possible to invert the real changes in LUC shares within pixels.

The northeastern China (NEC) covers about 15.3% of China's territory. It is the main bases for crop and wood productions in China, and has the largest wetland area compared with other regions. With the rapid transitions of socioeconomic environment, this region has experienced dramatic and complex changes in various LUC categories during the recent decades, making this region an ideal case for developing approaches of LUCC products. Through the comparisons with existing LUCC

products, we found that no LUCC products can comprehensively catch the actual changes in LUC area in the NEC during 1980-2020; therefore, it is also necessary to reconstruct a long-term and high-precision proportional LUCC dataset to accurately reflect the spatiotemporal patterns in major LUC types in the NEC. The objectives of this study are to: (1) construct a approach for tracking the changes of fractional shares of various LUC types by integrating multiple LUCC products, other geospatial datasets with statistical data using machine-learning and regular decision-tree methods; (2) evaluate the performance of this approach using the NEC as a case study area; (3) analyze the spatiotemporal patterns of LUCC in the NEC.

2. Study Area and Data Descriptions

2.1. Study Area

The northeastern China (NEC) is located between 118°-128°E and between 40°-48°N and has a total land area of 1.47×10^6 km^2 (Figure 1). The NEC includes Heilongjiang, Jilin, and Liaoning Provinces, as well as the eastern portion of the Inner Mongolia Province. This region has undergone massive LUCC since the 1980s, and is of great importance in ecological conservation, forest wood production, and food security in China [30]. Most of the region is characterized with a temperate climate with a small portion with a boreal (cold temperate) climate. Summer is short, dry and hot, while winter is long and cold. The mean annual air temperature is about 3.11 ℃, and mean annual precipitation is about 785 mm. Mean annual temperature increased by about 1.36 ℃ during 1980-2020. Precipitation mostly occurred in summer and decreases from about 1100 mm in the southeast to less than 300 mm in the southwest. This region has the largest wetland area in China but has greatly shrunk during the recent decades   [33], and also has the largest area of natural forest area. A large portion of the western NEC belongs to the TNSF afforestation project region; and thus large forest area has been planted since 1978. The NEC is also the most important wood and crop production bases in China. With the above conditions, rapid urbanization, increasing population, and economic development, the LUC area and types has been dramatically changed.
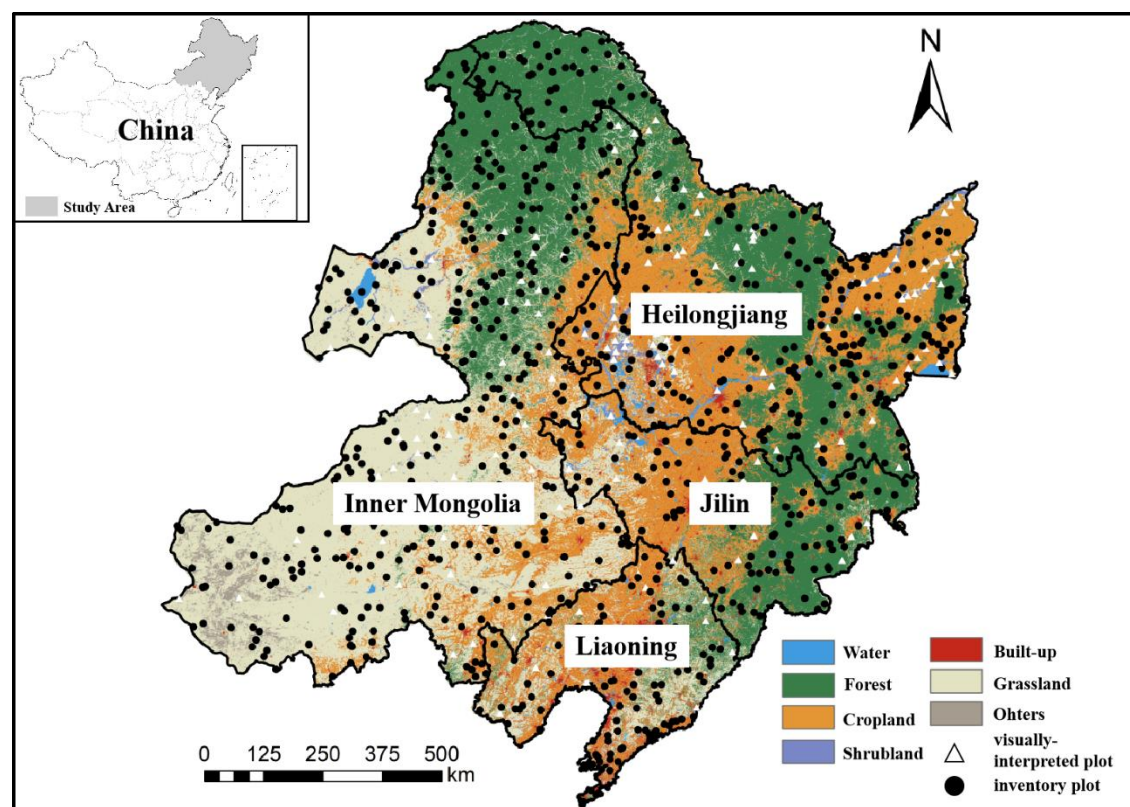

Figure 1. The study area and the distribution of major land use and cover types (Source: [17]) and the training & validation sample plots. Note: the triangle points are visually-interpreted plots at 1 km scale; the dark circle points are field investigated plots at 30 m scale (Source: [31])

2.2. Data Descriptions

2.2.1. Statistical Data

To develop the accurate LUCC area in the NEC, this study collected the inventory statistical data at provincial level. The land area of each land cover type during 1980-2020 was collected from the statistical yearbook data from each province. The annual cropland area and type data during 1980-2020 was retrieved from the provincial statistical yearbook (e.g., the data for Jilin Province is from http://tjj.jl.gov.cn/tjsj/tjnj/2009/ml/indexe.htm; accessed on July 25th, 2024). The forest area data for each 5 years during 1980-2020 were obtained from the National Forest Inventory (NFI) of China (https://www.stgz.org.cn/ldbggzpt/; accessed on July 25th, 2024). The 5-year NFI data was further linearly interpolated to annual data. The forest land from NFI was defined as the tree coverage is greater or equal to 20%, which includes tall forest, bamboo stand, dense shrubland (shrub coverage is greater than 30% in the arid and semi-arid region), tree nursery and cleared forest area due to fire disturbance and harvesting. The statistical data for other LUC types (grassland, shrubland, bare land, water body and built-up land) at the provincial level were collected from the First, Second and Third National Land Survey in 1997, 2007, and 2020, respectively. To eliminate the large fluctuation of inter-annual changes, statistical data was linearly fitted based on annual area and recalculated based on these fitted lines.

2.2.2. Geospatial Datasets

Many geospatial data were collected to assist the generation of the LUCC dataset. These include the existing LUCC products, long-term NDVI time series data, and other socioeconomic datasets (Table 1). The multiple LUCC products were collected to reconstruct the boundary and historical changes in each LUC category. These datasets include CLCD [17], ESRI LUCC product (https://livingatlas.arcgis.com/landcover/, accessed on July 25th, 2024), the CLUDA [16], NLCD (http://www.nesdc.org.cn/, accessed on July 25th, 2024), GLASS-GLC [32], and MODIS product [8]. We clipped the NEC area from these national or global level LUCC products. We also collected some literature data (non-spatial) for the evaluation of our approach, including the Wang_wetland [33], Mao_wetland [34] and Ye_grassland [35] data. Based on the NDVI data from AVHRR during1982 to 1999 (GIMMs NDVI) and MODIS during 2000 to 2020, previous researchers have developed multiple NDVI datasets at the 30 m, 1 km and 0.05º spatial resolution during 1981-2020 [27–29]. We first aggregated the 30 m NDVI dataset (1986-2023) to 1 km resolution for the period 1986-2020 and applied the change trends of the 0.05º NDVI dataset to extend the 1 km NDVI data for the period 1980-1985. This long-term NDVI data was used to extrapolate the change trend of land shares. All above geospatial datasets were downscaled or upscaled to the 1 km spatial resolution based on the neighborhood or average principles.

Table 1. Geospatial datasets used for land use and cover reconstruction or comparisons.

Datasets	Resolution	Time period	Sources
ESRI-LUCC	10 m	2017-2023	https://livingatlas.arcgis.com/landcover/
FROM-GLC	10 m	2017	[10]
NLCD	30 m	1980, 1990, 1995, 2000, 2005, 2010, 2015, 2020	http://www.nesdc.org.cn/
MODIS	500 m	2000-2020	https://modis-land.gsfc.nasa.gov/landcover.html
CLUDA	1 km	1980-2015	[36]
CLCD	30 m	1990-2020	[17]
GLASS-GLC	0.05°	1982-2015	[32]
GLC	1 km	1980-2100	[37]
Yu_cropland	5 km	1900-2016	[20]
Xia_forest (CFCD)	1 km	1980-2015	[18]
GLCLUC	30 m	2000-2020	[38]

GFC	0.05°	1982-2016	[39]
You_croptype	10 m	2017-2019	[40]
Mao_wetland	30 m	2015	[24,34]
NDVI	30 m	1986-2020	[29]
NDVI	0.05°	1981-2023	[28]

2.2.3. The Sampling Plot Datasets

We collected the all-season sample set data developed by the Finer-Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) project [31] to evaluate the performance of our approach in generating the LUCC products. In total, 2,453 validation plots were included for the NEC region. These samples mostly reflected the LUC types in 2015, and have undergone standardization and strict processing. These sampling plots were used for evaluating the performance of our developed datasets in reflecting the spatial patterns of LUC types.

To evaluate the performance of our approach in generating the LUCC shares at pixel level, we further collected sampling plots from the Google Earth Pro platform using the visual interpretation methods. The high-resolution (<5 m) images were chosen to conduct digitizing LUC shares within 1 km² pixels. The plots were chosen at the places with at least two periods high-resolution images were available during 2005-2020. The boundary different LUC types were drawn using polygons. The shares of different LUC types were calculated based on the polygon areas. The share changes between two time periods were calculated to represent the changes of LUCC shares during this period. The years, LUC types and changed shares were recorded. The developed LUC share datasets at pixel level were finally compared with these sampling plot data, and thus the LUCC areas and spatial patterns of the LUC share datasets were evaluated against the sampling plot data during the same time period. Due to very few available high-resolution images with at least two time periods in the NEC, we finally only identified 65 sampling plots.

3. Methods

To develop the integral LUCC dataset and match the statistical area for all LUC types, we separately generated the cropland, built-up land, forest, wetland, grassland, shrubland and bareland in sequence, and different approaches were developed for each LUC type due to different impact factors controlling the changes of these LUC types.

3.1. Hierarchical Assignment Method for Cropland Share Dataset

A hierarchical assignment method was applied to develop the cropland share maps during 1980-2020. This method was first proposed by [25] in the United States and was further applied in China [20]. To develop the gridded cropland share (%) data, we slightly modified this procedure by fusing the approach in [41]. Integrating existing land use products to produce new cropland maps involves a series of steps (Figure 2). First, the collected six cropland data products undergo uniform preprocessing. The first step is to determine the appropriate weight order for the input products. We used a fused method of accuracy information and expert judgment to establish the weight order of the input datasets (higher accuracy > lower accuracy, higher resolution > lower resolution). The second step is to label high-score cropland pixels. Each pixel is assigned different scores based on the weight order and various combinations, with the highest score being 21. When a pixel has a higher score, it is more likely to be labeled as a cropland grid. This is compared with the provincial cropland area (R) to identify areas smaller than R, which are designated as cropland grids and marked as T (True), while the remaining low-score pixels are labeled as P (Possible), with a total area marked as areaT. However, this area differs significantly from the provincial statistical data, necessitating further selection of suitable pixels for allocation.

The third step involves reallocating the remaining pixels. We need to identify pixels that are more likely to be cropland among the remaining pixels. Cropland cultivation primarily relies on human activity, so land near human settlements is more likely to be cultivated, while land farther from settlements is more prone to abandonment. Additionally, cropland close to lakes, rivers, and reservoirs is more likely to be irrigated than that further away from water sources [41]. When prioritizing, proximity to densely populated areas and watersheds should be considered, with the

weighting based on the distance from each grid cell to the nearest urban, rural, and water bodies, combining these two weights to generate an auxiliary distance grid. Next, within a provincial extent, all remaining cropland pixels were ranked in descending order based on new weight values. The total area of the cropland pixels with the highest weight value was summed (TA1) to match the provincial area. If area1 was less than the provincial area (RA), the cropland pixels with the second-highest weight value were counted, and their area (areaP2) was combined and compared with RA. If (TA1) was still less than RA, the iteration process turned to the third-highest weight and continued until their total area was closest to RA. This iterative thresholding process was conducted in each province of Northeast China to obtain provincial cropland maps. After the cropland distribution maps have been developed, the main crops within pixels were further identified through a similar iterative approach based on the high-resolution (10m) crop type distribution data　[40].

3.2. Built-Up Land and Water Body Share

The CLUDA data [36] is represented as a fractional share for each LUC type within each pixel at 1km spatial resolution. Our built-up land and water body distribution data were slightly modified based on the CLUDA because this dataset has been compared with many previous datasets and was consistent with the spatiotemporal change patterns of inventory data. The pixel-level built-up land and water body area were slightly reduced at which the sum area for cropland, built-up land and water body was greater than 100%.
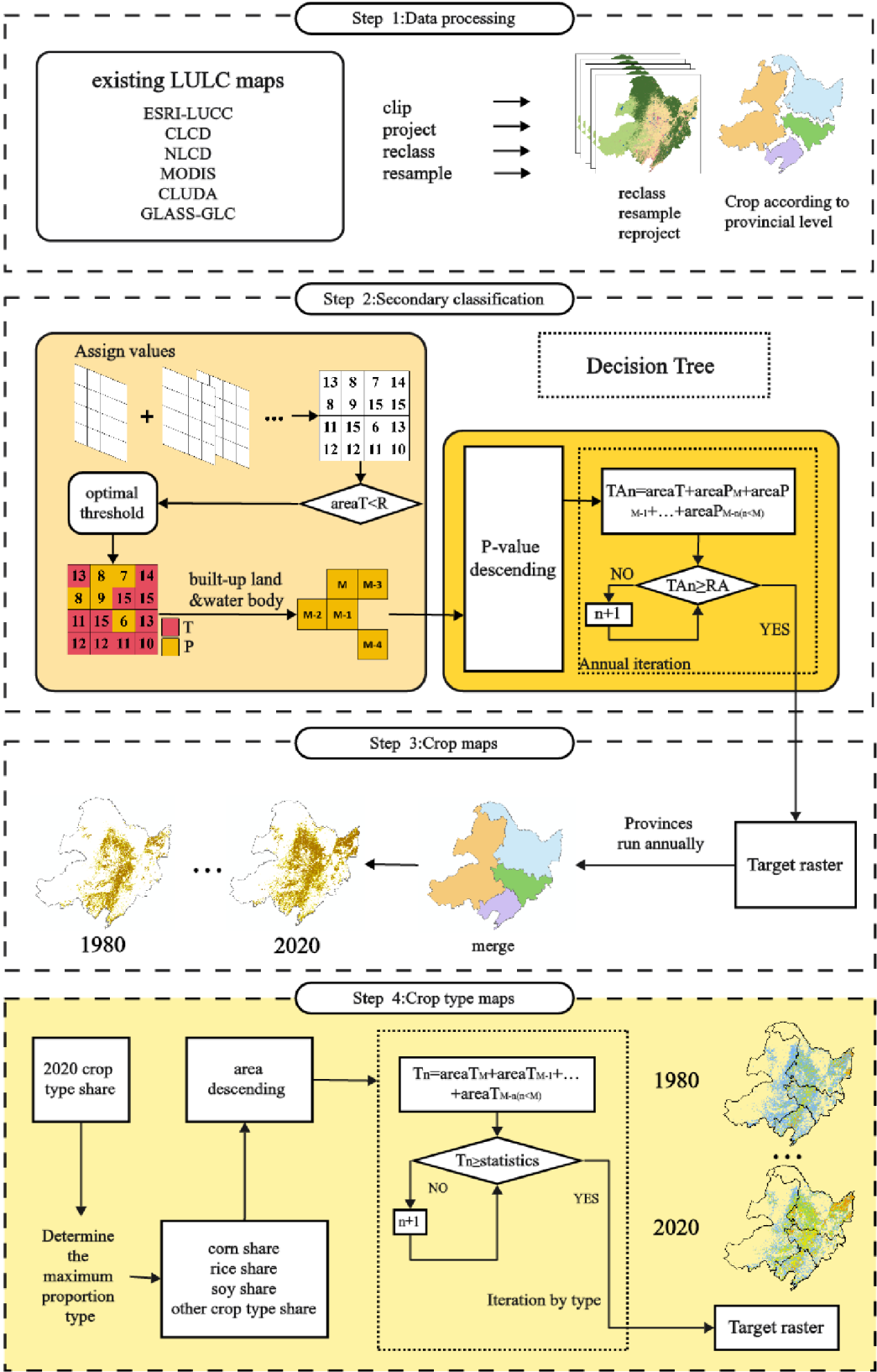
Figure 2. The dataset development procedure for fractional cropland area and crop types

3.3. Forest Share Data Reconstruction Method

After the cropland, built-up and water body datasets were generated, we further developed a procedure to reconstruct the forest area dataset (Figure 3). In this procedure, the first step is to generate the maximum forest distribution boundary. The LUCC products of ESRI-LUCC (2017-2023),

FROM-GLC (2017), CLCD, CLUDA, GLASS, NLCD and MODIS datasets were used to obtain the maximum forest cover boundary. After a series of processing such as reprojection, clipping, resampling, combination and aggregation, we generated the initial maximum forest distribution area during 1980-2020 at 1 km resolution. The pixels with forest cover share less than 10% was removed, and then the final maximum forest distribution boundary map was generated. The removal of non-forest area can reduce the interference and confounding caused by other land cover types.

At the second step, the annual NDVImax dataset during 1980-2020 was reconstructed. The combined monthly NDVI dataset at 1 km spatial resolution during 1980-2020 were applied to refit the change trends of forest share within forest-presence pixels. To eliminate the impact of anomalies, we first used the maximum value composites (MVC) method to composite the monthly NDVI data, and obtained the annual maximum NDVI dataset (NDVImax). To remove the unreasonable annual variations, we applied the SG filter (Savitzky-Golay Filter) to smooth the abnormal interannual variations of NDVImax at pixel level. The SG filter can not guarantee a continuous change trend of NDVImax during 1980-2020; therefore, to further establish a continuous change trend of NDVImax values, we fitted regression models at pixel level using the year as independent variable. We found that the polynomial equation ($y=ax^2+bx+c$) is the most suitable regression models for most pixels. Based on the fitted regression models, a new smooth-change annual NDVImax dataset during 1980-2020 was generated.

The third step is to invert forest share and its interannual variations. We assumed that the annual forest cover share changes can be represented by the annual NDVImax data. We first fitted the regression between regional change trends of annual NDVImax and inventory forest share during 1980-2020 using a polynomial equation: Forestshare = $-22.792\text{NDVImax}^2 + 34.77\text{NDVImax} - 12.916$, we found there was a significant correlation coefficient between them ($R^2 = 0.90$). However, we also found that this relationship is not suitable for individual pixels, and different regression models should be used to fit the forest area change. Therefore, we applied the random forest (RF) regressor algorithm to modify the regression models over spatial scale. The 5-year statistical provincial forest cover share change data were linearly interpolated to annual scale during 1980-2020. To obtain the change trend of provincial forest share, the RF regressor algorithm was iteratively run until the fitted forest share change pattern was close to the inventory forest share. After the RF regressor models have been determined, it was applied to simulate the pixel-level forest share changes using the NDVImax data as input.

At the final step, the additional forest share in the pixels greater than 100% by summing up cropland, built-up, water body and forest shares was subtracted.
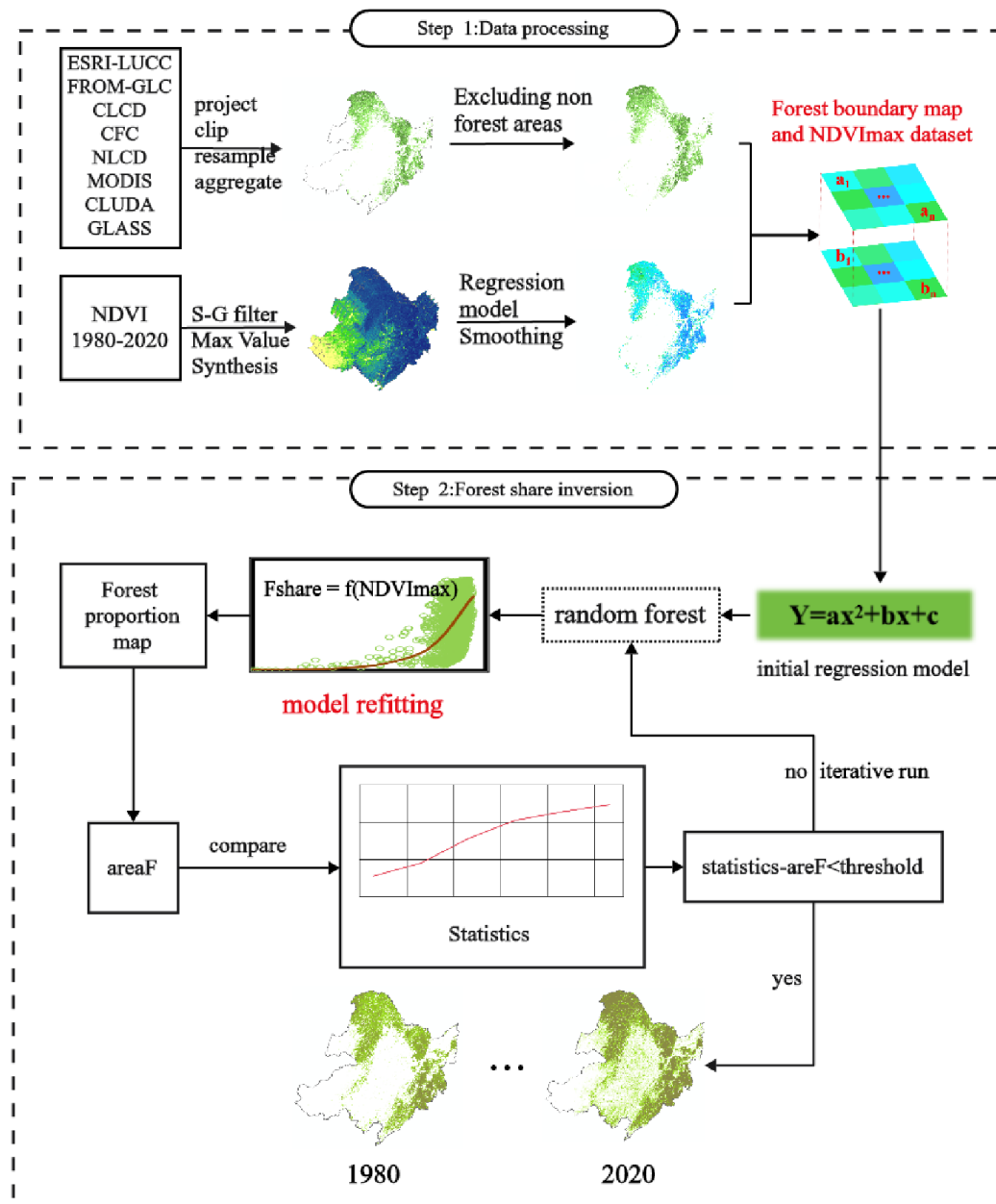
Figure 3. The forest share dataset development procedure.

3.4. Wetland Share Dataset Development

According to Niu et al. [21], China's wetland area has reduced by 33% from 1978 to 2008. According to Wang et al. [33] and Mao et al. [34], the wetland area has reduced by 30% during 1980-2015 and 33% during 1990-2019, respectively. Based on the differences in time periods, we finally calculated that wetland area has reduced by 34% from 1980-2020. In addition, based on their study year points (1980, 1990, 2000, 2013, 2015, 2019), we finally applied the linearly interpolation method to generate the annual wetland area during 1980-2020.

Before develop the geospatial wetland share dataset, we first calculated the maximum available land share for wetland by subtracting the sum values for cropland, built-up, water body and forest shares. In the wetland share dataset development procedure, the first step is to develop the wetland maximum boundary map and the current share (Figure 4). Similar to forest share development, the 8 existing LUCC products for the most recent year and Mao_Wetland [24] were used to generate the maximum wetland distribution boundary map. The wetland share in 2020 was generated on the basis of Mao_Wetland data, which is consistent with China's wetland inventory data. This dataset

was aggregated to 1 km spatial resolution and the wetland share was calculated. Finally, the maximum (potential) wetland boundary map and wetland share in 2020 were developed.

The second step is to retrospect the wetland share during 1980-2019. Because the wetland share was shrinking from 1980 to 2020, we need to expand the wetland area retrospectively from 2020 to 1980. The key problem is where would be the potential wetland area located in the history. Mao et al. [42], Wang et al. [33] and Luo et al. [43] noted that the reduction of wetland area in China was mostly due to conversion to cropland. Mao et al. [42] indicated that agricultural encroachment for food production is responsible for approximately 60% of natural wetland loss in China, 74.7% (11,778 km^2) of which occurred from 1990 to 2000. In addition, the conversion of natural wetlands to cropland has reached a maximum of 85.4% in the NEC. Furthermore, in a meta-analysis, Asselen et al. [44] also indicated that most of global wetlands were converted to cropland. Therefore, we also assume that the decreasing wetland in the NEC is primarily due to cropland expansion. In all pixels of maximum wetland distribution map, the increasing cropland area was assigned to the decreasing of wetland. Based on this rule, the initial wetland share was developed starting from 2020 to 1980.

However, based on the initial wetland share dataset, we found that the decline of wetland area from 1980 to 2020 was significantly greater than the actual decline area from inventory. To be consistent with the inventory data, we need to reduce the declining magnitude of wetland area. Similar to the cropland share data development method (the second step), we removed some unlikely wetland pixels from the initial wetland share dataset based on the declining order score rule. Through iteratively tuning the minimum score thresholds, we finally generated the wetland share dataset, which temporally consistent with the statistical wetland area.
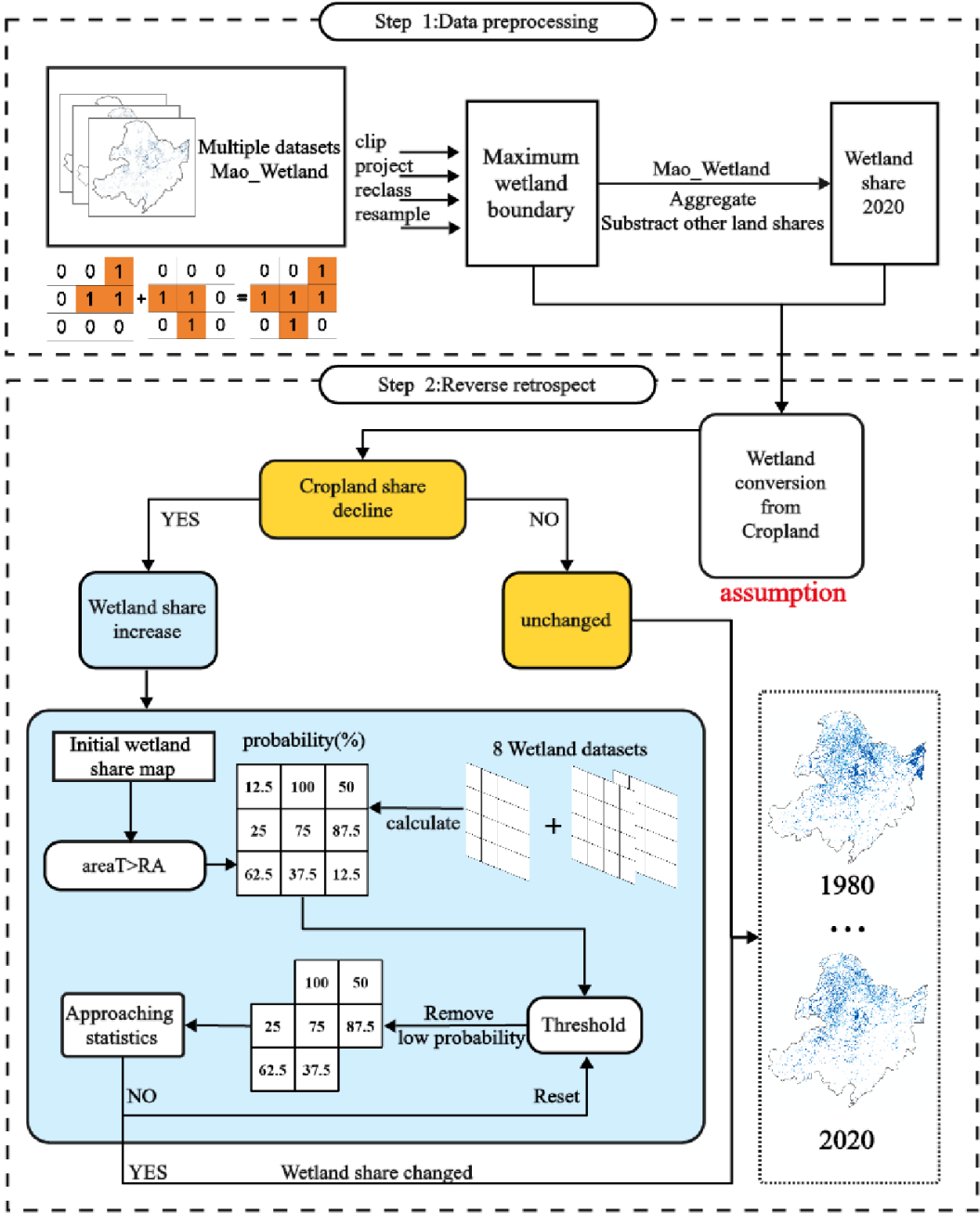
Figure 4. The wetland share dataset development procedure.

3.5. Reconstruction of Grassland, Shrubland and Bareland Shares

Before develop the geospatial grassland, shrubland, and bareland share datasets, we first calculated the maximum available land shares (remaining shares) for them by subtracting the sum values for cropland, built-up, water body, forest, and wetland shares. The statistical data for grassland and shrubland area from the National Land Survey were used for the constraint of the final area of these two LUC types.

To develop these datasets, the first step is to map the maximum distribution extents of grassland (Gmax), shrubland (Smax) and bareland (Bmax) (Figure 5). Similar to above procedures, the 6 LUCC products after preprocessing were overlaid, and pixels with presence of this LUC type were kept, while 0 was assigned for other pixels. Next, if multiple LUCC products have this LUC type in a pixel, then all their shares were kept. The share values of all products within a pixel will be ranked by a declining order and finally formed 6 initial share maps for each LUC type except for bareland, for

example, G1, G2, ...G6 (grassland); S1...S6 (shrubland). For bareland, no need to derive multiple maps of share since all left area will be assigned to it if bareland share present.

In the second step, the first iteration will be run using the initial share data (G1 and S1). The first share dataset for grassland (Gsh1) and shrubland (Ssh1) were calculated based on the initial shares and their ratios accounting for the remaining share (R). Based on the first share datasets, the total grassland and shrubland area was calculated and compared with the statistical data. If the area of the grassland or shrubland was not close to the statistical data, then the second-round iteration of calculation will implemented based on the G2 and S2 share dataset. Finally, we obtained the grassland and shrubland share datasets that are close to the statistical data. The left area (R-Gi-Si) will be assigned to bareland if it is present in the pixels.
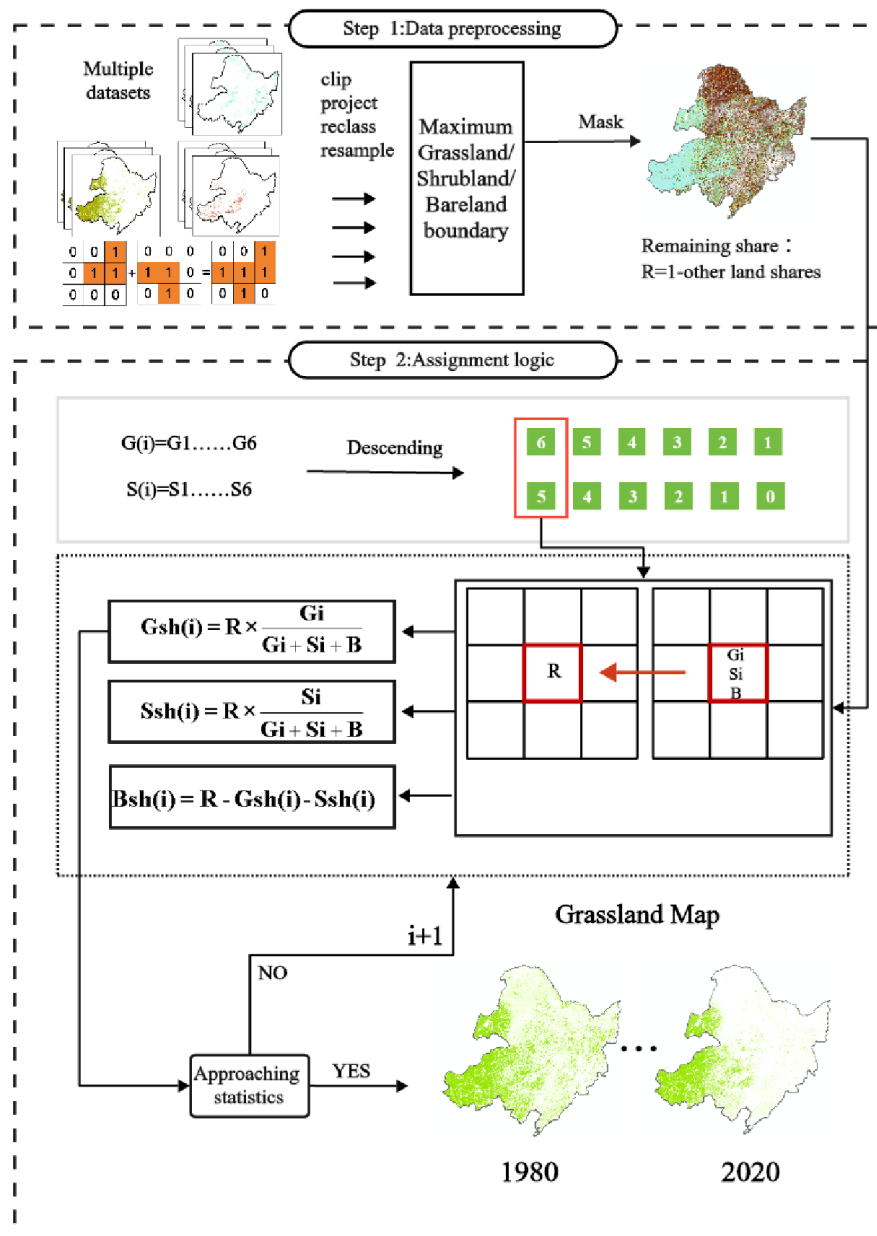

Figure 5. The grassland, shrubland and bareland share dataset development procedure.

3.6. Synthesis among All LUC Share Datasets

The shares for cropland, built-up, water body, forest, wetland, grassland, shrubland and bareland were summed up (SUM) for each year. For some pixels, the total shares were slightly less than 100%; therefore, we need to reassign the left shares (100%-SUM). Under this case, the left share was assigned to all existing LUC types within this pixel based on their proportions. For the pixels without any LUC types, the bare land share was assigned.

Therefore, all above developed LUC shares were slightly adjusted and formed the final products for all LUC types.

3.7. Accuracy Assessment and Intercomparisons

The collected 2543 and 65 sample plots were used to evaluate the performance of our developed LUCC datasets. We first chose the evaluation metrics of overall accuracy (OA), user accuracy (UA), producer accuracy (PA), and Kappa coefficient to evaluate the performance of spatial distribution patterns of various LUC types. Then, the performance of LUC shares was evaluated against the 65 visually-interpreted plots based on the correlation analysis using correlation coefficient (R^2) as a metric. All these metrics collectively offered a comprehensive evaluation of diverse facets of the model's performance [45].

4. Results

4.1. Accuracy Assessment

The accuracy assessment was first conducted using the 2453 validation sampling plots. The major LUC types in 2015 were extracted based on the share datasets for all LUC types. Based on the confusion matrix (Table 2), the producer and user accuracies were mostly greater than 80% for all LUC types, with lower accuracy for cropland and grassland and higher accuracy for forest and built-up land. The overall accuracy was 82%, and the Kappa coefficient was 0.77. Because the sample plot data from the FROM-GLC only reflected LUC types at 30 m spatial resolution, the lower accuracies for some LUC types may be caused by the inadequate representations for the major LUC types at the 1 km pixels. Therefore, the overall accuracy and Kappa coefficient can indicate that our developed datasets can accurately reflect the true spatial distribution patterns of LUC types in the NEC.

Due to few available high-resolution images for the NEC regions, we only visually-interpreted 65 sample plots for evaluating the performance of LUC share datasets. The correlation analysis indicated that the correlation coefficient (R^2) was 0.82 between visually-interpreted shares and our developed datasets for various LUC types (Figure 6). In addition, the spatial consistency was also compared using the 65 sample plots, and the result indicated that our classified LUC shares can match well with the visually-interpreted shares at spatial scale (Figure 7).

Table 2. The accuracy assessment confusion matrix for different land use and cover types.

| LUC types	cropland	forest	built-up	water	wetland	grassland	shrubland	Total	Producer accuracy	User accuracy
cropland	194	42	29	16	0	23	7	311	0.62	0.86
forest	10	839	12	29	10	26	21	947	0.89	0.89
built-up	2	8	338	4	0	0	0	352	0.96	0.80
water	0	0	6	276	9	0	0	291	0.95	0.76
wetland	2	8	5	9	107	21	0	152	0.70	0.82
grassland	16	41	30	20	5	173	12	297	0.58	0.70
shurbland	1	2	3	9	0	3	85	103	0.83	0.68
Total	225	940	423	363	131	246	125	2453	1.00	1.00
Overall accuracy = 82.02% Kappa = 0.77										
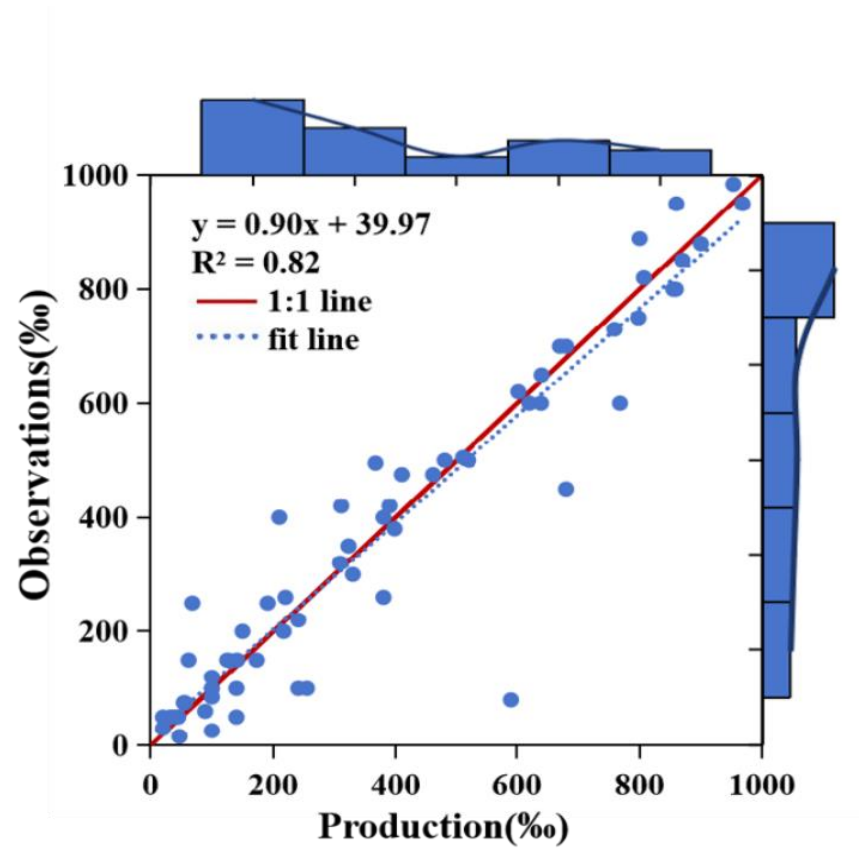
Figure 6. The correlations between the reconstructed and the visually-interpreted (observations) area shares (‰) for different land use and cover types within each pixel at 1 km spatial resolution.
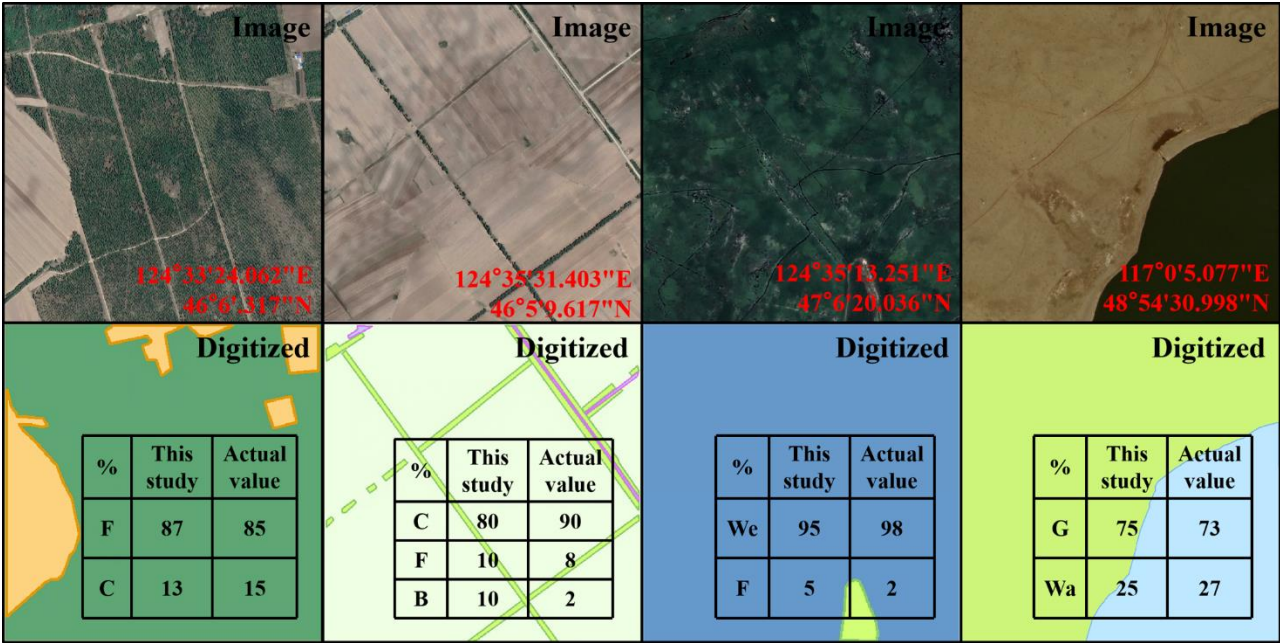

Figure 7. The evaluation of the developed land use and cover shares (%) within each pixel at 1 km spatial resolution against visually-interpreted shares based on high-resolution images. Note, F: forest share; B: built-up land share; C: cropland share; We: wetland share; G: grassland share; Wa: water body share.

4.2. Temporal change patterns of different land use and cover types

Based on the reconstructed datasets for LUC shares, the temporal change patterns of various LUC were analyzed (Figure 8). The results indicated cropland, forest, and built-up land area have increased by 72.36%, 47.42% and 150%, respectively from 1980 to 2020; while wetland, grassland,

shrubland and bareland have decreased by 39.04%, 43.67% and 68.73%, respectively. The largest increase of area was $1.71 * 10^5$ km² for forest due to the afforestation projects in the NEC and the largest reduction of area was $1.71 * 10^5$ km² for grassland. Cropland, built-up land forest showed faster increases since 2000. The results from LUC conversion matrix indicated that the increased cropland share from 1980 to 2020 was mostly owing to the conversion from grassland, followed by the conversion from wetland area. The increased forest share primarily came from the conversions of grassland and shrubland, mainly owing to the world largest TNSF project. The wetland area mainly converted to cropland due to reclamation. At provincial level, Heilongjiang Province has experienced the largest cropland expansion, with an increasing rate of 105.04%, while the least increase of 15.73% occurred in Liaoning Province. Forest share increased the most (64.17%) in Liaoning Province and the least (20.56%) in Jilin Province. Wetland share has reduced by 54.17% in Heilongjiang Province where has the highest wetland area, while has only reduced by 18.45% in Inner Mongolia. The grassland share has decreased the most (86.83%) in Heilongjiang and the least (24.51%) in Inner Mongolia.
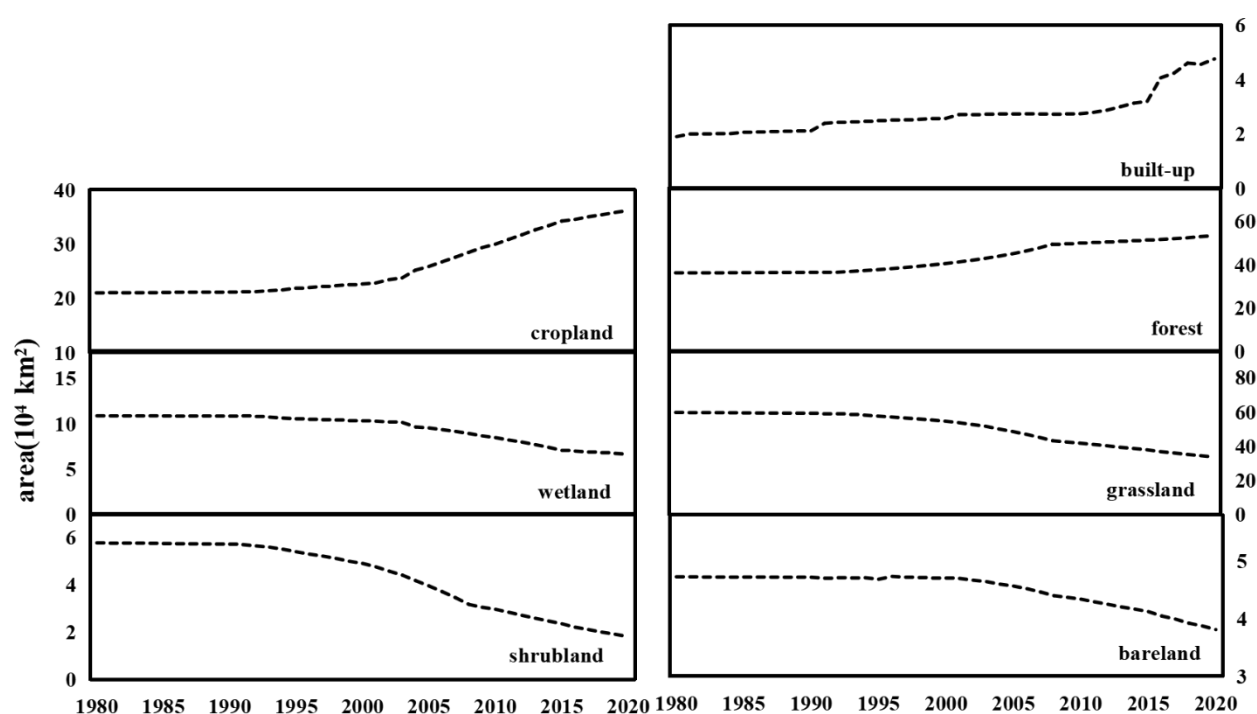

Figure 8. The total area (10^4 km²) of the all land use and cover types during 1980-2020 in the NEC.

4.3. Spatial Change Patterns in Land Use and Cover Types

The LUC shares in 1980, 2000 and 2020 were displayed in Figure 9. The area of cropland mainly distributed in the central and northeastern NEC (Songnen Plain and Liao River Plain). Cropland mainly distributed in the central and northeastern NEC. forest mainly distributed in the northwestern, north-central and southeastern NEC. Wetland mainly distributed in the northwestern, north-central and northeastern NEC, surrounding the rivers and lakes. The NEC has the largest wetland basins (Sanjiang Plain and Songnen Plain) in China. Grassland was mainly distributed in the southwestern China and scattered in the central NEC. The dense shrubland area was scarce in the NEC, mainly distributing in the northern and central NEC.

The expanded cropland shares mainly distributed in the central and northeastern NEC from 1980 to 2020 (Figure 10). The expansion primarily occurred during 2000-2020. The expansion of cropland generally displayed as a clustered and patchy pattern. Forest showed a region-wide expansion during 1980-2020, with the largest increases during 2000-2020 owing to the full implementation of the Three-North Shelterbelt Forest (TNSF) project and stricter national forest protection policy during this period. The largest declines of wetland shares mainly occurred in the northeastern (Sanjiang Plain) and central (Songnen Plain NEC during 1980-2020, with more declines

occurred during 2000-2020 owing to the rapid expansion of cropland. The declined grassland scattered in the entire NEC during 1980-2020 due to the expansions of forest and cropland, with smaller declines in the main distribution area in the southwest (Inner Mongolia). Shrubland area has experienced greater decline during 1980-2020 as compared with other LUC types because these areas are often also suitable for forest plantations, and thus were converted to forest in the afforestation projects.
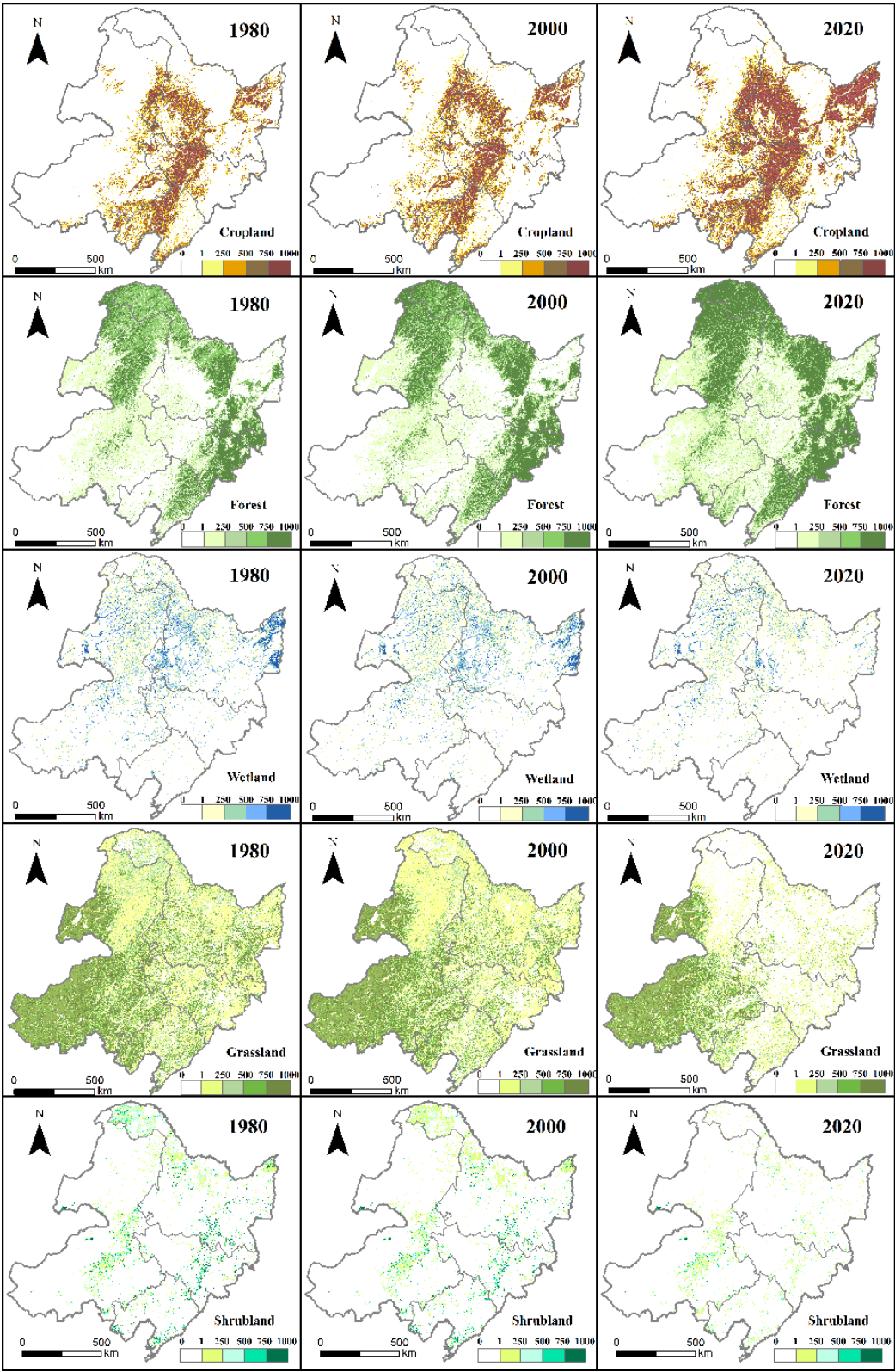
Figure 9. The spatial distribution of cropland, forest, wetland, grassland, and shrubland shares (‰) in 1980, 2000 and 2020 in the NEC.
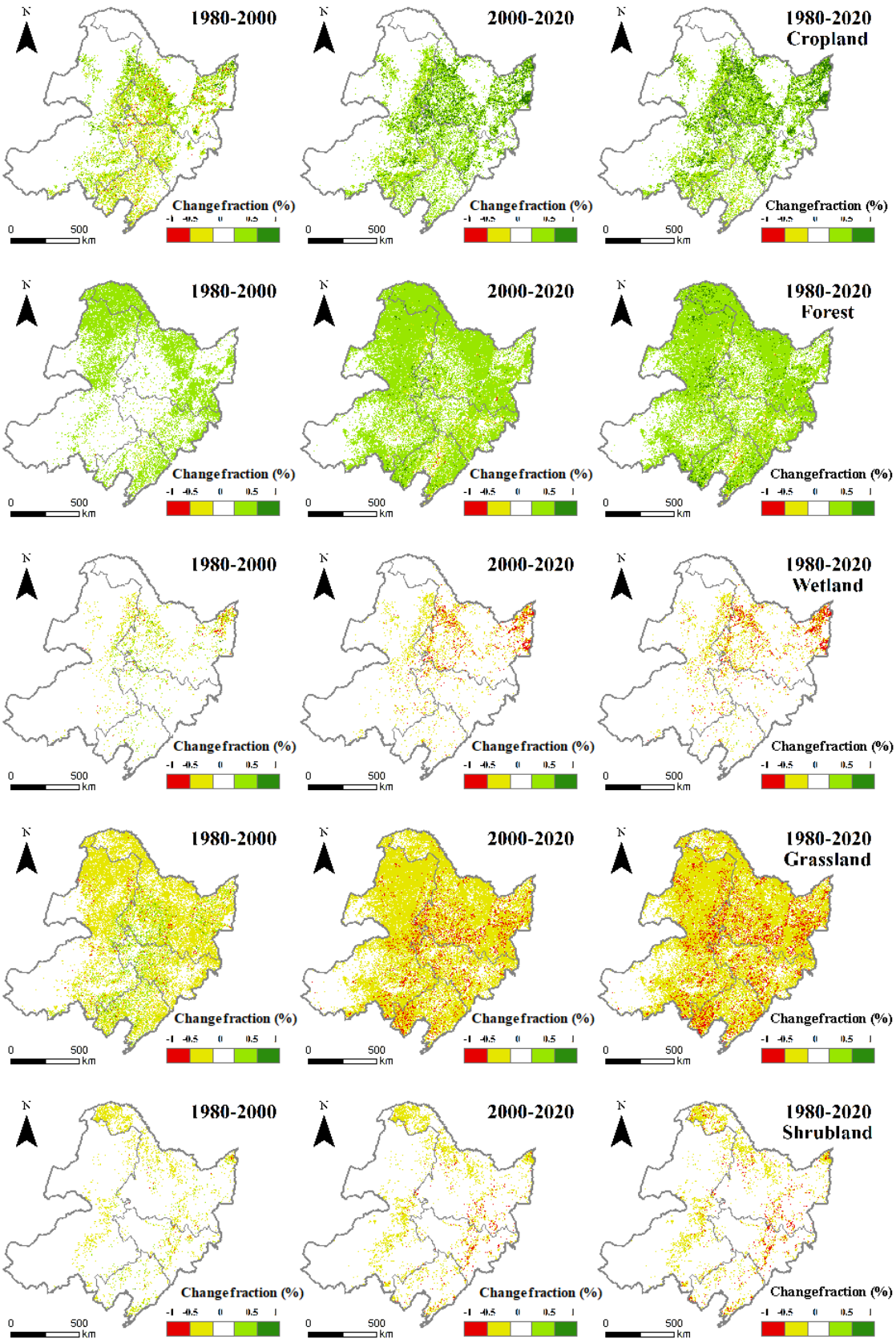
Figure 10. The spatial changes (%) of cropland, forest, wetland, grassland and shrubland shares during 1980-2000, 2000-2020 and 1980-2020 in the NEC.

4.4. The Comparisons with Existing LUCC Products

The magnitudes and temporal change patterns of our developed datasets for cropland, forest, wetland, and grassland shares were compared with other existing LUCC products and the statistical data to reflect the effectives of our approach (Figure 11). For cropland data, our dataset is close to the statistical data, and matched with most datasets in 2020; however, the change trends among different datasets were significantly different. Cropland area increased from 0% to 20% for most existing datasets, which are significantly lower than the change trend of the statistical data (69.77%). For forest area, some LUCC products can match the statistical forest area in 2020, e.g., the NLCD, CLUDA and CLCD; while other datasets generally showed a lower area in 2020. However, for the temporal change patterns from 1980 to 2020, only our dataset and Xia_Forest can match well with the statistical data. Other datasets only showed a slight increase or no change of forest area during this period. Most existing LUCC products showed significantly lower wetland area as compared with inventory data. The wetland area of all existing LUCC datasets (exclude Wang_wetland and Mao_wetland datasets, which have combined with inventory data) showed slight decrease (<5%) or no change during 1980-2020. For grassland, our dataset showed a decline of 41% during 1980-2020, while the NLCD and GLASS-GLC datasets showed a decline of 20.70% and 18.03%, respectively, and other datasets showed less than 10% decline. The inventory data from three times of National Land Survey indicated that grassland area has declined from 50.01 km^2 to 32.70 km^2 at a rate of 24.62%, which is lower than our result.

Based on the visually-interpreted sample plot data, we also compared the performance of our dataset at some pixels with two LUCC products with fractional values (Figure 12). From these three sampled pixels (1km^2), the digitized shares of our developed dataset can generally capture the actual shares of different LUC types, and the CLUDA dataset also match well with the actual data. In addition, the overall spatial patterns of these existing LUCC products were also compared with our dataset and we found a significant difference with our dataset (See Supplementary Figure S1-S5).
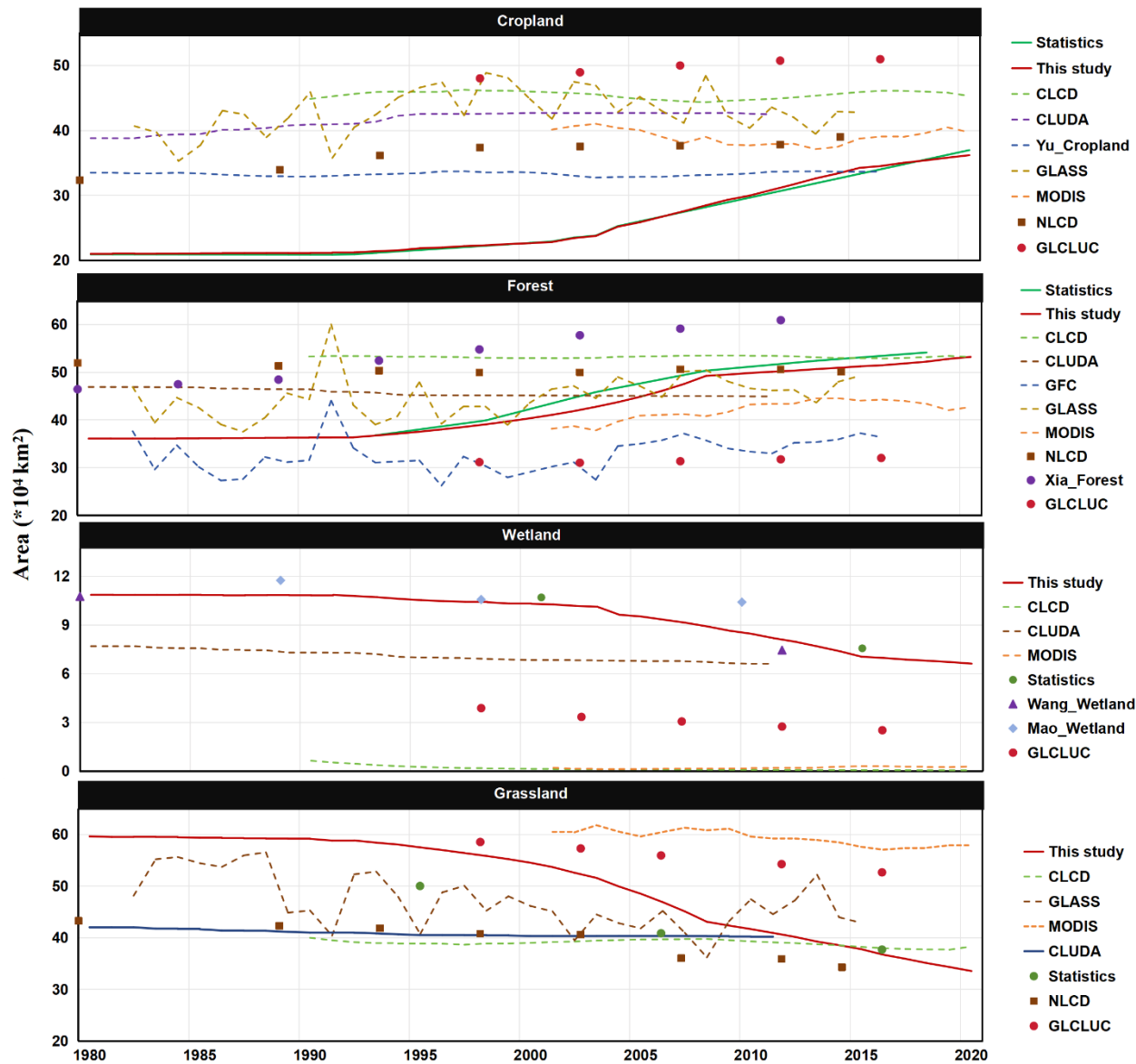
Figure 11. The intercomparisons of our developed LUC share dataset with two existing LUCC products and visually-interpreted data.
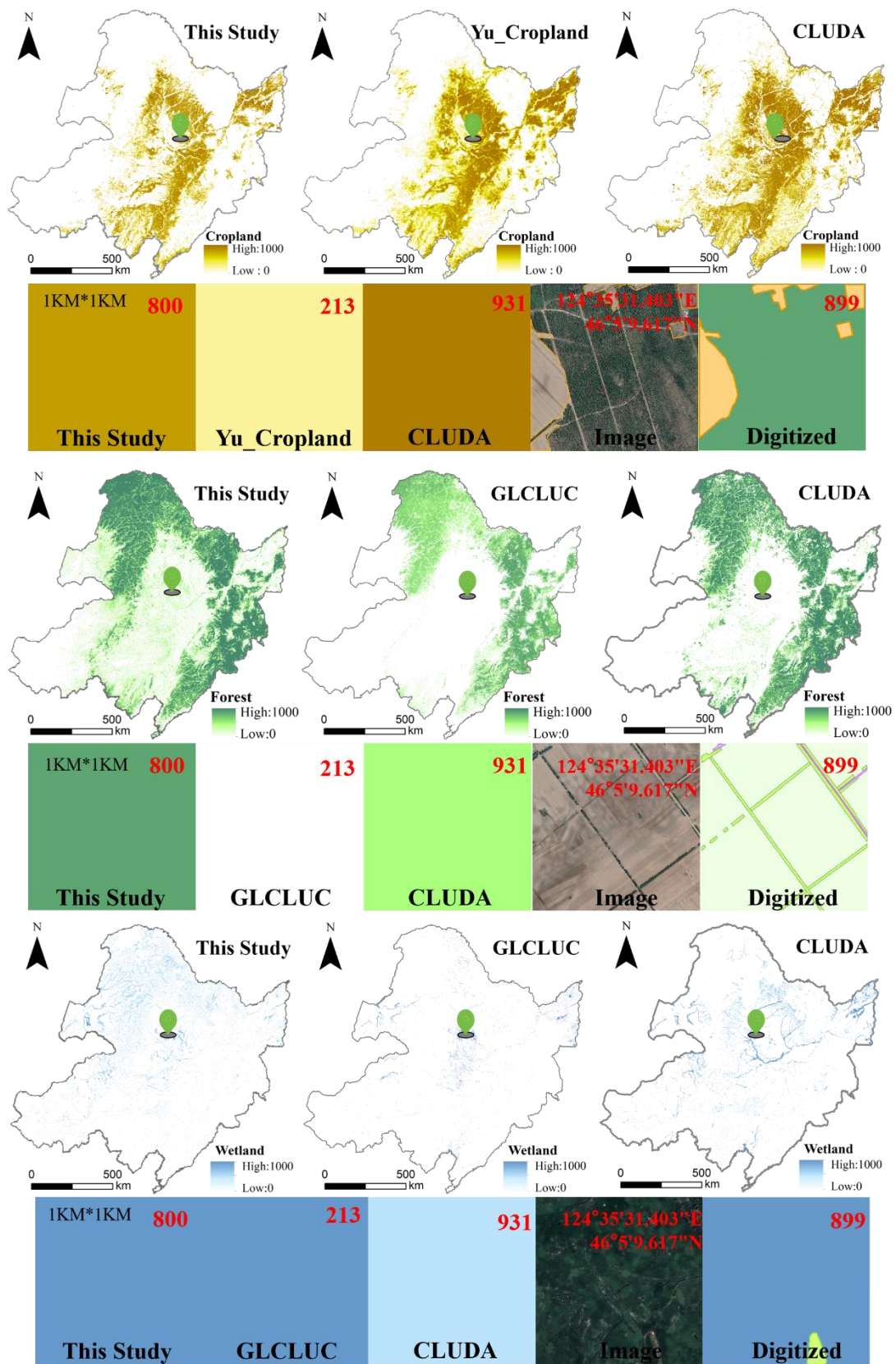
Figure 12. The evaluations of the LUC shares of our developed dataset against that of other LUCC products at the pixels with visually-interpreted LUC shares.

5. Discussion

5.1. The Effectiveness of Our Approach in Reflecting Spatiotemporal LUCC Patterns

The NEC has experienced extensive LUCC due to the climate change, urbanization, land reclamation, and economic development [17,33,35]. The NEC is the most important food production base in China, and crop yield is very high due to the fertile and moist soil conditions [20,40]. Due the high nutrients and favorable water condition, a large portion of wetland area has been reclaimed to the cropland to meet the increasing demand of cereal production since the 2000s [42]. A large part of the Three-North Shelterbelt Forest (TNSF) project, the world's largest afforestation project is located in the NEC region, which caused a rapid expansion of forest area and reduction of other land area, mostly grassland and shrubland where is suitable for tree planting [46,47]. In addition, the NEC was the major wood production base for China, causing large area of deforestation since the 1970s [48]. Since the 2000s, the Natural Forest Protection policy was fully implemented and the forest disturbance and loss has been gradually decreased. All above activities have driven a complex spatiotemporal LUCC pattern in this region. However, these phenomena have not realistically reflected in the current existing LUCC products [17,18,20].

Based on the statistical data and assisting datasets, our study has effectively integrated the existing LUCC products and developed a subpixel LUC share dataset that can accurately reflect the actual LUCC conditions in the NEC. Most existing datasets showed that forest area only slightly (<10%) increased in the NEC, which is obviously inaccurate as many studies have proved that the NEC was becoming greener and the forest area and biomass were increasing due to higher forest coverage [18,49,50]. Similarly, the temporal change patterns of other LUC types in most existing LUCC products were also not consistent with the statistical data, such as the grassland, cropland and wetland (Figure 11). Some LUCC products have tried to solve the inaccuracy in representing temporal change of individual LUC types, for example, Yu et al. [20,25] developed subpixel level cropland datasets for the entire China and USA based on multiple LUCC products and statistical data; Xia et al. [18] developed a pixel-level forest dataset for China by matching forest inventory data. These approaches can solve the misrepresentation of temporal changes for individual LUC types, but they are at the cost of losing accuracy for other LUC types. At present, few geospatial datasets can comprehensively consider the accuracies for all LUC types. Instead, our approach can effectively take account of the high accuracies of all LUC types as compared with the regional statistical data.

Through the comparisons with visually-interpreted LUC shares at each 1 km^2 pixel, we found that our classified LUC shares can generally match well (R^2 = 0.82) with the actual LUC shares, suggesting that our approach can provide high spatial accuracy. Further comparisons with other LUCC products indicated that our dataset can provide more detailed and accurate spatial representation for LUC shares (Figure 12).

5.3. The Reliability and Mechanisms of Our Approach

Although we found that most LUCC products can not reflect the actual temporal change trends, it does not mean that these datasets are wrong. We need to correctly explain the temporal patterns of the satellite-based products. Due to the limitation of pixel mixture, the actual changes of individual LUC types within a pixel are difficult to detect especially for the satellite images with coarser resolution. For example, forest is often defined as 20% tree cover within a certain area. Then if the tree cover is 30%, then this pixel is classified as forest; however, the left 70% area within this pixel is difficult to attribute to other LUC types. This further results in the failure to accurately track the temporal change of forest area within each pixel since either 30% or 100% tree coverage is regarded as forest, and then the final classified product showed no temporal changes of LUC types within this pixel. The pixel mixture issue can be partially solved for images with higher resolution, e.g., the spatial resolution lower than 5 m; however, there are few regional or global LUCC products at such high resolution. At current stage, we have to apply statistical approaches and combine the different characteristics of existing LUCC products to uncover the elements (various LUC types) within a blackbox (pixel). This is the reason why our study developed the approach to effectively integrate above information, and we have achieved this goal by successfully matching the actual temporal changes of all LUC types and more detailed representation of LUCC at subpixel level. The strengths of each LUCC dataset are integrated through the fusion, thereby enhancing the overall utility of the analysis.

In the hierarchical procedure of cropland dataset development, we used the resolution and distance weights to represent two distinct scales and assigned weights for each LUCC product, and

then a decision tree method was applied to determine the probability of being cropland through iteratively changing the threshold values. Our approach has combined the hierarchical methods from two previous studies   [20,51], and can integrate the strengths of all cropland products. The RF and other machine-learning methods have been widely used to fuse multiple LUCC datasets [17,32,52]. Our approach for developing forest dataset based on RF method was different from the previous method. The regression models were first fitted between annual mean NDVImax and statistical forest share at provincial scale. Then the RF regressor was applied to fit the regression model parameters through iteratively revising the parameter values until the calculated provincial forest share was close to the statistical data. Our approach can fit the models at pixel level and no need for training the RF algorithm using the sampling plot data since the pixel-level forest share data are difficult to obtain. For the wetland dataset development, a regular decision-tree method was applied. The probability of wetland at pixel level was first calculated based on multiple wetland datasets. Then different probability threshold values were applied to iteratively run the decision-making process until the wetland area was close to the statistical data. Most studies only applied the same method to fuse multiple LUCC datasets for all LUC types; however, the controlling factors for the changes of various LUC types are significantly different. Therefore, it is reasonable to apply different development methods. The comparison with visually-interpreted plot data indicated that our approaches can accurately reflect share changes of all LUC types within pixels, indicating that our approach is reliable.

5.3. Uncertainties and Outlooks

In this study, there are several uncertainties that could affect the accuracy of our approach. First, the provincial statistical/inventory data were used to constrain the spatiotemporal patterns of LUCC shares. In fact, the change patterns of LUCC shares could vary significantly even among pixels [41]; therefore the low spatial resolution (provincial) of statistical data could cause large spatial location uncertainties within this province. County-level data should be more suitable for narrow down the spatial uncertainty. In addition, statistical data were mostly extrapolated from plot-level sampling data, the sampling and extrapolation methods could significantly affect the quality of the provincial statistical data. In addition, there are several sources of statistical data and their magnitudes are not very different, especially for the grassland and wetland datasets. Although we chose the most official data sources, it inevitable brought either under- or over-estimation of LUCC shares.

Second, our fusion approach was based on multiple LUCC products. These datasets were derived from different methods and researchers, and have different spatial and temporal resolutions, which could cause large uncertainty and bias in calculating the probability or mapping the distribution boundary. In addition, the uncertainty in the NDVImax dataset could bring uncertainty in the developed forest datasets. Third, we only have a few sampling plots for evaluation of the LUC share at pixel level due to difficulty in available high-resolution images and digitizing work. The fewer validation plots and the possible visual interpretation errors could cause the some uncertainties for the results. To reduce the uncertainties, we will further expand the training and validation sampling plots, collect county-level statistical data as a constraint, and improve the fitting mechanisms for various LUC types in the near future.

From our intercomparisons, we found the temporal and spatial patterns were significantly different among the existing LUCC products. Many previous studies have applied these geospatial LUCC products to estimate carbon fluxes and other socioeconomic and ecological services under the background of LUCC. For example, Yu et al. [53] applied multiple LUCC datasets and a process-based model to estimate China's carbon sink due to LUCC; Chang et al. [54] assessed China's carbon stock based on the NLCD dataset; Li et al.   [55] also applied the NLCD dataset and a simple model to reexamine China's terrestrial ecosystem carbon balance. These direct applications of the LUCC products could result in large uncertainties in their results and conclusions. Through the comparisons, Yu et al. [53] also revealed that the carbon sink by using their own statistically-extrapolated LUCC product was significantly higher than that using multiple existing LUCC datasets. To avoid inappropriate assessments of the LUCC effects, new LUCC share datasets for the entire China and the globe are needed. Our approach has the prospect to be further applied to reconstruct the new LUCC datasets at both national and global scale.

6. Conclusions

To accurately and comprehensively reconstruct a LUCC dataset, our study introduced a simple approach. This approach took advantage of the strengths of existing LUCC products and integrated them with the statistical data using the regular decision tree, RF and hierarchical extrapolation methods. The share for each LUC type at subpixel level was reconstructed in sequence using different method and procedures. The accuracy evaluations for the location, share and changes of LUC types indicated that our approach can accurately match the temporal and spatial changes of all LUC types using the NEC as a case study region. Further comparisons with existing LUCC products also confirmed the higher accuracy of our developed LUCC dataset. Considering the inadequacy and inconsistency of most existing LUCC products in representations of the actual LUCC conditions, many previous studies may have significantly under- or over-estimate the LUCC consequences. The researchers need to be cautious when choose the existing LUCC datasets for studying the LUCC effects. And the most important and urgent thing is to reconstruct new LUCC share datasets with high spatial and temporal accuracy for all LUC types at national and global scale, and our approach has the prospect to do this work.

References

1. Findell, K.L.; Berg, A.; Gentine, P.; Krasting, J.P.; Lintner, B.R.; Malyshev, S.; Santanello Jr, J.A.; Shevliakova, E. The impact of anthropogenic land use and land cover change on regional climate extremes. *Nature Communications* **2017**, *8*, 989.

2. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K. Global consequences of land use. *Science* **2005**, *309*, 570-574.

3. Gibbard, S.; Caldeira, K.; Bala, G.; Phillips, T.J.; Wickett, M. Climate effects of global land cover change. *Geophys Res Lett* **2005**, *32*, 024550.

4. Chen, Y.; Ge, Y.; Heuvelink, G.B.; An, R.; Chen, Y. Object-based superresolution land-cover mapping from remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *56*, 328-340.

5. Cihlar, J. Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing* **2000**, *21*, 1093-1114.

6. 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* **2000**, *21*, 1303-1330.

7. Bartholome, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing* **2005**, *26*, 1959-1977.

8. Friedl, M.A.; Sulla-Menashe, 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.

9. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M. Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *103*, 7-27.

10. Chen, B.; Xu, B.; Zhu, Z.; Yuan, C.; Suen, H.P.; Guo, J.; Xu, N.; Li, W.; Zhao, Y.; Yang, J. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin* **2019**, *64*, 370-373.

11. 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.

12. Harper, K.L.; Lamarche, C.; Hartley, A.; Peylin, P.; Ottlé, C.; Bastrikov, V.; Martín, D.S.; Bohnenstengel, S.I. A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models. *Earth Syst Sci Data* **2023**, *15*, 1465–1499.

13. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. In *Global land use/land cover with sentinel 2 and deep learning*, 2021 IEEE international geoscience and remote sensing symposium IGARSS, 2021; IEEE: pp 4704-4707.

14. Ning, J.; Liu, J.; Kuang, W.; Xu, X.; Zhang, S.; Yan, C.; Li, R.; Wu, S.; Hu, Y.; Du, G. Spatiotemporal patterns and characteristics of land-use change in china during 2010–2015. *Journal of Geographical Sciences* **2018**, *28*, 547-562.

15. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W.; Zhang, S.; Li, R.; Yan, C. Spatiotemporal characteristics, patterns, and causes of land-use changes in china since the late 1980s. *Journal of Geographical sciences* **2014**, *24*, 195-210.

16. Xu, Y.; Yu, L.; Peng, D.; Zhao, J.; Cheng, Y.; Liu, X.; Li, W.; Meng, R.; Xu, X.; Gong, P. Annual 30-m land use/land cover maps of china for 1980–2015 from the integration of AVHRR, modis and landsat data using the BFAST algorithm. *Science China Earth Sciences* **2020**, *63*, 1390-1407.

17. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst Sci Data* **2021**, *13*, 3907-3925.

18. Xia, X.; Xia, X.; Chen, X.; Fan, L.; Liu, S.; Qin, Y.; Qin, Z.; Xiao, X.; Xu, W.; Yue, C., et al. Reconstructing long-term forest cover in China by fusing national forest inventory and 20 land use and land cover data sets. *Journal of Geophysical Research: Biogeosciences* **2023**, *128*, e2022JG007101.

19. Qin, Y.; Xiao, X.; Dong, J.; Zhang, G.; Shimada, M.; Liu, J. Forest cover maps of China in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *109*, 1-16.

20. Yu, Z.; Jin, X.; Miao, L.; Yang, X. A historical reconstruction of cropland in China from 1900 to 2016. *Earth Syst Sci Data* **2021**, *13*, 3203-3218.

21. Niu, Z.; Zhang, H.; Wang, X.; Yao, W.; Zhou, D.; Zhao, K.; Zhao, H.; Li, N.; Huang, H.; Li, C., et al. Mapping wetland changes in china between 1978 and 2008. *China Science Bulletin* **2012**, *57*, 2813–2823.

22. Gong, P.; Niu, Z.; Cheng, X.; Zhao, K.; Zhou, D.; Guo, J.; Liang, L.; Wang, X.; Li, D. China's wetland change (1990-2000) determined by remote sensing. *Science China Earth Sciences* **2010**, *53*, 1036–1042.

23. Xia, X.; Xia, J.; Chen, X.; Fan, L.; Liu, S.; Qin, Y.; Qin, Z.; Xiao, X.; Xu, W.; Yue, C. Reconstructing long-term forest cover in China by fusing national forest inventory and 20 land use and land cover data sets. *Journal of Geophysical Research: Biogeosciences* **2023**, *128*, 007101.

24. Mao, D.; Wang, Z.; Du, B.; Li, L.; Tian, Y.; Jia, M.; Zeng, Y.; Song, K.; Jiang, M.; Wang, Y. National wetland mapping in china: A new product resulting from object-based and hierarchical classification of landsat 8 oli images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *164*, 11-25.

25. Yu, Z.; Lu, C. Historical cropland expansion and abandonment in the continental US during 1850 to 2016. *Global Ecology and Biogeography* **2018**, *27*, 322-333.

26. Gong, P.; Li, X.c.; Wang, J.; Bai, Y.q.; Chen, B.; Hu, T.; Liu, X.; Xu, B.; Yang, J.; Zhang, W. Annual maps of global artificial impervious area (gaia) between 1985 and 2018. *Remote Sensing of Environment* **2020**, *236*, 111510.

27. Yang, J.; Dong, J.; Xiao, X.; Dai, J.; Wu, C.; Xia, J.; Zhao, G.; Zhao, M.; Li, Z.; Zhang, Y., et al. Divergent shifts in peak photosynthesis timing of temperate and alpine grasslands in China. *Remote Sensing of Environment* **2019**, *233*, 111395.

28. Li, H.; Cao, Y.; Xiao, J.; Yuan, Z.; Bai, X.; Wu, Y.; Liu, Y. A daily gap-free normalized difference vegetation index dataset from 1981 to 2023 in China. *Scientific Data* **2024**, *11*, 527.

29. Xu, X. A 10m year-by-year ndvi maximum dataset for China. *Resource and environmental science data registration and publication system* **2022**, *http://www.resdc.cn/*, DOI:10.12078/2022030801.

30. Mao, D.; He, X.; Wang, Z.; Tian, Y.; Zheng, H. Diverse policies leading to contrasting impacts on land cover and ecosystem services in northeast China. *Journal of Cleaner Production* **2019**, *240*, 117961.

31. Li, C.; Gong, P.; Wang, J.; Zhu, Z.; Biging, G.S.; Yuan, C.; Hu, T.; Zhang, H.; Wang, Q.; Li, X., et al. The first all-season sample set for mapping global land cover with Landsat-8 data. *Science Bulletin* **2017**, *62*, 508-515.

32. 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 Syst Sci Data* **2020**, *12*, 1217-1243.

33. Wang, Y.; Shen, X.; Lü, X. Change characteristics of landscape pattern and climate in marsh areas of northeast China during 1980−2015. *Earth and Environment* **2020**, *48*, 348-357.

34. Mao, D.; Wang, Z.; Luo, L.; Ren, C.; Jia, M. Monitoring the evolution of wetland ecosystem pattern in northeast China from 1990 to 2013 based on remote sensing. *Journal of Natural Resources* **2016**, *31*, 1253-1263.

35. Ye, Y.; Fang, X.Q. Spatial pattern of land cover changes across northeast China over the past 300 year. *Journal of Historical Geography* **2011**, *37*, 408–417.

36. Xu, Y.; Yu, L.; Peng, D.; Zhao, J.; Cheng, Y.; Liu, X.; Li, W.; Meng, R.; Xu, X.; Gong, P. Annual 30-m land use/land cover maps of China for 1980–2015 from the integration of AVHRR, MODIS and Landsat data using the BFAST algorithm. *Science China Earth Sciences* **2020**, *63*, 1390–1407.

37. Cao, B.; Yu, L.; Li, X.; Chen, M.; Li, X.; Hao, P.; Gong, P. A 1 km global cropland dataset from 10000 bce to 2100 ce. *Earth Syst Sci Data* **2021**, *13*, 5403-5421.

38. Potapov, P.; Hansen, M.C.; Pickens, A.; Hernandez-Serna, A.; Tyukavina, A.; Turubanova, S.; Zalles, V.; Li, X.; Khan, A.; Stolle, F., et al. The global 2000-2020 land cover and land use change dataset derived from the Landsat archive: First results. *Frontiers in Remote Sensing* **2022**, *3*, 856903.

39. Song, X.P.; Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Tyukavina, A.; Vermote, E.F.; Townshend, J.R. Global land change from 1982 to 2016. *Nature* **2018**, *560*, 639–643.

40. You, N.; Dong, J.; Huang, J.; Du, G.; Zhang, G.; He, Y.; Yang, T.; Di, Y.; Xiao, X. The 10-m crop type maps in northeast China during 2017–2019. *Scientific Data* **2021**, *8*, 1-11.

41. Zhang, C.; Dong, J.; Ge, Q. Mapping 20 years of irrigated croplands in China using modis and statistics and existing irrigation products. *Scientific Data* **2022**, *9*, 407.

42. Mao, D.; Luo, L.; Wang, Z.; Wilson, M.C.; Zeng, Y.; Wu, B.; Wu, J. Conversions between natural wetlands and farmland in China: A multiscale geospatial analysis. *Science of the Total Environment* **2018**, *634*, 550–560.

43. Luo, H.; Huang, F.; Zhang, Y. Space-time change of marsh wetland in Liaohe Delta area and its ecological effect. *Journal of Northeast Normal University* **2003**, *35*, 100-105.

44. Asselen, S.; Verburg, P.H.; Vermaat, J.E.; Janse, J.H. Drivers of wetland con-version: A global meta-analysis. *PloS one* **2013**, *8*, 381292.

45. Stehman, S. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* **1997**, *62*, 77-89.

46. Zheng, X.; Zhu, J.J. Estimation of shelter forest area in three-north shelter forest program region based on multi-sensor remote sensing data. *Chinese Journal of Applied Ecology* **2013**, *24*, 2257-2264.

47. Zhu, J.J.; Zheng, X. The prospects of development of the three-north afforestation program(TNAP) : On the basis of the results of the 40-year construction general assessment of the TNAP. *Chinese Journal of Ecology* **2019**, *38*, 1600-1610.

48. Liu, Z.; Wang, W.J.; Ballantyne, A.; He, H.S.; Wang, X.; Liu, S.; Ciais, P.; Wimberly, M.C.; Piao, S.; Yu, K., et al. Forest disturbance decreased in China from 1986 to 2020 despite regional variations. *Communications Earth and Environment* **2023**, *4*, 15.

49. Chen, C.; Park, T.; Wang, X.; Piao, S.; Xu, B.; Chaturvedi, R.K.; Fuchs, R.; Brovkin, V.; Ciais, P.; Fensholt, R., et al. China and India lead in greening of the world through land-use management. *Nature sustainability* **2019**, *2*, 122-129.

50. Zhu, Z.C.; Piao, S.L.; Myneni, R.B.; Huang, M.T.; Zeng, Z.Z.; Canadell, J.G.; Ciais, P.; Sitch, S.; Friedlingstein, P.; Arneth, A., et al. Greening of the earth and its drivers. *Nat Clim Change* **2016**, *6*, 791–795.

51. Zhang, C.; Dong, J.; Ge, Q. Mapping 20 years of irrigated croplands in China using MODIS and statistics and existing irrigation products. *Scientific Data* **2022**, *9*, 407.

52. Li, K.; Wang, J. A multi-source data fusion method for land cover production: A case study of the east European plain. *International Journal of Digital Earth* **2024**, *17*, 2339360

53. Yu, Z.; Ciais, P.; Piao, S.; Houghton, R.A.; Lu, C.; Tian, H. Forest expansion dominates China's land carbon sink since 1980. *Nature Communications* **2022**, *13*, 5374

54. Chang, X.; Xing, Y.; Wang, J.; Yang, H.; Gong, W. Effects of land use and cover change (LUCC) on terrestrial carbon stocks in China between 2000 and 2018. *Resources, Conservation and Recycling* **2022**, *182*, 106333.

55. Li, J.; Guo, X.; Chuai, X.; Xie, F.; Yang, F.; Gao, R.; Ji, X. Reexamine China's terrestrial ecosystem carbon balance under land use-type and climate change. *Land Use Policy* **2021**, *102*, 105275.