


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: lrp@iaesy.cn (R.L.); chengu1@zafu.edu.cn (G.C.)

Abstract: Land use and cover change (LUCC) is a key factor influencing global environmental and socioeconomic 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 cannot accurately reflect the spatiotemporal change patterns of LUCC at the 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 the spatial and temporal changes of the major LUC types by applying machine-learning, regular decision-tree and hierarchical assignment methods using 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 NEC, with an overall accuracy of 82%, Kappa coefficient of 0.77 and regression coefficient of 0.82. Further comparisons with existing LUCC datasets and statistical data also indicated the accuracy of our approach and datasets. Our approach unfolded the mixed-pixel issue of LUC types and integrated the strengths of existing LUCC products through multiple fusion processes. The analysis based on our developed dataset indicated that forest, cropland and built-up land area increased by 17.11×10^4 km², 15.19×10^4 km² and 2.85×10^4 km², respectively, during 1980–2020, while grassland, wetland, shrubland and bare land decreased by 26.06×10^4 km², 4.24×10^4 km², 3.97×10^4 km² and 0.92×10^4 km², respectively, in NEC. Our developed approach accurately reconstructed the shares and spatiotemporal patterns of all LUC types during 1980–2020 in NEC. This approach can be further applied to the entirety of China, and worldwide, and our products can provide accurate data supports for studying LUCC consequences and making effective land use policies.

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



Citation: Chen, Y.; Li, R.; Tu, Y.; Lu, X.; Chen, G. Comprehensive Representations of Subpixel Land Use and Cover Shares by Fusing Multiple Geospatial Datasets and Statistical Data with Machine-Learning Methods. *Land* **2024**, *13*, 1814. <https://doi.org/10.3390/land13111814>

Academic Editors: Kathryn Sheffield and Mohammad Abuzar

Received: 1 October 2024

Revised: 29 October 2024

Accepted: 30 October 2024

Published: 1 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

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 recent decades at regional, national and global scales [4,5], such as the Global Land Cover map (DISCover) for 1992 [6], 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], the European Space Agency Climate Change Initiative (ESA-CCI) land cover product for 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 data accuracy, 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], the annual China Land use/land cover datasets (CLUD-A) [16] and the China Land Cover Dataset (CLCD) [17]. Although these datasets have been validated with high accuracy, intercomparisons indicate that there is a large discrepancy among these datasets, and none of the spatiotemporal patterns of these datasets match well with 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 the present, and none of the datasets match the temporal change trends of the national statistical data. For example, Qin et al. [19] compared several LUCC products and indicated that the forest area of five datasets ranged from $174 \times 10^4 \text{ km}^2$ to $227 \times 10^4 \text{ km}^2$ in 2010. Yang and Huang [17] reported that forest area in China only increased by 4.34% during 1980–2019, significantly lower than the national forest inventory (NFI), which revealed a 77% increase from the period 1984–1988 (12.98% forest coverage) to the period 2014–2018 (22.96%). In addition, 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 datasets changed less than $\pm 10\%$ during 1980–2020, while reports have indicated that China's wetland area has been significantly reduced by 33% [21,22]. In addition, most of these existing datasets were only targeted at a single LUC type, while few studies have comprehensively addressed the spatiotemporal patterns for all LUC types. Many previous studies have applied these existing LUCC products to analyze or explain the spatiotemporal changes of vegetation cover, growth and ecological impacts. For example, Song et al. [23] estimated the canopy cover change for various LUC types at national, continental and global scales. Based on the LUCC products and other assisting data, Chen et al. [24] and Zhu et al. [25] stated that the Earth is becoming greener, partially driven by the vegetation recovery in China. In addition, Yu et al. [26], Chang et al. [27] and Li et al. [28] have either directly or indirectly applied these geospatial LUCC datasets to extrapolate or model the carbon balance in China. The results of all these previous studies heavily relied on the quality of the LUCC products and highlighted the importance of LUCC data quality [26]. 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. [18] reconstructed a new forest cover dataset (CFCD) from 1980 to 2015 by combining several existing LUCC datasets and NFI data; however, this approach matched only the temporal change patterns and sacrificed the spatial accuracy. To match statistical cropland area and change trends, Yu et al. [20] developed a subpixel-level cropland share dataset; however, this dataset was only targeted at cropland areas 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 pixel-based classification methods [20,29]. In the pixel-based approach, each pixel is treated as a binary value (either Boolean 0 or 1), i.e., each grid cell is completely occupied by a single LUC type [20]. This approach is more suitable for high-resolution images [30]. A small percentage of the pixels could be ignored based on this approach, resulting in an underestimation for temporal changes. Using forests as example, forest area in China is defined as tree coverage greater than 20% within a minimum area of 0.5 ha. Pixels with tree coverage ranging from 20% to 100% are regarded as forest area, which results in a failure to reflect the change of pixel-level tree coverage in the LUCC products and an underestimation of forest-area increase caused when the tree coverage increases from 20% 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. [31] and Niu et al. [21] only targeted matching forest, cropland, urban and wetland areas 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 [32–34]. Based on the existing vegetation indices and LUCC products, it is possible to invert the real changes in LUC shares within pixels.

Northeastern China (NEC) covers about 15.3% of China's territory. It is the main base for cereal and wood production in China and has the largest wetland area compared with other regions. With rapid transitions of socioeconomic environments, 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 to LUCC products. Through comparisons with existing LUCC products, we found that no LUCC product comprehensively catches the actual changes in LUC areas in NEC for 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 NEC. The objectives of this study are to (1) construct an approach for tracking the changes of fractional shares of various LUC types by integrating multiple LUCC products and other geospatial datasets with statistical data using machine-learning and regular decision-tree methods; (2) evaluate the performance of this approach using NEC as a case study area; (3) analyze the spatiotemporal patterns of LUCC in NEC.

2. Study Area and Data Descriptions

2.1. Study Area

Northeastern China (NEC) is located between 110°–135° E and between 38°–54° N and has a total land area of 1.47×10^6 km² (Figure 1). 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 [35]. 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 °C, and mean annual precipitation is about 785 mm. Mean annual temperatures increased by about 1.36 °C during 1980–2020. Precipitation mostly occurs 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 it has greatly shrunk during recent decades [36]; it also has the largest area of natural forest. A large portion of western NEC belongs to the Three-North Shelterbelt Forest (TNSF) afforestation project region; thus, large forest areas have been planted since 1978. NEC is also the most important wood and crop production base in China. With the above conditions, rapid urbanization, increasing population and economic development, LUC areas and types have been dramatically changed.

2.2. Data Descriptions

2.2.1. Statistical Data

To develop an accurate LUCC area in NEC, this study collected inventory statistical data at the provincial level. The land area of each land cover type for 1980–2020 was collected from the statistical yearbook data from each province. The annual cropland area and type data for 1980–2020 were retrieved from the provincial statistical yearbook (e.g., the data for Jilin Province are from <http://tjj.jl.gov.cn/tjsj/tjnj/>; accessed on 25 July 2024). The forest area data for every 5 years for 1980–2020 were obtained from the National Forest Inventory (NFI) of China (<https://www.stgz.org.cn/lbdbgzpt/>; accessed on 25 July

2024). The 5-year NFI data were further linearly interpolated to annual data. Forest land from the NFI was defined as tree coverage greater or equal to 20%, which includes tall forests, bamboo stands, dense shrubland (shrub coverage is greater than 30% in the arid and semi-arid region), tree nurseries and cleared forest areas 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 were linearly fitted based on annual area and recalculated based on these fitted lines.

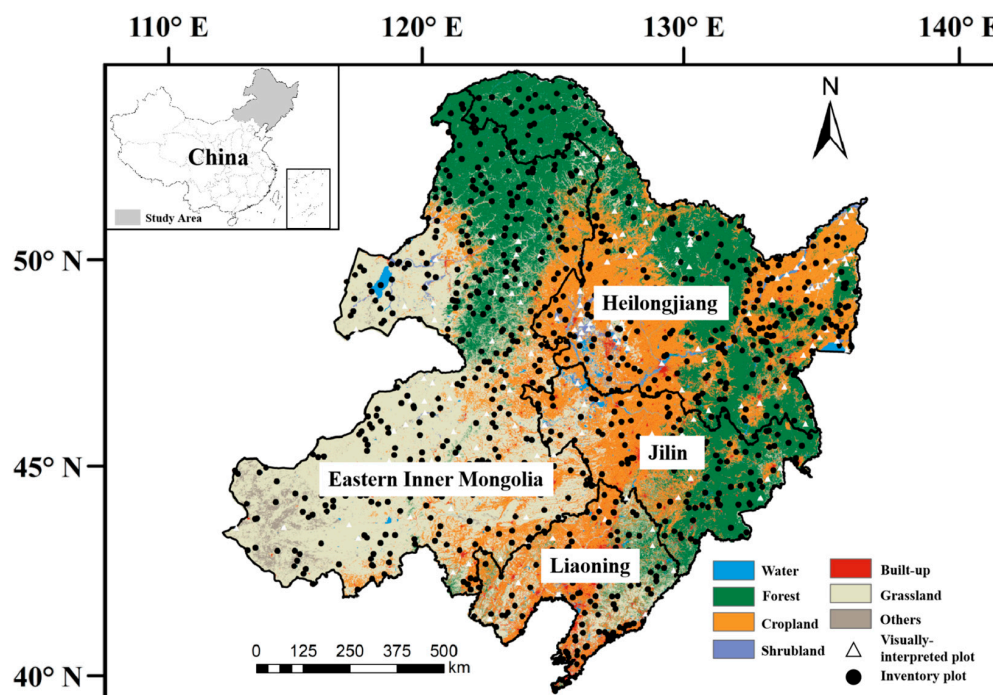


Figure 1. The study area and the distribution of major land use and cover types (Source: [17]) and the training and validation sample plots. Note: the triangle points are visually-interpreted plots at a 1 km spatial resolution; the dark circle points are field-investigated plots at a 30 m resolution (Source: [37]).

2.2.2. Geospatial Datasets

Many geospatial data were collected to assist in 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 (<https://livingatlas.arcgis.com/landcover/>, accessed on 25 July 2024), CLUDA [16], NLCD (<http://www.nesdc.org.cn/>, accessed on 25 July 2024), GLASS-GLC [38], and MODIS [8]. We clipped NEC's area from these national or global level LUCC products. We also collected some literature data (nonspatial) for the evaluation of our approach, including the Wang_wetland [36], Mao_wetland [39] and Ye_grassland [40] data. Based on the NDVI data from AVHRR from 1982 to 1999 (GIMMS NDVI) and MODIS from 2000 to 2020, previous researchers developed multiple NDVI datasets at 30 m, 1 km and 0.05° spatial resolutions for the period 1981–2020 [32–34]. We first aggregated the 30 m NDVI dataset (1986–2020) to a 1 km resolution for the period 1986–2020 and applied the change trends of the 0.05° NDVI dataset to extend the 1 km NDVI data for the period 1980–1985. These long-term NDVI data were used to extrapolate the change trend of land shares. All the 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/ (accessed on 25 July 2024)
FROM-GLC	10 m	2017	[10]
NLCD	30 m	1980, 1990, 1995, 2000, 2005, 2010, 2015, 2020	http://www.nesdc.org.cn/ (accessed on 25 July 2024)
MODIS	500 m	2000–2020	https://modis-land.gsfc.nasa.gov/landcover.html (accessed on 25 July 2024)
CLUDA	1 km	1980–2015	[16]
CLCD	30 m	1990–2020	[17]
GLASS-GLC	0.05°	1982–2015	[38]
GLC	1 km	1980–2100	[41]
Yu_cropland	5 km	1900–2016	[20]
Xia_forest (CFCD)	1 km	1980–2015	[18]
GLCLUC	30 m	2000–2020	[42]
GFC	0.05°	1982–2016	[23]
You_croptype	10 m	2017–2019	[43]
Mao_wetland	30 m	2015	[29,39]
NDVI	30 m	1986–2020	[34]
NDVI	0.05°	1981–2023	[33]

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 [37] to evaluate the performance of our approach in generating the LUCC products. In total, 2453 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 the 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 where at least two periods of high-resolution images were available during 2005–2020. The boundaries of different LUC types were drawn by hand 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 shares were recorded. The developed LUC share datasets at the pixel level were finally compared with these sampling plot data, and thus the LUCC areas and spatial patterns 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 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 reconstructed the cropland, built-up land, forest, wetland, grassland, shrubland and bare land 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 for 1980–2020. This method was first proposed by [30] 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 [44]. Integrating existing land use products to produce new cropland maps involves a series of steps (Figure 2). First, the collected six cropland datasets underwent uniform preprocessing. Then, the appropriate weight order for these input products was determined. 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 was to label high-score cropland pixels. Each pixel was 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. The pixels with higher scores were summed up and 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 (reference area, RA), necessitating further selection of suitable pixels for allocation.

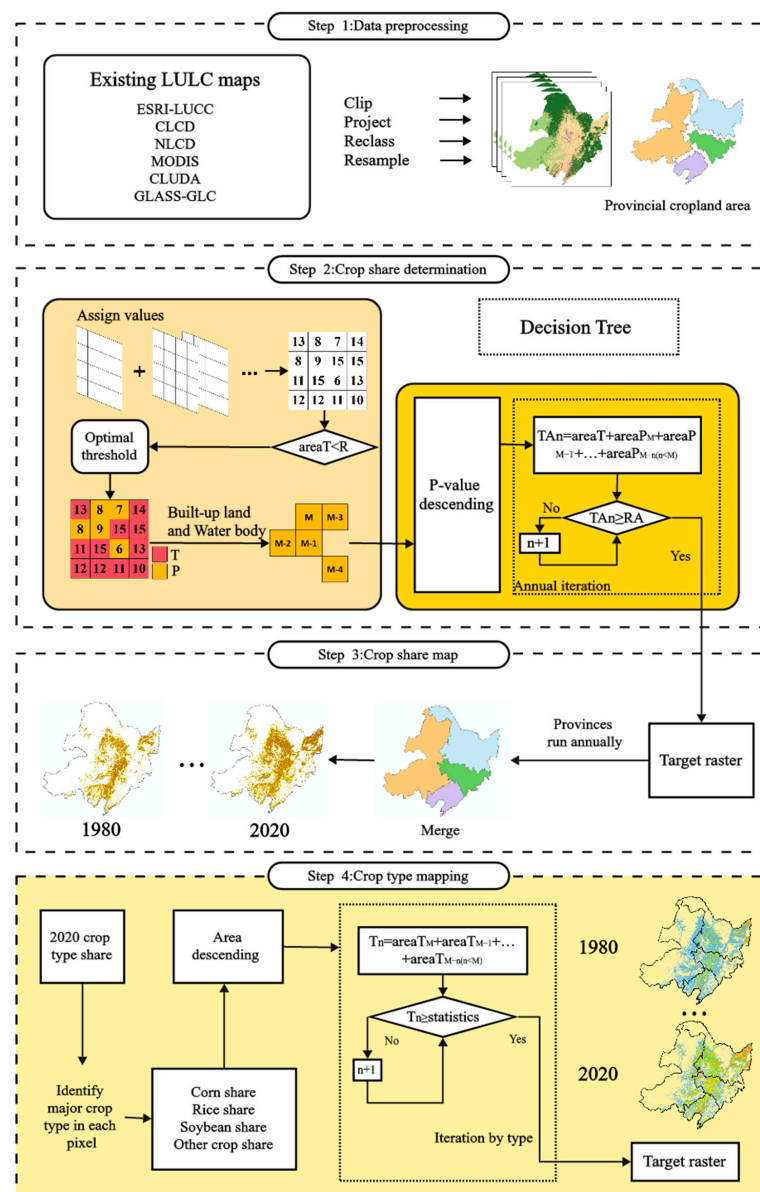


Figure 2. The dataset development procedure for fractional cropland area and crop types. Note: TA: summed total cropland area; RA: reference cropland area (provincial inventory data); m: iteration number; n: total iteration numbers; areaT: identified cropland area; areaP: possible cropland area.

The third step involved reallocating the remaining pixels. We needed to identify pixels that were 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 [44]. 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 body area. These two auxiliary weights were combined with the scoring weight to further determine the cropland assignment priority. All remaining cropland pixels were ranked in descending order based on the new combined weight values. The total area of the cropland pixels with the highest weight values was summed (TAn ; n is the total iteration numbers) to match the provincial area. If $area1$ was less than the RA , the cropland pixels with the second-highest weight value were counted, and their area ($areaP2$) was combined and compared with the RA . If TAn was still less than the RA , the iteration process turned to the third-highest weight and continued until the total area was closest to the RA . This iterative thresholding process was conducted in each province of NEC to obtain provincial cropland share maps. After the cropland distribution maps were developed, the main crops within pixels were further identified through a similar iterative approach based on the high-resolution (10 m) crop type distribution data [43].

3.2. Built-Up Land and Water Body Share

The CLUDA data [16] are represented as a fractional share for each LUC type within each pixel at a 1 km spatial resolution. Our built-up land and water body distribution data were slightly modified based on CLUDA because this dataset has been compared with many previous datasets and was consistent with the spatiotemporal change patterns of the inventory data. The pixel-level built-up land and water body areas were slightly reduced, in which the sum area for cropland, built-up land and water bodies was greater than 100%.

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 datasets of ESRI-LUCC (2017–2023), FROM-GLC (2017), CLCD, CLUDA, GLASS, NLCD and MODIS were used to obtain the maximum forest cover boundary. After a series of processes, such as reprojection, clipping, resampling, combination and aggregation, we generated the initial maximum forest distribution area for 1980–2020 at a 1 km resolution. The pixels with a forest cover share less than 10% were 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.

In the second step, the annual $NDVI_{max}$ dataset for 1980–2020 was reconstructed. The combined monthly NDVI dataset at a 1 km spatial resolution for 1980–2020 was 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 (MVCs) method to compose the monthly NDVI data and obtained the annual maximum NDVI dataset ($NDVI_{max}$). To remove the unreasonable annual variations, we applied the SG filter (Savitzky–Golay Filter) to smooth the abnormal interannual variations of $NDVI_{max}$ at the pixel level. The SG filter cannot guarantee a continuous change trend of $NDVI_{max}$ for 1980–2020; therefore, we further fitted regression models at the pixel level using the year as the independent variable. We found that the polynomial equation ($y = ax^2 + bx + c$) is the most suitable regression model for most pixels. Based on the fitted regression models, a new smooth-change annual $NDVI_{max}$ dataset for 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

NDVI_{max} data. We first fitted the regression between regional change trends of annual NDVI_{max} and inventory forest share for 1980–2020 using the polynomial equation $\text{Forestshare} = -22.792\text{NDVI}_{\text{max}}^2 + 34.77\text{NDVI}_{\text{max}} - 12.916$; 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 the spatial scale. The 5-year statistical provincial forest cover share change data were linearly interpolated to an annual scale for 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 were determined, they were applied to simulate the pixel-level forest share changes using the NDVI_{max} data as input.

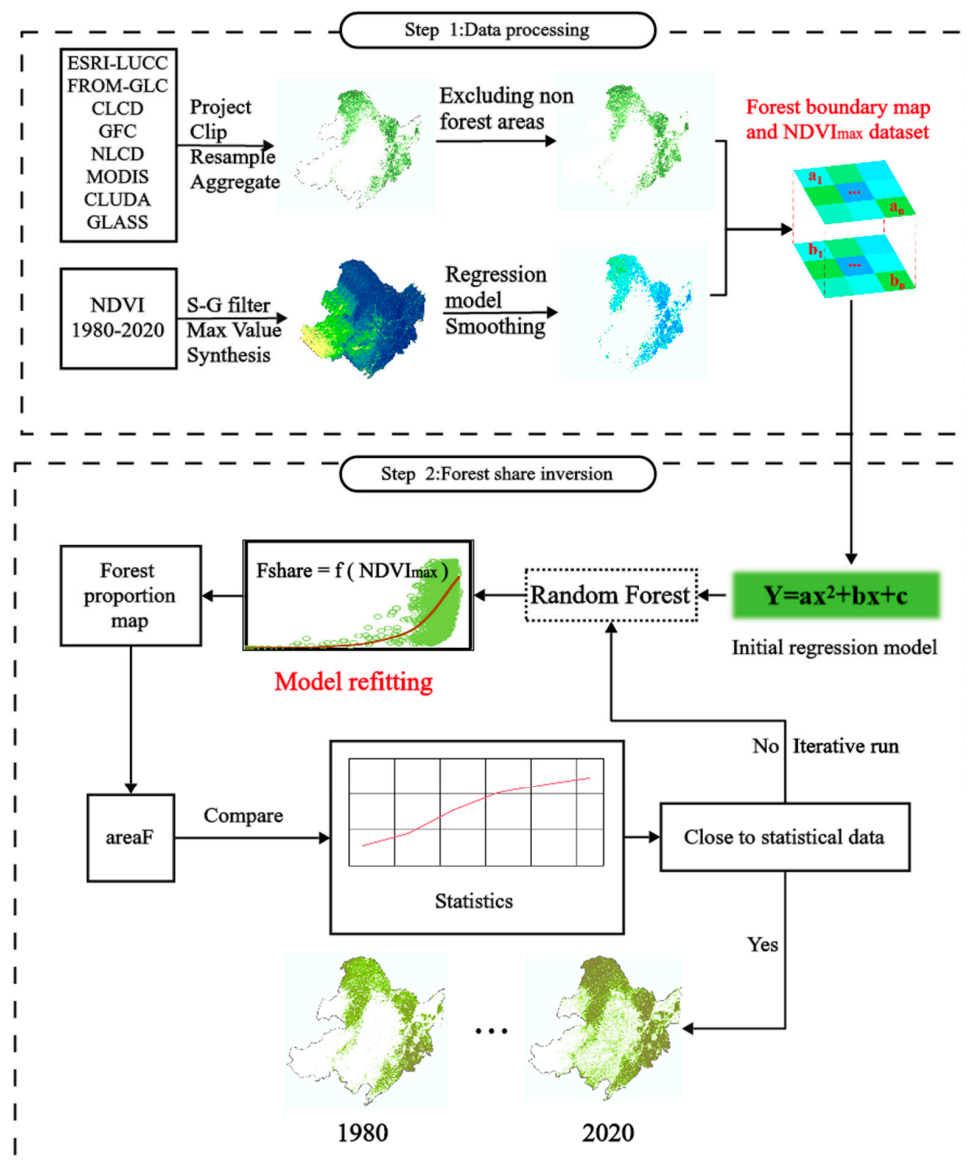


Figure 3. The forest share dataset development procedure. Note: areaF: the summed forest area after each iteration run; Fshare: the fitted forest share (%).

At the final step, the additional forest share in the pixels greater than 100% by summing up cropland, built-up area, water bodies and forest shares was subtracted.

3.4. Wetland Share Dataset Development

According to Niu et al. [21], China’s wetland area was reduced by 33% from 1978 to 2008. According to Wang et al. [36] and Mao et al. [39], the wetland area was 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 was reduced by 34% from 1980 to 2020. In addition, based on their study year points (1980, 1990, 2000, 2013, 2015, 2019), we finally applied the linear interpolation method to generate the annual wetland area for 1980–2020.

Before developing 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 [29] 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 are consistent with China’s wetland inventory data. This dataset was aggregated to a 1 km spatial resolution, and the wetland share was calculated. Finally, the maximum (potential) wetland boundary map and wetland share in 2020 were developed.

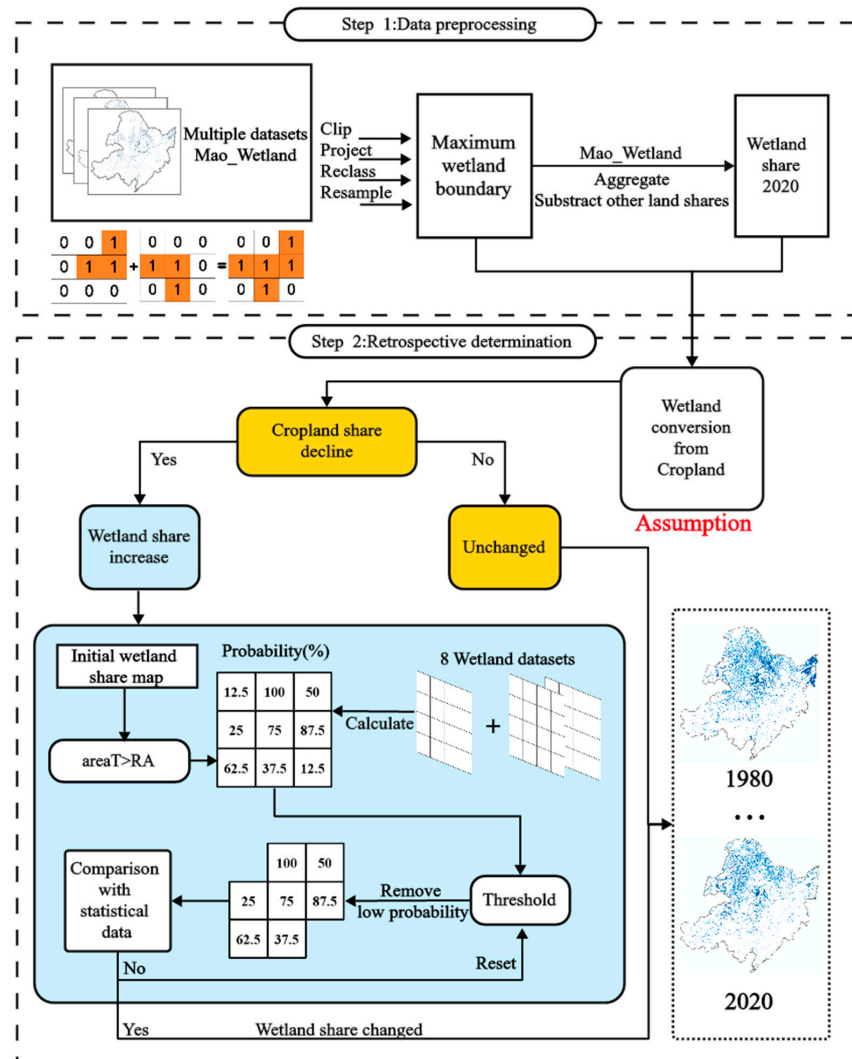


Figure 4. The wetland share dataset development procedure. Note: areaT: summed total wetland area; RA: reference wetland area (provincial inventory data).

The second step is to retrospect the wetland share for 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 the potential wetland area would be located in the history. Mao et al. [45], Wang et al. [36] and Luo et al. [46] noted that the reduction of wetland area in China was mostly due to conversion to cropland. Mao et al. [45] indicated that agricultural encroachment for food production is responsible for approximately 60% of natural wetland loss in China, 74.7% (11,778 km²) of which occurred from 1990 to 2000. In addition, the conversion of natural wetlands to cropland reached a maximum of 85.4% in NEC. Furthermore, in a meta-analysis, Asselen et al. [47] also indicated that most of global wetlands were converted to cropland. Therefore, we also assume that the decreasing wetlands in NEC are primarily due to cropland expansion. In all pixels of the maximum wetland distribution map, the increasing cropland area was assigned to the decreasing wetlands. 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 the inventory data. To be consistent with the inventory data, we needed to reduce the magnitude of the declining 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. By iteratively tuning the minimum score thresholds, we finally generated the wetland share dataset, which was temporally consistent with the statistical wetland area data.

3.5. Reconstruction of Grassland, Shrubland and Bare Land Shares

Before developing the geospatial grassland, shrubland, and bare land 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 (G_{max}), shrubland (S_{max}) and bare land (B_{max}) (Figure 5). Similar to above procedures, the 6 LUCC products after preprocessing were overlaid, and pixels with the presence of these LUC types were kept, while 0 was assigned for other pixels. Next, if multiple LUCC products had an LUC type in a pixel, then all their shares were kept. The share values of all products within a pixel were ranked by a declining order and finally formed 6 initial share maps for each LUC type except for bare land, for example, G_1, G_2, \dots, G_6 (grassland); S_1, \dots, S_6 (shrubland). For bare land, there was no need to derive multiple share maps since any area left was assigned to bare land.

In the second step, the first iteration was to run using the initial share data (G_1 and S_1). The first share datasets for grassland (G_{sh1}) and shrubland (S_{sh1}) were calculated based on the initial shares, with their ratios accounting for the remaining share (R). Based on the first share datasets, the total grassland and shrubland areas were 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 was implemented based on the G_2 and S_2 share datasets. Finally, we obtained the grassland and shrubland share datasets that were close to the statistical data. The left area ($R-G_{sh_i}-S_{sh_i}$) was assigned to bare land if it was present in the pixels.

3.6. Synthesis Among All LUC Share Datasets

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

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

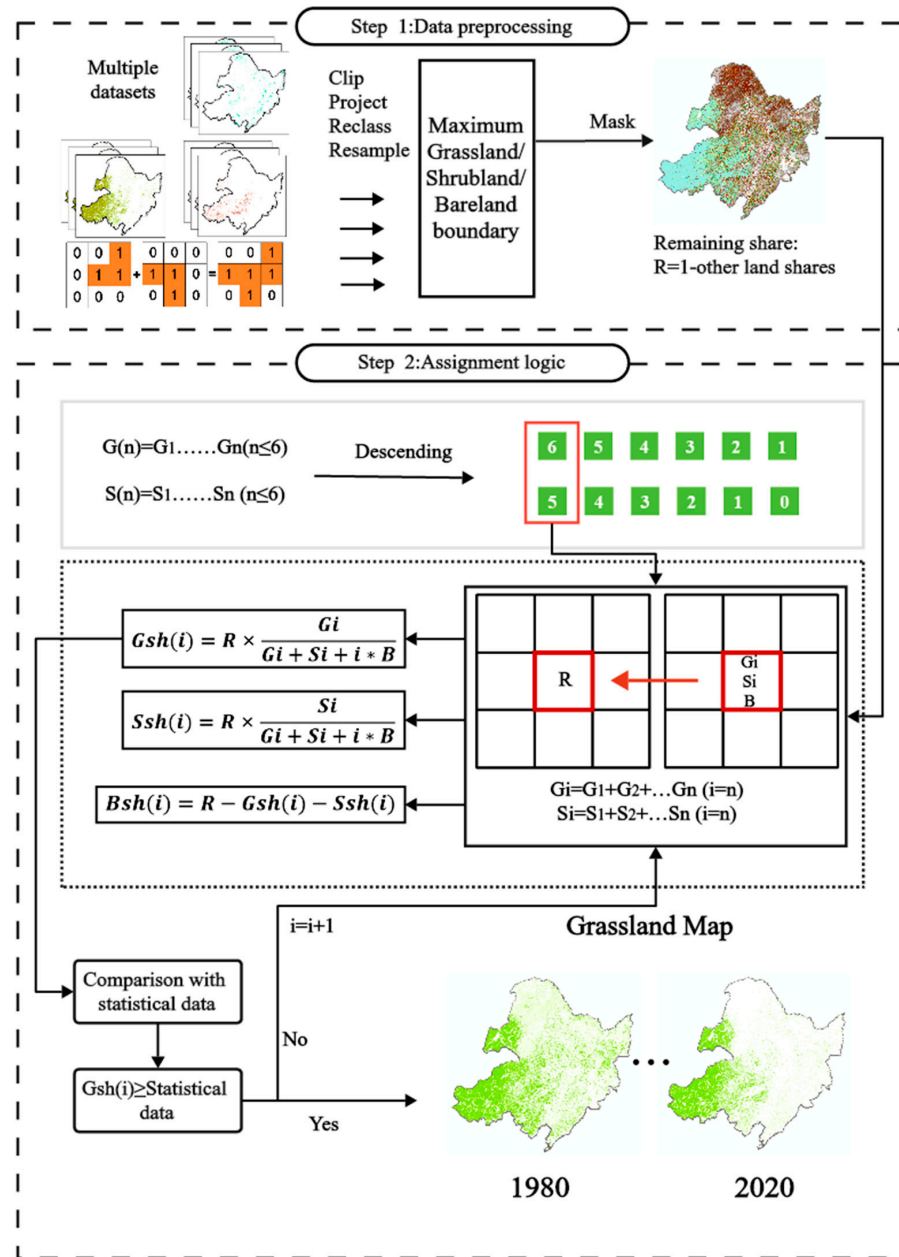


Figure 5. The grassland, shrubland and bare land share dataset development procedure. Note: R: the remaining area for allocation in each pixel; i: the iteration number; n: the existing dataset number; G_i , S_i and B_i : calculated relative fractions of grassland, shrubland and bare land, respectively, in each pixel in 2020 based on multiple existing LUCC products; Gsh, Ssh and Bsh: the generated grassland, shrubland and bare land shares.

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 [48].

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 a 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 pixel resolution. Therefore, the overall accuracy and Kappa coefficient indicate that our developed datasets can accurately reflect the true spatial distribution patterns of LUC types in NEC.

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

Land Cover	Cropland	Forest	Built-Up	Water	Wetland	Grassland	Shrubland	Total	Producer Accuracy	User Accuracy
Cropland	194	42	29	16	0	23	7	311	62%	86%
Forest	10	839	12	29	10	26	21	947	89%	89%
Built-up	2	8	338	4	0	0	0	352	96%	80%
Water	0	0	6	276	9	0	0	291	95%	76%
Wetland	2	8	5	9	107	21	0	152	70%	82%
Grassland	16	41	30	20	5	173	12	297	58%	70%
Shrubland	1	2	3	9	0	3	85	103	83%	68%
Total	225	940	423	363	131	246	125	2453	100%	100%

Overall accuracy = 82.02%; Kappa = 0.77.

Due to few available high-resolution images for NEC regions, we only visually-interpreted 65 sample plots for evaluating the performance of LUC share datasets at a 1 km spatial resolution. 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 match well with the visually-interpreted shares at the spatial scale (Figure 7).

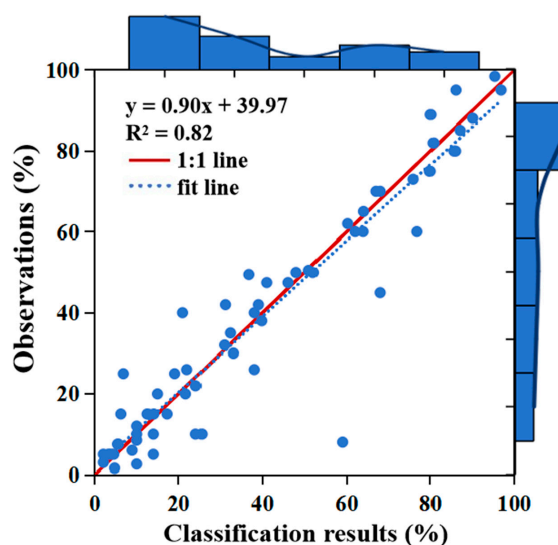


Figure 6. The correlations between the reconstructed and the visually-interpreted (observations) area shares (%) for different land use and cover types within each pixel at a 1 km spatial resolution.

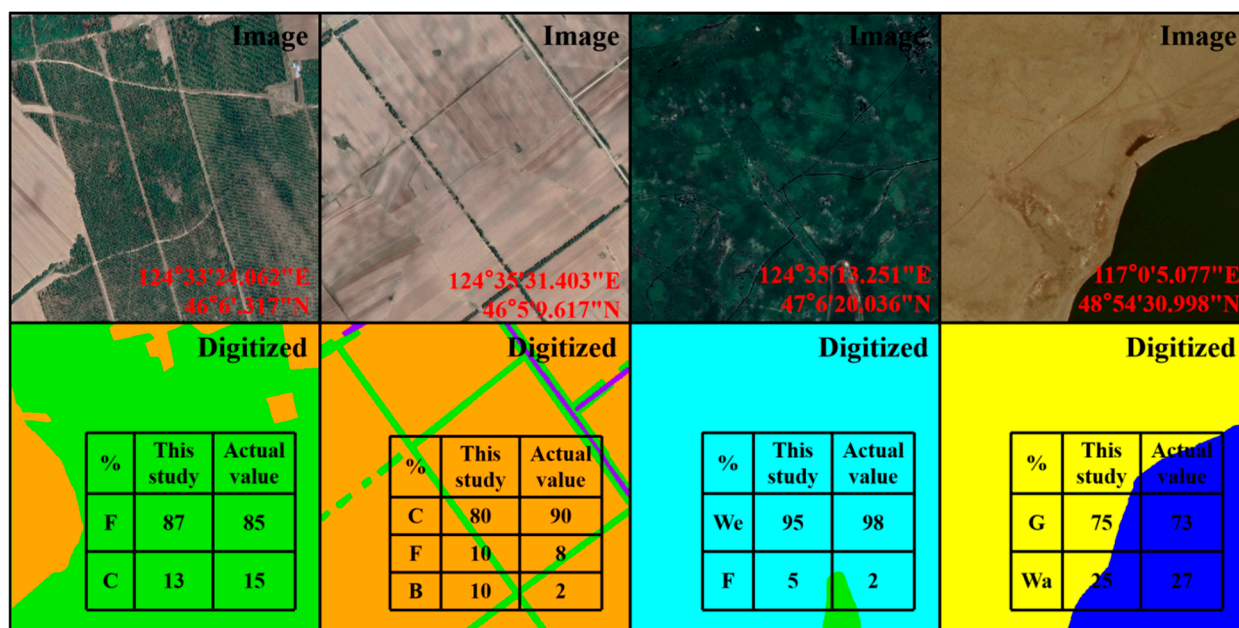


Figure 7. The evaluation of our developed land use and cover shares (%) within each pixel at a 1 km spatial resolution against visually-interpreted shares based on high-resolution images. Note: F: forest share (Green color region); B: built-up land share (Purple); C: cropland share (Orange); We: wetland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

4.2. Temporal Change Patterns of Different Land Use and Cover Types

Based on the reconstructed datasets, the temporal change patterns of various LUC areas were analyzed (Figure 8). The results indicated cropland, forest, and built-up land areas have increased by 72.36%, 47.42% and 150%, respectively, from 1980 to 2020, while wetland, grassland, shrubland and bare land have decreased by 39.04%, 43.67% and 68.73%, respectively. The largest increase of area was $1.71 \times 10^5 \text{ km}^2$ for forest due to the afforestation projects in NEC, and the largest reduction was $2.61 \times 10^5 \text{ km}^2$ for grassland. Cropland, built-up land and forest showed faster increases since 2000. The results from the LUC conversion matrix indicated that the increased cropland share from 1980 to 2020 was mostly owing to a conversion from grassland, followed by a conversion from wetland areas. The increased forest share primarily came from the conversions of grassland and shrubland, mainly owing to the world’s largest TNSF project. Wetland area was mainly converted to cropland due to reclamation. At the provincial level, Heilongjiang Province 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 reduced by 54.17% in Heilongjiang Province, where it has the highest wetland area, while it only reduced by 18.45% in Inner Mongolia. The grassland share decreased the most (86.83%) in Heilongjiang and the least (24.51%) in Inner Mongolia.

4.3. Spatial Change Patterns in Land Use and Cover Types

The LUC shares in 1980, 2000 and 2020 are displayed in Figure 9. The area of cropland is mainly distributed in central and northeastern NEC (Songnen Plain and Liao River Plain). Forests are mainly distributed in northwestern, north-central and southeastern NEC. Wetlands are mainly distributed in northwestern, north-central and northeastern NEC, surrounding the rivers and lakes. NEC has the largest wetland basins (Sanjiang Plain and Songnen Plain) in China. Grassland is mainly distributed in southwestern China and scattered in central NEC. Dense shrubland areas are scarce in NEC, mainly distributed in northern and central NEC.

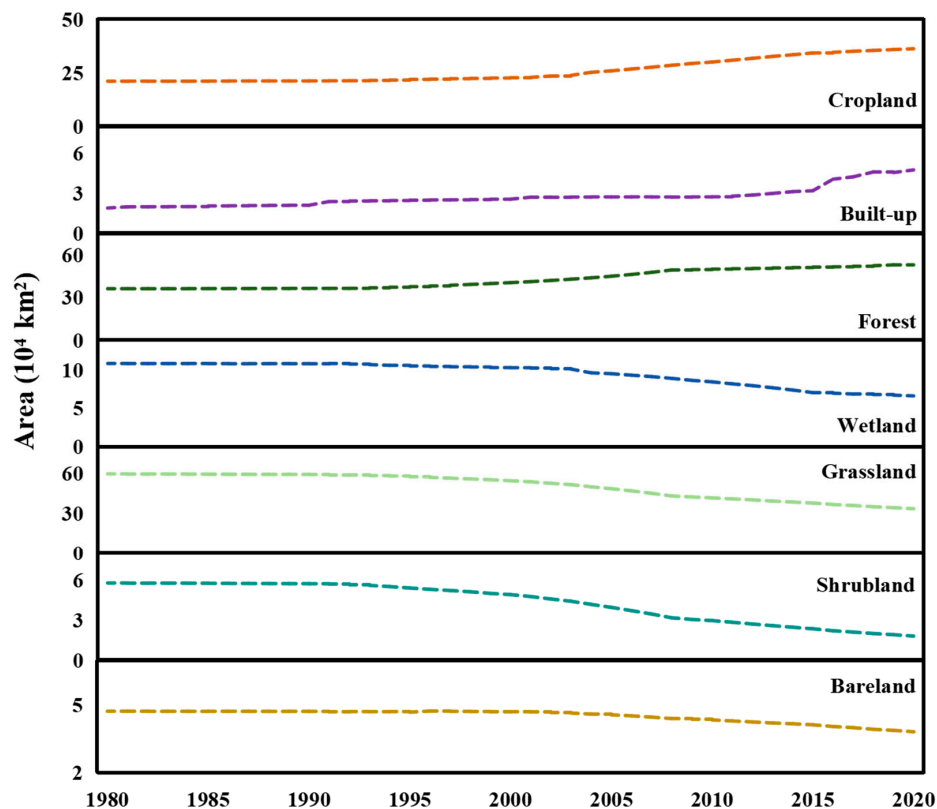


Figure 8. The total area (10^4 km²) of all land use and cover types for 1980–2020 in NEC.

The expanded cropland shares were mainly distributed in central and northeastern NEC from 1980 to 2020 (Figure 10). The expansion primarily occurred during 2000–2020. The expansion of cropland generally displayed 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 TNSF project and a stricter national forest protection policy during this period. The largest declines of wetland shares mainly occurred in northeastern (Sanjiang Plain) and central (Songnen Plain) NEC during 1980–2020, with more declines occurring during 2000–2020, owing to the rapid expansion of cropland. The decline of grassland was scattered throughout the entirety of NEC due to the expansions of forest and cropland, with smaller declines in the main distribution area in the southwest of Inner Mongolia. Shrubland areas experienced a greater decline during 1980–2020 as compared with other LUC types because these areas were often also suitable for forest plantations, and thus were converted to forest in the afforestation projects.

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 effectiveness 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 is significantly lower than the change trend of the statistical data (69.77%). For forest area, some LUCC products 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 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 the inventory data. The wetland area of all existing LUCC datasets (excluding Wang_wetland

and Mao_wetland datasets, which have been combined with the inventory data) showed slight decrease (<5%) or no change during 1980–2020. For grassland, our dataset showed a decline of 44% during 1980–2020, while the NLCD and GLASS-GLC datasets showed a decline of 20.70% and 18.03%, respectively, and other datasets showed a less than 10% decline. The inventory data from three of the National Land Surveys indicated that grassland area has declined from 50.01 km² to 32.70 km² at a rate of 24.62%, which is lower than our result.

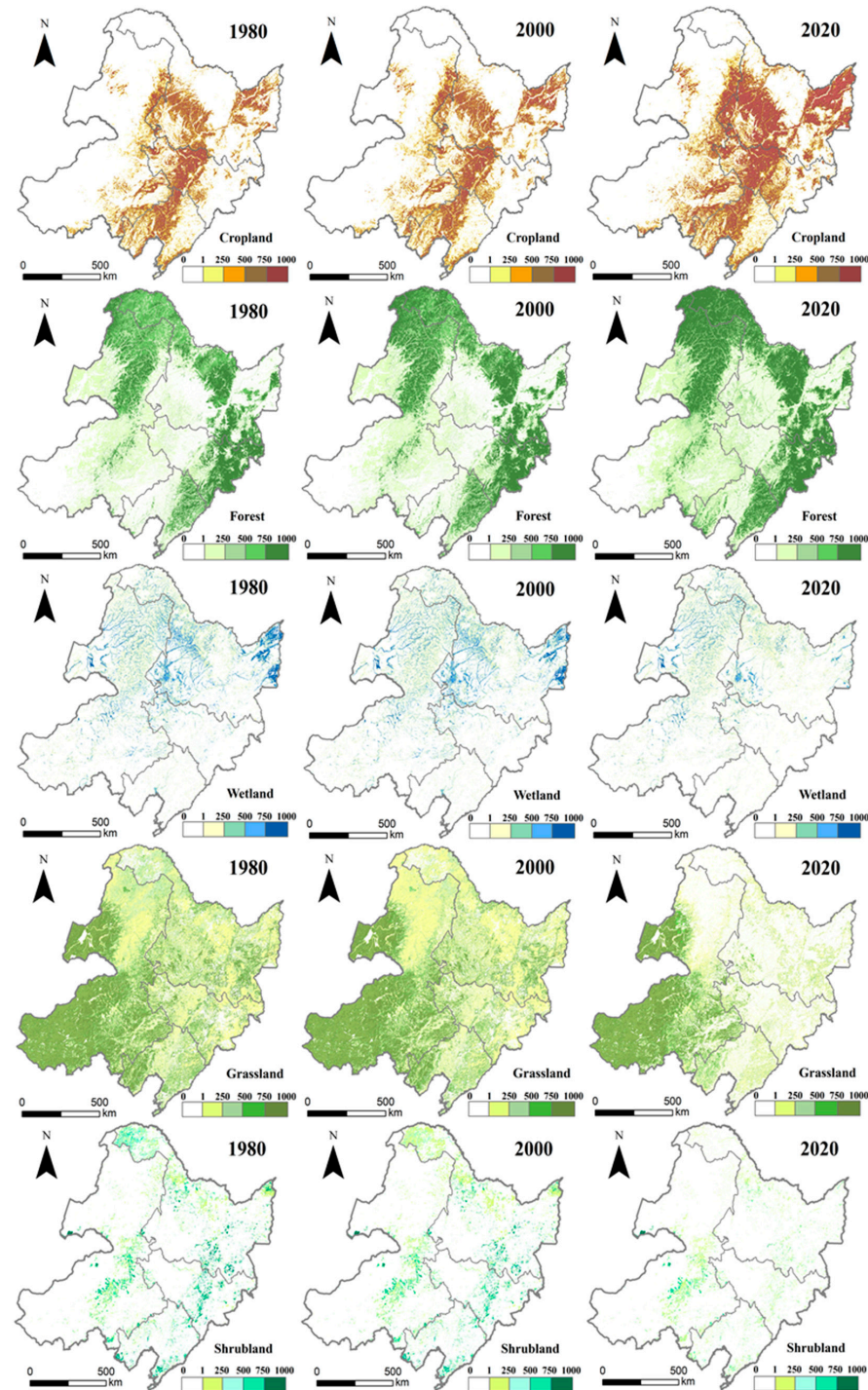


Figure 9. The spatial distribution of cropland, forest, wetland, grassland and shrubland shares (%) in 1980, 2000 and 2020 in NEC.

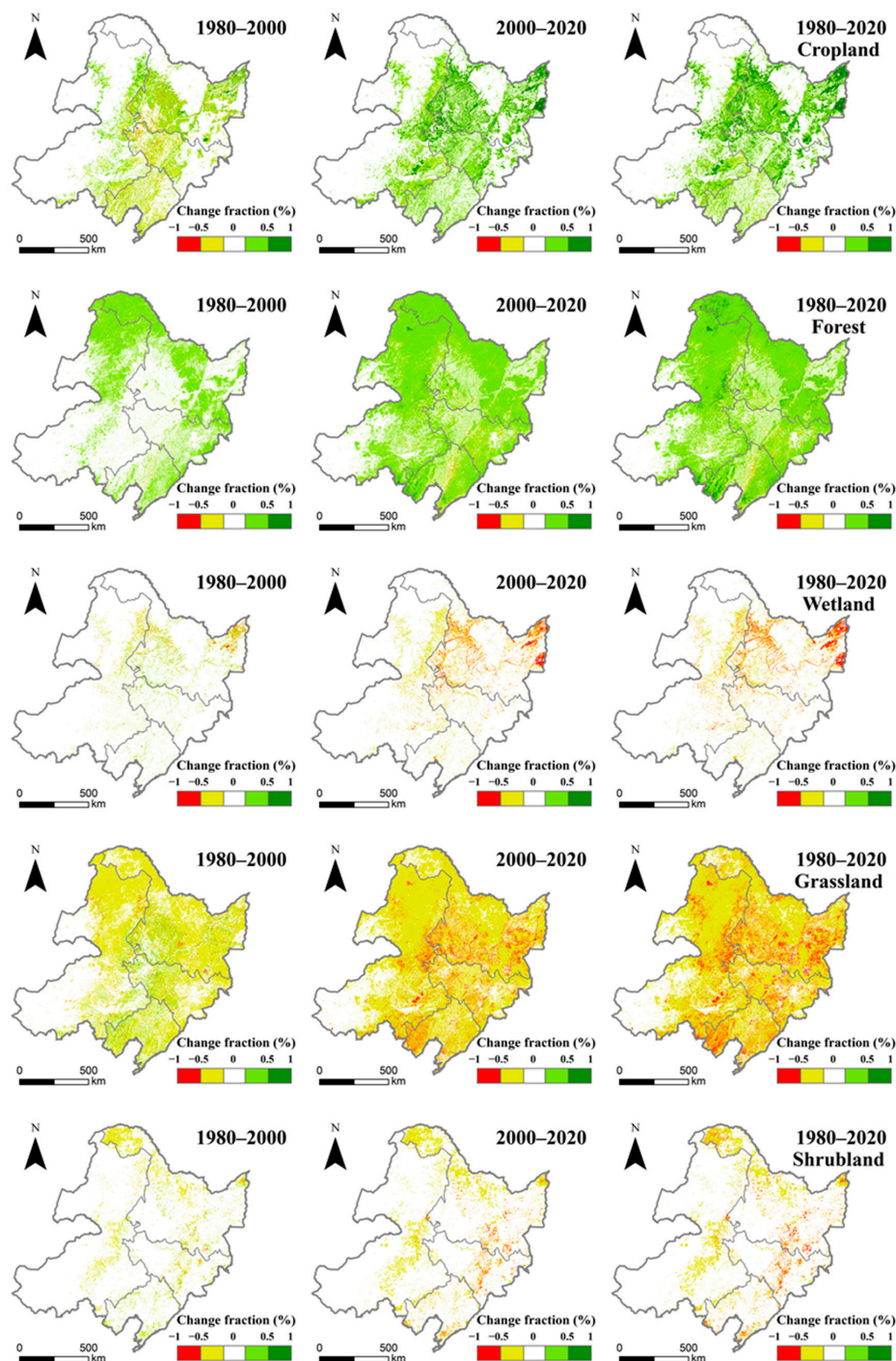


Figure 10. The spatial changes (%) of cropland, forest, wetland, grassland and shrubland shares for 1980–2000, 2000–2020 and 1980–2020 in NEC.

Based on the visually-interpreted sample plot data, we also compared the performance of our dataset for some pixels with two LUC products (Figure 12). From these three sampled pixels (1 km²), the digitized shares of our developed dataset generally capture the actual shares of different LUC types, and the CLUDA dataset also matches well with the actual data. In addition, the overall spatial patterns of these existing LUC products were also compared with our dataset, and we found a significant difference with our dataset (See Supplementary Figures S1–S5).

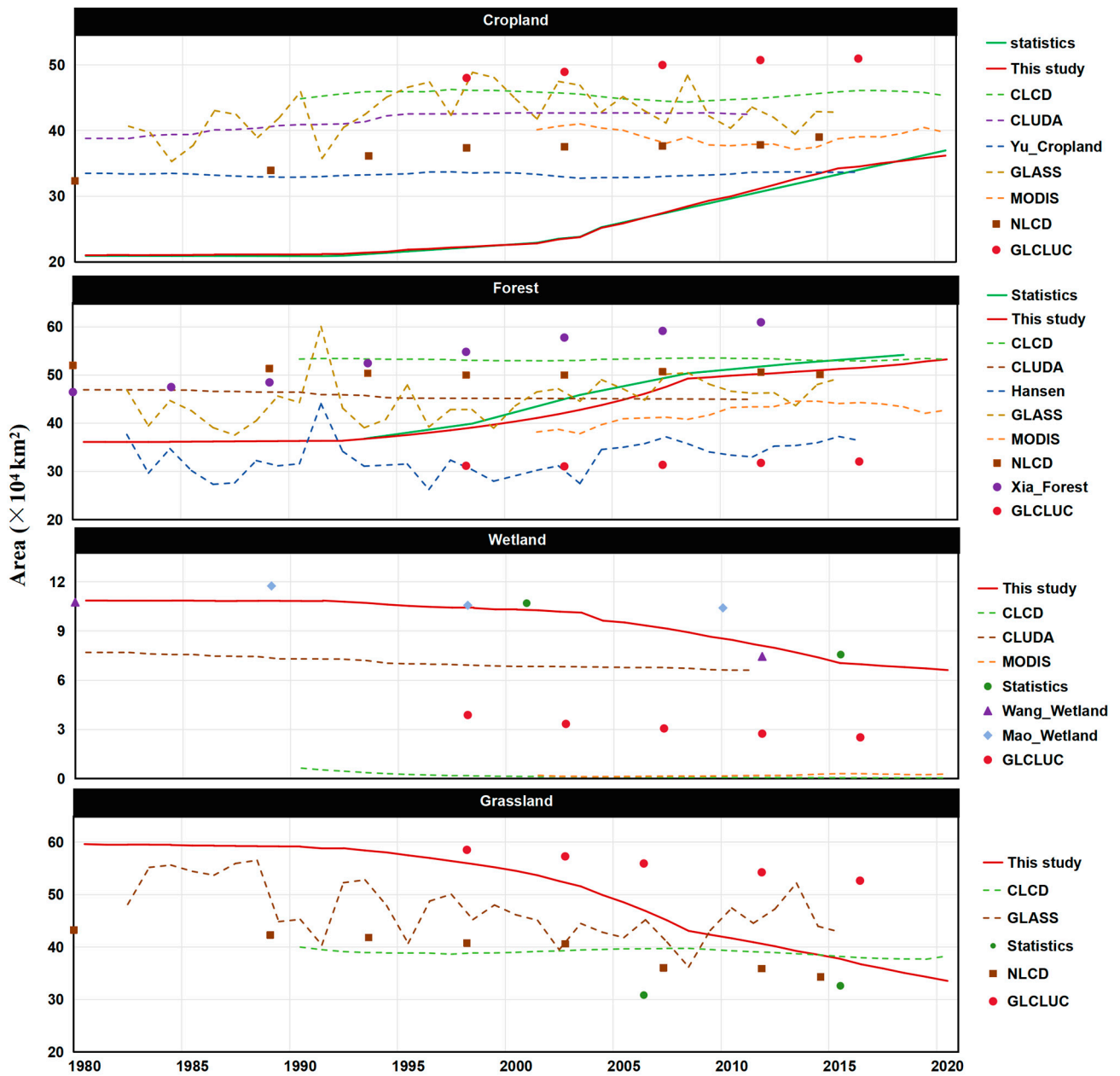


Figure 11. The intercomparisons of our developed LUC share dataset with two existing LUCC products and visually-interpreted data.

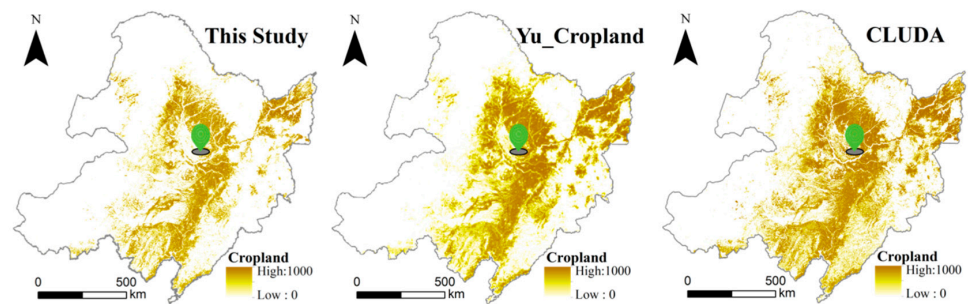


Figure 12. Cont.

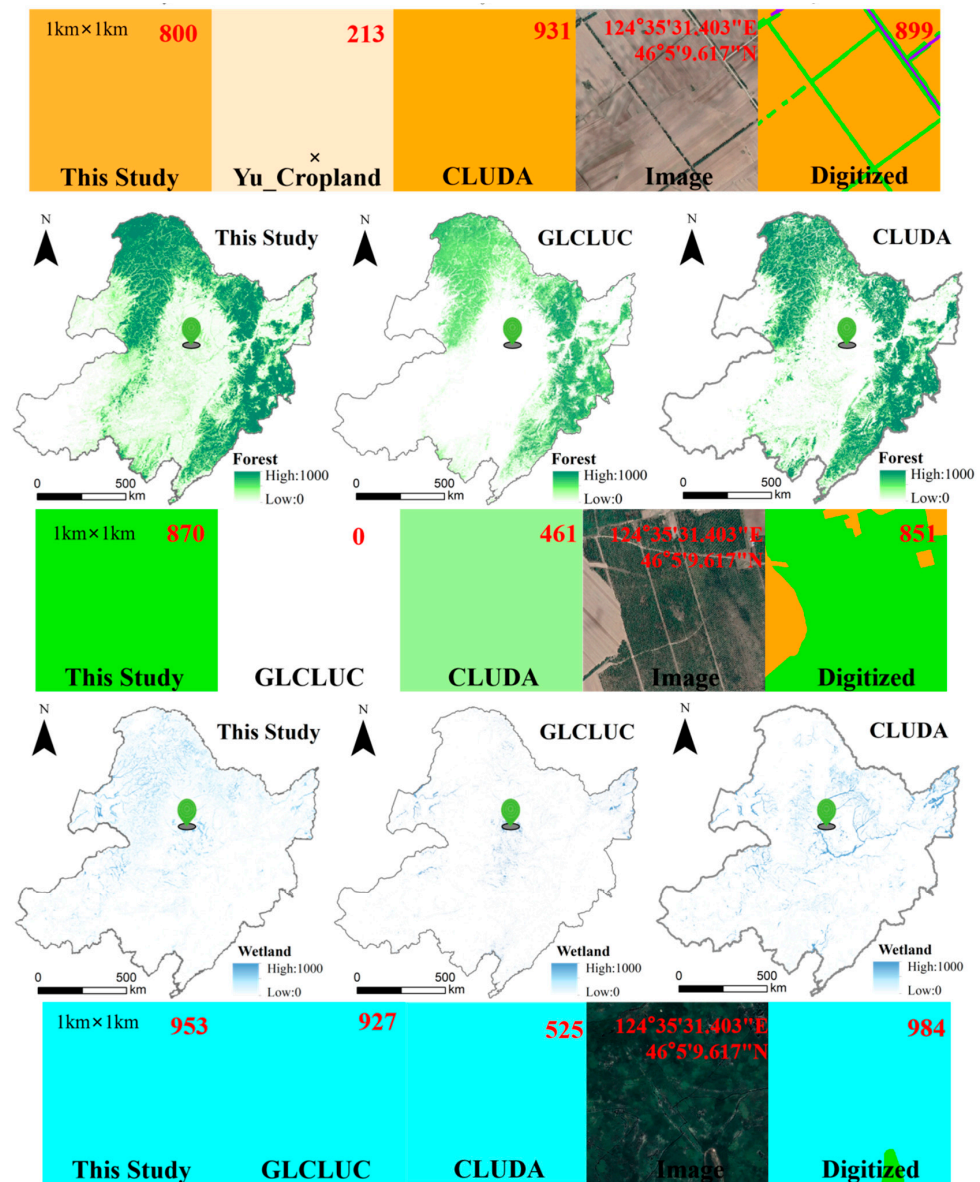


Figure 12. The comparisons of the LUC shares of our developed dataset with other existing LUCC products for the pixels with visually-interpreted LUC shares.

5. Discussion

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

NEC has experienced extensive LUCC due to climate change, urbanization, land reclamation and economic development [17,36,40]. NEC is the most important food production base in China, and crop yield is very high due to the fertile and moist soil conditions [20,43]. Due to the high nutrients and favorable water conditions, a large portion of wetland area has been reclaimed for cropland to meet the increasing demand of cereal production since the 2000s [45]. A large part of the 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 areas, mostly grassland and shrubland, where it is suitable for tree planting [49,50]. In addition, NEC was the major wood production base for China, causing large areas of deforestation since the 1970s [51]. Since the 2000s, the Natural Forest Protection policy was fully implemented, and forest disturbance and loss has been gradually decreased. All the above activities have driven a complex spatiotemporal LUCC pattern in this region. However, these phenomena have not realistically been reflected in the current existing geospatial LUCC products [17,18,20].

Based on the statistical data and assisting datasets, our study has effectively integrated multiple existing LUCC products and developed a subpixel LUC share dataset that can accurately reflect the actual LUCC conditions in NEC. Most existing datasets showed that forest area only slightly (<10%) increased in NEC, which is obviously inaccurate as many studies have proved that NEC was becoming greener and that forest area and biomass were increasing due to increasing forest coverage [18,24,25]. Similarly, the temporal change patterns of other LUC types in most existing LUCC products were also not consistent with the statistical data, such as 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. [20,30] developed subpixel level cropland datasets for all of China and the U.S.A. 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 come 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 comparisons with visually-interpreted LUC shares at each 1 km² pixel, we found that our classified LUC shares 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 a more detailed and accurate spatial representation for LUC shares (Figure 12).

5.2. The Reliability and Mechanisms of Our Approach

Although we found that most LUCC products cannot reflect 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 limitations of pixel mixture, the actual changes of individual LUC types within a pixel are difficult to detect, especially for satellite images with a coarser resolution. For example, forest is often defined as 20% tree cover within a certain area. Then, if the tree cover is 30%, this pixel is classified as forest; however, the remaining 70% of the area within this pixel is difficult to attribute to other LUC types. This further results in a failure to accurately track the temporal change of forest area within each pixel since both 30% or 100% tree coverage is regarded as forest, and thus the final classified product showed no temporal change of LUC type within this pixel. The pixel mixture issue can be partially solved for images with higher resolution, e.g., a spatial resolution lower than 5 m; however, there are few regional or global LUCC products at such high resolution. At the 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 an approach to effectively integrate the above information, and we have achieved this goal by successfully matching the actual temporal changes of all LUC types and a more detailed representation of LUCC at the subpixel level. The strengths of each LUCC dataset are integrated through their 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 it being cropland by iteratively changing the threshold values. Our approach combined the hierarchical methods from two previous studies [20,44] and can integrate the strengths of all cropland data products. RF and other machine-learning methods have been widely used to fuse multiple LUCC datasets [17,38,52]. Our approach for developing a forest dataset based on the RF method was different from the previous methods. The regression models were first fitted between annual mean $NDVI_{max}$ and statistical forest share at the provincial scale. Then, the RF regressor was applied to fit the regression model parameters by iteratively revising the parameter values until the calculated provincial forest share was

close to the statistical data. Our approach can fit the models at the pixel level and there is 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 the 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 [44]; therefore, the low spatial resolution (provincial) of the statistical data could cause large spatial location uncertainties within this province. County-level data should be more suitable for narrowing 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. Also, some LUC types in these sampling plots may be determined simply because they are registered as such types during the land survey, which could also cause large extrapolation errors of LUC area. Furthermore, there are several sources of statistical data and their magnitudes differed greatly, especially for the grassland and wetland datasets. Although we chose the most authoritative data sources, it inevitably brought either under- or over-estimation of LUCC shares.

Second, we made several extrapolations and assumptions during data development, which could result in some uncertainties in our data accuracy. For example, due to a lack of long-term and continuous inventory data for wetland area, we linearly interpolated the annual change of wetland area during 1980–2020 based on expert knowledge and five periods of data from different sources. Under the wetland protection policies, China's wetlands have increased during the recent decade; however, our developed dataset can not capture this increasing pattern due to lack of inventory data. In addition, we also made assumptions that most wetland area was converted to cropland and most forest area was converted from grassland/shrubland.

Third, 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 $NDVI_{max}$ dataset could bring uncertainty in the developed forest datasets. Furthermore, we only have a few sampling plots for evaluation of the LUC share at the pixel level due to a difficulty in available high-resolution images and digitizing work. The fewer validation plots and the possible visual interpretation errors could cause some uncertainties for the results. The correlation between our classified LUC shares and the visually-interpolated validation data also indicated that our approach slightly overestimated the actual LUC shares, partially due to the lack of enough plot-level data for training the mechanisms. To reduce these 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. [26] applied multiple LUCC datasets and a process-based model to estimate China's carbon sink due to LUCC; Chang et al. [27] assessed China's carbon stock based on the NLCD dataset; Li et al. [28] 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. [26] also revealed by using their own statistically-extrapolated LUCC product that the carbon sink was significantly higher than that using multiple existing LUCC datasets. To avoid inappropriate assessments of the LUCC effects, new LUCC share datasets for the entirety of China and the globe are needed. Although our study used a region in China as a case, the LUCC data quality was a universal issue for the entirety of China and many other regions/nations. For example, Yu et al. [30] indicated that the U.S. cropland area in multiple LUCC datasets was significantly different and the temporal changes patterns were not consistent with the inventory data. Our approach was based on the existing multiple LUCC products, statistical data and other assisting geospatial datasets, and our reconstruction methods were also widely used in geospatial data development. Therefore, our approach has the prospect of being further applied to reconstruct new LUCC datasets at both the national and global scale.

6. Conclusions

To accurately and comprehensively reconstruct an 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 the 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 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-estimated LUCC consequences. Researchers need to be cautious when choosing the existing LUCC datasets for studying 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 the national and global scale, and our approach has the prospect of achieving this goal.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/land13111814/s1>.

Author Contributions: Conceptualization, Y.C. and G.C.; methodology, Y.C. and G.C.; validation, Y.T., Y.C., R.L. and X.L.; formal analysis, Y.C., R.L. and G.C.; data curation, Y.C., R.L. and X.L.; writing—original draft preparation, Y.C.; writing—review and editing, G.C. and R.L.; visualization, Y.C.; supervision, G.C.; project administration, G.C. and R.L.; funding acquisition, G.C. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Joint Open Foundation of the Institute of Atmospheric Environment, China Meteorological Administration, Shenyang (Grant Number 2021SYIAEKFZD05), China National Key Research and Development Program (Grant Number 2023YFE0105100), Fundamental Research Funds of the Chinese Academy of Meteorological Sciences (Grant Number 2024Z001), and Overseas Expertise Introduction Project for Discipline Innovation (111 Project; Grant number D18008).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author (G.C.).

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

References

- Findell, K.L.; Berg, A.; Gentine, P.; Krasting, J.P.; Lintner, B.R.; Malyshev, S.; Santanello, J.A., Jr.; Shevliakova, E. The impact of anthropogenic land use and land cover change on regional climate extremes. *Nat. Commun.* **2017**, *8*, 989. [[CrossRef](#)] [[PubMed](#)]
- 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. [[CrossRef](#)] [[PubMed](#)]
- Gibbard, S.; Caldeira, K.; Bala, G.; Phillips, T.J.; Wickett, M. Climate effects of global land cover change. *Geophys. Res. Lett.* **2005**, *32*, 024550. [[CrossRef](#)]
- Chen, Y.; Ge, Y.; Heuvelink, G.B.; An, R.; Chen, Y. Object-based superresolution land-cover mapping from remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 328–340. [[CrossRef](#)]
- Cihlar, J. Land cover mapping of large areas from satellites: Status and research priorities. *Int. J. Remote Sens.* **2000**, *21*, 1093–1114. [[CrossRef](#)]
- 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. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [[CrossRef](#)]
- Bartholome, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [[CrossRef](#)]
- 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 Sens. Environ.* **2010**, *114*, 168–182. [[CrossRef](#)]
- 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 J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [[CrossRef](#)]
- 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. *Sci. Bull.* **2019**, *64*, 370–373.
- Zhang, X.; Liu, L.Y.; Chen, X.D.; 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 Syst. Sci. Data* **2021**, *13*, 2753–2776. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global land use/land cover with Sentinel 2 and deep learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 4704–4707.
- 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. *J. Geogr. Sci.* **2018**, *28*, 547–562. [[CrossRef](#)]
- 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. *J. Geogr. Sci.* **2014**, *24*, 195–210. [[CrossRef](#)]
- 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. *Sci. China Earth Sci.* **2020**, *63*, 1390–1407. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- 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. *J. Geophys. Res. Biogeosci.* **2023**, *128*, e2022JG007101. [[CrossRef](#)]
- 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 NFL. *ISPRS J. Photogramm. Remote Sens.* **2015**, *109*, 1–16. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- 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 Sci. Bull.* **2012**, *57*, 2813–2823. [[CrossRef](#)]
- 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. *Sci. China Earth Sci.* **2010**, *53*, 1036–1042. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- 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. *Nat. Sustain.* **2019**, *2*, 122–129. [[CrossRef](#)] [[PubMed](#)]
- Zhu, Z.C.; Piao, S.L.; Myneni, R.B.; Huang, M.T.; Zeng, Z.Z.; Canadell, J.G.; Ciais, P.; Sitch, S.; Friedlingstein, P.; Arneeth, A.; et al. Greening of the Earth and its drivers. *Nat. Clim. Chang.* **2016**, *6*, 791–795. [[CrossRef](#)]
- Yu, Z.; Ciais, P.; Piao, S.; Houghton, R.A.; Lu, C.; Tian, H. Forest expansion dominates China’s land carbon sink since 1980. *Nat. Commun.* **2022**, *13*, 5374. [[CrossRef](#)] [[PubMed](#)]
- 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. *Resour. Conserv. Recycl.* **2022**, *182*, 106333. [[CrossRef](#)]

28. 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. [[CrossRef](#)]
29. 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 J. Photogramm. Remote Sens.* **2020**, *164*, 11–25. [[CrossRef](#)]
30. Yu, Z.; Lu, C. Historical cropland expansion and abandonment in the continental US during 1850 to 2016. *Glob. Ecol. Biogeogr.* **2018**, *27*, 322–333. [[CrossRef](#)]
31. 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 Sens. Environ.* **2020**, *236*, 111510. [[CrossRef](#)]
32. 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 Sens. Environ.* **2019**, *233*, 111395. [[CrossRef](#)]
33. 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. *Sci. Data* **2024**, *11*, 527. [[CrossRef](#)] [[PubMed](#)]
34. Xu, X. A 10m Year-By-Year NDVI Maximum Dataset for China. *Resour. Environ. Sci. Data Regist. Publ. Syst.* **2022**. [[CrossRef](#)]
35. 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. *J. Clean. Prod.* **2019**, *240*, 117961. [[CrossRef](#)]
36. Wang, Y.; Shen, X.; Lü, X. Change characteristics of landscape pattern and climate in marsh areas of northeast china during 1980–2015. *Earth Environ.* **2020**, *48*, 348–357.
37. 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. *Sci. Bull.* **2017**, *62*, 508–515. [[CrossRef](#)]
38. 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. [[CrossRef](#)]
39. 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. *J. Nat. Resour.* **2016**, *31*, 1253–1263.
40. Ye, Y.; Fang, X.Q. Spatial pattern of land cover changes across Northeast China over the past 300 year. *J. Hist. Geogr.* **2011**, *37*, 408–417. [[CrossRef](#)]
41. Cao, B.; Yu, L.; Li, X.; Chen, M.; Li, X.; Hao, P.; Gong, P. A 1 km global cropland dataset from 10 000 bce to 2100 ce. *Earth Syst. Sci. Data* **2021**, *13*, 5403–5421. [[CrossRef](#)]
42. 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. *Front. Remote Sens.* **2022**, *3*, 856903. [[CrossRef](#)]
43. 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. *Sci. Data* **2021**, *8*, 41. [[CrossRef](#)] [[PubMed](#)]
44. Zhang, C.; Dong, J.; Ge, Q. Mapping 20 years of irrigated croplands in China using MODIS and statistics and existing irrigation products. *Sci. Data* **2022**, *9*, 407. [[CrossRef](#)] [[PubMed](#)]
45. 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. *Sci. Total Environ.* **2018**, *634*, 550–560. [[CrossRef](#)] [[PubMed](#)]
46. Luo, H.; Huang, F.; Zhang, Y. Space-time change of marsh wetland in Liaohe Delta area and its ecological effect. *J. Northeast. Norm. Univ.* **2003**, *35*, 100–105.
47. Asselen, S.; Verburg, P.H.; Vermaat, J.E.; Janse, J.H. Drivers of wetland con-version: A global meta-analysis. *PLoS ONE* **2013**, *8*, e81292. [[CrossRef](#)]
48. Stehman, S. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89. [[CrossRef](#)]
49. Zheng, X.; Zhu, J.J. Estimation of shelter forest area in Three-north Shelter Forest Program region based on multi-sensor remote sensing data. *Chin. J. Appl. Ecol.* **2013**, *24*, 2257–2264.
50. 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. *Chin. J. Ecol.* **2019**, *38*, 1600–1610.
51. 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. *Commun. Earth Environ.* **2023**, *4*, 15. [[CrossRef](#)]
52. Li, K.; Wang, J. A multi-source data fusion method for land cover production: A case study of the East European Plain. *Int. J. Digit. Earth* **2024**, *17*, 2339360. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.