Long-Term Changes of Land Use and Land Cover in the Yangtze River Basin from 1990–2020 Landsat Data

Junyuan Yao and Shuanggen Jin

Abstract

Economic development and climate change drive the land use and land cover (LULC) change globally. Annual robust maps of LULC are critical for studying climate change and land–climate interaction. However, the current existing methods for optimizing and expanding the publicly available China land cover data set (CLCD) are limited. In this article, 30-m annual LULC changes are obtained from 1990 to 2020 in the Yangtze River basin (YRB). The results show an overall accuracy rate of 82.66% and better performances on Geo-Wiki test samples when compared to similar products. Based on our 30-m annual LULC data set, the drastic LULC changes are found in YRB over a 30-year period, where impervious surface area more than tripled, cropland area decreased by 6.12%, and water area decreased by 6.09%. In addition, through the geographically and temporally weighted regression method, a fitting model with a goodness of fit of 0.91 well reveals that human activity plays a driving role in the LULC change of YRB.

Introduction

Land use and land cover (LULC) change has a close relationship with social and economic development, ecosystem carrying capacity, surface energy balance, and material circulation (Foley *et al.* 2005; Gibbard *et al.* 2005; Vorosmarty *et al.* 2010; Houghton *et al.* 2012; Haddeland *et al.* 2014; Findell *et al.* 2017). The Yangtze River basin (YRB) covers several major cities and nature reserves in China, which is of great importance in economic development and ecological conservation. Recently, there have been serious problems in YRB in terms of water environment pollution and ecological damage due to excessive reclamation and economic development (Yang *et al.* 2021a). Therefore, it is essential to map the LULC change in YRB to explore the processes and drivers within the basin over the past 30 years. Effective environmental governance and development planning depends heavily on accurate LULC products and effective quantitative analysis of LULC changes.

Remote sensing is the most efficient way to monitor large-scale LULC change. In recent years, the free access to a huge volume of remote sensing satellite data (e.g., AVHRR, MODIS, Landsat, and Sentinel-2) and high-performance cloud computing platforms such as Google Earth Engine (GEE) have greatly promoted large-scale and long-term remote sensing studies on LULC (Gorelick *et al.* 2017; Zhu *et al.* 2019). With the GEE platform, Qu *et al.* (2021) used the k-means method to optimize the sample quality and generated LULC products from 1992 to 2015 for three provinces in the Yangtze River delta region using the random forest (RF) classification method. Liu *et al.* (2020a) built a training sample based on OpenStreetMap data and used classifiers of RF and the classification and regression tree (CART) for LULC mapping in the middle YRB from 1987 to 2017. However, these works have built completely new training sample sets to train the classifiers, which was certainly a huge amount of work (Jin and Zhang 2016). In addition, these results are difficult to apply directly in our region due to uncertain stability.

In addition, a number of open-access LULC data sets were used in YRB LULC change. Chen *et al.* (2020) used the LULC map data from the Resource and Environment Data Cloud Platform of the Chinese Academy of Sciences and reclassified it for the same 20 LULC categories as IGBP-MODIS to examine the influence of land urbanization on meteorology and air quality in the Yangtze River delta. Yang *et al.* (2021b) selected the MOD12Q1 data product from 2001 to 2018 with a spatial resolution of 500 m and reclassified the LULC maps into eight dominant categories to analyze the influence of LUCC on net primary production in YRB and its causes. However, the spatial resolution of 1000 or 500 m is relatively coarse, which is not sufficient for fine-scale LULC monitoring due to the uncertainty inherent of coarse-resolution data (Sulla-Menashe *et al.* 2019). Recently, Landsat data, with their high resolution and long history, have been widely used in large-scale LULC mapping (Liu *et al.* 2020b). Leveraging the data and computing power of the GEE platform, many global land cover products have been released, such as FROM_GLC30 (finer-resolution observation and monitoring of global land cover) (Gong *et al.* 2013), GlobeLand30 (global land cover mapping) (Chen *et al.* 2015), and GLC_FCS30 (global land cover product with fine classification system) (Zhang *et al.* 2021). Compared to these global LULC products, the China Land Cover Dataset (CLCD) (Yang and Huang 2021) provided longer-time-series and higher-precision-validated LULC in the China's region for each year. Since its training samples are sufficient and reliable by combining stable samples extracted from China's land use/cover data sets (Liu *et al.* 2014), the accuracy of CLCD reaches a high level by both visual interpretation sample and third-party sample accuracy tests.

Despite the high-accuracy performance of CLCD, there are still several problems when applying it to the YRB. Considerable siltation occurs in the middle YRB and results in riverbed uplift and mudflats in many places that are more pronounced after the numerous construction of dams and water diversion for agriculture in the upper YRB (Chen *et al.* 2001; Yang *et al.* 2015). Unfortunately, the LULC in these areas is mostly misclassified as impervious surfaces in the CLCD. Further, Dongting Lake, Poyang Lake, and Middle River Bay are important wetland distribution areas in YRB formed by sediment accumulation due to the action of water flow and indirect coverage by water in different hydrological periods (Yang *et al.* 2005; Yang *et al.* 2018). These were also not represented well in the CLCD, in large part because it used the 50th-percentile value of each spectral band as the basic data

Junyuan Yao is with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, People's Republic of China; and Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, People's Republic of China.

Shuanggen Jin is with Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, People's Republic of China; and the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, People's Republic of China (sg.jin@yahoo.com)..

Contributed by Prasad S. Thenkabail, February 14, 2022 (sent for review April 13, 2022; reviewed by Nagesh Kumar D., Prasad S. Thenkabail).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 9, September 2022, pp. 573–581.
0099-1112/22/573–581

doi: 10.14358/PERS.22-00050R2

for training and classification, which made the classification results based on combinations of image elements at different times throughout the year. This approach undermined the recognition of land types like wetlands with varying within 1 year. In addition, there was also some misclassification in the mountain shadows of western Sichuan and northern Chongqing. Although the CLCD work flow can improve the overall classification accuracy, there are still some challenges for high-precision mapping of LULC in YRB. Therefore, it is necessary to correct the CLCD and obtain the high-accuracy LULC maps for large-scale YRB. This reclassification process should be an effective, efficient, economic, and operational approach.

In this article, 30-m annual LULC changes are obtained and analyzed in YRB, including (1) obtaining the 30-m annual LULC mapping of YRB from 1990 to 2020 by intercomparing the CLCD with thematic-class products (GFC global forest change, GSW global surface water, and GISA global impervious surface area) (Hansen *et al.* 2013; Pekel *et al.* 2016; Huang *et al.* 2021) and detecting and reclassifying the disputed areas of the CLCD in YRB by using the RF classifier with a smaller number of visually interpreted samples on GEE, (2) analyzing the process and trend of LULC change in YRB over 30 years with explaining the causes of change, and (3) proposing a time-series model of land use degree and exploring the drivers of the LULC change in YRB by using the geographically and temporally weighted regression (GTWR) model (Huang *et al.* 2010). The changes and drivers of the LULC in YRB can provide valuable information for local decision makers and stakeholders.

Data and Methods

Study Area

The Yangtze River Basin (about 1.8 million km2) is located between 90°33′ and 122°19′ E and 24°27′ and 35°54′ N in China, starting from the Qinghai-Tibet Plateau and going eastward into the East China Sea (Figure 1). Most areas in YRB belong to the subtropical monsoon climate zone with the average annual rainfall from 692 to 1611 mm and the average air temperature ranges from 9°C to 18°C (Zhang *et al.* 2019). Suitable climatic conditions have created a rich variety of vegetation types, and the differences in sea–land and elevation have resulted in many topographic features in the basin, such as mountains, plateaus, basins, hills, and plains. The large areas of plain and basin, as well as the rich water resource, have led to rapid industrial and agricultural development in YRB. Large cities in the basin, such as Shanghai, Nanjing, Wuhan, and Chongqing, have experienced rapid economic development in the past three decades, and urban expansion has brought about frequent land use modifications (Li *et al.* 2021). Human activities and climate change have threatened the unique ecological structure and have polluted water quality, which has led to the national policies of a 10-year ban on fishing and environmental remediation in the basin (Sun *et al.* 2017; Wu *et al.* 2021).

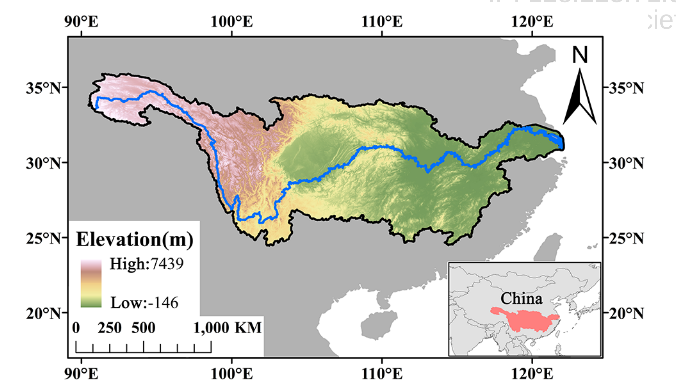

Figure 1. Study area of the Yangtze River basin.

Data

CLCD

The CLCD contains 30-m annual LULC in China from 1990 to 2019. CLCD's classification system includes nine major LULCs: cropland, forest, shrub, grassland, water, snow and ice, barren, impervious, and wetland. CLCD used sufficient training samples from combining stable samples extracted from China's land use/cover data sets and visually interpreted samples from satellite time-series data. The classification method was stable because they used spectrum, spectral index, phenology, and geographic location as input features to the RF classifier, and a postprocessing process was proposed incorporating spatial-temporal filtering and logical reasoning. The overall accuracy reached 79.31%, and two third-party verifications (Geo-Wiki and GLCVSS) achieved 54.57% and 65.46% accuracy, proving its superiority over similar products. The data of YRB were directly downloaded from https://doi.org/10.5281/zenodo.4417810, and the results in 2020 were obtained using the 2019 data as training.

Remote Sensing Data Sources

The GEE platform (https://earthengine.google.com) provides the Landsat data set of the U.S. Geological Survey (USGS, https://www.usgs.gov), and the Landsat surface reflectance data used in this study have conducted systematic atmospheric and terrain correction. Considering the quality of the Landsat data, we chose the Landsat 8 OLI data after 2013 and combined TM and ETM+ data before 2013. Images from June to September of each year were used to filter the production of training data. There are 157×8 Landsat scenes per year for the entire study area, and in practice, it will be less than this amount because of missing images. The Normalized Difference Vegetation Index (NDVI), Normalized Difference Built Index (NDBI), and Modified Normalized Difference Water Index (MNDWI) (Xu 2007; Szabó *et al.* 2016) are all calculated based on the original band, and elevation and slope are computed from digital elevation model (NASA Shuttle Radar Topography Mission Digital Elevation 30 m) data in GEE. The thematic-class products (GFC global forest change, GSW global surface water, and GISA global impervious surface area) are obtained on GEE and from http://irsip.whu.edu.cn. These thematic-class products are based on global Landsat data with a spatial resolution of 30 m with continuous updating. Among them, GISA uses a machine learning classification framework and postprocessing of luminous data to ensure the accuracy of the product. Similarly, GFC and GSW are calculated with the different spectral indices and fully take into account the influence of seasons.

Table 1. Definition of each category and sample distribution.

Class	Definition	Number
Cropland	Orchards and cropland, including paddy fields and dry fields	620
Forest	Forestry land used for growing trees and bamboo etc.	580
Shrub	Low shrubs and areas with low vegetation cover	137
Grass	All kinds of grasslands that grow mainly herbaceous plants, including shrub grassland and sparse forest grassland	102
Water	Land used for natural terrestrial waters and water conservancy facilities	680
Snow/ice	Land covered with snow year-round	140
Barren	Land that has not yet been used or that is difficult to use	304
Impervious	Urban and rural residential areas and other industrial, mining, and transportation land	496
Wetland	Land with flat and low-lying terrain, seasonal or year-round accumulation of water, and growing wet plants on the surface	146

Sample Data

The classification system in this article is the same as CLCD's classification system, including cropland, forest, shrub, grassland, water, snow and ice, barren, impervious, and wetland. Due to the excellent basis of CLCD, only a small number of samples are needed as training data. High-quality training and validation samples are evenly distributed in the study area and provided for supervised classification by combining the Google Earth HD map and GEE platform sample-making tools. In total, 1700 training samples and 1505 validation samples were selected independently in 1995, 2005, and 2015.

Anthropomorphic and Natural Data

Natural and social factors have a profound impact on LULC change. Seven major potential factors were selected in this study to test their impact on LULC change in YRB (Table 2), including population density (POP), gross domestic product (GDP), annual average precipitation (PRE), annual average temperature (TEM), net primary productivity (NPP), Normalized Difference Vegetation Index (NDVI), and digital elevation model (DEM). The data were downloaded from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences (https://www.resdc.cn).

Method

The process was implemented on the GEE platform and ArcGIS, shown in Figure 2. First, the categories of forest, water, and impervious surface in CLCD were compared with GFC, GSW, and GISA data year by year to derive the disputed area. Then the Landsat data of the disputed area were used to reclassify by the RF classifier. Next, through using visually interpreted and third-party validation samples, the classification accuracy was calculated and compared to verify the reliability of our LULC results; LULC change is also analyzed by quantifying the individual changes and interconversions of each category. Finally, through the process of the time-series analysis of land use degree, we transformed the LULC categories into 5-year land use degree data, and the natural and social drivers are analyzed by using GTWR model on ArcGIS.

Table 2. Anthropomorphic and natural data∗.

Data Name	Time	Spatial Resolution (m)
POP	1990, 1995, 2000, 2005, 2010, 2015	1000
GDP	1990, 1995, 2000, 2005, 2010, 2015	1000
PRE	1990, 1995, 2000, 2005, 2010, 2015	1000
TEM	1990, 1995, 2000, 2005, 2010, 2015	1000
NPP	2000, 2005, 2010	500
NDVI	1990, 1995, 2000, 2005, 2010, 2015	500
DEM	2000	30

POP= population density; GDP = gross domestic product; PRE = annual average precipitation; TEM = annual average temperature; NPP = net primary productivity; NDVI = Normalized Difference Vegetation Index; DEM = digital elevation model.

Data Processing and Classification

Image data sets and computing services on the GEE platform make the cloud removal process easy and efficient to execute by the CFmask algorithm (Zhu and Woodcock 2012). We selected data from June to September of each year because it is the vegetation growing season and the hydrological conditions are stable (Guo *et al.* 2008).

RF classification algorithm is used for processing large-scale and complex data (Belgiu and Drăguţ 2016) and has been widely used in LULC classification because it is good at overcoming the noise in the data and the overfitting problem of training (Na *et al.* 2010). Zhang *et al.* (2020) compared the contribution of the auxiliary feature vectors for RF classification in LULC studies (Zhang and Yang 2020) and showed that the accuracy of RF classification can be improved more by adding auxiliary feature vectors, such as spectral indices and elevation information. Taking into account the classification target category of this study and the physical geography of the study area, five auxiliary feature vectors (NDVI, MNDWI, NDBI, DEM, and slope) and the original spectral band are added as the input features to the RF classifier (Hoshikawa and Umezaki 2014). After experimental comparison, stable classification performance can be obtained when the number of decision trees of the RF classifier is set to 200.

Accuracy Assessment

To assess the accuracy of our results, we use a visually interpreted test set (1505 in total) and a third-party test set (Geo-Wiki) (4226 in total). Based on our results, the data (CLCD, GFC, GISA, and GSW) are verified to be very stable each year. The accuracy validation in 1995, 2005, and 2015 proves the stability of our products, which allows the comparison with other data products in 2015. In addition, we also validated our results with CLCD, FROM_GLC, and GLCFCS30 on the validation sample of Geo-Wiki. Furthermore, the accuracy of our results was validated by confusion matrixes, including producer's accuracy (PA), user's accuracy (UA), overall accuracy (OA), and kappa coefficients.

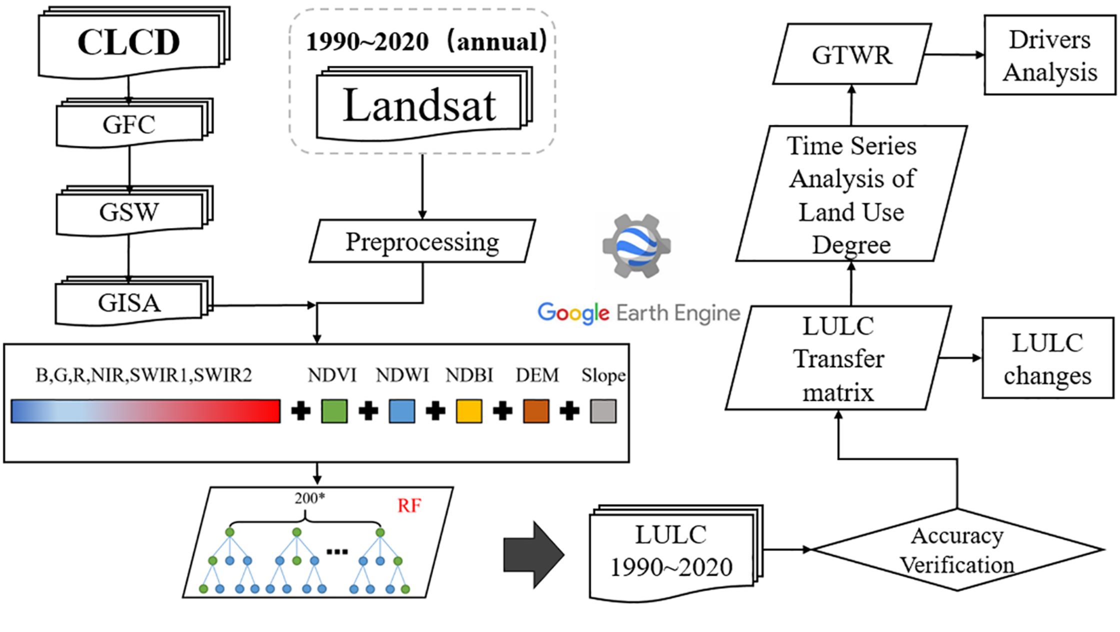

Figure 2. Methodology and flowchart.

Time-Series Model of Land Use Degree

The land use degree model is commonly used in LULC change analysis, which was developed from the original cropland use intensity model by Wang *et al.* (2010). Liu *et al.* (2020c) proposed a new quantitative analysis method for land use degree that has been applied in China (Liu *et al.* 2020c). This method divides the land use degree into four levels according to the balance state of the land cover under the influence of natural/social factors. The setting of the land use degree index is shown in Table 3.

The quantitative model of land use degree distinguishes the natural land cover and man-made land use by assigning different weight values to different land types so that the disordered LULC categories become orderly (Liu *et al.* 2020d). As shown in Figure 3, the LULC results within every 5 years (corresponding to the natural/social factors data) are considered as a whole. The land use degree indices over the five years are summed up to represent the overall contribution of different LULCs over the five years with reducing the effect of classification errors. In addition, our method provides more information when compared to the direct use of the land use index data for a single year since five to 20 indexes have more details than one to four indexes.

GTWR Model

The GTWR model is a classical model for the study of spatial heterogeneity in long time series. The GTWR model incorporates the time dimension into the geo-weighted regression model, which can obtain a better fit and make the estimation results more effective (Ma *et al.* 2018). Correlation calculations and multicollinearity tests are indispensable processes before being input into the GTWR model (Ran *et al.* 2019). Therefore, we calculated the Spearman correlation coefficients between the drivers and the land use degree data, which can respond to the degree of correlation between land use data and other spatial attribute data (Myers and Sirois 2004; Tran *et al.* 2010). In addition, due to different natural/social factors that exist with different temporal and spatial resolutions, the uniformization process is done before the correlation analysis. Finally, we select the factors with higher correlations and then exclude those with variance inflation factor (VIF) over 5 as dependent variables into the GTWR model. In this article, the base distance is set to the Chinese municipal administrative divisions of YRB considering the stability of the GTWR model. The GTWR model developed in this study depicts the quantitative spatial-temporal relationships of the LULC drivers, and its overall structure is described as follows:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_k \beta_k(u_i, v_i, t_i)\, x_{ik} + \varepsilon_i \quad (1)$$

where i ($i = 1, 2, \ldots, n$) denotes a city region; the dependent variable y_i refers to the LUD for each city; x_{ik} represents the driver factors; u_i, v_i, t_i are the longitude, latitude, and time, respectively; β_0 is the intercept value; β_k denotes a set of parameter values; and is the random error.

Table 3. Classification values of land use degree.

Types of Land	Uncultivated Land	Ecological Land	Agricultural Land	Construction Land
LULC	Barren, snow	Forest, shrub, grassland, wetland and water	Cropland	Impervious, including urban, residential area, and traffic land
Index value	1	2	3	4

LULC = c.

Results and Analysis

Classification Results and Accuracy Assessment

Partial LULC results are shown in Figure 4, and the study area was dominated mainly by forest (46.23%), cropland (28.40%), and grassland (18.18%). As in Figure 4, most of the natural forests in the basin are located in the middle and upper parts. The Sichuan Basin, the Central Plain, and the Yangtze River delta plain are the main agricultural areas. The grassland, barren, and snow land are located in the alpine areas of the Qinghai-Tibet Plateau in China. The downstream is rich in water resources, and the cities are scattered, with the largest urban agglomeration in the Yangtze River delta, including Shanghai and Nanjing. Finally, most of the wetlands are located near Taihu Lake, Poyang Lake, and Dongting Lake.

Good accuracy of the LULC classification results is an important prerequisite for subsequent computational analysis. Based on the validation sample of visual interpretation in 1995, 2005, and 2015, the overall accuracy of this study reached 80%–83%, and the kappa coefficient of this study is about 0.79, which proves that the validation results are stable and reliable. For each category, forest, snow/ice, and barren have the highest classification accuracy of around 90%, while cropland, shrubs, water, impervious surface, and grassland are relatively high with all accuracy of above 70% as well. The classification of wetland categories has also improved considerably after our efforts and reached 60.71%, which makes up for the lack of other data in this area. In the comparison with CLCD, the accuracy of some categories of our product remains similar, while the accuracy in cropland, barren, impervious surface, and wetland are all improved. Therefore, our results have advantages and make up for the shortcomings of the other products.

To verify the effect of our method on CLCD enhancement, we intercepted part of the result of CLCD and this study, shown in Figure 5. It can be seen by comparison with the Google HD map that there is an obvious misclassification of mountain shadows into the water in CLCD (Figure 5d). By the method of this article (after intercomparison and reclassification with GSW), the misclassified part is completely removed. On the Google HD map in Figure 5b, we can see that there

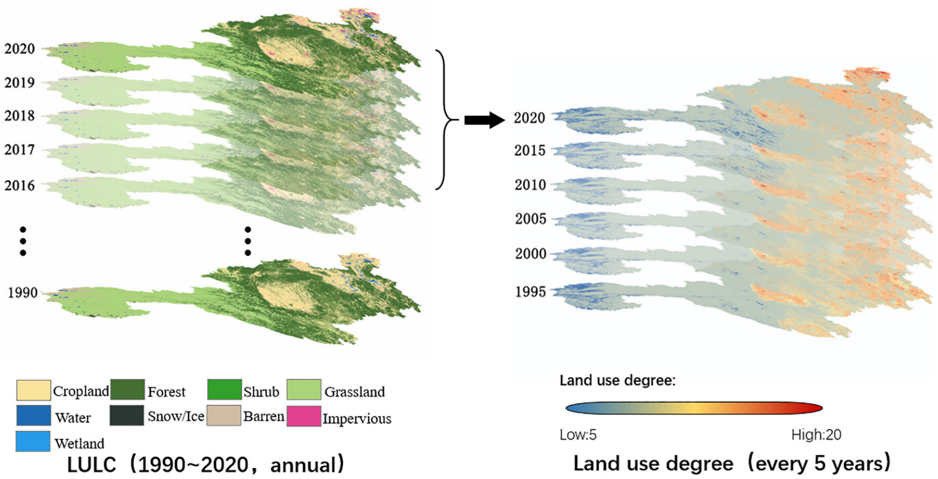

Figure 3. Time-series analysis of land use degree.

is a large intermittent wetland area that occurs during the dry seasons and is inundated during the wet seasons in the middle of the shoreline and the lake inundation area (Yang *et al.* 2020). This area was not well classified in the CLCD and was misclassified as cropland, which was partially corrected after our reclassification. However, it is still a bit flawed because some areas were not involved in the intercomparison, and, on the other hand, the spectral characteristics of wetlands are easily confused with the spectra of other land types (Mahdavi *et al.* 2018). Furthermore, by adding the sample of the exposed riverbed, it is also easy to see in Figure 5f that the area that was originally misclassified as impervious surface is well identified as barren. These comparisons show that our work has trustworthy accuracy and effectively improves the accuracy of CLCD in YRB.

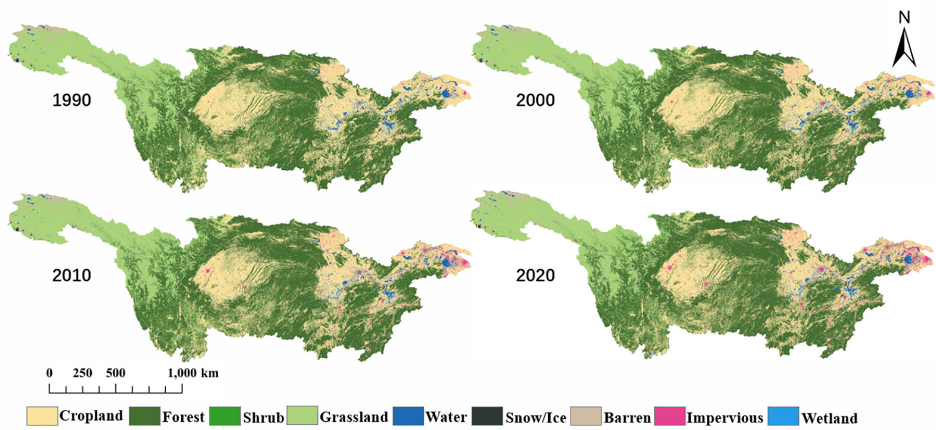

Figure 4. Land use and land cover mapping of the Yangtze River basin in 1990, 2000, 2010, and 2020.

Table 4. Validation of the results in this article based on visually interpreted test samples in 1995, 2005, and 2015.

	Cropland	Forest	Shrub	Grassland	Water	Snow/Ice	Barren	Impervious	Wetland	OA (%)
1995										
PA (%)	71.64	87.35	91.3	67.42	74.17	90.91	90.32	88.73	1	79.55
UA (%)	94.3	94.57	48.38	94.9	96.43	50	47.73	63	12.09	
2005										
PA (%)	80.39	86.95	89.29	70.1	76.16	95.24	91.59	85.86	96.67	82.31
UA (%)	93.45	96.38	40.32	94.91	93.57	50	55.68	85	31.87	
2015										
PA (%)	82.19	89.02	88	71.92	74.71	92	94.78	73.81	60.71	82.66
UA (%)	92.02	96.38	35.48	92.41	90.71	57.5	61.93	93	30.77	

PA = producer's accuracy; UA = user's accuracy; OA = overall accuracy.

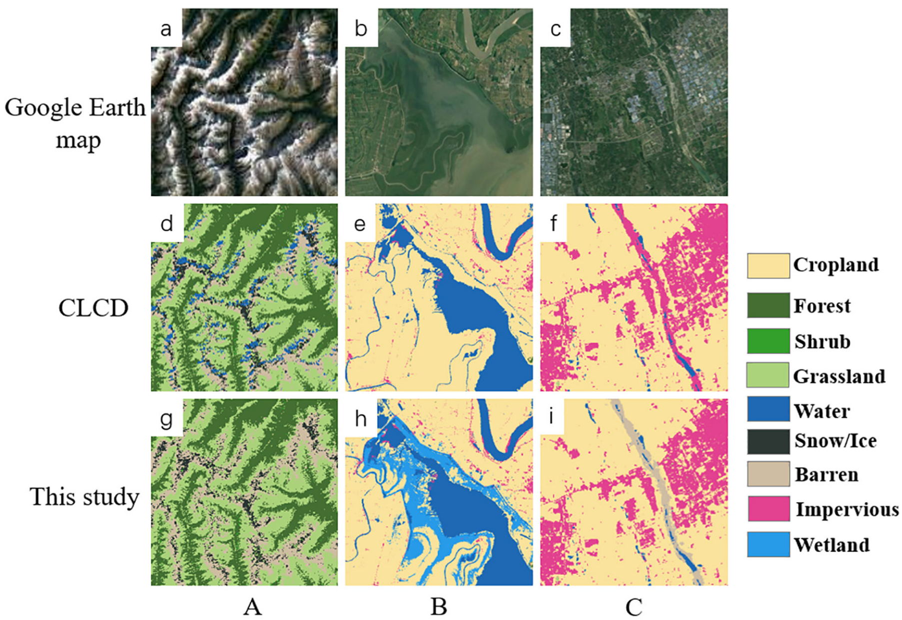

Figure 5. Google HD maps, China land cover data set (CLCD), and land use and land cover classification of this study.

Long-Time Spatial-Temporal Change Characteristics

The process and trend of LULC change are calculated for each category from 1990 to 2020 based on the LULC generated in this study (Figure 6). Impervious areas covered 4.28 million ha in 2020, sprawling unprecedentedly over the past 30 years and increasing more than three times relative to that in 1990, which is also consistent with GISA data. In general, impervious areas far exceed other categories in terms of the magnitude of change. Cropland area was dropped to 54.65 million ha in 2020 with a decrease of 6.12% when compared to 1990. The significant increase in forest land area by 4.04% (3.22 million ha) was due to China's positive response to the Grain for Green program (Robbins and Harrell 2014), especially in the middle and upper streams, while the lower stream remained stable (Xu *et al.* 2020). The area of surface water increased by 6.09% (0.2 million ha), especially after 1995, when the development of hydropower was proposed by the Ninth Five-Year Plan of China. The increasing reservoir and dam construction were some of the reasons accounting for the surface water extension (Ali *et al.* 2019). Barren declined slightly by about 9.00% until 1997, then trended steadily upward about 38.11% and grew significantly faster after 2015. After the adjustment, the area of the exposed riverbank upstream was classified in the barren category. YRB was flowing fast between 1990 and 2000, when a large amount of sediment was carried downstream by the current. With the construction of the water conservancy facilities afterward, the river flow velocity decreased, which reduced the amount of sand transported by the river (Chen *et al.* 2001). In particular, the completion of the Three Gorges Dam after 2015 caused more than 90% of the sediment to be retained in the upper basin of the Yangtze River (Yang *et al.* 2018), thus creating an elevated riverbed and increased barren. The wetlands have experienced dramatic changes (some of the fluctuations may also come from classification errors), with an increase of 0.03 million ha overall. The fragmentation of the wetland landscape in the basin is significant (Rui *et al.* 2017), and the construction of some wetland parks may account for some of the increase. Shrub decreased significantly by 50%, and the snow/ice land varied in a regular undulating pattern, covering an average of 0.33 million ha. Grassland continued to decline by 6.99% to 36.54 ha in 2020. Since grassland is located mainly on the Tibetan Plateau with less effect by human activities, the impact of climate change on vegetation in vulnerable areas is the main cause.

The years 1990, 2000, 2010, and 2020 are chosen to study the land transfer process, and 100 major transformations are selected to make Sankey diagrams (Figure 10). The vegetation cover area (forest, cropland, shrub, and grassland) remained largely unchanged over the years, but its internal transformations were frequent. There is a large mutual transfer of cropland and forestland, and large deforestation and afforestation occur during this time. Also, the imaging period error of the image itself and the different image quality (Landsat-8 has better image quality) have some influence. It can be observed that the main origin of impervious surface area growth is cropland, as is the case with water, which explains the decrease of cropland. This is due to the urban development that continues to occupy the surrounding cropland, and the policy of returning farmland to forests and lakes implemented in China is also a reason. The wetland area is small and relatively independent, and its transformation is not within the top 100, so it is not reflected in the figure.

Driving Force Analysis

LULC change is a complex process influenced by natural and social factors (Fox and Vogler 2005). In this article, the correlations between natural/social factors and land use degree were calculated with little overall fluctuation, and their average values are shown in Table 5. Among them, Correlation coefficients above 0.3 were considered to be relevant, while those above 0.6 were considered to be high. Then POP, TEM, PRE, and DEM were selected as independent variables input to the GTWR model because of the high correlation coefficients, where GDP was excluded due to its VIF value being greater than 5 in Table 6. Finally, our GTWR is set with land use degree as the dependent variable; POP, TEM, PRE, and DEM as the independent variables; the spatial dimension as the location of each municipality mass center; and the temporal dimensions as 1990, 1995, 2000, 2005, 2010, and 2015.

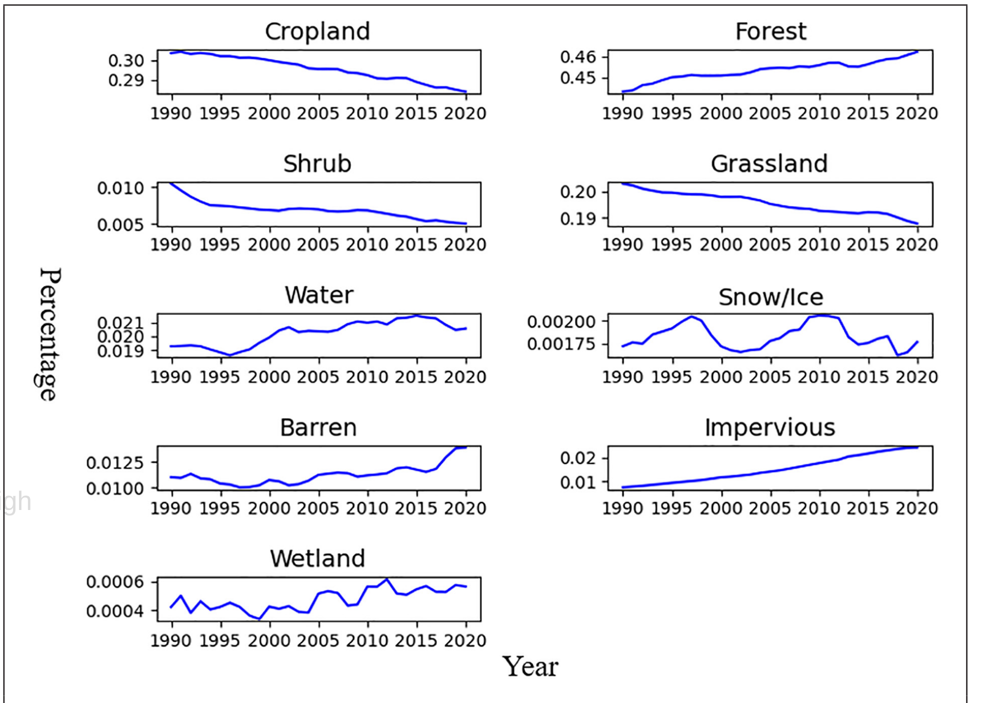

Figure 6. Statistics of land use and land cover changes in each class.

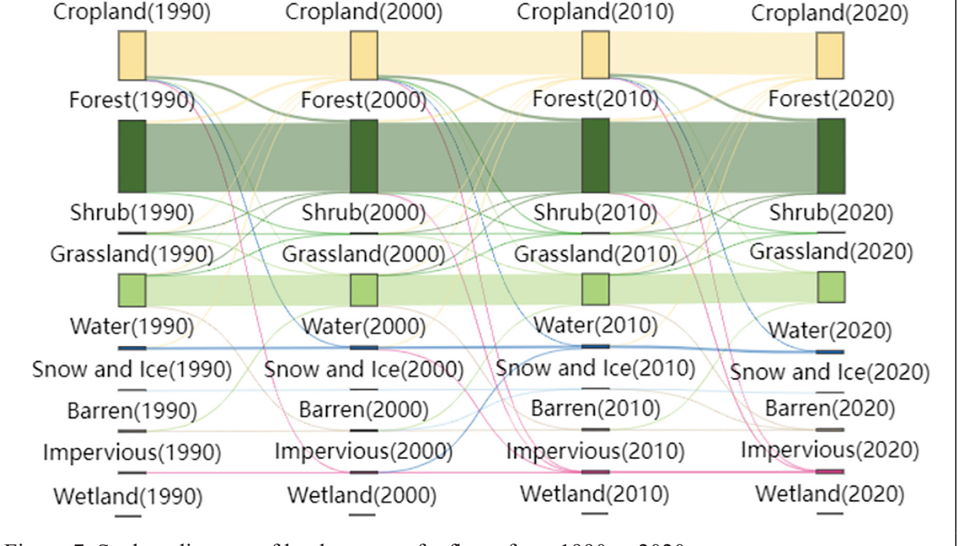

Figure 7. Sankey diagram of land use transfer flows from 1990 to 2020.

Table 5. Average of coefficients between all the factors.

	LUD	GDP	POP	NPP	TEM	PRE	NDVI	DEM
LUD								
GDP	0.58							
POP	0.63	0.93						
NPP	0.15	0.43	0.38					
TEM	0.51	0.77	0.78	0.45				
PRE	0.33	0.52	0.52	0.44	0.64			
NDVI	0.04	0.01	−0.06	0.04	0.01	−0.20		
DEM	−0.57	−0.84	−0.84	−0.44	−0.87	−0.60	−0.03	

LUD = Land use degree; GDP = gross domestic product; POP = population density; NPP = net primary productivity; TEM = annual average temperature; PRE = annual average precipitation; NDVI = Normalized Difference Vegetation Index; DEM = digital elevation model. The *p*-values for all significance tests are much less than 0.001.

Table 6. Variance inflation factor validation for different factors.

Factors	GDP	POP	NPP	TEM	PRE	NDVI	DEM
VIF	5.04	3.81	1.13	1.15	2.82	1.04	2.95

GDP = gross domestic product; POP = population density; NPP = net primary productivity; TEM = annual average temperature; PRE = annual average precipitation; NDVI = Normalized Difference Vegetation Index; DEM = digital elevation model; VIF = variance inflation factor.

We finally obtained a GTWR model with a goodness of fit of 0.91, an Akaike information criterion of 1281.36, and a spatial-temporal distance ratio of 0.2688. After a comparison of goodness of fit, the model outperformed the geographically weighted regression model (0.81) and the ordinary least squares model (0.79).

Using the university kriging interpolation, the GTWR fitting results are presented visually, and the coefficients of the respective variables are selected for the earliest year (1990) and the latest year (2015), as shown in Figure 8. From the distribution of POP coefficient values, we can see that human activity plays a driving role in the LULC of YRB, being the most obvious especially in the middle region, and the no-man's-land in the highland region can be disregarded. This result is also consistent with the results of the driving analysis of LULC changes in the Savannah River and Muga watersheds (Zurqani *et al.* 2018; Belay and Mengistu 2019). Meanwhile, natural factors also have a strong influence on LULC changes by affecting the growth of natural vegetation (Hu and Hu 2019). The DEM, on the contrary, is negatively correlated with land use degree in YRB, indicating that the plains at low elevation are suitable for land development and utilization, while the higher the elevation is, the more difficult it is to develop the area. However, the effect of PRE on land use degree is more complex, with a large amount of rainfall occurring in the middle and lower reaches of YRB, so, on the one hand, PRE explains well the development of urban areas in the middle and lower reaches, while, on the other, the forested land with lower land use degree within these areas shows a negative drive. The TEM shows an overall positive drive, but there is some counter-drive in the downstream and mid-basin regions, and the negative drive has expanded in the mid-basin.

Discussion

Accuracy Comparison with Other LULC Products

To better validate the accuracy of our results, we intercompared with CLCD, FROM_GLC, and GLCFCS30 using the Geo-Wiki validation samples. Since Geo-Wiki does not include water and wetland in the validation points of YRB, we excluded the statistics of these two in the table. As with the visual interpretation sample validation, the overall accuracy of our results in Geo-Wiki was higher than that of the other three products (CLCD: 57.91%; FROM_GLC: 54.76%; GLCFCS30: 52.72%) at 62.45%. In general, it appears that the accuracy of our results is better than that of the other products in all categories. Although

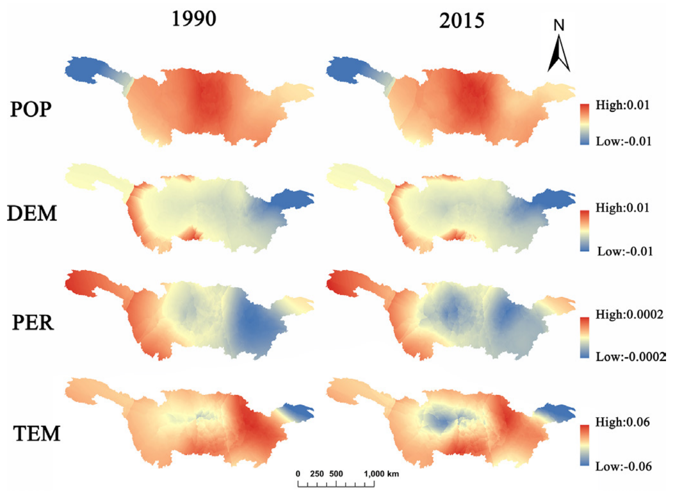

Figure 8. Coefficient distributions of the respective variables in the geographically and temporally weighted regression model.

the accuracy is lacking in shrubs and grasses, it is still excellent when compared to FROM_GLC and GLCFCS30.

Trends in LULC Change

Landsat images in 1990 and 2020 were selected and did change area coloring to highlight the change areas and to show the trends of LULC changes in YRB. As shown in Figure 9, five typical LULC changes are compared. Figure 9A illustrates the urban expansion of Shanghai, which has tripled in size in 20 years, with the rapid expansion of its main urban area and surrounding satellite cities taking over areas of previously cropland. Figure 9B shows the rapid development of fisheries around Chaohu Lake, which is an important component of the new cropland. Figure 9C illustrates the open-pit mines from 2011 and shows the spatial detail at the 30-m resolution scale. Figure 9D is the transformation of cropland in the bend of the Yangtze River, which occupies the original mudflats and wetlands. Figure 9E shows the increase in water surface area upstream of the dam before and after the Three Gorges Water Conservancy Project.

Limitations and Future Work

With the GEE platform, we made improvements to the CLCD in YRB using three thematic products (GFC, GISA, and GSW). This method has been proven to greatly reduce the workload when compared to retraining the classifier for the whole basin and to obtain good

Table 7. Comparison of mapping accuracy based on Geo-Wiki test samples for this study, CLCD, FROM_GLC, and GLCFCS30.

	Geo-Wiki							
	Cropland	Forest	Shrub	Grassland	Snow/ Ice	Barren	Impervious	OA (%)
	This study							
PA (%)	60.65	72.86	33.33	25.33	100	72.72	69.15	62.45
UA (%)	76.50	84.33	2.17	28.02	3.85	2.95	51	
	CLCD							
PA (%)	58.83	68.93	33.33	25.32	100	72.72	65.90	57.81
UA (%)	78.99	82.61	2.45	29.67	3.84	3.21	49.57	
	FROM_GLC							
PA (%)	62.24	66.57	11.11	18.46	100	50	61.71	54.76
UA (%)	54.02	79.86	4.80	28.99	1.92	11.44	41.5	
	GLCFCS30							
PA (%)	45.71	73.06	5.88	19.60	66.67	75	60.12	52.72
UA (%)	70.41	64.49	0.44	30.82	3.85	4.46	49	

CLCD = China land cover data set; OA = overall accuracy; PA = producer's accuracy; UA = user's accuracy.

accuracy. However, CLCD data are missing for products from 1986 to 1989 due to the commercial preorder acquisition plan of Landsat 5 before 1990, which further limited its availability in China before 1990 (Loveland and Dwyer 2012). In addition, based on the higher classification accuracy of the results in this article, we derived a more credible land distribution and change in YRB from 1990 to 2020. However, some driver data are not openly available, which will be the main work of our next study. In general, in the future, we would like to complement and expand the pre-1990 LULC products by combining other sensor data and collect more detailed driving factors to deeply understand the driving forces of LULC changes in YRB.

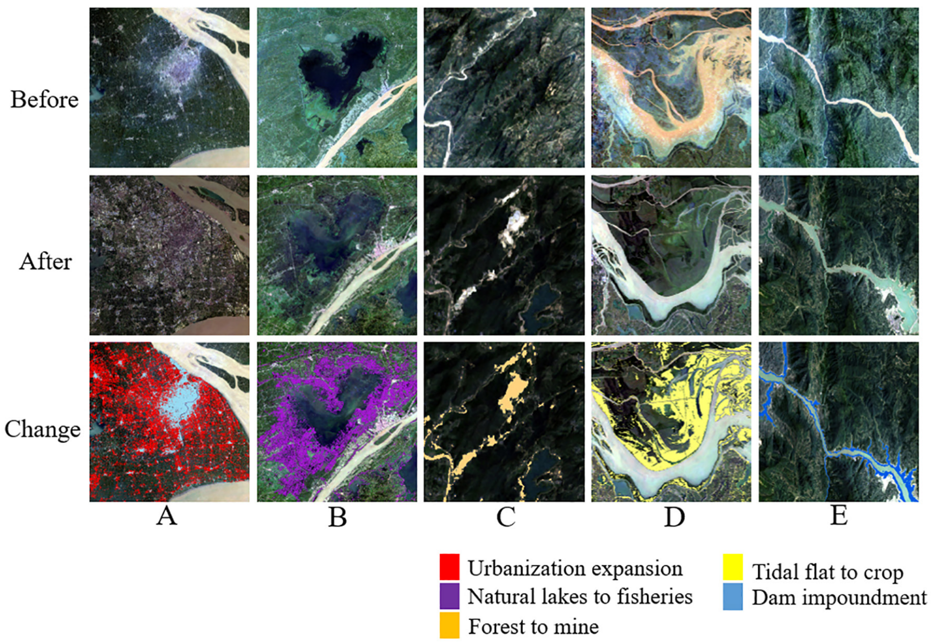

Figure 9. Land use and land cover change sample of the Yangtze River basin.

Conclusions

The YRB is of great ecological and economic significance to China. Continuous and accurate LULC mapping of YRB is important for both fine-resolution monitoring and sustainable development within the basin, and it is also a basic parameter for studying the ecological environment and climate change in the basin. In this article, we propose an optimization algorithm based on the open-access CLCD data set and produce annual 30-m LULC maps of YRB from 1990 to 2020. The results show an improvement of accuracy of about 82.66%, which is higher than CLCD's 77.21% and two other global LULC products. Similarly, in the third-party validation sample, Geo-Wiki, the results of this article also achieved higher precision when compared to the other three LULC products. In addition, the LULC changes dramatically in YRB between 1990 and 2020. The impervious surface has more than tripled, and cropland is decreasing and converting to the impervious surface, forestland, and water. Using the GTWR model, we found that anthropogenic activities play an important role in driving LULC change within YRB, while natural factors do the opposite, with both DEM and PRE factors limiting the improvement of land use degree. Therefore, it is necessary to develop a rational way for sustainable development. Finally, the annual 30-m LULC data from 1990 to 2020 in this article will be well combined with hydrological data for deeper exploration of the environment and climate change for YRB.

Acknowledgments

This work was supported by the Strategic Priority Research Program Project of the Chinese Academy of Sciences under grant XDA23040100, the Jiangsu Natural Resources Development Special Project under grant JSZRHYKJ202002, and the Shanghai Leading Talent Project under grant E056061. The data used in this study are available in Science Data Bank at https://www.scidb.cn/s/nIrqUf (doi:10.11922/sciencedb.j00116.00008).

References

Ali, R., A. Kuriqi, S. Abubaker and O. Kisi. 2019. Hydrologic alteration at the upper and middle part of the Yangtze River, China: Towards sustainable water resource management under increasing water exploitation. *Sustainability* 11(19):5176..

Belay, T. and D. A. Mengistu. 2019. Land use and land cover dynamics and drivers in the Muga watershed, Upper Blue Nile basin, Ethiopia. *Remote Sensing Applications: Society and Environment* 15:100249.

Belgiu, M. and L. Drăguţ. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114:24–31.

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

Chen, T., M. Li, L. Luo, S. Deng, R. Zhou and D. Chen. 2020. Simulating the effects of land urbanization on regional meteorology and air quality in Yangtze River Delta, China. *Applied Geography* 120:102228.

Chen, Z., J. Li, H. Shen and W. Zhanghua. 2001. Yangtze River of China: Historical analysis of discharge variability and sediment flux. *Geomorphology* 41(2–3):77–91.

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

Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily and H. K. Gibbs. 2005. Global consequences of land use. *Science* 309(5734):570–574.

Fox, J. and J. B. Vogler. 2005. Land-use and land-cover change in montane mainland southeast Asia. *Environmental Management* 36(3):394–403.

Gibbard, S., K. Caldeira, G. Bala, T. J. Phillips and M. Wickett. 2005. Climate effects of global land cover change. *Geophysical Research Letters* 32(23):1–4.

Gong, P., J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu and S. Liu. 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* 34(7):2607–2654.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202:18–27.

Guo, H., Q. Hu and T. Jiang. 2008. Annual and seasonal streamflow responses to climate and land-cover changes in the Poyang Lake basin, China. *Journal of Hydrology* 355(1–4):106–122.

Haddeland, I., J. Heinke, H. Biemans, S. Eisner, M. Florke, N. Hanasaki, M. Konzmann, F. Ludwig, Y. Masaki, J. Schewe, T. Stacke, Z. D. Tessler, Y. Wada and D. Wisser. 2014. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences* 111(9):3251–3256.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz and T. R. Loveland. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342(6160):850–853.

Hoshikawa, K. and M. Umezaki. 2014. Effects of terrain-induced shade removal using global DEM data sets on land-cover classification. *International Journal of Remote Sensing* 35(4):1331–1355.

Houghton, R. A., J. I. House, J. Pongratz, G. R. van der Werf, R. S. DeFries, M. C. Hansen, C. Le Quéré and N. Ramankutty. 2012. Carbon emissions from land use and land-cover change. *Biogeosciences* 9(12):5125–5142.

Hu, Y. and Y. Hu. 2019. Land cover changes and their driving mechanisms in Central Asia from 2001 to 2017 supported by Google Earth Engine. *Remote Sensing* 11(5):554.

Huang, B., B. Wu and M. Barry. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science* 24(3):383–401.

Huang, X., J. Li, J. Yang, Z. Zhang, D. Li and X. Liu. 2021. 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019. *Science China Earth Sciences* 64(11):1922–1933.

Jin, S. and T. Zhang. 2016. Terrestrial water storage anomalies associated with drought in southwestern USA from GPS observations. *Surveys in Geophysics* 37(6):1139–1156.

Li, W., L. Li, J. Chen, Q. Lin and H. Chen. 2021. Impacts of land use and land cover change and reforestation on summer rainfall in the Yangtze River basin. *Hydrology and Earth System Sciences* 25(8):4531–4548.

Liu, C., W. Li, G. Zhu, H. Zhou, H. Yan and P. Xue. 2020d. Land use/land cover changes and their driving factors in the northeastern Tibetan Plateau based on geographical detectors and Google Earth Engine: A case study in Gannan Prefecture. *Remote Sensing* 12(19):3139.

Liu, D., N. Chen, X. Zhang, C. Wang and W. Du. 2020a. Annual large-scale urban land mapping based on Landsat time series in Google Earth Engine and OpenStreetMap data: A case study in the middle Yangtze River basin. *ISPRS Journal of Photogrammetry and Remote Sensing* 159:337–351.

Liu, H., P. Gong, J. Wang, N. Clinton, Y. Bai and S. Liang. 2020b. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth System Science Data* 12(2):1217–1243.

Liu, H., M. Zheng, J. Liu and X. Zheng. 2020c. Sustainable land use in the trans-provincial marginal areas in China. *Resources, Conservation and Recycling* 157:104783.

Liu, J., W. Kuang, Z. Zhang, X. Xu, Y. Qin, J. Ning, W. Zhou, S. Zhang, R. Li and C. Yan. 2014. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *Journal of Geographical Sciences* 24(2):195–210.

Loveland, T. R. and J. L. Dwyer. 2012. Landsat: Building a strong future. *Remote Sensing of Environment* 122:22–29.

Ma, X., J. Zhang, C. Ding and Y. Wang. 2018. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems* 70:113–124.

Mahdavi, S., B. Salehi, J. Granger, M. Amani, B. Brisco and W. Huang. 2018. Remote sensing for wetland classification: A comprehensive review. *GIScience and Remote Sensing* 55(5):623–658.

Myers, L. and M. J. Sirois. 2004. Spearman correlation coefficients, differences between. In *Encyclopedia of Statistical Sciences*, edited by Myers, L. and M. J. Sirois., 1–2. New York: Wiley.

Na, X., S. Zhang, X. Li, H. Yu and C. Liu. 2010. Improved land cover mapping using random forests combined with Landsat thematic mapper imagery and ancillary geographic data. *Photogrammetric Engineering and Remote Sensing* 76(7):833–840.

Pekel, J. F., A. Cottam, N. Gorelick and A. S. Belward. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540(7633):418–422.

Qu, L. A., M. Li, Z. Chen and J. Zhi. 2021. A modified self-adaptive method for mapping annual 30-m land use/land cover using Google Earth Engine: A case study of Yangtze River Delta. *Chinese Geographical Science* 31(5):782–794.

Ran, Q., Y. Hao, A. Xia, W. Liu, R. Hu, X. Cui, K. Xue, X. Song, C. Xu, B. Ding and Y. Wang. 2019. quantitative assessment of the impact of physical and anthropogenic factors on vegetation spatial-temporal variation in northern Tibet. *Remote Sensing* 11(10):1183.

Robbins, A. S. and S. Harrell. 2014. Paradoxes and challenges for China's forests in the reform era. *China Quarterly* 218:381–403.

Rui, S., P. Yao, W. Wen, B. Yue and L. Gang. 2017. Assessment of wetland ecosystem health in the Yangtze and Amazon River Basins. *International Journal of Geo-Information* 6(3):81.

Sulla-Menashe, D., J. M. Gray, S. P. Abercrombie and M. A. Friedl. 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment* 222:183–194.

Sun, R., P. Yao, W. Wang, B. Yue and G. Liu. 2017. Assessment of wetland ecosystem health in the Yangtze and Amazon River Basins. *ISPRS International Journal of Geo-Information* 6(3):81.

Szabó, S., Z. Gácsi and B. Balázs. 2016. Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Landscape and Environment* 10(3–4):194–202.

Tran, C. P., R. W. Bode, A. J. Smith and G. S. Kleppel. 2010. Land-use proximity as a basis for assessing stream water quality in New York State (USA). *Ecological Indicators* 10(3):727–733.

Vorosmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann and P. M. Davies. 2010. Global threats to human water security and river biodiversity. *Nature* 467(7315):555–561.

Wang, S. -Y., J. -S. Liu, and T. -B. Ma. 2010. Dynamics and changes in spatial patterns of land use in Yellow River Basin, China. *Land Use Policy* 27(2):313–323.

Wu, J., S. Zeng, L. Yang, Y. Ren and J. Xia. 2021. Spatiotemporal characteristics of the water quality and its multiscale relationship with land use in the Yangtze River Basin. *Remote Sensing* 13(16):3309.

Xu, H. 2007. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27(14):3025–3033.

Xu, L., G. Yu, Z. Tu, Y. Zhang and N. E. Tsendbazar. 2020. Monitoring vegetation change and their potential drivers in Yangtze River Basin of China from 1982 to 2015. *Environmental Monitoring and Assessment* 192(10):642.

Yang, H. F., S. L. Yang, K. H. Xu, J. D. Milliman, H. Wang, Z. Yang, Z. Chen and C. Y. Zhang. 2018. Human impacts on sediment in the Yangtze River: A review and new perspectives. *Global and Planetary Change* 162:8–17.

Yang, H., X. Zhong, S. Deng and H. Xu. 2021b. Assessment of the impact of LUCC on NPP and its influencing factors in the Yangtze River basin, China. *Catena* 206:105542.

Yang, H. F., S. L. Yang, K. H. Xu, J. D. Milliman, H. Wang, Z. Yang, Z. Chen and C. Y. Zhang. 2018. Human impacts on sediment in the Yangtze River: A review and new perspectives. *Global and Planetary Change* 162:8–17.

Yang, J. and X. Huang. 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth System Science Data* 13(8):3907–3925.

Yang, L., L. Wang, D. Yu, R. Yao, Q. He, S. Wang and L. Wang. 2020. Four decades of wetland changes in Dongting Lake using Landsat observations during 1978–2018. *Journal of Hydrology* 587:124954.

Yang, S. L., K. H. Xu, J. D. Milliman, H. F. Yang and C. S. Wu. 2015. Decline of Yangtze River water and sediment discharge: Impact from natural and anthropogenic changes, *Scientific Reports* 5:12581.

Yang, S. L., J. Zhang, J. Zhu, J. P. Smith, S. Dai, A. Gao and P. Li. 2005. Impact of dams on Yangtze River sediment supply to the sea and delta intertidal wetland response. *Journal of Geophysical Research: Earth Surface* 110(F3):1–12.

Yang, X., F. Meng, P. Fu, Y. Zhang and Y. Liu. 2021a. Spatiotemporal change and driving factors of the eco-environment quality in the Yangtze River Basin from 2001 to 2019. *Ecological Indicators* 131:108214.

Zhang, D., X. Liu and P. Bai. 2019. Assessment of hydrological drought and its recovery time for eight tributaries of the Yangtze River (China) based on downscaled GRACE data. *Journal of Hydrology* 568:592–603.

Zhang, F. and X. Yang. 2020. Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment* 251:112105.

Zhang, X., L. Liu, X. Chen, Y. Gao, X. Xie and J. Mi. 2021. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data* 13(6):2753–2776.

Zhu, Z. and C. E. Woodcock. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment* 118:83–94.

Zhu, Z., M. A. Wulder, D. P. Roy, C. E. Woodcock, M. C. Hansen, V. C. Radeloff, S. P. Healey, C. Schaaf, P. Hostert, P. Strobl, J. -F. Pekel, L. Lymburner, N. Pahlevan and T. A. Scambos. 2019. Benefits of the free and open Landsat data policy. *Remote Sensing of Environment* 224:382–385.

Zurqani, H. A., C. J. Post, E. A. Mikhailova, M. A. Schlautman and J. L. Sharp. 2018. Geospatial analysis of land use change in the Savannah River Basin using Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation* 69:175–185.