ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202402.00215

Land use and cover change and influencing factor analysis in the Shiyang River Basin, China postprint

Authors: ZHAO Yaxuan

Date: 2024-02-21T14:54:47+00:00

Abstract

Land use and cover change (LUCC) is the most direct manifestation of the interaction between anthropological activities and the natural environment on Earth's surface, with significant impacts on the environment and social economy. Rapid economic development and climate change have resulted in significant changes in land use and cover. The Shiyang River Basin, located in the eastern part of the Hexi Corridor in China, has undergone significant climate change and LUCC over the past few decades. In this study, we used the random forest classification to obtain the land use and cover datasets of the Shiyang River Basin in 1991, 1995, 2000, 2005, 2010, 2015, and 2020 based on Landsat images. We validated the land use and cover data in 2015 from the random forest classification results (this study), the high-resolution dataset of annual global land cover from 2000 to 2015 (AGLC-2000-2015), the global 30 m land cover classification with a fine classification system (GLC_{FCS30}), and the first Landsat-derived annual China Land Cover Dataset (CLCD) against ground-truth classification results to evaluate the accuracy of the classification results in this study. Furthermore, we explored and compared the spatiotemporal patterns of LUCC in the upper, middle, and lower reaches of the Shiyang River Basin over the past 30 years, and employed the random forest importance ranking method to analyze the influencing factors of LUCC based on natural (evapotranspiration, precipitation, temperature, and surface soil moisture) and anthropogenic (nighttime light, gross domestic product (GDP), and population) factors. The results indicated that the random forest classification results for land use and cover in the Shiyang River Basin in 2015 outperformed the AGLC-2000-2015, GLC_{FCS30}, and CLCD datasets in both overall and partial validations. Moreover, the classification results in this study exhibited a high level of agreement with the ground truth features. From 1991 to 2020, the area of bare land exhibited a decreasing trend, with changes primarily occurring in the middle and lower reaches of the basin. The area of grassland initially decreased and then increased, with changes occurring mainly in the upper and middle reaches of the basin. In contrast, the area of cropland initially increased and then decreased, with changes occurring in the middle and lower reaches. The LUCC was influenced by both natural and anthropogenic factors. Climatic factors and population contributed significantly to LUCC, and the importance values of evapotranspiration, precipitation, temperature, and population were 22.12%, 32.41%, 21.89%, and 19.65%, respectively. Moreover, policy interventions also played an important role. Land use and cover in the Shiyang River Basin exhibited fluctuating changes over the past 30 years, with the ecological environment improving in the last 10 years. This suggests that governance efforts in the study area have had some effects, and the government can continue to move in this direction in the future. The findings can provide crucial insights for related research and regional sustainable development in the Shiyang River Basin and other similar arid and semi-arid areas.

Full Text

Land Use and Cover Change and Influencing Factor Analysis in the Shiyang River Basin, China

Yaxuan Zhao[1], Bo Cao[1],[2]*, Linwei Sha[1], Jinquan Cheng[1], Xuanru Zhao[1], Weijin Guan[1], Baotian Pan[1],[2]

[1]Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730030, China
[2]Shiyang River Basin Scientific Observing Station, Lanzhou University, Lanzhou 730030, China

---

Abstract

Land use and cover change (LUCC) represents the most direct manifestation of interactions between human activities and the natural environment at Earth's surface, exerting significant impacts on both environmental conditions and socioeconomic systems. Rapid economic development and climate change have triggered substantial alterations in land use and cover patterns. The Shiyang River Basin, situated in the eastern Hexi Corridor of China, has experienced pronounced climate change and LUCC over recent decades. This study employed random forest classification to generate land use and cover datasets for the Shiyang River Basin for the years 1991, 1995, 2000, 2005, 2010, 2015, and 2020 based on Landsat imagery. We validated the 2015 land use and cover data from our random forest classification results against three existing datasets—the high-resolution annual global land cover dataset from 2000 to 2015 (AGLC-2000-2015), the global 30 m land cover classification with a fine classification system (GLC_{FCS30}), and the first Landsat-derived annual China Land

Cover Dataset (CLCD)—as well as against ground-truth classification results to evaluate classification accuracy. Furthermore, we examined and compared spatiotemporal LUCC patterns across the upper, middle, and lower reaches of the basin over the past 30 years, and utilized the random forest importance ranking method to analyze influencing factors based on both natural (evapotranspiration, precipitation, temperature, and surface soil moisture) and anthropogenic (nighttime light, gross domestic product (GDP), and population) variables.

The results demonstrated that our random forest classification for 2015 outperformed the AGLC-2000-2015, GLC_{FCS30}, and CLCD datasets in both overall and partial validations, exhibiting strong agreement with ground-truth features. From 1991 to 2020, bare land area displayed a decreasing trend, with changes concentrated primarily in the middle and lower reaches. Grassland area initially declined before increasing, with changes mainly occurring in the upper and middle reaches. Conversely, cropland area first increased then decreased, with changes focused in the middle and lower reaches. LUCC was influenced by both natural and anthropogenic factors, with climatic variables and population contributing significantly. The importance values for evapotranspiration, precipitation, temperature, and population were 22.12%, 32.41%, 21.89%, and 19.65%, respectively. Policy interventions also played a crucial role. Land use and cover in the Shiyang River Basin exhibited fluctuating changes over the 30-year study period, with ecological conditions improving during the last decade. This suggests that governance efforts have been effective and should continue. These findings provide critical insights for related research and regional sustainable development in the Shiyang River Basin and other comparable arid and semi-arid regions.

Keywords: land use and cover classification; land use and cover change (LUCC); climate change; random forest; accuracy assessment; three-dimensional sampling method; Shiyang River Basin

Citation: ZHAO Yaxuan, CAO Bo, SHA Linwei, CHENG Jinquan, ZHAO Xuanru, GUAN Weijin, PAN Baotian. 2024. Land use and cover change and influencing factor analysis in the Shiyang River Basin, China. Journal of Arid Land, 16(2): 246–265. https://doi.org/10.1007/s40333-024-0071-6

---

1. Introduction

The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) confirmed that anthropogenic activities have caused unprecedented climate warming, with global average surface temperatures rising by 1.1°C above pre-industrial levels [?]. Climate change has produced irreversible impacts including glacier melting, increased extreme weather events, crop yield reductions, shortened growing seasons, wetland degradation, and desertification [?, ?, ?, ?, ?, ?, ?]. These impacts have progressively reshaped land use and cover patterns [?], significantly affecting Earth's surface systems, particularly regarding biodiversity, surface and subsurface runoff, carbon cycling and storage, agricultural land use, urban resource availability, soil salinization, and transportation- and pollution-related ecological problems [?, ?, ?, ?, ?, ?, ?, ?, ?, ?].

Current land use and cover classification methods include visual interpretation, single-band thresholding, spectral relationship analysis, object-oriented approaches [?, ?, ?], and machine learning techniques [?]. Research indicates that random forest classification—a machine learning method—outperforms other approaches in both accuracy and operational efficiency [?, ?]. Xu et al. [?] employed multi-data fusion, temporal change detection, and random forest to develop the high-resolution annual global land cover dataset from 2000 to 2015 (AGLC-2000-2015). Zhang et al. [?] combined high-quality training data from the Global Spatial Temporal Spectra Library (GSPECLib) on Google Earth Engine with random forest classification to create the global 30 m land cover product with a fine classification system (GLC_{FCS30}) from 1985 to 2020 at 5-year intervals. Yang and Huang [?] utilized random forest on Google Earth Engine, based on the China land use/cover dataset, satellite time-series imagery, and visual interpretation samples from Google Earth, to produce the first Landsat-derived annual China Land Cover Dataset (CLCD), comprising 30 m annual land cover and change dynamics across China from 1990 to 2019.

To effectively address ecological problems caused by LUCC, analyzing its influencing factors is essential [?]. Current research on LUCC drivers remains inadequate, relying primarily on qualitative analyses such as logistic regression, multiple linear regression, and principal component analysis [?, ?, ?]. However, these methods are relatively simplistic, depend on linear models, and are less effective at capturing complex driving mechanisms [?]. In contrast, random forest—an ensemble machine learning approach—effectively handles highly correlated data and multidimensional features while combating overfitting. It can manage numerous quantitative and qualitative explanatory variables and systematically rank input variable importance, thereby overcoming certain limitations of traditional methods. Random forest demonstrates strong performance in both classification and regression tasks [?, ?, ?, ?] and has been widely applied in land use and cover analysis [?, ?, ?].

The Shiyang River Basin occupies a unique geographical position at the intersection of the eastern monsoon region, the arid northwest, and the Qinghai-Tibet Plateau [?]. The upper Qilian Mountains serve as important ecological barriers in western China, while the lower plains function as river corridors separating the Badain Jaran and Tengger Deserts [?, ?]. Researchers analyzing multiple aspects of the basin have concluded that it has been in an unhealthy state over recent decades [?, ?]. Current land use and cover classification studies in the Shiyang River Basin employ unsupervised and supervised classification combined with visual interpretation, while LUCC driver analyses rely on qualitative, principal component, and logistic regression methods [?, ?, ?, ?, ?]. Despite numerous studies, limitations remain: LUCC time series are insufficiently long, different classification methods yield inconsistent results complicating accuracy assessment, and factor analyses often depend on qualitative approaches or simplistic linear models that inadequately capture complex driving mechanisms.

Therefore, this study aimed to explore spatiotemporal LUCC patterns and influencing factors while comparing differences across the upper, middle, and lower reaches of the Shiyang River Basin over the past 30 years. Compared to previous research, our improvements include: a longer temporal coverage, adoption of a three-dimensional (3D) sampling method to enhance classification accuracy, and explicit subdivision of the study area into three sub-regions for more detailed analysis. Analyzing LUCC in the Shiyang River Basin is crucial for regional sustainable development and provides a theoretical foundation for studies in similar arid and semi-arid regions.

2.1 Study Area

The Shiyang River Basin (101°07 –104°15 E, 37°07 –39°28 N) is located in the eastern portion of the arid region of Northwest China, within the eastern Hexi Corridor at the northern foot of the Qilian Mountains [**?**] (Fig. 1). The basin covers an area of 4.1$×10^{4}$ km^2 and contains eight rivers. The region experiences an annual average temperature of 6.5°C, annual precipitation ranging from 50–600 mm, and potential evaporation of 700–2600 mm [**?**, **?**]. The total population is approximately 2.16$×10^{6}$ persons [**?**], with a gross domestic product (GDP) of 3.03$×10^{10}$ USD [**?**]. Elevation varies from 1245 to 5214 m, with higher terrain in the south and lower terrain in the north. The upper and middle reaches are divided based on the 2000 m contour and land use/cover integrity, while the boundary between middle and lower reaches follows the Minqin County borderline. The three sub-regions (upper, middle, and lower reaches) cover areas of 1.2$×10^{\{4\},1.3×10}\{4\}, and 1.6×10^{4}$ km^2, respectively.

2.2 Data Sources and Preprocessing

More than 40 Landsat images at 30 m spatial resolution were acquired from the USGS (https://earthexplorer.usgs.gov/) for the period 1991–2020. The selected years were 1991, 1995, 2000, 2005, 2010, 2015, and 2020, focusing primarily on June–September imagery. Adjacent years were used as references when image quality was poor (Table 1). All images underwent radiometric calibration and atmospheric correction. The classification results were validated against three domestically published datasets—AGLC-2000-2015, GLC_{FCS30}, and CLCD [**?**, **?**, **?**]—which were reclassified to match our classification system. Gaofen-1 (GF-1) imagery from 2015, obtained from the China Center for Resources Satellite Data and Application (https://data.cresda.cn), was used for partial validation.

The Normalized Difference Vegetation Index (NDVI) product dataset (1998–2019) was obtained from the Resource and Environment Science and Data Center (https://www.resdc.cn) at 1 km spatial resolution and annual temporal resolution. Evapotranspiration (mm), precipitation (mm), temperature (°C), surface soil moisture (m^3/m^3), nighttime light, GDP ($×10^{6}$ USD), and population (persons/km^2) were employed to analyze LUCC drivers (Table 2). Annual values were used for all factors. Original evapotranspiration, temperature, and precipitation data were at monthly resolution; we calculated annual averages for temperature and annual totals for evapotranspiration and precipitation for each study year. Surface soil moisture data were originally at daily resolution; we performed format conversion and calculated annual averages from daily values. Nighttime light, GDP, and population data were originally at annual resolution. For GDP, 2019 data were used for 2020 due to data availability (Table 2). After obtaining data for all corresponding years, we adopted a uniform coordinate system and performed cropping to ensure consistency.

2.3.1 Land Use and Cover Classification and Accuracy Assessment

Land use and cover classification was performed using random forest classification in ENVI 5.3, with results subsequently refined through visual interpretation using Google Earth. The classification system was based on frameworks from the Chinese Academy of Sciences, the United States Geological Survey, and the FROM-GLC database [?, ?, ?]. Considering the unique characteristics of the Shiyang River Basin, the final system comprised eight land use and cover types: bare land, grassland, cropland, forest, wetland, impervious surface, water body, and glacier (Table 3).

A 3D sampling method was employed to select region of interest (ROI) samples following Yang et al. [?] (Fig. 2). This approach applied different band combinations, vegetation fractional coverage, and Google Earth views to original images to generate a 3D terrain scene, created 3D samples, projected them onto original images, and made further adjustments to obtain final samples. The separability of ROI samples exceeded 1.88. Vegetation fractional coverage was computed in ArcGIS using Landsat images to calculate NDVI, followed by a pixel-wise binary model to determine pure vegetation and pure soil pixel values, then applying the following formulas [?]:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

where NDVI is the Normalized Difference Vegetation Index, NIR is near-infrared band reflectance, and RED is red-light band reflectivity.

$$VFC = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}}$$

where VFC represents vegetation fractional coverage, NDVI_{soil} is the NDVI value for pure bare ground pixels, and NDVI_{veg} is the NDVI value for pure vegetation pixels. The upper and lower NDVI thresholds were determined using a 5% confidence interval to obtain NDVI_{soil} and NDVI_{veg}.

Accuracy validation of the 2015 land use and cover data from our random forest classification (Landsat_{RFC}) and the three comparative datasets (AGLC-2000-2015, GLC_{FCS30}, and CLCD) was conducted through overall and partial validations. For overall validation, 601 verification points were uniformly distributed throughout the basin (Fig. 1). Visual interpretation using Google Earth established ground-truth classifications, and classification accuracy rates (Eq. 3) were calculated by comparing each dataset's results against these reference points. For partial validation, three 100.0 km^2 verification regions were selected in each of the upper, middle, and lower reaches. Ground-truth classifications were derived by applying random forest classification to GF-1 imagery and correcting with Google Earth (Kappa coefficient >0.93). Overall accuracy (Eq. 4) and Kappa coefficient (Eq. 5) were then calculated for all four datasets using ENVI.

$$Ar = \frac{N_c}{601} \times 100\%$$

where Ar is the classification accuracy rate (%) and N_c is the number of correctly classified verification points.

$$OA = \frac{\sum_{i=1}^{n} N_{ii}}{N} \times 100\%$$

where OA represents overall accuracy (%), n is the number of classes, N_{ii} is the number of correctly classified pixels, and N is the total number of samples.

$$k = \frac{N\sum_{i=1}^{n} N_{ii} - \sum_{i=1}^{n} N_{i+}N_{+i}}{N^2 - \sum_{i=1}^{n} N_{i+}N_{+i}}$$

where k is the Kappa coefficient, and N_{i+} and N_{+i} represent the sums of class i in the classified and validation data, respectively.

Additionally, field verification was conducted at 29 points to evaluate the Landsat_{RFC} results for 2020 (Fig. 3).

2.3.2 LUCC Analysis

We calculated the area of each land use and cover type for each year (1991, 1995, 2000, 2005, 2010, 2015, and 2020) in ArcGIS. Based on these data, we analyzed the proportional coverage of each type in 2020 and the spatiotemporal change characteristics from 1991 to 2020. Furthermore, we employed a land use transition matrix to analyze conversions between different land use and cover types. This two-dimensional matrix captures the dynamic conversion relationships in the same region across different time periods, reflecting changes in the quantity and direction of land use and cover [?, ?]. The formula is expressed as:

$$S_{mn} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{bmatrix}$$

where S_{mn} is the area (km^2) of land use and cover type m in the previous period converted to type n in the later period, and k is the number of land use and cover types.

2.3.3 Analysis of Influencing Factors of LUCC

Random forest is an ensemble learning algorithm proposed by Breiman [?] that combines multiple decision trees. The method employs bootstrap resampling, randomly selecting multiple samples with replacement from the original training set to create new training subsets, then generating numerous classification trees to form a random forest. Classification results for new data are determined through majority voting across all trees. Feature importance is assessed by observing each feature's contribution to individual trees, calculating average contributions, and comparing contribution levels [?, ?].

In LUCC driver analysis, scholars typically use multi-year averages of factors such as Digital Elevation Model (DEM), precipitation, temperature, population, GDP, evapotranspiration, distance to rivers and roads, nighttime light, and surface soil moisture as independent variables [?, ?, ?, ?, ?]. We argue that changes in these factors drive LUCC, allowing more accurate consideration of multiple factor impacts. The analysis was completed using the Sklearn library from Python's random forest package. We used the difference in each influencing factor between the beginning and end years of a period as independent variables; for example, the difference between 2020 and 2015 precipitation values was used for analyzing 2015–2020 LUCC drivers. Due to data limitations, surface soil moisture and population data were included starting in 2000. The dependent variable was constructed using land use and cover classification results from period beginning and end years in an 'xy' format, where 'x' represents the initial year classification and 'y' represents the final year classification. Importance values of influencing factors were obtained for six periods (1991–1995, 1995–2000, 2000–2005, 2005–2010, 2010–2015, and 2015–2020) and averaged to derive comprehensive results for 1991–2020.

3.1 Land Use and Cover Classification and Accuracy Assessment Results

The Landsat_{RFC} datasets from 1991 to 2020 were generated using random forest classification and visual interpretation according to the classification system in Table 3. Notable land use and cover changes occurred during this period (Fig. 4). Overall accuracy and Kappa coefficients were calculated for all seven years, with overall accuracy exceeding 91.10% and Kappa coefficients greater than 0.84. Generally, Kappa values above 0.75 indicate high model effectiveness and accuracy [?], confirming the reliability of our classification results.

Comparative validation of Landsat_{RFC} (2015) against the three other datasets revealed superior performance (Table 4). Landsat_{RFC} achieved a classification accuracy rate of 92.01%, outperforming all comparative datasets. In partial validation across upper, middle, and lower reaches, Landsat_{RFC} (2015) also achieved the highest accuracy, with overall accuracy values of 76.05%, 67.66%, and 84.33%, respectively. Field verification in 2020 showed high agreement between Landsat_{RFC} and ground-truth features, confirming high classification accuracy.

3.2 Land Use and Cover Pattern in 2020

In 2020, bare land dominated the Shiyang River Basin, covering 49.05% of the total area, followed by grassland (33.90%), cropland (10.48%), and forest (5.23%) (Fig. 5). In the upper reaches, grassland, forest, and bare land accounted for 98.21% of the area, with proportions of 72.13%, 18.51%, and 7.57%, respectively. In the middle reaches, bare land (41.57%), grassland (33.17%), and cropland (22.61%) comprised the major land types. The lower reaches were dominated by bare land (84.42%), with grassland (7.43%) and cropland (7.13%) as secondary types (Fig. 5). As shown in Figures 4 and 6, bare land was primarily distributed in the lower reaches; grassland was concentrated in the upper and middle reaches; cropland and impervious surfaces were mainly in the middle reaches; forest and glacier were predominantly in the upper reaches; wetland was primarily found in the lower reaches; and water bodies were relatively evenly distributed across all three sub-regions.

3.3 Spatiotemporal Analysis of LUCC During 1991–2020

NDVI serves as a crucial remote sensing indicator for vegetation monitoring and land use/cover assessment [?]. Understanding NDVI variations is paramount for targeted land use and cover restoration and conservation efforts [?]. Analysis revealed an overall increasing NDVI trend in the Shiyang River Basin from

1998 to 2019 (Fig. 7), indicating positive ecological development. Among sub-regions, the upper reaches exhibited the fastest rate of change, more than double that of the entire basin.

LUCC dynamics are illustrated in Figures 8 and 9. From 1991 to 2020, bare land decreased by 3408.7 km^2, primarily in the middle (-2672.2 km^2) and lower (-998.5 km^2) reaches, though it increased by 262.0 km^2 in the upper reaches. Despite bare land being predominantly distributed in the lower reaches (Fig. 6), the most significant changes occurred in the middle reaches. Grassland area decreased by 1122.3 km^2 from 1991 to 2010, mainly in the upper reaches, then increased by 2438.0 km^2 from 2010 to 2020, with gains concentrated in the upper and middle reaches. Cropland area increased by 1649.6 km^2 from 1991 to 2010, then decreased by 618.5 km^2 from 2010 to 2020, with changes primarily in the middle and lower reaches. Forest area increased by 769.8 km^2 overall, mainly in the upper reaches. Impervious surface area increased by 365.7 km^2, primarily in the middle reaches. Wetland and glacier areas decreased by 67.1 km^2 and 9.9 km^2, respectively, with wetland loss concentrated in the lower reaches and glacier reduction in the upper reaches. Water body area decreased by 9.7 km^2 from 1991 to 2000 (mainly -7.4 km^2 in upper reaches), then increased by 13.0 km^2 from 2000 to 2020, with gains primarily in the lower reaches.

The land use and cover transfer diagram (Fig. 10) reveals that decreased bare land was primarily converted to grassland, followed by cropland. Cropland expansion from 1991 to 2010 resulted mainly from conversion of bare land and grassland, while post-2010 reductions were attributed to conversion back to bare land and small amounts of impervious surface. Grassland transformed into bare land and cropland from 1991 to 2010, but increased after 2010 through conversion from bare land. Continuous impervious surface expansion primarily resulted from conversion of bare land and cropland. Decreasing glacier area was mainly transformed into bare land. Water body changes involved mutual conversion with bare land. Generally decreasing wetland area was converted to bare land. The overall increasing forest trend resulted from conversion from grassland.

3.4 Analysis of the Influencing Factors of LUCC

The random forest importance ranking method assessed factor importance across the entire basin and three sub-regions (Fig. 11). Basin-wide, LUCC was primarily influenced by precipitation (32.41%), evapotranspiration (22.12%), temperature (21.89%), and population (19.65%). In the upper reaches, drivers included precipitation, population, evapotranspiration, and temperature. In the middle reaches, population became the primary factor, exceeding natural factors in importance. In the lower reaches, surface soil moisture (32.96%) was the most influential factor, followed by population, precipitation, and temperature. Overall, precipitation exerted greater influence than evapotranspiration and temperature, while among anthropogenic factors, population had the most significant impact, particularly in the middle reaches.

---

4. Discussion

4.1 Analysis of the Influencing Factors of LUCC from the Aspect of Land Use and Cover Types Our random forest classification revealed distinct change trends among land use and cover types (Fig. 8). Bare land area decreased overall as various crops were cultivated and urban construction expanded onto bare land to meet human needs and pursue rapid economic development [?]. Fluctuating climate conditions (Fig. 12) caused mutual conversion between bare land and grassland, particularly at their boundaries. Climate change indirectly affects vegetation cover [?], and before the 2010 turning point, accumulated precipitation and temperature created more favorable water and thermal conditions that promoted plant growth, increased evapotranspiration, and facilitated conversion of bare land to grassland.

Cropland area initially increased then decreased, with a turning point coinciding with that of grassland. Forest area showed an upward trend. The "grain-oriented" ideology proposed in the late 1990s and rural tax reduction policies implemented in 2006 promoted reclamation of abandoned land, increasing grain production and converting grassland to cropland [?, ?]. Additionally, China's Western Development Strategy and Grain for Green Program, initiated in 2000, transformed steep croplands (slopes >25°) and bare land into forest and grassland to restore ecosystems [?].

Impervious surface area continued to increase, driven by population growth and urban development. Field inspections in 2015 revealed numerous photovoltaic installations to meet growing energy demands [?]. Water body area initially decreased, reaching a minimum around 2000, then increased, while wetland area decreased. Excessive water resource exploitation during rapid economic development reduced water area [?], but the 2007 "Restoration Plan for the Shiyang River Basin" approved by the Chinese government initiated management and protection measures that increased water area [?]. However, continued human activities slowly reduced water body area in the middle reaches. Wetland decrease may be attributed to warming climate [?]. Glacier area decreased steadily due to climate change, consistent with Qilian Mountains trends [?].

Our land use and cover type trends, except for forest, align with existing research [?, ?, ?, ?]. While some studies found declining forest area [?, ?], ours showed increasing forest, consistent with Yang and Huang's [?] national-scale analysis of China from 1985–2019, validating our classification accuracy despite discrepancies likely arising from classification standard and study area variations.

4.2 Analysis of the Influencing Factors of LUCC at the Sub-Regional Level The random forest importance ranking method identified core influencing factors across the basin and sub-regions. A distinguishing feature of this study is using inter-annual differences rather than constants as variables. Precipitation, temperature, evapotranspiration, and population emerged as main drivers (Fig. 11), with policies exerting significant impacts. These findings align closely with previous Shiyang River Basin studies identifying climate, population, topography, and policy as primary LUCC factors [?, ?, ?, ?, ?], suggesting high reliability.

In the whole basin and upper/middle reaches, precipitation, temperature, evapotranspiration, and population were primary drivers. Climate change affected bare land-grassland interconversion, glacier reduction, and water body changes, while population influenced impervious surface expansion and cropland changes, with policies also playing roles [?, ?, ?, ?, ?, ?]. In the middle reaches, cropland and impervious surface accounted for 69.51% and 84.73% of their respective total basin areas, and population impact on LUCC was greater than in other sub-regions.

In the lower reaches, surface soil moisture, temperature, precipitation, and population were key factors. Bare land, grassland, and cropland occupied large proportions, with bare land comprising 85.80% of the sub-regional area. The limited grassland and cropland coverage restricted soil moisture retention and evaporation reduction capacity [?]. Recent ecological governance initiatives such as "Ant Forest" have alleviated land degradation and desertification [?], while climate change enhanced soil moisture retention. Increased rainfall and decreased evaporation raised surface soil moisture, facilitating conversion from bare land to grassland and cropland [?, ?]. Human activities constrained impervious surface expansion and cropland changes. Furthermore, the 2002 promotion of Minqin Liangucheng Nature Reserve to National Nature Reserve status significantly enhanced vegetation cover, mitigated desertification, and improved the ecological environment [?], contributing to bare land reduction and grassland increase after 2000.

---

5. Conclusions

This study analyzed the Shiyang River Basin using ArcGIS, ENVI, Python programming, and random forest classification to generate land use and cover datasets from 1991–2020. The classification process integrated Google Earth imagery, band combinations, and vegetation fractional coverage to enhance accuracy. We analyzed the 2020 spatial distribution, temporal changes, and transitions of land use and cover types from 1991–2020, and assessed natural and anthropogenic factor influences using random forest importance ranking.

1. **Classification Accuracy**: Random forest classification of land use and cover in the Shiyang River Basin from 1991–2020 demonstrated high ac-

curacy, with overall accuracy exceeding 91.10% and Kappa coefficients greater than 0.84 for all years. Landsat_{RFC} outperformed the three comparative datasets (AGLC-2000-2015, GLC_{FCS30}, and CLCD) in both overall and partial validations.

2. **Dominant Land Types**: Bare land, grassland, and cropland were the main land use and cover types, accounting for over 90.00% of the basin area. The upper reaches were dominated by grassland, forest, and bare land, while the middle and lower reaches were dominated by bare land, grassland, and cropland. The lower reaches featured an exceptionally high bare land proportion (84.42% in 2020).

3. **Temporal Trends**: Bare land area decreased overall, with changes concentrated in the middle and lower reaches. Grassland area initially decreased then increased, while cropland area initially increased then decreased. Grassland changes occurred mainly in the upper and middle reaches, whereas cropland changes were in the middle and lower reaches. Forest and impervious surface areas increased primarily in the upper and middle reaches, respectively. Wetland and glacier areas decreased in the lower and upper reaches, respectively. Water body area initially decreased then increased across all sub-regions.

4. **Driving Factors**: LUCC was influenced by both natural and anthropogenic factors. Precipitation, evapotranspiration, temperature, and population were primary drivers in the upper and middle reaches, while surface soil moisture, population, precipitation, and temperature dominated in the lower reaches. Policy factors also significantly influenced LUCC in the Shiyang River Basin.

These findings deepen understanding of LUCC trends and drivers in the Shiyang River Basin, providing scientific significance for regional sustainable development and a foundation for related research in similar arid and semi-arid regions.

Conflict of Interest

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

Acknowledgements

This work was financially supported by the Central Government to Guide Local Technological Development (23ZYQH0298), the Science and Technology Project of Gansu Province (20JR10RA656, 22JR5RA416), and the Science and Technology Project of Wuwei City (WW2202YFS006).

Author Contributions

Conceptualization: ZHAO Yaxuan, CAO Bo, PAN Baotian
Data curation: ZHAO Yaxuan, CHENG Jinquan
Formal analysis: ZHAO Yaxuan
Funding acquisition: CAO Bo, GUAN Weijin
Investigation: ZHAO Yaxuan, SHA Linwei
Methodology: ZHAO Yaxuan, SHA Linwei, ZHAO Xuanru
Supervision: CAO Bo, PAN Baotian
Validation: ZHAO Yaxuan, GUAN Weijin
Visualization: ZHAO Xuanru, GUAN Weijin
Writing – original draft: ZHAO Yaxuan
Writing – review & editing: CAO Bo, CHENG Jinquan

All authors approved the manuscript.

References

Alhamid A K, Akiyama M, Aoki K, et al. 2022. Stochastic renewal process model of time-variant tsunami hazard assessment under nonstationary effects of sea-level rise due to climate change. *Structural Safety*, 99: 102263, doi: 10.1016/j.strusafe.

Anderson J R, Hardy E E, Roach J T, et al. 1976. A land use and land cover classification system for use with remote sensor data. *Geological Survey Professional Paper*, 964, doi: 10.3133/pp964.

Bai B, Yue P, Zhang Q, et al. 2023. Changing characteristics of ecosystem and water storage under the background of warming and humidification in the Qilian Mountains, China. *Science of the Total Environment*, 893: 164959, doi: 10.1016/j.scitotenv.

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

Breiman L. 2001. Random forests. *Machine Learning*, 45: 5–32.

Cao B, Pan B T, Wen Z L, et al. 2019. Changes in glacier mass in the Lenglongling Mountains from 1972 to 2016 based on remote sensing data and modeling. *Journal of Hydrology*, 578: 124010, doi: 10.1016/j.jhydrol.2019.124010.

Chen J D, Gao M, Cheng S L, et al. 2022. Global 1 km$×$1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. *Scientific Data*, 9: 202, doi: 10.1038/s41597-022-01346-3.

Chen Y, Yang G J, Zhou L H, et al. 2020. Quantitative analysis of natural and human factors of oasis change in the tail of Shiyang River over the past 60 years. *Acta Geological Sinica (English Edition)*, 94(3): 637–645.

Cui F Q, Wang B J, Zhang Q, et al. 2021. Climate change versus land-use change—What affects the ecosystem services more in the forest-steppe ecotone? *Science of the Total Environment*, 759: 143525, doi: 10.1016/j.scitotenv.2020.143525.

Center for International Earth Science Information Network, Columbia University. 2018. Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). [2023-05-06]. https://doi.org/10.7927/H45Q4T5F.

Du X D, Jin X B, Yang X L, et al. 2014. Spatial pattern of land use change and its driving force in Jiangsu Province. *International Journal Environmental Research and Public Health*, 11(3): 3215–3232.

Ebabu K, Taye G, Tsunekawa A, et al. 2023. Land use, management and climate effects on runoff and soil loss responses in the highlands of Ethiopia. *Journal of Environmental Management*, 326: 116707, doi: 10.1016/j.jenvman.2022.116707.

Gao H S, Li Z M, Pan B T, et al. 2016. Fluvial responses to late Quaternary climate change in the Shiyang River drainage system, western China. *Geomorphology*, 258: 82–94.

Gong P, Wang J, Yu L, et al. 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.

Gule T T, Lemma B, Hailu B T. 2023. Implications of land use/land cover dynamics on urban water quality: Case of Addis Ababa City, Ethiopia. *Heliyon*, 9(5): e15665, doi: 10.1016/j.heliyon.2023.e15665.

Guo B, Liu Y F, Fan J F, et al. 2023a. The salinization process and its response to the combined processes of climate change–human activity in the Yellow River Delta between 1984 and 2022. *Catena*, 231: 107301, doi: 10.1016/j.catena.2023.107301.

Guo B, Lu M, Fan Y W, et al. 2023b. A novel remote sensing monitoring index of salinization based on three-dimensional feature space model and its application in the Yellow River Delta. *Geomatics, Natural Hazards and Risk*, 14(1): 95–116.

Hamed Fahmy A, Amin Abdelfatah M, El-Fiky G. 2023. Investigating land use land cover changes and their effects on land surface temperature and urban heat islands in Sharqiyah Governorate, Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, 26(2): 293–306.

Han T, Xie Y W, Jiang Y Y. 2015. The land use change characteristics and its driving force analysis of Shiyang river basin in northwest China. *Proceedings Volume 9815, Multispectral Image Processing and Pattern Recognition 2015 (MIPPR 2015): Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*. 98150N (14 December 2015). https://doi.org/10.1117/12.2205775.

Hao Y Y, Xie Y W, Ma J H, et al. 2017. The critical role of local policy effects in arid watershed groundwater resources sustainability: A case study in the Minqin oasis, China. *Science of the Total Environment*, 601–602: 1084–1096.

Hatfield J L, Prueger J H. 2015. Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10: 4–10.

Hu X Q, Jin Y Z, Ji L H, et al. 2018. Land use/cover change and ITS eco-environment effect in Shiyang River Basin. *IOP Conference Series: Earth Environment Science*. Volume 191. The 4th International Conference on Water Resource and Environment (WRE 2018). 17–21 July, 2018. Kaohsiung City, Taiwan, China.

Ibarrola-Rivas M J, Unar-Munguia M, Kastner T, et al. 2022. Does Mexico have the agricultural land resources to feed its population with a healthy and sustainable diet? *Sustainable Production and Consumption*, 34: 371–384.

IPCC (Intergovernmental Panel on Climate Change). 2021. *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, USA: Cambridge University Press.

Jia J, Su X, Zhang J, et al. 2023. Spatial-temporal characteristics of land use changes in landslide-damaged areas of the Loess Plateau in Tianshui City from 1985 to 2020. *Journal of Soil and Water Conservation*, 37(4): 195–204. (in Chinese)

Kang S C, Zhang Q G, Zhang Y L, et al. 2022. Warming and thawing in the Mt. Everest region: A review of climate and environmental changes. *Earth-Science Reviews*, 225: 103911, doi: 10.1016/j.earscirev.2021.103911.

Li C, Zhong Y, Zhang W. 2017. Study on spatial-temporal variation of Hong Lake based on VIIRS and MODIS images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W7: 797–802.

Liao Z M, Wang X, Zhang Y L, et al. 2024. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China. *Ecological Indicators*, 158: 111584, doi: 10.1016/j.ecolind.2024.111584.

Liu J Y. 1996. *Remote Sensing Macraoscopic Survey and Dynamic Study of China's Resources and Environment*. Beijing: China Science and Technology Press, 4–40. (in Chinese)

Liu X Q. 2021. Research on the ecological construction and health assessment of Shiyang River. *Shaanxi Water Resources*, (4): 7–9. (in Chinese)

Liu Y Y, Chen G J, Meyer-Jacob C, et al. 2021. Land-use and climate controls on aquatic carbon cycling and phototrophs in karst lakes of southwest China. *Science of the Total Environment*, 751: 141738, doi: 10.1016/j.scitotenv.2020.141738.

Lu Z, Wang D Q, Deng Z D, et al. 2021. Application of red edge band in remote sensing extraction of surface water body: a case study based on GF-6 WFV data in arid area. *Hydrology Research*, 52(6): 1526–1541.

Ma Z M, Kang S Z, Zhang L, et al. 2008. Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. *Journal of Hydrology*, 352(3–4): 239–249.

Meng Y R, Yang M X, Liu S, et al. 2021. Quantitative assessment of the importance of bio-physical drivers of land cover change based on a random forest method. *Ecological Informatics*, 61: 101204, doi: 10.1016/j.ecoinf.2020.101204.

Mi X T. 2021. Evaluation of river health in the upper reaches of Shiyang River. MSc Thesis. Lanzhou: Lanzhou University. (in Chinese)

Pal M, Mather P M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4): 554–565.

Resende Vieira F, Christofaro C. 2024. Contributions of the vegetation index (NDVI) in water quality prediction models in a semi-arid tropical watershed. *Journal of Arid Environments*, 220: 105122, doi: 10.1016/j.jaridenv.2024.105122.

Rodriguez-Galiano V F, Ghimire B, Rogan J, et al. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67: 93–104.

Serra P, Pons X, Saurí D. 2008. Land-cover and land-use change in a Mediterranean landscape: A spatial analysis of driving forces integrating biophysical and human factors. *Applied Geography*, 28(3): 189–209.

Sun Q. 2021. Land use change and multi-scenario simulation in Luochuan County based on Markov-Flus model. MSc Thesis. Xi'an: Chang'an University. (in Chinese)

Tong L, Kang S Z, Zhang L. 2007. Temporal and spatial variations of evapotranspiration for spring wheat in the Shiyang river basin in northwest China. *Agricultural Water Management*, 87(3): 241–250.

Viedma O, Moreno J M, Güngöroglu C, et al. 2017. Recent land-use and land-cover changes and its driving factors in a fire-prone area of southwestern Turkey. *Journal of Environmental Management*, 197: 719–731.

Wang H J, Zhang B, Xia C, et al. 2020. Using a maximum entropy model to optimize the stochastic component of urban cellular automata models. *International Journal of Geographical Information Science*, 34(5): 924–946.

Wang H Q, Zhang M S, Zhu H, et al. 2012. Hydro-climatic trends in the last 50years in the lower reach of the Shiyang River Basin, NW China. *Catena*, 95: 33–41.

Wang Q, Wang H J, Chang R H, et al. 2022. Dynamic simulation patterns and spatiotemporal analysis of land-use/land-cover changes in the Wuhan metropolitan area, China. *Ecological Modelling*, 464: 109850, doi: 10.1016/j.ecolmodel.2021.109850.

Wang Q Z, Guan Q Y, Lin J K, et al. 2021. Simulating land use/land cover change in an arid region with the coupling models. *Ecological Indicators*, 122: 107231, doi: 10.1016/j.ecolind.2020.107231.

Wang R M, Ding J L, Ge X Y, et al. 2023. Impacts of climate change on the wetlands in the arid region of Northwestern China over the past 2 decades. *Ecological Indicators*, 149: 110168, doi: 10.1016/j.ecolind.2023.110168.

Wang Y C, Zhao J, Fu J W, et al. 2019. Effects of the Grain for Green Program on the water ecosystem services in an arid area of China—Using the Shiyang River Basin as an example. *Ecological Indicators*, 104: 659–668.

Wei S T, Luo M, Zhu L F, et al. 2023. Using object-oriented coupled deep learning approach for typical object inspection of transmission channel. *International Journal of Applied Earth Observation and Geoinformation*, 116: 103137, doi: 10.1016/j.jag.2022.103137.

Wei W, Xie Y W, Shi P J, et al. 2017. Spatial temporal analysis of land use change in the Shiyang River Basin in arid China, 1986–2015. *Polish Journal of Environmental Studies*, 26(4): 1789–1796.

Wondie M, Schneider W, Katzensteiner K, et al. 2016. Modelling the dynamics of landscape transformations and population growth in the highlands of Ethiopia using remote-sensing data. *International Journal of Remote Sensing*, 37(23): 5647–5667.

Wu H, Lin A Q, Xing X D, et al. 2021a. Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. *International Journal of Applied Earth Observation and Geoinformation*, 103: 102475, doi: 10.1016/j.jag.2021.102475.

Wu L Z, Ma X F, Dou X, et al. 2021b. Impacts of climate change on vegetation phenology and net primary productivity in arid Central Asia. *Science of the Total Environment*, 796: 149055, doi: 10.1016/j.scitotenv.2021.149055.

Wu R, Wang J Y, Zhang D C, et al. 2021c. Identifying different types of urban land use dynamics using Point-of-interest (POI) and Random

Forest algorithm: The case of Huizhou, China. *Cities*, 114: 103202, doi: 10.1016/j.cities.2021.103202.

Wu Z G, Zhu D Y, Xiong K N, et al. 2022. Dynamics of landscape ecological quality based on benefit evaluation coupled with the rocky desertification control in South China Karst. *Ecological Indicators*, 138: 108870, doi: 10.1016/j.ecolind.2022.108870.

Xian X Q, Zhao H X, Guo J Y, et al. 2023. Estimation of the potential geographical distribution of a new potato pest (*Schrankia costaestrigalis*) in China under climate change. *Journal of Integrative Agriculture*, 22(8): 2441–2455.

Xie H L, He Y F, Xie X. 2017. Exploring the factors influencing ecological land change for China's Beijing–Tianjin–Hebei Region using big data. *Journal of Cleaner Production*, 142: 677–687.

Xie X F, Wu T, Zhu M, et al. 2021. Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecological Indicators*, 120: 106925, doi: 10.1016/j.ecolind.2020.106925.

Xu X C, Li B J, Liu L P, et al. 2021. The global 2000–2015, 30 m resolution land cover mapping year by year. *National Remote Sensing Bulletin*, 25(09): 1896–1916. (in Chinese)

Xu L, Gao G Y, Wang X F, et al. 2023. Distinguishing the effects of climate change and vegetation greening on soil moisture variability along aridity gradient in the drylands of northern China. *Agricultural and Forest Meteorology*, 343: 109786, doi: 10.1016/j.agrformet.2023.109786.

Xue D X, Zhou J J, Zhao X, et al. 2021. Impacts of climate change and human activities on runoff change in a typical arid watershed, NW China. *Ecological Indicators*, 121: 107013, doi: 10.1016/j.ecolind.2020.107013.

Yang C, Wu G F, Ding K, et al. 2017. Improving land use/land cover classification by integrating pixel unmixing and decision tree methods. *Remote Sensing*, 9(12): 1222, doi: 10.3390/rs9121222.

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

Yang Y W, Tian Y C, Zhang Q, et al. 2022b. Impact of current and future land use change on biodiversity in Nanliu River Basin, Beibu Gulf of South China. *Ecological Indicators*, 141: 109093, doi: 10.1016/j.ecolind.2022.109093.

Yu Y, Guo B, Wang C L, et al. 2022a. Carbon storage simulation and analysis in Beijing-Tianjin-Hebei region based on CA-plus model under dual-carbon background. *Geomatics, Natural Hazards and Risk*, 14(1): 2173661, 10.1080/19475705.2023.2173661.

Yu Z P, Yan T H, Liu X R, et al. 2022b. Urban land expansion, fiscal decentralization and haze pollution: Evidence from 281 prefecture-level cities in China. *Journal of Environmental Management*, 323: 116198, doi: 10.1016/j.jenvman.2022.116198.

Zhang C X, Wang X M, Li J C, et al. 2021a. The impact of climate change on aeolian desertification in northern China: Assessment using aridity index. *Catena*, 207: 105681, 10.1016/j.catena.2021.105681.

Zhang D C, Liu X P, Wu X Y, et al. 2019. Multiple intra-urban land use simulations and driving factors analysis: A case study in Huicheng, China. *GIScience & Remote Sensing*, 56(2): 282–308.

Zhang J G, Lan Z L, Li H W, et al. 2023a. Coupling effects of soil organic carbon and moisture under different land use types, seasons and slope positions in the Loess Plateau. *Catena*, 233: 107520, doi: 10.1016/j.catena.2023.107520.

Zhang M M, Tan S K, Zhang C, et al. 2023b. Assessing the impact of fractional vegetation cover on urban thermal environment: A case study of Hangzhou, China. *Sustainable Cities and Society*, 96: 104663, doi: doi.org/10.1016/j.scs.2023.104663.

Zhang X, Liu L Y, Chen X D, et al. 2021b. 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.

Zhao C H, Zhang H B, Wang H Y, et al. 2023. Analysis of changes in the spatiotemporal characteristics of impervious surfaces and their influencing factors in the Central Plains Urban Agglomeration of China from 2000 to 2018. *Heliyon*, 9(8): e18849, doi: 10.1016/j.heliyon.2023.18849.

Zhao R F, Chen Y N, Shi P J, et al. 2013. Land use and land cover change and driving mechanism in the arid inland river basin: A case study of Tarim River, Xinjiang, China. *Environment Earth Sciences*, 68: 591–604.

Zhou Y, Li X H, Liu Y S. 2020. Land use change and driving factors in rural China during the period 1995–2015. *Land Use Policy*, 99: 105048, 10.1016/j.landusepol.2020.105048.

Zhu G F, Guo H W, Qin D H, et al. 2019. Contribution of recycled moisture to precipitation in the monsoon marginal zone: Estimate based on stable isotope data. *Journal of Hydrology*, 569: 423–435.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.