Spatiotemporal Evolution and Prediction of Land Use in Mining Areas of Semi-Arid Regions Postprint

Authors: Liu Chang

Date: 2022-01-26T18:27:39+00:00

Abstract

The land use patterns in mining areas of semi-arid regions are undergoing tremendous changes under mining disturbances. Taking the Datong mining area in Shanxi Province, one of the eight major coal production bases in China, as the research object, this study analyzes the spatiotemporal changes of land use types from 1985 to 2015 and the driving factors influencing land use changes, and constructs an RF (Random Forest, RF)-FLUS (Future Land Use Simulation, FLUS) model to simulate and predict future land use changes in mining areas of semi-arid regions. The results show that: (1) From 1985 to 2015, the area of forestland, cropland, and water bodies in the mining area decreased, while the area of grassland and construction land increased. (2) The distribution of forestland and grassland is significantly influenced by climate and distances to water systems and facility points; the distribution of cropland is significantly influenced by climate, elevation, and distances to water bodies and residential points; the most important influencing factor for water body distribution is precipitation; the distribution of construction land is mainly influenced by production capacity and distance to facility points. (3) Both the FLUS model and the RF-FLUS model exhibit high fitting accuracy, but the RF-FLUS model demonstrates higher accuracy than the FLUS model and yields results closer to actual land pattern changes. (4) According to the RF-FLUS model predictions for land use changes in the mining area in 2025, forestland, grassland, and cropland within the mining area all show a declining trend, with little change in the rate of decline; water bodies remain unchanged, while construction land and other types (bare land and unused land) maintain a stable upward trend. This study provides a favorable scientific basis for exploring the complex dynamic evolution mechanisms of land patterns in mining areas, exploring optimization paths for small-scale land resources, and promoting healthy regional ecological development.

Full Text

Spatio-temporal Evolution and Prediction of Land Use in Mining Areas of Semi-arid Regions

LIU Chang[1], **ZHANG Hong**[2,3], **ZHANG Xiaoyu**[2,3], **YANG Guoting**[2], **LIU Yong**[1,3]

[1]Institute of Loess Plateau, Shanxi University, Taiyuan 030006, Shanxi, China
[2]College of Environment and Resource Sciences, Shanxi University, Taiyuan 030006, Shanxi, China
[3]Shanxi Laboratory for Yellow River, Taiyuan 030006, Shanxi, China

Abstract

Land use and land cover change (LUCC) plays a critical role in regional land planning and ecological environmental protection. In mining areas, LUCC induced by human activities such as coal mining has intensified human-environment conflicts. This study examines the spatio-temporal patterns of LUCC and their driving factors in the Datong mining area of Shanxi Province —a major coal production base in China—from 1985 to 2015. Furthermore, we developed a Random Forest-Future Land Use Simulation (RF-FLUS) model to predict future land use patterns in this semi-arid mining region. The results reveal that: (1) From 1985 to 2015, forestland, cropland, and water bodies decreased, while grassland and construction land increased. (2) Climate conditions, elevation, and proximity to water systems and facilities significantly influenced the distribution of forestland, grassland, and cropland; precipitation was the primary factor affecting water body distribution; and coal production capacity along with distance to facilities were the main determinants of construction land distribution. (3) Both FLUS and RF-FLUS models demonstrated high simulation accuracy, with the RF-FLUS model achieving superior Kappa and Overall Accuracy (OA) indices. (4) Predictions for 2025 suggest continued decline in forestland, grassland, and cropland, stable water areas, and continued expansion of construction land and other land types. This study provides scientific insights for understanding the complex dynamic mechanisms of land use evolution in mining areas, optimizing land resource management at local scales, and promoting sustainable regional ecological development.

Keywords: coal mining area; land use change; model prediction; random forest model; driving factors

Introduction

Rapid urbanization has led to significant land cover and land use changes (LUCC) [**?**], which can cause ecological environmental degradation and gradually destabilize regional ecosystems [**?**]. Research on LUCC has evolved from phenomenological description to mechanistic analysis, and from qualitative to

quantitative simulation [?]. Land use change models can analyze the causes of LUCC, explore landscape evolution patterns, and predict future land use demands and distributions [?], making the study of LUCC dynamics and improvement of prediction capabilities a current research focus.

Land use change models primarily include: (1) Empirical statistical models such as multiple regression and geographically weighted regression [?]; (2) Cellular Automata (CA) models like SLEUTH [?]; (3) Machine learning models such as neural networks and genetic algorithms [?]; and (4) Multi-agent system models [?]. Empirical statistical and machine learning models effectively integrate remote sensing data to establish relationships between land changes and spatio-temporal factors based on historical patterns. However, these models often assume constant driving forces, making them unsuitable for long-term predictions and prone to overfitting [?]. CA models, based on discrete spatial units, better handle spatial information and are suitable for simulating complex geographic patterns [?], though early studies using logistic regression for transition rules offered computational convenience but insufficient accuracy [?]. Yang et al. [?] proposed using artificial intelligence and machine learning algorithms as transition rules, improving accuracy but without weight output capabilities.

As researchers and decision-makers demand higher precision for land classification, future pattern simulation, and suitability probability assessment, the Random Forest model has gained widespread application in land use research due to its error-balancing capabilities, moderate complexity, and ability to analyze driving factor importance [?]. For instance, Zhang et al. [?] combined Random Forest with CA models to simulate land use changes in Dongguan while analyzing factor importance for different land types. Ma et al. [?] applied Random Forest classification to land use mapping in complex terrain regions of Qinghai Province. Chen et al. [?] leveraged Random Forest advantages with CA models to study complex nonlinear urban spatial evolution through importance discrimination of influencing factors.

Land use change is a geographic process influenced by both natural conditions and socioeconomic factors. In mining areas, factors vary significantly due to differences in location, topography, mining methods, scale, and socioeconomic levels [?]. The Datong mining area lies in a semi-arid region where limited natural carrying capacity, combined with the persistent temporal, spatial, and intensive disturbance characteristics of coal resource extraction, not only occupies and destroys substantial land but also exacerbates the fragile ecological environment. For example, Duan et al. [?] found that underground mining altered plant community structure and vegetation coverage in arid desert regions. Shi et al. [?] documented ecological problems such as groundwater decline and surface subsidence caused by intensive coal mining. Zou et al. [?] demonstrated that mining intensified soil erosion in the Weibei mining area of the Loess Plateau.

To reveal land use change patterns under natural conditions and human impacts in mining areas, this study introduces Random Forest methods to improve model accuracy and more accurately predict future land use patterns.

This approach has theoretical significance for understanding the complex dynamic evolution mechanisms of land use patterns and practical importance for exploring land resource management optimization and implementing ecological regulation strategies in mining areas.

1.1 Study Area Overview

The Datong mining area is located in the southwestern part of Datong City, Shanxi Province, between 39°43 -40°10 N and 112°31 -113°14 E. The terrain consists of gentle hills with a monoclinal structure, with elevations ranging from 1031 to 1964 m. The Shili River and "U"-shaped valleys develop throughout the region. Situated in a fragile semi-arid zone, the mining area experiences large diurnal temperature variations, concentrated summer precipitation, dry climate throughout the year, and annual evaporation far exceeding rainfall, resulting in limited water area.

The main coal-bearing strata are the Jurassic and Carboniferous-Permian systems. Jurassic coal seams have been almost fully exploited except for the southwestern corner. Carboniferous-Permian mined-out areas are mainly distributed in the eastern and southern parts. With low harmful element content, high tar yield, and high calorific value, Datong coal is an important high-quality thermal coal production base in China, primarily using underground mining methods. The study area location is shown in Figure 1.

1.2 Data Sources and Processing

Data types include land use data and influencing factor data. Influencing factors comprise natural factors, socioeconomic factors, and distance factors (Table 1). Land use data were interpreted from Landsat time-series remote sensing imagery and reclassified into six categories according to national standards. DEM data were obtained from the Chinese Academy of Sciences Resource and Environmental Science Data Center (http://www.resdc.cn), from which slope and aspect were extracted. Temperature and precipitation data were obtained from the China Meteorological Data Network (http://data.cma.cn) and spatially interpolated from station data. Coal production capacity was obtained through field surveys. Population data were refined to 30m resolution based on reference [**?**]. Distance factors were calculated using the Euclidean distance algorithm in ArcGIS 10.5, with point-of-interest data obtained through web crawler technology.

1.3 Research Methods

1.3.1 FLUS Model

The FLUS model couples a System Dynamics Model with a Cellular Automata model, introducing an Artificial Neural Network (ANN) to construct transition rules based on the original CA framework. This approach offers advantages

in simulating competition and conversion among various land use types while integrating human activities and natural effects [**?**].

The model consists of two components:

First, the ANN module trains sample data to construct spatial occurrence probabilities for regional land use types. ANN advantages include iterative learning and fitting of complex relationships between input data and training targets [**?**]. In the input layer, neurons correspond to input variables; output layer neurons represent occurrence probabilities of specific land types within grid cells, where higher values indicate greater likelihood of target land type occurrence [**?**]. The probability of land type k occurring in grid cell d at time i is calculated as:

$$p_{i,k}^{d} = \frac{1}{1 + e^{-\mathrm{net}_{j,k}^{d}}}$$

where $\mathrm{net}_{j,k}^{d} = \sum w_{j,k} \cdot Sigmoid(\mathrm{net}_j)$ represents the activation connection between input and hidden layers; $w_{j,k}$ denotes adaptive weights between output and hidden layers; and $Sigmoid(\mathrm{net}_j)$ is the activation function.

Second, an adaptive inertia competition mechanism based on roulette wheel selection integrates CA spatial operations to adjust discrepancies between macro-level land demand and current land quantities to achieve target values, addressing limitations in handling complex interactions among land types [**?**]. The comprehensive suitability probability is calculated as:

$$P_{i,k}^{d} = p_{i,k}^{d} \times \mathrm{Inertia}_{k}^{i} \times \Omega_{i,k}^{d}$$

where $\mathrm{Inertia}_{k}^{i}$ represents the adaptive inertia coefficient for land type k at time i, determined by the difference between demand and current quantity; and $\Omega_{i,k}^{d}$ denotes neighborhood effects. Transition costs derived from historical land use data and empirical knowledge are represented as $sc_{m,k}$.

1.3.2 Random Forest Model

Random Forest is an ensemble intelligent algorithm generating multiple decision trees with advantages in data mining, transition rule accuracy, and stability. It overcomes ANN overfitting issues by training on bootstrap samples [**?**]. For each of n decision trees with m variables, 2/3 of samples serve as in-bag data for model training while 1/3 serves as out-of-bag (OOB) data for error estimation and validation, with smaller OOB error indicating higher accuracy. This sampling approach reduces correlation among decision trees, providing a foundation for more accurate simulation and prediction of land use changes in the Datong mining area.

1.3.3 Mean Gini Decrease Method

The mean Gini decrease method was employed to calculate and analyze the importance of different factors influencing land use patterns in the Datong mining area. This method ranks decision trees by Gini Impurity, where feature importance directly reflects participation significance in model calculations. Each feature' s importance is a floating-point number between 0 and 1, with larger values indicating greater influence [**?**]. By traversing all tree nodes and summing Gini coefficient decreases for each selected feature variable, the method quantifies each variable' s impact on the final pattern.

1.3.4 Model Accuracy Validation

Model accuracy was validated using confusion matrices and Kappa coefficient tests [**?**] for quantitative analysis. Higher diagonal percentages in confusion matrices indicate greater accuracy.

2 Results

2.1 Spatio-temporal Characteristics of Land Use Patterns

From 1985 to 2015, the Datong mining area experienced dramatic land use transformations (Figure 2). Forestland, cropland, and water bodies decreased, while grassland and construction land increased. Specifically, during 1985–1995 when the mining area entered a mature stage with mechanized production and expanded mining scale, cropland decreased by approximately 3.09% and was mainly scattered at the mining area' s edge. Concurrently, the implementation of the Grain-for-Green policy made grassland the dominant land use type, accounting for 55.93% of total area. Water bodies decreased from 1.04% to 0.35%.

During 1995–2005, comprehensive production transformation and scaled mining operations accelerated urban industrialization, characterized by substantial construction land increase from 7.09% to 12.7%. Cropland, forestland, grassland, and water areas all decreased due to human activities, while other land types increased slightly by 0.29%.

From 2005–2015, the mining area entered a transition period with government policy support. Land use changes showed various types converting to grassland, with cropland-to-grassland transfer reaching 72.58% and forestland-to-grassland conversion exceeding 40.75 km^2. Water bodies predominantly transferred to other types (94.35%), with minimal conversion to grassland. Other land use types showed minimal changes.

2.2 Land Use Transfer Analysis Across Historical Periods

Land use transfer matrices (Table 2) reveal distinct patterns across periods. During 1985–1995, mining activities and human livelihoods caused large-scale

water body conversion to forestland, grassland, and cropland, with some construction land transferring to forestland. From 1995–2005, large-scale mining expansion, deforestation for mining, and road construction for coal transportation extensively occupied cropland and forestland, with primary transfers to grassland. During 2005–2015, water bodies predominantly transferred to other types (94.35%), with minor cropland and water area conversions to grassland.

Overall, the trend from 1985–2015 shows various land use types converting to grassland, with forestland-to-grassland transfer being the most significant.

2.3 Analysis of Land Use Change Drivers

Analysis of variable importance reveals distinct driving factors for different land use types (Figure 4). Forestland and grassland distributions were primarily influenced by climate conditions and distances to water systems and facilities. Cropland distribution was affected by elevation, climate, and distances to water bodies, facilities, and residential points—primarily because arid region agriculture requires convenient irrigation, and farmers prefer locations near settlements. Precipitation was the most critical factor for water body distribution. Construction land distribution was mainly determined by coal production capacity and distance to facilities; higher production capacity increases various infrastructure including factories, roads, and worker accommodations, while proximity to service facilities (restaurants, schools, hospitals, entertainment venues) directly determines construction land distribution. Other land types were primarily influenced by distance to water systems, as areas near water are easily transformed.

2.4 Land Use Pattern Simulation and Future Prediction

The RF-FLUS model was applied to simulate land use changes in the Datong mining area. Using 2005 as the baseline year, the model simulated 2015 land use patterns, showing high consistency with actual distributions (Figure 5). Confusion matrix validation (Table 3) yielded Kappa coefficients of 0.892 for RF-FLUS and 0.851 for FLUS, with overall accuracy (OA) of 0.921 and 0.896 respectively, demonstrating that the RF-FLUS model achieved higher accuracy and better reflected actual land use patterns.

Based on the 2015 baseline, the RF-FLUS model predicted 2025 land use distribution (Figure 6). Results indicate that forestland, grassland, and cropland will continue declining, with decreases of 14.97 km^2, 130.75 km^2, and 157 km^2 respectively. Water areas will remain stable, while construction land and other land types will increase steadily, with construction land expanding from 12.7% to 15.71% (increase of 26.26 km^2) and other types increasing by 44.12 km^2.

3 Conclusions

This study examined spatio-temporal land use pattern changes across different stages in the Datong mining area, compared simulation accuracy advantages of different land use models, analyzed primary driving factors, and conducted predictive analysis, yielding the following conclusions:

(1) The Jurassic coal seams in the Datong mining area have been almost fully exploited, with Carboniferous-Permian mining concentrated in the northern and southern regions. Under mining influence, land use patterns underwent dramatic transformation from 1985–2015, characterized by increases in grassland, construction land, and other land types, while forestland and cropland initially increased then decreased, with water bodies being largely encroached upon. Land use change in mining areas represents a dynamic spatio-temporal evolution process driven by resource extraction.

(2) Different land use types in the mining area have distinct driving factors. Forestland and grassland are primarily influenced by climate and distances to water systems and facilities; cropland distribution is affected by multiple factors including elevation, climate, and distances to water bodies, facilities, and settlements; precipitation is the most important factor for water bodies; construction land is driven by production capacity and distance to facilities; and other land types are mainly influenced by distance to water systems. Mineral resource development constitutes the primary driving force of land use change, while natural factors such as climate and topography also exert significant influence.

(3) The RF-FLUS model, by integrating Random Forest algorithms, overcomes overfitting issues in neural networks and leverages Random Forest' s high accuracy advantages, providing more accurate predictions of future land use patterns. This approach offers important reference value for simulating land use changes in small-scale mining areas and understanding their complex dynamic evolution mechanisms.

References

[?] Li Shengpeng, Liu Jianling, Lin Jin, et al. Spatial and temporal evolution of habitat quality in Fujian Province, China based on the land use change from 1980 to 2018[?]. Chinese Journal of Applied Ecology, 2020, 31(12): 4080-4090.

[?] Liu Jiyuan, Zhang Zengxiang, Zhuang Dafang, et al. A study on the spatial temporal dynamic changes of land use and driving forces analyses of China in the 1990s[?]. Geographical Research, 2003, 22(1): 1-12.

[?] Veldkamp A, Lambin E F. Predicting land use change[?]. Agriculture, Ecosystems and Environment, 2001, 85(1): 1-6.

[?] Li Baojie, Gu Hehe, Ji Yazhou. Simulation of land use change in coal mining area under different scenarios based on the CLUE-S model: A case study of

jiawang mining area in Xuzhou city[?]. Tropical Geography, 2018, 38(2): 274-281.

[?] Liu Yanwen, Liu Chengwu, He Zongyi, et al. Spatio-temporal evolution of ecological land and influence factors in Wuhan urban agglomeration based on geographically weighted regression model[?]. Chinese Journal of Applied Ecology, 2020, 31(3): 987-998.

[?] Sun Yizhong, Yang Jing, Song Shuying, et al. Modeling of multilevel vector cellular automata and its simulation of land use change[?]. Acta Geographica Sinica, 2020, 75(10): 2164-2179.

[?] Xu Hongtao, Chen Chunbo, Zheng Hongwei, et al. Correlation analysis and adaptive genetic algorithm based feature subset and model parameter optimization in salinization monitoring[?]. Journal of Geo-information Science, 2020, 22(7): 1497-1509.

[?] Wang Ziyang, Shi Peiji, Zhang Xuebin, et al. Simulation of Lanzhou urban land expansion based on multi-agent model[?]. Chinese Journal of Applied Ecology, 2021, 32(6): 2169-2179.

[?] Dai Erfu, Ma Liang. Review on land change modeling approaches[?]. Progress in Geography, 2018, 37(1): 152-162.

[?] Zhou Chenghu, Ou Yang, Ma Ting, et al. Theoretical perspectives of CA based geographical system modeling[?]. Progress in Geography, 2009, 28(6): 833-838.

[?] Li Xia, Liu Xiaoping. Case-based cellular automaton for simulating urban development in a large complex region[?]. Acta Geographica Sinica, 2007, 75(10): 1097-1109.

[?] Yang Qingsheng, Li Xia. Calibrating urban cellular automata using genetic algorithms[?]. Geographical Research, 2007, 26(2): 229-237.

[?] Zhang Dachuan, Liu Xiaoping, Yao Yao, et al. Simulating spatio-temporal change of multiple land use types in Dongguan by using random forest based on cellular automata[?]. Geography and Geo-Information Science, 2016, 32(5): 29-36.

[?] Ma Huijuan, Gao Xiaohong, Gu Xiaotian. Random forest classification of Landsat 8 imagery for the complex terrain area based on the combination of spectral, topographic and texture information[?]. Journal of Geo-Information Science, 2019, 21(3): 359-371.

[?] Chen Kai, Liu Kai, Liu Lin, et al. Urban expansion simulation by random forest based cellular automata: A case study of Foshan City[?]. Progress in Geography, 2015, 34(8): 937-946.

[?] Bian Zhengfu, Zhang Yanping. Land use changes in Xuzhou coal mining area[?]. Acta Geographica Sinica, 2006, 61(4): 349-358.

[?] Duan Yufeng, Zhang Yuxiu, Yu Chuang. Effects of the underground coal mining on the dynamic changes of vegetation in arid desert area[?]. Acta Ecologica Sinica, 2020, 40(23): 8717-8728.

[?] Shi Xiaoqiong, Yang Zeyuan, Zhang Yanna, et al. Reviews of influence by high-intensity coal mining on ecological environment in Northern Shanxi[?]. Coal Technology, 2016, 35(1): 314-316.

[?] Zou Yajing, Yan Qingwu, Tan Xueling, et al. Evaluation of soil erosion and driving factors analysis in Weibei mining area[?]. Arid Land Geography, 2019, 42(6): 1387-1394.

[?] Ye T T, Zhao N Z, Yang X C, et al. Improved population mapping for China using remotely sensed and points of interest data within a random forests model[?]. Science of the Total Environment, 2019, 658: 936-946.

[?] Liu X P, Liang X, Li X, et al. A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects[?]. Landscape and Urban Planning, 2017, 168: 94-116.

[?] Li X, Anthony Gar-On Yeh. Neural network based cellular automata for simulating multiple land use changes using GIS[?]. International Journal of Geographical Information Science, 2002, 16(4): 323-343.

[?] Wu F L. Calibration of stochastic cellular automata: The application to rural-urban land conversions[?]. International Journal of Geographical Information Science, 2002, 16(8): 795-818.

[?] Hadi M, Siva K B, Jamal B T, et al. Validation of CA-Markov for simulation of land use and cover change in the Langat Basin, Malaysia[?]. Journal of Geographic Information System, 2012, 4(6): 548-556.

[?] Sun Hongchao, Zhang Zhengxiang. Changes of landscape pattern vulnerability of Songhua River Basin in Jilin Province and its driving forces[?]. Arid Zone Research, 2019, 36(4): 1005-1041.

[?] Qin Qirui, Li Xuemei, Chen Qingwei, et al. Estimation of future land use change in the Tianshan mountainous area based on FLUS model[?]. Arid Zone Research, 2019, 36(5): 1270-1279.

[?] Mariana B, Lucian D. Random forest in remote sensing: A review of applications and future directions[?]. Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31.

[?] Zhang Jingdu, Mei Zhixiong, Lyu Jiahui, et al. Simulating multiple land use scenarios based on the FLUS model considering spatial autocorrelation[?]. Journal of Geo-Information Science, 2020, 22(3): 531-542.

Figure 1: Figure 1

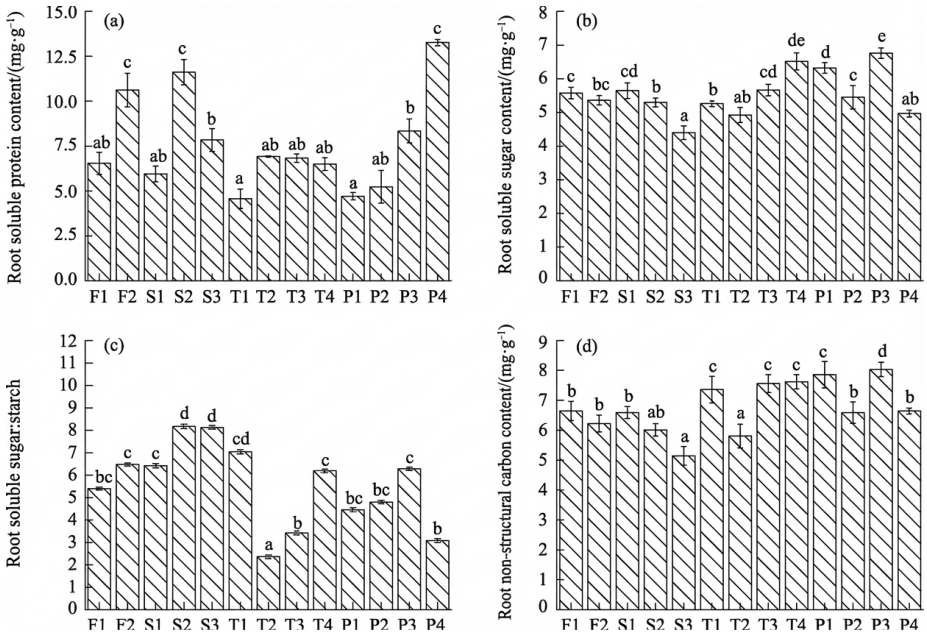

Figure 2: Figure 3

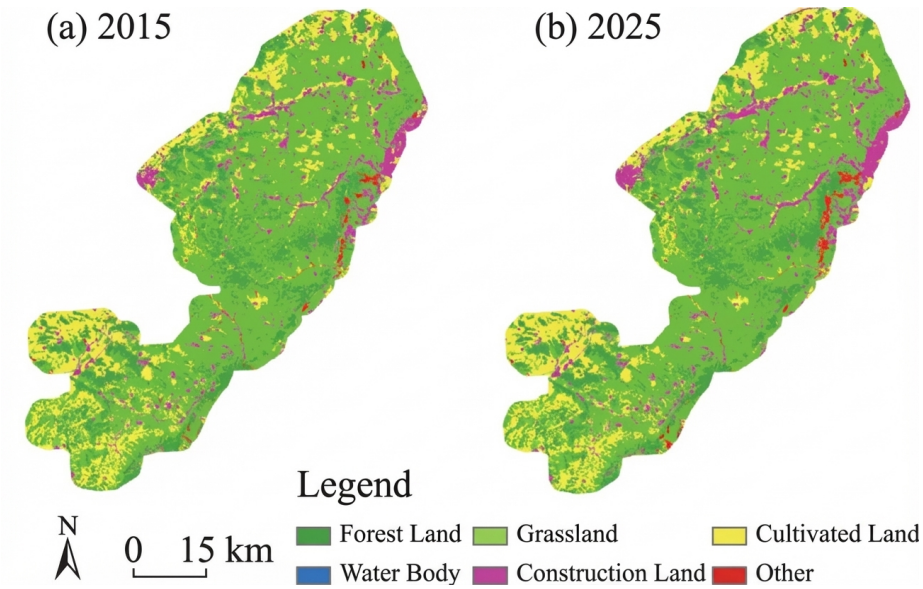

Figure 3: Figure 6

Figures

Source: ChinaXiv —Machine translation. Verify with original.