Postprint: Geographical Detector-Based Analysis of Urban Expansion and Its Driving Factors in Urumqi City

Authors: Zhao Yongyu

Date: 2021-12-14T00:00:00+00:00

Abstract

Taking Urumqi City in Xinjiang, an arid region city in northwestern China, as an example, this study utilized three periods of Landsat TM and OLI remote sensing imagery data from 2000, 2010, and 2020. Employing the random forest classification method and based on 250 m$\times$250 m grid units, the area ratio of each land use type within each grid unit was calculated. The spatial characteristics of land cover/land use (LUCC) in Urumqi from 2000 to 2020 were analyzed, and the geographic detector model was applied to further reveal its driving factors. The spatiotemporal variation characteristics and response factors of land cover/land use in Urumqi from 2000 to 2020 were quantitatively analyzed and evaluated. The results show that: (1) In the past 20 years, Urumqi has experienced rapid urbanization, with a substantial increase in construction land area. Based on the built-up area in 2000, the city has continuously spread and developed toward the north, northeast, northwest, and southeast directions, mainly through outward expansion, while also exhibiting a trend of intensive infill growth within the city. (2) Although green space experienced some fluctuations in the past 20 years, it showed an overall increasing trend; bare land has consistently shown a continuous decreasing trend. (3) The dominant factor influencing the spatiotemporal variation of land cover/land use in Urumqi is the Normalized Difference Vegetation Index (NDVI), which has the highest contribution rate. In the interactive factor detection results, when NDVI and elevation act together, they exhibit a two-factor enhancement with the greatest explanatory power. Monitoring the dynamic changes in land use in Urumqi over the past 20 years can provide valuable insights for the current urban development of Urumqi and also serve as a reference for future urban planning directions.

Full Text

Quantitative Analysis of Urban Expansion and Influencing Factors in Urumqi City Based on Geographical Detector

ZHAO Yongyu[1], Alimujiang KASIM[1,2], GAO Pengwen[1], LIANG Hongwu[1]

[1]School of Geographic Science and Tourism, Xinjiang Normal University, Urumqi 830054, Xinjiang, China
[2]Research Center for Urbanization Development of Silk Road Economic Belt, Xinjiang Normal University, Urumqi 830054, Xinjiang, China

Abstract: Taking Urumqi City in the arid region of northwest China as a case study, this paper analyzes the spatial characteristics of land cover/land use (LUCC) in Urumqi from 2000 to 2020 using Landsat TM/OLI remote sensing imagery from three periods (2000, 2010, and 2020). The analysis employs a random forest classification method based on 250 m × 250 m grid cells. A geographical detector model is further applied to reveal the underlying driving factors, enabling quantitative evaluation of the spatiotemporal change characteristics and response factors of LUCC over the past two decades. The results demonstrate that: (1) Urumqi has experienced rapid urbanization, with construction land area increasing substantially. Based on the 2000 built-up area, the city has expanded continuously toward the north, northeast, northwest, and southeast, primarily through outward expansion while also exhibiting infill growth within the urban interior. (2) Green space area has fluctuated to some extent but shows an overall increasing trend, while bare land has consistently decreased throughout the study period. (3) The geographical detector identifies the normalized difference vegetation index (NDVI) as the dominant factor influencing LUCC spatiotemporal variation, with the highest contribution rate. Interactive factor detection reveals that NDVI and elevation (DEM) exhibit dual-factor enhancement when combined, producing the greatest explanatory power.

Keywords: land use; random forest algorithm; grid cell; geographical detector; Urumqi City

1 Study Area Overview

Urumqi City is located in northwestern China, at the northern foothills of the central Tianshan Mountains and the southern edge of the Junggar Basin, situated in the hinterland of the Eurasian continent. It serves as an important gateway for China' s westward opening and the capital city of the core area of the Silk Road Economic Belt under the "Belt and Road" Initiative. The urban area is surrounded by mountains on three sides, adjacent to the Turpan region in the southeast, bordering the Changji Hui Autonomous Prefecture to the east and west, and neighboring the Bayingolin Mongol Autonomous Prefecture to the south. The city presents a narrow, elongated "Y"-shaped valley basin topography, with higher terrain in the north and south and lower in the northwest, averaging approximately 800 m in elevation. This unique geography combines characteristics of both plain and mountain cities. In recent years, Urumqi' s urbanization process has accelerated, establishing the city as the core political, cultural, and financial center of Xinjiang, with its radiating influence on the entire Xinjiang region and Central Asia increasing daily. Leveraging its unique geographical and locational advantages, Urumqi has gradually developed into a new comprehensive oasis city.

2 Data Sources and Research Methods

This study analyzes the precise, micro-scale spatiotemporal change characteristics of Urumqi in different directions and regions from a grid perspective. It employs the random forest algorithm, which offers higher classification accuracy and superior performance, combined with a land use transfer matrix to quantitatively analyze the substitution and mutual transformation processes among various land use types.

2.1 Data Sources

This research utilized Landsat TM (Thematic Mapper) and OLI (Operational Land Imager) remote sensing imagery (Table 1) from three periods (2000, 2010, and 2020) obtained from the United States Geological Survey (http://www.usgs.gov/). The images were acquired in July and August to minimize seasonal variation effects on land cover changes. The spatial resolution is 30 m, and weather conditions were favorable during acquisition. Meteorological data were obtained from the China Meteorological Data Sharing Service Network, including annual average precipitation, average temperature, monthly precipitation, sunshine hours, evaporation, and daily average wind speed for Urumqi from 2000 to 2020. Digital elevation model (DEM) data with 90 m $\times$ 90 m resolution were obtained from the Geospatial Data Cloud to calculate slope, terrain relief, and river network density. Normalized difference vegetation index (NDVI) data were derived from the Chinese annual vegetation index spatial distribution dataset provided by the Resource and Environmental Science Data Center of the Chinese Academy of Sciences. Socioeconomic data, including population and gross domestic product (GDP) grid interpolation data, were obtained from the Xinjiang Statistical Yearbook and the Resource and Environmental Science Data Center of the Chinese Academy of Sciences.

2.2.1 Land Cover/Land Use Classification

Due to the common "same object with different spectra, different objects with same spectrum" phenomenon in arid region land classification, extracting urban land use from medium-low resolution remote sensing imagery presents certain difficulties. Confusion between bare land and construction land frequently occurs in classification results. This study not only examines the overall spatiotemporal changes of land use in Urumqi but also focuses on fine-scale changes within different urban regions, necessitating extremely high classification accuracy. Considering the characteristics of land use types in Urumqi, the original land use categories were consolidated into four types (Table 2): green space, construction land, water bodies, and bare land (including unvegetated mountain areas). This consolidation aims to improve classification accuracy by reducing the number of land use categories.

2.2.2 Random Forest Classification

Random forest classification is an ensemble learning method that combines multiple weak classifiers based on decision trees to produce a strong classifier forest with accurate results. Proposed by Breiman, this machine training algorithm based on decision trees determines final classification results through voting among multiple trees. The algorithm requires no data dimensionality reduction, is suitable for high-dimensional data, and is less prone to overfitting problems. Compared with traditional machine learning algorithms such as support vector machines, decision trees, and artificial neural networks, random forest demonstrates superior performance in classification efficiency and accuracy. Consequently, an increasing number of scholars have applied this algorithm to extract high-precision classification results from remote sensing imagery. This study employs the random forest algorithm for land use classification based on remote sensing imagery data.

2.2.3 Grid Cell Method

The grid cell method establishes a vector grid platform using ArcGIS, with the study area boundary as the extent, to create grids according to research requirements. Land use classification results derived from remote sensing imagery are transferred into corresponding grid cells, and the percentage area of each land use type within each grid cell is calculated. When land use types cross multiple grid cell boundaries, previous studies often prioritized assigning such pixels to the grid cell with larger area coverage, which causes discrepancies between actual and allocated areas. This study converts the random forest classification results into vector data and intersects them with the established vector grid, dividing patches crossing multiple grid boundaries into smaller fragments (At). The area of each land use type within each grid cell is calculated and summed, then the proportion (Pt) within the corresponding grid cell is computed using the formula:

$$P_t = \frac{\sum_{i=1}^{n} A_{t,i}}{A_g} \times 100\%$$

where P_t is the percentage area of land use type t in each grid cell; $A_{t,i}$ is the area of land use type t after being divided into small fragments (m^2); n is the number of fragments of land use type t clipped within each grid cell; and A_g is the actual area of each grid cell (m^2).

Considering the study area size, computational efficiency, and research objectives, after repeated experiments, each vector grid cell was set to 250 m × 250 m, generating approximately 38,000 grid cells.

2.2.4 Geographical Detector

The core feature of the geographical detector is detecting spatial stratified heterogeneity to reveal underlying driving forces. This emerging statistical method can effectively screen spatial heterogeneity among variable factors. The factor detector quantifies the degree to which different variables influence the spatiotemporal variation heterogeneity of land use types and detects their impact magnitude using the formula:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}$$

where L is the number of types of influencing factors; N_h and N are the sample numbers of type h and the entire region, respectively; σ^2 is the variance of influencing factors; and q is the detection power value of a certain factor on land use types. The value range of q is [0,1], where a larger q value indicates greater influence of the factor on the spatial distribution of land use types, and vice versa.

Interaction detection evaluates whether the combined effect of different influencing factors changes their explanatory power on land use types. The interaction relationships between two factors can be classified into several types (Table 3).

3 Results and Analysis

3.1 Overall Land Cover/Land Use Change in Urumqi

Based on the land use dynamic degree in different periods, the magnitude of change for various land use types follows the order: construction land > green space > water bodies > bare land. From 2000 to 2020, the dynamic changes of different land use types showed large fluctuations, with construction land and green space increasing at rates of 22.51% and 1.16%, respectively, while bare land and water bodies decreased at rates of 0.24% and 0.95%, respectively. The primary reason is that with the acceleration of urbanization in Urumqi, urban land area has continuously increased, with construction land encroaching on surrounding bare land, farmland, and grassland, causing continuous reduction in these areas.

3.2 Land Cover/Land Use Transfer Matrix Characteristics

The increase or decrease in land use type areas cannot intuitively reflect the conversion processes between different land use types. A transfer matrix (Figure 4) was employed to analyze the mutual transformation processes among various types in Urumqi from 2000 to 2020. Over the past 20 years, significant conversions occurred among bare land, green space, and construction land. The transfer area to construction land followed the order: bare land > green space > water bodies. All four land use types experienced both large-area transfers in and out simultaneously (Table 5). The substantial increase in construction land area is evident from multi-temporal remote sensing imagery. Throughout the study period, all land use types underwent certain mutual transformations. From 2000 to 2020, conversions among bare land, green space, and construction land were particularly prominent. Green space area also continued to grow, primarily transferred from bare land, which is closely related to the "barren hill greening" project implemented in Urumqi in recent years. The greening coverage rate in built-up areas has increased from 21.8% to 36.16%, significantly increasing green space in many areas of Urumqi, such as the key greening transformation areas of Yamalik Mountain, Hongguang Mountain, and Spider Mountain.

3.3 Driving Factors of Land Cover/Land Use Spatiotemporal Distribution

The construction land area in Urumqi has consistently shown an increasing trend, with significant variations in different directions and regions. Based on the 2000 built-up area, expansion has spread toward the north, northeast, northwest, and southeast, primarily through outward expansion while also exhibiting infill growth within the urban interior (Figure 5). This pattern results from multiple overlapping factors. In recent years, Urumqi has implemented regional integration policies such as "Urumqi-Changji Integration," "Urumqi-Changji Metropolitan Circle," and "Integration of Corps and Locality," promoting increasing economic exchanges with Changji City to the northwest, Shihezi City, Wujiaqu City (a corps city) to the north, and Fukang City to the northeast. The administrative merger of Miquan City in the northeast into Urumqi' s jurisdiction in 2007 has further driven built-up area expansion in the northeast direction. The urban road system also exerts a "pulling" effect on urban expansion, with national highways, provincial roads, Lianhuo Expressway, Tuwu-Dawu Expressway, and Wuchang Expressway promoting construction land expansion toward the northwest, northeast, and southeast.

Factor detection results (Table 6) show that the explanatory power of different factors on land use spatiotemporal distribution from 2000 to 2020, in descending order, is: NDVI > DEM > GDP > annual average temperature > precipitation > sunshine hours. NDVI, as the dominant factor influencing land use spatiotemporal variation in Urumqi, has an average contribution rate of 40.82%, followed by DEM with an average contribution rate of 33.94%. This indicates that vegetation cover conditions and topographic factors are the primary drivers of land cover/land use distribution patterns in Urumqi. The distribution of oases and topographic conditions determine the land use pattern. The interactive detection results (Figure 6) show that all factor interactions exhibit enhancement effects, indicating significant coordination and correlation among factors. When NDVI and DEM act together, they show dual-factor enhancement with the greatest explanatory power.

4 Conclusions

This study utilized remote sensing data to analyze the spatiotemporal change patterns of land cover/land use in Urumqi from 2000 to 2020, reflecting the dynamic changes of urban expansion in Urumqi. Most existing studies on land cover/land use change detection focus on overall changes and type conversions within study areas, with relatively few fine-scale quantitative analyses of urban interiors. Therefore, this research adopts a grid perspective to precisely and microscopically analyze the spatiotemporal change characteristics of Urumqi in different directions and regions. Using the random forest algorithm with higher classification accuracy and superior performance, combined with a land use transfer matrix, this study quantitatively analyzes the substitution and mutual transformation processes among various land use types, yielding the following conclusions:

(1) Urumqi has experienced rapid urbanization, with construction land area increasing substantially. Based on the 2000 built-up area, the city has expanded continuously toward the north, northeast, northwest, and southeast, primarily through outward expansion while also exhibiting infill growth within the urban interior. Green space has fluctuated to some extent but shows an overall increasing trend, while bare land has consistently decreased. The long-term "barren hill greening" project, implemented in key areas such as Yamalik Mountain, Hongguang Mountain, and Spider Mountain, has significantly increased green space in many regions of Urumqi.

(2) The geographical detector identifies the main influencing factors causing spatiotemporal variation in land cover/land use. NDVI emerges as the dominant factor with the highest contribution rate, followed by DEM. The distribution of oases and topographic conditions determine the land cover/land use distribution pattern in Urumqi. Interactive detection results show that all factor interactions exhibit nonlinear enhancement and dual-factor enhancement, indicating significant coordination and correlation among factors. When NDVI and DEM act together, they show dual-factor enhancement with the greatest explanatory power. NDVI represents the most important driving factor for land cover/land use changes in Urumqi, posing the greatest threat to land cover/land use change. This reflects Urumqi' s vulnerability as a typical arid oasis city, with its ecosystem doubly affected by natural and human factors in arid regions. The human-land relationship is extremely sensitive, and urban development re-

lies on oasis distribution with water-dependent characteristics, preventing Urumqi from becoming a mega-city and limiting it to aggregating within oases according to its ecological carrying capacity.

References

[1] Liu Jinyong, Kong Fanhua, Yin Haiwei, et al. Land use change and its effects on ecosystem service value in Jinan City of Shandong Province, east China[J]. Chinese Journal of Applied Ecology, 2013, 24(5): 1231-1236.

[2] Gao J, Li F, Gao H, et al. The impact of land use change on water-related ecosystem services: A study of the Guishui River Basin, Beijing, China[J]. Journal of Cleaner Production, 2016, 163(Suppl. 1): S139-S148.

[3] Shen G, Yang X C, Jin Y X, et al. Land use changes in the Zoige Plateau based on the object oriented method and their effects on landscape patterns[J]. Remote Sensing, 2019, 12(1): 14.

[4] Liu Yaru, Wang Cong, Yan Lijiao. Impacts of land use change on ecosystem services in the agricultural area of North China Plain: A case study of Shangqiu City, Henan Province, China[J]. Chinese Journal of Applied Ecology, 2018, 29(5): 1597-1606.

[5] Genbatu Ge, Shi Z J, Zhu Y J, et al. Land use/cover classification in an arid desert oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms[J]. Global Ecology and Conservation, 2020, 22: e00971.

[6] Meghan Blumstein, Jonathan R Thompson. Land use impacts on the quantity and configuration of ecosystem service provisioning in Massachusetts, USA[J]. Journal of Applied Ecology, 2015, 52(4): 1009-1019.

[7] Lü Chen, Lan Xiuting, Sun Wei. A study on the relationship between natural factors and population distribution in Beijing using geographical detector[J]. Journal of Natural Resources, 2017, 32(8): 1385-1397.

[8] Wang Jinfeng, Xu Chengdong. Geodetector: Principle and prospective[J]. Acta Geographica Sinica, 2017, 72(1): 116-134.

[9] Zhang M Y, Zhang L, Ren X L, et al. Effect of land use and land cover change on the changes in net primary productivity in karst areas of southwest China: A case study of Huanjiang Maonan Autonomous County[J]. Journal of Resources and Ecology, 2020, 11(6): 606-616.

[10] Wang Dongchen, Du Peijun, Su Hongjun, et al. Land use/land cover changes in Datong based on remote sensing data[J]. Journal of Arid Land Resources and Environment, 2015, 29(7): 68-75.

[11] Zhu Zengyun, Kasimu Alimujiang. Prediction of land use landscape pattern in Hutubi County based on CA-Markov model[J]. Ecological Science, 2020, 39(1): 136-145.

[12] Schirpke U, Kohler M, Leitinger G, et al. Future impacts of changing land use and climate on ecosystem services of mountain grassland and their resilience[J]. Ecosystem Services, 2017, 26: 79-94.

[13] Chen Leishi, Zhao Junsan, Li Yi, et al. On land use classification by means of machine learning based on multi-source remote sensing fusion[J]. Journal of Southwest China Normal University (Natural Science Edition), 2018, 43(10): 103-111.

[14] Li Ping, Wu Manqiao, Zeng Lianming. On application of support vector machine technology to land use monitoring[J]. Bulletin of Surveying and Mapping, 2010(8): 28-30, 67.

[15] Kasimu Alimujiang, Aimaitia Yusupujiang. Study on spatial structure of urban greenbelt landscapes in Urumqi City based on ALOS satellite image[J]. Areal Research and Development, 2012, 31(2): 86-89.

[16] Li Guoqing, Huang Jinghua, Liu Guan, et al. A study of the landscape fragmentations of land cover structure based on Landsat8 remote sensing image: A case study of Mata Watershed in Yan'an, Shaanxi Province[J]. Remote Sensing for Land and Resources, 2020, 32(3): 121-128.

[17] Zhao Haili, Wang Qiwen, Zhu Lixiang, et al. Analysis of spatial temporal evolution and influencing factors of health resources in underdeveloped areas based on geodetectors[J]. Arid Land Geography, 2021, 44(2): 594-603.

[18] Zhao Lingdi, Li Shasha, Zhao Zhibo, et al. Temporal and spatial evolution and regional difference analysis of urban ecological efficiency in arid and semiarid areas[J]. Arid Land Geography, 2020, 43(2): 449-457.

[19] Wu Zhijie, Zhao Shuhe. A study of enhanced index based built-up index based on Landsat TM imagery[J]. Remote Sensing for Land & Resource, 2012, 93(2): 50-55.

[20] Zheng Xu, Wei Lemin, Guo Jianjun, et al. Driving force analysis of water yield in inland river basins of arid areas based on geo-detectors: A case of the Shule River[J]. Arid Land Geography, 2020, 43(6): 1477-1485.

[21] Ghulam Abduwasit, Ghulam Oghlan, Maitiniyazi Maimaitijiang, et al. Remote sensing based spatial statistics to document tropical rainforest transition pathways[J]. Remote Sensing, 2015, 7(5): 5993-6016.

[22] Wang Na, Li Qiangzi, Du Xin, et al. Identification of main crops based on the univariate feature selection in Subei[J]. Journal of Remote Sensing, 2017, 21(4): 519-530.

[23] Wang Xuewei, Wang Shijun, Song Yang, et al. Changchun land use spatio-temporal variation under the transportation elements driving[J]. Economic Geography, 2015, 35(4): 155-161.

[24] Sexton J O, Urban D L, Donohue M J, et al. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record[J]. Remote Sensing

of Environment, 2013, 128: 246-258.

[25] Zhu Yongsen, Zeng Yongnian, Zhang Meng. Extract of land use/cover information based on HJ satellites data and object-oriented classification[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(14): 258-265.

[26] Maitiniyazi Maimaitijiang, Kasimu Alimujiang. Spatio-temporal change of Urumqi urban land use and land cover based on grid cell approach[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(1): 210-216.

[27] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[28] Dong Shishi, Huang Zhexue. A brief theoretical overview of random forests[J]. Journal of Integration Technology, 2013, 2(1): 1-7.

[29] Fang Chuanglin. Strategic thinking and spatial layout for the sustainable development of urban agglomeration in northern slope of Tianshan Mountains[J]. Arid Land Geography, 2019, 42(1): 1-11.

[30] Zhang Weichun, Liu Hongbin, Wu Wei. Classification of land use in low mountain and hilly area based on random forest and Sentinel-2 satellite data: A case study of Lishi Town, Jiangjin, Chongqing[J]. Resources and Environment in the Yangtze Basin, 2019, 28(6): 1334-1343.

[31] Gu Xiaotian. Research on land use/land cover classification based on machine learning in the Huangshui River Basin[D]. Xining: Qinghai Normal University, 2018.

[32] Han Haiqing, Wang Xuhong, Niu Linzhi, et al. The LUCC characteristics and driving forces of cultivated land in the five Central Asian countries from 1992 to 2015[J]. Chinese Journal of Agriculture, 2021, 29(2): 325-339.

[33] Wang Shanshan, Chen Xi, Duan Hanming, et al. Study on response of land surface temperature to land use and land cover change using remote sensing data: A case on Urumqi, China[J]. Journal of Desert Research, 2012, 32(3): 878-884.

[34] E Ziji, Kasimu Alimiujiang, Maitiniyazi Maimaitijiang, et al. Temporal and spatial variations of urban land cover/land use based on grid element in northwest arid area of China: A case of Kashgar City in Xinjiang[J]. Arid Land Geography, 2018, 41(3): 625-633.

[35] Baolangtijiang Haziya, Wu Zhaopeng, Chen Xuegang, et al. Analysis of landscape pattern change and driving force in Urumqi City[J]. Ecological Science, 2018, 37(1): 62-70.

[36] Yang Ya' nan, Kasimu Alimujiang. Relationship between regional transportation advantage and development of regional economy in Xinjiang under the background of the national strategy of the Belt and Road[J]. Arid Land Geography, 2017, 40(3): 680-691.

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