Study on land change based on principal component analysis

Jing Shen, Dongfan Piao*
College of Geography and Ocean Sciences, Yanbian University, Yanbian 133002, China

*Corresponding author: piaodongfan@ybu.edu.cn

Abstract. The Loess Plateau is the area with the most serious soil erosion and the most fragile ecological environment in China and even in the world. The deterioration of ecological environment may be closely related to unreasonable land use. In order to study the characteristics and driving factors of land use in small watershed of Loess Plateau, based on the land use data, DEM data and social data of Wuding River Basin in 1980, 1990, 2000, 2010 and 2018, this paper uses ArcGIS platform, land transfer matrix and land use dynamic degree to analyze them, and uses principal component analysis to discuss land use driving forces.

Keywords: Loess Plateau, Wuding River Basin, Driving forces

1. Introduction
In recent years, with the development of social science and technology, the increasing strengthening of human activities, people's material and spiritual needs are also rising, correspondingly, people are constantly changing the nature of land use, so that land resources have been seriously damaged, causing a series of immeasurable environmental problems. At present, land use / land cover has become a focus of scholars at home and abroad. It has become a new trend to study land use change and its effects from a regional perspective, which has been highly valued by many international organizations. As an important aspect of landscape dynamics, land use plays a very important role in the formation of soil erosion. Over the past few years, there have been a series of ecological and environmental problems in the Loess Plateau, such as soil erosion, broken terrain and vegetation degradation. This paper will use the principal component analysis method to explore the characteristics and driving forces of land use change in Wuding River Basin of the Loess Plateau, and put forward relevant governance suggestions.

2. General situation and research methods of the study area
2.1. Overview of the study area
Wuding River Basin, a tributary of the Yellow River, is located in the southern edge of Mu Us sandy land and the northern edge of the Loess Plateau. Wuding River Basin belongs to temperate continental monsoon climate, rainy in summer, dry and cold in winter. The average annual precipitation is 350 mm, and the average annual evaporation is 1100 mm. The precipitation is mainly concentrated in summer, and the distribution is uneven. It is characterized by heavy rain and drought, and it is easy to form flood or drought.
Since 1949, a series of measures have been taken to prevent soil erosion in Wuding River Basin. Remarkable achievements have been made in controlling soil erosion, improving ecological environment and realizing sustainable development. With the change of land use and land use, it is necessary for us to have a deeper understanding of these land use changes.

2.2. Research methods

2.2.1. Land transfer matrix. Land transfer matrix is a classical method to study the direction and quantity of land use transfer, which can reveal the evolution process of land use pattern. Land use change is expressed by land use transfer matrix, in which the sum of row elements represents the area of this kind of land before transfer, and the sum of column elements represents the area of this kind of land after transfer. The land use transfer matrix of 1980-1990, 1990-2000, 2000-2010 and 2010-2018 can be obtained by using the tool module of ArcGIS software and office software. At the same time, the change rate of land use transfer can be obtained more conveniently.

2.2.2. Principal component analysis. Because the driving force of land use change needs to consider many factors, there are many methods to extract the influential factors. The common methods include factor analysis, grey relational clustering and principal component analysis. However, because the influencing factors of land use change are multivariable, according to the previous research results, it is necessary to use principal component analysis to reduce the dimension. The driving force of land use change is better.

The main principle of principal component analysis is dimension reduction. Specifically, by processing multi-dimensional data of different orders of magnitude, we can find variables reflecting common information, delete closely related repeated variables, and establish new variables, so that these new variables are independent of each other and reflect the original information as much as possible. In addition, on the premise of ensuring the accuracy of the research, it can simplify the complex research and greatly improve the research efficiency. In general, the mathematical treatment is to assume that there are n samples and each sample has m variables, thus forming a matrix of order n × M,

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \cdots & a_{mn}
\end{bmatrix}
\]

Note that the original variable is \(a_1, a_2, \cdots, a_n\). Let the new variable after dimension reduction be \(Y_1, Y_2, \cdots, Y_p\) \((p \leq m)\), then

\[
\begin{align*}
    Y_1 &= l_{11}a_1 + l_{12}a_2 + \cdots + l_{1m}a_m \\
    Y_2 &= l_{21}a_1 + l_{22}a_2 + \cdots + l_{2m}a_m \\
    & \quad \vdots \\
    Y_p &= l_{p1}a_1 + l_{p2}a_2 + \cdots + l_{pm}a_m
\end{align*}
\]

Represents the coefficient of the linear combination, \(Y_1, Y_2, \cdots, Y_p\). They are not related to each other, \(Y_i\) \((i = 1, 2, \ldots, p)\) is the largest variance in all linear combinations of \(a_1, a_2, \cdots, a_m\).

2.2.3. Kaiser Meyer Olin and Bartlett test. Kaiser Meyer Olin (KMO) is one of the validity indexes for principal component analysis, which is used to compare simple correlation coefficient and partial correlation coefficient between variables. Commonly used KMO measurement standard: KMO is close to 1, the correlation is strong, the number of partial phase relationship is far less than the simple correlation coefficient. Generally, KMO is above 0.9, which is very suitable for factor analysis; KMO
is between 0.8 and 0.9, which is more suitable for factor analysis; KMO is between 0.7 and 0.8, which is suitable; KMO is between 0.6 and 0.7, which is acceptable; KMO is between 0.5 and 0.6, which is very poor; KMO is less than 0.5, which is not suitable.

Bartlett test is used to test the correlation between the variables in the correlation matrix, whether it is the unit matrix, that is, whether each variable is independent. In factor analysis, if the original hypothesis is rejected, it means that factor analysis can be done. If the original hypothesis is not rejected, it means that these variables may provide some information independently and are not suitable for factor analysis. Generally, when the significance probability of Bartlett’s ball test statistical value $p < 0.05$, it is suitable to run the principal component analysis. Using the above two methods, we can see whether the selected factors can be used for principal component analysis.

3. Conclusion

3.1. Land transfer matrix

3.1.1. Change analysis of land use area. From the changes of land use area in the five periods, it can be concluded that great changes have taken place in the land use mode in the Wuding River Basin during the study period: 1. The area of grassland has been increasing year by year, and it will basically occupy the dominant position in 2018, which may be due to the effect of returning farmland to forest and grassland; 2. The area of cultivated land has been decreasing, accounting for 20.32% until 2018; 3. The area of water area is increasing. The area is also decreasing year by year; 4. The area of forest land has been increasing; 5. The unused land is decreasing and gradually transformed into other types of land; 6. The area of construction land has increased a lot, mainly because the residential land has increased year by year.

3.1.2. Land use transfer matrix analysis. The transfer matrix of land use change can clearly and intuitively show the complex conversion process between land use types in each period. Since 2000, the area of cultivated land transferred to woodland has increased by more than 40%, especially in recent years.

3.2. Driving force analysis of land use pattern change

3.2.1. Determination of principal component index. The main factors affecting land use change are natural, economic, environmental, policy and other factors. In order to follow the principles of data desirability and feasibility, this study takes the data from 2000 to 2018 as the research unit, selects socio-economic driving factors and natural factors as the basic data, and uses the correlation analysis method to screen out the following nine factors, using SPSS. The software standardized the data and selected a principal component whose eigenvalue was greater than 1. It can be seen from table 6 that there are different degrees of correlation among the nine indicators.

| Variable number | Variable name                         | Company                  |
|-----------------|---------------------------------------|--------------------------|
| V1              | Total Population                       | ten thousand people      |
| V2              | GDP                                   | RMB100mn                 |
| V3              | NDVI                                  |                          |
| V4              | Annual average temperature            | Temperature              |
| V5              | Non agricultural population           | ten thousand people      |
| V6              | Per capita GDP                        | RMB100mn                 |
| V7              | Investment in fixed assets            | Ten thousand yuan        |
| V8              | Total industrial output value         | Ten thousand yuan        |
| V9              | Rural per capita income               | element                  |
3.2.2. **Principal component analysis.** Through SPSS software, the correlation coefficient matrix of driving factors, eigenvalue and contribution rate of each component, and principal component load matrix are calculated by principal component analysis. According to the cumulative contribution rate greater than based on the principle of 70% and eigenvalue greater than 1, two principal components are extracted from the driving factors. The cumulative contribution rate of the first and second components reaches 97.517%. Therefore, the purpose of analysis can be achieved by selecting the first principal component as the total population at the end of the year and the second principal component as GDP. The two principal components can reflect most of the information of the original driving factors.

On this basis, the load matrix of the two principal components is calculated. According to the principle that the closer the value is to 1, the greater the degree of correlation, it can be concluded that the first principal component is related to GDP, rural per capita income, total industrial output value, etc.; the second principal component is significantly related to the year-end total population, average annual temperature, and non-agricultural population. Therefore, we can see that the two principal components reflect the population size, economic development level and climate factors, indicating that these factors mainly drive the land use change in Wuding River Basin.

3.3. **Kaiser Meyer Olin (KMO) and Bartlett test results**

KMO statistics is between 0 and 1. When the square sum of simple correlation coefficients among all variables is far greater than the square sum of partial correlation coefficients, the KMO value is close to 1. The closer the KMO value is to 1, the stronger the correlation between variables, and the more suitable the original variables for factor analysis. In this analysis, KMO value reaches 0.685, which proves that this variable is suitable for principal component analysis.

This paper studies the characteristics and driving forces of land use change in the past 40 years from 1980 to 2018 from the perspective of time and space. Through principal component analysis, it can be seen that the driving forces of land use change in Wuding River Basin are the total population at the end of the year and GDP, which shows that population growth and economic development have a great impact on the change of land use structure. In addition, policy support is also needed. It also plays an important role in land use change.

**References**

[1] Li Lijuan, Yang Junwei, Jiang Dejuan, Li Jiuyi, Hou Xiyong. Spatiotemporal change of land use in Wuding River Basin in 1990s [J]. Geography research, 2005 (04): 527-534 + 656

[2] Wei He Jie, Zhang Yanfang, Dong Xiaobin, Hu Xiaohuan, Wang Xuechao, Lu Nachuan, Zhao Yanan. Vegetation cover change and its carbon sequestration effect in Wuding River Basin from 2000 to 2013 [J]. Soil and water conservation bulletin, 2016,36 (01): 44-50

[3] Wang Simeng, Huang Chang. Monitoring and evaluation of natural ecological environment quality based on remote sensing and GIS: a case study of Wuding River Basin [J]. Geography of arid areas, 2018,41 (01): 134-141