Environment, Resource and Ecology Journal (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/erej.2025.090115
ISSN 2616-3756 Vol. 9 Num. 1

Decision Support for Population Regulation in Arid and Semi-Arid Rural Areas Based on Big Data Methods: A Case Study of Inner Mongolia

Gao Xiang, Wu Xiaoguang, Xu Jingxin, Zheng Wenjing, Yun Hao

Inner Mongolia Autonomous Region Institute of Territorial Spatial Planning, Hohhot, 010013, Inner Mongolia Autonomous Region, China

Keywords: Arid and Semi-Arid Areas, Rural Areas, Human Activities; Environmental Factors, Pattern Clustering, Internal Connection

Abstract: Arid and semi-arid regions are widely distributed across the globe, and their governance is a key component in achieving the United Nations Sustainable Development Goals, attracting significant international attention. Understanding the characteristics of rural population dynamics in such regions and exploring their development pathways can help effectively guide various production and living activities, thereby promoting harmonious coexistence between humans and nature. Using the Inner Mongolia Autonomous Region as a case study, this paper integrates multi-source spatiotemporal big data and employs geostatistical and machine learning methods to investigate the intrinsic relationships between human activities and environmental factors from multiple dimensions. Based on the data analysis, three main conclusions are drawn. First, rural human activities in the arid and semi-arid areas of Inner Mongolia exhibit spatial heterogeneity and are broadly associated with six environmental factors. Second, based on three indicators—resident population distribution, human activity intensity, and resident population migration—the town-level units within the study area can be classified into three types: shrinking towns, stagnant towns, and developing towns. Third, significant differences exist in the development levels of environmental factors across units with different human activity patterns. Among these, economic, transportation, and public service factors show the greatest disparities, while climatic and locational factors exhibit relatively smaller differences. The findings of this study offer localized insights from China to support research and decision-making in arid and semi-arid regions.

1. Introduction

Arid and semi-arid regions are typically defined as areas with annual precipitation of less than 500 millimeters. These regions account for approximately one-third of the Earth's land surface and are characterized by fragile ecosystems, facing a range of issues such as desertification, soil erosion, and water scarcity. These challenges intensify the impacts of climate change on both the natural environment and human society.

In China, arid and semi-arid regions have become key areas for the advancement of ecological civilization. Currently, these regions cover more than 3 million square kilometers across China, accounting for over one-third of the country's total land area, and are primarily concentrated in the Western and northern China [1]. Among them, the Inner Mongolia Autonomous Region, as a vital ecological security barrier in northern China and a critical zone for environmental governance, contains 130,667 square kilometers of arid and semi-arid land, representing 11.04% of its total area. This region faces prominent issues such as land desertification and soil erosion, and its effective management is crucial for achieving national sustainable development goals.

Human activities in rural areas have a significant impact on the ecosystems of arid and semi-arid regions. Research has shown that the influence of human activity factors on vegetation is comparable to that of climatic factors, primarily reflected in several aspects. First, irrational land use practices—for example, in the Manas River Basin, excessive agricultural reclamation has led to ecological problems such as lake desiccation, grassland degradation, and soil salinization [2]. Second, improper water resource management—as seen in northwest China, population growth has encroached on ecological water use, resulting in an increase of 6.5 million hectares of desertified land in the Shiyang River Basin [3]. Third, overburdened agricultural activities—a study of the Tarim River Basin found that habitat quality in agricultural and settlement areas is relatively low, and the expansion of oasis agriculture and residential zones, such as through grazing, has been a major driver of native habitat loss and fragmentation [4]. These findings highlight the importance of deeply exploring human activity patterns in arid and semi-arid rural areas and guiding various production and living behaviors based on environmental carrying capacity, thereby promoting development in a manner that supports harmony between humans and the land.

In terms of measuring human activity characteristics, researchers generally employ either direct or composite measurement methods. The former directly uses relevant indicators of human activity to characterize the associated features, such as population density, population growth rate, and urbanization rate [5]. The latter uses composite indicators such as per capita ecological footprint and vegetation coverage rate. For example, a study on the ecological footprint, ecological carrying capacity, and footprint depth in Xilin Gol League used an improved three-dimensional ecological footprint model for evaluation, analyzing the temporal and spatial evolution characteristics of these indicators to reflect the changes in human activity in the region.

Building upon the examination of human activities, some scholars have further explored the intrinsic relationship between these activities and environmental changes in arid and semi-arid regions. Several studies have shown that both climate change and human activities jointly influence ecological processes in these regions, with human activities often playing a dominant role [7]. Meanwhile, emerging issues such as water resource scarcity, rapid urbanization, and excessive exploitation of energy resources continue to hinder the sustainable development of the ecological environment, with human activities often driven by factors such as government policies and economic development [8].

Meanwhile, existing studies still have two main shortcomings. First, the characterization of human activities is largely based on macro-level statistical and survey data, which can only reflect the overall human activity situation in a region. Second, regarding the environmental differences corresponding to different human activity patterns, existing research often focuses on a limited number of associated factors, lacking systematic analysis.

Based on the above context, this study utilizes spatiotemporal big data on human activities and machine learning methods to address three key questions: 1) Do human activities in arid and semi-arid rural areas exhibit differentiated patterns? What social characteristics do these differences reflect?2) Behind these differentiated patterns, are there deeper economic, social, and environmental imbalances? 3) How can data analysis conclusions support precise governance decision-making? To address these questions, the study analyzes the spatial distribution characteristics of human activities in arid and semi-arid rural areas of Inner Mongolia, conducts cluster analysis of human activity patterns, and examines the heterogeneity of geographical units under different human activity patterns in terms of economic, social, and environmental factors. Ultimately, the study aims to provide decision support for territorial spatial governance in arid and semi-arid rural areas.

2. Data and Method

2.1 Study area

The scope of this study covers the arid and semi-arid rural areas of Inner Mongolia Autonomous Region, which are primarily located in the northern-central part of the region (Figure 1). Given that the study focuses on rural areas, urban districts, including the streets and offices in county seats, were excluded from the research scope. The total area of the study area is 130,667 km^2, accounting for 11.04% of the total land area of the entire region. This study uses towns as the basic analytical unit, with a total of 45 towns. There are significant differences in land area and population development levels among the towns.

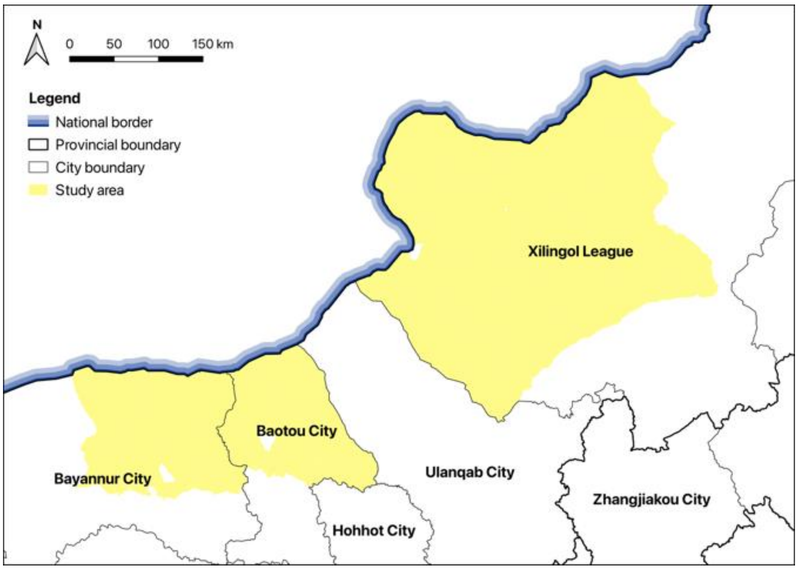


Figure 1: Study Area

2.2 Research data

2.2.1 Human activity indicators

This study uses three population big data sets provided by Baidu Map's "Huiyan" service: resident population distribution, human activity intensity, and resident population migration.

(1) Resident population distribution

By continuously analyzing the location information of all users over a period of time (six months in this study), and employing methods such as identifying nighttime stay locations, the residence of users is determined. The number of users residing in each grid is then aggregated to obtain the resident population for that grid. The resident population data used in this study was obtained on November 30, 2023, with a grid size of 100 meters by 100 meters.

(2) Human activity intensity

By aggregating the cumulative number of active users in each grid for each hour, the human activity intensity for each grid during that period is obtained. This study uses hourly human activity intensity data from November 23, 2023, to November 30, 2023, with a grid size of 100 meters by 100 meters.

(3) Resident population migration

The total population moving from grid A (origin) to grid B (destination) during a certain period is aggregated to determine the total scale of resident population migration between grid A and grid B. This study calculates the cumulative total migration of resident populations between grids from November 30, 2022, to November 30, 2022, with a grid size of 100 meters by 100 meters.

Based on the above data, this study further calculates the following three human activity indicators (Table 1).

Table 1: Human Activity Indicators

Name	Abbr.	Definition	Calculation
Resident population size	RES	The total number of resident population in towns	$\mathrm{RES} = \sum_i \mathrm{r_i}$ (1) $\mathrm{r_i}$ refers to resident population size in grid i
Average human activity intensity	ACT	The average intensity of human activities in each grid within the town	$\mathrm{ACT} = \frac{\sum_n \mathrm{act_n}}{\mathrm{N}}$ (2) N refers to the total number of grids within the town, $\mathrm{act_n}$ refers to the 7-day hourly average human activity intensity for grid n
Population emigration rate	FR_{out}	The proportion of the resident population in the town moving out to other areas relative to the total resident population of the town	$FR_{out} = \frac{FLOW_{out}}{RES}$ (3) $FLOW_{out}$ refers to the size of population emigration, refers to the resident population size

2.2.2 Environmental factors

This study selects six environmental factors related to human activities (Tab.2). The Pearson correlation coefficient results indicate that there is no strong correlation between the factors, meaning each factor has independent representational ability.

(1) Per capita gross national product (GDP)

Economic development level is a key factor influencing human activity, the per capita GDP of each town is calculated as follows.

$$GDP_i = \frac{G_i}{RES} \tag{4}$$

Where is the total GDP of the town, obtained by aggregating data from open-source GDP grid distribution data.

(2) Transportation remoteness (TRS)

Transportation remoteness is defined as the cost of accessing regional transportation services for a research unit. In this study, the indicator for town i represents the straight-line distance between the town's location and the nearest highway entrance/exit. The location data for the highway entrances/exits were obtained by the author through online methods.

(3) Location conditions (LOC)

Location conditions reflect the degree to which a town is influenced by the proximity to nearby central cities. In this study, "it" represents the straight-line spatial distance between town i and the nearest regional central city (prefecture-level city urban area).

(4) Medical service level (SRV)

This study uses the Two-Step Floating Catchment Area method (2SFCA)[6] to calculate the level of general medical services in each town. The general hospital POI data were obtained by the author through online methods.

First, calculating the supply ratio of the medical facility j,

$$R_j = \frac{S_j}{\sum_{k\in\{d_{kj}\leq d_0\}} G(d_{kj}, d_0) P_k} \quad (5)$$

Where is the resident population of grid k within the effective service range of healthcare facility j ($d_{kj} \leq d_0$)? The maximum service radius of the medical facility is set to 50 km in this study. What is the Euclidean distance from grid k to facility j? What is the capacity of facility j? For simplification, this study assumes that the scale of all the general hospitals involved is similar. What is the Gaussian function considering spatial friction, and what is its calculation method?

$$G(d_{kj}, d_0) = \begin{cases} \frac{e^{-\left(\frac{1}{2}\right)\times\left(\frac{d_{kj}}{d_0}\right)^2} - e^{-\left(\frac{1}{2}\right)}}{1-e^{-\left(\frac{1}{2}\right)}}, & if \quad d_{kj} \leq d_0 \\ 0, & if \quad d_{kj} > d_0 \end{cases} \quad (6)$$

Second, calculating the facility service level for grid i,

$$A_i = \sum_{l\in\{d_{il}\leq d_0\}} G(d_{il}, d_0) R_l \quad (7)$$

Where represents the supply ratio of facility l.

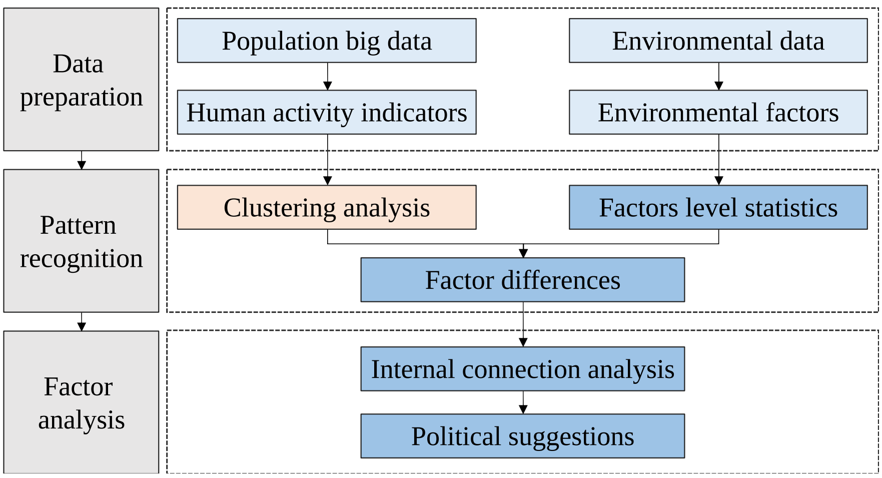
(5) Educational service level (EDU)

The calculation principle for the service level of primary schools is the same as that for healthcare facilities. The primary school POI data were obtained by the author through online methods. The search radius for educational facilities is set to 8 km.

(6) Annual average temperature (TMP)

This study aggregates open-source temperature raster data to obtain the annual average temperature for each town.

2.3 Research framework



Figuer 2: Research framework

This study is generally divided into three steps: data preparation, relationship modeling, and factor analysis (Figuer 2). In the data preparation stage, population big data and environmental data are collected. After data cleaning and unitization, human activity indicators and environmental factors at the town level are formed. In the pattern recognition stage, on one hand, the classic correlation coefficient calculation method in statistics is used to calculate the correlation between each pair of human activity indicators and environmental factors, and the impact of environmental factors on human activity is preliminarily explored. On the other hand, the K-means clustering algorithm is applied to divide all towns into three types based on three population characteristic factors. In the factor analysis stage, the importance of each environmental factor is first identified based on the random forest model results. Then, the ranking of importance is integrated with the correlation analysis results from the previous step to explore the possible causes of differences in human activity patterns.

2.4 Clustering methods for human activity patterns

Based on three human activity characteristic factors, this study uses the K-means method to summarize and identify the human activity patterns of each town. The algorithm includes the following steps.

(1) Initialization

Randomly select data points as the initial cluster centers.

(2) Assignment

For each data point in the dataset, calculate its distance to each cluster center and assign the data point to the cluster represented by the nearest cluster center. The distance metric is usually Euclidean distance. For a data point and cluster center c_k, the Euclidean distance formula is:

$$d(x, c_k) = \sqrt{\sum_{i=1}^{n} \left(x_i - c_{k_i}\right)^2} \tag{8}$$

(3) Update

Recalculate the center of each cluster. The cluster center is the mean of all data points within the cluster. The update formula for the j-th feature of the cluster center in cluster is:

$$c_{kj} = \frac{1}{|S_k|} \sum_{x \in S_k} x_j \tag{9}$$

Where is the set of data points in cluster k, is the number of data points in cluster k, and is the j-th feature value of data point in cluster k.

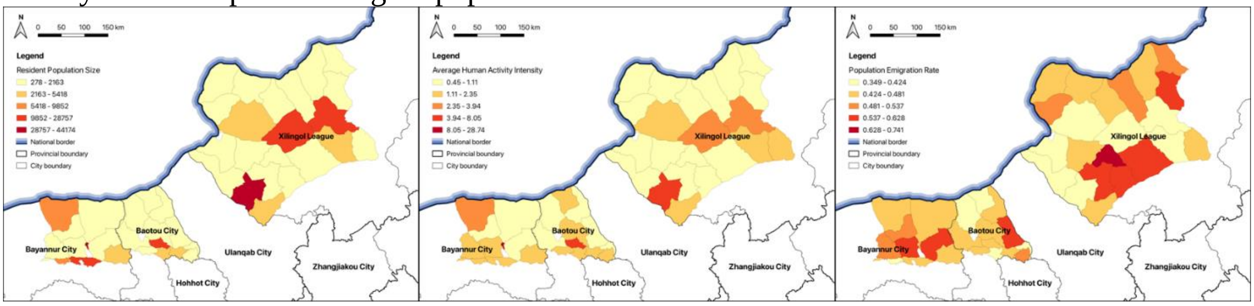
(4) Iteration

Repeat the assignment and update phases until the cluster centers no longer change, or the preset number of iterations is reached.

3. Results

3.1 Human activity characteristics

The distribution of human activity factors within the study area (Figure 3) exhibits three basic characteristics. First, the value distribution follows a "long tail" pattern, meaning that a small number of towns have high values, while most towns have values in the middle to lower range. Second, there is a certain degree of spatial autocorrelation. Third, there is a correlation between the spatial distribution of resident population and human activity intensity, meaning that towns with higher human activity intensity generally have a larger resident population, while towns with lower activity intensity tend to experience higher population outflow rates.



a) Resident Population. b) Average Human Activity Intensity. c) Population Emigration Rate

Figure 3: Distribution of Human Activities.

3.2 Human activity patterns

Based on the clustering results, the 45 towns within the study area are divided into three categories (Figure 4). Among them, the number of towns in the growth stage is relatively small, with the majority facing stagnation or decline.

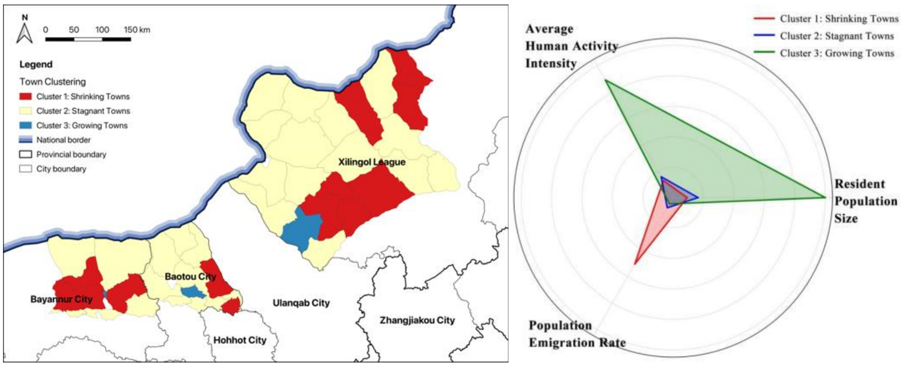


Figure 4: a) Spatial Distribution of Town Clustering. b) Human Activity Patterns in Three Types of Towns

(1) Cluster 1: Shrinking Towns

There are 13 towns in this category. The human activity characteristics of these towns are very low resident population and activity intensity, along with a high population outflow rate, indicating that these towns are experiencing continuous population loss. On the one hand, the intensity of production and daily activities will continue to decrease, which will reduce the impact on the fragile ecological environment; on the other hand, these towns may also face problems such as declining economic vitality, reduced public service levels, and village hollowing. Therefore, it is necessary to guide population urbanization effectively while ensuring basic public services and the quality of the living environment.

(2) Cluster 2: Stagnant Towns

There are 29 towns in this category. The human activity characteristics of these towns are low resident population, low activity intensity, and low population outflow rate, indicating that the population size in most towns within the study area is relatively stable, with low economic and social activity intensity and slow development. The population dynamics in these towns are relatively small, mainly due to a certain net population outflow, with natural growth and mechanical growth offsetting each other.

(3) Cluster 3: Growing Towns

This is the least common category, with only three towns: Saihantala Town in Xilin Gol League, Bailingmiao Town in Baotou, and Hailutu Town in Bayannur. The resident population and activity intensity are relatively high, and the population outflow rate is relatively low, indicating that these towns are in the growth stage. Growing towns are generally the central towns in their regions, with stronger industrial and population attraction.

3.3 Environmental factor differences

Based on the clustering results, this study further analyzes the differences in the distribution of various economic, social, and environmental factors across different clusters (Fig.5). The results show that the environmental factors corresponding to different human activity patterns generally exist at different levels of development in the geographical units.

(1) Per Capita Gross Domestic Product (GDP)

From the perspective of value distribution, the per capita GDP level is highest in shrinkage-type towns, followed by stagnation-type towns, while growth-type towns have the lowest per capita GDP with the smallest internal disparity. The possible reason for this phenomenon is that in shrinkage-type towns, the scale of agricultural and pastoral operations is relatively higher, and with a smaller population base, the per capita GDP is higher. In contrast, growth-type towns have fewer per capita production factors, resulting in a lower per capita GDP.

Previous studies have shown that residents in towns with higher per capita GDP have stronger migration ability and willingness, enabling them to move to central cities with better living conditions and public service levels. This view is also validated in this study, where residents in shrinkage-type towns with higher per capita GDP tend to have stronger migration abilities and are more likely to move to central cities with better living conditions and public services. Conversely, in towns with lower per capita income, residents are more likely to stay and engage in agricultural and pastoral work, with weaker migration capabilities.

(2) Geographic location

The two factors of location conditions and transportation remoteness exhibit similar value distribution levels across the three clusters. From the perspective of transportation remoteness, the median values for shrinkage-type and stagnation-type towns are quite close. However, compared to most shrinkage-type towns, more stagnation-type towns have higher transportation remoteness values.This suggests that towns with poor transportation conditions tend to have relatively insufficient development momentum. In contrast, growth-type towns are generally closer to highway exits, making transportation more convenient compared to the other two types.

From a location condition perspective, stagnation-type towns are generally the furthest from central cities, while shrinkage-type and growth-type towns have similar location conditions. This is because proximity to central cities not only injects development momentum into towns but may also facilitate the urbanization migration of rural populations.

(3) Climate condition

The median annual average temperature of shrinkage-type and stagnation-type towns shows little difference, but the overall temperature variation in stagnation-type towns is greater than that in shrinkage-type towns. Growth-type towns have both a higher median and overall level of annual average temperature compared to the other two types. Given that the study area is relatively cold, this phenomenon may be attributed to the fact that a higher average annual temperature can improve the comfort of residents' living conditions.

(4) Public service level

The importance of medical facility service levels and educational facility service levels ranks fifth and sixth, respectively, with relatively weaker influence. Compared to educational facility service levels, the importance of medical facility service levels is stronger, which may be due to the fact that the majority of towns in the study area have an aging and shrinking population, resulting in a more urgent demand for medical facilities than for educational ones.

Regarding the distribution of factor values across the three types of towns, both healthcare and education levels show similar trends. Overall, the pattern is that shrinkage-type towns have the lowest levels, stagnation-type towns are in the middle, and growth-type towns have the highest. In shrinkage-type towns, both healthcare and education levels are the poorest, making it more difficult to access basic public services. This phenomenon has been verified in urban studies, where research shows that urban shrinkage significantly increases the difficulty of accessing public services, such as healthcare and education, for vulnerable groups. In stagnation-type towns, basic public service facilities have generally been provided according to rural community and urban life circle requirements, but in the future, attention should be given to avoiding the inefficiency or idleness of service facilities due to the smaller population. Growth-type towns generally have public service facilities with higher service capacity and construction standards, which not only meet the needs of local residents but also radiate to surrounding towns.

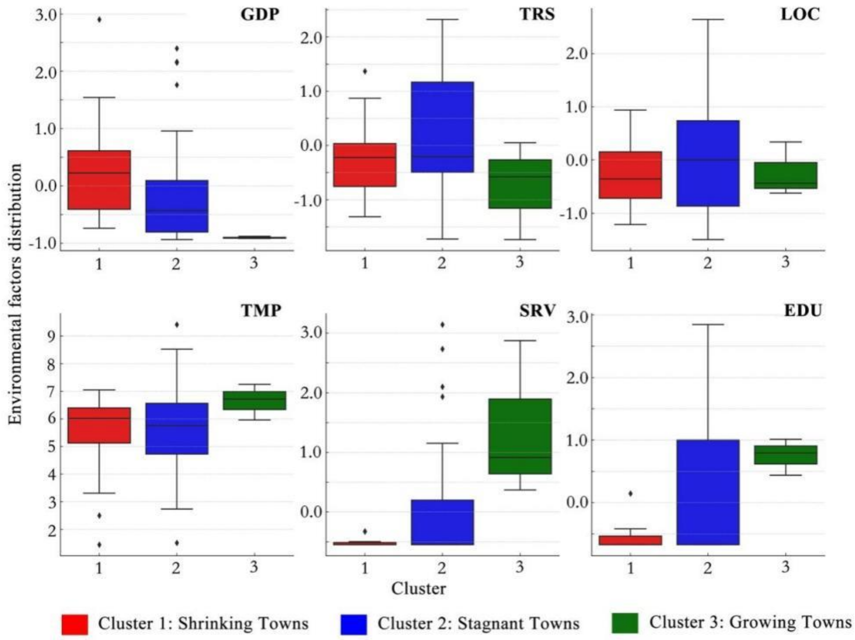


Figure 5: Distribution of Environmental Factors in Different Town Categories

4. Discussion

Based on the previous analysis, human activity governance policies for arid and semi-arid rural areas should generally follow three principles.

(1) Ordered population control

Considering that human activity is a significant cause of environmental degradation in arid and semi-arid areas, population management in such regions should be strengthened. For areas with the potential for migration, policies should be implemented to guide people towards surrounding regions with stronger resource and environmental carrying capacities and better economic and social development foundations. For areas that are not yet capable of supporting migration, population growth should be strictly controlled. This should be done in conjunction with ecological protection and restoration projects to effectively reduce environmental impact.

(2) Active industrial optimization

On the one hand, the role of industry as a "guiding force" for population urbanization should be leveraged. Regional industrial layout should be guided, with the development of related industries such as agricultural and livestock product processing and tourism, in addition to the dominant industries like agriculture and animal husbandry. This can effectively improve the living standards of residents in arid and semi-arid areas, creating conditions for voluntary urbanization. On the other hand, efforts should be made to control the environmental impact of production processes to avoid further environmental deterioration.

(3) Rational facility support

While guiding population migration, the configuration of basic public service facilities should be ensured to improve the quality of life for residents. Additionally, the regional transportation infrastructure should be optimized, making it a crucial tool for guiding population urbanization. This should be aligned with economic and industrial strategies to achieve precise and targeted guidance for rural population urbanization.

For the three types of towns with different human activity patterns, this study further proposes the following targeted planning recommendations.

(1) Shrinking towns

For shrinking towns, the trend of population shrinkage should be acknowledged, and efforts should be made to guide residents towards urbanization while ensuring their quality of life. Specific measures include:

Optimizing and upgrading the industrial structure: Actively develop environmentally friendly industries to strengthen residents' economic conditions.

Improving the balance of facility distribution: Considering that shrinking towns often have dispersed populations, it is essential to enhance the balance of facility distribution, ensuring that basic living needs are met without neglecting certain areas.

Maintaining the road system: Ensure basic public transportation services are available by maintaining and improving the road infrastructure.

Improving building quality: Enhance the capacity to resist harsh weather conditions, improving the overall quality of the living environment.

(2) Stagnant towns

Stagnant towns are the most numerous and widely distributed and thus the focus of management. Development strategies include:

Developing advantageous industries: Promote high-value-added, low-pollution industries.

Optimizing public service facilities: Appropriately scale public service facilities based on actual population needs, ensuring that facilities are built to meet these demands.

Improving road and transport infrastructure: Focus on improving connections between rural settlements and surrounding transportation networks, such as highways and regional traffic facilities, to enhance transportation conditions.

(3) Growing towns

Growing towns are the growth hubs in arid and semi-arid rural areas and should be leveraged as stepping stones for rural population urbanization. Specific measures include:

Developing labor-intensive industries: Utilize the advantages of local resources and labor force to promote agricultural product processing, textiles, and other industries to improve residents' income levels and strengthen migration capacity.

Consolidating and enhancing public service functions: Strengthen the role of growing towns as regional public service hubs, improving service quality while also ensuring sufficient capacity in public service facilities. The development of public service systems should cater to both urban and rural life circles.

Enhancing road infrastructure: Further develop the transportation system and rural public transport,ensuring the provision of infrastructure to support economic and social development and population urbanization.

5.Conclusion

This study integrates multiple sources of big data and various machine learning methods to accurately characterize the human activity features in the arid and semi-arid rural areas of Inner Mongolia Autonomous Region. It explores the differences in environmental factors corresponding to different human activity patterns and subsequently proposes targeted governance recommendations for each pattern. Compared with existing research, this study also focuses on the intrinsic relationship between human activities and the socio-economic and natural environment, yielding several consistent conclusions. At the same time, by adopting emerging data and a systematic perspective, this study provides more in-depth and comprehensive research findings.

Specifically, the innovations of this study can be summarized in three aspects. First, a fine-grained and multidimensional characterization of human activity in the study area was conducted. Second, based on population size, activity intensity, and population migration, the study classified the towns into three different human activity patterns, thereby portraying the urbanization patterns in ecologically fragile regions. Third, the study further identified the levels of environmental factors corresponding to different human activity patterns and, based on this, proposed practical and valuable recommendations. The conclusions of this study contribute to exploring the path for ecological civilization construction in arid and semi-arid regions.

References

[1] Zhang, T.; Chen, Y. The Effects of Landscape Change on Habitat Quality in Arid Desert Areas Based on Future Scenarios: Tarim River Basin as a Case Study. Front. Plant Sci. 2022, 13. https://doi.org/10.3389/fpls.2022.1031859.

[2] Li X.; He X.; Yang G.; Liu H.; Long A.; Chen F.; Liu B.; Gu X. Land use/cover and landscape pattern changes in Manas River Basin based on remote sensing. ijabe 2020, 13 (5), 141–152. https://doi.org/10.25165/j.ijabe. 20201305. 4783.

[3] Li, C.; Wang, Y.; Wu, X.; Cao, H.; Li, W.; Wu, T. Reducing Human Activity Promotes Environmental Restoration in Arid and Semi-Arid Regions: A Case Study in Northwest China. Science of the Total Environment 2021, 768, 144525. https://doi.org/10.1016/j.scitotenv.2020.144525.

[4] Li, C.; Cai, G.; Du, M. Big Data Supported the Identification of Urban Land Efficiency in Eurasia by Indicator SDG 11.3.1. ISPRS International Journal of Geo-Information 2021, 10 (2), 64. https://doi.org/10.3390/ijgi10020064.

[5] Wu, J.; Yang, X. Assessment of Ecological Carrying Capacity in Xilingol League Based on Three-Dimensional Ecological Footprint Model. Sustainability 2025, 17 (1), 128. https://doi.org/10.3390/su17010128.

[6] Hamerly, G.; Elkan, C. Learning the k in K-Means. In Advances in Neural Information Processing Systems; MIT Press, 2003; Vol. 16.

[7] Wang, K.; Deng, L.; Shangguan, Z.; Chen, Y.; Lin, X. Sustainability of Eco-Environment in Semi-Arid Regions: Lessons from the Chinese Loess Plateau. Environmental Science & Policy 2021, 125, 126–134. https://doi.org/10.1016/j.envsci.2021.08.025.

[8] McGrail, M. R. Spatial Accessibility of Primary Health Care Utilising the Two Step Floating Catchment Area Method: An Assessment of Recent Improvements. Int J Health Geogr 2012, 11 (1), 50. https://doi.org/10.1186/1476-072X-11-50.