Identifying spatio-temporal patterns of the young population changes in mainland China from 2010 to 2020

2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) | 979-8-3503-0351-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/Agro-Geoinformatics59224.2023.10233565

Nianshuo Wen
School of Remote Sensing and Information Engineering
Wuhan University
Wuhan, China
wennianshuo@163.com

Xingran Luo
School of Remote Sensing and Information Engineering
Wuhan University
Wuhan, China
lxr3157433608@163.com

Binbin Lu*
School of Remote Sensing and Information Engineering
Wuhan University
Wuhan, China
binbinlu@whu.edu.cn

***Abstract*—The negative population growth within China has become a new trend due to long-term low fertilities and aging conditions. The shrink of the young population could lead to a series of negative impacts on the future population size, age structure, and even social structure. Therefore, it is important to explore the spatio-temporal changes of the young population. In this study, we adopted the 6th and 7th census data collected in 2010 and 2020, respectively, to analyze the spatio-temporal patterns of young population changes. On the one hand, we explore the changes of the young population and its proportion via exploratory data analysis techniques. On the other hand, we identified the spatial pattern of young population changes through autocorrelation analysis. Besides, we performed the Geodetector tool to explore the spatially heterogeneous features in these changes. Results show that: 1) the exploratory analysis reveals significant spatial heterogeneity of the young population and its proportion; 2) the change of young population and its proportion presented positive autocorrelation, and a trend of migration from northwest to southeast, i.e. migration from the interior area to the coastal areas; 3) socio-economic factors are differently associated with the change of the proportion of young population, among which factors concerning working-age population, GDP, and higher education present stronger explanatory power compared with the other factors.**

Keywords—Young population, pattern analysis, spatial autocorrelation, Geodetector, spatial heterogeneity

I. INTRODUCTION

The young population constitutes a dynamic and pivotal segment of society, embodying the promise of a nation's future. Gaining a comprehensive understanding of their distribution and spatio-temporal patterns within a country holds immense value for comprehending demographic trends, social development, and urban planning.

China has witnessed rapid economic growth in recent decades. However, alongside urbanization and modernization, the nation is confronted with a pressing population situation. According to the National Bureau of Statistics, China's population in 2022 experienced the first instance of negative growth by 850,000 since 1962. The population decline coincides with a persistently low fertility rate. The 7th national census data indicates that China's total fertility rate in 2020 stood at a mere 1.3 [1] .

While it has been posited that declining fertility naturally accompany industrialization and urbanization, China faces a significant challenge characterized by a substantial decrease in fertility rates amid incomplete industrialization and modernization. In fact, data from the World Bank reveals that, since the mid-1990s China's total fertility rate has consistently remained lower than that of countries within the same income category, and notably lower than that of high-income countries overall [2]. Prolonged low fertility will give rise to a host of adverse effects on future population size, age structure, and even social structure [3]. Consequently, investigating the distribution of China's young population assumes significant importance.

This study aims to identify the spatio-temporal distribution patterns of the young population within the region. Leveraging spatial autocorrelation and Geodetector methods, we seek to explore and analyze the heterogeneity of the young population and its underlying influencing factors.

II. DATA

In this study, we utilized the census data from the 6th and 7th national population censuses conducted in 2010 and 2020, respectively. These census data were obtained from the official statistical bulletins published by the statistical bureaus at the county level. Additionally, we collected economic and social data of each county corresponding to the same years. The socio-economic data were sourced from the statistical yearbooks published by local government authorities for the years 2010 and 2020.

III. METHOD

A. Moran's I

Spatial autocorrelation reflects the correlation of an indicator in space [4]. Moran's I is frequently used for measuring spatial autocorrelation, which is able to represent the linear correlation between the observed values of an indicator and its spatial lag terms. Global index is calculated through the equation below:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij}\,(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij}} \quad (1)$$

where I is the index; x_i is the observed value of the indicator x of the ith sample; $\bar{x}$ is the mean of x; and w_{ij} is the weight between the ith and jth samples [5].

With variance normalized, the value of I ranges from - 1 to 1. If $I > 0$, it indicates that the indicator has a positive global autocorrelation; If $I < 0$, it indicates that the autocorrelation is negative; If $I = 0$, it indicates that there is no autocorrelation [6].

Moran's I presents the global autocorrelation in the entire region. The autocorrelations in local areas, however, are various. To detect local autocorrelation and identify clusters

979-8-3503-0351-3/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Makerere University Library. Downloaded on August 05,2026 at 15:17:06 UTC from IEEE Xplore. Restrictions apply.

and outliers of young population, Local Moran's I was used in this study. The local index is calculated as follows:

$$I_i = \frac{z_i - \bar{z}}{\sigma^2} \sum_{j=1, j\neq i}^{n} w_{ij}\left(z_j - \bar{z}\right) \quad (2)$$

where I_i is the local Moran's index of the ith sample ($i = 1,2,\ldots,n$), z_i and z_j are the values in the ith and jth samples respectively ($j \neq i$), n is the size of the samples, $\bar{z}$ is the mean of the variables in the total region, σ^2 is the variance of the variables in the total region, w_{ij} is the weight between the ith and jth samples [7].

B. Geodetector

Geodetector is a set of statistical methods which detects spatial stratified heterogeneity (http://geodetector.org/) and reveals the driving forces of spatial differentiation of the dependent variable. Its basic idea is that an independent variable prominently influences the dependent variable if the spatial pattern of the independent variable is similar to that of the dependent variable [8].In this study, three methods of Geodetector were used.

Factor detector calculates power determinants of independent variables. The determinants are described through value q, which is expressed as follows:

$$\mathrm{q} = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST} \quad (3)$$

where $h = 1,2,\ldots,L$, L is the number of strata of each independent variable, N is the total number of samples, N_h is the number of samples in stratum h, σ^2 and σ_h^2 are the variance in the total region and that in stratum h respectively, SST is the total variance, SSW is the sum of variance of each stratum. The value $q \in [0,1]$. A higher q value indicates a stronger association of the independent variable with the dependent variable and a higher explanatory power accordingly [9].

Interaction detector calculates the power determinant of the interaction between two independent variables, and detects whether the explanatory power of the two variables increases or decreases when the they come into an interaction. Table I presents the types of interactions of two independent variables.

Ecological detector, based on the F-value test, explores whether the influence of interactions is a significant difference. The F-value test is expressed as follows:

$$F = \frac{N_{X_1}\left(N_{X_2} - 1\right)SSW_{X_1}}{N_{X_2}\left(N_{X_1} - 1\right)SSW_{X_2}} \quad (4)$$

$$SSW_{X_1} = \sum_{h=1}^{L_1} N_h \sigma_h^2, SSW_{X_2} = \sum_{h=1}^{L_2} N_h \sigma_h^2 \quad (5)$$

where N_{X_1} and N_{X_2} are the sample sizes of independent variables X_1 and X_2, respectively; SSW_{X_1} and SSW_{X_2} are the quadratic sum of the stratum formed by X_1 and X_2, L_1 and L_2 represent the sample number of X_1 and X_2. The null hypothesis H_0 is that $SSW_{X_1} = SSW_{X_2}$. When H_0 is rejected at the level of significance of α, it is considered that there is a significant difference in the effect of X_1 and X_2 on the spatial distribution of the dependent variable [9].

In this study, the autocorrelation analysis and mapping were performed by ArcGIS 10.8. Spatial heterogeneity analysis was performed by R using the package of GD. The original data underwent classification, utilizing the optimal approach determined by the "optidisc" function from the GD package.

IV. Result and Analysis

A. Exploratory Analysis

We conducted an exploratory analysis to examine the changes in the young population and its proportion from 2010 to 2020 at the county level (Fig. 1). The findings revealed notable heterogeneity in the distribution of the young population. As depicted in Fig. 1 (a), there was a significant increase in the young population in two types of counties: a) counties with a traditionally high local young population, such as Henan and Shandong Provinces, and b) regional central cities, which were economically developed. Conversely, counties with a significant decrease in the young population were concentrated in a) Northeast China, b) middle Inner Mongolia, c) Southwest China, and d) Jiangxi and Anhui Provinces, which were economically disadvantaged. Overall, this distribution suggests a clear siphon effect on the migration of the young population across the nation, characterized by a movement from economically disadvantaged prefectures to economically developed ones. Fig. 1 (b) illustrates the spatial distribution of the change in the proportion of the young population. In general, the distribution of the proportion mirrored that of the overall population distribution.

To examine the joint pattern of young population and its proportion, we classified all samples into distinct types based on the directional changes in both the young population and its proportion within each county (Fig. 2). The analysis revealed that counties could be grouped into two primary types: those where both the young population and its proportion increased, and those where both decreased. Among the 2,248 counties with available data, approximately 42.9% (965 counties) belonged to the latter type, exhibiting simultaneous decreases in both the young population and its proportion. These counties were geographically dispersed throughout China, with relatively fewer occurrences in Henan, Shandong, the southern coastal provinces, and the Qinghai-Tibet region. These observations highlight the polarization trend in the distribution of the young population

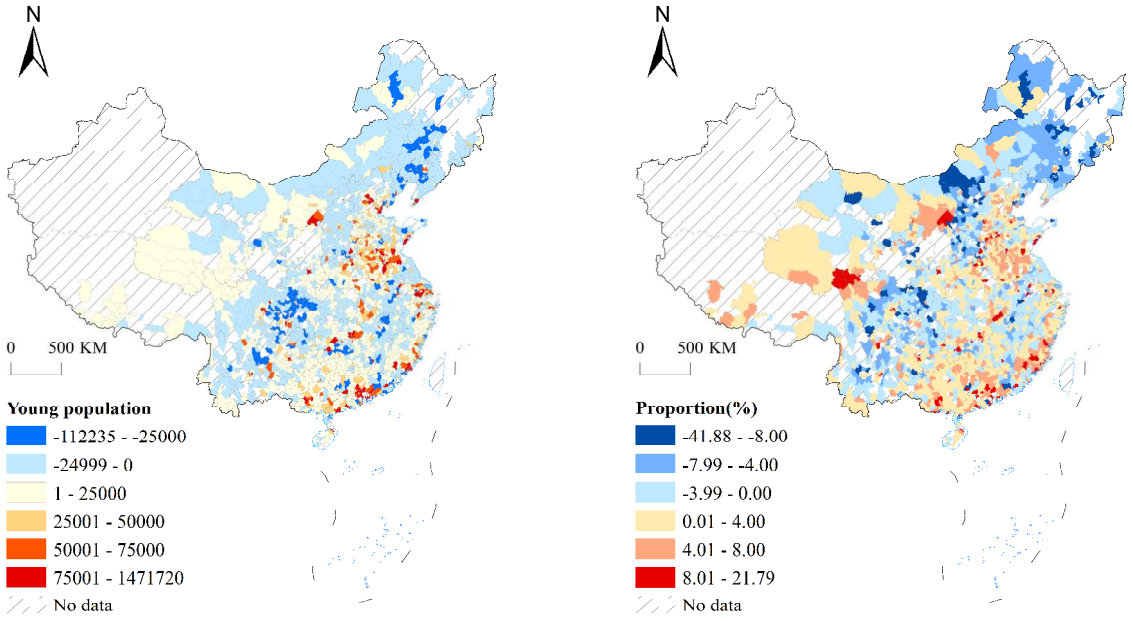
TABLE I Types of Interaction Defined in Geodetector

Description	Interaction
$q(X_1 \cap X_2) < \mathrm{Min}(q(X_1), q(X_2))$	Weakened, nonlinear
$\mathrm{Min}(q(X_1), q(X_2)) < q(X_1 \cap X_2) < \mathrm{Max}(q(X_1), q(X_2))$	Weakened, single factor nonlinear
$q(X_1 \cap X_2) > \mathrm{Max}(q(X_1), q(X_2))$	Enhanced, double factors
$q(X_1 \cap X_2) = q(X_1) + q(X_2)$	Independent
$q(X_1 \cap X_2) > q(X_1) + q(X_2)$	Enhanced, nonlinear

TABLE II Global Autocorrelation Analysis Result

Indicators	Moran's I	Z score	P value
Population	0.133	27.545	<0.01
Proportion	0.231	42.990	<0.01

Authorized licensed use limited to: Makerere University Library. Downloaded on August 05,2026 at 15:17:06 UTC from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 1 Spatial distribution of the change of young population (a) and its proportion (b) of each county from 2010 to 2020.

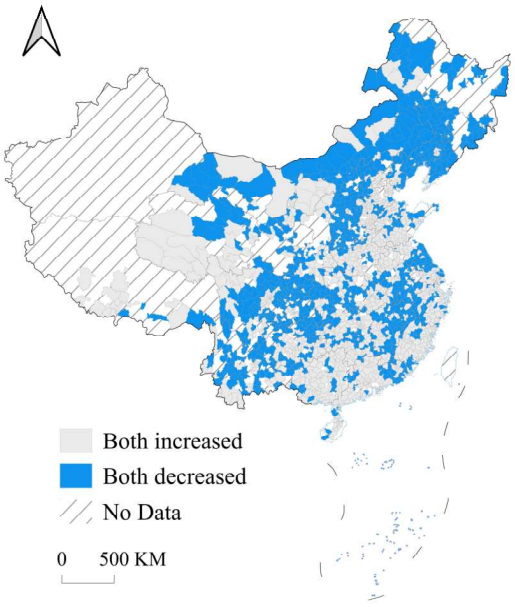


Fig. 2 Counties divided into 2 types according to their change of young population and its proportion.

TABLE III EXPLANATORY POWER OF SOCIO-ECONOMIC FACTORS

Variable	q-statistics	p-value
The change in GDP (V1)	0.0881	0.0000
The change in GDP per capita (V2)	0.0207	0.0004
The proportion of added value of secondary industry GDP in 2010 (V3)	0.0445	0.0000
The change in the proportion of public financial expenditure to GDP (V4)	0.0588	0.0000
The proportion of the population with higher education to local total population (V5)	0.0637	0.0000
The change in the working-age population (V6)	0.1245	0.0000
The change in the proportion of the working-age population to the local total population (V7)	0.0535	0.0000
The urbanization rate in 2010 (V8)	0.0258	0.0000

and underscore the nationwide nature of negative population growth among the youth.

B. Spatial Pattern Identification

We calculated the Global Moran's I indices for both the young population and its proportion to investigate their global autocorrelation, as outlined in Table II. The resulting Moran's I index for the young population was determined to be 0.133, while the index for the proportion stood at 0.200. Significantly, both indices demonstrated statistical significance, indicating the presence of a weak positive autocorrelation.

To further explore the autocorrelation at the county level, we calculated the Local Moran's I indices, and the results are illustrated in Fig. 3. Fig. 3 (a) specifically showcases the clusters and outliers of the young population. The analysis revealed that high-high clusters were primarily concentrated in: a) certain regions of Henan, Shandong, Anhui, and Jiangsu provinces, b) western Hunan, c) national megacities such as Beijing, Shanghai, Guangzhou, and Shenzhen, along with their surrounding areas, and d) coastal counties and their surrounding regions in Fujian Province. Conversely, the low-low clusters were concentrated in: a) Northeast China, b) middle Inner Mongolia and Shanxi, c) Southwest China, and d) southern Anhui and Northeast Jiangxi. Fig. 3 (b) illustrates the clusters and outliers of the proportion of the young population. Overall, the distribution of proportion clusters aligns closely with that of the population clusters, indicating a similar spatial pattern with expanding clusters. However, there was a distinct characteristic observed in the proportion clusters compared to the population clusters. Notably, certain

Authorized licensed use limited to: Makerere University Library. Downloaded on August 05,2026 at 15:17:06 UTC from IEEE Xplore. Restrictions apply.

high-high clusters emerge in Western China, particularly in the Qinghai-Tibet Area. This finding suggests that although the young population in the Qinghai-Tibet Area was smaller compared to other contemporaneous regions, the local proportion of young population was significantly higher. This observation is consistent with the findings depicted in Fig. 2, further reinforcing the regional variation in the proportion of young population within China.

In general, high-high clusters predominantly occurred in coastal areas, whereas low-low clusters were observed in the interior regions. This spatial pattern suggests that the migration of the young population is primarily from the interior areas towards the coastal regions. The distribution of clusters aligns well with the findings obtained from the exploratory analysis.

C. Spatially Heterogeneous Features Analysis

Based on the socio-economic data, we constructed eight variables for the application of the Geodetector method. These variables can be categorized into two types: economic factors and social factors. The economic factors encompassed four variables, i.e. the change in Gross Domestic Product (GDP) (V1), the change in GDP per capita (V2), the proportion of added value of the secondary industry GDP in 2010 (V3), and the change in the proportion of public financial expenditure to GDP (V4). On the other hand, the social factors consisted of four variables, i.e. the change in the proportion of the population with higher education to the local total population (V5), the change in the working-age population (V6), the change in the proportion of the working-age population to the local total population (V7), and the urbanization rate in 2010 (V8). These variables were selected to capture key dimensions of both economic and social factors, facilitating a comprehensive analysis of their influences on the spatio-temporal distribution of the young population.

1) Factor Detector

Using the factor detector method of Geodetector, we determined the q-value for each socio-economic factor with respect to the change in the proportion of the young population from 2010 to 2020 (Table III). Generally, these factors demonstrated weak explanatory power in relation to the observed changes. The result aligns with that obtained by Li *et al.* [10]. Among the factors examined, V6 (change in the working-age population) exhibited the highest q-value of 0.1245, indicating that the working-age population is more influential on the proportion of the young population. Additionally, the q-values for V1 (change in GDP) and V5 (proportion of population with higher education) exceeded 0.06, suggesting that these factors also exerted noticeable effects on the proportion of the young population.

TABLE IV THE RESULT OF ECOLOGICAL DETECTOR

	V1	V2	V3	V4	V5	V6	V7	V8
V1								
V2	Y[a]							
V3	Y	Y						
V4	Y	Y	Y					
V5	Y	Y	Y	Y				
V6	Y	Y	Y	Y	Y			
V7	Y	Y	Y	Y	Y	Y		
V8	Y	Y	Y	Y	Y	Y	Y	

[a] The category Y means a significant difference, while N means insignificance.

2) Interaction Detector

While analyzing the explanatory power of individual factors is valuable, it is crucial to recognize that the spatial pattern of the change in the proportion of the young population is jointly influenced by the interactions of multiple factors [8]. Table IV presents the results of interaction detection. All the q-value of interactions of double factors were significantly higher than that of a single factor, which were manifested as double-factor or nonlinear enhancement. Among the nonlinear interaction, the q-value of the interaction of $V7 \cap V6$ resulted to the highest, which was 0.3031. The interaction of $V6 \cap V1$ ranked the second, and able to explain 0.2587 of the spatial-pattern of the proportion. The interactions of $V6 \cap V5$, $V6 \cap V2$, and $V2 \cap V1$ exceeded 0.2. The result indicates that the proportion was more possible to be influenced by interactions among multiple factors rather than a single or several dominant factors. On the other hand, it reveals a stronger synergistic influence between working-age population with factors concerning economy or education than the other factors.

3) Ecological Detector

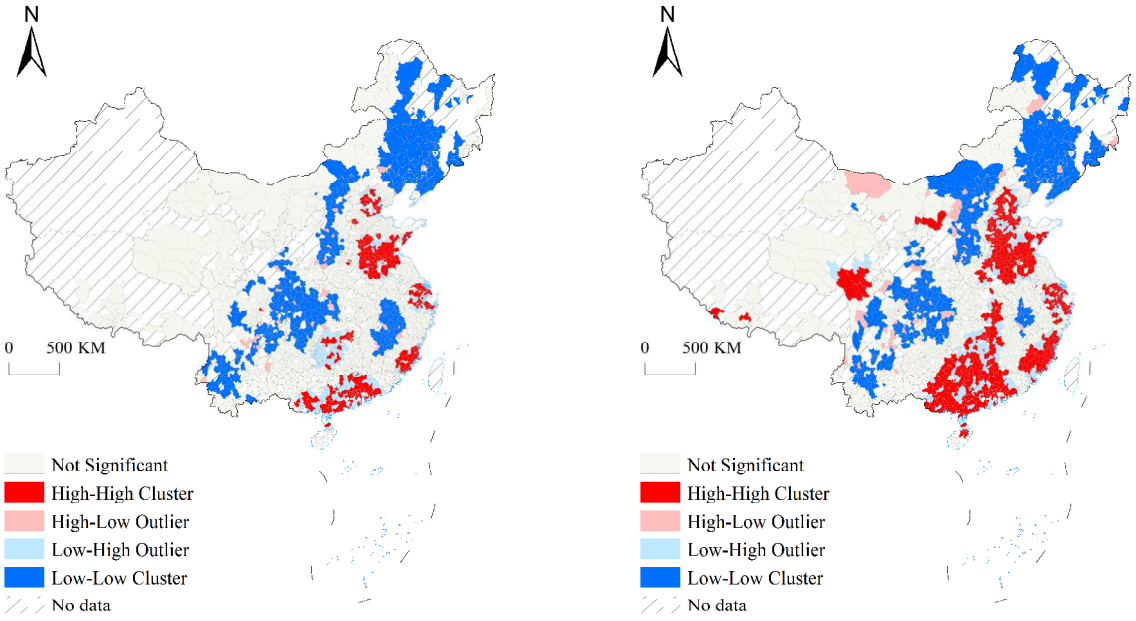
To determine the significance of the interactions, the ecological detector module was utilized. The results were

TABLE V RESULT OF INTERACTION DETECTION

	V1	V2	V3	V4	V5	V6	V7	V8
V1								
V2	0.2034							
V3	0.1239[a]	0.1241						
V4	0.1503	0.1184	0.1000					
V5	0.1717	0.1244	0.1562	0.1388				
V6	0.2587	0.2076	0.1971	0.1671	0.2481			
V7	0.1842	0.1031	0.0906	0.1413	0.1101	0.3031		
V8	0.1448	0.0678	0.0732	0.1162	0.1545	0.1586	0.0837	

[a] The interactions of $V3 \cap V1$, $V4 \cap V3$, $V6 \cap V4$, $V7 \cap V3$, and $V7 \cap V5$ were double-factor enhancement. The others were nonlinear enhancement.

Authorized licensed use limited to: Makerere University Library. Downloaded on August 05,2026 at 15:17:06 UTC from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 3 Local autocorrelation of young population (a) and its proportion (b).

classified into two categories Y and N (Table V). The result reveals that all interactions demonstrated significant differences in terms of their influence on the observed patterns.

V. Conclusion

In this study, we identified the spatio-temporal patterns of young population in mainland China. The main conclusions drawn in the study were as follows:

1) Exploratory analysis reveals a significant spatial heterogeneity in the distribution of the young population and its proportion, which exhibited similarities to a well-known geographic phenomenon, i.e. siphon effect around regional megacities.

2) The analysis of changes in young population and its proportion from 2010 to 2020 indicated a weak positive global autocorrelation. Local autocorrelation analysis presents a migration trend from the northwest to the southeast, illustrating a movement of the young population from the interior areas towards the coastal areas.

3) The socio-economic factors exhibited a weak association with the changes in the proportion of young population. Among these factors, those related to working-age population, GDP, and higher education showed relatively stronger explanatory power. Moreover, the interactions between socio-economic factors were significant influences and had a higher explanatory power than any single factor alone. Notably, interactions involving factors related to working-age population and factors concerning GDP or higher education demonstrated relatively higher explanatory power.

ACKNOWLEDGMENT

This study is jointly funded by the National Natural Science Foundation of China (NO. 42071368, No. U2033216) and Hubei Provincial Key Research and Development Program (2021BAA185)

References

[1] M. Chi, S. Xie, “Analysis on the Causes of Fertility Rebound in Europe since the 21st Century and Its Enlightenment to China,” Journal Article, vol. 44(4), 2022, pp. 80-90.

[2] Z. Zhu, G. Zhao, “Higher Education and Chinese Women's Fertility,” Population Journal, vol. 44(2), 2022, pp. 16-31.

[3] X. Liu, G. Wang, “Re-analysis of Chinese Fertility Level and the Related Social Development Indexes in the New Period,” Academic Exploration, 2022(7), pp. 67-68.

[4] X. Wang, X. Gao, “The spatial distribution and evolution characteristics of the Muslim population in China,” World Regional Studies, 2021, vol. 30(6), pp. 1297-1307.

[5] Y. Zhang, X. Wang, “A Study on Population Spatial Distribution in China from the Perspective of Urban Agglomeration,” Population & Economics, 2020(3), pp. 1-13.

[6] Z. Gao, G. Chang, X. Wang, J. Lyu. “Changes of Population Spatial Distribution and Its Influencing Factors in Shiyang River Basin in Recent 30 Years,” Journal of Fujian Normal University(Natural Science Edition), 2021, vol. 37(1), pp. 80-87.

[7] G. Tepanosyan, L. Sahakyan, C. Zhao, A. Saghatelyan, “The application of Local Moran's I to identify spatial clusters and hot spots of Pb, Mo and Ti in urban soils of Yerevan,” Applied Geochemistry: Journal of the International Association of Geochemistry and Cosmochemistry, 2019, vol. 104, pp. 116-123.

[8] J. Shi, R. Jin, W. Zhu. “Quantification of Effects of Natural Geographical Factors and Landscape Patterns on Non-point Source Pollution in Watershed Based on Geodetector: Burhatong River Basin, Northeast China as An Example,” Chinese Geographical Science, 2022, vol. 32(4), pp. 707-723.

[9] W. Cheng, H. Xi, S. Celestin. “Application of geodetector in sensitivity analysis of reference crop evapotranspiration spatial changes in Northwest China,” Sciences in Cold and Arid Regions, 2021, vol. 13(4), pp. 314-325.

[10] J. Li, D. Lu, C. Xu, Y. Li, M. Chen. “Spatial heterogeneity and its changes of population on the two sides of Hu Line,” Acta Geographica Sinica, 2017, vol. 72(1), pp. 148-160.

Authorized licensed use limited to: Makerere University Library. Downloaded on August 05,2026 at 15:17:06 UTC from IEEE Xplore. Restrictions apply.