scientific reports

OPEN

A study on the spatial distribution and driving factors of traditional villages–a case study of the Beijing-Tianjin-Hebei region in China

Peng Peng[1✉], Yuechen Zhang[1,2], Jiajie Zhao[1,2], Yan Rong[1,2], Limin Zhao[1,2] & Jinyu Wu[1,2]

Traditional villages encompass abundant cultural resources, possessing significant humanistic and socio-economic value. This study used ArcGIS 10.8 and GeoDa software, took "Chinese Traditional Villages" as the research object, and analyzed their spatial autocorrelation and spatial heterogeneity through methods such as nearest neighbor analysis, Moran's Index, and kernel density estimation. Additionally, by combining techniques like Geodetector, ordinary least squares (OLS), and geographically weighted regression (GWR), the study selected the 10 most correlated driving factors and examined the drivers of their spatial distribution characteristics. The results are as follows: (1) The spatial distribution of villages exhibits significant clustering characteristics. (2) Rainfall and GDP per capita are the two important explanatory variables influencing the spatial distribution of villages. (3) The clustering phenomenon of villages is positively correlated with temperature, rainfall, and tourism resources, while negatively correlated with population density and overall economic level.

Keywords Traditional villages, Driving factors, Spatial heterogeneity, Beijing-Tianjin-Hebei region

"Chinese Traditional Villages"[1] not only preserve rich historical information and unique cultural landscapes but also embody unique folk customs and historical evolution. Furthermore, these villages serve as material carriers of historical memory, emotional bonds of cultural identity, and key cornerstones of sustainable rural development. The protection of these villages is closely linked to the United Nations Sustainable Development Goals (SDG 11)[2], "Sustainable Cities and Communities." By maintaining the cultural diversity, ecological integrity, and social cohesion of traditional villages, this effort can contribute a Chinese solution to building inclusive, safe, and resilient human settlements worldwide.

The rapid economic development and accelerated urbanization[3] have led to an increasingly severe imbalance between urban and rural development[4]. A large number of traditional villages now face a systemic crisis, characterized by population decline[5], land fragmentation, and ecological degradation. Consequently, their numbers have sharply decreased, placing them on the brink of extinction. In this context, it is highly significant to explore ways to integrate the cultural preservation of traditional villages with sustainable development goals. This research aims to promote the sustainable development of rural communities while protecting the historical context and offering insights that could facilitate the achievement of global sustainable development goals.

In the context of globalization and urbanization, villages worldwide face similar challenges. For example, villages in the Italian Alps have experienced excessive population loss and changes in spatial structure[6], while rural India has suffered severe damage due to natural and human factors[7]. By comparing these international cases, researchers can not only better understand and address the problems in the Beijing-Tianjin-Hebei region but also verify and enrich their research methods. This analysis demonstrates the global significance of traditional village protection in the Beijing-Tianjin-Hebei region and emphasizes the feasibility of adopting similar methods in other areas. It provides an important reference for the protection and development of villages in developing countries.

At present, the study of traditional villages has shown a trend of multidisciplinary integration, integrating perspectives from geography[8], history[9], ecology[10], and others. The research methodology is now different from the traditional one that focused on theoretical studies[11,12], tourism resources[13], and historical artifacts[14]. With the improvement of research methodologies and techniques, current scholars are more inclined to

[1]School of Architecture & Art, Shijiazhuang Tiedao University, Shija-Zhuang 050043, Hebei, China. [2]Yuechen Zhang, Jiajie Zhao, Yan Rong, Limin Zhao and Jinyu Wu contributed equally to this work. ✉email: pengpeng20240325@163.com

adopt a macroscopic perspective to study traditional villages. They use advanced methods such as Geographic Information System (GIS) technology[15], spatial data analysis (in some cases performed with GeoDa Software)[16], Space Syntax[9], and GWR models[17] to deeply analyze the spatial distribution and evolution of traditional villages. This macro perspective echoes the "core-periphery" theory, which explores how resources and development opportunities are unevenly distributed among different regions, forming core and peripheral areas of development. This distribution thus affects the survival and development of traditional villages.

However, existing research still has significant limitations. First, in terms of research scale, most results rely excessively on a single division method, such as administrative divisions, topography, geomorphology[18], or watersheds[19]. It is common to take national[20], provincial[21], and city (county) levels[22] or specific terrain areas[23] as research objects but to ignore the integrity of regions with special geographical and cultural characteristics, such as the Beijing-Tianjin-Hebei area. Affected by the natural geographical barrier of the Yanshan-Taihang Mountains, the distribution and development of traditional villages in this region exhibit unique regional linkages, and existing research has failed to fully consider this key factor. Second, at the level of influencing factor analysis, factors such as the natural environment, socio-economics[24,25], and historical culture[26] are often discussed separately, lacking systematic research on the interaction mechanisms of multi-dimensional factors. Furthermore, an effective multi-factor coupling analysis model has not yet been established.

This study focuses on 310 national traditional villages in the Beijing-Tianjin-Hebei region and uses ArcGIS technology to integrate 10 key driving factors, such as physical geography, historical culture, and social economy, to build a multidimensional spatial database. Research methods, including spatial autocorrelation analysis, OLS, GWR, and Geodetector models, are used to explore the driving mechanism of spatial differences in traditional villages. This aims to overcome the scale and factor analysis limitations of traditional research and provide more comprehensive and accurate theoretical support and practical bases for the protection and sustainable development of traditional villages.

This study not only provides a theoretical explanation for the spatial differences of traditional villages in the Beijing-Tianjin-Hebei region but also contributes to the specific goals of "Protect the World's Cultural and Natural Heritage" and "Inclusive and Sustainable Urbanization" in SDG 11 through differentiated protection strategies. It provides a scientific basis for policymaking in similar regions.

Study area and data
Study area
The year 2024 marks the 10th anniversary of the coordinated development strategy for Beijing, Tianjin, and Hebei being designated as a national priority. Over the past decade, China has continuously promoted this strategy. Recently, with the nation's heightened focus on traditional villages, research into these communities has attracted significant attention from experts and scholars. The traditional villages in the Beijing-Tianjin-Hebei region are geographically connected, maintain close interpersonal relationships, demonstrate regional cohesion, and share a common culture. These villages are valuable historical and cultural assets, developed over centuries, and they reflect the unique spiritual values of the Chinese people as well as a rich regional folk culture[27].

Historically, academic research has primarily concentrated on Hebei Province. However, solely analyzing Hebei does not adequately capture its regional characteristics. Studying Beijing, Tianjin, and Hebei as an integrated whole and using the Taihang and Yanshan mountains as natural geographical boundaries will deepen our understanding of the region's overall dynamics. Moreover, the divisions used in scientific research differ from administrative boundaries. It is not entirely appropriate to segment regions based solely on administrative criteria. A comprehensive consideration of multiple factors, including history, culture, and human geography, is essential.

The Beijing-Tianjin-Hebei region, situated in the North China Plain and bordered by the Yanshan Mountains to the north, the Taihang Mountains to the west, and Bohai Bay to the east, features diverse topography including plains, mountains, plateaus, and hills. The plains cover approximately 37% of the region's total area and act as the primary center for population and economic activities. Known as China's "Capital Circle," the region includes Beijing, Tianjin, and 11 prefecture-level cities in Hebei Province. Key areas such as the capital Beijing, the municipality of Tianjin, and the cities of Baoding and Langfang in Hebei Province are central to its coordinated development efforts. As of October 2024, the Ministry of Housing and Urban–Rural Development, along with other departments, has recognized 310 villages in the Beijing-Tianjin-Hebei region in six batches of "Chinese Traditional Villages."

Data processing
According to the list of 8,155 traditional villages released, there are 310 villages in this region, accounting for 3.8% of the total. These villages are distributed across 13 municipal administrative regions (Fig. 1). Table 1 shows the data sources used in the study.

Methodology
Research framework
This study collected the latitude and longitude data of 310 traditional villages in the Beijing-Tianjin-Hebei region and used ArcGIS 10.8 for data analysis. First, nearest neighbor analysis was applied to characterize spatial distribution patterns; second, the Gini coefficient and Lorenz curve were used to assess spatial distribution uniformity; third, kernel density estimation and standard deviation ellipse were employed to examine spatial layout characteristics; finally, Moran's Index was used to analyze spatial aggregation.

Subsequently, ArcGIS 10.8 was utilized to explore the overall correlations between natural, social, and cultural factors and the distribution of traditional villages. The global influence of each factor on the distribution of traditional villages was compared through OLS and Geodetector's factor detection to evaluate the importance
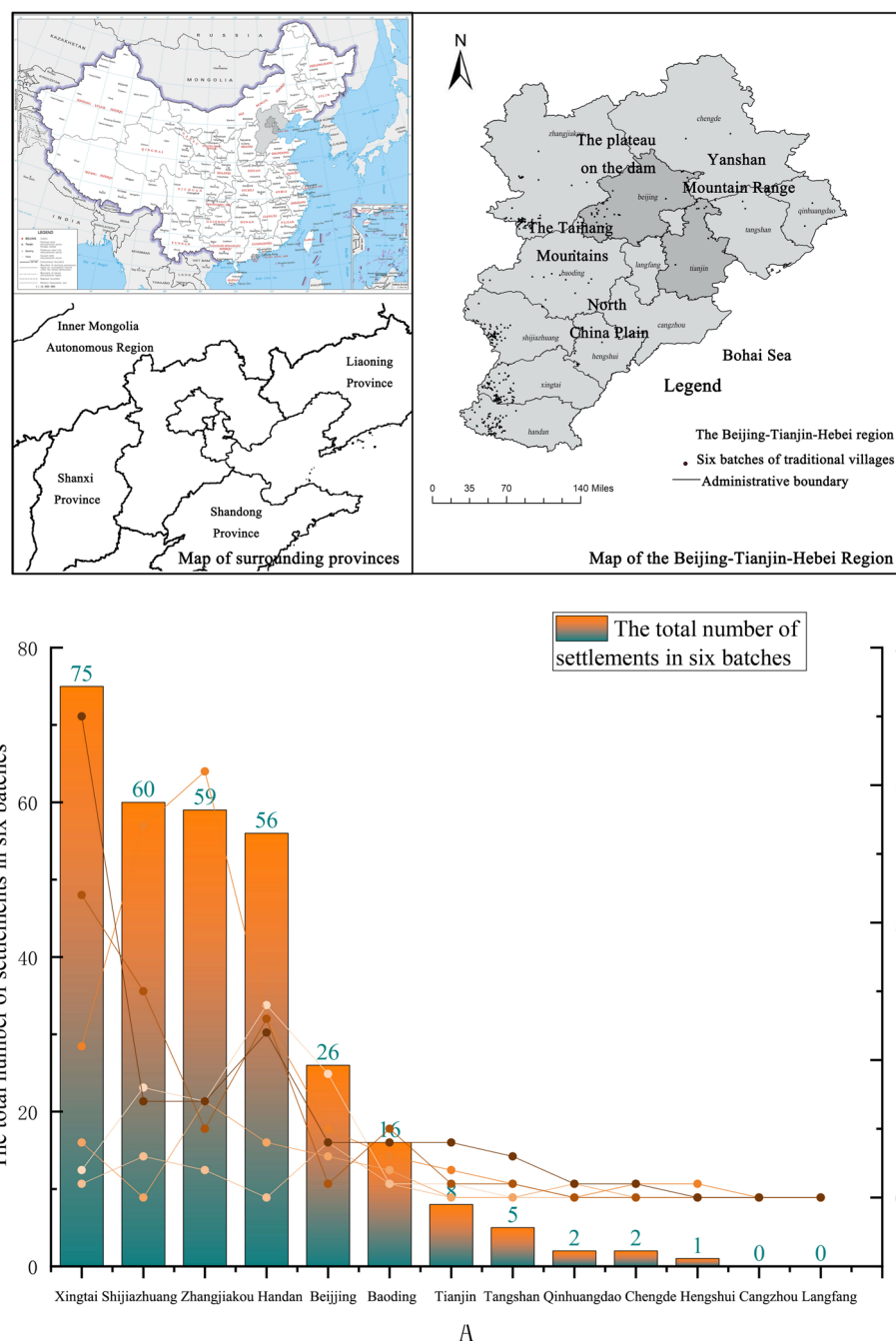
Fig. 1. Distribution of the Beijing-Tianjin-Hebei region and six batches of traditional villages.Figures were drawn by authors, using ArcGIS 10.8(Ministry of Natural Resources,China.http://bzdt.ch.mnr.gov.cn/index.html).

of each factor. GWR was then used to analyze the spatial heterogeneity of these factors, revealing how their effect intensities varied with spatial location. Meanwhile, Geodetector's interaction detector was employed to assess synergistic effects between factors and to identify types of interactions, enabling a comprehensive analysis of the driving mechanisms of multiple factors on the distribution of traditional villages.

This approach ultimately explored the spatial distribution patterns in the Beijing-Tianjin-Hebei region in depth, providing a scientific basis for formulating protection and development policies for traditional villages. The research framework is shown in Fig. 2.

|	Data sources
"Chinese Traditional Villages"	https://www.gov.cn/zhengce/zhengceku/2023-03/21/content_5747708.htm
Gross domestic product (GDP) and gross domestic product per capita (GDP per capita)	The National Bureau of Statistics (https://www.stats.gov.cn/)
Digital Elevation Model (DEM)	Geospatial Data Cloud Platform (https://www.gscloud.cn/)
River data	Ministry of Water Resources of the People's Republic of China (http://szy.mwr.gov.cn/)
Climate data	National Earth System Science Data Center (https://www.geodata.cn/)
Demographic data	Resource and Environmental Science and Data Platform (https://www.resdc.cn/)
A-level tourist attractions data	Hebei Provincial Department of Culture and Tourism (https://whly.hebei.gov.cn/) Beijing Municipal Bureau of Culture and Tourism (https://whlyj.beijing.gov.cn/) Tianjin Municipal Bureau of Culture and Tourism (https://whly.tj.gov.cn/)

Table 1. Data sources.
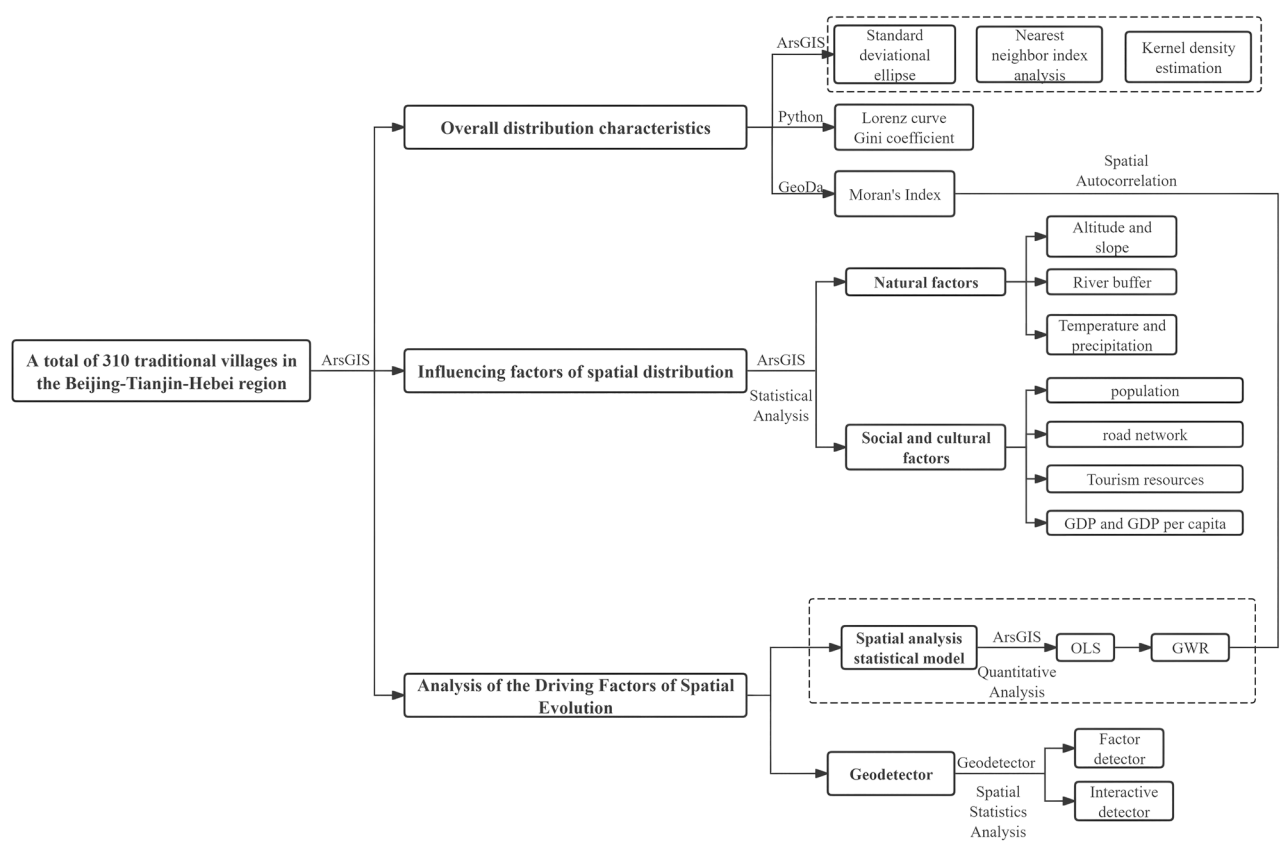
Fig. 2. Technical route(Figure was drawn by authors) .

Methods
Standard deviation ellipse
The standard deviation ellipse is an ellipse centered on the mean of the data points and generated with the standard deviation of the data points on each coordinate axis as the major and minor axes. In the field of data analysis, how to better understand the distribution of data has always been an important issue. The standard deviation ellipse algorithm is an effective visualization method that can help us better understand the distribution of data, find outliers, determine reliable data ranges, etc[28]. The study employs the standard deviation ellipse in the "Directional Distribution" tool of ArcGIS 10.8, and inputs data from four periods sequentially for result analysis. This approach facilitates the analysis of changes in the spatial distribution of traditional villages in the Beijing-Tianjin-Hebei region across different time periods.

The formula for the standard deviation the x'-axis and y'-axis is:

$$SDE_X = \sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \quad SDE_y = \sqrt{\sum_{i=1}^{n} \left(y_i - \overline{Y}\right)^2}$$

Nearest neighbor index analysis

The nearest neighbor index serves as an indicator for measuring the spatial distribution patterns of geographic features. By comparing the actual data to that of a completely random distribution pattern using the nearest neighbor distance, the spatial distribution patterns of geographic elements can be categorized as clustered, uniform, or random[29].

The calculation formula is:

$$ANN = \frac{\overline{Do}}{\overline{DE}}$$

Among them, $\overline{Do}$ is the average between the measured elements and their nearest neighbor centroid distance:

$$\overline{Do} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$\overline{D}_E = \frac{0.5}{\sqrt{\frac{n}{A}}}$$

Lorenz curve and the Gini coefficient

The Lorenz curve is a graphical tool used to measure the degree of evenness in a distribution[30]. The greater the curvature of the Lorenz curve, the more uneven the distribution is. Conversely, the smaller the curvature, the more even the distribution becomes. The study uses the Lorenz curve to graphically illustrate the distribution of traditional villages among cities in the Beijing-Tianjin-Hebei region.

The Gini coefficient, proposed by statistician Corrado Gini and based on the Lorenz curve, measures the degree of inequality. It serves as a statistical tool for assessing the fairness of social income distribution[31]. According to international practice, a Gini coefficient of 0 indicates perfect income equality. A coefficient between 0 and 0.2 signifies highly equitable income distribution. Values between 0.2 and 0.3 indicate relatively equitable distribution, while those between 0.3 and 0.4 suggest a reasonably fair distribution. A coefficient between 0.4 and 0.5 indicates a large income gap, and a value greater than 0.5 signifies a huge income disparity. Traditionally, the Gini coefficient is used to measure the degree of inequality in income distribution, but in our study, we applied it to explore whether the distribution of the number of traditional villages among different cities in the Beijing-Tianjin-Hebei region is equal.

Moran's index

The Moran's index is divided into global Moran's index and local Moran's index. The global Moran's index of a region is analyzed to assess whether spatial clusters or outliers exist in the study area[32]. When global analysis indicates autocorrelation, a local Moran's index analysis is required to identify specific outliers or clustered areas.

The study investigates the formation mechanism of traditional villages among adjacent counties by using the ROOK method in GeoDa software to create a proximity weights matrix. This matrix defines the connections between areal features (counties) and analyzes the spatial pattern.

The calculation formula is:

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

Kernel density estimation

Kernel density estimation is a nonparametric statistical method used to estimate the probability density function underlying a data sample. According to the first law of geography[33], each event can be considered a core element. In areas with a high number of events, the correlation among these core elements is stronger. Kernel density estimation embodies this concept through a function and visually represents it using relevant software[34]. The study employed kernel density analysis in ArcGIS 10.8 under "Spatial Analysis" to reveal the clustering of villages using the default search radius. It divided them into nine categories according to the natural breaks method and estimated the number and distribution characteristics of traditional villages in the Beijing-Tianjin-Hebei region.

Functional formula:

$$f_n = \frac{1}{nh} \sum_{i=1}^{n} K(X - X_i)$$

OLS and GWR

OLS and GWR can explain and predict the impact of variables on the dependent variables. Unlike OLS regression, which has a constant fitting coefficient, the fitting coefficient in GWR models is a function of the geographic location of the sample point.

OLS: The most widely used least squares method, suitable for linear regression models[35]. The goal is to find a linear function that minimizes the sum of squared residuals between all observations and the predicted values of the linear function.

$$\text{OLS formula: } y_i = \beta_o + \sum_{k-1}^{m} \beta_k x_{ik} + \in [1, n]$$

GWR: In 1996, Professor A. Stewart Fotheringham of the University of St Andrews in the United Kingdom formally proposed the GWR mode[36].

$$\text{GWR formula: } y_i = \beta_o (u_i, v_i) + \sum_{k-1}^{m} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i$$

Geodetector

The Geodetector method is a statistical approach that reveals the driving forces behind geographical phenomena by detecting the hierarchical heterogeneity of these phenomena. It can explain the spatial heterogeneity of geographical elements and analyze the interactions among different factors[37].

The Geodetector is used to explore the spatial heterogeneity of the dependent variable Y. The Y value, as the dependent variable, changes with the variation in one or more explanatory variables. Therefore, the selection of the Y value should be related to the research objectives and capable of reflecting the spatial variation phenomenon in the study area. This study aims to explore the spatial distribution characteristics of traditional villages. For the selection of the Y value, an indicator is needed that can reflect the distribution pattern of the villages. Kernel density analysis effectively represents the distribution characteristics of villages. The study ultimately uses kernel density as the dependent variable and multiple driving factors as X values to evaluate the influence of each driving factor (X value) on kernel density (Y value).

Results

Overall distribution pattern and characteristics

Types of spatial distribution

There are three types of spatial distribution of villages: clustered distribution, random distribution and uniform distribution. As shown in Fig. 3, R = 0.427 < 1, the average observed distance is 7.369 km, the expected average distance is 17.252 km, and the absolute value of the Z-score is 19.29. The results show that the spatial distribution of villages in this region is a clustered distribution.

Distribution uniformity of traditional villages

The Lorenz curve deviates significantly from the even equilibrium line and exhibits a greater degree of curvature, indicating that the distribution of villages is uneven (Fig. 4). There are 250 villages in Xingtai, Shijiazhuang,
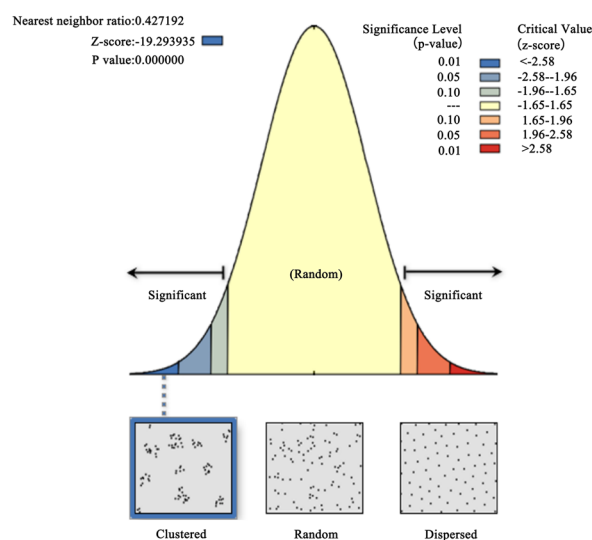
Fig. 3. Nearest neighbor exponential analysis.Figure was drawn by authors, using ArcGIS 10.8
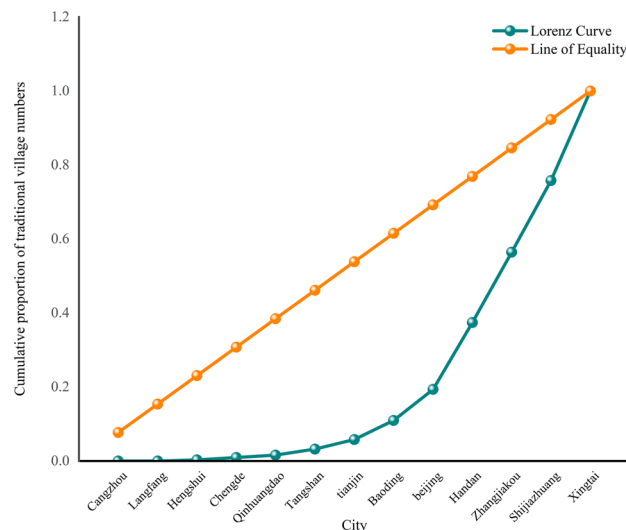
Fig. 4. Lorentz curve. Figure was drawn by authors, using Origin 2024.

Zhangjiakou, and Handan in Hebei Province, representing 81% of the total. In contrast, the remaining eight cities have fewer than 20 villages, indicating significant differences in their distribution.

Global Moran's Index
Hebei Province has 167 counties; Tianjin has 16 districts; Beijing has 16 districts. There are 199 county-level government districts in this region, of which 42 have traditional villages.

Global spatial autocorrelation can determine the spatial correlation of villages[38]. The global Moran's Index is 0.174, demonstrating that the overall distribution of villages in the Beijing-Tianjin-Hebei region is positively correlated. A p-value of 0.01 and a z-value of 3.74 indicate that the distribution in the area is not random but shows significant clustering characteristics.

Local Moran's index
Since the Global Moran's Index reflects the spatial differences of the entire region, it is difficult to show the spatial characteristics and relationships within this region[39]. Therefore, the analysis uses the Local Moran's Index to further assess the local spatial characteristics of the spatial distribution of traditional villages. The region mainly exhibits high-high types, high-low types, and low–high types (Fig. 5). The specific descriptions are as follows:

The high-high (H–H) type belongs to the "double-high" area with a high density of villages itself and high density in the surrounding areas, and exhibits strong spatial correlation among traditional villages. It is mainly distributed in Pingshan County, Luquan County, and the Jingxing Mining Area in Shijiazhuang, Xindu County and Shahe County in Xingtai, and Shexian County in Handan.

The low–high (L–H) type refers to areas with a low density of villages themselves and a high density in the surrounding areas. It is mainly distributed in Yangyuan County, Xuanhua County, and Zhulu County in Zhangjiakou, Zanhuang County and Yuanshi County in Shijiazhuang, and Xiangdu District in Xingtai.

The high-low (H–L) type refers to areas with a high density of villages itself and a low density of villages in the surrounding areas. It is mainly distributed in Zunhua County in Tangshan and Huailai County in Zhangjiakou.

In general, the spatial correlation of village density in the Beijing-Tianjin-Hebei region varies greatly, showing a certain degree of clustering and imbalance in spatial distribution. The overall characteristics are more obvious "high-high" clustering, while "low–high" and "high-low" clusters have a narrower distribution range.

Spatial layout of traditional villages
The standard deviation ellipse is applied to analyze the spatial patterns and evolutionary dynamics of geographical elements[40]. The ratio of the X-axis to the Y-axis is called the oblateness. The greater the oblateness, the more significant the directionality of the distribution trend of villages. Analyses show that after the Yuan Dynasty, the directionality of the distribution trend of villages became increasingly pronounced. Before the Ming Dynasty, the length of the short axis increased over time, and the centripetal force of the ellipse decreased, showing that the distribution of villages became more dispersed. After the Ming Dynasty, the length of the short axis decreased over time, and the centripetal force of the ellipse increased, and the distribution of villages became more concentrated. As shown in Fig. 6(a), the distribution of villages has exhibited a clear trend of gradual stabilization since the Ming Dynasty. Over time, the directionality of village distribution has become more apparent, the degree of dispersion has tended to stabilize, and villages have gradually been distributed along the Taihang Mountains.

The kernel density estimation method was employed to examine the clustering and distribution characteristics of the villages. The overall distribution is uneven (Fig. 6(b)), with most villages distributed in the northern and central areas of the Beijing-Tianjin-Hebei region, while other areas are relatively scattered and have low density.
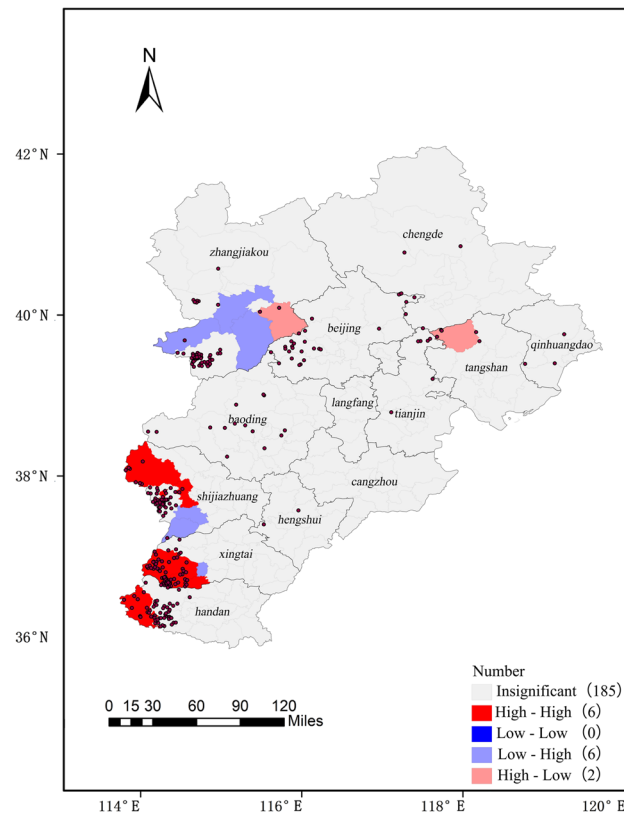
Fig. 5. Moran exponent.Map was drawn by authors, using GeoDa.
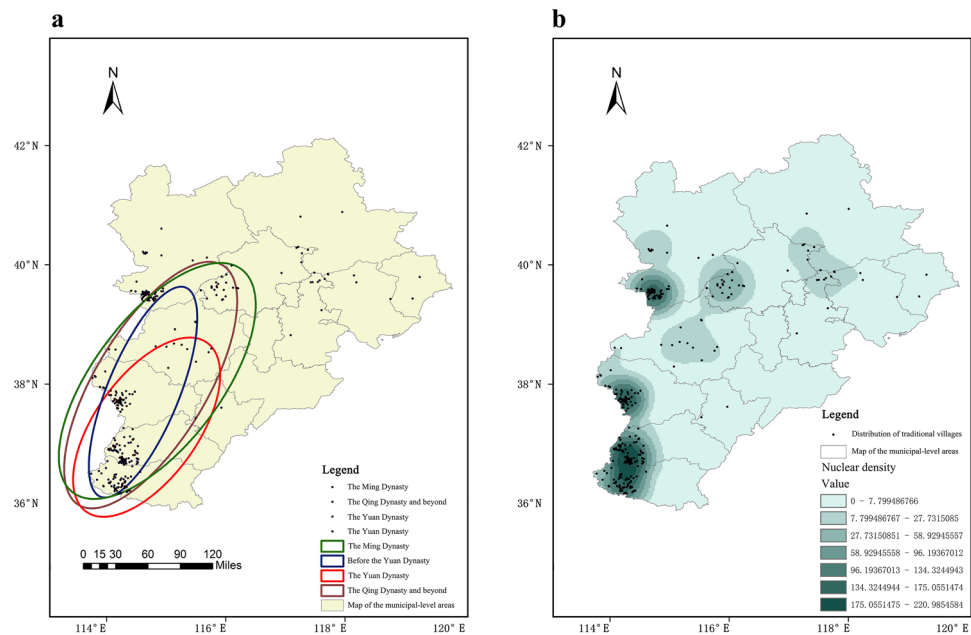
Fig. 6. (**a**) standard deviation ellipse; (**b**) kernel density estimation.Maps were drawn by authors, using ArcGIS 10.8

The overall spatial layout is mainly composed of three high-density areas and three medium-density areas. The three high-density areas are located in the west of Shijiazhuang, at the junction of Xingtai and Handan, and south of Zhangjiakou. The medium-density area is positioned in the central part of the region.

Factors affecting spatial distribution

Influencing factor indicators

Based on the experience of previous researchers and in combination with the specific attributes of the study area, this study constructed a system of independent variables comprising ten indicators, including both natural and socio-economic factors, to serve as the basis for selecting influencing factor indicators, as shown in Table 2.

Relationship between spatial distribution and natural factors

Altitude   Altitude affects climate conditions, vegetation cover, land use, and human activities. Studies have shown that changes in altitude are key determinants of climate suitability and ecological diversity. Therefore, altitude is an important indicator of the potential for human settlement and agricultural development.

The Yanshan Mountains run east–west and the Taihang Mountains run north–south; together, these mountain ranges converge to form the backbone of the Beijing-Tianjin-Hebei area, creating its most significant mountain landscape. This study selected altitude as the key indicator to analyze the impact of natural factors on village distribution.

The study used ArcGIS10.8 to extract the altitude data of villages and analyzed the average, minimum, maximum altitude and height difference of 310 villages to study the distribution pattern of villages at different altitude levels.

As shown in Fig. 7(a), villages are mainly distributed along the Yanshan-Taihang mountain range, while those in the North China Plain are more sparsely distributed. The results indicate that the average altitude of the 310 villages is 512 m, with an altitude range from the lowest at 4 m to the highest at 1620 m, totaling a difference of 1616 m. There are 42 villages below 200 m, accounting for 13.5% of the total, and these are relatively scattered across the plains. There are 130 villages between 200 and 500 m, representing 42% of the total, and 138 villages between 500 and 2000 m, accounting for 44.5% of the total, with a higher concentration in the hilly highlands. The villages are predominantly located in the hilly and mountainous areas of the Yanshan-Taihang Mountains in the northwest, yet they are relatively dispersed. The results indicate that village distribution diminishes with increasing altitude.

Slope   Slope is a key topographic factor that affects land use type, agricultural sustainability, and soil erosion. Farming difficulty and soil erosion impact human settlement and ecological balance. Therefore, slope is an important indicator for assessing land suitability and environmental risks.

The slope of cultivated land is categorized into five levels, namely $\leq 2°$, $2° \sim 6°$, $6° \sim 15°$, $15° \sim 25°$, and $> 25°$. Different slope levels have different impacts on agricultural land use efficiency.

As the elevation increases, the number of villages gradually declines (Fig. 7(b)). In this area, nearly half of the villages have relatively steep average slopes. A total of 90 villages, accounting for 29.03%, are situated in regions with gentle slopes ($\leq 2°$) and generally have no soil erosion. A total of 181 villages are located in regions with relatively gentle slopes ($2° \sim 15°$), accounting for 58.39%, and are characterized by mild to moderate soil erosion, necessitating enhanced soil and water conservation efforts. As the average slope gradually increases, the problem of soil erosion becomes increasingly severe, and the number of villages also decreases. When the slope exceeds 15°, these villages account for 12.58% of all villages; the issue of soil erosion is already significant, and a series of comprehensive measures needs to be taken to prevent and control soil erosion. When the slope is greater than 25°, the strategy of gradually converting farmland to forestland should be considered.

Villages are predominantly located in regions with level terrain and gentle slopes. These areas provide favorable conditions for housing construction, production activities, agricultural operations, and daily life. For areas with steeper terrain, corresponding measures should be taken to minimize disturbance to the ecosystem and promote its recovery.

Factor	Indicator	Description of indicators
Natural Factor	Altitude (X_1)	Calculation of study unit mean elevation from DEM raster data
	Slope (X_2)	Calculation of the average slope of the study unit from the slope raster data
	River buffer (X_3)	Analyze the buffer zone of the five-level water system according to the national water system classification
	Temperature (X_4)	Calculation of the mean annual air temperature of the study unit from the mean annual air temperature raster data
	Rainfall (X_5)	Calculation of mean annual rainfall in the study unit from mean annual precipitation raster data
Social and Cultural Factors	Population (X_6)	Population figures from statistical yearbooks
	GDP (X_7)	Obtaining GDP by county based on official government statistics
	GDP per capita (X_8)	Ratio of GDP to number of resident population
	Tourism resources (X_9)	According to the latest statistics of A-level tourist attractions released by the Tourism Bureau
	Transportation conditions (X_{10})	Analysis of highway buffer zone settings in the Beijing-Tianjin-Hebei region

Table 2. Indicators of influencing factors.
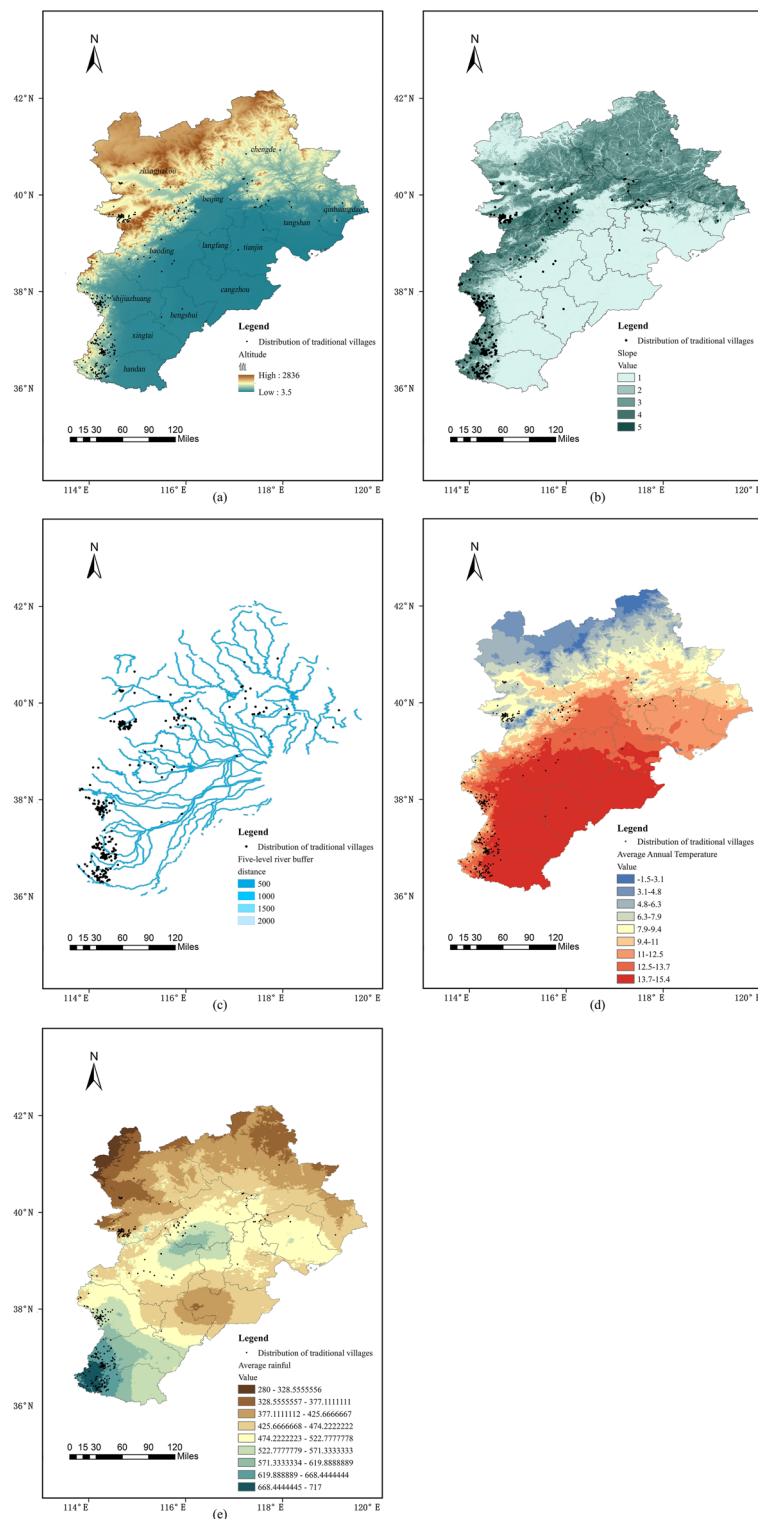
Fig. 7. (**a**) Elevation; (**b**) Slope; (**c**) River buffer; (**d**) Temperature; (**e**) Rainfall. Maps were drawn by authors, using ArcGIS 10.8

River factors   Rivers play a decisive role in the location of traditional villages, especially in areas with high water dependence and agricultural production. Rivers provide the necessary water resources to support agricultural irrigation and also affect flood risk management in the region. Therefore, it is necessary to use river factors as a key indicator for assessing the spatial distribution of villages.

According to relevant regulations, China's water systems are divided into the following categories[41]: First-level water systems, mainly covering China's large rivers, such as the Yangtze River and the Yellow River; second-

level water systems, which usually refer to the main tributaries of the first-level water systems; third-level water systems, which mainly refer to the tributaries of the second-level water systems; fourth-level water systems, which usually refer to the tributaries of the third-level water systems, such as small rivers and streams; fifth-level and lower water systems, which mainly refer to the tributaries of the fourth-level water systems, such as ditches and artificial waterways. These water systems are relatively small in scale, and their main functions include irrigation, drainage, and flood prevention.

Beijing and Tianjin both belong to the Haihe River Basin, and all water systems in Beijing belong to the Haihe River Basin, with 91% of Hebei Province also belonging to this basin. The Luanhe River System is the second largest river system in Hebei Province.

The Beijing-Tianjin-Hebei region is frequently hit by floods. Floods in Beijing are mainly caused by the three major water systems: the Yongding River, Beiyun River, and Chaobai River[42]. Historical statistics show that the Yongding River had 81 breaches, 59 overflows, and 9 diversions in the 835 years from AD 1115 to 1949. From 1470 to 1948, the region experienced 43 floods that lasted for more than two years consecutively, totaling 122 years, with the longest consecutive flood record being 7 years. Tianjin, as the estuary where the five major river systems of the Haihe River converge, is a typical flood-prone area. According to historical documents, from AD 146 to 1854, there were 341 years affected by waterlogging, of which 85 were major floods, 50 years were characterized by drought but mainly waterlogging, and 7 years experienced severe drought and flooding. Particularly worth mentioning is the flood in 1939, which submerged Tianjin for a long time, lasting for 2 months, with the water depth in the streets reaching 2 m. Hebei Province has long been plagued by a variety of natural disasters, including floods, waterlogging, and droughts, with their severity and frequency being especially prominent. Due to the extremely uneven distribution of rainfall in the province, large-scale floods and droughts often occur alternately, significantly impacting the lives and production of local people.

By constructing a river buffer zone analysis model for this region, it was found (Fig. 7(c)) that there are 41 villages within a 500-m buffer distance, 14 villages within a 1,000-m buffer distance, 8 villages within a 1,500-m buffer distance, and 14 villages within a 2,000-m buffer distance. Additionally, the number of villages with a buffer distance greater than 2,000 m is 230, representing 74.2% of the total. These data show that most villages are located far from rivers. According to historical documents, these areas are prone to flooding. Therefore, maintaining a sufficient distance between villages and rivers is crucial to mitigate the impact of waterlogging disasters.

<u>Temperature</u>   Temperature suitability is a key factor in determining crop types and planting times, while extreme temperature conditions are particularly restrictive to agricultural production. The study utilized GIS technology to overlay temperature data with village location data, visually displaying the relationship between temperature and village distribution, and thus exploring its impact on spatial distribution.

The annual mean temperature in this region ranges from -1.5 °C to 15.4 °C (Fig. 7(d)). Extreme temperatures have a significant restrictive impact on agricultural production. The study found that 64 villages were situated in regions where the average annual temperature was below 10 °C, accounting for 20.65% of the total. Temperature conditions in most other villages are relatively suitable. Low-temperature environments significantly negatively affect crop growth[43], which may cause the growth to be slowed or even stagnant, thereby affecting yield and quality. Suitable temperature conditions promote crop growth and development, while inappropriate temperatures can hinder it.

<u>Rainfall</u>   Precipitation directly affects agricultural irrigation demand, water availability, and flood risk management. In areas with limited water resources or uneven precipitation, it becomes an important factor affecting land use and agricultural planting choices. Therefore, analyzing precipitation is of great significance for studying the spatial distribution of villages in the Beijing-Tianjin-Hebei region.

The study utilized ArcGIS 10.8 to statistically analyze and calculate the annual average precipitation in each region, and matched these data with the geographical locations of the villages to evaluate the impact of precipitation on the distribution of villages.

As shown in Fig. 7(e), the annual precipitation in this region ranges from 280 to 717 mm. Among these, 10 villages are located in the semi-arid zone with an average annual rainfall of between 200 and 400 mm, representing 3.21% of the total. Most of the remaining villages are distributed in the semi-humid zone. The data in the figure indicate that rainfall affects the distribution of villages. Insufficient rainfall may lead to water shortages and have a negative impact on agricultural production, while excessive rainfall may enhance the potential for natural disasters such as floods.[44]

Relationship between spatial distribution and social and cultural factors
Population
According to survey data, the population size of the Beijing-Tianjin-Hebei region has reached 110.37 million. The total population of the four cities of Beijing, Tianjin, Shijiazhuang, and Baoding is 58.2579 million, accounting for over 50% of the region's population. In this region, the urban population totals 75.73 million, and the permanent population's urbanization rate has climbed to 68.61%. At the same time, the ethnic minority population is 4.71 million, representing 4.27% of the total population. The western region has a relatively sparse population distribution due to its rugged terrain, particularly in the plateau and mountainous areas north and west of the Yanshan-Taihang Mountains.

A total of 127 traditional villages were identified within the population range of 400,000 to 500,000 (Fig. 8(a)). In the population range of less than 200,000, only one traditional village was found, while in the population range exceeding 1,000,000, there are only two traditional villages. In areas with small populations, the shortage of labor and capital restrict the construction and maintenance of villages. Relatively speaking, densely populated
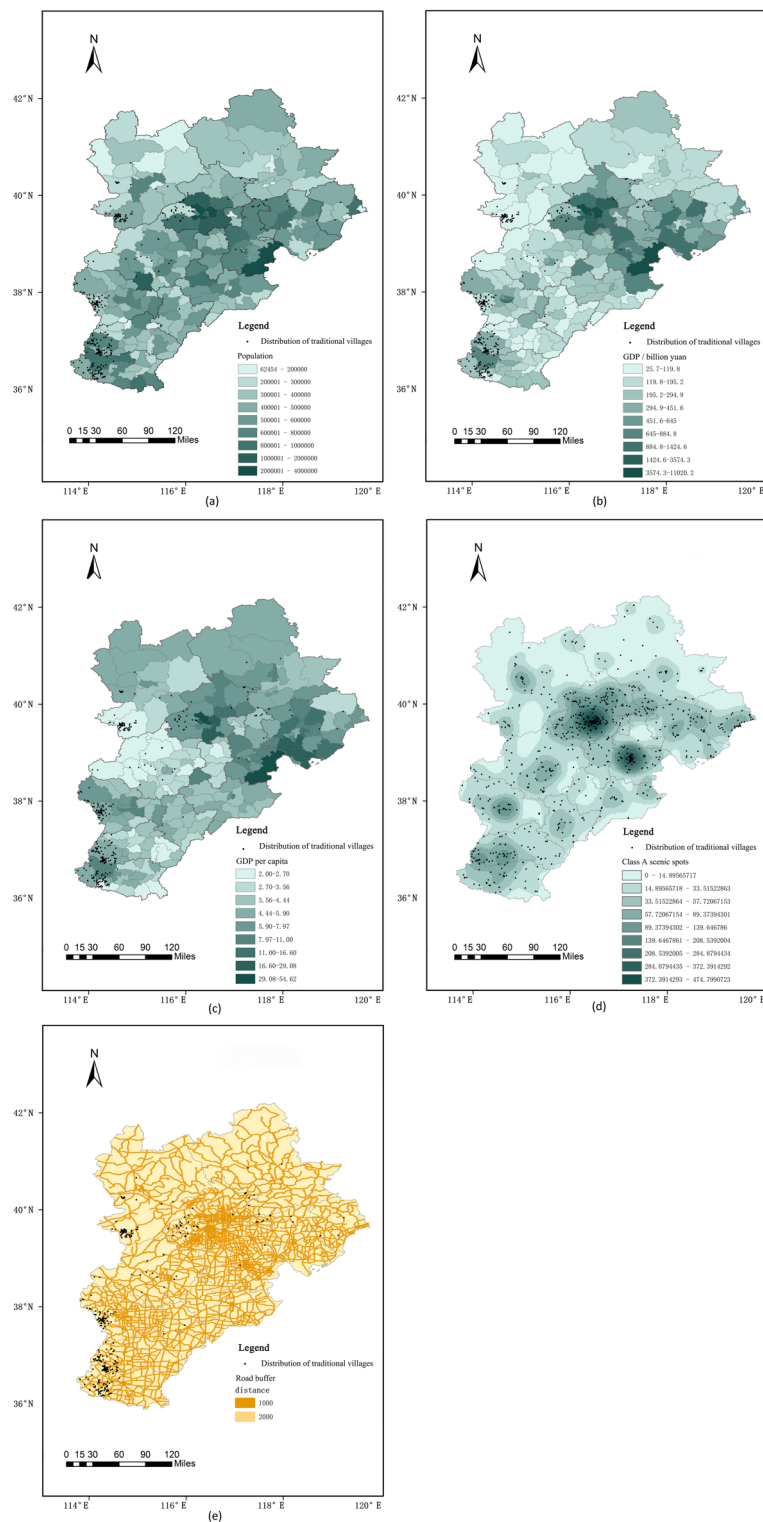
Fig. 8. (**a**) Population; (**b**) GDP; (**c**) GDP per capita; (**d**) Tourism resources; (**e**) Transportation conditions. Maps were drawn by authors, using ArcGIS 10.8

areas exhibit a higher degree of modernization, causing a comparatively smaller number of villages. Areas with a moderate population size are more beneficial to the protection and development of villages. Such areas provide the necessary labor and community functions, helping to preserve traditional villages without the adverse effects of excessive modernization.

Based on the above analysis, extreme values of population density impact the geographical distribution. In densely populated areas, urbanization tends to be more pronounced, such as in all of Beijing and Tianjin, as

well as in parts of Tangshan, Shijiazhuang, and Qinhuangdao in Hebei Province. In these areas, the acceleration of urbanization and modernization has had a substantial impact on traditional villages. Especially in Beijing, the national capital, the acceleration of industrialization and urbanization has intensified the conflict between village protection and development. In contrast, in sparsely populated, extremely low-density areas, such as in Jingxing County, a mountainous area within Shijiazhuang, most villages are surrounded by mountains, and villagers are scattered in basins at the foot of the mountains. These villages were formed as a result of historical wars and famines and have remained isolated from the outside world. Due to the difficult living conditions, sparse population, and scattered locations, these traditional villages have been well preserved.

GDP

GDP, as a core indicator for measuring the economic prosperity and level of development in a country or region, is of great significance for economic analysis of the region (Fig. 8(b)). The number of traditional villages in the lower GDP range (2.57 billion to 29.49 billion yuan) is relatively large, amounting to 231, representing 74.52% of the total. In the higher GDP range (88.48 billion to 1.10202 trillion yuan), the number of villages decreases significantly to 3. By observing the graphical data, we can clearly see that with the continuous growth of GDP, the number of villages has dropped significantly. Economic growth is usually accompanied by accelerated industrialization and urbanization, which leads to a substantial number of rural people migrating to cities, thereby causing the gradual decline and even extinction of traditional villages. In addition, economic growth may also lead to over-exploitation of natural resources and environmental damage. This will further increase the difficulty of protecting traditional villages and cause relatively greater damage to them.

GDP per capita

GDP per capita, as a key indicator for measuring a country or region's economic development level and the affluence of its inhabitants, is a more accurate and representative measure than total GDP[45]. As shown in Fig. 8(c): the GDP per capita distribution across counties and districts in the Beijing-Tianjin-Hebei region is as follows: 90 traditional villages are located in the lower range (GDP per capita between 20,000 yuan and 44,400 yuan), representing 29.03% of the total; 219 traditional villages are located in the medium-to-high range (GDP per capita between 44,400 yuan and 110,000 yuan), accounting for 70.97% of the total; and no traditional villages are found in the higher range (GDP per capita exceeding 110,000 yuan). The results clearly demonstrate that regional economic imbalances have influenced the distribution pattern.

Tourism resources

Exploring the distribution of tourism resources within the village will help reveal the potential for tourism development and promote both its sustainable development and cultural heritage protection. There are two opposing views on the protection of cultural heritage. One view is that tourism development is a key way to revitalize the culture of traditional villages; the other view is that tourism development has a negative impact on the cultural protection of traditional villages[46] and may lead to the utilitarianization of interpersonal relationships in the village, the commercialization of culture, and the degradation of cultural landscapes. However, in actual operation, there is complementarity between the cultural protection of traditional villages and the use of tourism resources. As long as a balance can be found between the two, tourism can fully leverage its positive role in the protection and utilization of traditional village culture if properly managed.

According to the latest statistics released, there are 853 A-level tourist attractions, including 526 in Hebei Province, 104 in Tianjin, and 223 in Beijing[47–49]. As shown in Fig. 8(d), after an in-depth analysis of these data, we can clearly observe that Tianjin and Beijing exhibit extremely high tourism resource intensity, with Hebei Province also displaying relatively high resource intensity.

Most traditional villages are well-preserved, boasting rich tourism and cultural resources with significant potential for tourism development. These villages are unique in architectural style and layout and also hold significant value in folk culture and historical heritage. The development of tourism can effectively promote the protection and inheritance of these village cultures, allowing more people to understand and appreciate this unique cultural heritage[50]. However, due to various factors such as natural erosion, human destruction, and a lack of effective protection measures, the original appearance of some of these traditional villages has gradually become blurred. At the same time, some precious folk cultures are gradually disappearing, leading to a loss of their original tourist appeal. Consequently, it is difficult for these areas to effectively protect and inherit village culture through the development of tourism.

Transportation conditions

Road construction has significantly improved the transportation accessibility of traditional villages, which has far-reaching significance for enhancing village infrastructure. With the gradual convenience of transportation, the villagers' travel conditions have significantly improved, freeing them from the constraints of previous transportation inconveniences[51]. The introduction of modern resources and facilities will gradually promote economic prosperity, and convenient transportation also promotes the vigorous development of tourism, contributing to the overall economic prosperity.

As shown in Fig. 8(e), the majority of villages in the Beijing-Tianjin-Hebei region are effectively located within the buffer zones of national, provincial, and county roads. Specifically, within the 1,000-m buffer zone, a total of 139 villages are covered; within the 2,000-m buffer zone, the number of covered villages is 60. These data clearly demonstrate the key role of transportation conditions in enhancing village accessibility and promoting economic development.

Limitations in selecting influencing factor indicators
Regarding the research parameter system, the article mainly focuses on directly quantifiable geographical and climatic factors, as well as some economic and cultural factors, but it does not fully cover all factors affecting regional development. In particular, factors such as land use changes, historical policy interventions, and urbanization trends also affect the long-term development of the region. The study should also discuss qualitative factors such as intangible heritage, which cannot be quantified but have an impact on village development. For example, the impact of religious culture, historical policies, and land use changes could enrich the analysis of cultural impacts, while also facilitating discussion on these non-quantifiable factors.

Religious culture is one of the cultural elements, but because it covers a narrow scope and cannot constitute an important factor in analyzing the spatial distribution of traditional villages, this article does not include religious culture among the driving factors. Among the 310 traditional villages in the Beijing-Tianjin-Hebei region, only a very small number are located around temples or Taoist temples, and some villages also have small temples for villagers to worship. For example, the "Shousheng Temple" in Wangli Village (formerly known as Wanger Village) in Jingxing County, Hebei Province, is a small temple that was spontaneously built by villagers during the Ming Dynasty.

Regarding historical policies, the author has paid attention to this aspect during his research. The most distinctive policy can be traced back to the early Ming Dynasty. Because Beijing was sparsely populated after the war, the court began to implement an immigration policy. According to historical records, large-scale immigration activities were organized several times during the Ming Dynasty. These activities were mainly in Hongdong County, Pingyang Prefecture, Shanxi Province, known as the "Hongdong Immigration." Some people moved to Beijing, Tianjin, and 129 counties in Hebei. Some of the traditional villages in Beijing, Tianjin, and Hebei originated from migrations from surrounding provinces in the early Ming Dynasty. However, such historical records are too subjective and difficult to quantify in research. In our study, we did not explore these factors in depth, but we recognize their importance. Future research will attempt to explore how these cultural factors specifically affect regional development and residents' lives through case studies and other methods to supplement the existing quantitative analysis.

In terms of land use, northern villages differ from southern villages. In the Yellow River Basin, northern villages are mainly concentrated settlements where farmland and farmhouses are closely connected. The fertility or barrenness of farmland influences the distribution and development of villages. Compared with the fertile soil in the south, the quality of the land in the north is relatively poor, and the population in the villages is not as large as in the south. Taking Zhengding County in Hebei Province during the Ming and Qing Dynasties as an example, most villagers would live in villages with fertile land, which offered higher income and profit. The amount of land cultivated was closely related to the amount of silver earned. Similarly, this type of research is not easy to quantify, and since the Beijing-Tianjin-Hebei region is located in the north, the differences within the region are not pronounced.

Analysis of spatial heterogeneity of traditional villages

Geodetector analysis
<u>Univariate analysis</u>   Ten independent variables were selected for analysis, and their scores are as follows: the altitude (X1) in the traditional village area is divided into five levels; the slope (X2) in the traditional village area is divided into five levels; the buffer distance between the traditional village and the river (X3) is classified into five divisions at intervals of 0.5 km; the transportation network analysis (X10) is divided into three categories with a buffer zone of 1 km; temperature (X4), rainfall (X5), population (X6), GDP (X7), GDP per capita (X8), and tourism resources (X9) are divided into nine categories according to natural breakpoints.

The factor detectors are ranked by their explanatory power: rainfall (X5) > GDP per capita (X8) > transportation conditions (X10) > tourism resources (X9) > slope (X2) > population (X6) > GDP (X7) > river buffer (X3) > temperature (X4) > altitude (X1)[37].

The smaller the p-value, the stronger the evidence that the independent variable X impacts the dependent variable Y, indicating higher reliability for this inference. As shown in Fig. 9(a), rainfall X5 (0.003), population X6 (0.074), and transportation conditions X10 (0.071) are influencing factors with relatively small p-values, indicating a significant effect on the spatial distribution of villages. Altitude (X1), slope (X2), river buffer (X3), temperature (X4), and GDP (X7) exhibited relatively weaker explanatory power.

<u>Factor interaction detection</u>   In the process of exploring the relationships among multiple factors, there are varying degrees of connection between these factors. To accurately assess the roles of these factors in explaining spatial differences, this study applied the interactive detection module within the Geodetector. The study focused on two specific factors, A and B, comparing their impacts on traditional villages. It aimed to determine whether these factors act independently on the spatial distribution of villages or interact in a joint manner.

As an effective analytical tool, the interaction detector can help us identify interactions between different explanatory variables[37]. To thoroughly explore the interactions between these factors, interaction detectors were further utilized for more precise analysis, as shown in Fig. 9(b). The interaction between GDP and population (q = 0.90) is the foremost driving force, followed by the interaction between GDP and rainfall (q = 0.88), the interaction between population and river buffer (q = 0.85), the interaction between GDP and river buffer (q = 0.79), and the interaction between transportation conditions and population (q = 0.77).

According to the analysis results, spatial heterogeneity is influenced by the interactions among multiple factors, including population, economy, and environment. The interaction effects are significantly stronger than those of any single factor, revealing the key driving mechanisms behind the formation of spatial patterns.
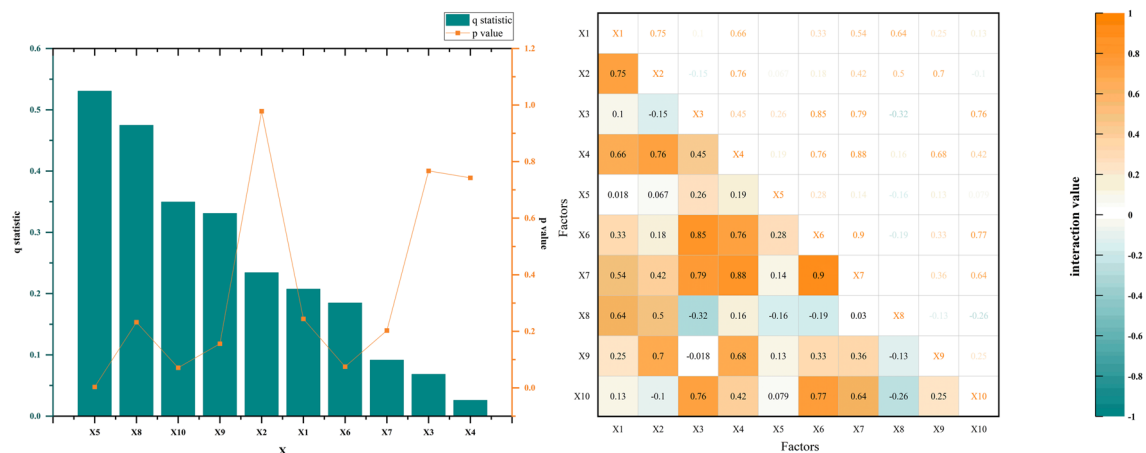
Fig. 9. Geodetector factor detection results: (**a**) single factor detection results; (**b**) interaction detection results. Figures were drawn by authors.

	Pearson's r	p-value	R-Square (COD)
Altitude (X_1)	0.17122	< 0.0001	0.02931
Slope (X_2)	0.25673	0.74132	0.06591
River buffer (X_3)	-0.13953	0.68222	0.01947
Temperature (X_4)	0.05627	< 0.0001	0.00317
Rainfall (X_5)	0.45026	< 0.0001	0.20274
Population (X_6)	-0.04535	0.01299	0.00206
GDP (X_7)	-0.03101	0.00923	0.00096
GDP per capita (X_8)	-0.01718	0.02530	0.00029
Tourism resources (X_9)	0.23507	0.00489	0.05526
Transportation conditions (X_{10})	-0.01685	0.00889	0.00028

Table 3. Results of OLS analysis.

OLS and GWR model analysis

The study employs the OLS method to quantify global spatial relationships and utilizes the GWR model to analyze spatial heterogeneity by calculating the local regression coefficients (LRC) that relate the spatial distribution of local regions to influencing factors[52].Before conducting the GWR analysis, the spatial autocorrelation of the kernel density was assessed to establish the presence of spatial dependence.
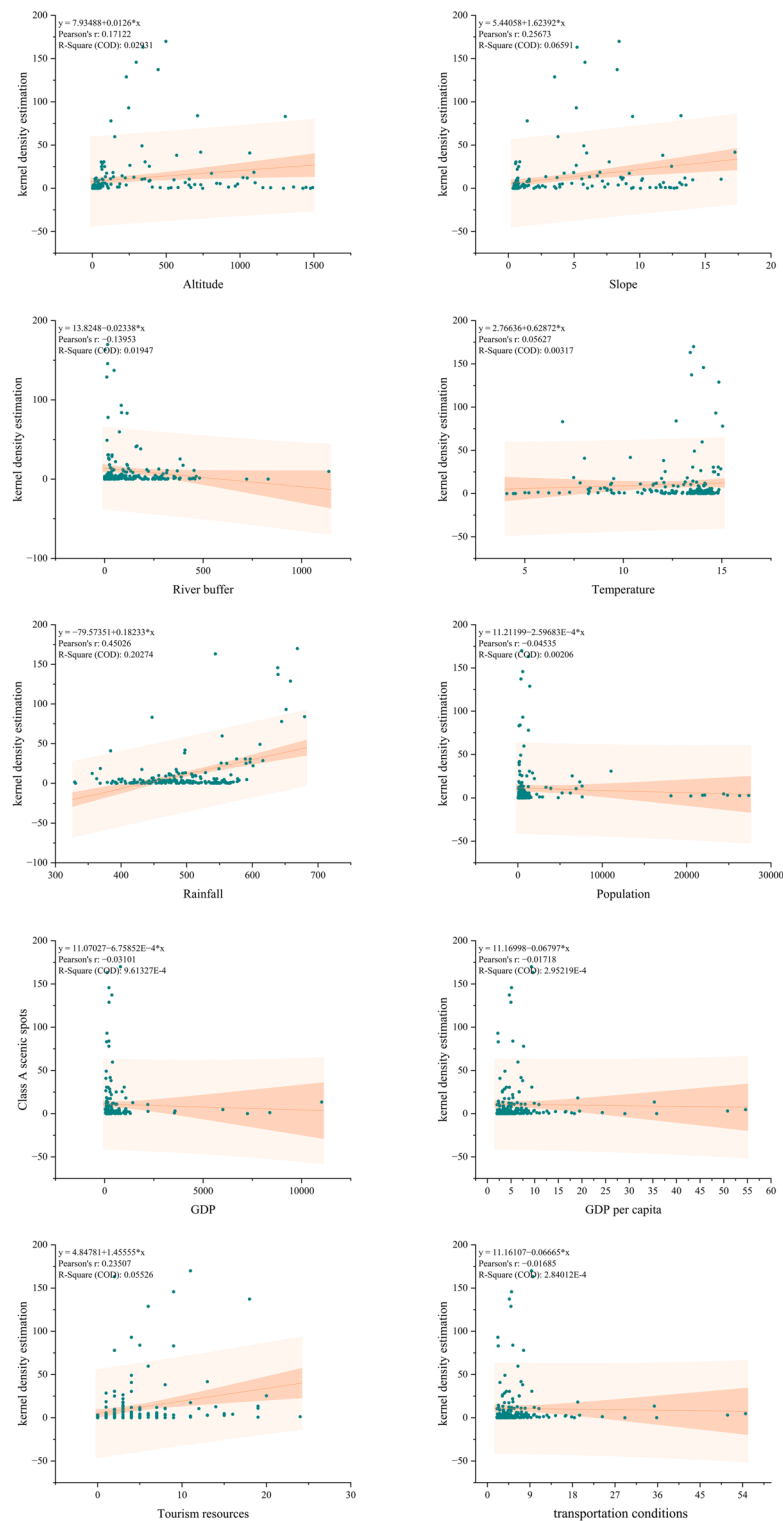
OLS analysis
The study uses the kernel density estimate of villages as the dependent variable Y, and altitude (X1), slope (X2), river buffer (X3), temperature (X4), rainfall (X5), population (X6), GDP (X7), GDP per capita (X8), tourism resources (X9), and transportation conditions (X10) as independent variables X.

The Pearson correlation coefficient (r value) measures linear correlation, ranging from -1 to 1[53]. R-squared (COD) refers to the indicator for determining the goodness of fit of the linear regression line.

There are significant correlations between traditional village kernel density estimates and eight factors. Altitude (X1), temperature (X4), rainfall (X5), and tourism resources (X9) (p < 0.05, r > 0) are positively correlated with the kernel density of the region. These factors significantly influence the distribution of village kernel density. Population (X6), GDP per capita (X8), GDP (X7), and transportation conditions (X10)(p < 0.05, r < 0) are negatively correlated with the kernel density of the region. The slope (X2) and river buffer (X3)(p > 0.05) show no significant correlation with the kernel density in this region. The analysis results do not account for spatial heterogeneity. R-squared (COD) generally fails to adequately explain the spatial distribution, indicating that further analysis is required. Based on linear regression analysis, two non-significantly related influencing factors were eliminated. A local spatial regression model based on GWR was further constructed, as shown in Table 3 and Fig. 10.

GWR analysis
There are significant spatial clustering and autocorrelation in the spatial distribution of villages (Moran's Index = 0.174 > 0, Z = 3.74). The GWR model increases the R^2 value (OLS: 0.449; GWR: 0.473) and reduces the bandwidth and AICc values (OLS: 1766.78; GWR: 1752.50), confirming the influence of spatial heterogeneity.

Fig. 10. OLS linear regression analysis. Figures were drawn by authors.

Because the VIF values of altitude, GDP per capita, and transportation conditions are greater than 7.5, there is multicollinearity that needs to be addressed. Temperature, rainfall, tourism resources, population, and GDP were selected as factors influencing the local spatial heterogeneity of the spatial distribution, as shown in Fig. 11. The average absolute value of the LRC for each influencing factor indicates their overall contribution to the protection of the cultural heritage. The local spatial impact intensity of the five factors is as follows: temperature has the largest overall contribution ($|LRC| = 2.03837$), followed by tourism resources ($|LRC| = 1.79494$), rainfall ($|LRC| = 0.20169$), GDP ($|LRC| = 0.00430$), and population ($|LRC| = 0.00047$). Additionally, by combining the
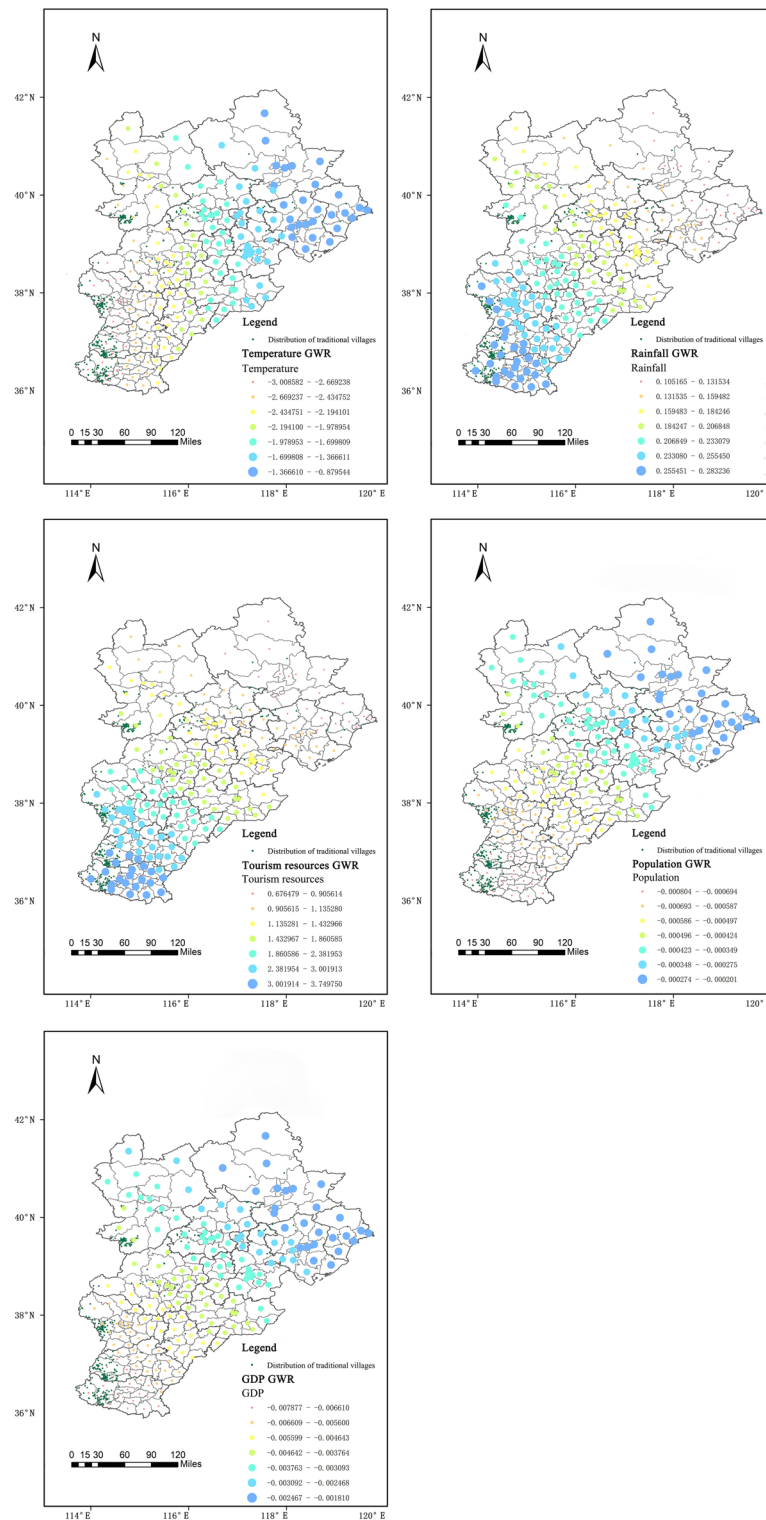
Fig. 11. Spatial distribution of LRC calculated by the GWR model. Maps were drawn by authors, using ArcGIS 10.8

LRC values with spatial distribution, the influencing factors are categorized into two types: positively correlated areas (LRC > 0) and negatively correlated areas (LRC < 0), as illustrated in Fig. 8.

According to the analysis of the temperature regression coefficient diagram, temperature has a significant positive impact. Temperature factors affect the economic foundation of villages by influencing the growth cycle of crops, becoming one of the key factors in village distribution. However, this impact varies in different regions, with its negative impact gradually decreasing from northeast to southwest. There are three high-density clusters

City	Construction year				Village origins		Topography	
Xingtai	Yuan	Ming	Qing	Republic of China	Hebei migration Shanxi migration Sichuan migration	automatic formation	Hills hilly area	
Shijiazhuang	Yuan	Ming	Qing		Hebei migration Shanxi migration	automatic formation	hills hilly area basin	
Handan	Yuan	Ming	Qing		Hebei migration Shanxi migration Henan migration	automatic formation	hills hilly area	
Zhangjiakou	Yuan	Ming	Qing		Hebei migration Shanxi migration	automatic formation border garrison war migration	hills hilly area basin river	

Table 4. Village information.

of traditional villages in the southwest region, and the negative impact of temperature on the area is relatively weak.

Analysis of the distribution of scenic spots in the region shows that there is a positive correlation between the distribution of scenic spots and villages, with the degree of influence decreasing from southwest to northeast. High-value areas are focused on Handan, Xingtai, and Shijiazhuang, while low-value areas are focused on Chengde, and Qinhuangdao. There are high-density clusters of villages in the southwest, whereas regions in the north, such as Zhangjiakou and Chengde, exhibit relatively simpler tourism development and protection measures, with their historical heritage not being as rich as that of Handan and other places.

The analysis of the rainfall regression coefficient diagram shows that rainfall promotes the clustering distribution of traditional villages, with this effect decreasing from southwest to northeast. The impact on the three high-density clustered areas of villages is higher than that in other areas. The GDP regression coefficient diagram shows that there is a negative correlation between economic growth and the distribution of villages, with the degree of influence decreasing from northeast to southwest. The regions with the highest degree of influence are mainly Qinhuangdao and Chengde, which are economically at a medium level, while the regions with lower degrees of influence are mainly in Handan and Xingtai, which are economically less developed and more suitable for the distribution of villages. The more developed the economy is, the less favorable it is for the distribution of villages, and overdevelopment may cause damage to villages. Similarly, increases in population density are unfavorable to the distribution of villages, with the degree of this impact decreasing from northeast to southwest, indicating a negative correlation.

Discussion

Traditional villages emerge from the prolonged interplay between human activities and the natural environment. They record information on the social economy, politics, culture, and folk customs of the time from multiple perspectives[54]. As villages with a long history, these villages not only contain rich traditional resources but also demonstrate significant historical, cultural, and economic value.

Across the vast land of China, different cultural and historical backgrounds have given rise to a variety of village types. On the Qinghai-Tibet Plateau[55], Tibetans, influenced by Tibetan Buddhism for thousands of years, have developed a typical social structure. In the western Sichuan plain[56], Linpan villages demonstrate how to adjust living spaces to achieve thermal comfort according to different climatic conditions, providing important guidance for the reconstruction and restoration of villages in the region. Villages in the Bohai Rim region exhibit significant spatial differences in distribution and degrees of dynamic change[57], while villages in the southeast coastal area are primarily influenced by cultural factors[18].

Particularly, the traditional villages in the Beijing-Tianjin-Hebei region are rich, diverse, and have distinct cultural characteristics. As a high-density cluster of villages, they demonstrate high research value. However, previous studies on villages in this region have been limited to small scopes or a single perspective, lacking comprehensive and integrated analysis. Therefore, this study focuses on analyzing the distribution patterns of villages and their influencing factors, to provide theoretical support for the conservation and inheritance of villages.

Comprehensive analysis of driving factors
Historical culture is the core factor
Shijiazhuang, Xingtai, Handan, and Zhangjiakou, four cities where traditional villages account for more than 15%, were selected as research objects. Statistical analysis was conducted by collecting data from various districts and counties in these cities (Table 4). Most of the villages originated in the Yuan, Ming, and Qing dynasties, and some can even be traced back over 600 years. For these villages with a long history, the initial reasons for their establishment are often difficult to verify accurately through historical documents, and only limited information can be obtained through historical relics such as inscriptions and temples preserved in the villages. During a survey of some villages in Jingxing County, Shijiazhuang, it was found that a few villages still retained some inscriptions, most of which date from the Qing Dynasty, while those from the Ming Dynasty are relatively few. In addition, there are temples in a few villages. Most of these temples are built by the villagers themselves for worship. For relatively new villages, most were formed naturally or due to population migration, and only a few were due to wartime migration. In terms of topographical features, most of these villages are located in hilly and mountainous areas, while there are fewer basins and rivers. The historical background of war and escaping from

chaos, as well as the geographical environment, have jointly shaped the characteristics of traditional villages and nurtured their unique cultural customs, thus forming traditional villages with regional characteristics[58]. At the same time, traditional villages have nurtured many precious cultural heritage items in the course of their long-term development. These cultural heritage items have evolved into valuable wealth over time, which is also the value of the inheritance and protection of villages[59].

The important role of natural socio-economic factors

With the continuous advancement of modernization, the protection and development of traditional villages also face certain challenges in the pursuit of economic benefits and the implementation of large-scale urban expansion[60]. When GIS was used to correlate the geographic information of natural elements with the distribution characteristics of villages, the analysis revealed that the distribution was significantly affected by natural elements. When analyzing altitude, rivers, slopes, climate, etc., it was found that in areas with favorable natural conditions, the number of villages is greater, and the density is higher, with most exhibiting a clustered pattern of spatial distribution[61]. Using GWR to analyze the GDP of each county in the region, the analysis shows that economic growth is negatively correlated with the distribution of villages. Areas with a high density of traditional villages are mainly those with relatively underdeveloped economies, which reflects, to some extent, that the spatial distribution of villages is partially affected by the level of economic development in the region[60]. Using Geodetector for interaction analysis, the analysis indicates that economic impact is inseparable from population impact. Both exhibit significant non-linear growth[62].

Protection and development strategies

Traditional villages in the Beijing-Tianjin-Hebei region are influenced by a variety of factors, including the natural environment and socioeconomic factors. Based on the importance of these influencing factors for the spatial distribution pattern, this study proposes corresponding protection strategies. Research analysis shows that population and economic factors are negatively correlated with village distribution. The three high-density areas and three medium-density areas in the overall spatial layout are all located in areas with lower economic development levels. Examples include Jingxing County in Shijiazhuang, as well as Xingtai and Handan in Hebei Province. There is a positive correlation between the natural environment and village distribution. The areas where traditional villages are located have rich cultural heritage and pleasant natural conditions, which have ensured the ongoing development and clustering of villages over the past hundreds of years. Based on these influencing factors, this study proposes corresponding protection strategies.

In highly urbanized areas such as Beijing and Tianjin, a strategy that prioritizes conservation and development as a complementary approach should be adopted. Economic development has led to the concentration of population and resources in cities, leading people to migrate from rural areas, forming a "hollowing out" phenomenon. During the investigation, we found that the aging problem in traditional villages has worsened and the demographic structure of rural communities has changed significantly. Therefore, in these economically prosperous areas, it is particularly important to adopt a strategy centered on protecting traditional villages, supplemented by moderate and controlled development.

In some areas with low economic development, backward economic conditions have led to villages being inadequately protected. For such villages, we rationally plan and restore them without damaging the overall form, improving their livability and environmental quality, and enhancing living comfort. At the same time, by introducing advanced technology and modern concepts, we nurture new vitality. We use the rich historical culture and natural resources of the countryside to vigorously develop tourism, realize the dynamic and organic combination of traditional elements and modern functions, and pragmatically achieve "activated" utilization rather than passive preservation.

In some high-density areas, traditional villages are well preserved, but indicators such as population, river distribution, transportation network, and total GDP show that villages have issues such as population decline and reduced economic vitality. Based on the analysis of geographic detectors (interaction between GDP and population (q = 0.90); population and river buffer (q = 0.85); GDP and river buffer (q = 0.79); transportation conditions and population (q = 0.77)), it is evident that the spatial heterogeneity of villages is influenced by the interaction of comprehensive factors such as population, economy, and environment. When formulating protection strategies, the impact of comprehensive factors, including population density and economic environment, should be comprehensively considered.

Limitations

In addition, this study still has some shortcomings. For example, when using Kernel density estimation to analyze the macro spatial pattern of traditional village distribution, the default bandwidth of ArcGIS 10.8 was used, which utilizes the spatial variable of the "Silverman empirical rule" to calculate the default search radius. Although it is suitable for large-scale areas for macro pattern exploration with large sample sizes, it still has limitations: the lack of quantitative bandwidth comparison and verification may lead to insufficient local cognition. The next step will be to conduct a comparative analysis based on two commonly used quantitative methods in analyzing bandwidth, the Silverman rule of thumb and the cross-validation method, to further optimize the bandwidth selection. When applying the Moran's Index method for analysis, ROOK adjacency is used instead of other methods, such as Queen adjacency. Although this study investigates the formation mechanism of adjacent counties of traditional villages, counties are closed surface elements, and their spatial interaction depends on the actual geographical border relationship. The study selected the ROOK rule to construct a weight matrix through topological adjacency, which strictly limits "only counties that share boundaries as neighbors" to define the basis of spatial connection. However, there are still limitations: different methods should be selected for different research objectives. Compared with Queen adjacency, ROOK adjacency is more suitable for administrative

division characteristics, avoids interference from subjective distance parameters, and can ensure that the conclusions reflect the real boundary driving mechanism. This selection method is somewhat subjective, and when selecting a method, an in-depth analysis of the research objectives should be conducted to obtain more accurate conclusions.

Conclusion

This study systematically analyzes 310 "Chinese Traditional Villages" in the Beijing-Tianjin-Hebei region, expanding the understanding of traditional village research and exploring their spatial distribution and driving mechanisms. The following are the main conclusions and their connections to the global SDGs:

(1) The traditional villages in the Beijing-Tianjin-Hebei region are clustered around "three high-density areas" and "three medium-density areas," which are closely related to geographical, economic, and social factors. This distribution reflects the cultural diversity and ecological integrity of the region and is also an important manifestation of social cohesion. This echoes the vision of "Sustainable Cities and Communities" advocated by the United Nations SDG 11.

(2) By using a geodetector, the study revealed the significant impact of rainfall and per capita GDP on the spatial distribution of traditional villages. The interactions between these factors, such as total GDP and population ($q = 0.90$) and total GDP and rainfall ($q = 0.88$), indicate that the synergistic effect of economic conditions and natural resources has a decisive influence on village distribution. The study used the GWR model to reveal the significant spatial heterogeneity of the distribution of traditional villages and their influencing factors. For example, the positive impact of temperature and tourism resources on village distribution, and the negative correlation between population and GDP, the local differences in these factors emphasize the need for region-specific strategies. This multi-dimensional interaction mechanism not only provides a new perspective for understanding the spatial distribution of traditional villages, but also provides a scientific basis for formulating regionally targeted protection and development strategies.

(3) The study provides a broader perspective for understanding and addressing the challenges of globalization and urbanization through comparisons with the Italian Alps and rural India. The effectiveness of this approach highlights its potential application in rural conservation and development in other developing countries.

The research results not only provide specific strategies for village protection in the Beijing-Tianjin-Hebei region but also provide a reference for sustainable development and cultural heritage protection in similar regions around the world.

Data availability

The datasets used and/or analyzed in the current study are not publicly available because this experiment was done collaboratively and the results belong to the team, but are available on demand from the corresponding author.

References

1. Yan, P., Cheng, K. & Yang, J. T. Cultural Space of the Traditional Settlement in Hehuang Area. *Qinghai Huazhong Arch.* **39**, 117–121. https://doi.org/10.13942/j.cnki.hzjz.2021.07.024 (2021).
2. Cf, O. *Transforming our world: the 2030 Agenda for Sustainable Development* (United Nations, 2015).
3. Li, B. Z. & Yao, R. M. Urbanisation and its impact on building energy consumption and efficiency in China. *Renew. Energy* **34**, 1994–1998. https://doi.org/10.1016/j.renene.2009.02.015 (2009).
4. Wang, Y., Geng, H. & Ouyang, G. H. In: *Proc. 51st ISOCARP Congress.* (2015).
5. B., S. R. & B., H. J. Future Aging of Rural Population: Migration and Backflow. *Journal of Minzu University of China (Philosophy and Social Sciences Edition)* **51**, 113–124, https://doi.org/10.15970/j.cnki.1005-8575.2024.05.013 (2024).
6. Steinicke, E., Cede, P. & Löffler, R. IN-migration as a new process in demographic problem areas of the alps. ghost towns vs. amenity settlements in the alpine border area between Italy and Slovenia. *Erdkunde* **66**, 329–344. https://doi.org/10.3112/erdkunde.2012.04.04 (2012).
7. Huang, C. & Dong, C. Preservation and rehabilitation of the dwellings of Ohkay Owingeh traditional Indian village. *Indus. Constr.* **47**, 44–48 (2017).
8. Li, M. X., Yan, Y., Ying, Z. Y. & Zhou, L. Measuring Villagers' Perceptions of Changes in the Landscape Values of Traditional Villages. *Isprs Inter. J. Geo-Inform.* https://doi.org/10.3390/ijgi13020060 (2024).
9. Lin, Z. R., Liang, Y. & Liu, X. H. Study on spatial form evolution of traditional villages in Jiuguan under the influence of historic transportation network. *Heritage Sci.* https://doi.org/10.1186/s40494-024-01153-0 (2024).
10. Zhou, Y. C., Liu, M., Xie, G. H. & Liu, C. Q. Landscape Ecology Analysis of Traditional Villages: A Case Study of Ganjiang River Basin. *Appl. Sci.-Basel* https://doi.org/10.3390/app14020929 (2024).
11. Wang, D. G., Zhu, Y. J., Zhao, M. F. & Lv, Q. Y. Multi-dimensional hollowing characteristics of traditional villages and its influence mechanism based on the micro-scale: A case study of Dongcun Village in Suzhou China. *Land Use Policy* https://doi.org/10.1016/j.landusepol.2020.105146 (2021).
12. Fang, Y. G. & Liu, J. S. Cultural landscape evolution of traditional agricultural villages in North China-Case of Qianzhai Village in Shandong Province. *Chin. Geogra. Sci.* **18**, 308–315. https://doi.org/10.1007/s11769-008-0308-x (2008).
13. Zhang, L., Zhou, S. Y. & Guo, Y. L. Study on the evolution of the spatial structure and driving force of traditional village tourism in South Anhui province. *Front. Earth Sci.* https://doi.org/10.3389/feart.2024.1461292 (2024).
14. Ge, H. Y. et al. Study on space diversity and influencing factors of Tunpu settlement in central Guizhou Province of China. *Heritage Sci.* https://doi.org/10.1186/s40494-022-00719-0 (2022).
15. Dandan, S. & Kyungjin, Z. Analysis of spatial distribution characteristics and driving factors of ethnic-minority villages in China using geospatial technology and statistical models. *J. Mt. Sci.* **21**, 2770–2789. https://doi.org/10.1007/s11629-023-8164-9 (2024).
16. Ma, H. D. & Tong, Y. Q. Spatial differentiation of traditional villages using ArcGIS and GeoDa: A case study of Southwest China. *Ecol. Inform.* https://doi.org/10.1016/j.ecoinf.2021.101416 (2022).

17. Li, C. X. & Wu, K. N. Driving forces of the villages hollowing based on geographically weighted regression model: a case study of Longde County, the Ningxia Hui Autonomous Region China. *Nat. Hazards* **89**, 1059–1079. https://doi.org/10.1007/s11069-017-3008-y (2017).
18. Lu, X. X., Peng, Z., Zhou, Y. C., Xie, Y. Q. & Chen, Z. J. A study on the spatial distribution characteristics and driving factors of traditional villages in the southeast coast of China: A case study of Puxian. *Fujian. Plos One* https://doi.org/10.1371/journal.pone.0303746 (2024).
19. Guo, Y. L., Rao, J. Y., Huang, J. & Zhu, Y. L. Spatial and Temporal Evolution and Conservation Significance of Traditional Villages in the Yangtze River Basin China. *Buildings* https://doi.org/10.3390/buildings14103249 (2024).
20. Chen, W. X. et al. Identifying the spatial differentiation factors of traditional villages in China. *Heritage Sci.* https://doi.org/10.1186/s40494-023-00994-5 (2023).
21. Long, T. X., Isik, C., Yan, J. L. & Zhong, Q. K. Promoting the sustainable development of traditional villages: Exploring the comprehensive assessment, spatial and temporal evolution, and internal and external impacts of traditional village human settlements in hunan province. *Heliyon* https://doi.org/10.1016/j.heliyon.2024.e32439 (2024).
22. Liu, S., Bai, M., Yao, M. & Huang, K. Identifying the natural and anthropogenic factors influencing the spatial disparity of population hollowing in traditional villages within a prefecture-level city. *PLoS ONE* https://doi.org/10.1371/journal.pone.0249448 (2021).
23. Chen, X. H., Xie, W. Z. & Li, H. B. The spatial evolution process, characteristics and driving factors of traditional villages from the perspective of the cultural ecosystem: A case study of Chengkan Village. *Habitat Inter.* https://doi.org/10.1016/j.habitatint.2020.102250 (2020).
24. Xu, J., Yang, M. S., Hou, C. P., Lu, Z. L. & Liu, D. Distribution of rural tourism development in geographical space: a case study of 323 traditional villages in Shaanxi China. *Eur. J. Remote Sens.* **54**, 318–333. https://doi.org/10.1080/22797254.2020.1788993 (2021).
25. Huang, Y. H. & Xue, Q. M. Spatio-Temporal distribution characteristics and driving factors of traditional villages in the Yellow River Basin. *PLoS ONE* https://doi.org/10.1371/journal.pone.0303396 (2024).
26. Nie, Z. Y., Chen, C., Pan, W. & Dong, T. Exploring the Dynamic Cultural Driving Factors Underlying the Regional Spatial Pattern of Chinese Traditional Villages. *Buildings* https://doi.org/10.3390/buildings13123068 (2023).
27. Y., Z. D. Research on Coordinated Preservation and Development of Traditional Villages in Beijing-Tianjin-Hebei Region. *J. Beijing Univ. Civil Eng. Arch.* **33**, 1–5 (2017).
28. Lefever, D. W. Measuring Geographic Concentration by Means of the Standard Deviational Ellipse. *American Journal of Sociology* **32**, 88–94. http://www.jstor.org/stable/2765249 (1926).
29. Gatrell, A. C., Bailey, T. C., Diggle, P. J. & Rowlingson, B. S. Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Trans. Inst. Br. Geogr.* **21**, 256–274. https://doi.org/10.2307/622936 (1996).
30. Lorenz, M. O. Methods of Measuring the Concentration of Wealth. *Publ. Am. Stat. Assoc.* **9**, 209–219. https://doi.org/10.2307/2276207 (1905).
31. Gini, C. Measurement of Inequality of Incomes. *Econ. J.* **31**, 124–126. https://doi.org/10.2307/2223319 (1921).
32. Moran, P. A. P. Notes on Continuous Stochastic Phenomena. *Biometrika* **37**, 17–23. https://doi.org/10.2307/2332142 (1950).
33. Tobler, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **46**, 234–240. https://doi.org/10.2307/143141 (1970).
34. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **33**, 1065–1076 (1962).
35. Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London* **222**, 309–368. http://www.jstor.org/stable/91208 (1922).
36. Brunsdon, C., Fotheringham, A. S. & Charlton, M. E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **28**, 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x (1996).
37. Wang, J. F. & Xu, C. G. Geodetector: Principle and prospective. *Acta Geogr. Sin.* **72**, 116–134 (2017).
38. Xiong, C. S., Tan, R. & Qiu, W. Z. Zoning of high standard farmland construction based on local indicators of spatial association. *Trans. Chin. Soc. Agri. Eng.* **31**, 276–284 (2015).
39. Mu, H. B. T. P. R. T. & Zhao, Y. Dynamic Analysis of Mechanism of Financial Development and Urbanization on Urban-Rural Income Gap. *Stat. Dec.* https://doi.org/10.1354/j.cnki.tjyjc.2018.05.021 (2018).
40. Bai, B. & Zhao, Z. Q. Measuring the spatial concentration in China's manufacturing: A multi-elliptical model. *Sci. Geogr. Sinica* **44**, 50–58. https://doi.org/10.1324/j.cnki.sgs.20220658 (2024).
41. Dong, Y. H. & Wang, X. L. Method of River Segmentation with Five Subzones and Its Application to the Changjiang River Segmentation. *J. Yangtze River Sci. Res. Inst.* **34**, 1–6 (2017).
42. Hu, X. J., Wu, J. D., Ye, Z. H. & Su, M. Investigation and analysis of influencing factors of the "2012.7.21" rainstorm and flood disaster in Beijing. *China Flood and Drought Management* **22**, 1–3+30, https://doi.org/10.16867/j.cnki.cfdm.2012.06.001 (2012).
43. Huo, Z. G. et al. Impacts of Climate Warming on Crop Diseases and Pests in China. *Sci. Agri. Sin.* **45**, 1926–1934 (2012).
44. Wu, S. F., Wu, P. T., Feng, H., Niu, W. Q. & Zhang, H. Fuzzy Analysis on Rainfall Adaptability to Crop Growth in Yangling Area. *J. Irrigation Drainage* https://doi.org/10.1352/j.cnki.ggps.2002.01.006 (2002).
45. Shang, G. Z. H. The Impact and Mechanism of Population Urbanization on China's GDP. *J. South-Central Minzu Univ.* https://doi.org/10.1989/j.cnki.42-1704/C.20240516.01 (2024).
46. Ma, M. Y. & Tang, J. X. Nonlinear impact and spatial effect of tourism urbanization on human settlement environment: Evidence from the Yellow River Basin China. *J. Cleaner Prod.* https://doi.org/10.1016/j.jclepro.2023.139432 (2023).
47. Tourism, H. P. D. o. C. a. *The List of A-Grade Tourist Attractions in Hebei Province*, <https://whly.hebei.gov.cn/c/2025-04-23/576582.html> (2025).
48. Tourism, T. M. B. o. C. a. *The List of A-Grade Tourist Attractions in Tianjin Municipality*, <https://whly.tj.gov.cn/> (2025).
49. Tourism, B. M. B. o. C. a. *The List of A-Grade Tourist Attractions in Beijing Municipality*, <https://whlyj.beijing.gov.cn/> (2025).
50. Habibulloev, S., Han, F., Bakhtiyorov, Z., Xuankai, M. A. & Nuhzor, O. Factors influencing sustainable development in Eco-tourism settlements: A comparative analysis. *Heliyon* https://doi.org/10.1016/j.heliyon.2024.e26454 (2024).
51. Wei, X. & Quan, T. The low-carbon influencing mechanism of spatial characteristic parameters in China's residential communities. *J. Asian Arch. Building Eng.* **20**, 88–100. https://doi.org/10.1080/13467581.2020.1800470 (2021).
52. Oshan, T. M., Li, Z. Q., Kang, W., Wolf, L. J. & Fotheringham, A. S. MGWR: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *Isprs Inter. J. Geo-Inform.* https://doi.org/10.3390/ijgi8060269 (2019).
53. Li, T., Li, C. K., Zhang, R., Cong, Z. & Mao, Y. Spatial Heterogeneity and Influence Factors of Traditional Villages in the Wuling Mountain Area, Hunan Province, China Based on Multiscale Geographically Weighted Regression. *Buildings* https://doi.org/10.3390/buildings13020294 (2023).
54. Liu, C. & Xu, M. Characteristics and Influencing Factors on the Hollowing of Traditional Villages-Taking 2645 Villages from the Chinese Traditional Village Catalogue (Batch 5) as an Example. *Inter. J. Environ. Res. Public Health* https://doi.org/10.3390/ijerph182312759 (2021).
55. Zhang, Y., Baimu, S., Tong, J. & Wang, W. Geometric spatial structure of traditional Tibetan settlements of Degger County, China: A case study of four villages. *Front. Arch. Res.* **7**, 304–316 (2018).
56. Zhang, L. L. et al. Impacts of Spatial Components on Outdoor Thermal Comfort in Traditional Linpan Settlements. *Inter. J. Environ. Res. Public Health* https://doi.org/10.3390/ijerph19116421 (2022).

57. Yang, R., Liu, Y. S., Long, H. L. & Qiao, L. Y. Spatio-temporal characteristics of rural settlements and land use in the Bohai Rim of China. *J. Geog. Sci.* **25**, 559–572. https://doi.org/10.1007/s11442-015-1187-6 (2015).
58. Chen, K. Y. Analysis on the Formation and Inheritance of Traditional Characteristic Villages: Reflection on the Field Investigation of Xijiang Qianhu Miao Village in Guizhou and Zhaji Ancient Village in Anhui. *Urbanism Arch.* **20**, 82–85. https://doi.org/10.19892/j.cnki.csjz.2023.09.19 (2023).
59. Yu, Z. Y. & Zhao, S. Research on Spatial Characteristics and Influence Mechanism of Traditional Villages and Intangible Cultural Heritage in Southwest China. *Areal Res. Devel.* **43**, 90–96 (2024).
60. Fang, Y. L., Lu, H. Y., Huang, Z. F. & Zhu, Z. G. Spatiotemporal Distribution of Chinese Traditional Villages and Its Influencing Factors. *Econ. Geog.* **43**, 187–196. https://doi.org/10.1595/j.cnki.jjdl.2023.09.020 (2023).
61. Miao, Y. K., Luo, P. J. & Chang, J. Application of GIS in Chinese Traditional Village Research. *Indus. Constr.* https://doi.org/10.13200/j.gyjzG20011801 (2021).
62. Ju, X. X. et al. Spatial Distribution Characteristics and Influencing Factors of Traditional Villages in Zhejiang, Anhui, Shaanxi Yunnan Provinces. *Econ. Geograph.* **42**, 222–230. https://doi.org/10.15957/j.cnki.jjdl.2022.02.024 (2022).

Acknowledgements

Author contributions

PP and YC: Writing. PP: review and editing; writing—original draft; supervision; resources; project administration. PP and YC: Visualization, Software. JJ and YC: Investigation, formal analysis.YR, LM, and JY: Data curation, conceptualization. All authors reviewed the manuscript.

All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.