RESEARCH

Spatial heterogeneity analysis on distribution patterns and underlying determinants of historical rural settlements: a case study in Taihang Baxing

Sinan Yuan[1,2†], Nanshuo Mu[1,2†], Jinni Bai[1,2] and Han Zhang[1,2*]

Abstract

As vital carriers of cultural heritage, rural areas are facing the threats of degradation and urbanization during rapid urbanization processes. In recent years, China has formulated and implemented a series of national policies, regulations, and heritage protection lists to safeguard and perpetuate the traditional patterns and historical features of rural heritage. However, these rural heritage listings exhibit significant macro-level distributional imbalances and generally lack consideration of regional characteristics. This study uses the Taihang Baxing (THBX) in China as a case study, initially employing historical local documents and maps as base data. It utilizes the stability in the evolution of Chinese characters to compare historical and contemporary village names, filtering out existing historical rural settlements (HRS) to construct a dataset. Further, a multi-scale geographically weighted regression (MGWR) model is applied to analysis the impacts of natural, economic, and social factors on the concentration of HRS (CHRS). The study reveals that the CHRS in the region is markedly characterized by clustering, with evident spatial heterogeneity in the impacts of various factors. Slope (SLP), total annual precipitation (TP), and potential crop yield (PCY) significantly influence the CHRS globally, while elevation (ELV), distance to ancient path (DAP) and surface runoff (DSR), gross domestic product (GDP), and road density (RD) have significant local impacts. This research offers a methodological framework for HRS data compilation and provides theoretical and methodological references for national and local levels of rural heritage protection, optimization of rural spatial patterns, prediction of rural evolution pathways, and implementation of China's rural revitalization strategy.

Keywords Historical rural settlements, Rural heritage, Distribution patterns, Spatial heterogeneity, MGWR

[†]Sinan Yuan and Nanshuo Mu have contributed equally to this work.

*Correspondence:
Han Zhang
zhanghan0122@tju.edu.cn
[1] School of Architecture, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, China
[2] Key Laboratory of Information Technology for Architectural Cultural Heritage, Ministry of Culture and Tourism, No. 92 Weijin Road, Nankai District, Tianjin 300072, China

Introduction

Rural settlements carry a myriad of critical functions related to the production and daily lives of rural residents, emerging from the long-term interaction between humans and their natural environments [1, 2]. Over time, these interactions have imbued rural settlements with archaeological values reflecting historical, cultural, and social significance [3]. The rapid development of the global economy has induced profound changes in urban-rural dynamics, leading to a global phenomenon of rural decline and increasing concern for the simultaneous erosion of rural heritage. In 2017, ICOMOS proposed that

Yuan *et al. Heritage Science*   (2024) 12:395

Page 2 of 22

all rural areas could be interpreted as heritage sites and encouraged nations to establish rural heritage inventories at various levels to enhance the protection of these heritages [4]. As a major agricultural country with a large rural population and extensive geographic area, China has approximately 580 million people distributed across millions of undeveloped villages. Since the economic reforms initiated in 1978, China has accumulated a vast and diverse array of rural heritages [5]. During this period, China's urbanization has continuously advanced, marked by a rapid urban-rural transformation. The urbanization rate increased from 17.9% in 1978 to 63.9% in 2020 [6]. Under the development philosophy of "strong cities, weak countryside," rural settlements generally face challenges such as population outflow, resource scarcity, and economic frailty [7, 8]. Issues such as depopulation and aging within rural residential areas have become prevalent in China, posing challenges to the preservation and development of rural heritage [9].

In 2012, China initiated the "Traditional Villages Survey," introducing the concept of Traditional Chinese Villages (TCV) to better protect China's rural heritage. TCV are defined as villages that were established early on, possess a wealth of traditional resources, and have significant historical, cultural, scientific, artistic, social, and economic values worthy of preservation [10]. The launch of the TCV survey sparked extensive research on rural settlements in China, accumulating significant findings in areas such as spatial form [11], landscape imagery [12], value assessment [13], archaeological preservation [14, 15], and the impacts of tourism development [16–18]. Additionally, numerous studies have analyzed the spatial distribution patterns of TCV at the regional level, summarizing their characteristics and identifying key areas for heritage conservation [19], leading to contributions to recognizing and safeguarding important rural heritage conservation areas [20].

Numerous studies have begun to focus on how different location conditions influence the distribution patterns of rural heritage, summarizing the underlying determinants. The migration and evolution of rural settlements are influenced by a complex interplay of natural, economic, and social factors [1, 21]. Understanding these intricate coupling relationships remains one of the key challenges in the field of rural settlement research [22]. The exploration of large-scale, selforganizing variable factors in the evolution of rural settlement patterns has become a necessary component in achieving the revitalization of rural China. This encompasses both theoretical inquiries and practical applications aimed at enhancing rural areas' sustainability and cultural integrity [23, 24].

While the TCV directory has generated significant discourse on rural heritage, research into the macro distribution and influencing factors of TCV remains subject to further scrutiny. The selection mechanism and process for TCV impose limitations that result in blind spots in their spatial distribution, leading to significant local clustering and overall extreme imbalance [25]. Studies on the macro spatial distribution focusing on the TCV can only to a limited extent reveal the correlation between the distribution characteristics of the currently listed TCV and influencing factors such as the current natural environment and social conditions. Some studies have begun to focus on historical rural settlements (HRS) with larger sample sizes as their research subjects. These efforts focus on studying spatial distribution and influencing factors, aiming to accurately reconstruct the real state of rural heritage distribution on a national or regional scale, and to elucidate the relationships with specific external environmental conditions [26–28]. The findings from these dataset construction efforts reveal that China still has a vast amount of valuable rural heritage that has not yet been fully explored. However, due to the difficulty in obtaining historical data, acquiring a more comprehensive database of HRS samples has become one of the obstacles hindering current research on the macro distribution of rural heritage.

Furthermore, on the methodological level, the development of geographic information technology and geospatial big data has facilitated the quantitative and multi-scale analysis of distribution patterns and variable factors concerning rural heritage. The integration of various types of geospatial big data provides auxiliary support to field research in the study of rural heritage. Moreover, statistical analysis methods such as regression models [29], Geodetector [30, 31], and panel data models [32] have also been incorporated into rural heritage research. These studies have more broadly identified potential quantitative variables that could impact the distribution of rural settlements, deepening the understanding of geospatial big data in rural heritage research and preliminary explaining the underlying determinants of rural heritage distribution from a mathematical and logical perspective.

Rural settlements exhibit distinct spatial attributes, with their evolution driven by nonlinear dynamics and spatial non-stationary. Traditional quantitative analysis methods often overlook these characteristics, which can lead to results that deviate from reality [33]. Consequently, scholars have employed traditional Geographic Weighted Regression (GWR) models to study the spatial differentiation of traditional villages across various scales, including nationwide in China [34], the Wuling Mountain area [35], and the middle reaches of the Yangtze River urban agglomeration [36]. These studies have confirmed that the spatial distribution of rural heritage

is significantly influenced by geographic location, and the effectiveness of driving factors shows noticeable spatial heterogeneity [37]. However, a limitation of the GWR is its assumption that all variables operate at the same spatial scale. Factors influencing the spatial distribution of rural settlements may exhibit similarities at certain spatial scales, significant differences may arise beyond those scales [38]. While GWR can account for spatial non-stationary, it fails to capture the scale differences in spatial heterogeneity of various influencing factors [39]. The Multi-Scale Geographic Weighted Regression (MGWR) model addresses these shortcomings by not only analyzing spatial heterogeneity effectively but also by automatically adjusting the regression bandwidth for different influencing factors. This adjustment allows for a more accurate reflection of the spatial variations of these factors [37], thereby providing more reliable results in the analysis of spatial heterogeneity [40].

In conclusion, the increasingly refined integration of spatial datasets on rural heritage with spatial statistics facilitates a deeper understanding of rural heritage research and comprehensive conservation planning. In this study, HRS is defined as rural settlements established before the Republic of China era (1912 A.D.) that have been preserved to the present day. TCV falls within the definition of HRS, and therefore, TCV is also considered a component of HRS. The objective is to develop more objective and applicable methods for extracting HRS data and to broadly examine the spatial heterogeneity of different influencing variables. Using the Taihang Baxing (THBX) area as a case study, we analysis the spatial distribution characteristics of HRS, evaluate the impacts of underlying driving factors through correlation analysis and spatial regression models, and test for spatial heterogeneity.

The main research questions addressed in this study are as follows:

- How to objectively obtain geographic information about existing HRS?
- What are the distribution patterns and underlying determinants of HRS?
- What is the spatial heterogeneity of the roles played by variables on the distribution patterns of HRS?

By addressing these issues, the study aims to provide key insights into the distribution and determinants of rural heritage. These insights will offer guidance for incorporating historical and cultural heritage into regional development planning. Additionally, the findings will contribute a macro-level theoretical reference for national strategies aimed at rural revitalization, thus supporting efforts to preserve and enhance rural areas in the context of their unique historical and cultural values [3].

Materials and methods
Research framework
The research framework of this study is illustrated in Fig. 1. This study began by collecting and filtering geographic information on HRS, thereby creating a dataset for the THBX. Subsequent steps involved collecting various variables to test their validity for the analysis, culminating in the selection of the MGWR to analysis spatial heterogeneity. The findings related to these research questions are discussed in Sections "Analysis of HRS spatial distribution patterns", "Applicability analysis of selected variables", and "Analysis of spatial heterogeneity in the effects of variables". Finally, Section "Discussion" serves as a retrospective overview of the study, discussing the insights gained, the limitations encountered, and the prospects for future research.

Study area
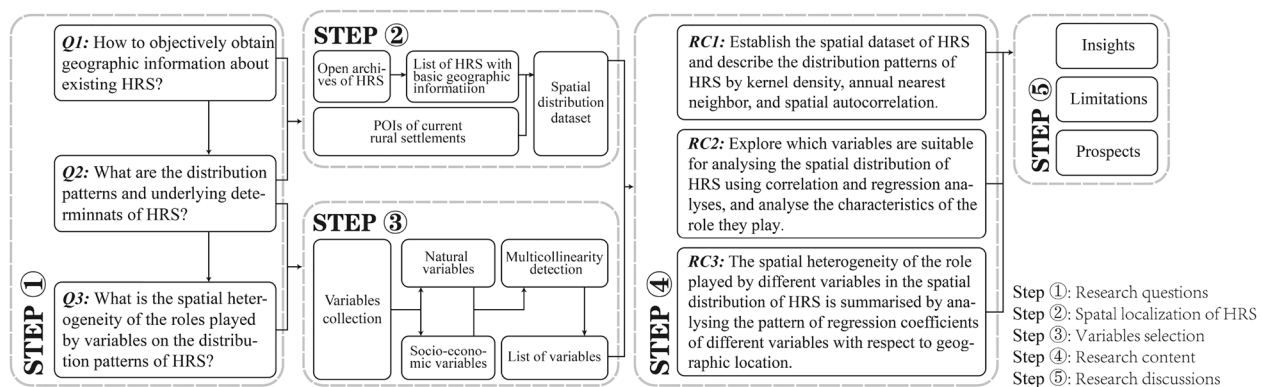The Taihang Mountains, located in northern China, extend from the northeast to the southwest and act as

Fig. 1 Research framework

Yuan *et al. Heritage Science*   (2024) 12:395

Page 4 of 22

the geographical boundary between the North China Plain and the Loess Plateau. This range also marks the transition zone between the second and third terraces of China's topography. Characterized by its complex terrain and challenging access, the region features eight valleys formed by river erosion, which serve as crucial roads connecting the east and west sides of the Taihang Mountains. Historically known as the Taihang Baxing (THBX), the word 'Xing' refers to ancient paths. These include, from north to south, the Jundu Xing, Puyin Xing, Feihu Xing, Jing Xing, Fukou Xing, Bai Xing, Taihang Xing, and Zhiguan Xing, as depicted in Fig. 2a. Alongside the Silk Road and the Grand Canal, the Taihang Baxing are an integral part of China's route-based cultural heritage [41, 42]. As history has progressed to the present, THBX no longer solely represents the eight ancient paths; it has come to represent the area where these paths are located, becoming a cultural symbol with regional attributes. The military, cultural, economic, and transport features that have emerged along the THBX have fostered the development of uniquely characteristic settlements clusters in this region.

As is shown in Fig. 2b, the overall distribution of TCV across the country is characterized by a higher concentration in the south than in the north, and more in the east than in the west. The THBX stands out as an important and unique high-density cluster of TCV in the north. This suggests that the settlement characteristics of this region are representative of northern Chinese settlements.

Consequently, as shown in Fig. 2c, the study area encompasses parts of Beijing, Hebei Province, Henan Province, and Shanxi Province, spanning four provinces and municipalities. The boundaries of the specific study area are primarily determined by the administrative boundaries, covering a total of 120 county-level administrative units. Notably, 11 of these counties have been selected for the national list of "Demonstration Counties for Concentrated and Continuous Protection of Traditional Villages," which ensures the effectiveness of our subsequent HRS information extraction efforts [43–45].

Collection of HRS information and geographic location
The initial step of this study involves proposing a method for selecting HRS within the THBX, focusing on compiling an HRS dataset along with corresponding geospatial information. Figure 3 illustrates the workflow for capturing the spatial locations of HRS. The criteria for selecting
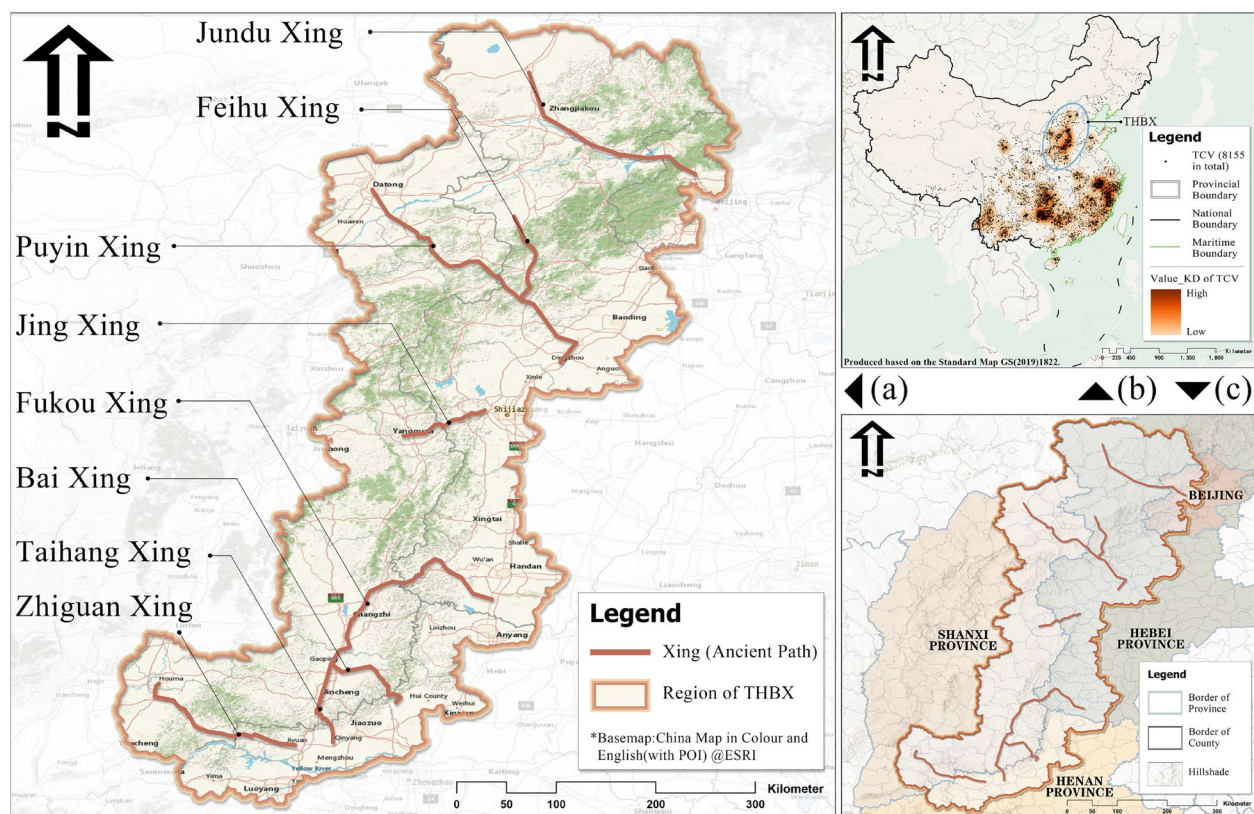

Fig. 2 a Profile of THBX (including the position of Eight Ancient Paths). **b** Provinces involved in THBX. **c** Distribution of TCV nationwide
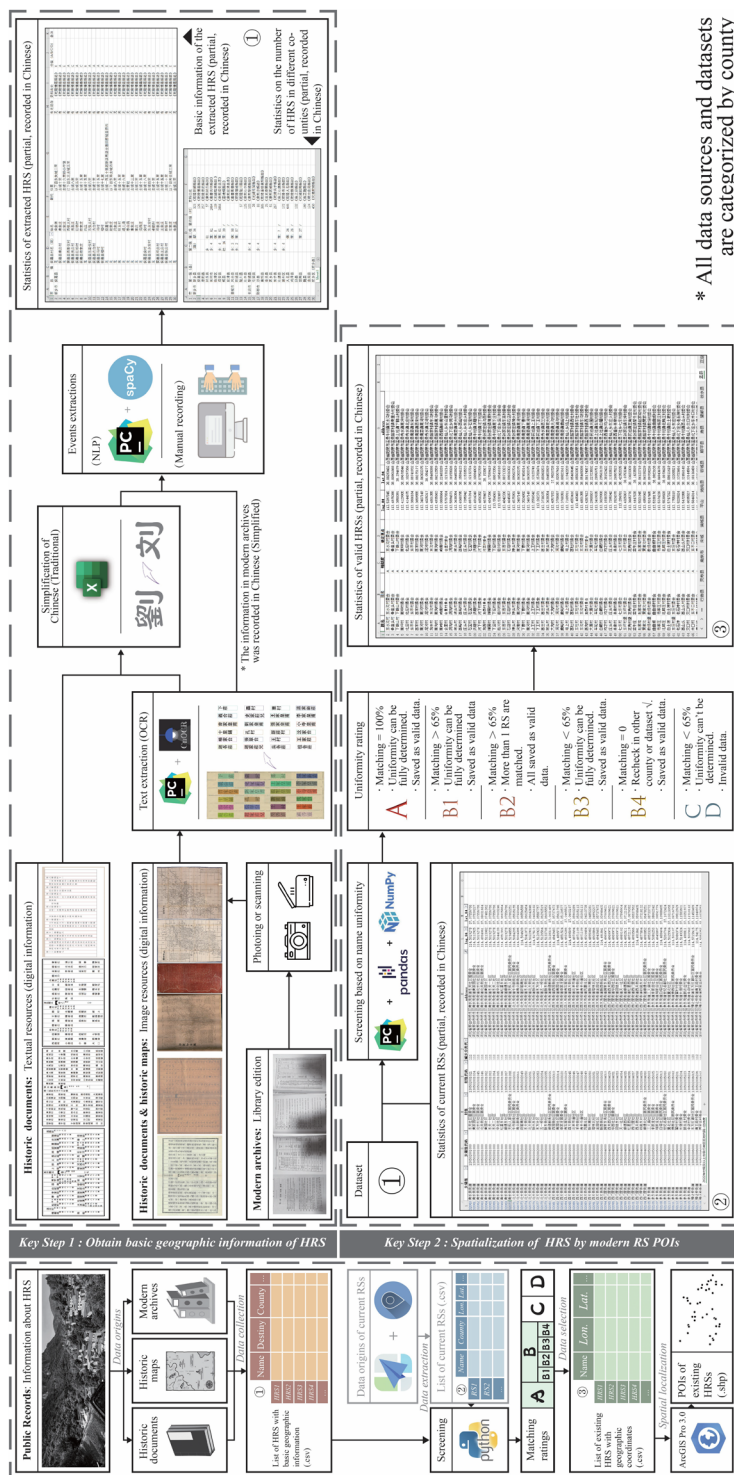
Yuan *et al. Heritage Science*   (2024) 12:395

Page 5 of 22

Fig. 3 Technical route for obtaining HRS spatial positioning and the details of the key steps

Yuan *et al. Heritage Science*   (2024) 12:395

Page 6 of 22

HRS are defined as rural settlements established prior to the founding of the Republic of China (1912 A.D.) that have persisted to the present day. Research in the field of historical toponymy has indicated that place names serve as linguistic evidence of a region's historical civilization, providing a link between the present and the past. The inheritance displayed by place names supports archaeological and heritage studies [46]. In China, a Chinese character has a more direct connection to its represented meaning compared to an English word and has maintained a stable correspondence between symbols through historical evolution (from traditional to simplified Chinese). Therefore, by matching recorded HRS names with the names of contemporary rural settlements, it is possible to identify existing HRS [47]. This method leverages the linguistic and historical continuity inherent in place names to bridge historical and contemporary data, enhancing the accuracy and depth of rural heritage studies.

Three types of historical information data are accessible for research: historical documents, historical maps, and modern documents, all formally recorded and categorized by administrative divisions. Martínez de la Fuente has demonstrated that historical documents contain crucial information on HRS and possess significant potential for studies in spatial distribution [28]. After collection and organization, the historical data for counties encompassed by the THBX and their associated historical administrative regions can be categorized into image data and text data. For image data, Optical Character Recognition (OCR) systems analysis the visual patterns of text characters within document images, converting them into machine-readable text. The CnOCR package, a Python-based tool, excels in recognizing and extracting Chinese textual information, thus it is employed to extract text from image data [48]. To facilitate smoother expression in subsequent research, the extracted traditional Chinese text is converted into simplified Chinese. Following this, the SpaCy package, a tool within the Natural Language Processing (NLP) domain, is utilised to extract place names associated with HRS from the entire text corpus, and these are cataloged by county [49]. The text produced through automated processes undergoes a further stage of manual proofreading to ensure its accuracy and reliability. This integrated approach of combining advanced computational tools with meticulous human verification ensures the integrity and usability of the data for further analysis.

For the information and geographical coordinates of modern rural settlements, the publicly accessible POIKit tool was utilized to extract Points of Interest (POI) data from Amap, including place names [50, 51]. Subsequently, using the Pandas and Numpy packages in Python

[52, 53], the place names from the HRS dataset were matched with those of modern villages to determine the precision of the matches. Since the information for HRS is aggregated by county, the data for modern rural settlements were also filtered by county during the matching process. Following this, the matching results were graded, and entries classified as Grade A or Grade B were saved as valid data. Finally, the HRS dataset was processed using ArcGIS Pro (Version 3.0.1) to determine the spatial locations of HRS, thus enabling subsequent spatial analysis. This approach not only ensures the accuracy of the data integration but also provides a robust foundation for analyzing the geographical distribution and continuity of historical and contemporary rural settlements.

Ultimately, we identified 15,208 existing HRS within the THBX, each equipped with detailed geographical information, including latitude and longitude (Counties table (with the number of HRS) included in THBX is in Appendix 1). The number of HRS identified significantly exceeds the number of TCV listed in the national catalog for the THBX, which totals 778. Subsequently, using ArcGIS Pro, we created a point dataset for HRS and plotted their distribution based on the standard map approved by the Ministry of Natural Resources of China, with the approval number GS(2019)1822, as shown in Fig. 4.

Analytical methods
Spatial analysis
Traditionally, methods for describing the spatial distribution of rural settlements typically employ administrative regions as controlling variables, calculating the density of rural settlements within each administrative area [54].
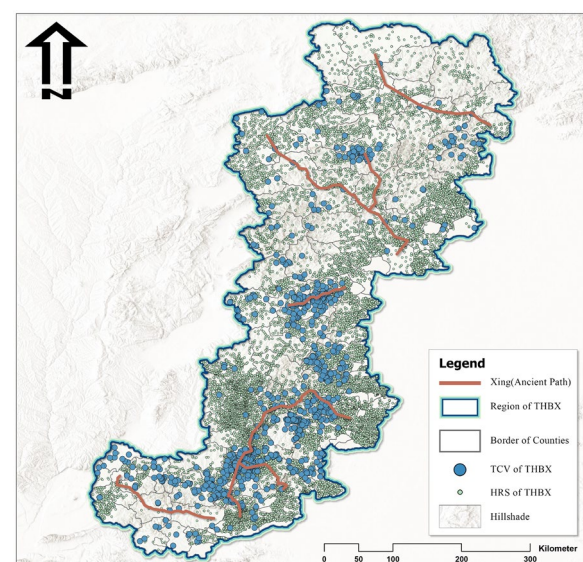

Fig. 4 Spatial dataset of HRS in THBX (compared with TCV)

Yuan *et al. Heritage Science*   (2024) 12:395

Page 7 of 22

However, this approach is unsuitable for characterizing the CHRS, as HRS emerged during a period characterized by central authority's limited control and reach. The appearance of HRS often depends on specific geographic conditions being favorable, and their development demonstrates marked adaptability and self-organizing characteristics. Therefore, this study opts for kernel density estimation to represent the CHRS [54]. Moreover, to further determine the spatial distribution patterns of HRS, the Average Nearest Neighbor (ANN) method is utilized to assess whether HRS are clustered, randomly distributed, or dispersed, analyzing from both a global and topological perspective [55].

Quantitative analysis of the underlying determinants spatial distribution patterns will focus on the mathematical relationships between various variables. This study requires the extraction of results from kernel density analyses conducted at HRS locations to develop variables that describe spatial distribution. Following the extraction, testing for spatial dependency of the data is essential. Spatial autocorrelation analysis aids in understanding the latent structures within spatial datasets, thereby detecting spatial dependencies. The global Moran's I index assesses the extent to which similar attribute values tend to cluster together or display spatial dispersion, providing a comprehensive measure of spatial autocorrelation across the entire study area. On the other hand, the local Moran's I employs a more nuanced approach to identify clusters of high or low attribute values within specific areas, revealing spatial hot-spots, cold spots, and outliers. Concurrently, Local Indicators of Spatial Association (LISA) maps facilitate the visual interpretation of these characteristics [56].

Statistical analysis
HRS exhibit heterogeneity in their spatial distribution determinants and are influenced by geographic location changes [57]. This study employs correlation and regression analyses to discuss the impact of various variables on the spatial distribution patterns of HRS, while also accounting for changes in spatial locations. The Pearson correlation coefficient quantifies the strength and direction of the linear relationship between two continuous variables, thus it is utilized to interpret the numerical relationships among variables and the spatial patterns of HRS distribution [54]. In Geodetector, the factor detector and interaction detector use spatial statistical methods to test the influence weight of variables on the spatial distribution patterns of HRS and to detect the impact of combinations of two variables on this proportion [58].

Regression analysis is better suited for exploring the relationships among multiple variables. The early Ordinary Least Squares (OLS) is a global regression model that assumes constant relationships between variables across the entire study area, which is ineffective at capturing spatial variations [59]. GWR addresses this limitation by allowing relationships to vary locally, but it still uses a single bandwidth to smooth spatial data, which may not fully capture multi-scale processes [60]. Therefore, this study adopts MGWR, allowing the analysis of relationships across multiple spatial scales to more accurately understand the spatial heterogeneity of different variables' impacts [61]. The software version used for MGWR is 2.2. Additionally, to confirm the applicability of MGWR, the results from OLS and GWR analyses are used as supplementary comparisons.

Selections of variables and data sources
Variable collection
In the context of correlation and regression analyses, the outcomes of kernel density analysis for HRS are employed as the dependent variable, denoted as the concentration of HRS (CHRS). This study further investigates the independent variables that exert an influence on this measure. The spatial distribution of rural settlements is affected by both physical geography and socioeconomic factors [62, 63]. Table 1 displays the results of the collected independent variables, including detailed information.

Firstly, terrain has a controlling effect on the selection of rural settlement sites [64]. From a macro perspective, elevation (ELV), slope (SLP), and topographical relief (TR) determine the fundamental conditions for regional agricultural production and transportation [64, 65].

Secondly, agricultural production activities influenced by regional climate and production resources subsequently affect the spatial distribution of rural settlements [66]. Temperature and precipitation are basic indicators of regional climate. Research from Guangjin has shown that annual average temperature (AAT) and total annual precipitation (TP) significantly impact the distribution of rural settlements [67], with heterogeneity across different regions; thus, they are included as independent variables. Relevant data often come from suitable open-access spatial datasets. As for production resources, this study focuses on the influence of water resources and arable land resources [68] on the distribution of rural settlements. Access to water sources is crucial for drinking, irrigation, and other daily uses. Fertile soil enhances agricultural productivity, which is the primary economic activity of rural settlements, ensuring food production and agricultural trade.

Thirdly, the current state of regional development also determines the macro distribution and individual development of rural settlements. Population density (PD), gross domestic product (GDP), and road density

Yuan *et al. Heritage Science*   (2024) 12:395

Page 8 of 22

Table 1 Description and pre-processing of variables

Variable types	Variable categories	Variable names	Variable description	Abbreviations	Units	Sig.	VIF
Dependent variable	–	Concentration of HRS	Density of HRS within a specified area.	CHRS	HRS/km^2	–	–
Independent variable	Natural indicators	Elevation	The average elevation of the entire HRS area.	ELV	m	0.000	8.869
		Slope	The average slope within each $30m^2$ area across the entire HRS region.	SLP	°	0.000	*14.046*
		Topographical relief	Change in elevation per km^2.	TR	m/km^2	0.000	*14.063*
		Total precipitation	Total precipitation sum over 12 months within a specified area.	TP	mm	0.000	9.206
		Annual average temperature	Average temperature over 12 months within a specified area.	AAT	°F	0.000	1.327
		Distance from the Ancient Paths	Minimum Euclidean distance from HRS to the Ancient Path.	DAP	m	0.000	1.227
		Distance from surface runoff	Minimum Euclidean distance from HRS to surface runoff.	DSR	m	0.000	1.122
	Socio-economic indicators	Potential crop yield	Maximum yield of cereal crops that can be produced per hectare of land annually.	PCY	kg/ha	0.000	2.273
		Gross domestic product	Gross Domestic Product(GDP) per km^2.	GDP	$10,000CNY/km^2$	0.000	3.878
		Population density	Number of people per km^2.	PD	$Person/km^2$	0.000	4.449
		Road density	Total length of roads per km^2.	RD	km/km^2	0.000	2.259
		Concentration of large cities	Density of large cities within a specified area.	CLC	LC/km^2	0.000	2.627
		Concentration of cultural heritage units	Density of cultural heritage units within a specified area.	CCHU	CHU/km^2	0.000	1.580

(RD) respectively represent the community, economy, and transportation development of a region. Higher population density is usually associated with more developed infrastructure and services. Rural areas with higher GDP usually have better economic opportunities and development potential. As for road density, rural commerce and the flow of people often rely on it [5, 69, 70]. It should be noted that regional development can contribute to the development and retention of rural settlements, which is beneficial for the preservation of HRS. On the other hand, urbanization brought by development can also lead to the disappearance of rural settlements [71]. Furthermore, considering the unique influence of the eight ancient paths in the THBX area, this study also employs the distance to the ancient paths (DAP) to analyze their impact on HRS.

Finally, this study introduced the concentration of large cities (CLC) and the concentration of national cultural heritage units (CCHU) to further explore the potential risks of urbanization and over-development of HRS. Among them, large cities are defined as urban area point data of city-level administrative units. The spatial distribution dataset of national cultural heritage units can be publicly accessed.

Data sources

The quantitative data sources are shown in Table 2. Firstly, three composite data sources are introduced. Among them, the ASTER Global Digital Elevation Model is a global digital elevation data product jointly released by the National Aeronautics and Space Administration (NASA) and the Ministry of Economy, Trade, and

Table 2 Data source description

Name	Data references	Access time
ASTER Global Digital Elevation Model	[72]	2020
Climatic Research Unit Time-Series datasets version 4.07	[73]	2022
China GDP spatial distribution kilometer grid dataset	[74]	2020
National district and county point dataset	[50]	2022
Map of National Key Cultural Relics Protection Units	[75]	2019
OpenStreetMap dataset	[76]	2023
LandScan Global population dataset	[77]	2022
"Potential crop yield in China" dataset	[78]	2010

Yuan *et al. Heritage Science* (2024) 12:395

Page 9 of 22

Industry of Japan (METI), completed based on the observations of NASA's TERRA satellite. This data source can provide quantitative data for ELV, SLP, and TR of THBX. The Climatic Research Unit Time-Series (CRU TS) is one of the most widely used climate datasets, supported by the NERC Centres for Atmospheric Science, UK. CRU TS provides monthly data covering the land surface from 1901 to 2022 at a resolution of 0.5° (approximately 1 km). The quantitative data for AAT and TP of THBX in 2022 used in this study are provided by this dataset. OpenStreetMap (OSM) is a free and editable world map containing numerous data such as roads, waterways, and buildings, created and maintained by nearly 5 million registered users and over 1 million map contributors from around the world using free tools and software. The quantification of variables RD and DSR relies on the open data provided by OSM. Moreover, this study also employed various other data sources that support only individual variables. Specific citation information also can be found in Table 2.

Data pre-processing

Multicollinearity reduces the overall interpretability of a model; therefore, it is necessary to examine multicollinearity among variables before proceeding with further statistical analyses. The Variance Inflation Factor (VIF) provides a quantified metric that indicates how much a predictor variable's variance is inflated due to its linear relationship with other predictor variables. This metric assists in removing outliers from the data, thereby enhancing the interpretability and stability of the regression model [79]. A VIF value greater than 10 indicates significant multicollinearity, suggesting that the coefficient estimates may be unreliable. Variables with VIF values between 5 and 10 also warrant caution as they may still pose a mild risk of multicollinearity. Given that multicollinearity is a property of synergy among multiple variables, outlier variables often do not appear in isolation. Thus, potential multicollinearity can be mitigated by selectively removing or transforming these outlier variables and retesting.

The VIF diagnostics within this study revealed that the VIF values for SLP and TR were 14.046 and 14.063, respectively, indicating significant multicollinearity. The VIF values for ELV and AAT were also notable at 8.869 and 9.206, respectively. Indeed, the calculation of TR inherently reflects some aspects of the SLP characteristic of HRS, and variations in ELV also influence AAT. Consequently, TR and AAT were removed from the analysis, and VIF diagnostics were performed again. The results indicated that no multicollinearity issues remained among the remaining variables, thereby clearing the path for further statistical analysis.

Results

Analysis of HRS spatial distribution patterns

Using ArcGIS Pro for kernel density estimation, the CHRS in the THBX was mapped as shown in Fig. 5, allowing further exploration of spatial distribution patterns. The characteristics of HRS distribution at the regional level in THBX are as follows:

HRS are more densely distributed in the southern part of the THBX than in the north. Within the southern range of Fukou Xing—Taihang Xing, two prominent high-density areas emerge, with the highest value reaching 0.44 /km^2. There is a significant correlation between the spatial position of the Xing (Ancient Paths) and the high-density CHRS. The density of HRS is higher along the sides of the ancient paths than in the central part, which may be related to the central zone of the Taihang Mountains. The terrain of the Taihang Mountains is too steep, making it difficult to establish villages or protect them from natural disasters. Further analysis reveals that HRS on the western or northern endpoints of each Xing show a punctuated clustering distribution. Although high values are present throughout the THBX, there is a substantial difference between high and low values. On the other hand, the highly clustered areas of HRS on the eastern or southern endpoints exhibit a distinct striped or areal distribution, with a generally higher number and more uniform spread.

The results of ANN analysis, as displayed in Fig. 6a, show an nearest neighbor ratio of 0.649771 and a z-score of −82.626356, both indicating that the likelihood of a random CHRS is less than 1%, with clustering characteristics evident in the distribution. In THBX, the average
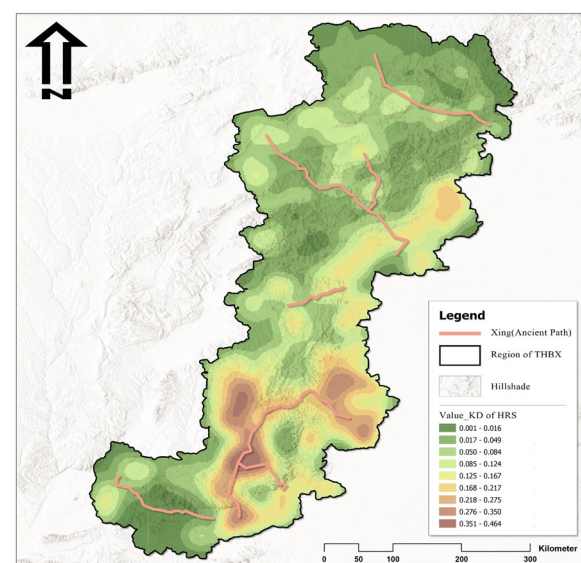

Fig. 5 Kernel density analysis results of HRS in THBX

Yuan *et al. Heritage Science*   (2024) 12:395

Page 10 of 22

observed distance among HRS is 1300.3771 ms, compared to an expected average distance of 2001.2842 ms. This clustering pattern of HRS suggests the importance and potential value of investigating underlying determinants in subsequent studies. However, the study of determinants focuses on the relationships between different variables. Therefore, it requires a statistically significant and quantified set of variables (in this study, the kernel density estimates of HRS in THBX are used as the dependent variable, CHRS). This approach allows for a deeper understanding of the factors influencing HRS distribution and contributes to more targeted and effective spatial planning and development strategies within the region.

In addition to the ANN analysis, spatial autocorrelation analysis was conducted to examine the kernel density values at locations of HRS [80]. From the global spatial analysis, the global Moran's I for the kernel density attribute of HRS in THBX is recorded at 0.967717 with a z-score of 66.006809, indicating a significant clustering trend in the overall distribution of this attribute value (Fig. 6b). Consequently, the kernel density attribute values of HRS in THBX fulfill the prerequisites for further spatial statistical investigations. The Local Indicators of Spatial Association (LISA) map, generated from the analysis of local spatial autocorrelation results (Fig. 6c), allows for more precise determination of the distribution types of HRS within THBX. Viewed across the entire region, most local Moran's I calculations for HRS show statistical significance. The predominant distribution patterns for HRS are high/high clustering and low/low clustering. Low/high and high/low clustering patterns are mainly found in transitional areas between the two aforementioned clusters. The findings from the spatial autocorrelation analysis indicate that the kernel density values of HRS exhibit spatial dependency, which validates their use as numerical values for CHRS in subsequent statistical analyses. This demonstrates the significance of considering spatial relationships to better understand and model the distribution patterns of rural settlements effectively.

Applicability analysis of selected variables
Univariate applicability assessment
This study initially explores the suitability of selected variables for explaining the CHRS in the THBX from three
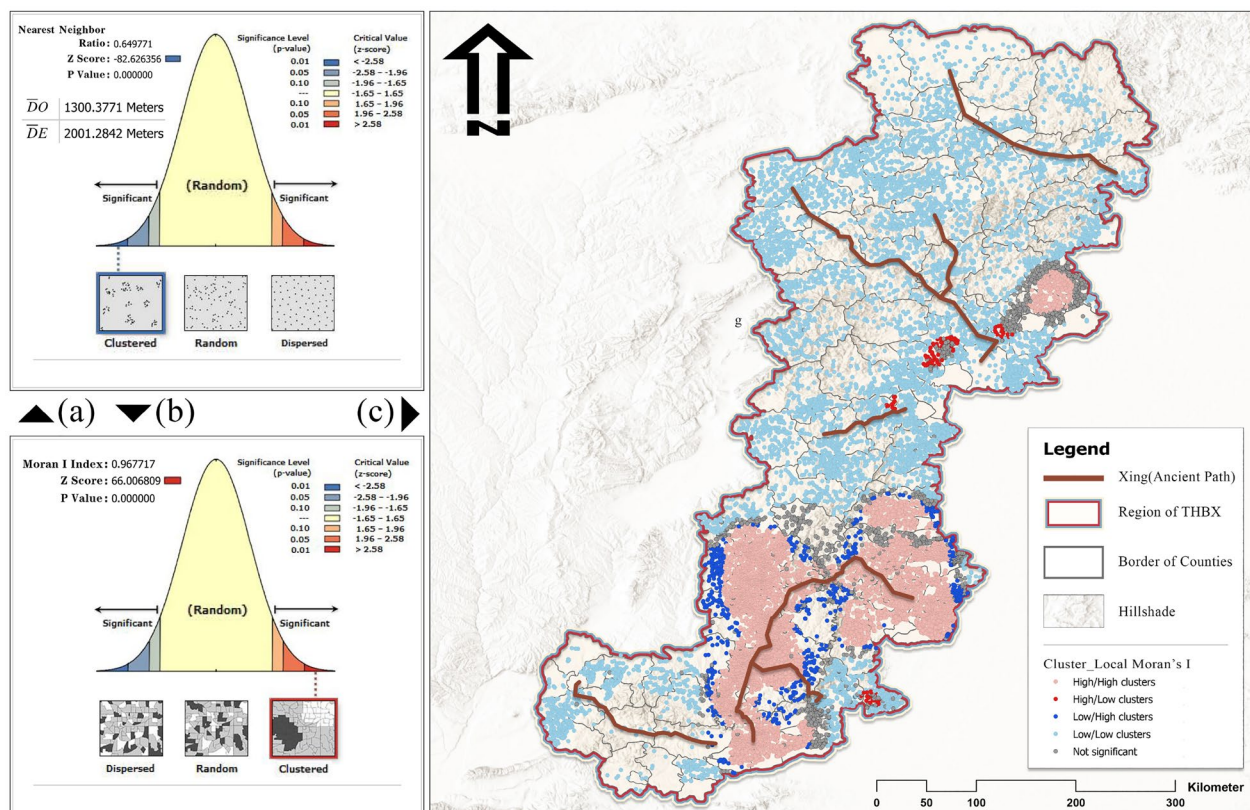

Fig. 6 **a** Average Nearest Neighbor analysis for HRS in THBX. **b** Global Moran's I analysis for HRS in THBX. **c** LISA clustering results of HRS in THBX based on Local Moran's I

perspectives (Fig. 7). Firstly, Pearson correlation analysis is employed to examine the similarity in numerical trends among variables. As depicted in Fig. 7a, under the influence of single factors, the driving effects of different variables vary. Significant correlations are widely present among the variables. Notably, all natural geographic variables show significant correlations with CHRS. However, the correlations between RD, CCHU and CLC with CHRS are not significant. Across the entire Pearson analysis, both higher and lower values of coefficients are found in the analysis of correlations among independent variables. The highest correlation coefficient between CHRS and the GDP is 0.29, and the lowest is −0.28 of DAP, with other coefficients having absolute values less than 0.25. Among them, ELV, SLP, DSR, and CLC exhibit negative correlations with CHRS, whereas TP, PCY, GDP, and RD show positive correlations. The correlation coefficients between the selected variables and CHRS in this study are too low to fully explain the relationships therein [81]. This suggests the need for further investigation, potentially incorporating additional variables or exploring more complex statistical models to capture the nuances of these relationships better.

Secondly, the factor detection tool of the Geodetector generates q-values, which indicate the proportion of the sample variance that an independent variable can explain (q * 100%). This can be interpreted as a method of expressing a variable's explanatory power or weight. The factor detection results, as shown in Fig. 7c, are significant for all variables. The variables with the highest influence weights are DAP at 0.34, CLC at 0.28, and CCHU at 0.32. Notably, the influence weight of CCHU is evidently stronger than what was observed in the correlation analysis. However, no variable is able to explain more than 50% of the CHRS.

This result highlights the complexity of factors influencing the CHRS in the THBX. The significant q-values suggest that while these variables have a measurable impact, the CHRS is likely influenced by a combination of multiple factors. The lack of any single variable explaining a majority of the variance indicates the potential need for a multifaceted approach in further research, integrating various socio-economic and environmental factors to gain a more comprehensive understanding of CHRS.

The results of the univariate analyses indicate that the influence of single variables alone is insufficient to fully explain the spatial distribution patterns of HRS. A significant feature of the interaction detector in Geodetector is its ability to detect interactions between two factors, thereby revealing potential synergistic or antagonistic effects that may not be apparent when considering single factors alone (Fig. 7b). The detection results for each combination are characterized by two features: bivariate enhancement (55, 63.64%) and nonlinear enhancement
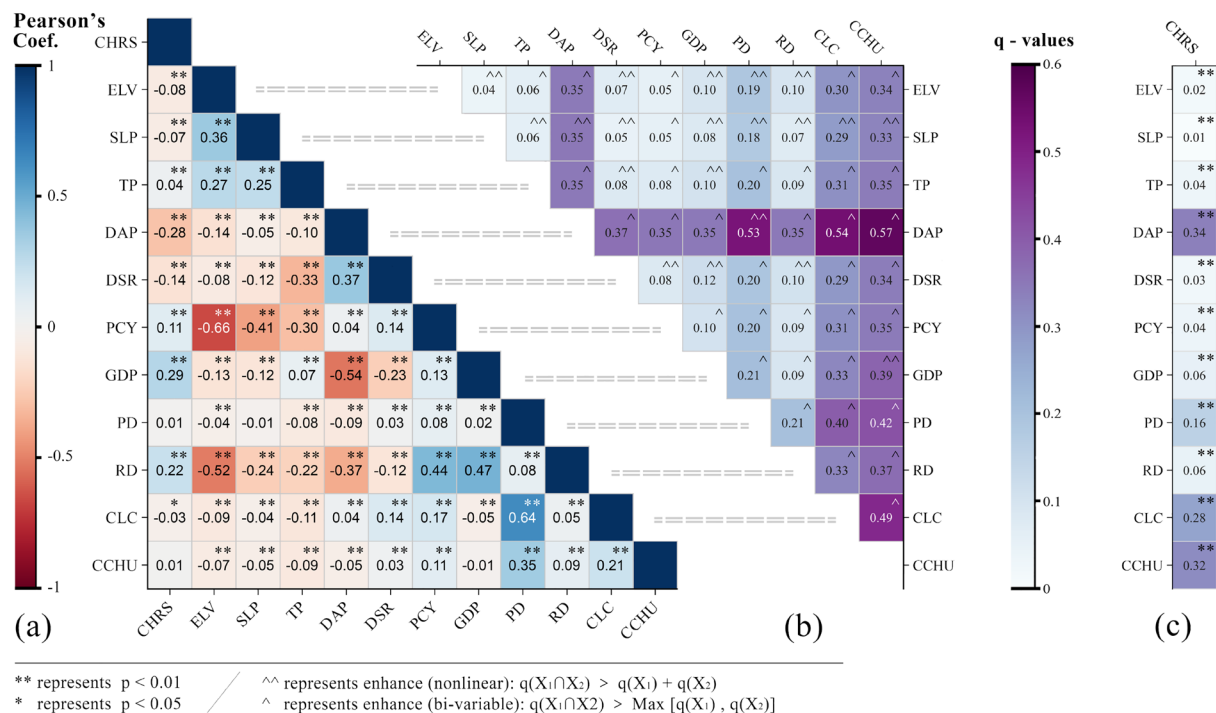

Fig. 7 Pearson correlation analysis. **b** Results of interaction detector by Geodetector. **c** Results of factor detector by Geodetector

Yuan *et al. Heritage Science*   (2024) 12:395

Page 12 of 22

(20, 36.36%). This suggests that the combined explanatory power of any two variables on CHRS is better than that of individual variables. Among the variable combinations exhibiting nonlinear enhancement, the combination of RD and DAP has the best explanatory power for the CHRS in THBX, with a q-value of 0.527. Subsequently, three other combinations-CCHU with GDP (q=0.387), DAP with SLP (q=0.353), and CCHU with SLP (q=0.332)-also explain over 30% of the CHRS. For combinations exhibiting bivariate enhancement, more than 50% of the CHRS can be explained by the combination of CCHU and DAP (q=0.570, the highest value in the interaction detection results) or CLC and DAP (q=0.539). Additionally, there are 19 pairs of variable combinations that have a bivariate enhancement effect (34.55%), explaining 30% to 50% of the CHRS. Although some bivariate combinations still possess weaker explanatory power (q<0.3), the results from the interaction detector in Geodetector still indicate that a multivariate integrated analysis can more effectively explain the CHRS in THBX.

Comparative analysis of multivariate regression results

In this section, we conducted spatial heterogeneity analysis using three models: OLS, GWR, and MGWR, and compared their fitting results with data based on four indicators: Optimization criterion (corrected Akaike information criterion, *AICc*), determination coefficients (R^2), adjusted determination coefficients ($Adj.R^2$), and residual sum of squares (*RSS*). In order to facilitate interpretation and comparison of the local parameter outputs of the models, we standardized the data for each variable [82]. This method also effectively guards against potential multicollinearity risks. The results are shown in Table 3. Lower values of *RSS* and *AICc* indicate better model performance, whereas R^2 and $Adj.R^2$ exhibit the opposite behavior. In addition, Table 3 also displays the bandwidth values (Bw) of the independent variables in the model analysis results, representing the usage of distance metrics for each independent variable. In the three multivariate regression models, the OLS model exhibits the highest *AICc* value (245933.163), with the lowest R^2 (0.238) and $Adj.R^2$ (0.237), and a slightly higher *RSS* value (5542.798) compared to GWR. This suggests that the OLS model provides the least satisfactory explanation for the CHRS in THBX due to the presence of spatial non-stationarity in the relationship. Subsequently, the GWR is utilised. Its *AICc* value (32248.356) demonstrates a notable decrease, with R^2 (0.342) and $Adj.R^2$ (0.324) slightly surpassing those of the OLS model. However, the GWR's *RSS* value (5616.460) is the highest, likely attributable to considerable variations in the spatial scales of different variables, indicating the need for multiscale analysis.

Table 3 Model comparison of fitting results between OLS, GWR and MGWR

Metrics	OLS	GWR	MGWR	
AICc	245933.163	32248.356	29734.703	
R^2	0.238	0.342	0.634	
Adj R^2	0.237	0.324	0.622	
RSS	5542.798	5616.460	4865.900	
			ELV: 484	(3.18%)
			SLP: 4344	(28.56%)
			TP: 6129	(40.30%)
			DAP: 307	(2.02%)
			DSR: 1243	(8.17%)
			PCY: 853	(5.61%)
Bandwidth (%)	–	406 (2.67%)	GDP: 90	(0.59%)
			PD: 15207	(99.99%)
			RD: 202	(1.33%)
			CLC: 15207	(99.99%)
			CCHU: 13224	(86.95%)
			Intercept: 65	(0.42%)

MGWR demonstrates the best performance, with the lowest *AICc* value (29734.703) and *RSS* value (4865.900), while achieving the highest R^2 (0.634) and $Adj.R^2$ (0.622). These findings suggest that the MGWR outperforms both OLS and GWR in explaining the CHRS in THBX, providing a more effective insight into the influencing patterns.

In terms of bandwidth selection, the GWR adopts a single bandwidth (406, 2.67%), whereas the MGWR exhibits a significant disparity in bandwidth selection for different independent variables. It ranges from 90 (0.59%) to 15207 (99.99%), showing a large variation, which indicates that a single bandwidth cannot accurately reveal the scale effects of various factors. Different influencing factors have different ranges of effects on the CHRS in THBX. The bandwidths of GDP (Bw=90), RD (Bw=202), DAP (Bw=307), ELV (Bw=484), PCY (Bw=853) and DSR (Bw=1243) account for 0.42% to 8.17% of the total sample size, indicating their effects operate at a micro-scale, locally. Then, the bandwidths of SLP (Bw=4344) and TP (Bw=6129) account for 28.56% to 40.30% of the total sample size, indicating that their effects exhibit limited spatial non-stationarity. Lastly, the bandwidths of CCHU (Bw=13224), RD (Bw=15207), and CLC (Bw=15207) account for 86.95% to 99.99% of the total sample size, indicating their effects are at a global scale, with little spatial heterogeneity, which indicates that their effects are all at the global scale, with minimal spatial heterogeneity. Additionally, using the same kernel function and bandwidth parameters may exacerbate collinearity among variables, sometimes called concurvity. By permitting bandwidth variation, variables can undergo diverse

transformations, thereby circumventing the induction of multicollinearity/concurvity [82]. Hence, during model computations, we conducted thorough local multicollinearity checks using the local condition numbers of GWR and MGWR, without encountering any issues. In summary, we will further interpret the results of MGWR.

Overall interpretation of MGWR results
Table 4 summaries the local parameter estimates of MGWR, including "Bandwidth (%)", "Sig. HRS (%)" (the proportion of effective coefficients of HRS to the total number of HRS when P ≤ 0.05), "+" (the proportion of significant positive coefficients of Sig. HRS), and "-" (the proportion of significant negative coefficients of Sig. HRS). The results indicate that the coefficients of PD and CLC, with the bandwidth accounting for 99.99% of the total sample size, are completely insignificant, which also represents a type of global consistency. The remaining independent variables exhibit significance for over 60% of HRS, which holds statistical significance in terms of interpretation. Initially, the entirely significant independent variables include SLP, TP, and PCY, displaying a globally consistent effect feature. The coefficients of TP and PCY are all positive, while that of SLP is negative. Additionally, although the effect of CCHU with bandwidth proportion close to the global scale only explains part of the HRS (66.79%), all significant coefficients exhibit negative impact effects. Subsequently, the significant coefficients of the remaining influencing factors exhibit different effect patterns, with some impacting Sig. HRS positively while others negatively. It is worth noting that, although the effect patterns exhibited by significant coefficients vary, they still show clear tendencies. Whether positive or negative, there will definitely be one type of feature accounting for over 75% of the total.

Figure 8 depicts the maximum, minimum, and average values of the local parameters of effective HRS samples under the interpretation of different variables. Obviously, the fluctuations in the minimum and maximum values of each variable are larger than the mean value. The local parameters of CCHU, SLP, and TP hardly change and their values are close to 0. The range of variation of the local parameters of ELV, PCY, and DSR is larger, but the mean values are still around 0. It is worth noting that the local parameters of RD, GDP, and DAP have the largest variation and produce the highest and lowest values in the variation. The overall interpretation of the MGWR results indicates that the effects of different variables on CHRS are nonlinear, which implies spatial heterogeneity in the distribution of local parameters. Although we can preliminarily characterize the effects of different variables, further research on the spatial distribution of their local parameters is still needed.

Analysis of spatial heterogeneity in the effects of variables
In the MGWR output, the local parameter of Intercept represents the influence of different geographical locations on the dependent variable when other independent variables are fixed. To some extent, the MGWR intercept can capture factors uniquely determined by geographical location, which influence the CHRS in THBX but are not included in the model. Notably, in this study, the intercept has the smallest bandwidth (Bw=65, 0.42%), indicating a micro-scale variable. Moreover, it shows global significance (Sig. HRS=15208, 100%), highlighting the high sensitivity of THBX's HRS to geographical changes. In Fig. 9, the global distribution of the intercept's local parameters is displayed, as well as the global distribution of standardized residuals. Following Sullivan, we
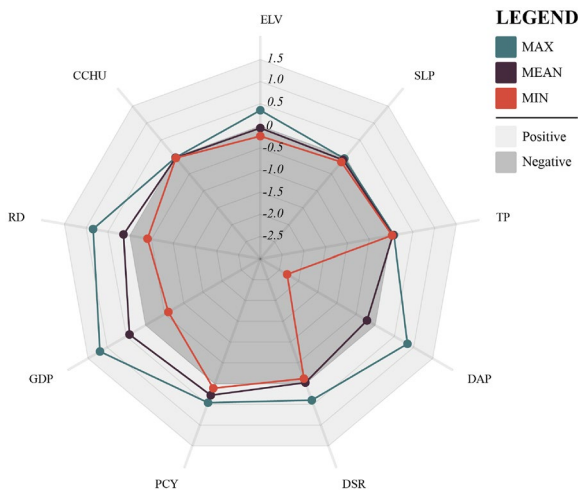
Table 4 Parameter estimation using MGWR

Variables	Bandwidth (%)	Sig. HRS (%)	+%	−%
ELV	484 (3.18%)	9216 (61.60%)	21.76%	78.24%
SLP	4344 (28.56%)	15208 (100.00%)	0.00	100.00%
TP	6129 (40.30%)	15208 (100.00%)	100.00%	0.00
DAP	307 (2.02%)	10644 (69.99%)	0.16%	99.84%
DSR	1243 (8.17%)	9937 (65.34%)	16.06%	83.94%
PCY	853 (5.61%)	15208 (100.00%)	100.00%	0.00
GDP	90 (0.59%)	11148 (73.30%)	99.62%	0.38%
PD	15207 (99.99%)	0 (−)	−	−
RD	202 (1.33%)	9842 (64.72%)	88.16%	11.84%
CLC	15207 (99.99%)	0 (−)	−	−
CCHU	13224 (86.95%)	10158 (66.79%)	0.00	100.00%
Intercept	65 (0.42%)	15208 (100.00%)	0.00	100.00%

[a] Sig. HRS(%) refers to the number and proportion of samples in the HRS dataset that meet the significance requirements in spatial statistics, where *P* is greater than 0.05

Fig. 8 MGWR local parameters of each variable

Yuan *et al. Heritage Science*   (2024) 12:395

Page 14 of 22

should focus on positions where the standardized residuals exceed three times the standard deviation, as these locations may contain outliers [83]. The standardized residuals in THBX have a global distribution of no more than 0.5 standard deviations, suggesting that there are no prominent outliers or outlying points in the MGWR results. Thus, the model offers a good fit to the data, and the explanatory variables employed effectively account for the spatial variation of the dependent variable.

Figure 10 presents the distribution of significant HRS samples and the spatial patterns of local parameters in the THBX across various explanatory factors. It is evident that the use of variable bandwidth offers a more comprehensive and precise view in spatial data exploration. Additionally, the standardization of data enables more meaningful comparisons of specific values of local parameters across different variable factors, aiding in a more intuitive summarisation of the determinants and magnitudes of these factors.

Figure 10a displays the significant samples and spatial distribution of local parameters for ELV. The local parameter range for ELV is between [−0.231,0.353], suggesting a predominant negative effect on the CHRS in THBX with a minor positive influence as well. In terms of spatial distribution, the positive impacts are predominantly found in the northernmost part of THBX, specifically around Jundu Xing and its vicinity, primarily located in Shanxi Province. Furthermore, certain HRS in the low-elevation areas on the southeastern side of THBX also demonstrate a positive effect from ELV. ELV demonstrates detrimental impacts on two strip-shaped regions:

the Taihang Xing - Jing Xing region and the Feihu Xing - Puyin Xing region. It is noteworthy that the local parameters of the samples at the outer edges of the strip exhibit slightly higher than those in the interior. The areas of positive influence may be associated with the Ming Great Wall. The THBX area is closely linked to the Zhenbao Town defense zone. Many of the fortifications in Zhenbao Town were built along the Taihang Mountains, while the passes were constructed over significant paths, such as Juyong Pass, Zijing Pass, Daoma Pass, Gu Pass, Dongyang Pass, Tianjing Pass, and Hengling Pass [84]. The Military Capital Ridge in THBX falls within the Juyong Pass defense area of the Ming Great Wall, a region acting as the closest pass city to the capital, Beijing. As a result, a large number of military settlements were established around this area, typically located at higher altitudes on sites that are easy to defend but difficult to attack [85].

The spatial pattern of significant samples and local parameters for SLP is shown in Fig. 10b. The data summary indicates a significant negative influence of SLP on all HRS in THBX at a global level. The range of local parameter values is [−0.151, −0.048], with less variability than ELV. This indicates that in mountainous terrain areas, the clustering trend of HRS in THBX is inhibited with the increase of SLP. The southeastern part of THBX experiences the most pronounced negative impact, while the low-impact zone is located in the eastern part of the central region. With a bandwidth of 4344 for SLP, its influence scale is closer to medium-range, indicating weak spatial heterogeneity and a smoother transition of global local parameters. Due to the more complex terrain and greater variation in slope, the southeastern part of THBX exhibits the most significant negative impact. Conversely, the eastern central region, which is the alluvial fan plain of the Hutuo River and close to Shijiazhuang city in Hebei province, has a relatively gentle slope, resulting in a lesser impact on CHRS in this area.

The results in Fig. 10c indicate that TP can also explain the distribution characteristics of all HRS in the THBX. However, the local parameter variation range for TP is [0.025,0.067], suggesting that its impact on the CHRS is consistently positive, but the specific parameter values are smaller and the variations are very slight. The area with the most pronounced impact is the southeastern part of THBX, while the least pronounced area is on the western side of the Fukou Xing—Bai Xing region. Notably, this area also has the highest density of CHRS. With TP's bandwidth at 6129, which falls within a medium-scale effect range, its spatial heterogeneity is comparably weak, leading to a smoother transition of local parameters across different regions. The southeastern part of THBX, close to Zhengzhou City in Henan province, is influenced by
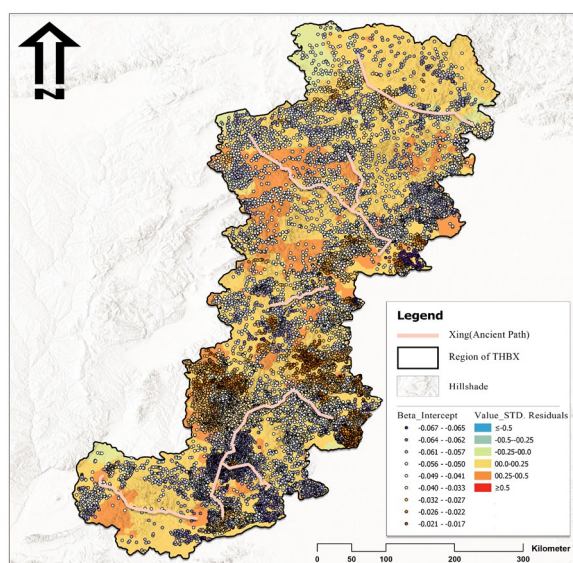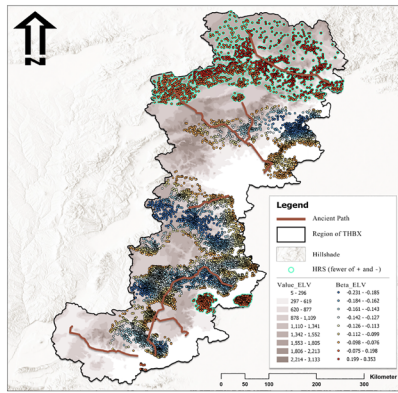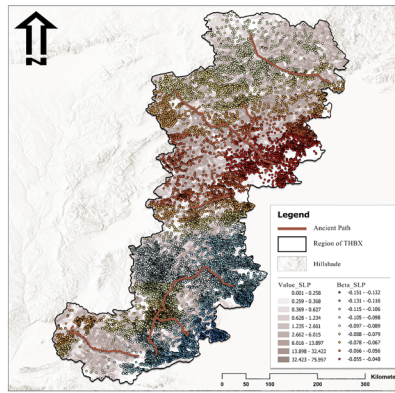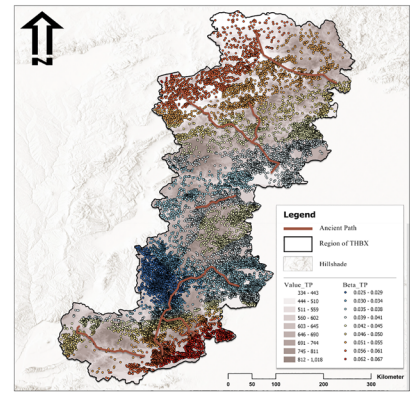

Fig. 9 Spatial distribution of standardized (STD) residuals and local parameters (Beta) of intercept in MGWR results
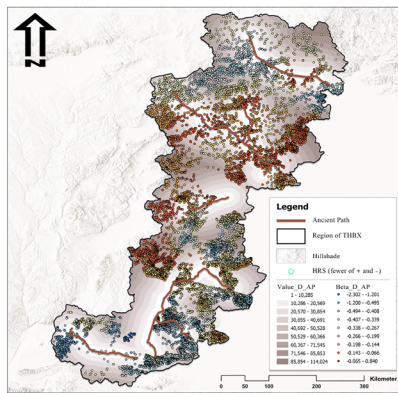
(a) *ELV*: Bw=484, Sig. HRS=9216(61.60%)
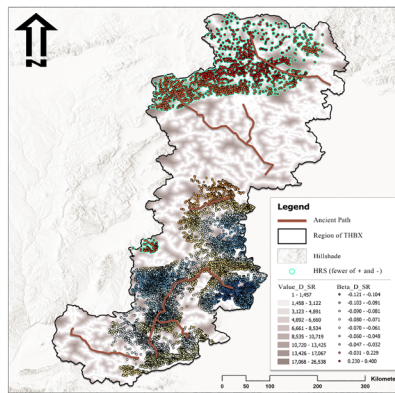
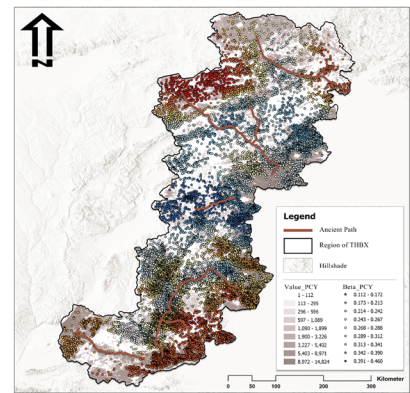(b) *SLP*: Bw=4344, Sig. HRS=15208(100.00%)

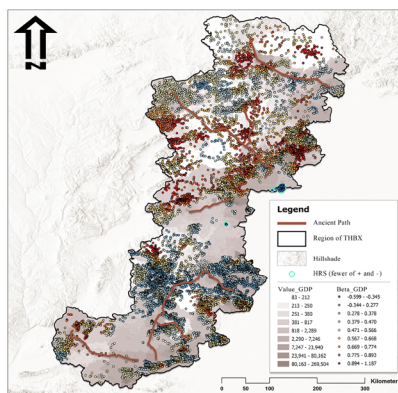(c) *TP*: Bw=6129, Sig. HRS=15208(100.00%)

(d) *DAP*: Bw=307, Sig. HRS=10644(69.99%)
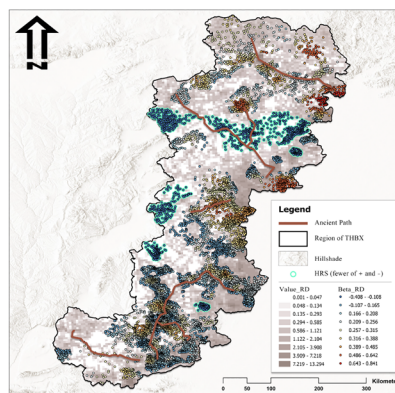
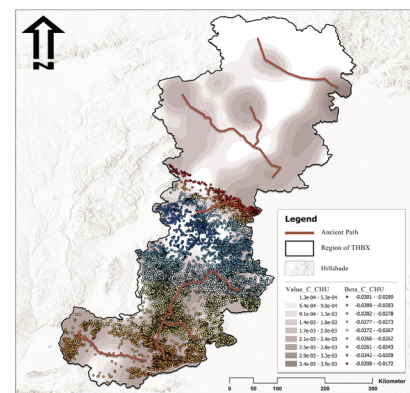(e) *DSR*: Bw=1243, Sig. HRS=9937(65.34%)

(f) *PCY*: Bw=853, Sig. HRS=15208(100.00%)

(g) *GDP*: Bw=90, Sig. HRS=11148(73.30%)

(h) *RD*: Bw=202, Sig. HRS=9842(64.72%)

(i) *CCHU*: Bw=13224, Sig. HRS=10158(66.79%)

Fig. 10 The spatial distributions of local coefficients for influencing factors of HRS distribution in THBX

Yuan *et al. Heritage Science*   (2024) 12:395

Page 16 of 22

the Yellow River alluvial fan plain [86]. Henan has historically been one of China's most populous provinces, hence the relatively high number of settlements in the area. Additionally, the higher ELV and complex terrain in the southeast block much of the summer monsoon, resulting in higher TP than nearby areas, which has the greatest impact on CHRS in this area. Conversely, the area west of Fukou Xing to Bai Xing within THBX is the area with the densest concentration of settlements, influenced by the mountains with lower TP, making it the least affected area by TP.

Figure 10d, e display the spatial patterns of significant samples and local parameters for DAP and DSR, elucidating the relationship between the CHRS in THBX and the main linear elements. In the case of DAP, it is evident that the majority of Sig.HRS distributions are negatively correlated with DAP. The local parameter range for DAP is [−2.302,0.840], which stands as the most extensive and variable among all independent variable local parameters, emphasizing its significant impact. This suggests that being far from the ancient path is unfavorable for the CHRS. Interestingly, the CHRS at the intersections of the three ancient paths in the south are positively impacted by DAP. This suggests that overly accessible geographic conditions might pose potential risks to the HRS in this location. The eight ancient paths not only played a crucial role as transportation routes in ancient times, such as during the Qin Dynasty when they served as vital routes for imperial travel, but also as main postal roads during the Ming and Qing dynasties. Many of these ancient paths still form part of the routes of modern highways today [84].

For DSR, its local parameter variation range is [−0.121, 0.400], with both the numerical value and the degree of change being smaller than DAP. However, similar to DAP, the distribution of the majority of Sig.HRS is negatively correlated with DSR. This suggests that surface runoff majorly shows positive influence on the aggregation of HRS. Spatially, the HRS in the Jing Xing—Taihang Xing region is segmented into three clusters with negative local parameters, having lower values internally and relatively higher values at the periphery. However, in the northwestern part of the Jundu Xing— Puyin Xing area, the distribution of some HRS shows a positive correlation with DSR, indicating that surface runoff has caused damage to their aggregation and development. Moreover, the bandwidths for DAP and DSR are 307 and 1243, respectively, with DAP exhibiting more pronounced spatial heterogeneity compared to DSR.

As shown in Fig. 10f, PCY has a significant positive effect on all CHRS in THBX. The local parameter values of PCY vary between [0.112−0.460], indicating a moderate impact intensity. The HRS in the southeast

and northwest parts of THBX are most affected by PCY, while the local parameters around Jing Xing are relatively lower. Although the bandwidth of PCY is 853, falling into the semi-microscopic scale, it still shows a significant global trend in THBX, with higher values in the north and south and lower in the central part. The southeastern and northwestern areas, which are the most positively impacted areas mentioned above, comprise the Yellow River alluvial fan plain and the Datong Basin, respectively [86]. Both areas feature geological conditions conducive to agriculture, leading to a concentration of agriculturally oriented rural settlements.

The results in Fig. 10g display the significant samples and spatial patterns of the local parameters for GDP. The local parameters of GDP vary between −0.599 and 1.187, showing a large variation range, with mostly positive values. This indicates that GDP mainly has a positive impact on the CHRS in THBX. A minority of HRS negatively impacted by GDP are predominantly located near the eastern endpoint of the Puyin Xing, where GDP growth might deter CHRS. Given a bandwidth of 90, GDP operates on an extremely micro scale, just slightly larger than the intercept. Hence, the high-value regions of GDP's local parameters in THBX display a scattered distribution, with pronounced variations in influence across different areas. Additionally, influenced by the river systems originating from the Taihang Mountains, numerous alluvial fans have formed along the eastern side of the mountains, bringing fertile soil to the North China Plain. Historically, these lands have been densely populated, forming a north-to-south urban belt along the Taihang Mountain range with cities like Beijing, Baoding, Shijiazhuang, Xingtai, Handan, Anyang, Hebi, and Xinxiang located on these alluvial fans. Consequently, the GDP within the THBX area is generally higher on the eastern side than on the western side.

In Fig. 10h, regarding the driving characteristics explained by RD, we found that its local parameter variation range is [−0.408,0.841], mostly positive, indicating that it mainly exerts a positive influence on the aggregation of Sig.HRS. Among them, a small part of the HRS affected negatively by RD mainly gathers in the middle of the Feihu Xing - Puyin Xing region and the west side of the Jing Xing - Fukou Xing region. Due to the bandwidth being 202, RD is a driving factor at the microscale. Consequently, there is considerable variability in the influence among different Sig.HRS aggregation areas. Lastly, in Fig. 10i, the local parameter variation range for CCHU is [−0.0301,−0.0172], indicating that the CHRS in THBX is suppressed by the enhancement of CCHU. Compared to the determinants mentioned earlier, the negative impact generated by CCHU is extremely weak. Additionally, CCHU is significant only for the HRS in the southern

Yuan *et al. Heritage Science*   (2024) 12:395

Page 17 of 22

part of THBX, with the most negative impact occurring at Jing Xing—Fukou Xing. t is worth noting that the analysis related to CCHU is a discussion focused on the early development determinants of HRS. It serves to elucidate the CHRS rather than probing into their future developmental trends.

Discussion

The primary objective of this study is to catalog the HRS that persist today in the Taihang Baxing, establish a GIS-based geospatial data set for these rural settlements, and analysis their spatial distribution characteristics. Additionally, the study examines the factors influencing the distribution of these HRS and the underlying determinants. The research involved the compilation of HRS recorded in the archival materials of the THBX, followed by a comparison with the names of modern villages. Based on various levels of confidence, a total of 15,208 HRS have been identified and preserved to date. The study first employed kernel density analysis to explore the clustering characteristics of these HRS. Subsequently, using the MGWR, it analysed the spatial heterogeneity of the distribution patterns and influencing factors of these historical settlements.

Insights

One of the key focuses of this research is the compilation and analysis of historical documents, and the continuity characteristics of Chinese place names. This led to the development of a novel process for identifying and organizing historical settlements. The research reveals that the CHRS in the THBX generally exhibits a pattern of greater density in the south and sparser distribution in the north, aligning along a north–south axis following the Taihang Mountains. Two primary clustering cores of HRS have formed near the Fukou Xing, Bai Xing, and Taihang Xing, mainly located at the intersections of these three Xings and the eastern end of the Fukou Xing. A secondary core has developed in the central areas of the Feihu Xing and Jundu Xing. Previous studies, such as those by Hui on TCV in the Taihang Mountains, analyzed the distribution characteristics of traditional villages [26]. However, since TCV represent only a subset of HRS, the clustering characteristics identified in those studies show notable differences from the findings of this research. A more comprehensive dataset typically yields clearer and more accurate statistical results regarding settlement distribution. The study continues to explore the underlying driving factors of HRS distribution through deeper analysis using the MGWR.

The analysis using the MGWR indicates that natural factors have consistently been one of the critical influences on the spatial distribution of villages. The ELV demonstrates an alternating pattern of significant and non-significant influences on the spatial CHRS, predominantly exerting a negative impact with clear spatial heterogeneity. The SLP has a universally significant negative impact on the CHRS. Conversely, the TP exhibits a positive influence on the CHRS, although the changes are very subtle. In the initial stages of village formation, natural geographical factors often play a crucial restrictive role in site selection, with environments conducive to survival and labor typically hosting numerous HRS. This finding aligns with the conclusions of Lai, who focused on TCV as his study subjects [37]. Although traditional villages are a subset of HRS, they exhibit similar characteristics and conclusions in some aspects of the research.

The analysis reveals that the DAP has a significant negative correlation with the CHRS, exhibiting the most pronounced effect. The DSR impacts HRS in a manner similar to DAP, but its influence is less significant. These results underscore a clear dependency of HRS clustering on both Xing and water systems, with a stronger reliance on the former. This suggests that transportation conditions play a more critical role in settlement site selection than water systems. Similar findings are echoed in the research by Zhang Xinrui, who also noted the correlation between settlement distribution and proximity to paths and water systems in his study of HRS around the Feihu Xing [27]. Feng Wenlan's study on the distribution patterns of rural settlements in the upper reaches of the Min River further supports the greater dependency of settlements on transportation networks compared to water systems [87]. This research in the THBX reaffirms these conclusions and further identifies that DAP provides a stronger explanatory power for village clustering than water systems, which exhibit greater spatial heterogeneity.

The analysis shows that the PCY significantly and positively affects the clustering of HRS on a global scale. GDP also exhibits a positive impact in most regions; however, in high-value areas, its influence is dispersed and occurs at a microscale. RD also impacts at a microscale and displays a dispersed influence globally. The effect of RD on the clustering of HRS varies significantly across different regions. In the rapid process of urbanization, the preservation of HRS depends on the contemporary socioeconomic and demographic developments. However, the impact of these factors varies by region, necessitating targeted protection and development planning tailored to specific local conditions.

The CCHU have an extremely slight negative impact on HRS clustering, with significant effects observable only in the southern part of the THBX. This contrasts with Hongwei's findings, where CCHU was positively

Yuan *et al. Heritage Science*   (2024) 12:395

Page 18 of 22

correlated with the concentration of TCV [25]. The difference might be due to TCV selections being influenced by local cultural preservation units, which are more numerous and place greater emphasis on cultural heritage in certain areas, resulting in more approved traditional villages. However, in a more extensive and comprehensive HRS dataset, the influence of cultural preservation units on village clustering appears to be minimal [25].

The study on the variable factors influencing the CHRS in the THBX reveals the drivers behind the settlement patterns, which reflect the area's natural environmental conditions and regional socio-economic development levels [30, 87]. In this study, the subject is the HRS within the THBX area that still exist today, with modern external environmental factors as influences. The objective is to analyze the current developmental state to summarize the external environmental characteristics that affect the preservation of rural heritage. The results of this study can provide methodological and theoretical references for the excavation of rural heritage and the evolution of regional rural patterns. Small towns are a significant driving force behind urbanization and rural-focused development in China [88]. Therefore, analyzing the distribution characteristics and variable factors of HRS can provide theoretical guidance for the comprehensive development of China's contemporary rural settlement system in terms of structure and function.

Limitations and prospects

It must be acknowledged that this study has certain limitations. First, there are issues related to the sources of HRS data. China, with its extensive historical background, has rural settlements that have developed over thousands of years. However, when attempting to organize and excavate historical data, we encounter a lack of uniform standards in the documentation process. For instance, some historical county annals provide detailed records of village names, while others offer only brief mentions. Consequently, this study could not compile an entirely comprehensive dataset of HRS. Secondly, the process of matching HRS with modern place names presents challenges. Despite narrowing the administrative boundaries from provinces to counties, some rural settlements still share names or have undergone name changes. In this study, such rural settlements have been marked with different confidence levels, and only those with a higher degree of match were included in the statistical analysis. Revealing and restoring HRS is a labor-intensive and lengthy task. It is hoped that future research can continuously refine the dataset created in this study to bring the data on HRS closer to their true state. By comparing historical settlements with contemporary villages, or the distribution of villages across different periods, it is possible to summarize the patterns of village formation and development, as well as to better predict the trajectory of rural evolution and development.

In Huang and Oliveira's research, the correlation between clan culture and the internal spatial form with the development of HRS has been emphasized, though it still remains on an individual scale [89, 90]. Therefore, conducting extensive research and quantification of the spatial form of HRS and traditional culture, and introducing more detailed variables, will help enhance the completeness of the regression model in this study. A more comprehensive regression analysis will support the development of a machine learning-based, multivariate model for predicting the spatial distribution of HRS. This model can be cross-validated with the HRS spatial information acquisition method proposed in this study, advancing the exploration and protection of rural heritage.

Conclusion

As China's recognition of the value of rural heritage increases, research related to HRS is increasingly encouraged. China, with one of the longest histories of agriculture, provides a favorable environment for the preservation and development of HRS. In this study, we first established an objective methodology for acquiring spatial information on HRS. This involved extracting information about HRS from historical records and matching it with modern rural settlement POI to obtain a spatial distribution dataset for HRS. Subsequently, using this dataset, we analyzed the spatial distribution patterns of HRS in the THBX and employed the MGWR to quantitatively assess the impact of natural geographic and socio-economic factors on the CHRS.

This study has validated the correlation between various factors and the formation of spatial distribution patterns of HRS. Our results from the THBX indicate that the CHRS is the outcome of complex interactions among multiple factors. Additionally, the influence of these factors on the spatial distribution patterns of HRS is non-linear and significantly affected by geographical location changes. The methodology for obtaining HRS data proposed in this study will robustly support heritage exploration efforts in various regions. Furthermore, HRS represent a living heritage, which requires not only preservation but also development. The impact of different factors on the spatial distribution patterns of HRS may indicate potential advantages and disadvantages for their future development. This study also validates the adaptability of the chosen factors in understanding and analyzing HRS.

The findings of this study provide methodological support for the discovery and preservation of rural heritage,

as well as insights for optimizing rural spatial patterns and predicting the evolutionary paths of rural areas. Future work should include expanding the scope of the study and examining a broader range of factors to gain a more comprehensive understanding of the logic behind the preservation and CHRS. Further research could focus on cluster analysis of heritages within different regions that have already been studied, which would produce more regionally specific and detailed research outcomes. For the THBX, the next steps could involve verifying the accuracy of the existing HRS dataset and conducting further studies on non-significant or unique areas identified in the spatial regression model analysis. This approach will enhance the precision and relevance of the research, contributing significantly to regional planning and heritage conservation efforts.

Appendix A
Appendix 1: Counties (with the number of HRS) included in THBX

Province	City	County						
Beijing	–	Changping(68)	Mentougou(90)	Fangshan(105)	Yanqing(51)			
Hebei	Zhangjiakou	Urban District(38)	Zhangbei(79)	Shangyi(44)	Chongli(36)	Chicheng(99)	Wanquan(85)	Huaian(142)
		Huailai(117)	Zhuolu(100)	Yu County(202)	Yangyuan(172)	Xuanhua(161)		
	Baoding	Urban District(3)	Fuping(72)	Laiyuan(166)	Laishui(153)	Yi County(165)	Dingxing(168)	Rongcheng(52)
		Qingyuan(98)	Wangdu(68)	Boye(39)	Anguo(118)	Dingzhou(31)	Quyang(177)	Tang county(119)
		Xushui(178)	Shunping(87)	Anxin(6)	Mancheng(117)			
	Shijiazhuang	Urban District(5)	Pingshan(122)	Lingshou(98)	Xingtang(138)	Zhengding(35)	Luancheng(92)	Yuanshi(92)
		Luquan(92)	Jingxing(158)	Zanhuang(145)				
	Xingtai	Urban District(1)	Shahe(288)	Xingtai(361)	Neiqiu(101)	Lincheng(43)		
	Handan	Urban District(93)	Wuan(433)	She County(269)	Ci County(219)	Linzhang(230)	Chengan(119)	Handan(62)
Henan	Anyang	Urban District(1)	Linzhou(249)	Anyang(261)				
	Xinxiang	Weihui(176)	Huixian(314)	Huojia(53)				
	Jiaozuo	Urban District(13)	Xiuwu(96)	Wuzhi(161)	Boai(134)	Wen County(77)	Qinyang(187)	Mengzhou(93)
	Jiyuan	Urban District(15)						
	Luoyang	Mengjin(83)	Xinan(5)					
	Sanmenxia	Mianchi(10)	Yima(1)					
Shanxi	Jinzhong	Shouyang(143)	Xiyang(203)	Heshun(114)	Zuoquan(189)	Yushe(159)		
	Changzhi	Urban District(106)	Qin County(151)	Lucheng(126)	Zhangzi(79)	Changzhi(179)	Huguan(145)	Pingshun(59)
		Licheng(206)	Wuxiang(389)	Xiangyuan(410)	Tunliu(290)			
	Jincheng	Urban District(0)	Qinshui(144)	Gaoping(455)	Lingchuan(349)	Zezhou(436)	Yangcheng(92)	
	Yuncheng	Wenxi(103)	Xia County(86)	Xinjiang(11)	Jiang County(44)	Yuanqu(9)		
	Linfen	Xiangfen(6)	Quwo(55)	Houma(0)	Yicheng(56)			
	Yangquan	Urban District(33)	Yu County(164)	Pingding(100)				
	Xinzhou	Fanzhi(141)	Wutai(182)					
	Shuozhou	Huairen(72)	Ying County(142)					
	Datong	Urban District(158)	Yanggao(125)	Guangling(86)	Lingqiu(145)	Hunyuan(38)	Datong(110)	Tianzhen(102)

Yuan *et al. Heritage Science*   (2024) 12:395

Page 20 of 22

Appendix B

Appendix 2: Abbreviations used in this paper (ranking from A-Z)

Abbreviations	Full name
AICc	Corrected Akaike Information Criterion
ANN	Average Nearest Neighbor
CCHU	Concentration of Cultural Heritage Unities
CHRS	Concentration of Historical Rural Settlements
CLC	Concentration of Large Cities
DAP	Distance to Ancient Paths
DSR	Distance to Surface Runoff
ELV	Elevation
GDP	Gross Domestic Product
GWR	Geographic Weighted Regression
HRS	Historical Rural Settlements
LISA	Local Indicators of Spatial Association
MGWR	Multi-Scale Geographic Weighted Regression
OLS	Ordinary Least Squares
PCY	Potential Crop Yield
PD	Population Density
POI	Points of Interest
RD	Road Density
RSS	Residual Sum of Squares
SLP	Slope
TCV	Traditional Chinese Villages
THBX	Taihang Baxing
TP	Total Precipitation
TR	Topographical Relief
VIF	Variance Inflation Factor

Availability of data and materials
Data can be provided upon request from the corresponding author.

Declarations

References
1. Wang Y, Liu Y, Li Y, Li T. The spatio-temporal patterns of urban-rural development transformation in China since 1990. Habitat Int. 2016;53:178–87.
2. Zhu S, Kong X, Jiang P. Identification of the human-land relationship involved in the urbanization of rural settlements in Wuhan city circle, china. J Rural Stud. 2020;77:75–83.
3. Silva A, Roders A. Cultural heritage management and heritage (impact) assessments. In: International Conference on Facilities Management, Procurement Systems and Public Private Partnership, 2012;pp. 375–382
4. ICOMOS: ISCCL-ICOMOS-IFLA Principles Concerning Rural Landscapes as Heritage. https://rm.coe.int/1680083746, Last accessed on 10 April 2024 (2017)
5. Li H, Yuan Y, Zhang X, Li Z, Wang Y, Hu X. Evolution and transformation mechanism of the spatial structure of rural settlements from the perspective of long-term economic and social change: A case study of the Sunan region. China J Rural Stud. 2022;93:234–43.
6. Long H, Tu S, Ge D, Li T, Liu Y. The allocation and management of critical resources in rural China under restructuring: problems and prospects. J Rural Stud. 2016;47:392–412.
7. Li Y, Jia L, Wu W, Yan J, Liu Y. Urbanization for rural sustainability-rethinking China's urbanization strategy. J Clean Prod. 2018;178:580–6.
8. Yang R, Zhang J, Xu Q, Luo X. Urban-rural spatial transformation process and influences from the perspective of land use: A case study of the Pearl River Delta region. Habitat Int. 2020;104: 102234.
9. Zhou M, Chu S, Du X. Safeguarding traditional villages in China: the role and challenges of rural heritage preservation. Built Heritage. 2019;3:81–93.
10. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Notice of the Ministry of Housing and Urban-Rural Development, the Ministry of Culture, the National Cultural Heritage Administration and the Ministry of Finance on conducting traditional village surveys. https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201204/20120423_209619.html, Last accessed on 12 May 2024 (2012)
11. Tao W, Chen H. Jieyong: spatial form and spatial cognition of traditional village in syntactical view: a case study of Xiaozhou village. Guangzhou Acta Geographica Sinica. 2013;68(2):209–18.
12. Liu P, Dong S. Study of landscape-image of Chinese ancient village. Geogr Res. 1998;17(1):32–9.
13. Wang Y, Guo H, Yang L. The discussion on evaluation and sustainable development models of traditional villages in Western Mountainous area of Bejing: A case study of traditional villages in Mentougou district of Beijing. Scientia Geographica Sinica. 2006;06:735–42.
14. Zhao Y, Zhang J, Li N, Liang L. The study on conservation evaluation system and method on the historic cultural towns & villages: A case study of the first group famous historic and cultural towns & villages in China. Scientia Geographica Sinica. 2006;06:4497–505.
15. Wang Y, Yang L, Guo H. The discussion on traditional villages conservation and its tourism development models in Western Mountain of Beijing: A case study of Mentougou district. J Mt Sci. 2006;24(4):466–72.

Yuan *et al. Heritage Science*   (2024) 12:395

Page 21 of 22

16. Li P, Wang Q, Ryan C. The impacts of tourism on traditional villages: a case study of Mt.Qiyun, Anhui province. Tourism Tribune. 2012;27(04):57–63.

17. Lu S, Lu L, Wang L, Wang Y, Liang D, Yang Z. Temporal characteristics of tourist flows to ancient villages: a case study of two world cultural heritage, Xidi village and Hongcun village. Scientia Geographica Sinica. 2004;02:250–6.

18. Liu D, Hu J, Chen J, Xu X. The study of spatial distribution pattern of traditional villages in China. China Popul Resour Environ. 2014;24(04):157–62.

19. Li Y, Yao W, Zhang Y, Peng W. Spatial distribution characteristics of traditional villages in China. China Cultural Heritage. 2020;04:51–9.

20. Yagoub M, Al Yammahi AA. Spatial distribution of natural hazards and their proximity to heritage sites: case of the United Arab Emirates. Int J Disaster Risk Reduct. 2022;71: 102827.

21. Kohler F, Marchand G, Negrão M. Local history and landscape dynamics: a comparative study in rural Brazil and rural France. Land Use Policy. 2015;43:149–60.

22. Li H, Song W. Evolution of rural settlements in the Tongzhou district of Beijing under the new-type urbanization policies. Habitat Int. 2020;101: 102198.

23. Liu Y, Liu Y, Zhai R. Geographical research and optimizing practice of rural hollowing in China. Acta Geogr Sin. 2009;64(10):1193–202.

24. Gong J, Jian Y, Chen W, Liu Y, Hu Y. Transitions in rural settlements and implications for rural revitalization in Guangdong province. J Rural Stud. 2022;93:359–66.

25. Zhou H. Theory and Methodology of Traditional Village Value Rating Classification. Beijing: Science Press; 2022.

26. Hui Z, Jiaqi C, Shaoying X, Jingjing L, Xiaochao Y. Spatio-temporal distribution and evolutionary characteristics of traditional villages in Taihang Mountains area. City Plan Rev. 2020;44(08):90–7.

27. Zhang X, Yuan L. A preliminary study on historical evolution and spatial characteristics of traditional settlements along the Feihu passage based on GIS. J Landscape Res. 2021;002:17–23.

28. Fuente JL, Infante-Amate J, Travieso E. Historical changes in Mediterranean rural settlements (Southern Spain, 1787–2019). J Rural Stud. 2024;106: 103217.

29. Yu C, Li G, Cao Y, Wang J, Fang X, Zhou L, Liu Y. Distinct types of restructuring scenarios for rural settlements in a heterogeneous rural landscape: application of a clustering approach and ecological niche modeling. Habitat Int. 2020;104: 102248.

30. Yang R, Xu Q, Long H. Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization. J Rural Stud. 2016;47:413–24.

31. Qu Y, Jiang G, Yang Y, Zheng Q, Li Y, Ma W. Multi-scale analysis on spatial morphology differentiation and formation mechanism of rural residential land: a case study in Shandong province. China Habitat Int. 2018;71:135–46.

32. Song W, Chen B, Zhang Y. Land-use change and socio-economic driving forces of rural settlement in China from 1996 to 2005. Chin Geogra Sci. 2014;24:511–24.

33. Zhou X, Wang Z, Liu Y, Cheng Y, Zhou Y, Zhang J. Factors influencing the evolution of rural settlements based on MGWR: a case study of Haikou city. Trop Geogr. 2023;43(8):1599–610.

34. Li J, Wang X, LI X. Spatial distribution characteristics and influencing factors of Chinese traditional villages. Econ Geogr. 2020;40(02):143–53.

35. Yu J, Tang S, Chen Y. Spatial differentiation and influencing factors of traditional villages in Wuling Mountain area based on GWR. J Hubei Univ (Nat Sci). 2021;43(04):367–76.

36. Ma Y, Huang Z. Research on spatial distribution and accessibility of the traditional villages in the urban agglomeration on the middle reaches of the Yangtze River based on GWR model. Hum Geogr. 2017;32(04):78–85.

37. Lai F, Dayu Z. Study on spatial differentiation characteristics and influencing factors of traditional villages in North China based on MGWR model. Chin Landscape Archit. 2022;38(10):56–61.

38. Shen Y, Karimi K. The economic value of streets: mix-scale spatio-functional interaction and housing price patterns. Appl Geogr. 2017;79:187–202.

39. Song J, Yu H, Lu Y. Spatial-scale dependent risk factors of heat-related mortality: a multiscale geographically weighted regression analysis. Sustain Cities Soc. 2021;74: 103159.

40. Yu H, Fotheringham AS, Li Z, Oshan T, Wolf LJ. On the measurement of bias in geographically weighted regression models. Spatial Statistics. 2020;38: 100453.

41. Li M, Cao Y, Li G. An approach to developing and protecting linear heritage tourism: the construction of cultural heritage corridor of traditional villages in Mentougou district using GIS. Int J Geoheritage Parks. 2023;11:607–23.

42. Zhang Z. Analysis of the cultural heritage characteristics of the "Eight Taihang Xing" line. Academy. 2012;16(6):27–31.

43. Ministry of Housing and Urban-Rural Development of the People's Republic of China: 2024 Traditional Villages Concentrated Protection and Utilization Demonstration Counties (Municipalities and Districts) List. https://www.mohurd.gov.cn/xinwen/gzdt/202404/20240402_777360. html, Last accessed on 12 May 2024 (2024)

44. Ministry of Housing and Urban-Rural Development of the People's Republic of China: 2023 Traditional Villages Concentrated Protection and Utilization Demonstration Counties (Municipalities and Districts) List. https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/202304/20230403_771022.html, Last accessed on 12 May 2024 (2023)

45. Ministry of Housing and Urban-Rural Development of the People's Republic of China: 2022 Traditional Villages Concentrated Protection and Utilization Demonstration Counties (Municipalities and Districts) List. https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/202203/20220328_765372.html, Last accessed on 12 May 2024 (2022)

46. Cacciafoco FP, Cavallaro F. Place names: approaches and perspectives in toponymy and toponomastics. Cambridge: Cambridge University Press; 2023.

47. Wang S-Y. William: the Chinese language. Sci Am. 1973;228(2):50–63.

48. Breezedeus: CnOCR. https://github.com/breezedeus/cnocr, Last accessed on 10 May 2024 (2024)

49. Explosion: SpaCy. https://github.com/explosion/spaCy, Last accessed on 10 May 2024 (2024)

50. Amap Software: AMap LBS. https://lbs.amap.com/, Last accessed on 10 May 2024 (2023)

51. Civitasv: AMapPoi. https://github.com/Civitasv/AMapPoi, Last accessed on 10 May 2024 (2024)

52. Pandas development team: Pandas. https://github.com/pandas-dev/pandas, Last accessed on 10 May 2024 (2024)

53. NumPy: NumPy. https://github.com/numpy/numpy Last accessed on 12 May 2024 (2024)

54. Li Y, Long H, Liu Y. Spatio-temporal pattern of China's rural development: a rurality index perspective. J Rural Stud. 2015;38:12–26.

55. Liu Y, Chen M, Tian Y. Temporal and spatial patterns and influencing factors of intangible cultural heritage: ancient Qin-Shu roads. Western China Heritage Sci. 2022;10(1):201.

56. Liu J, Liu Y, Yan M. Spatial and temporal change in urban-rural land use transformation at village scale-a case study of Xuanhua district, North China. J Rural Stud. 2016;47:425–34.

57. Liu W, Xue Y, Shang C. Spatial distribution analysis and driving factors of traditional villages in Henan province: a comprehensive approach via geospatial techniques and statistical models. Heritage Sci. 2023;11(1):185.

58. Wang J-F, Li X-H, Christakos G, Liao Y-L, Zhang T, Gu X, Zheng X-Y. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. Int J Geogr Inf Sci. 2010;24(1):107–27.

59. Chumney EC, Simpson KN. Methods and designs for outcomes research. Bethesda: American Society of Health-System Pharmacists (ASHP); 2006.

60. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression. J R Stat Soc: Series D (The Statistician). 1998;47(3):431–43.

61. Fotheringham AS, Yang W, Kang W. Multiscale geographically weighted regression (MGWR). Ann Am Assoc Geogr. 2017;107(6):1247–65.

62. Bian J, Chen W, Zeng J. Spatial distribution characteristics and influencing factors of traditional villages in China. Int J Environ Res Public Health. 2022;19(8):4627.

63. Jia A, Liang X, Wen X, Yun X, Ren L, Yun Y. Gis-based analysis of the spatial distribution and influencing factors of traditional villages in Hebei province. China Sustain. 2023;15(11):9089.

64. Zhang X, He J, Deng Z, Ma J, Chen G, Zhang M, Li D. Comparative changes of influence factors of rural residential area based on spatial econometric regression model: a case study of Lishan Township, Hubei province. China Sustain. 2018;10(10):3403.

Yuan *et al. Heritage Science*   *(2024) 12:395*

Page 22 of 22

65. Yu H, Luo Y, Liu S, Wang Y, Yang Y, Liu W. The influences of topographic relief on spatial distribution of mountain settlements in three Gorges area. Environ Earth Sci. 2015;74:4335–44.
66. Li Y, He J, Yue Q, Kong X, Zhang M. Linking rural settlements optimization with village development stages: a life cycle perspective. Habitat Int. 2022;130: 102696.
67. Tian G, Qiao Z, Zhang Y. The investigation of relationship between rural settlement density, size, spatial distribution and its geophysical parameters of China using Landsat TM images. Ecol Model. 2012;231:25–36.
68. Chen Z, Liu Y, Feng W, Li Y, Li L. Study on spatial tropism distribution of rural settlements in the Loess Hilly and Gully region based on natural factors and traffic accessibility. J Rural Stud. 2022;93:441–8.
69. Smailes PJ, Argent N, Griffin TL. Rural population density: its impact on social and demographic aspects of rural communities. J Rural Stud. 2002;18(4):385–404.
70. Jia K, Qiao W, Chai Y, Feng T, Wang Y, Ge D. Spatial distribution characteristics of rural settlements under diversified rural production functions: a case of Taizhou. China Habitat Int. 2020;102: 102201.
71. Liu Y, Lu S, Chen Y. Spatio-temporal change of urban-rural equalized development patterns in China and its driving factors. J Rural Stud. 2013;32:320–30.
72. NASA, METI, AIST, Japan Spacesystems, and U.S., Japan ASTER Science Team: ASTER Global Digital Elevation Model V003. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/ASTER/ASTGTM.003, Last accessed on 12 May 2024 (2018)
73. Harris I, Osborn T, Jones P, Lister D. Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset, Scientific Data. https://doi.org/10.6084/m9.figshare.11980500, Last accessed on 12 May 2024 (2020)
74. Xu X. China GDP Spatial Distribution Kilometer Grid Dataset. Institute of Geographic Sciences and Natural Resources Research. https://doi.org/10.12078/2017121102, Last accessed on 12 May 2024 (2017)
75. NCHA: Map of National Key Cultural Relics Protection Units. http://gl.ncha.gov.cn/, Last accessed on 12 May 2024 (2023)
76. OSM Contributors: OpenStreetMap. https://www.openstreetmap.org, Last accessed on 12 May 2024 (2023)
77. Sims K, Reith A, Bright E, Kaufman J, Pyle J, Epting J, Gonzales J, Adams D, Powell E, Urban M, Rose A. LandScan Global Population 2022 Dataset. Oak Ridge National Laboratory. https://doi.org/10.48690/1529167, Last accessed on 12 May 2024 (2023)
78. Xu X, Liu L, Cai H. China Agricultural Potential Dataset. Institute of Geographic Sciences and Natural Resources Research. https://doi.org/10.12078/2017122301, Last accessed on 12 May 2024 (2017)
79. Kock N, Lynn G. Lateral collinearity and misleading results in variance-based sem: an illustration and recommendations. J Assoc Inf Syst. 2012;13:546–80.
80. Xie Z, Yan J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. J Transp Geogr. 2013;31:64–71.
81. Iversen GR, Gergen M. Statistics: the conceptual approach. Berlin: Springer; 1997.
82. Oshan TM, Li Z, Kang W, Wolf LJ, Fotheringham AS. Mgwr: a Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS Int J Geo Inf. 2019;8(6):269.
83. Sullivan JH, Warkentin M, Wallace L. So many ways for assessing outliers: what really works and does it matter? J Bus Res. 2021;132:530–43.
84. Wang S. A brief discussion on the eight taihang passages and their vicissitudes in the history. Geogr Res. 1997;16(01):68–76.
85. Zhang Y, Li S, Tan L, Zhou J. Distribution and integration of military settlements' cultural heritage in the large pass city of the great wall in the ming dynasty: A case study of juyong pass defense area. Sustainability. 2021;13(13):7166.
86. Yan Z, Zhao Y, Jiao H, Kou Y. The geological control of the yellow river alluvial fan on central plain cities. Earth Sci Front. 2010;17(06):278–85.
87. Feng W, Li A, Zhou W. Spatial pattern of rural settlements in the upper reaches of the Minjiang River: a case study in Maoxian county. Sichuan J Mountain Sci. 2007;4:146–54.
88. Gu C, Li Y, Han SS. Development and transition of small towns in rural China. Habitat Int. 2015;50:110–9.
89. Huang Y, Li E. Culture heritage utilization evaluation of historic villages: 10 cases in the Guangzhou and Foshan area, Guangdong province, China. J Asian Archit Build Eng. 2023;22(5):2994–3005.
90. Oliveira V. Morphological research in planning. Urban design and architecture. Berlin: Springer; 2021.

Publisher's Note