The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W2-2023
ISPRS Geospatial Week 2023, 2–7 September 2023, Cairo, Egypt

IMPACTS OF GEOSPATIAL FACTORS ON VILLAGE SITE SELECTION ALONG THE GREAT WALL WATCHTOWER BASED ON LOGISTIC REGRESSION: A CASE STUDY IN GUBEIKOU SECTION, BEIJING, CHINA

Tianyue Yang[1, 2], Xian Guo[1, 2] *, Jie Jiang[1, 2]

[1]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China - 2108570021126@stu.bucea.edu.cn, (guoxian, jiangjie) @bucea.edu.cn

[2]Key Laboratory of Urban Spatial Information, Ministry of Natural Resources of the People's Republic of China, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

KEY WORDS: Watchtower, The Great Wall, Village Site Selection, Heritage Conservation Planning.

ABSTRACT:

The Great Wall was listed as a World Heritage Site in 1987. In recent years, the Chinese government placed significant emphasis on the preservation of the Great Wall and its invaluable resources. However, it remains challenging to quantitatively analyze the impact of multiple factors and further identify the Great Wall resource sites. In this paper, a spatial quantitative analysis framework with binary logistic regression is implemented for the impact factors of village site selection along the Great Wall watchtowers. Elevation, slope, orientation, water distance, road distance, distance-to-watchtowers were selected as candidates. 177 villages were analyzed, and the model calculation was assessed with the confusion matrix and ROC curve. Youden index is used to determine the optimal threshold for the uncertain villages, subsequently. And we classified villages into three categories: villages with strong, medium and weak relevance with Great Wall resource sites. Experimental results show that the strong and medium relevance villages accounted for 68.36%. These villages are primarily located in the central part of the study area, characterized by flat terrain and closely aligned with the road network. Village density decreases with distance from the watchtowers. Additionally, out of the 40 uncertain villages, 5 were confirmed as villages while the remaining were classified as non-villages. The conclusions of this study can serve as a reference for evaluating and predicting suitable conservation and planning initiatives for the Great Wall and its surrounding villages in the future.

1. INTRODUCTION

The construction of The Great Wall can be dated to the eighth century BC as an old Chinese military defense. Listing it as a major historical and cultural site at the national level, the Chinese government attaches great attention to the conservation of the Great Wall by conducting investigations and value assessments of the Great Wall resources. The Plan of Beijing Municipality for Conservation & Development of The Great Wall Cultural Belt (2018-2035) was announced to enhance the preservation and development of Great Wall resources. This plan designates a range of approximately 5000 square kilometers for the Great Wall Cultural Belt (as shown in Figure 1), as well as assigning core areas, radiating areas, and resource sites. Among them, resource sites are defined as the villages that have evolved from the command centers, passes, castles, and forts along the Great Wall, which play a core role in the inheritance and development of the Great Wall culture.

However, after thousands of years of changes, the Great Wall villages which evolved from these structures may have been intermingled with some villages. Many historical and cultural information have been lost or damaged (Chen et al., 2020), which has negative impacts on the protection of the Great Wall. It is crucial to identify and categorize Great Wall resource sites, so as to implement precise preservation and sustainable development. To achieve this, spatial analysis methods will be employed to explore the relationship between villages and the Great Wall and interpret the spatial distribution pattern of the resource sites is the main target of this study. The outcome will support the precise protection of village culture within the Beijing Great Wall Cultural Belt.

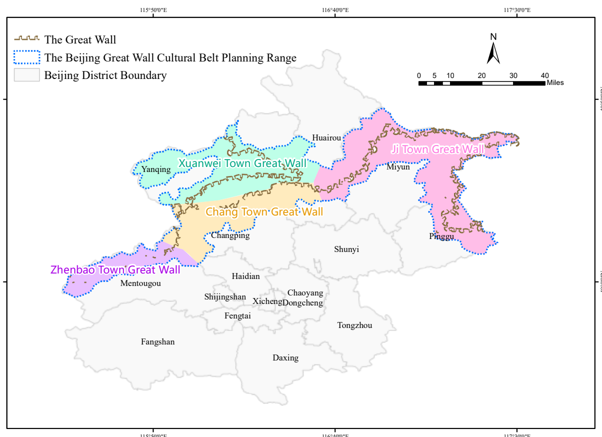


Figure 1. The range of The Beijing Great Wall Cultural Belt

It has been widely reported that the general site selection of villages mainly depends on geographical factors (Qi, Lu, Han, Ma, & Yang, 2022). The villages neighbouring to the Great Wall resource sites, also have a spatial relationship with the Great Wall that affects where they are located. As the Great Wall is composed of a variety of defensive buildings, including walls, watchtowers, gates, bastions, camp towns, guard stations, and beacon towers, the watchtowers are the most essential and symbolic elements among them. They were responsible for commanding, observing, delivering messages, storing weapons and food supplies for the defending soldiers, and also providing shelter for them.

Some studies discuss village spatial forms and patterns in the fields of village culture (Wang & Chiou, 2019; Xu & Wang,

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-375-2023

2021), rural tourism development (Bian, Chen, & Zeng, 2022), etc. There are a few of researches on the site selection of military settlements on the Great Wall. The function of castle-based military settlements in the military defence system of the Ming Great Wall (Du et al., 2021) and the geographic distribution of the villages in relation to the military settlements along the wall has been studied (Cao & Zhang, 2018). It remains challenging to quantitatively analyze the impact of multiple factors, such as topography, geography, and distance-to-watchtowers, and further identify the Great Wall resource sites.

In this paper, by integrating Great wall resource sites datasets, DEM, remote sensing images and other auxiliary data sets, we implemented a spatial quantitative analysis framework for the village site selection in Gubeikou section with a binary logistic regression and Youden index. Firstly, a forward stepwise regression was used to screen for impact factors associated with village site selection. Subsequently, a binary logistic regression method was used to quantify the relevance between village site selection and impact factors. Based on the results, the villages were classified according to their relevance with the Great Wall resource sites; Finally, the confusion matrix was used to calculate the model prediction accuracy, ROC curve to evaluate the model accuracy, and the Youden index was used to obtain the village siting threshold, which is an essential parameter for screening out the villages as Great Wall resource sites and the actual condition of uncertain villages.

This study focused on a 5 kilometers buffer zone around the Miyun watchtowers (No. 253 to No. 397) and Beiqi watchtowers (No. 1 to No. 3) as the study area in the Great Wall Cultural Belt. Elevation, slope, orientation, water distance, road distance and distance-to-watchtowers were selected as impact factors. 177 villages within the study area were analyzed quantitatively, and the model calculation results were validated with historical data.

The reminder of this paper is organized as follows. Section 2 describes the study area and data. Section 3 introduces the method of quantitative analysis and accuracy evaluation of the impact factors of village site selection model. Section 4 shows the experimental results, followed by analysis and discussion in Section 5. Finally, conclusions and future research directions are presented in Section 6.

2. STUDY AREA AND DATA

2.1 Study Area

Gubeikou town is the only national historic and cultural town on the Great Wall Cultural Belt, located in the northeast part of Beijing and under the jurisdiction of Miyun District. It occupies an area of roughly 85.82 square kilometers and is located between 40°36'38"N and 40°42'23" N and 117°03'58"E and 117°17'30"E. The Great Wall runs through the northwest and southeast directions of Gubeikou town, consisting of three Beiqi watchtowers (No. 1 to No. 3) and 144 Ming watchtowers (No. 253 to No. 397). The watchtowers are distributed on the north and east sides of Gubeikou town along the border with Hebei Province, with roughly equal distances between adjacent watchtowers, as shown in Figure 2.

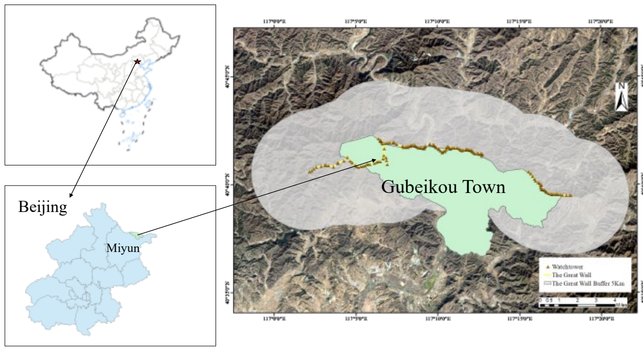


Figure 2. Study area

2.2 Datasets

2.2.1 Sample set of villages

This paper focuses on the villages that have naturally developed around the watchtowers of the Great Wall. These villages are described as "village names in the Tianditu(World map) and obvious building site outlines in remote sensing images". Then, using Tianditu (World map) and Amap, visual interpretation was used to calibrate the villages in ArcGIS. 177 villages were obtained after sketching the outline of the core area, recording names and information about related attributes, and taking village points from the central mass point (the center point with dense buildings) of their villages. In order to build the regression model, geographic data of non-village points were also collected, as village site selection is a binary classification problem. 911 non-village points were randomly selected within the study area. Among the 177 villages and 911 non-village points, 75% of the samples were randomly chosen as training samples, while the remaining 25% were selected as validation samples. The distribution of village and non-village samples in the study area can be seen in Figure 3.

2.2.2 Geographical data

(1) The dataset pertaining to the Great Wall watchtowers, castles, and beacon towers, as well as Beijing's zoning map, county boundaries, town boundaries, and land use classification maps were provided by Prof. Ding He's team at Beijing University of Civil Engineering and Architecture;
(2) 30m resolution remote sensing images of Beijing, captured by the Landsat sensor at a resolution of 30 meters, were obtained on January 24, 2020;
(3) 30m digital elevation model data, acquired by ASTER sensors, acquired in 2009;
(4) National road network, building outline, railroad and water system data from the national geographic information public service platform Tianditu (World map).

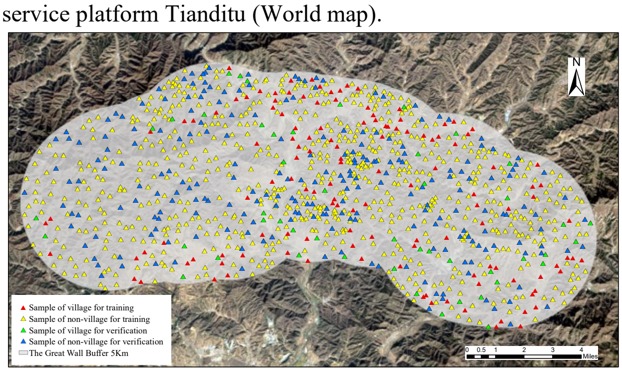


Figure 3. Spatial distribution of village and non-village samples in the study area

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-375-2023 | © Author(s) 2023. CC BY 4.0 License.

3. METHOD

3.1 Overall methodology and process

As outlined in Figure 4, the modelling process comprises five steps:

(1) Impact factor selection: Initial selection of impact factors based on fieldwork, literature review, and collection of geographic and historical data. A database of geographic and cultural elements was established.

(2) Model building: Village samples were selected for training, and a model was built using binary logistic regression. Forward stepwise regression was used to identify important variables associated with village site selection and create the optimal model that quantitatively expressed the relationship between impact factors and village site selection.

(3) Model accuracy test: The validation data was entered into the model built from the training data, and the accuracy of the model was checked by calculating the confusion matrix and ROC curve of the validation model.

(4) Optimal threshold calculation: Youden index is used to determine the optimal threshold for discussing of the uncertain villages based on the ROC curve.

(5) Investigation of geographical factors: The impact weight and trend of various geographical factors on village site selection were investigated in accordance with the findings from the model that quantitatively expresses the relationship between village site selection and impact factors. The model was used to explain the principle of geographical factors on village site selection, screening out the Great Wall resource sites (Mandal and Mandal, 2018), and use the optimal threshold to speculate out the actual condition of uncertain villages.

3.2 Selection of impact factors

Numerous villages are scattered around the Great Wall near Gubeikou town, having evolved from ancient forts and transformed through migration and expansion. Field research revealed that the Great Wall resources, such as walls and gates, are distributed among the Great Wall's surrounding villages（as shown in Figure 5). Interpretation of remote sensing images (as shown in Figure 6) also shows that the spatial distribution of Ming dynasty passes can still be roughly inferred based on the building and street layouts within many castles. This can be inferred that the selection of village sites in the northern mountainous region of Beijing was influenced by the construction of the Ming Great Wall and its distinctive military defence function, with distance-to-watchtowers playing a pivotal role.

In this paper, various variables such as the distribution of watchtowers and the administrative boundaries of the townships were integrated. Data from multiple sources, including the Local Chronicles of Miyun District and the Miyun District People's Government website were collected and reviewed. Using ArcGIS, a 5 kilometers buffer zone was established along the line connecting the watchtowers, with villages in this area selected as study objects.

The impact factors considered in this study include elevation, slope, orientation, water distance, road distance, and distance-to-watchtowers. Raster layers of elevation, slope, and orientation were derived from the analysis of Digital Elevation Model (DEM) data for the study area. Orientation was categorized into sunny and shady sides, with "east, southeast, south, and southwest" representing sunny sides and "north, northeast, west, and northwest" indicating shady sides. To assess transportation, water source, and military impact conditions, the straight-line distances between grid points for roads, water sources, and watchtowers were calculated using a 100m*100m grid.

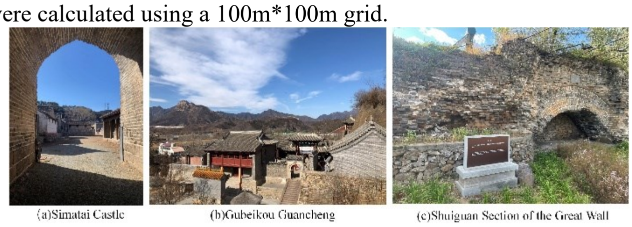


Figure 5. Field research found that the Great Wall resource sites are located in the villages of Gubeikou town

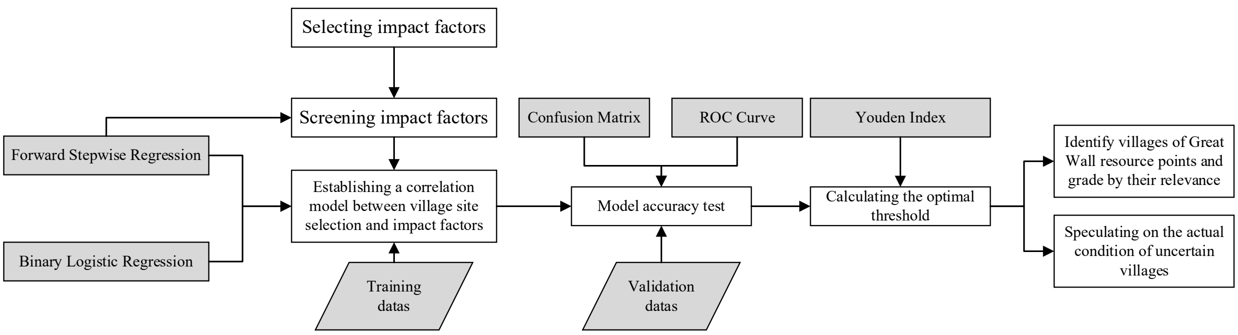


Figure 4. Flow chart of the implementation and analyzing methods

Table 1. Binary logistic regression optimal model results

Independent variable	B	S.E.	Wald	df	Sig.	Exp(B)	95%C.I.for EXP(B) Lower Bound	95%C.I.for EXP(B) Upper Bound
$Elevation$ (E)	-.004	.002	5.324	1	.021	.996	.992	.999
$Slope$ (S)	-.245	.033	54.698	1	.000	.783	.733	.835
$Orientation$(O)	1.365	.253	29.104	1	.000	3.917	2.385	6.434
$Road\ distance$(RD)	-.840	.302	7.761	1	.005	.432	.239	.779
$Distance\_to\_watchtowers$ (DtW)	.307	.095	10.539	1	.001	1.360	1.129	1.637
α	.983	.650	2.283	1	.131	2.671		

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-375-2023 | © Author(s) 2023. CC BY 4.0 License.

Figure 6. A castle evolved into a settlement can be interpreted with the remote sensing image (Geyu Castle)

3.3 Quantitative analysis methods

3.3.1 Forward stepwise regression

The forward stepwise regression was used to screen for impact factors. The input factors $X \in \{x_1, x_2, \ldots x_n\}$ will be introduced into the model one by one, and establish univariate regression model with the dependent variable Y, respectively. The formulation can be denoted as:

$$Y = \beta_i X_i + \epsilon, \quad i = 1,2,\ldots,n \tag{1}$$

X_i is used to the univariate model to obtain estimates of the parameters, which is found by applying the F-test to the equations $F_{max}^{(1)} = max\left\{F_1^{(1)}, F_2^{(1)}, \ldots, F_n^{(1)}\right\}$.

3.3.2 Binary logistic regression

Binary logistic regression was used to analyze the impact of multiple factors on the village site selection (Zheng et al., 2021). The equation is created by substituting into with the samples of "Village" and "Non-village" as the dependent variable and each impact factor $x_1, \ldots, x_i$ as the independent variable. The formula is as follows:

$$p = \frac{e^{\alpha + \sum_{k=1}^{k} \beta_k x_k}}{1 + e^{\alpha + \sum_{k=1}^{k} \beta_k x_k}} \tag{2}$$

3.3.3 Model accuracy validation based on ROC curve and confusion matrix

The accuracy was assessed with the confusion matrix and ROC curve (Hanley and McNeil, 1982). The confusion matrix (Sammut and Webb, 2011) includes four situations: true positive sample (TP), false positive sample (FP), true negative sample (TN), and false negative sample (FN). The Sensitivity, Precision, Specificity, and Accuracy of model can be calculated using these four variables.

3.3.4 Optimal threshold selection based on the Youden index

Youden index (Fluss et al., 2005) was used to determine the optimal threshold for discussing of the uncertain villages based on the ROC curve.

$$Youden\ index = max(Sensitivity + Specificity - 1) \tag{3}$$

This index is frequently used to evaluate the accuracy of imbalanced binary classification models in terms of sample size. The actual number of villages in this case is much less than the number of non-villages, and the Youden index is applicable to this type of sample distribution.

4. MODEL RESULTS

4.1 The relative influence of the impact factors

Figure 7 illustrates the kernel density statistics of the six impact factors. The slope of village samples is mainly concentrated between 0 and 10 degrees (Figure 7a). Non-village samples show higher density on the shady side, indicating the importance of sunny sides in village site selection (Figure 7b). Village samples have elevations mostly between 200 and 400 meters, while non-village samples are concentrated between 300 and 600 meters (Figure 7c); The villages are typically located within 0.5 km of roads (Figure 7d). The distance between the village samples and water source mainly concentrates between 0.5 and 2 kilometers. However, the sparse sample density between 0 and 0.5 kilometers indicates that villages are not usually located close to water sources(Figure 7e); Both villages and non-villages show a uniform density distribution within the 0-5 kilometers from the watchtowers (Figure 7f). This indicates that villages are typically situated at a specific distance from the Great Wall.

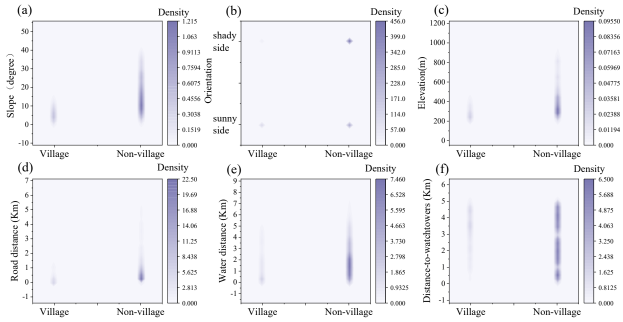


Figure 7. Kernel density statistics of impact factors

4.2 Binary logistic regression model results

The regression analysis employed a stepwise approach to establish five models, with each model introducing new impact factors based on the previous one. The final model in the last step is the optimal model (Abdelmutalab et al., 2016). The results showed that the factors of elevation, slope, orientation, road distance, and distance-to-watchtowers were retained in the model. The weight of each variable, after five iterations, can be found in Table 1. The exponential regression coefficient, denoted by Exp(B), signifies the change in the odds of the dependent variable when there is a one-unit increase in the independent variable. It can be regarded as the indicator of the relationship between village location and the factors surrounding the Great Wall. In this context, Exp(B) greater than 0 indicates a positive relevance between the specific factor and village location, while a value less than 0 implies a negative relevance. The final regression equation for the model of Great Wall village location is as follows:

$$P = \frac{e^{0.983 - 0.245*S + 1.365*O - 0.004*E - 0.84*RD + 0.307*DtW}}{1 + e^{0.983 - 0.245*S + 1.365*O - 0.004*E - 0.84*RD + 0.307*DtW}} \tag{4}$$

The significance of the optimal model in the Hosmer and Lemeshow test for the training model was 0.7 (as shown in Table 2), which is greater than 0.05. This indicates that the equation fits well and demonstrates a good overall fit between the predicted and observed values.

Table 2. Hosmer and Lemeshow test results for the optimal model

Step	Chi-square	df	Sig.
5	5.528	8	.700

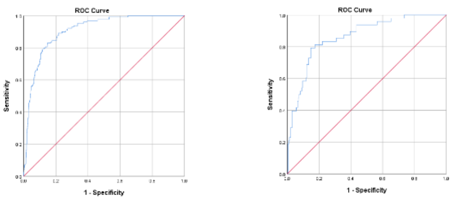
4.3 Model validation

According to the ROC curve, the accuracy of 0.863 for the validation model (Figure 8), which corresponds to an 86.3% model prediction success rate for the validation sample (Table 3). The curve characteristics imply that the model has reference value for village site selection exploration when 0.7<AUC<0.9(Greiner et al., 2000).

Table 3. The area under the ROC curve

	Area	Std. Error	Asymptotic Sig. b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Training samples	.909	.012	.000	.885	.933
Validation samples	.863	.028	.000	.808	.918

In addition, the Youden index was used to determine the optimal threshold value for the model in the ROC curve, and samples with feature values greater than 0.227 were labelled as "village" and those with feature values less than 0.227 as "non-village". There were 36 true positive samples (TP), 34 false positive samples (FP), 196 true negative samples (TN), and 12 false negative samples (FN). And the accuracy of the model calculated by the confusion matrix is 83.5% (Table 4).



(a) (b)

Figure 8. ROC curve: (a)Training model (b)Validation model

Table 4. Confusion matrix at the optimal threshold

TP	TN	FP	FN	Sensitivity	Precision	Specificity	Accuracy
36	196	34	12	.750	.514	.852	.835

5. ANALYSIS AND DISCUSSION

5.1 Impact factors of village site selection

5.1.1 Elevation

The model results (Table 1) show a negative relevance between elevation and village site selection, indicating that for every 1-meter increase in elevation, the probability of village site selection decreases by 9.96% (Exp(B) = 0.996). Additionally, the majority of the village samples in the training area have elevations ranging from 200 meters to 400 meters. This is due to steep terrain, limited resources, and inconvenient living conditions in mountainous regions. Therefore, it is preferable to develop village sites at lower elevations.

5.1.2 Slope

There is a negative relevance between slope and village site selection, with each 1-degree increase decreasing the probability of village site selection by 7.83% (Exp(B) = 0.783) (Table 1). The slope of the village sample in the training area is mainly distributed between 0 and 10 degrees. This pattern can be attributed to the study area being located in a mountainous region, where the slopes of the mountain peaks become steeper as elevation increases. The excessively steep slopes pose challenges for agricultural and construction activities, leading people to choose locations with more favourable slopes for living.

5.1.3 Orientation

Village site selection and orientation are positively correlated, with a higher probability of selecting a sunny site compared to a shady site (Exp(B) = 3.917) (Table 1). This observation can be explained by the better climate of sunny side, especially in the agricultural field, where sufficient sunlight is the second most critical factor in addition to water conditions. The kernel density map depicting the distribution of orientation provides evidence supporting the aforementioned conclusion.

5.1.4 Road distance

The regression analysis shows a negative correlation between road distance and village site selection. For every 1 km increase in road distance, the probability of village site selection decreases by 4.32% (Exp(B) = 0.432) (Table 1). Roads play a crucial role in the development, productivity, and daily activities of village inhabitants. Considering the strategic significance of the Great Wall in military defence and its placement in mountainous terrain, the surrounding area can be difficult to access. Consequently, the construction of roads becomes crucial in enabling residents to lead a normal life and maintain connections with the outside world.

5.1.5 The distance-to-watchtowers

The distance-to-watchtowers and village site selection are positively correlated, as evidenced by a 13.6% increase in the probability of village site selection for every 1 km increment (Exp(B) = 1.360) (Table 1). In practice, villages are primarily located within a range of 0 km to 5 km, with a concentrated distribution between 2.5 km and 4.5 km. This pattern is influenced by the strategic placement of watchtowers along the Great Wall for defence purposes. Watchtowers are strategically placed on mountain peaks to enhance surveillance capabilities. Due to the living conditions on mountain peaks were challenging, the proximity to watchtowers posed a significant threat to the safety of the local population. Therefore, selecting village sites at a reasonable distance from the watchtowers is crucial to ensure the security of the inhabitants.

5.1.6 Water distance

The distance between villages and river systems is not the main factor influencing village site selection, contrary to previous understanding. In the Gubeikou area, village site selection takes into account both flood vulnerability and accessibility to water diversion. This ensures a reliable water supply for sustenance and economic activities. Instead of being directly located near water sources, villages in this area are often situated on higher ground. This strategic decision strikes a balance between reducing the risk of floods and maintaining accessibility to water resources. Also, a destructive flood in the Chaohe River in the 1960s severely damaged the inhabitants and completely destroyed the towns of Gubeikou region and the surrounding villages. Consequently, the residents of Gubeikou were forced to relocate, and in order to prevent future flood damage, a ditch and canal system was constructed in Chaoguan Village for water storage. This system serves as a protective measure during floods and also enables the stored water resources to be reused for irrigation, promoting overall water resource recycling.

5.2 Relevance between villages and the Great Wall

Elevation, slope, and distance-to-watchtowers are three of the six factors that are associated with the location of the Great Wall. The Great Wall is strategically positioned on high-elevation mountains with steep slopes for military defence purposes. Watchtowers are integral to the defence system, and their proximity to settlements plays a vital role in assessing their relationship. Based on the regression equation (Equation 4), we classified villages into three categories: villages with strong relevance (0.573-0.892), medium relevance (0.307-0.573) and weak relevance (0-0.307) with Great Wall resource sites. The red dots represent weak relevance villages, the yellow dots represent medium relevance of villages, the green dots represent the strong relevance villages, and the distribution is shown in Figure 9.

Among the total villages, 61 villages (34.46%) were classified as having strong relevance to Great Wall resource sites, as depicted in Figure 9. These villages are predominantly located in the central area of Gubeikou town. There were also 60 villages (33.9%) categorized as having medium relevance. In the mountainous regions of the east and west, characterized by high elevation, steep slopes, limited water resources, and challenging transportation, there are 56 villages (31.64%) identified as having weak relevance. The combined percentage of strong and medium relevance villages is 68.36%. The concentration of villages in the central part of the study area, with a distribution pattern closely aligned with the road network, can be attributed to the preference for flat terrain. This choice is driven by the need for suitable conditions for agricultural cultivation and the construction of roads. Another essential component of human production and daily existence is transportation and food. Therefore, topography and geographical location are the primary factors that people consider during village site selection.

The distance-to-watchtowers is a key factor in village site selection, as observed in the denser distribution of villages further away from the center of watchtowers. Among the villages, those with strong and medium relevance are more prevalent compared to weak relevance villages. Due to the geographical characteristics of the Great Wall, which is primarily situated on mountain ridges, there are fewer scattered villages in the eastern and western hilly regions. During the Ming Dynasty, the Gubeikou region experienced the establishment of distinctive military towns along the Great Wall. These historic villages exhibit various cultural expressions, including military defence, temples, transit, and tombs. In contrast, the majority of villages in the area are relatively small and are thought to have originated from forts that served as training grounds for soldiers and storage facilities for grain during times of war. This transformation can be attributed to the evolution of traditional villages, which have disappeared with the loss of the military function of the watchtowers, and have been replaced by modern villages for living and agricultural development.

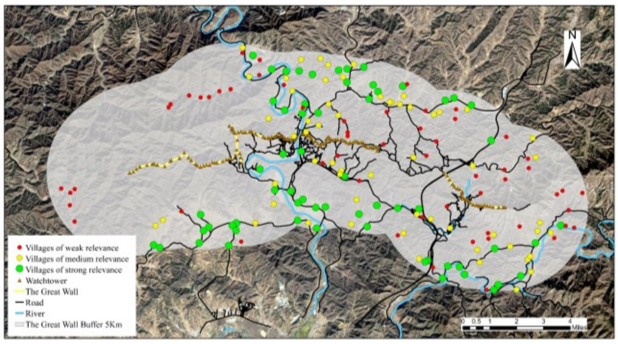


Figure 9. Villages were classified according to their relevance to Great Wall resource sites. The red dots represent weak relevance villages, the yellow dots represent medium relevance of villages and the green dots represent the strong relevance villages.

5.3 Discussion of "uncertain villages"

The planning of villages along the Great Wall plays a vital role in the establishment of the Great Wall National Cultural Park. Through geographical analysis, we can anticipate the future development of various types of villages. Some villages with favourable ecological conditions and historical resources are suitable for tourism development. However, further investigation is required for remote and small villages. If these villages encounter challenges such as deteriorating infrastructure and population decline, potential measures such as village relocation and consolidation may need to be considered. Due to the delay in updating the underlying data, a particular category of villages—described in this work as "uncertain villages"—exists that have names that are visible on the map but lack clear building sites in the remote sensing images. In order to explain the reasons for the "disappearance" of these villages from remote sensing images from a geographical perspective, this paper attempted to speculate the status of " uncertain villages" using the model established above.

40 "uncertain villages" candidates were firstly manually identified according to the principle that they have visible names on the map but lack clear building sites in the remote sensing images, using Tianditu (World map) and remote sensing image. Our results indicate that the results of five villages, namely Qingshanzi (0.485), Dongling (0.390), Jinshangen (0.377), Shaligou (0.367), and Lijiagou (0.263), exceeded the threshold. These villages are labelled red in Figure 10. Four of these villages were classified as having medium relevance to the Great Wall resource sites, while “Lijiagou (0.263)” having weak relevance. According to the spatial distribution of the rest 35 candidates presented in black dots, these areas are predominantly found in mountainous and gully locations, away from the Great Wall and the communities situated along the tide river and the plain area, and identified as extinct villages or non-village areas according to our model. Such locations often have unfavourable terrain for agriculture, transportation, and human habitation. And these villages lack unique qualities like military settlements and cultural centers with the Great Wall feature to encourage tourism development and have only limited connections to the Great Wall. Furthermore, the outmigration of young individuals in search of employment opportunities has exacerbated the population loss in these villages (Wilson et al., 2018). Consequently, the combined effects of shifting focus and declining population over time have resulted in the gradual disappearance of these villages.

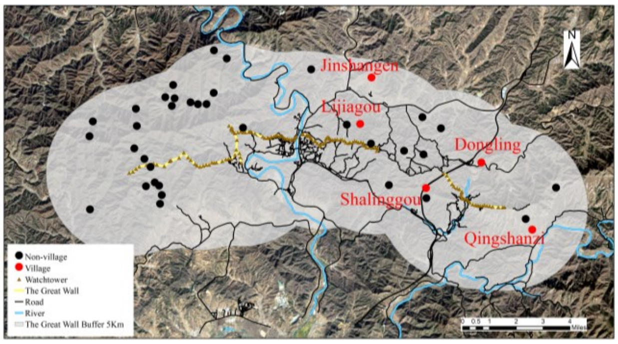


Figure 10. Results of using the model to speculate on the actual situation of uncertain villages. The red dots represent possible villages and the black dots are identified as non-villages.

6. CONCLUSIONS AND FUTURE WORK

This study investigated the impact factors of village site selection neighbouring the watchtowers of the Great Wall in Gubeikou town. The analysis included 177 villages located within a 5-kilometer radius of both the three Beiqi watchtowers (No. 1 to No. 3) and the 144 Ming watchtowers (No. 253 to No. 397) along the Great Wall. A comprehensive spatial quantitative analysis framework was implemented, considering factors such as elevation, slope, orientation, water distance, road distance, and distance-to-watchtowers. The accuracy was assessed with the confusion matrix and ROC curve. And Youden index was used to determine the optimal threshold for discussing of the uncertain villages based on the ROC curve. Finally, the villages were analyzed by classification according to the relevance to the Great Wall resource sites. The findings indicate that:

1.The watchtowers of the Great Wall have a significant impact on the village site selection, resulting in a distinctive spatial distribution pattern where nearby villages form a belt-like arrangement surrounding the watchtowers. The density of villages gradually increases from the inner region towards the outer region, with the watchtowers acting as the focal points. This indicates that the traditional settlements originally situated in close proximity to the watchtowers have gradually disappeared over time, as the military significance of the watchtowers diminished. These settlements have been replaced by modern villages that serve residential and agricultural purposes. To overcome the challenges posed by the unfavourable geographic conditions near the watchtowers, people have opted to relocate and expand towards more favourable plains located at a greater distance.

2.Elevation, slope, orientation, road distance, and distance-to-watchtowers were identified as the primary factors influencing village site selection in the study area. Among the villages located within the 5-km buffer zone around the Great Wall watchtowers, 68.36% villages showed strong or medium relevance with these factors. These villages were characterized by flat terrain, ample sunlight, well-established road access, and were considered suitable for human habitation. Interestingly, the impact of water distance on village location was found to be relatively limited. This could be attributed to factors such as the risk of flooding and the presence of table fields and ditches. In these cases, villagers tend to choose intermediate locations that strike a balance between proximity to water sources and maintaining a safe distance.

3.We further performed a speculation on the actual situation of uncertain villages by optimal threshold calculated from Youden index. As a result, five out of 40 candidates were classified as villages. Non-villages and extinct villages are found to be located in mountainous areas, distant from the Great Wall and other villages. These regions are characterized by steep terrain and lack distinctive features associated with the Great Wall. The decline in population, due to outmigration and changing priorities, has contributed to the gradual disappearance of these villages over time.

Undoubtedly, this study has certain limitations that warrant further research. Firstly, the inclusion of additional natural factors such as vegetation, soil, and climate, as well as socio-economic factors like population distribution and GDP, would enhance the site selection modelling. Moreover, the analysis conducted in this study was predominantly from a macroscopic perspective, and incorporating factors from a microscopic perspective, such as the internal morphology of villages and house layouts, would be beneficial. Above generalization can effectively extend the application of the proposed framework to the analysis of specific small-scale scenarios, such as villages and communities, as well as the villages planning and protection in the entire Great Wall National Cultural Park.

ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Foundation of China (Grant 52178029), the Pyramid Talent Training Project of Beijing University of Civil Engineering and Architecture (JDYC20200322), and the Fundamental Research Funds for Beijing University of Civil Engineering and Architecture (X20044).

REFERENCES

Abdelmutalab, A., Assaleh, K., El-Tarhuni, M., (2016). Automatic modulation classification using hierarchical polynomial classifier and stepwise regression, *2016 IEEE Wireless Communications and Networking Conference. Doha, Qatar*, 1–5.

Bian, J., Chen, W., & Zeng, J. (2022). Spatial Distribution Characteristics and Influencing Factors of Traditional Villages in China. *International Journal of Environmental Research and Public Health*, *19*(8), 4627.

Cao, Y., & Zhang, Y. (2018). The fractal structure of the Ming Great Wall Military Defense System: a revised horizon over the relationship between the Great Wall and the military defense settlements. *Journal of Cultural Heritage*, 33, 159-169.

Chen, X., Xie, W., Li, H., (2020). The spatial evolution process, characteristics and driving factors of traditional villages from the perspective of the cultural ecosystem: A case study of Chengkan Village. *Habitat International*, 104, 102250.

Du, Y., Chen, W., Cui, K., Guo, Z., Wu, G., Ren, X. (2021). An exploration of the military defense system of the Ming Great Wall in Qinghai Province from the perspective of castle-based military settlements. *Archaeological and Anthropological Sciences*, 13, 1-18.

Fluss, R., Faraggi, D., Reiser, B., (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47, 458-472.

Greiner, M., Pfeiffer, D., Smith, R.D., (2000). Principles and practical application of the receiver-operating characteristic

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-375-2023

analysis for diagnostic tests. *Preventive veterinary medicine*, 45, 23-41.

Hanley, J.A., McNeil, B.J., (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Mandal, S., Mandal, K., (2018). Modeling and mapping landslide susceptibility zones using GIS based multivariate binary logistic regression (LR) model in the Rorachu river basin of eastern Sikkim Himalaya, India. *Modeling Earth Systems and Environment*, 4, 69-88.

Qi, J., Lu, Y., Han, F., Ma, X., & Yang, Z. (2022). Spatial distribution characteristics of the rural tourism villages in the Qinghai-Tibetan Plateau and its influencing factors. *International Journal of Environmental Research and Public Health, 19*(15), 9330.

Sammut, C., Webb, G.I., (2011). Encyclopedia of machine learning. Springer Science & Business Media.

Wang, H.-f., & Chiou, S.-c. (2019). Study on the sustainable development of human settlement space environment in traditional villages. *Sustainability, 11*(15), 4186.

Wilson, G.A., Hu, Z., Rahman, S., (2018). Community resilience in rural China: the case of Hu Village, Sichuan Province. *Journal of rural studies*, 60, 130-140.

Xu, Q., & Wang, J. (2021). Recognition of values of traditional villages in southwest China for sustainable development: A case study of Liufang Village. *Sustainability, 13*(14), 7569.

Zheng, Z., Morimoto, T., Murayama, Y., (2021). A GIS-based bivariate logistic regression model for the site-suitability analysis of parcel-pickup lockers: a case study of Guangzhou, China. *ISPRS International Journal of Geo-Information*, 10, 648.