

RESEARCH

Open Access



Mapping landscape in Longshan period's hierarchical society (3000–2000BCE) of North Loess Plateau: from archaeological predictive model to GIS spatial analysis

Jianxin Cui^{1*}

Abstract

On the North Loess Plateau of China, city civilization, social complexity, and stratification emerged during the Longshan period (3000–2000BCE). Based on Geographic Information System (GIS) analysis and archaeological predictive model, we conducted a comparative analysis of environmental characteristics between sites and non-sites, ordinary and walled city sites, as well as large and smaller city sites. Initially, we developed a Binary Logistic Regression (BLR) model to predict the locations of archaeological sites from this period. Our findings indicate a high predictive accuracy of the model, demonstrating a clear environmental preference by the people of the Longshan Period. The presence or absence of the site was found to be influenced by various factors, including temperature, elevation, river distance, and precipitation. Furthermore, we discovered that walled cities had higher environmental requirements compared to ordinary sites. Terrain and land use played a more significant role in shaping prehistoric cities than climate. Lastly, the landscape in the Shimao site, which served as a most crucial and largest settlement centers on the North Loess Plateau (NLP), resembled that of other minor walled cities. Due to its abundant grassland, Shimao relied more on animal husbandry rather than agriculture. The combination of agriculture and animal husbandry has promoted the urbanization processes.

Keywords Longshan period, Environmental variables, Walled city, GIS, BLR

Introduction

The Longshan culture, a late Neolithic cultural type, was widespread in the middle and lower Yellow River regions, which can be divided into two periods based on cultural characteristics: the early Longshan culture (3000–2500 BCE) and the late Longshan culture (2500–2000 BCE). In ancient Chinese literature, this period is referred to as the age of 'ten thousand states,' indicating a society of chiefdoms [1]. During this time, population density increased,

leading to settlement differentiation and social stratification. The Longshan period exhibited evidence of complex social organization, urbanization, specialized crafts and technologies, including bronze metallurgy, jade carving, standardized pottery tiles, and competitive interaction [2]. Regional centers emerged, constructing walls with political, religious, and economic functions.

Previous research has indicated a significant cultural advancement during the Longshan period on the NLP [3]. The number of Longshan cultural sites has notably increased compared to the preceding Yangshao culture (ca. 5000–3000 BCE). Archaeologists believe that the increase of the site number reflects population growth and cultural prosperity under favorable climate conditions. While previous studies have explored the spatial

*Correspondence:

Jianxin Cui
cuijx@snnu.edu.cn

¹ Northwest Institute of Historical Environment and Socio-Economic Development, Shaanxi Normal University, Xi'an 710119, Shaanxi, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

distribution of sites in relation to factors like altitude and distance to the river [4], no research has yet been focused on the connection between human activities and comprehensive geographical landscapes during this time. The use of spatial analysis method in GIS has become popular for analyzing the relationship between human activities and landscapes, and has been successfully applied in various regions such as Eastern Iberia, Crete, Romania, India, and Mongolia [5–9]. The academic community is now focusing on a wider range of landscape factors, including water flow, humidity index, and soil quality [10]. These successful research endeavors also inspire us to investigate the natural foundations of the cultural prosperity during the Longshan period.

Archaeological predictive models have been widely used, particularly in the study of prehistoric period sites [11–16]. In the past, numerous models were employed in Cultural Resource Management (CRM), but in recent years, there has been a growing focus on the crucial role of models in landscape simulation and research [6, 17–19]. There are two methods for constructing an archaeological predictive model.

The first method is known as "inductive" or "data drive." Typically, BLR or MaxEnt statistical methods are utilized to determine the correlation between known archaeological sites and environmental variables [20, 21]. Although both MaxEnt and BLR models require the removal of strongly collinear independent variables, there is a difference in how they handle the following process. Once entered into the model, the MaxEnt model does not screen the independent variables, whereas BLR can further screen the independent variables based on statistical results. It removes factors that are not significant relation to the dependent variable. As a result, the BLR method provides a more concise final result.

The second method, known as 'deductive' or 'theory-driven,' involves the use of Artificial Intelligence (AI) and fuzzy logic models. AI models, such as one-class classifiers, have shown promise in detecting buried archaeological sites [22]. However, this method does not provide a quantitative relationship between sites and environmental variables. Fuzzy logic systems can be used to analyze settlement preferences [23]. But this approach involves a complex calculation process, particularly when establishing fuzzy sets and selecting parameters [10]. The final result is influenced by the expert's experience. The weight overlay method is employed to establish models in GIS and the weight of each variable is determined based on the knowledge or assumptions of experts [24]. In conclusion, most of the current deductive models rely on expert knowledge, which may introduce subjectivity and uncertainty.

Using the BLR model, this paper aims to address the following questions based on the high predictive accuracy: What were the landscape preferences of the Longshan people? Which environmental factor has the most significant impact on the site-present or absent?

Archaeological background

Existing archaeological site prediction models and GIS spatial analyses often overlook the inherent differences among sites, treating them as equal entities. However, during the Longshan period, the society on the NLP experienced increasing complexity and the emergence of a hierarchical structure. By 2300 BCE, the region had established complex trade networks and hierarchical social systems, accompanied by the increase of settlement number [1]. These sites can be categorized into at least three levels. Based on the presence or absence of stone walls, the sites can be divided into walled cities and ordinary settlements. The latter are generally smaller in size, suggesting that the walled enclosures reflected a degree of regional importance. Walled cities are further classified as either large central city sites Shimao or small to medium-sized city sites, based on site size and the quantity and value of unearthed artifacts. The recently discovered Shimao site (400 ha, ca. 2300–1800 BCE), a Neolithic walled city on the NLP, has sparked significant interest in academia regarding the Longshan period's social organization, hierarchy, and human-nature relationship [25–28]. It was not only one of the most crucial and largest settlement centers of the NLP but also in northern China. The presence of numerous exquisite jade artifacts, large stone statues, bronze ware, bone and pottery indicates that it was likely served as a civilized center with political, economic, and religious functions [29, 30].

Around Shimao, there are numerous smaller walled cities of various sizes situated on hilltops, alongside a significant number of ordinary settlement sites. This paper aims to examine the environmental variables of this hierarchical social structure. Specifically, our study aims to analyze the environmental factors that contribute to the presence of walled cities and determine if there are any notable distinctions between these cities and ordinary sites. Furthermore, we seek to investigate if spatial data in GIS can unveil significant environmental differences between Shimao (the mega city) and the intermediate or minor city sites.

Data and methods

Study area

According to administrative boundaries, the NLP spreads over two regions: the Yulin Region and the Yan'an Region. This article focuses on Yulin Region, which is the northernmost region of Shaanxi. Yulin is

confined by 36°57′– 39°34′N latitude and 107°28′– 111°15′E longitude. The area covers 43,113 km², accounting for 45% of Shaanxi region. The main watersheds in Yulin include the Kuye, Wuding, and Tuwei River and the areas of these three river basins are 8706km², 30260km² and 3294km², respectively. There are higher altitudes in the northwest compared to the southeast, with an average altitude ranging from 650 to 1500 m (Fig. 1). This area exhibits two types of geomorphology. The terrain in the northern portion of the aeolian plateau is characterized by its flatness, while the river valley in this area is broad but relatively short. The landscape primarily consists of fixed and semi-fixed sand dunes, interspersed with saline-alkali beaches and lakes. Moving towards the southernmost edge zone, mobile sand belts can be observed. On the other hand, the terrain in the southern loess hilly region displays a high relief, featuring dense gullies and alternating loess ridges and loess tablelands. This area is a transition zone between monsoon and non-monsoon regions.

The annual average temperature ranges from 8 to 12 °C, and the annual mean precipitation is between 350 and 600 mm, mostly occurring in summer [31].

The regression model

Binary Logistic Regression (BLR) has been utilized to establish the statistical relationship between independent variables and the dependent variable [32]. It is useful where the dependent variable is dichotomous (e.g., succeed/fail, live/die, site-present and site-absent). Assumptions and statistical requirements are not strictly adhered to. Normality of the variables is not necessary. However, the independent variables in the model should be non-collinear. In this study, the collinear test was conducted by the SPSS21 software before running the model (Additional file 2: Table S1). In comparison to other methods used for assessing relationships between predictive and dependent variables, BLR has shown reliability [33].

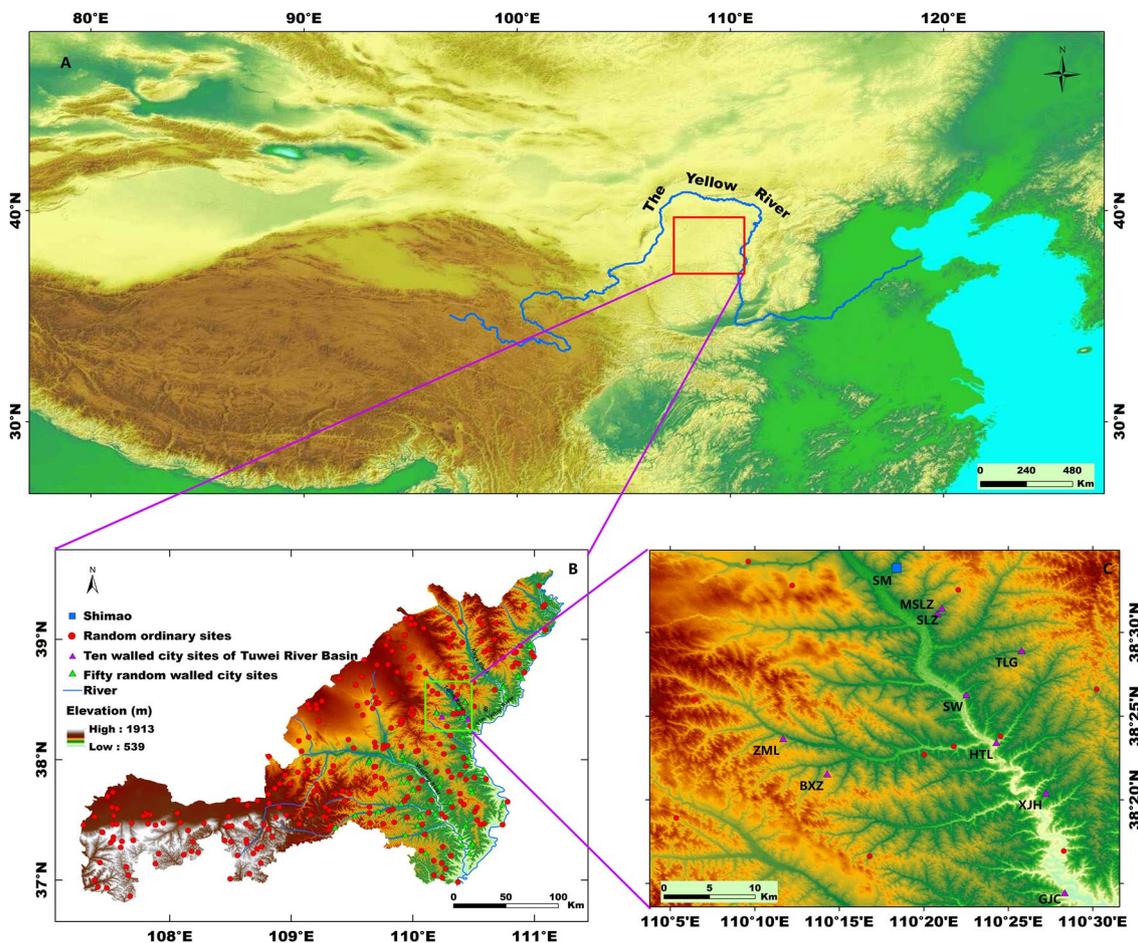


Fig. 1 A Location of the study area, China. B Map of the study area and the 270 random selected sites are shown. C Ten walled sites of Tuwei River Basin discussed in the paper

Dependent variable

The BLR model requires two types of location-specific input data: site-present and site-absent data. The first type can be obtained from known archaeological sites, while the second consists of random points generated by ArcGIS10.2 software. For this study, a total of 338 Longshan period sites have been incorporated into the GIS system. The dependent variable comprised 270 (80%) of the 338 sites with accurate location data and 270 random points were taken as the non-site locations. The model was then tested using the remaining 68 known archaeological sites and the same number of random points.

Predictive variables

Table 1 provides detailed information about the variables used in the model. The Aster DEM data with a resolution of 30 m was downloaded from the <https://www.gscloud.cn/>. This data was utilized to generate variables such as elevation, slope, relief (within 300 m), and aspect. The annual average precipitation, temperature data with a resolution of 500 m, and land cover data with a resolution of 1 km were obtained from the National Earth System Science Data Infrastructure of China. The river data was digitized from the topographic maps. Furthermore, the Distance to the river variable was calculated based on Euclidean distance tool in ArcGIS10.2. The Vegetation cover variable was computed from the Landsat 8 image using ENVI 4.8 software. All data was resampled to a resolution of 1 km. Over the past few thousand years, the geological background of this region appears tectonically stable with a single material composition, namely, loess. Thus, rock uplift has limited influence on landform evolution on the NLP. The erosion process dominated the landform evolution process during this period [34]. Although the soil erosion changed the specific values of elevation, slope, and aspect subtly. For example, the

downward erosion rate was 0.2 cm/ka in the NLP [35]. It would not change the landform's spatial distribution pattern. Therefore, the landform features put into the model was extracted from the modern DEM data. However, the reconstructed 0.5°×0.5° mid-Holocene temperature and precipitation data are not precise enough for small-scale regional studies [36]. Therefore, the paleoclimate raster data was derived by subtracting climate fluctuation values from modern data. The detailed reconstruction processes refer to supplementary materials (Additional file 4: Text S1 and Additional file 3: Table S2).

The environmental variables comparison of different social hierarchy sites

Fifty ordinary sites and 50 walled sites were randomly selected from the 338 sites using the ArcGIS10.2 software. To ensure reliable statistics, the buffer zone concept was often applied to environmental analysis of archaeological sites [37]. The average value of the environmental data within a 5-km radius of these sites was calculated. The calculated results were then exported to SPSS 21 for descriptive statistics analysis. Then, the potential environmental differences between the walled cities and the ordinary sites were examined. To further demonstrate the environmental disparities related to social hierarchy, the paper focused on the Tuwei River Basin, where the largest site, Shimao, is located. The precise geographic locations of ten walled cities of this area were obtained from published sources [29]. Subsequently, we compared the environmental variables of Shimao with those of the middle and small walled sites.

To enhance the understanding of the relationship between land use patterns and the distribution of sites among various social classes, this study has integrated more detailed land classification data. As there is no available prehistoric land use data, the 30-m resolution

Table 1 Variable names and descriptions

Variable name	Scale	Description
STATE (dependent)	Nominal	Variable with the values 1 = site presence and 0 = site absence
Elevation	Ratio	The height above sea level is measured in meters
Slope	Ratio	Slope in degree
Aspect	Ratio	Aspect in degree
Relief300	Ratio	An index to show the geomorphology feature, with the neighborhood 300m*300m
Precipitation	Ratio	A climate index in millimeter
Temperature	Ratio	A climate index in degree
Distance to water	Nominal	Distance to a water feature; It was classified into 6 classes (500, 1000,2000,5000,10,000,50,000 m are the intervals). In the model, Class one which is less than and equal to 500 m is the reference for the dummy variable settings)
Vegetation cover	Nominal	Five categories of Fractional vegetation cover (FVC) and 10%, 30%, 45%, 60% are the intervals
Land cover	Nominal	1 = farmland and 0 = non-farmland

land classification data from China in 1985 was utilized. It is important to note that significant efforts have been made since the 1950s to manage and reclaim sand land in the local area. Through the construction of numerous irrigation channels, previously barren land has been transformed into cultivable land. In order to minimize the impact of human activities, we have employed a model to modify the cultivated land in 1985 and have also revised the distribution of cultivated land and sandy land in another submitted paper [38]. This revised land use data can serve as a substitute for prehistoric land use, which was less influenced by human activity.

Results

Model output: human landscape selection preference

To identify the environmental variables that would have a significant effect on the model, potential variables were screened in SPSS21 (Table 1). The backward elimination (wald) method was used to eliminate the least significant variable at each iteration. A significance level (Sig.) of 0.1 was set when testing the variable’s correlation using Wald statistics [33]. The retained variables and their correlation coefficients are used to establish the model (Table 2). B represents the partial regression coefficient and S.E. represents the standard error. Wals is used to assess the impact of the independent variable on the dependent variable. A higher Wals or a smaller corresponding Significance (Sig.) indicates a more significant impact. Sig. denotes the significance level, represented by the P-value. Df represents the degrees of freedom. EXP (B) represents the odds ratio (OR), which quantifies the extent of an independent variable’s effect on the dependent variable.

The formula is as follows:

$$Z = -1.567 + (\text{Precipitation} * 0.008) + (\text{Elevation} * -0.004) + (\text{Temperature} * 0.405) + (\text{Distance (1)} * -0.292) + (\text{Distance (2)} * -0.803) + (\text{Distance (3)} * -0.578) + (\text{Distance (4)} * 0.116) + (\text{Distance (5)} * -18.786)$$

The following equation was used to convert the regression output to a probability score [24] and a predictive map was generated by this equation (Additional file 1: Figure S1). Here, e represents the base of the natural log system, Z is the optimal linear fit result of predictive variables.

$$\text{Prob(S)} = \frac{1}{1 + e^{-z}}$$

After evaluating the accuracy using two methods, it is shown that the model demonstrates high prediction accuracy (Additional file 5: Text S2) and can effectively explain the environmental preferences of human sites. Out of the ten variables initially included in the database, only four variables have demonstrated statistical significance: Temperature, Elevation, Precipitation, and Water Distance. Both Temperature and Precipitation exhibit positive coefficients, indicating that areas with higher temperature and precipitation values are more likely to contain a site. Specifically, a one-unit increase in temperature leads to a 1.5-unit increase in the probability of site presence (Table 2). The impact of precipitation, however, is relatively smaller compared to temperature. Additionally, there is a slight decrease in the probability of site presence as elevation increases.

The relationship between site presence and the Distance to the river is complex. The Distance to the river data was divided into six distinct classes: ≤500 m, 500–1000 m, 1000–2000 m, 2000–5000 m, 5000–10,000 m, and ≥50,000 m. In the model, this variable was taken as a dummy variable which are created to represent each category, allowing for easier interpretation of regression

Table 2 Significant environmental variables identified as being associated with site location

	B	S. E	Wals	df	Sig.(P)	Exp(B)
Precipitation	0.008	0.005	2.548	1	0.110	1.008
Elevation	− 0.004	0.001	15.131	1	0.000	0.996
Temperature	0.405	0.233	3.030	1	0.082	1.500
Distance			8.252	5	0.143	
Distance (1)	− 0.292	0.241	1.467	1	0.226	0.747
Distance (2)	− 0.803	0.299	7.226	1	0.007	0.448
Distance (3)	− 0.578	0.440	1.725	1	0.189	0.561
Distance (4)	0.116	0.660	0.031	1	0.861	0.123
Distance (5)	− 18.786	10263.906	0.000	1	0.999	0.000
constant	− 1.567	3.472	0.204	1	0.652	0.209

results [39]. In this case, the reference category is the distance between 0 and 500 m. The influence of other categories on the probability of site presence were compared to this reference level. Table 2 shows that only Distance (2) (1000–2000 m) significantly differs from the reference level. This suggests that the probability of site presence decreases as the distance from water increases. The last category ($\geq 50,000$ m) has a small number of observations data, which may explain why the difference from the reference level is not significant. However, it is known that there are no site-present instances within this distance range. Therefore, it can serve as a useful indicator for predicting non-site locations.

The regression model is not significantly affected by the variables that were removed. To evaluate the

relationship between site distribution and the removed variables, we plotted the frequency distribution of known sites in the training set across various environmental variables. The distribution of the sites is skewed towards low-slope and low-relief regions. Additionally, a significant proportion of sites face east and south (67.5–247.5). The vegetation cover of known locations is primarily centered on Class 2 and Class 3 ($10\% < FVC < 30\%$ and $30\% < FVC < 45\%$). The sites are located in both farmland and non-farmland areas (Fig. 2), and there are also numerous non-site points within the farmland. Although these variables play a role in determining the site’s location, they are not essential for the site’s existence. There is no statistically significant difference between sites and

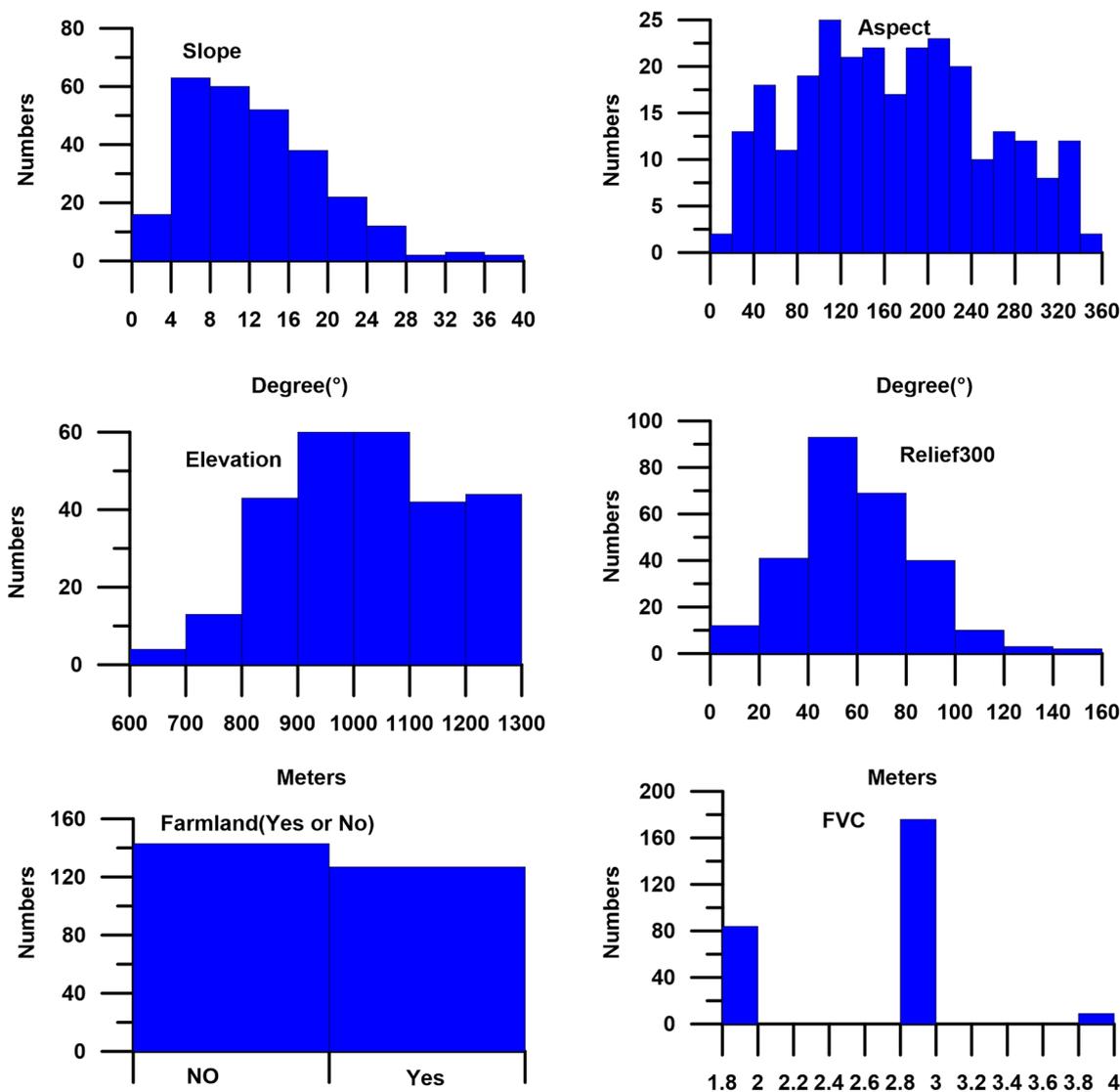


Fig. 2 Frequency distribution maps of known sites for the eliminated variables

non-sites in terms of these environmental factors, leading to their exclusion from the model (Fig. 2).

Through a comparative analysis of the values of two environmental variables, namely the variable included in the model and the variable excluded, for both site and non-site locations, we can determine which variable has a greater discriminatory potential. To illustrate this, we take temperature (the input variable) and slope (the removal variable) as the example. And the variable values obtained from site and non-site were compared. While there is some overlap in temperature values between sites and non-sites, the majority of non-sites have significantly lower temperatures compared to sites (Fig. 3). Therefore, temperature can effectively differentiate the two groups. A significance test was performed using SPSS21 software, which showed a significant differentiation between sites and non-sites (Table 3). On the other hand, the slope values of sites and non-sites completely overlap, making differentiation these two categories unattainable. Similarly, other variables have been excluded from the model for the same reason.

Environmental variables comparison on the walled city and ordinary sites

During the Longshan period, some sites became fortified with stone walls. The inter-regional climatic change, population pressure, and intergroup conflict perhaps have caused the construction of these walled cities [1]. The locations of 50 walled sites from the Longshan period were collected. These sites are usually located on hilltops near rivers, and stone walls were constructed on the top of natural cliffs to protect the settlements. The walled sites were often surrounded by smaller non-walled sites, indicating that these fortifications may have protected for the people of the surrounding area [1]. Therefore, these stone-walled sites have a crucial role in society. Here, we will investigate what types of environmental conditions supported the prehistoric walled city system and whether they differed from ordinary Longshan sites. For comparison, these environment data were divided into three categories (climate, terrain, and land use) and Excel bar graphs were generated (Fig. 4).

The walled city and ordinary sites have minor differences in temperature and precipitation. The former has a slightly higher mean temperature of 0.11 °C and lower precipitation of 0.99 mm. In terms of terrain

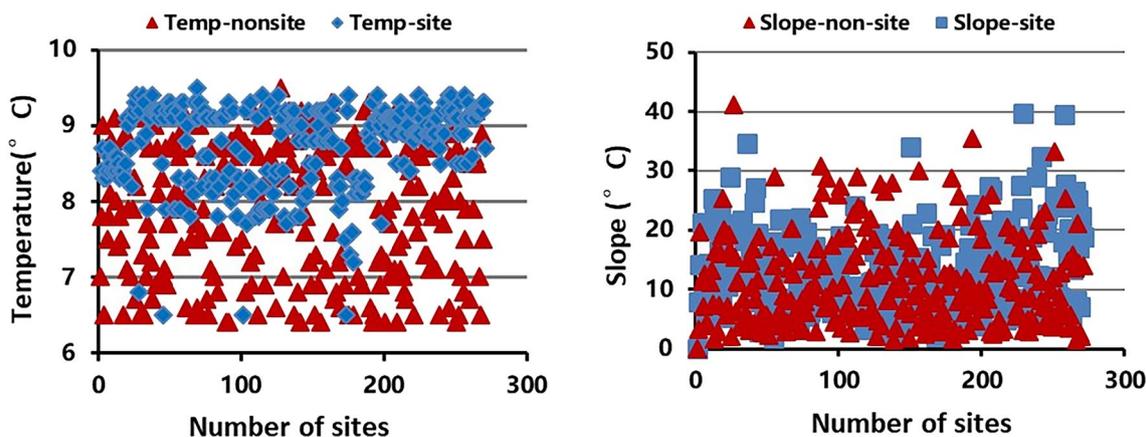


Fig. 3 Scatterplots of Temperature and Slope values extracted for the site and non-site location

Table 3 Temperature Independent Samples T-Test for sites and non-sites

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	97.217	.000	12.286	538	.000	-.7841	.0638	-.9094	-.6587
Equal variances not assumed			12.286	457.118	.000	-.7841	.0638	-.9095	-.6587

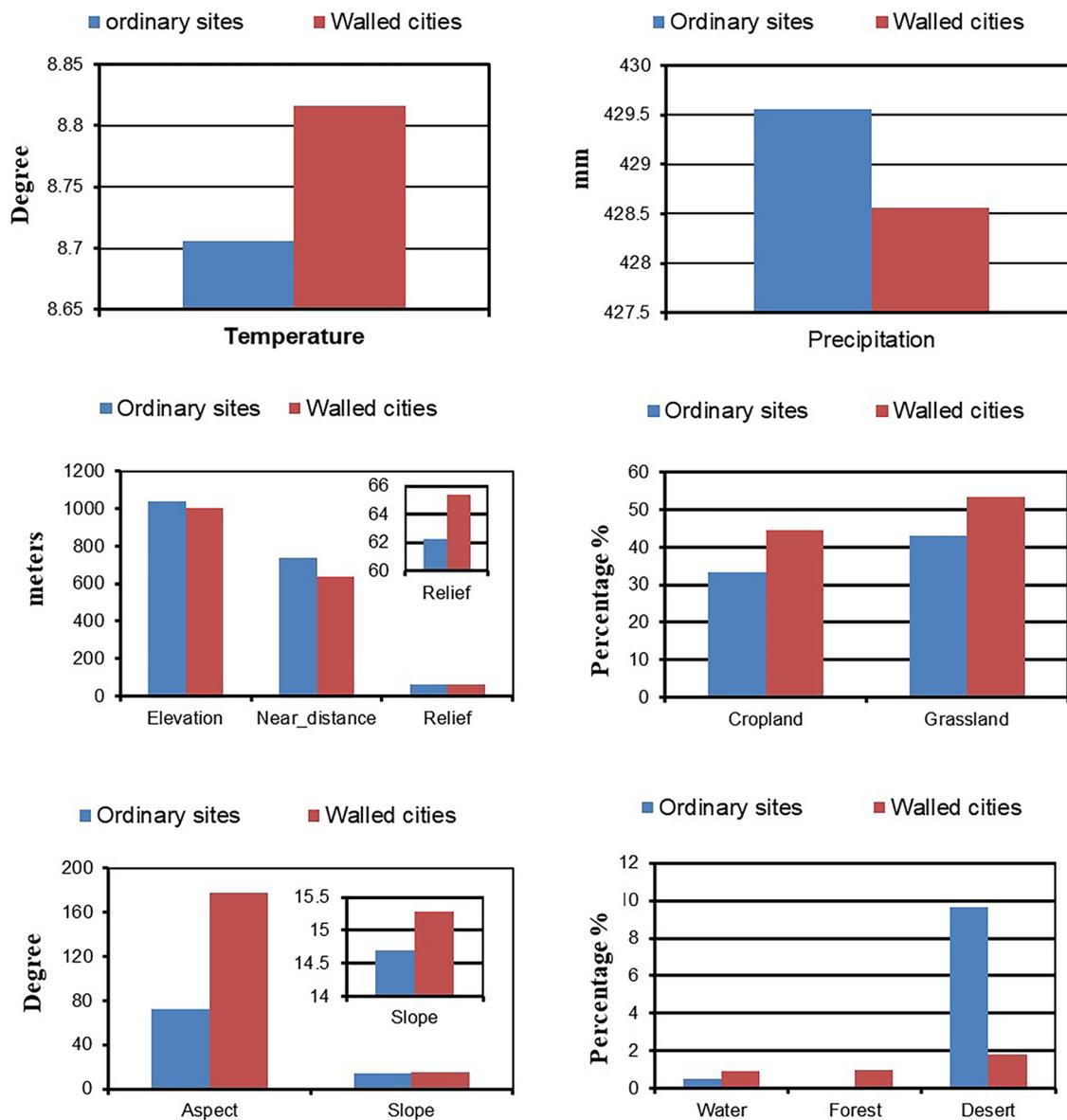


Fig. 4 Environmental variables comparison between the walled city and the ordinary sites during the Longshan period

factors, the walled city sites have lower elevation and shorter distance to the water compared to the ordinary sites. However, the walled city has higher slope and relief, which is advantageous for defense. Additionally, walled cities prefer southern-facing locations, while ordinary sites prefer north-facing locations, indicating significant aspect differences. The preference for south-facing locations may be due to factors such as better solar exposure, warmer climate, or cultural beliefs. All these factors suggest that larger sites, particularly city sites, have stricter criteria for location selection compared to ordinary sites (Fig. 4). This is because city sites serve

multiple functions, including political, economic, and military defense. As a result, the duration of city sites is extended, and environmental factors carry more weight in site selection.

The land use of various site types demonstrates significant diversity. Both walled cities sites and ordinary sites have a proportion of cropland area, with the former having a higher proportion. Additionally, walled city sites have more abundant grassland resources, which were considered valuable for farmers or nomads. It is worth noting that some ordinary sites are located within 5 km of deserted lands. Despite a short climate reversal

during the Longshan period, we can conclude that these locations were not the best choice for the sites due to their susceptibility to desertification. The cultural layers of several Longshan sites are situated just above the desert, indicating that the ordinary sites were sometimes located in suboptimal areas for survival. These places might have served as temporary dwellings with relatively low environmental requirements. Furthermore, out of the ten walled city sites in the Tuwei River Basin, two were surrounded by broad-leaved forests, while no forests were found around the 50 ordinary sites. This suggests that the presence of forests was a deliberate choice made by humans, who utilized wood as a source of fuel, building material, and tools.

Environment variables comparison between the largest city site Shimao and the other walled cities

The recently excavated Shimao site has uncovered a highly advanced city center in the NLP, previously considered a fringe region of Chinese civilization [29, 30]. The emergence of the first civilizations often took place in

areas with favorable geography for intensive agriculture [40–42]. These civilizations originated from agrarian communities that produced sufficient food to support cities, leading to the development of social hierarchies based on factors such as gender, wealth, and division of labor. In order to demonstrate the hierarchical structure of Longshan society, we conducted a comparison of environmental data between Shimao (SM) and other walled cities. However, no significant differences were found in these variables (Fig. 5).

Discussion

Settlement preference, the subsistence strategies and the war

The use of an archaeological predictive model allows for the identification of site location ‘s environment preferences. Despite the potential background noise caused by diverse environmental landscapes, this model enables us to explore the relationship between landscapes and human environmental preferences. The BLR model demonstrated that people in the NLP relied heavily on several

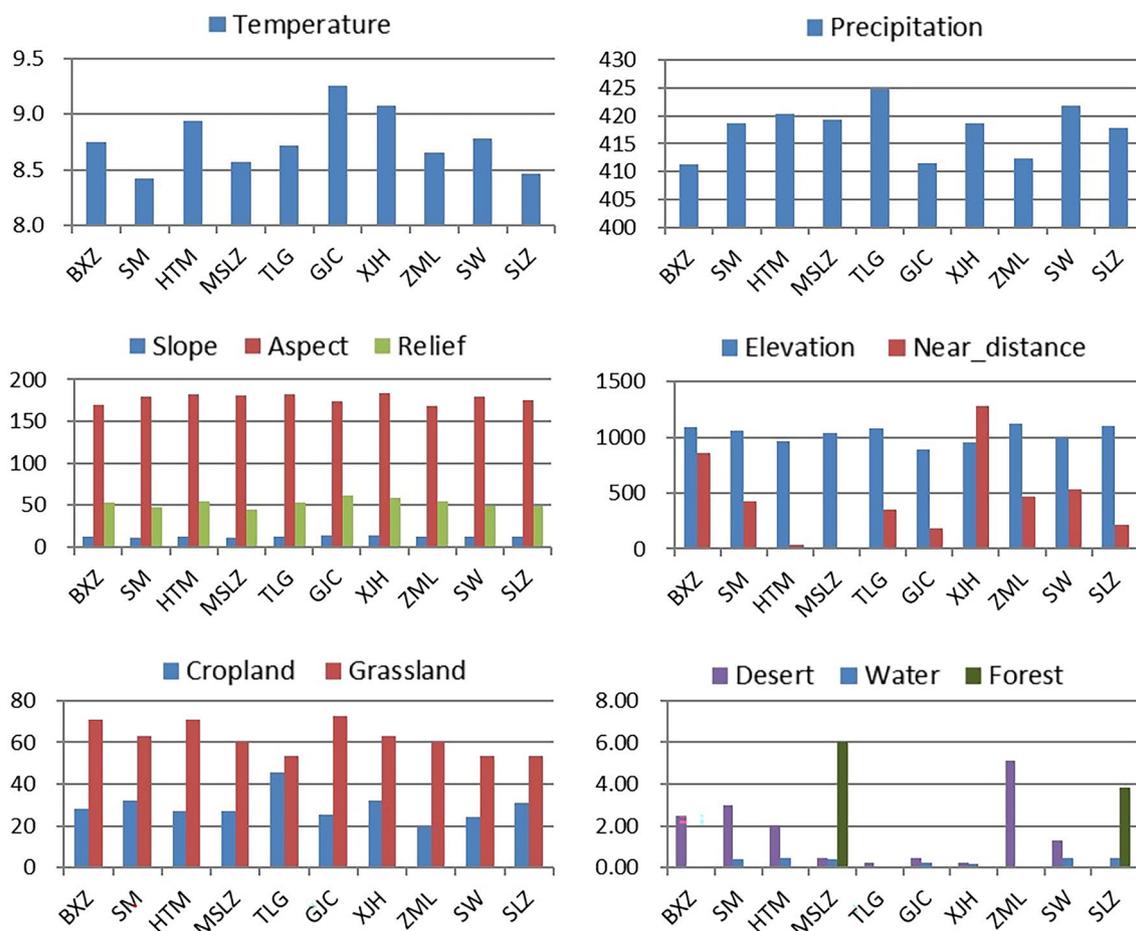


Fig. 5 Environmental variables for the 10 walled cities

key factors including Temperature, Distance to river, Precipitation, and Elevation when faced with diverse environmental conditions. By analyzing the predicted map of archaeological sites, a clear division between high and low probability areas can be observed (Additional file 1: Figure S1). The eastern part of the study area shows a high potential for archaeological sites. It was found that areas with higher average annual temperature, closer proximity to water sources, higher precipitation, and lower altitude were considered as optimal environments for human site selection.

The model excluded variables such as Aspect, Slope, Farmland, and Vegetation cover due to their limited ability to distinguish between sites and non-sites. During the Longshan period, many ordinary residences were partially underground and may not have required significant heating sources, thus eliminating the Aspect factor. The relationship between vegetation coverage and human site selection is intricate, and further research is necessary to fully understand it. Moreover, the model assumes that agriculture was the primary focus during the Longshan period, implying a strong correlation between arable land distribution and the existence of sites. Additionally, slope should also be a significant factor limiting human survivability, since steeper slopes are not conducive to agricultural production. However, the model's results do not align with the expected outcomes, which suggests that human survival was not significantly impacted by the slope factor during that period.

First, this discrepancy could be attributed to the location of the study area in the ecological edge zone, where the influx of cattle and sheep has led to changes in traditional production methods and the emergence of animal husbandry as a predominant activity [43, 44]. Previous studies have also indicated that compared with Yangshao culture (5000–3000 BCE), there was a tendency for the population during the Longshan period to migrate towards areas with steeper slopes [45]. This suggests that human living spaces expanded beyond flat agricultural areas, with the utilization of mountain and grassland resources to enhance animal husbandry. The presence of milk collecting at multiple Longshan period sites in the research region further supports this notion [46, 47]. The diachronic statistical analysis of ancient human bone collagen carbon isotopes in Shaanxi, Shanxi, and other regions reveals a peak followed by a decline approximately 4500–4000 years ago. This suggests that the shift in carbon isotope values was more likely due to the consumption of C3-fed beef, lamb, and milk rather than the introduction of wheat [48, 49]. These protein sources would have enhanced the nutrition and fitness of our ancestors, contributing to their survival and the resilience of human society. The coexistence of agriculture

and animal husbandry has played a crucial role in stimulating economic development and maintaining cultural prosperity within a relatively short period of time.

Second, steep slopes are a very good natural defense system [37]. The sites are usually located on hilltops near rivers, and stone walls were constructed on the top of natural cliffs to protect the settlements. The construction of these fortifications may have been the result of intergroup conflict, which was at least partially caused by climatic fluctuation and deterioration of the ecosystem, coupled with population pressure [1]. Shimao, a significant settlement center, encompassed a range of defensive structures in its eastern gate. These structures included a well-protected gate with guardhouses, two gate towers, and curtain walls with bastions and a corner tower [29]. The study in the Plateau-plain Transition Zone of NE Romania also indicated that prehistoric communities preferred to place their settlements for defensive purposes on hilltops, or in the close proximity of a steep slope [37]. It can be seen that the ecological transition zone was the forefront of conflicts between different groups of people, and wars occurred more frequently. Therefore, the landform of settlement location broke through the distribution rule of traditional agricultural sites.

The drivers of early urbanization and Shimao's environment preference

The emergence of a mixed agriculture and animal husbandry economy has contributed to the process of urbanization. As the population in Inner Mongolia migrates southward, there has been an increase in population pressure on the NLP area [50, 51]. This population growth has resulted in intensified competition for resources, greater social complexity, and the development of cities with defensive functions. Previous studies have associated urbanization with advancements in agricultural technology [42, 52]. Improved agricultural techniques have led to increased food production, enabling the support of larger populations and driving urbanization. However, during that time, agricultural production in the NLP region was low and unstable. Based on ¹⁵N enrichment, research indicated a significant decline in millet farmlands in the study area from the late Neolithic to the early Bronze Age (5000–3000 BP) [48]. The introduction of animal husbandry has allowed the region to sustain a larger population and increased human resilience to natural disasters. This suggests that urbanization was not limited to the most developed agricultural regions. As long as there was sufficient means of subsistence, even ecological margins could undergo urbanization.

In terms of agricultural production, the Shimao site does not possess any environmental or resource

advantages compared to smaller walled cities. However, archaeological excavations have revealed that the Shimao site held the highest social status, indicating that social stratification was more evident in social relations rather than in agricultural production. The significant proportion of grassland of Shimao provided a basis for the development of a mixed pastoral-agricultural economy, as the growing population faced the challenge of limited land resources. The introduction of cattle and livestock from the west further supported the animal husbandry industry, enabling the grasslands to produce meat and milk for human consumption. The discovery of a substantial number of bone needles in the city suggests that the inhabitants were skilled in crafting bone objects, such as needles for sewing animal hides [53]. This new subsistence strategy enhanced productivity by reducing dependence on agriculture and laid the economic foundation for the growth of major cities. It also implies that agricultural food in large cities or centers of civilization may not have been produced independently. During this period, the trade network likely played a crucial role in supplying the needs of the Shimao people. Many grassland communities in Chinese history adopted this model and relied on international trade for essential production materials. However, frontier cities faced greater challenges in achieving sustainable development compared to cities in the Central Plains, which enjoyed better environments and abundant resources.

Factors that impact the model accuracy

In archaeological landscape research, researchers typically focus on topographic and hydrological variables [22, 54, 55]. However, climate factors are rarely incorporated into the model due to the lack of available paleoclimate grid data. This paper aims to address this gap by utilizing quantitative paleoclimate data which was reconstructed from modern raster data by subtracting the difference between past and modern. Currently, this method is still relatively simple. In the future, collaboration with climate simulation scholars is necessary to develop regional paleoclimate grid data and enhance the model's accuracy. Verhagen et al. have identified various issues related to the predictive accuracy. These issues include the relevance of environmental input data, the lack of temporal and/or spatial resolution, the use of spatial statistics, the testing of predictive models, and the need to incorporate social and cultural input data [56]. Taken the environmental input data as an example, the relationship between the spatial resolution of environmental input data and predictive accuracy has yet to be thoroughly examined. Excessively high spatial resolution

can impact overall operational efficiency and require more advanced computer hardware. On the other hand, if the resolution is too low, it may result in the smoothing out of spatial differences in environmental variables. In addition, due to the static nature of the point data, the model cannot take into consideration the small-scale temporal dimension of human behavior and movements [57]. In conclusion, to improve the predictive accuracy of the model, it is essential to enhance the research on the input data.

Conclusions

This study presents a statistical and predictive approach based on GIS for analyzing the relationship between humans and their surroundings. Using the BLR model, this research reveals that Temperature was the most significant factor in determining the presence or absence of sites. Distance to water, Precipitation, and Elevation were also found to be influential factors. Other variables were excluded from the analysis as they had limited ability to distinguish between sites and non-sites.

The Longshan period was characterized by high social complexity. In order to understand the hierarchical structure of this period, this paper analyzed the differences in environmental variables between walled cities and ordinary settlements using GIS and statistical tools. The analysis reveals that middle/small walled city sites were preferred over locations with more favorable terrain and land resources. This suggests that once the minimum threshold for site presence was met, climate conditions were no longer the primary factor for city site selection. However, unlike smaller walled cities, the Shimao site, which was one of the largest prehistoric city centers in north China during the Longshan period, did not have a larger amount of cropland available in its vicinity. This suggests that Shimao may have served as a center for military defense, rituals, or commerce, rather than as an agricultural production center.

Unlike the urbanization process supported by agriculture in the plain region, urbanization in border areas has emerged due to population pressure and the introduction of animal husbandry. Diversified subsistence strategies compensated for the shortage of agricultural production and ensured a stable economic income. This production mode has laid the foundation for the development of grassland cities throughout history.

In conclusion, this study conducted a quantitative analysis of the environmental context for site presence and city development. Additionally, the analysis of environmental variables offered valuable insights into the hierarchical structure of the social system. In the future, it is

crucial to conduct more comprehensive investigations on the data of the input model. This will not only enhance the predictive accuracy but also deepen our overall understanding of the relationship between humans and nature.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40494-024-01194-5>.

Additional file 1: Prediction surface to show the probability and prediction value.

Additional file 2: Table S1. Collinear test of the predictive variables.

Additional file 3: Table S2. Comparison of paleoclimate and current data.

Additional file 4: Quantitative paleoclimatology reconstruction results.

Additional file 5: The model prediction surface generation and model validation results.

Acknowledgements

Thanks Professor Sun Zhouyong, Shao Jing and Dr. Di Nan for their help in the field survey. We also thank the anonymous reviewers and the editor for their constructive comments and suggestions.

Author contributions

CJX conceptualization, methodology, formal analysis, data interpretation, prepared figures and tables, writing—original draft, writing—review and; editing.

Funding

This research was supported by the Major Project of the Key Research Base of Humanities and Social Sciences of the Ministry of Education, China (22JJJD770053), and the National Natural Science Foundation of China (41571190).

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The author reports there are no competing interests to declare.

Received: 7 November 2023 Accepted: 23 February 2024

Published online: 05 March 2024

References

- Liu L, Chen X. The archaeology of China: from the late paleolithic to the early Bronze Age. Cambridge: Cambridge University Press; 2012.
- Demattè P. Longshan-Era urbanism: the role of cities in predynastic China. *Asian Perspect*. 1999;38:119–53.
- Wang W, Guo X. Research on the settlement pattern and society during Longshan and Xia periods. *Archaeol Cult Relics*. 2016;216:52–9.
- Hu K, Mo D, Mao L, Li S, Wang H. Spatial analysis and landscape significance of mankind settlement sites in Wuding River Basin in Mid-Holocene. *Scientia Geographica Sinica*. 2011;31:415–20.
- Yubero-Gómez M, Rubio-Campillo X, López-Cachero FJ, Esteve-Gràcia X. Mapping changes in late prehistoric landscapes: a case study in the Northeastern Iberian Peninsula. *J Anthropol Archaeol*. 2015;40:123–34.
- Spencer C, Bevan A. Settlement location models, archaeological survey data and social change in Bronze Age Crete. *J Anthropol Archaeol*. 2018;52:71–86.
- Nicu IC, Romanescu G. Effect of natural risk factors upon the evolution of Chalcolithic human settlements in Northeastern Romania (Valea Oii watershed). From ancient times dynamics to present days degradation. *Z Geomorphol*. 2016;60:1–9.
- Raj U, Sinha NK, Tewari R. National-scale inventory and management of heritage sites and monuments: advantages and challenges of using geospatial technology. *Curr Sci India*. 2017;113:1934–47.
- Dal Zovo C, Parceró-Oubiña C, César González-García A, Güimil-Fariña A. Mapping human mobility and analyzing spatial memory: palimpsest landscapes of movement in the Gobi-Altai Mountains. *Mongolia J Anthropol Archaeol*. 2023;71: 101516.
- Jochim MA. Dots on the map: issues in the archaeological analysis of site locations. *J Archaeol Method Th*. 2023;30:876–94.
- Finke PA, Meylemans E, Van De Wauw J. Mapping the possible occurrence of archaeological sites by Bayesian inference. *J Archaeol Sci*. 2008;35:2786–96.
- Perakis KG, Moysiadias AK. Geospatial predictive modelling of the Neolithic archaeological sites of Magnesia in Greece. *Int J Digit Earth*. 2011;4:421–33.
- Oonk S, Spijker J. A supervised machine-learning approach towards geochemical predictive modelling in archaeology. *J Archaeol Sci*. 2015;59:80–8.
- Nicu IC, Miha-Pintilie A, Williamson J. GIS-based and statistical approaches in archaeological predictive modelling (NE Romania). *Sustainability*. 2019;11:5969.
- Yaworsky PM, Vernon KB, Spangler JD, Brewer SC, Codding BF. Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. *PLoS ONE*. 2020;15: e0239424.
- Christianne LF, Konstantinos PT. Detecting variability: a study on the application of bayesian multilevel modelling to archaeological data. Evidence from the Neolithic Adriatic and the Bronze Age Aegean. *J Archaeol Sci*. 2021;128:1–14.
- Di Leo P, Bavusi M, Corrado G, Danese M, Giammatteo T, Gioia D, Schiattarella M. Ancient settlement dynamics and predictive archaeological models for the Metapontum coastal area in Basilicata, southern Italy: from geomorphological survey to spatial analysis. *J Coast Conserv*. 2018;22:865–77.
- Bradley EJ, Smith GM, Nussear KE. Ecological niche modeling and diachronic change in Paleoindian land use in the northwestern Great Basin, USA. *J Archaeol Sci Rep*. 2022;45: 103564.
- Demján P, Dreslerová D, Kolář J, Chuman T, Romportl D, Trnka M, Lieskovský T. Long time-series ecological niche modelling using archaeological settlement data: Tracing the origins of present-day landscape. *Appl Geogr*. 2022;141: 102669.
- Noviello M, Cafarelli B, Calculli C, Sarris A, Mairota P. Investigating the distribution of archaeological sites: Multiparametric vs probability models and potentials for remote sensing data. *Appl Geogr*. 2018;95:34–44.
- Li L, Li Y, Chen X, Sun D. A prediction study on archaeological sites based on geographical variables and logistic regression—a case study of the Neolithic Era and the Bronze Age of Xiangyang. *Sustainability*. 2022;14(23):15675.
- Sharafi S, Fouladvand S, Simpson I, Alvarez JAB. Application of pattern recognition in detection of buried archaeological sites based on analysing environmental variables, Khorramabad Plain. *West Iran J Archaeol Sci-Rep*. 2016;8:206–15.
- Jaroslawn J, Hildebrandt-Radke I. Using multivariate statistics and fuzzy logic system to analyse settlement preferences in lowland areas of the temperate zone: an example from the Polish Lowlands. *J Archaeol Sci*. 2009;36:2096–107.
- Canilao MAP. Weight of evidence predictive modelling and potential locations of ancient gold mining settlements in Benguet in the 16th to 18th centuries. *Philipp J Sci*. 2017;146:187–92.
- Sun Z, Shao J, Shao A, Kuang N, Qu F, Liu X. Shimao site of Shenmu county, Shaanxi Province. *Archaeology*. 2013;1248:15–24.
- Dai L, Balasse M, Yuan J, Zhao C, Hu Y, Vigne J-D. Cattle and sheep raising and millet growing in the Longshan age in central China: stable isotope investigation at the Xinzhai site. *Quatern Int*. 2016;426:145–57.
- Lü Z. Shimao ancient city: early human civilization development and environment choices. *J Hist Geogr*. 2016;31(63–8):139.

28. Cui JX, Sun ZY, Burr GS, Shao J, Chang H. The great cultural divergen- ce and environmental background of Northern Shaanxi and its adjacent regions during the late Neolithic. *Archaeol Res Asia*. 2019;20: 100164.
29. Sun Z, Shao J, Liu L, Cui J, Bonomo MF, Guo Q, Wu X, Wang J. The first Neolithic urban center on China's north Loess Plateau: the rise and fall -of Shimao. *Archaeol Res Asia*. 2018;14:33–45.
30. Jaang L, Sun Z, Shao J, Li M. When peripheries were centres: a preliminary study of the Shimao-centred polity in the loess highland. *China Antiq*. 2018;92:1008–22.
31. Ding Z, Yao S. Assessing the ecological effectiveness of Sloping Land Conversion Programme to identify vegetation restoration types: a case study of Northern Shaanxi Loess Plateau. *China Ecol Indic*. 2022;145: 109671.
32. Kvamme KL. A predictive site location model on the high plains: an ex- ample with an independent test. *Plains Anthropol*. 1992;37:19–40.
33. Vaughn S, Crawford TA. predictive model of archaeological potential: an example from northwestern Belize. *Appl Geogr*. 2009;29:542–55.
34. Xiong LY, Li SJ, Hu GH, Wang K, Chen M, Tang GA. Past rainfall-driven erosion on the Chinese loess plateau inferred from archaeological e-vidence from Wucheng City. *Shanxi Commun Earth Environ*. 2023;4:1–8.
35. Chen W. The erosional rates of the loess region in the Wudinghe Riv-er catchment. *Arid Land Geography*. 1989;12:25–30.
36. Chen W, Xiao A, Braconnot P, Ciais P, Viovy N, Zhang R. Mid-Holocene high-resolution temperature and precipitation gridded reconstructions over China: implications for elevation-dependent temperature changes. *Earth Planet Sci Lett*. 2022;593: 117656.
37. Mihu-Pintilie A, Nicu IC. GIS-based landform classification of Eneolithic archaeological sites in the plateau-plain transition zone (NE Romania): habitation practices vs. flood hazard perception. *Remote Sens*. 2019;11:915.
38. Kong X, Cui J. Variables and validation data analysis to improve the prehistoric cultivated land predictive precision of Yulin, northern Shaanxi, China. *Land*. 2024;13(2):153.
39. Garavaglia SB, Dun AS, Hill BM. A smart guide to dummy variables: four applications and a macro. Paper presented at the Proceedings of the Northeast SAS Users Group Conference. 1998; Pittsburgh, PA, USA.
40. Weiss H. Excavations at Tell Leilan and the origins of north Mesopotami- an cities in the third millennium B.C. *Paléorient*. 1983;9:39–52.
41. Algaze G. The Mesopotamian Advantage: Initial Social Complexity in So- uthwestern Asia. *Curr Anthropol*. 2001;42:199–233.
42. Styring AK, Charles M, Fantone F, Hald MM, McMahon A, Meadow RH, et al. Isotope evidence for agricultural extensification reveals how the world's first cities were fed. *Nature Plants*. 2017;3:17076.
43. Hu S, Zhang P, Yuan M. A study on the faunal remains from the Huoshiliang site in Yulin. Shaanxi Province *Acta Anthropologica Sinica*. 2008;27:232–48.
44. Hu S, Yang T, Yang M, Shao J, Di N. Research on faunal remains from the miaoliang site in jingbian county, Northern Shaanxi on the formation of animal husbandry in china. *Quat Sci*. 2022;42:17–31.
45. Feng X. Settlement, subsistence strategies and environment background of Hetao region from Yangshao to Longshan periods. Master thesis. 2019.
46. Hu S, Yang M, Sun Z, Shao J. Research on faunal remains excavated from 2012 to 2013 of Shimao Site, Shenmu county. Shaanxi province *Archaeology and Cultural Relics*. 2016;216:109–21.
47. Yang M, Hu S, Guo X, Wang W, Yang T. Research on the faunal remain- s from Muzhuzhuliang Site in Shenmu city Shaanxi. *Acta Anthropologica Sin-ica*. 2017;41:394–404.
48. Sheng P, Shang X, Zhou X, Storozum M, Yang L, Guo X, et al. Feeding Shimao: archaeobotanical and isotopic investigation into early urbanism (420–0~3000 BP) on the Northern Loess Plateau. *China Environ Archaeol*. 2021;26:1–15.
49. Cheung C, Zhang H, Hepburn JC, Yang D, Richards MP. Stable isotope and dental caries data reveal abrupt changes in subsistence economy in ancient China in response to global climate change. *PLoS ONE*. 2019;14:e0218943.
50. Tian G, Tang X. A study on the man-environment relation in the Daihai area of inner-Mongolia. *Collect Essays Chin Hist Geogr*. 2001;16:4.
51. Han J. Expansion and influence of Laohushan culture. *Cult Relics Cent China*. 2007;133:20.
52. Bogaard A, Fraser R, Heaton THE, Wallace M, Vaiglova P, Charles M, et al. Crop manuring and intensive land management by Europe's first farmers. *PNAS*. 2013;110:12589–94.
53. Lu Z, Hui R, Lü W, Sun Z, Shao J, Di N, Liu J. Observation and research on textile unearthed from the Shimao Site in Shenmu county, Shaanxi Province. *Archaeology*. 2023;1441:106–20.
54. Kempf M, Günther G. Point pattern and spatial analyses using archaeological and environmental data—a case study from the Neolithic Carpathian Basin. *J Archaeol Sci Rep*. 2023;47: 103747.
55. Dimuccio LA, Ferreira R, Batista A, Gameiro C, Zambaldi M, Cunha L. Predictive spatial analysis for a critical assessment of the preservation potential of Palaeolithic record in the Leiria region (central Portugal). *Quatern Int*. 2023;668:44–62.
56. Verhagen P, Kamermans H, Leusen MV. Chapter 2-the future of archaeological predictive modelling. In: Kamermans H, Leusen MV, Verhagen P, editors. *Archaeological prediction and risk management: alternatives to current practice*. Leiden: Leiden University Press; 2009. p. 19–25.
57. Verhagen JWHP. Chapter 2- spatial analysis in archaeology: moving into new territories. In: Bubenzer O, Siart C, Forbriger M, editors. *Digital geoarchaeology: new techniques for interdisciplinary human- environmental research*. Cham: Springer; 2017. p. 11–25.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in pub- lished maps and institutional affiliations.

Jianxin Cui CJX's speciality is environmental archaeology, with a recent focus on Landscape Archaeology and the application of Remote Sensing (RS) and Geographic Information Systems (GIS) methods in Archaeology. She is particularly interested in studying the Neolithic period of the North Loess Plateau of China, specifically the interactions between politics, culture, and climate. Her research is interdisciplinary, as it involves the use of various data sources and technical methods.