

Predicting the Distribution of Yangshao Culture Sites by Integrating Natural Environment and Settlement Spatial Organization Characteristics

JIATING LI

Zhengzhou University

JIAN LI

lijian5277@163.com

Zhengzhou University

JIN ZHU

Zhengzhou University

HAOLIN LI

Zhengzhou University

HAO CUI

Zhengzhou University

Article

Keywords: Luoyang Basin, spatial autocorrelation, site distribution, Maxent model, Yangshao culture

Posted Date: October 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11313161/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Title: Predicting the Distribution of Yangshao Culture Sites by Integrating Natural Environment and Settlement Spatial Organization Characteristics

Author: Jiating Li¹, Jian Li², Jin Zhu¹, Haolin Li², Hao Cui²

¹School of Archaeology and Cultural Heritage, Zhengzhou University, Zhengzhou 450001, China

²School of Geo-Science and Technology, Zhengzhou University, Zhengzhou 450001, China

***Corresponding author:** Jian Li

Email: lijian5277@163.com

This research received no external funding.

Abstract: The Luoyang Basin is a dense concentration area of Yangshao culture sites in the Central Plains. This study uses collected Yangshao sites in the basin to examine their settlement distribution and its environmental drivers by applying global and local spatial autocorrelation analyses to reveal clustering patterns, and by constructing a Maxent model with factors such as elevation, slope, temperature, and precipitation to identify key controlling factors and suitable ranges. The results show that: (1) Yangshao-period sites in the study area exhibit spatial autocorrelation with pronounced clustering. Stagewise Global Moran's I indicates that the floodplains along the Yi and Luo rivers and the terrace on the northern bank of the Luo River represent persistent high-high hotspots, reflecting a long-term preference for particular landscapes. (2) The average AUC of the Maxent model is 0.914; elevation contributes 68.6%, the minimum temperature of the coldest month 11.9%, and precipitation in the wet season 7.1%; together these three factors account for over 87%. (3) Highly suitable zones concentrate at elevations of 100–300 m on secondary terraces, with the coldest-month minimum temperature $> -7.5^{\circ}\text{C}$, within 500–1500 m of waterways, and on low-slope ($< 3^{\circ}$) flat land. (4) The potentially highly suitable areas show a “multi-core and river-aligned” pattern, mainly located in the Yi–Luo river belt and near the front edge of the loess terrace on the Luo River's northern bank, where several large Yangshao settlements are concentrated. By integrating spatial structure characteristics with environmental factors in the Maxent model, the study reveals their joint driving role in prehistoric settlement siting and provides a reference analytical framework for predicting the potential distribution of prehistoric settlements.

Keyword: Luoyang Basin; spatial autocorrelation; site distribution; Maxent model; Yangshao culture

Introduction: In the production-poor prehistoric periods, people largely relied on locally sourced resources from their surroundings for daily livelihoods. Consequently, settlement site selection was substantially constrained and shaped by the surrounding environment. Analyzing the relationship between settlement distributions and environmental factors can illuminate the human–environment relations of the time and the degree to which people understood and exploited natural resources. The Luoyang Basin is located in western Henan Province, comprising the entirety of Luoyang City and part of Gongyi City, situated in the middle reaches of the Yellow River, surrounded by mountains on all sides with generally higher elevations to the

west and lower elevations to the east. Water resources are provided by the Yi River and the Luo River within the basin. This area is archaeologically rich and has yielded numerous well-documented settlement sites, such as Wangwan, Suyang, Zhuali, and Shuanghuai Tree sites. Based on the traces and phenomena uncovered at these settlements, archaeologists have formed an initial understanding of the evolution of Yangshao culture in the Luoyang Basin and have reached a consensus that the Luoyang Basin represents a core region for the emergence of civilization in the Central Plains^[1]. In recent years, owing to the abundance of prehistoric material in the region, research on prehistoric settlements has expanded beyond merely their spatial positions to also emphasize the relative positioning of sites and the connections between settlement clusters. From 1997 to 2007, the Luoyang Basin's eastern-central area was surveyed for pre-Qin sites as part of the Sino-Australian-American-Italian joint archaeological project, coordinated with the Chinese Academy of Sciences, uncovering 456 pre-Qin sites in the area^[2]. The Luoyang City Institute of Archaeology subsequently conducted regional re-surveys to address gaps in the project. Building on these survey records, many scholars have employed geographic information systems (GIS) to conduct spatial analyses of prehistoric sites in the Zhengzhou–Luoyang region, examining environmental factors^[3], human–environment relations, and related topics. Bi Shuo Ben and colleagues were among the earliest to integrate methods such as spatial clustering, spatial autocorrelation, and geographically weighted regression into studies of prehistoric settlements, establishing a multivariate quantitative framework for analyzing human–environment relationships that combines site locations with topography and hydrology^[4-6].

Maximum Entropy (MaxEnt) models derive potential distribution probabilities from existing data and environmental variables. They have been widely applied in studies of species distributions^[7], ecological assessment^[8-9], and disaster early warning^[10], and can be leveraged in settlement studies to evaluate prehistoric settlement environments and predict potential sites. Internationally, research on MaxEnt and settlement archaeology is relatively mature. For example, one project combined MaxEnt with 172 known open-air Paleolithic sites to predict the spatial distribution of open-air Paleolithic sites in Nagpur and Chandrapur districts, Maharashtra, India, revealing elevation, distance to rivers, and geology as the dominant predictors^[11]. Another study used a model with 11 environmental variables and known site data to predict Iron Age sites in Kurdistan^[12]. In China, domestically, applications of MaxEnt to settlement site distribution or site prediction remain comparatively scarce^[13]. Leveraging the rich archaeological resources of the Luoyang Basin, this study aims to integrate spatial clustering patterns with MaxEnt-based environmental suitability modelling to examine Yangshao-period settlement distribution and the adaptive strategies of Yangshao communities. By combining spatial analysis preferences with predictive siting, the research seeks to provide quantitative evidence for prehistoric settlement archaeology in the Luoyang Basin and offer a new perspective on prehistoric human–environment relations, while also guiding the search and discovery of other prehistoric sites.

Materials and methods

Overview of the Study Area

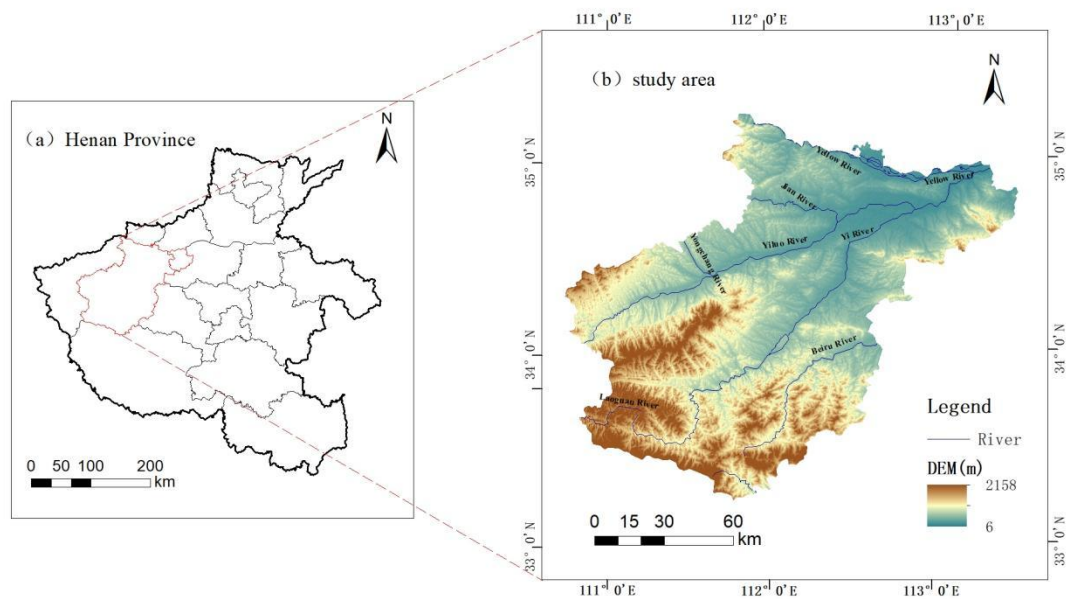


Fig. 1 Location map of the study area

The Luoyang Basin, also known as the Yi River Valley Plain, Yi Luo Plain, or Yi Luo Basin, is located in the western part of Henan Province (Figure 1), in the heart of the Central Plains, China. It lies between 34°30'N–35°52'N and 112°18'E–113°07'E, at the southeastern edge of the Loess Plateau and at a transitional zone between the Second and Third Terraces of Chinese terrain. The basin encompasses the modern Luoyang urban area and parts of Yanshi District, Yi Luo River area, as well as portions of Mengjin District, Xin'an County, Yiyang County, and Gongyi City. The Zhou Mountain ranges Zoushan and Mangshan, connected to the Qinling Mountains to the west, lie at the western and northern margins of the basin, with the southeastern and southern parts hosting the Wan'an Mountain chain as a continuation of the Songshan range. The basin's interior is relatively flat, with dense river networks and abundant water sources; the Luo River and Yi River traverse the basin from southwest to northeast, joining to form the Yi-Luo River in its eastern part and ultimately feeding into the Yellow River^[14]. Surrounded by mountains on three sides and bordered by a river on one side, the Luoyang Basin forms a natural geographic unit. The region lies in a cultural sequence where the Yangshao culture overlies the Peiligang culture and underlies the Longshan culture, reflecting a clearly continuous cultural lineage through crucial stages of social differentiation and civilization origins. Since the 1920s–1930s, archaeologists have conducted extensive work in this area, uncovering important Yangshao-era settlements such as Wangwan^[15], Shuanghuaishu^[16], and Suyang^[17], which provide abundant material for the study of prehistoric settlements in the region.

Data Sources and Processing

Settlement Site Data

The site point dataset used in this study was compiled from multiple sources: the China Cultural Relics Map Atlas – Henan Volume, and the Luoyang Basin eastern–central Pre-Qin site regional survey (1997–2007), complemented by local archaeology gazetteers, excavation reports, and archaeological surveys produced by relevant local authorities. These sources were integrated to construct a prehistoric settlement site database for the Luoyang Basin.

Dating and quality control: The eastern–central Luoyang Basin survey (1997–2007) provides relatively clear dating for recorded sites; dating for sites documented in the Cultural Relics Map Atlas or those that have been excavated is taken from published excavation materials. For sites with limited cultural facies details, chronology is inferred by cross-referencing artifacts and features with nearby sites. During data screening, surveyed sites with small areas, insufficient remains, or poorly described cultural facies were excluded. In the initial coordinate harmonization, site coordinates were converted to decimal degrees and stored as a three-column table: site, longitude, latitude. All data are stored in CSV format to facilitate downstream analysis.

Environmental Data

Prehistoric settlement location is strongly constrained by natural environment. Candidate factors include elevation, slope, temperature, and precipitation, among others. We extracted Holocene-like climate variables (19 bioclimatic variables) from WorldClim (worldclim.org) and obtained study-area DEM data from the Geospatial Data Cloud (GSCloud, gscloud.cn) of the Chinese Academy of Sciences. Terrain derivatives (slope, aspect) were derived from the DEM. All data were originally in raster format.

Data harmonization: All rasters were reprojected to GSC_WGS_1984, then cropped to the study extent using the study-area boundary polygon to ensure consistent pixel size, spatial resolution, and extent. Finally, all factor rasters were exported to ASC format suitable for MaxEnt input via ArcGIS.

Research Methods

We develop a hybrid modeling framework that integrates natural environment variables with settlement spatial organization features to predict site suitability and potential distribution of prehistoric settlements. The overall workflow is illustrated in Figure 2. **Step 1: Quantify settlement clustering.** We first grid the study area at a fixed resolution and compute the number of sites per grid cell. We then apply global and local spatial autocorrelation analyses (e.g., Global Moran's I and Local Indicators of Spatial Association, LISA) to characterize the overall degree of settlement aggregation and to classify local clustering types (e.g., high–high, low–low, hot spots). **Step 2: construct predictor sets.** We compile two classes of predictors. Natural environment predictors include elevation, slope, aspect, temperature, and precipitation (derived from WorldClim 19 bioclimatic variables and the DEM-based terrain derivatives). The second class comprises settlement-space organization predictors derived from the local clustering results (e.g., cluster membership, local density, and

spatial interaction indicators reflecting neighborhood effects). Step 3: modeling and prediction. The two predictor sets are entered into a maximum entropy (MaxEnt) model to estimate site suitability across the study area. The model outputs a continuous suitability surface, from which we identify and map regions with high potential for undiscovered prehistoric settlements.

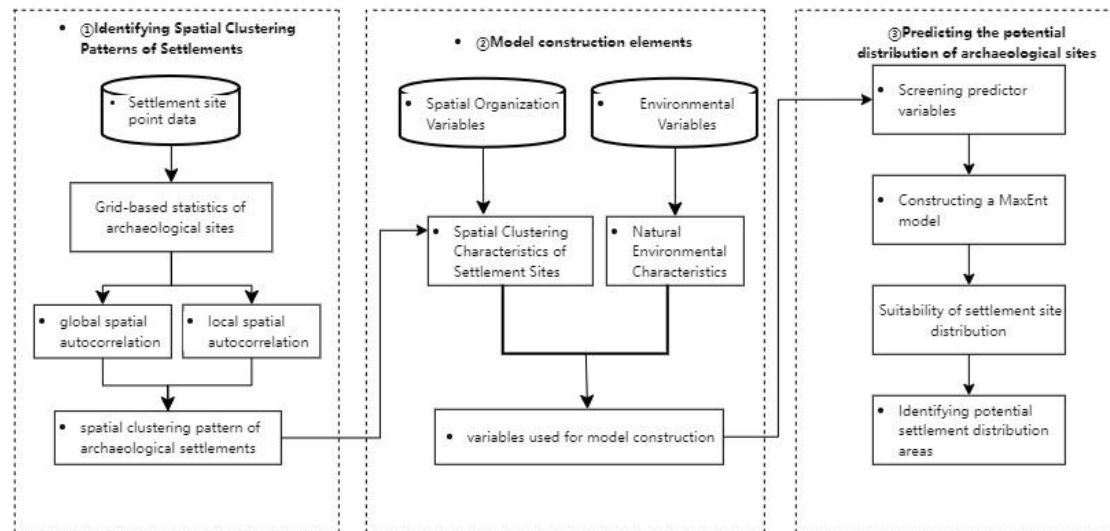


Figure 2. Framework for predicting the potential distribution of Yangshao cultural sites.

Identification of Settlement Spatial Patterns Based on Spatial Autocorrelation Analysis

Spatial autocorrelation analysis is used to test whether geographic attributes exhibit spatial dependence^[18]. In the study area, we construct a 5 km × 5 km regular grid^[19], count the number of site records within each grid cell, and use these grid counts as the attribute values. Global Moran's I and Local Moran's I are then computed to assess the overall spatial pattern and local spatial associations of settlement distribution, respectively.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

The Moran's I statistic uses n (number of grid cells), x_i and x_j (site counts in cells i and j), $\bar{x}$ (mean site count), w_{ij} (spatial weight between cells i and j), and S_0 (sum of all weights). Moran's $I > 0$ indicates spatial clustering of site counts, Moran's $I < 0$ indicates dispersion, and Moran's $I \approx 0$ indicates a random distribution.

$$I_i = \frac{x_i - \bar{x}}{S^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x})$$

The local Moran's I is denoted as I_i , where I_i represents the local spatial autocorrelation statistic for grid cell i , and S^2 is the variance of the site counts across grid cells; other notation follows the same definitions as above. Based on the configuration of the target grid cell and its neighboring cells, local spatial associations are categorized into four types: high–high, low–low, high–low, and low–high. In

particular, high–high clusters indicate grid cells with high site counts that are surrounded by neighboring cells with similarly high counts, which helps identify core areas of settlement concentration.

Construction of Settlement Spatial Organization Features

Site distribution is determined not only by natural environmental conditions such as terrain, hydrology, and climate, but also by spatial proximity and clustering relationships among settlements. This study constructs site-prediction features from two aspects: natural environment features and settlement spatial organization features. Natural environment features include variables such as elevation, slope, distance to waterways, and temperature and precipitation, which are used to describe the natural geographic conditions relevant to prehistoric settlement production, daily life, and resource acquisition.

Settlement spatial organization features include settlement core density, spatial proxy variables constructed from Voronoi (Thiessen) polygons, and spatial clustering types. Settlement core density describes the degree of concentration of sites within a given neighborhood, and its computation is defined accordingly. Voronoi polygons are used as spatial proxy variables to characterize the influence domains of settlements, for example by capturing the area or boundary relationships of the polygons, which can supplement the explanatory power of kernel-density descriptions.

$$D(s) = \frac{1}{h^2} \sum_{i=1}^m K\left(\frac{d_{is}}{h}\right)$$

$D(s)$ is the settlement kernel density at location s ; d_{is} is the distance between site i and location s ; h is the bandwidth; K is the kernel function. A larger density value indicates that the distribution of sites within the neighborhood around s is more concentrated. When extracting kernel-density features from known site locations, the contribution of the site itself to the density at its own location is not counted.

Thiessen (Voronoi) polygons derived from settlement points represent the spatial partitioning pattern and balancing relationships among settlements. Using each site as a center, the study area is partitioned by planar Euclidean distance. Distances from any location within a polygon to the center site, in relation to distances to other sites, are used to infer the theoretical hinterland and sphere of influence of individual settlements and to assess spatial competition and adjacency among settlements. The high–high (hotspot) spatial type is constructed from local spatial autocorrelation results; grid cells showing a significant high–high pattern are treated as core areas of settlement clustering, thereby addressing the spatial clustering type. All predictive features are computed in the same coordinate system, spatial extent, spatial resolution, and raster grid structure.

Potential distribution of archaeological sites predicted by a MaxEnt model

The Maximum Entropy Model (MaxEnt), based on the principle of maximum entropy, uses known archaeological site points and their corresponding predictive features to estimate the suitability of site distribution at different locations in the study area. Its basic formula is:

$$P_{(x)} = \frac{1}{Z} \exp \left[\sum_{k=1}^q \lambda_k \int k(x) \right]$$

In the formula, $P(x)$ is the suitability of the archaeological site distribution at spatial position x ; $f_k(x)$ is the value of the k -th predictive feature at location x ; λ_k is the corresponding parameter; Z is the normalization constant. This paper inputs both natural environmental features and settlement spatial organization features into the MaxEnt model, and its expression form is:

$$P_{(x)} = F[E(x), S(x)]$$

In the formula, $E(x)$ represents the natural environmental features at location x ; $S(x)$ represents the settlement spatial organization features; F denotes the mapping from the predictive features learned by the MaxEnt model to the suitability of site distribution. The model outputs the suitability of site distribution in logistic form, with values between 0 and 1. The larger the value, the more the natural environment and spatial organization features of that location resemble known sites, indicating higher potential distribution suitability. Based on the model output, classify the suitability levels of site distribution and identify potential site distribution areas.

Area Under the Receiver Operating Characteristic Curve (AUC) is used to evaluate the model's predictive performance. AUC ranges from 0 to 1; higher values indicate stronger discrimination between site samples and background locations. An AUC close to 0.5 suggests discrimination near random chance.

Results

Spatial clustering characteristics of the site distribution

Using ArcMap 10.8 for spatial analysis, based on previous results of spatial studies in the Zheng-Luo region, during the construction of the fishnet grid a cell size of $5 \text{ km} \times 5 \text{ km}$ was chosen. The sites of the Yangshao culture were separately analyzed for the early, middle, and late phases using global spatial autocorrelation analysis. The results are shown in Table 1.

Table 1. Results of global spatial autocorrelation analyses for three periods of the Yangshao Culture in the Luoyang Basin.

Period	Early	Middle	Late
Moran's I	-0.001496	0.376965	0.551528
Z	-0.399319	14.716651	21.375536
P	0.689658	≈ 0	≈ 0

According to the results of the global spatial autocorrelation analysis in Table 1, the global Moran's I index for the Yangshao Early period is 0.001496, with a Z-score of 0.399319 and $P = 0.689658$, which does not pass the 0.05 significance level test. This indicates that the settlements in the Yangshao Early -Luoyang Plain show no significant spatial clustering or dispersion and exhibit an overall random distribution pattern; during this period, settlements had not formed a regional concentrated distribution. The Moran's I index for the Yangshao Middle period rises to 0.376965, with $Z = 14.716651$ and P value approaching 0, indicating significance at

conventional levels and a clearly clustered settlement pattern. In the Yangshao Late period, the Moran's I index further increases to 0.551528, with $Z = 21.375536$ and P value approaching 0, indicating a higher degree of clustering than in the Middle period.

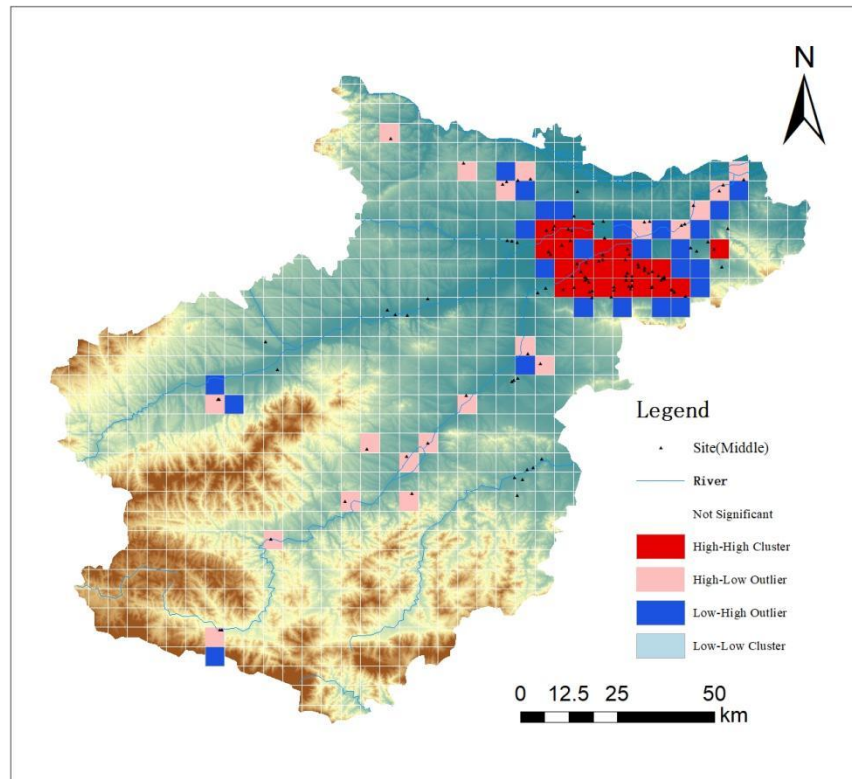
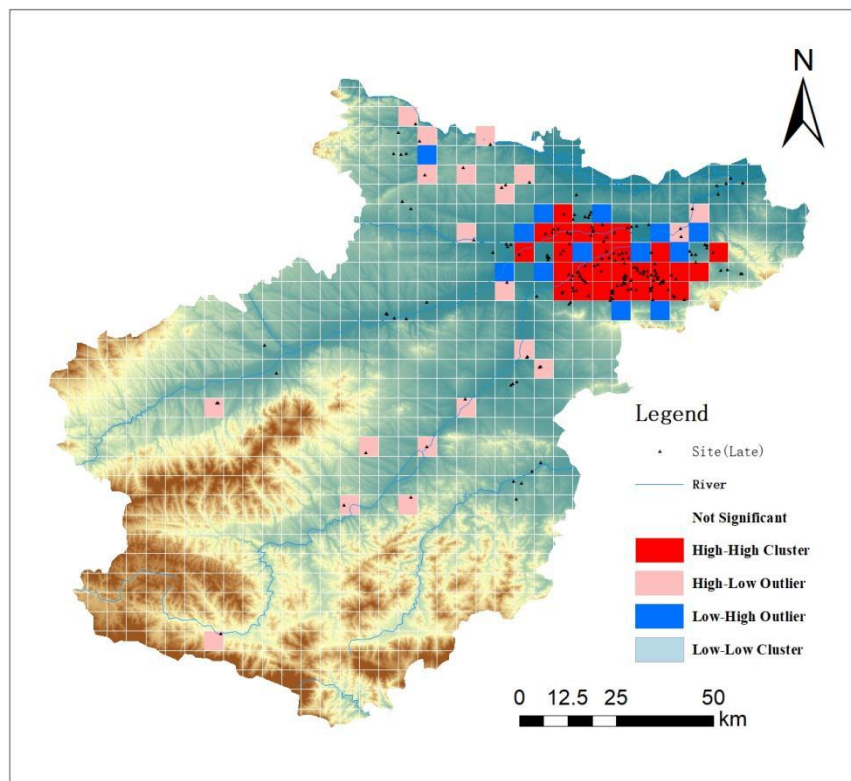


Figure 3(a): Local spatial autocorrelation analysis results for the Middle Yangshao period.



• Figure 3(b): Local spatial autocorrelation analysis results for the Late Yangshao period.

Based on the results of the global autocorrelation analysis, Local Indicators of Spatial Association (LISA) were used to further analyze the clustering types of settlements for the Yangshao mid- and late periods. The results are shown in Figure 3(a). The Yangshao Middle period exhibits a significant high–high (HH) and low–low (LL) local autocorrelation pattern, mainly distributed in the central–western part of the Luoyang Basin. The riverine deltas near the confluence of the Yi and Luo rivers constitute the principal clustering area, forming continuous clusters where high-density sites are surrounded by neighborhoods of similarly high density. Fewer and more scattered settlement points are distributed in the eastern part of the basin. Figure 3(b) shows that the Yangshao Late period still exhibits a significant HH and LL local autocorrelation pattern. The high–high clustering areas largely align with those of the Middle period, occurring mainly near the Yi and Luo rivers. However, compared with the Middle period, although many new settlement points appear in the Late period and the overall area of distribution expands, the high–high clustering appears to “tighten” at a regional scale, with continuous, connected distributions in the river valleys.

Integrating the above results, in terms of temporal evolution, the spatial clustering of Yangshao settlements in the Luoyang Basin shows a trend from an initial random distribution to increasingly pronounced clustering in the Middle and Late periods. After the Middle period, settlements began to form groups, and the degree of clustering intensified further in the Late period. The high–high clusters coincide closely with rivers such as the Yi and Luo, indicating that favorable water and soil conditions drove the concentrated settlement pattern; surrounding mountains largely exhibit low–low clustering, constrained by topography, making large-scale settlement development difficult. This suggests that, within the study time frame, human site selection was not random but constrained by geomorphology, water resources, and other environmental factors.

Natural Environmental Characteristics of Archaeological Site Distribution

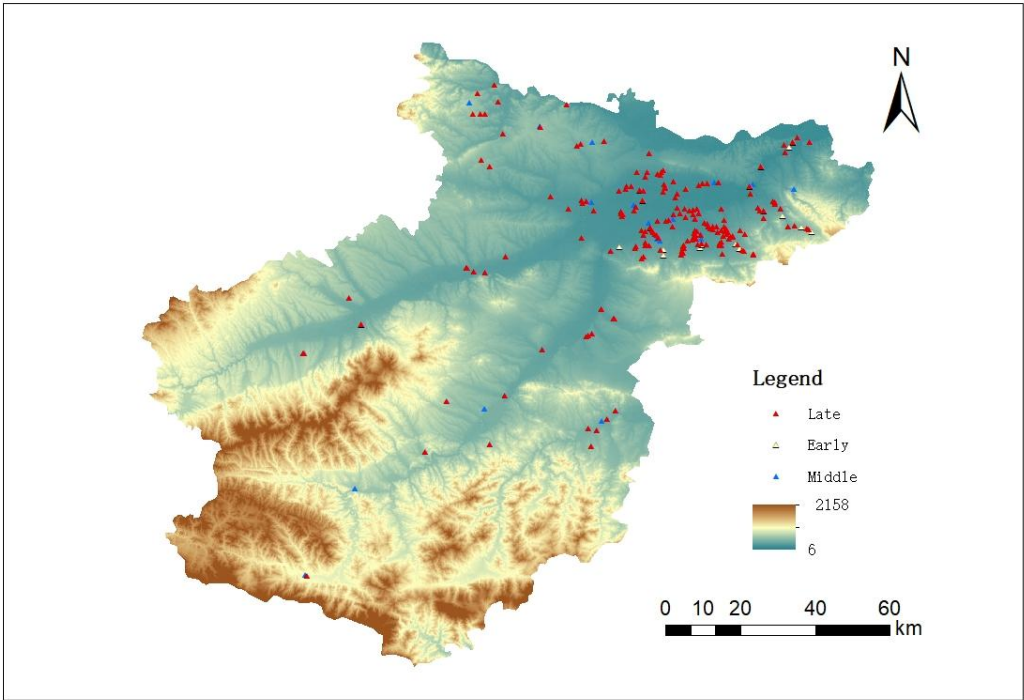


Fig4. Schematic Map of the Elevation Distribution of Sites of the Three Phases of Yangshao Culture

The distribution of Yangshao culture sites by period is presented in Figure 4. Overlay the Yangshao culture settlements of the Luoyang Basin for the three phases with digital elevation model data , and use ArcGIS software to perform extraction and analysis. This yields the elevations (Table 2), slopes (Table 3), and distances to water (Table 4) for the early, middle, and late Yangshao settlements.

Table 2 Altitude distribution statistics of Yangshao sites in different periods

Elevation (m)	Early	percentage (%)	Middle	percentage (%)	Late	percentage (%)
100~200	9	45%	60	51. 28%	113	59. 8%
200~300	9	45%	33	28. 2%	48	25. 4%
300~400	1	5%	13	11. 1%	16	8. 46%
400 以上	1	5%	11	9. 4%	12	6. 34%

The statistical results of site elevations are shown in Table 2. Overall, the elevation distribution of sites in the study area exhibits a clear “lowland clustering” feature, which becomes progressively stronger over time. Sites in all periods are predominantly distributed at low elevations, and the clustering tendency in low-elevation intervals becomes more pronounced as time progresses.

Table 3 Slope distribution statistics of Sites of the Three Phases of Yangshao Culture

slope (°)	小于3° (个)	percentage (%)	3-6° (个)	percentage (%)	6-10° (个)	percentage (%)	10° 以上 (个)	percentage (%)
Early	5	25%	8	40%	6	30%	1	5%
Middle	46	39.3%	51	43.6%	14	12%	6	5.1%
Late	71	37.6%	71	37.6%	38	20.1%	9	4.8%

Overlay the three phases of Yangshao settlement sites in the Luoyang Basin with slope maps, and classify slope suitability according to previous studies on livable slopes ^[20]: slopes below 3° fall within the range suitable for human habitation, 3°–6° are relatively suitable for human habitation, and 6°–10° are less suitable for human habitation. The slope distribution of the three Yangshao phases in the study area is shown in Table 3.

Table 4. Statistics of distribution and distance to water in different periods of the Yangshao Culture

Distance (m)	Early	percentage (%)	Middle	percentage (%)	Late	percentage (%)
0~500	15	75%	66	56.4%	107	56.6%
500~1000	3	15%	22	18.8%	36	19%
1000~1500	1	5%	11	9.4%	24	12.7%
1500~2000	0		12	10.26%	12	6.3%
2000~2500	0		4	3.42%	2	1.1%
2500~3000	0		1	0.85%	1	0.53%
3000~3500	0		1	0.85%	1	0.53%
3500 以上	0				6	0.32%

Water resources from hydrological systems supply water for both production and domestic use, representing a resource with a high degree of dependence. The horizontal distances from the Early, Middle, and Late Yangshao settlement sites to the nearest watercourses were computed using the Near tool in the Analysis Tools > Proximity toolbox of ArcGIS, and the results are presented in Table 4.

From the statistics above, the artifact sites of the Yangshao culture show a clear lowland bias in their elevation distribution. From the Early to the Late period, the proportion of sites located at an elevation of 100–200 m rises from 45% to 59.8%. In contrast, the middle period shows a higher-elevation interval accounting for 10% in the Early period, rising to 20.5% in the Middle period, and then declining in the Late period, with the overall proportions remaining comparatively low. In terms of slope, site distributions for all periods are mainly concentrated in the 3°–6° and <3° slope ranges, indicating a pronounced preference for flat terrains; settlements are predominantly located in flatter areas. This distribution pattern provides a basis for subsequent MaxEnt model variable-contribution analyses. Regarding horizontal

proximity to water, sites across all periods are highly concentrated within 1 km of rivers; within the 0–1 km range, more than 75% of all sites are clustered, with the 0–500 m interval representing the absolute core distribution zone, corroborating the special status of the confluence of the Yi River and Luo River as a core settlement area.

Prediction Using the MaxEnt Model

ArcGIS 10.8 was used to perform correlation analysis between 19 climate-related variables and the terrain factors described above. When the absolute value of the correlation coefficient between environmental variables exceeded 0.85, only the variable with the higher contribution to the MaxEnt model was retained in order to reduce data redundancy and minimize the impact of environmental variable collinearity on the MaxEnt model. Variables with contributions less than 1% were then removed. Ultimately, seven environmental variables were selected for constructing the MaxEnt model (Table 5), namely: isothermality, minimum temperature of the coldest month, precipitation seasonality, precipitation of the wettest quarter, elevation, slope, and distance to water.

Table 5. Information on Environmental Variables Used in MaxEnt Model Construction

Data types	Environmental factors	Percent contribution	Permutation importance
climatic variables	bio1		
	bio2		
	*bio3	3.6	1.6
	bio4		
	bio5		
	*bio6	11.9	19
	bio7		
	bio8		
	bio9		
	bio10		
	bio11		
	bio12		
	bio13		
	bio14		
	bio15	3	2.6
	bio16	7.1	5.6
	bio17		
	bio18		
	bio19		

topographic variables	*elev	68.6	57.8
	*slope	3.3	10.7
	dist	2.4	2.8

*Environmental variables for building the MaxEnt model.

By running the MaxEnt model with 10 repeated trials, model accuracy was evaluated using the AUC value from the ROC curve, where higher AUC within the 0–1 range indicates greater accuracy. The results indicate that the average AUC for predicting the potential geographic distribution of Yangshao-period sites in the Luoyang Basin using MaxEnt is 0.914, which, according to standard evaluation criteria, corresponds to an "extremely accurate" prediction. This suggests that the MaxEnt-based predictions of the Luoyang Basin Yangshao-period site distribution are credible.

Descriptive statistics indicating a lowland bias are quantitatively validated by the MaxEnt model. The model results show that among all environmental factors, elevation contributes the most, followed by minimum temperature of the coldest month and the mean temperature of the wet season, Figure 5.

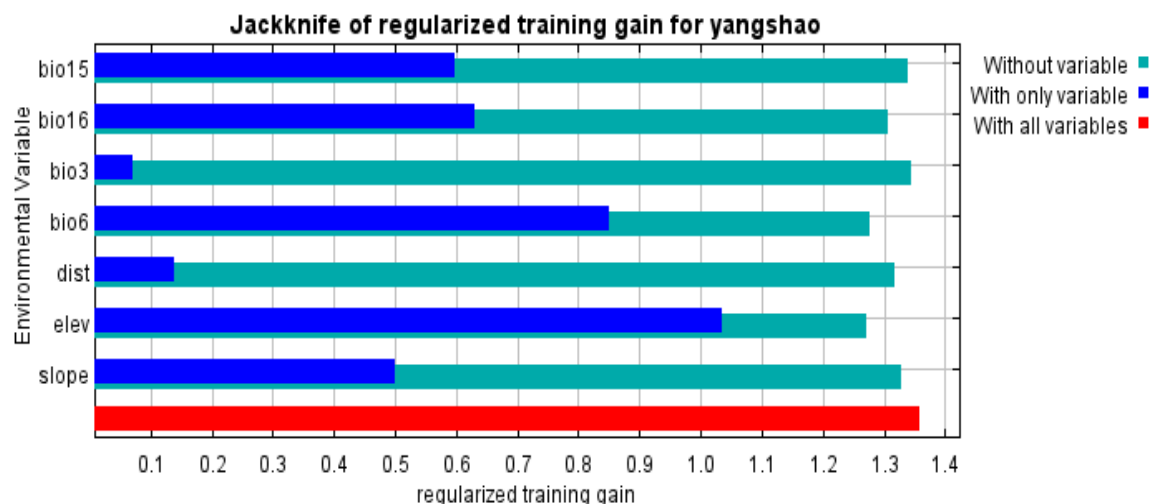


Fig5. Jackknife Plot of Regularized Training Gain for the MaxEnt Model

Using the seven environmental factors obtained from the above screening, the response curves were used to identify the suitable environmental ranges for Yangshao settlement site selection. Based on the threshold of minimum training presence (0.5) to distinguish suitable versus unsuitable, the corresponding ranges of environmental factors are relatively favorable for human production and livelihood. Specifically: the suitable altitude range is approximately 100–300 m; bio3 suitability is mainly concentrated within 4.5–27.5; bio6 suitable range is about -8.5–6.5 °C; bio15 suitable range is 76–88; bio16 suitable range is 38–60 mm, with a water-distance suitability range of 50–1,200 m; the slope suitability range is 0–8°, with the optimal range at 2–4°. The response curves for the environmental factors are shown in Figure 6.

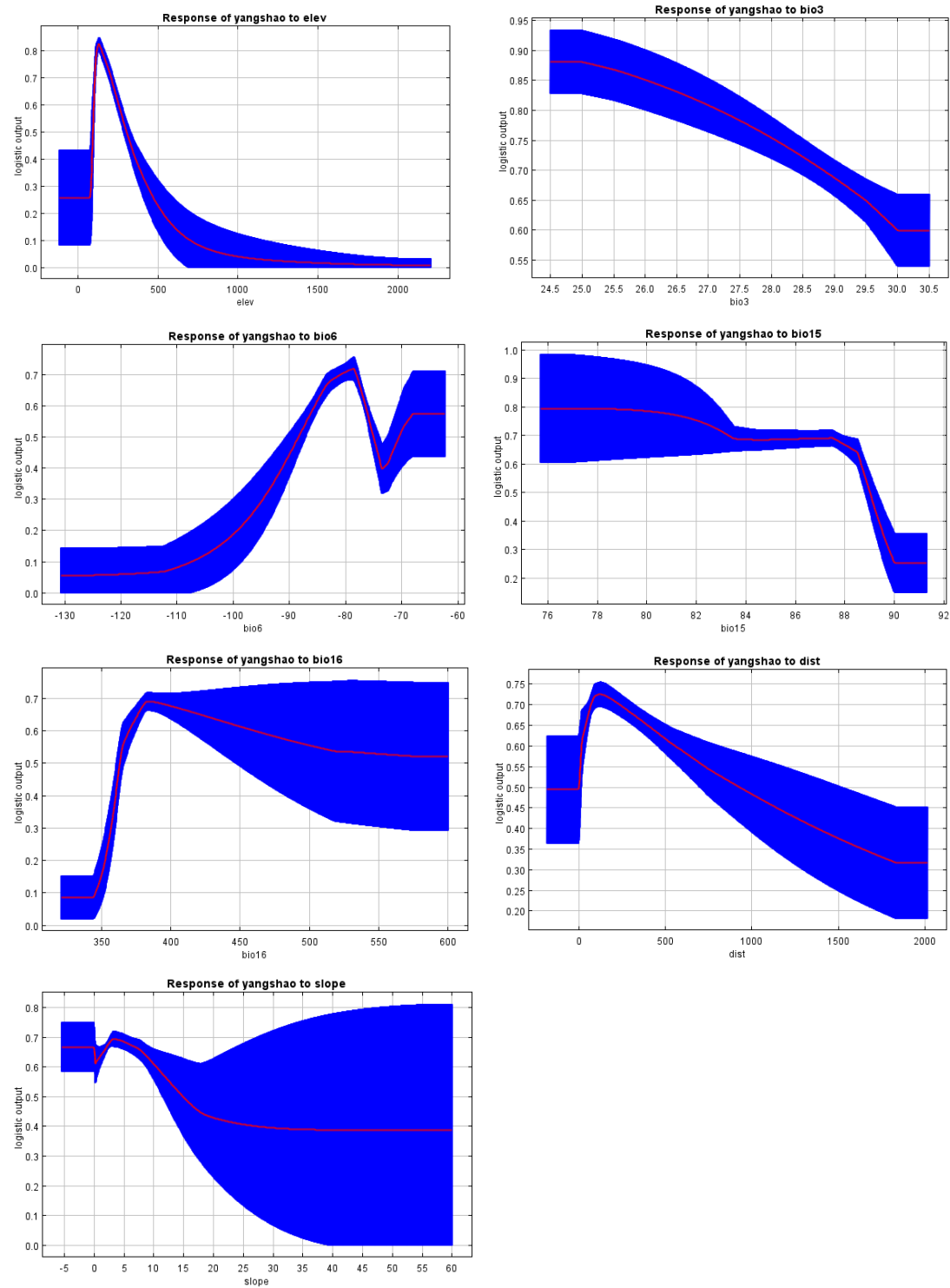


Fig6. Response Curves of Environmental Variables

Taken together, the environment exerted an important influence on the choice of settlement locations by early peoples, with elevation playing a decisive role. The Neolithic settlements exhibit a high degree of spatial continuity and clustering, concentrated in a loess-covered area at elevations of about 100–200 meters near water sources. The terrain there is flat and expansive, and the soil is loose. In light of the development trajectory of the Yangshao culture and its principal production activities,

the prehistoric people living near this area primarily needed to meet water requirements for daily drinking, cooking, watering livestock, and for pottery production and construction. The Yangshao culture relied primarily on millet and glutinous millet for agriculture, which favors loose, fertile, well-drained soils for crop cultivation. Alluvial river valleys and loess plateaus provide relatively superior water and soil resources, and the soft soils facilitate early farming tools such as bone and stone hoes^[21-22]. Furthermore, high-suitability zones display a blocky distribution at the basin's outlet and at the confluence of small tributaries, indicating that ease of water access remained a key consideration in site selection. Although the dependence on environmental factors varied slightly across periods, people tended to move toward higher elevations or locations farther from water, the overall pattern not undergoing drastic change.

Incorporating Settlement Spatial Organization Features into Prediction

To investigate the influence of settlement spatial organization features on site distribution prediction, this study constructed a baseline model containing only natural environmental variables (Model A) and an enhanced model that includes spatial organization variables (Model B). Evaluation results based on 10-fold cross-validation are shown in Table 6.

Table 6. Comparison of the Models' Prediction Results.

evaluation metrics	Model A	Model B
AUC(training)	0.9136	0.9502
AUC (test)	0.8966	0.9461
TSS	0.628±0.045	0.749±0.056
Training gain	1.3579	1.8829
Test gain	1.3329	1.9392

In terms of model performance, Model A's training AUC and test AUC are 0.9136 and 0.8966, respectively, indicating good predictive capability. By contrast, Model B's training AUC and test AUC rise to 0.9502 and 0.9461, with a Δ AUC of 0.0495. Additionally, Model B exhibits a lower AUC standard deviation than Model A, suggesting more stable cross-validation performance and stronger generalization. Regarding gain metrics, Model B's test gain is also higher than that of Model A. Taken together, introducing spatial-organization variables yields higher prediction accuracy and robustness for locating site distributions.

Table 7. Information on model factors incorporating settlement spatial organization characteristics.

Data types	variables	Percent contribution	Permutation importance
Environmental variables	bio3	0.8	0.3
	bio6	1.5	0.6

	biol5	1.6	3.3
	biol6	1.2	4.6
	elev	1.8	2.6
	slope	1.7	3.1
	dist	2.4	2.4
spatial			
organization	Kernel	75.4	79
variables	Thiessen	13.2	4.2

In Model A, natural environmental variables dominate decisively. The contribution rate of elevation reaches 68.64%, with permutation importance at 57.84%, making it the core driving factor. The second-ranking contributors are the minimum temperature of the coldest month (11.92%) and precipitation in the wettest season (7.10%). In Model B, however, with the inclusion of spatial-organization variables, the pattern of variable contributions changes substantially. The contribution rate of spatial-organization kernel-density variables together totals 75.4%, and their permutation importance reaches 79%, making them the primary information source for the model. Meanwhile, the previously dominant contribution of elevation drops to 2.4%, and its permutation importance falls to 2.4%. (Table 7) There appears to be strong collinearity between kernel-density (spatial-organization) variables and natural-environment variables like elevation. In the Luoyang Basin, areas with highly clustered settlements often co-locate with low-elevation river terrace surfaces; thus, spatial-organization variables statistically absorb the explanatory power originally attributed to elevation and other environmental factors.

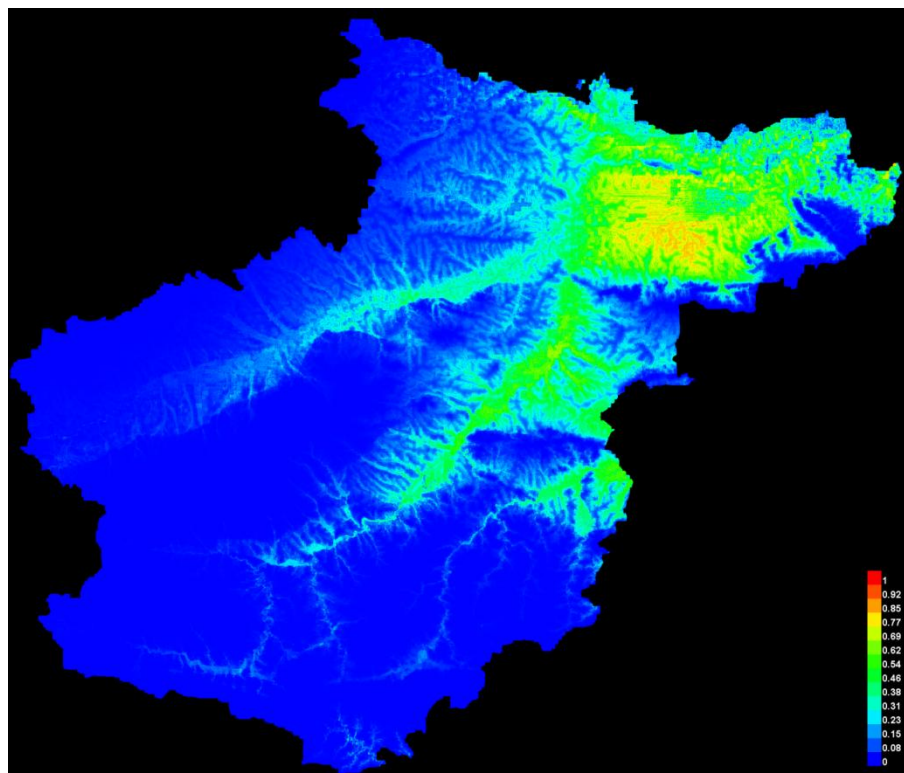


Figure 8(a). Predicted site suitability map for archaeological sites using MaxEnt Model A

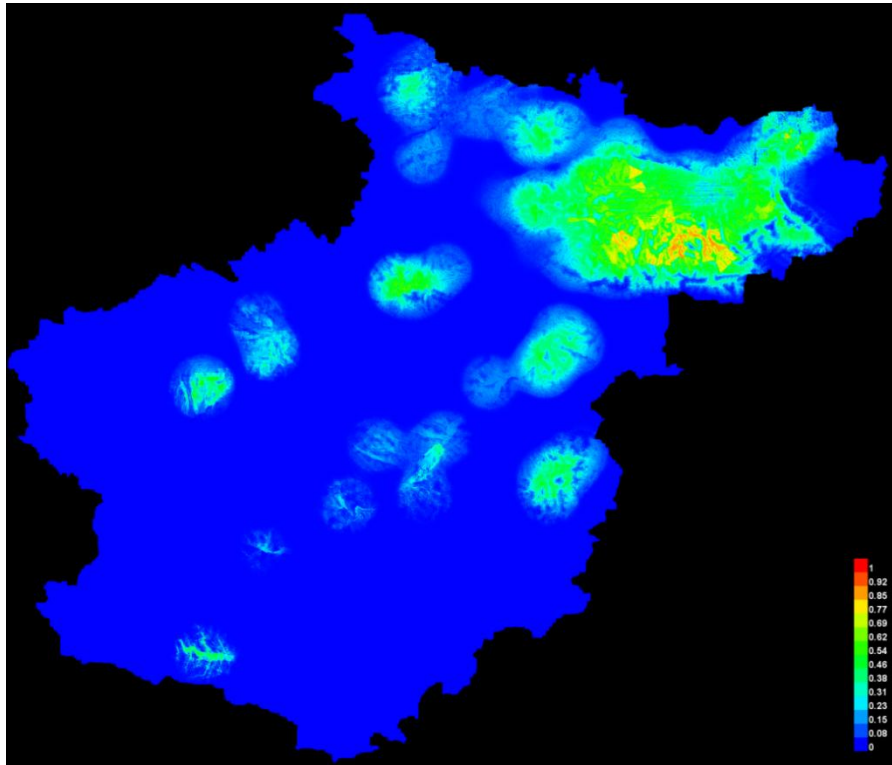


Figure 8(b). Predicted site suitability map for archaeological sites using MaxEnt Model B

From the spatial prediction distribution maps, Model A's highly suitable zones are mainly banded along the low-elevation terraces of the Yi-Luo River and its tributaries, showing clear influence from terrain and hydrological conditions (Figure 8(a)). In contrast, Model B's predicted high-suitability areas exhibit a stronger centralized clustering pattern: the high-probability zones are not only confined to naturally suitable areas but also markedly converge toward and contract around known central settlements and densely populated settlement areas (Figure 8(b)). This indicates that, when delineating potential site distributions, Model B can consider natural environmental suitability while also internalizing the ancient reliance on existing social networks.

The preceding spatial-feature analysis indicates that the distribution of Yangshao culture settlements in the study area generally evolved from an early random pattern toward increasingly pronounced centralization in the middle and late periods. To further quantify how this spatial clustering characteristic affects actual site-location decisions, this study conducts a staged comparison of predictive models for the Yangshao culture across early, middle, and late periods, as shown in Table 8.

Table 8: Comparison of Optimized Model Prediction Results for Different Periods of Yangshao Culture

	Early	Middle	Late
AUC of the environmental-factor	0.967	0.920	0.912

model			
AUC for the model with spatial covariates	0.998	0.959	0.955
Δ AUC	0.031	0.039	0.043

From the model accuracy comparison, the baseline model using only natural environmental variables (the environmental-factor model) demonstrates good predictive ability across all three periods, with AUC values of 0.967, 0.920, and 0.912, respectively. However, as time progresses, the explanatory power of the environmental-factor model shows a slight downward trend, suggesting that relying solely on natural environmental factors is insufficient to fully explain the more complex site-selection logic of the middle-to-late Yangshao settlements. After incorporating spatial-organization variables, the AUC values for the three periods rise to 0.998, 0.959, and 0.955, respectively. Notably, the improvement in AUC (Δ AUC) exhibits a clear stage-wise increase: smallest in the early period, larger in the middle period, and largest in the late period. This pattern aligns well with the previous spatial-feature analysis showing progressively stronger clustering.

Of particular interest is the mid-to-late Yangshao periods, where, after integrating spatial-organization variables, the mid-period kernel-density contribution is 83.1 and the Thiessen-polygon contribution is 5; for the late period, kernel-density contribution is 78.1 and Thiessen-polygon contribution is 12.7. This result indicates that during the middle period, people tended to settle within existing settlement clusters, with spatial clustering being pronounced. The increase in the Thiessen-polygon contribution suggests that by the late Yangshao period, the spatial boundaries and spheres of influence between settlements became more delineated. Taken together, the Yangshao middle period shows clustering with relatively weak inter-settlement territorial competition, while the late period exhibits further increased clustering, with more settlements and still high clustering intensity. Moreover, the delineation of sites' territories and spatial competitive relationships becomes more evident, which is consistent with the overall development trajectory of the Yangshao culture.

Discussion

Mechanisms Linking Environmental Factors to Settlement Location

The above analytical conclusions indicate that Yangshao-period settlement location did not adhere to a single fixed environmental preference, and the mechanisms by which environmental factors influence settlement choice exhibit clear stage-dependent differences. The core lies in how climate-driven changes in the surrounding environmental constraints alter the relationship between habitable space and suitability.

In the early Yangshao period, the constraints imposed by environmental factors on settlement choice were relatively weak. At that time, the number of sites was small, and habitable spaces such as low-elevation valleys and gentle terraces were readily available. Variables such as elevation, slope, and distance to rivers showed little variation, and the habitable area was broad, indicating that early populations could select settlement points within a relatively large area. Low-elevation flat zones could already meet basic water supply and agricultural needs, yielding substantial flexibility

in environmental selection, with limited filtering effects from geographic and climatic factors.

During the Holocene warm-humid optimum, the overall warming and moisture increase improved hydrothermal conditions in high-elevation areas, enhanced valley terrace soils' fertility and moisture retention, and expanded habitable and cultivable space, thereby weakening environmental constraint effects. In terms of spatial distribution, the share of sites in low-elevation core zones continued to rise, while new sites appeared at higher elevations. Newly established settlements preferentially occupied more favorable gentle valleys, with terraces and gentle slopes accounting for a growing share in total settlement. This does not mean that altitude and slope were irrelevant; rather, once basic water supply and arable conditions were met, environmental choice shifted from a "concentration in the most optimal location" to an "elastic expansion across multiple geomorphological units." Regional studies document that graben basins and tectonically formed mountains and hills provided abundant favorable settlement resources, and the Holocene warm-humid climate further supported habitation. In some areas, even high-elevation hills and hillsides—with flat summit terrains and natural defensive advantages—became new settlement options^[23].

With the decline of the mid-Holocene warm period, the climate became drier and cooler, and climate fluctuations intensified. High-resolution paleoclimatic records indicate a clear climate shift in the mid-Holocene (roughly 6.0–5.0 cal ka BP): multiple cooling events and monsoon weakening occurred, followed by the end of the Holocene warm period^[24]. In elevation terms, sites in the 100–200 m low-elevation interval accounted for 59.8% of sites, indicating renewed clustering toward the optimal river-valley zones. Slope selection also trended toward flatter terrains, with only a minority of settlements expanding into slightly steeper marginal areas; Yangshao populations thus moved again toward low-latitude and low-elevation river valleys in the late period.

Taken together, the mechanisms by which environmental factors influence Yangshao settlement choices can be summarized as follows: in the early period, habitable space was abundant and environmental constraints were weak; in the middle period, the warm-humid phase improved water-heat conditions, constraints weakened, and choice became more elastic, with settlements modestly expanding toward higher elevations and secondary geomorphologies; in the late period, climate fluctuations tightened constraints and settlements re-concentrated in low-elevation river valleys. This process is not simply environmental determinism but the result of the joint action of environmental conditions and social organizational capacity. People were evidently aware, at least from early on, that environmental changes affected their production and daily life. When the overall environment was favorable, there was proactive exploration and development beyond core aggregation zones; during climatic perturbations, deliberate constructions such as ring ditches and, later, artificial lakes and wetlands helped concentrate populations in the suitable areas.

Clustering Patterns and the Process of Yangshao Social Complexity

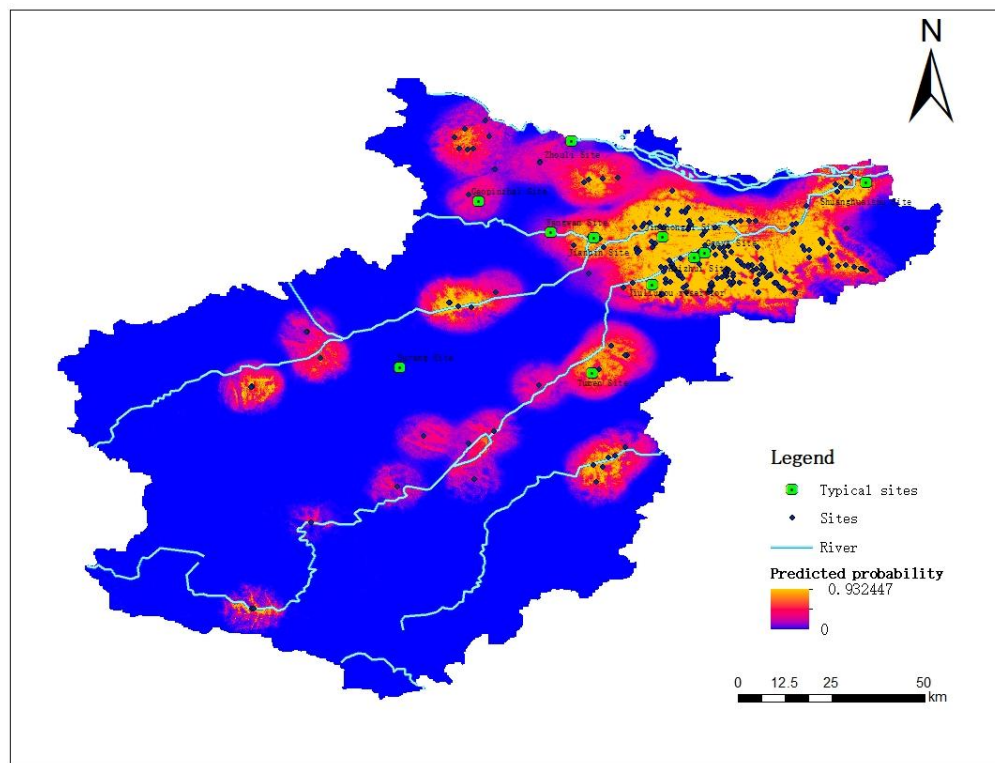


Figure 9. Comparison between high-suitability areas in the predicted distribution map and existing Yangshao cultural sites.

Synthesizing the outcomes of spatial autocorrelation and Maxent models, the Luoyang Basin Yangshao site pattern transitions from initial random dispersion to a mid-to-late period of strong clustering into settlement complexes, a trend that is more evident in the predictive maps (Figure 9) produced by the models.

Early Yangshao culture shows relatively low developmental intensity, social equality with no clear social classes. Settlements are predominantly small, households rely on the natural environment, and spatially the settlements show no strong inter-site correlation, exhibiting a dispersed banded distribution. This is consistent with the early period's relatively high AUC scores in the predictive models, which indicate that settlement location was highly dependent on natural environmental conditions, and even when spatial organization variables are added, their influence on model predictions remains limited.

In the middle Yangshao period, cultural features become evident: with distinctive painted pottery signals, the cultural influence radiates across nearly half of China, corresponding to a central “heart” of the Central Plains under a “multi-petal flower” pattern. Settlements increase markedly in population; house building and agricultural economy become mature and prosperous, and productivity rises. Burial goods, house construction, and other aspects show clear social stratification. Central settlements show pronounced hierarchical differentiation, with large sites such as Shuanghuashu and Suyang appearing.

In the late Yangshao period, the process of social complexity accelerates significantly, entering a civilizational stage. Population scales continue to grow,

agriculture and handicraft industries develop rapidly, social stratification becomes explicit, forming a pyramid-like hierarchy of core, secondary, and ordinary settlements. The region's existing central sites continue to develop, with the emergence of a super-large central settlement like Shuanghuashu. Surrounding peripheral sites in Xishan, Daheshi, and other centers, along with functional sites for pottery and jade production, collectively constitute a regional central settlement complex. This regional agglomeration strengthens, and during climatic fluctuations this “grouping” pattern is reinforced by advantageous location and internal coordination, contributing to defense and disaster resilience.

Under the joint influence of social complexity progression and the spatial distribution of settlements, site location is no longer determined solely by natural suitability. Social factors such as population competition and resource allocation shape the pattern of settlements. Natural suitability remains the foundation, but political power and community organization further filter and reshape actual settlement distribution. On the one hand, core low-lying, gently sloped terrains remain the preferred choice, providing a stable material base for large central settlements. At the same time, driven by population pressure, social competition, and social stratification, central settlements occupy the more suitable zones, while a minority of ordinary settlements must relocate to higher slopes or slightly less favorable environmental conditions at the margins. On the other hand, as central settlement influence expands into a regional super-settlement, and forms a pyramid relationship with scattered surrounding settlements, the Yangshao mid-to-late period witnesses the emergence of organized regional political entities ^[25], yielding settlement groups with shared cultural undercurrents, clear internal divisions of labor, and close cultural exchange. Regional agglomeration effects strengthen, and this “grouping” mode can even provide defense and disaster resilience during climate perturbations thanks to favorable location and internal coordination.

Limitations and Prospects

This study integrates spatial organization features with MaxEnt modeling to address not only how settlements are distributed, but also why these distributions occur by linking distribution patterns to their underlying drivers. The combined approach demonstrates that spatial structure and environmental/contextual constraints jointly shape settlement patterns, enabling a more holistic interpretation that couples pattern recognition with driver analysis.

Nevertheless, several limitations warrant emphasis. First, the paleoclimate and paleotopography reconstructions employed are approximations of the ancient environment and carry inherent uncertainties. The derived variables used in the analyses—kernel density, Thiessen (Voronoi) tessellations, and spatial autocorrelation—are themselves contingent on site-distribution data and thus introduce additional uncertainty when used to predict site locations. It is essential to distinguish clearly between the predictive utility of archaeological models and their inferential interpretations: although MaxEnt performs well in prediction, directly attributing its outputs to causal mechanisms of site selection requires grounding in archaeological context and social development.

Second, the available site data reflect known discoveries and survey results and thus contain data gaps, with stronger survey intensity typically occurring in population-concentration zones. Such survey bias can systematically influence model training and interpretation. Therefore, conclusions should be tempered by considering model assumptions, the errors inherent in ancient environmental reconstructions, and the broader archaeological background to avoid over-extending quantitative inferences about past human behavior.

Looking forward, incorporating richer archaeological materials and more precise data and methodologies could further refine the framework. Potential enhancements include integrating cost-path distance and transportation-network analyses to portray inter-site spatial relationships more accurately. Applying the methodological framework to other regions (e.g., the upper Wei River and the Zheng-Luo area) would test the regional generality of the observed “strengthened spatial organization constraint” pattern and help archaeologists design more targeted site surveys. In this way, the combined use of spatial organization metrics and predictive modeling can meaningfully inform both theory and practice in archaeological research.

Reference

1. Wenming, Y. Archaeology toward the 21st Century. Sanqin Publishing House, Xi'an, 2017.
2. Institute of Archaeology, Chinese Academy of Social Sciences. China – Australia – America – Iran Joint Archaeological Team of the Yiluo River Basin. Pre-Qin Period Sites in the Central and Eastern Luoyang Basin: 1997 – 2007 Regional Systematic Survey Report. Beijing: Science Press, 2019.
3. Liu J. GIS-supported Settlement Archaeology Research. PhD thesis, China University of Geosciences (Beijing), Beijing, 2007.
4. Bi S., Wan L., Shen X., et al. Spatial autocorrelation analysis of prehistoric settlement distribution characteristics in the Zheng – Luo region. *Science of Surveying and Mapping*. **43**, 87 – 94 (2018).
5. Yan L., Zhang Y., Lu P., et al. Research on the scale hierarchy of early settlements based on the K-means clustering method. *Areal Research and Developmen*. **39**, 176-180 (2020).
6. Bi S., Ji H., Chen C., et al. Application and analysis of geographic detectors in the study of prehistoric settlement human – land relationships. *Journal of Geo-information Science*. **17**, 118-127 (2015).
7. Liu J., Wang L. Settlement archaeology research supported by spatial analysis techniques. *Remote Sensing Information*. **3**, 51 – 53+94 (2006).
8. Xing D., Hao Z. Maximum entropy principle and its application in ecological research. *Biodiversity Science*.; **19**, 295 – 302 (2011).
9. Li F., Li L., Wu G., et al. Habitat suitability assessment of snow leopard in Qilian Mountains of Qinghai based on maximum entropy model. *Acta Ecologica Sinica*. **43**, 2202-2209 (2023).
10. Yang T., Ma L. Research on drought hazard risk in Sichuan Province based on maximum entropy principle. *Plateau and Mountain Meteorology Research*. **41**, 103-107 (2021).
11. Begade, S. A Predictive Model for the Identification of Open-Air Palaeolithic Sites in Nagpur and Chandrapur Districts, Maharashtra, India. *J. Archaeol. Sci. Rep*. **69**, 105541 (2026).
12. Rostami, M., Javanmardzadeh, A., Saed Mucheshi, A. Predictive Modelling of Iron Age Sites: A Case Study of the Eastern Kurdistan Region, Iran. *Parseh J. Archaeol. Study*. **9**, 95-120 (2026).
13. Wu, R., Gan, S., Yu, L., et al. Research on predictive distribution of archaeological sites in Pakistan based on maximum entropy model. *Science of Surveying and Mapping*. **46**, 96-103 (2021).
14. Liu, J. "Sky Eye" seeing through ancient rivers and mountains: Settlement and environmental archaeology in the Luoyang Basin supported by remote sensing and GIS. *World Heritage*. **8**, 49-51 (2015).
15. Archaeological Fieldwork Team of Peking University. Brief report on the excavation of the Wangwan site in Luoyang. *Archaeology*. **4**, 175-178 (1961).
16. Gu, W., Wang, X., Hu, Y., et al. The Shuanghuaisu Neolithic site in Gongyi City, Henan Province. *Archaeology*. **7**, 27-48+2 (2021).

17. Ren, G. Brief report on the excavation of the Xiacun area at the Suyang site in Yiyang, Henan Province. *Cultural Relics of Central China*. **5**, 23-38 (2021).
18. Wang, J., Liao, Y. & Liu, X. Spatial Data Analysis Tutorial (Science Press, Beijing, 2019).
19. Bi, S. et al. Spatial autocorrelation analysis of prehistoric settlement distribution characteristics in the Zheng-Luo region. *Science of Surveying and Mapping*. **43**, 87-94 (2018).
20. Gao, M. et al. Spatiotemporal evolution and human-environment relationships of early cultural sites from the Longshan to Xia-Shang periods in Henan Province, China. *npj Heritage Sci*. **13**, 74 (2025).
21. Pang, X. & Gao, J. Investigation of agricultural economy in the process of civilization in the Central Plains region. *Agricultural Archaeology*. **4**, 1-13+26 (2006).
22. Zheng, J. Excavation and research of the Yangshao culture in the Central Plains region. *Cult. Relics Cent. China* **2**, 9-21 (1996).
23. Lu, P. et al. Relationship between settlement distribution and regional tectonic structure in the circum-Songshan area during 9000 – 3000 aBP. *Acta Geographica Sinica* **69**, 738-746 (2014).
24. Wu, W. et al. 5.5 cal kaBP climate event, population growth, condition constraints and the emergence of prehistoric complex societies in China. *Scientia Sinica(Terrae)*. **48**, 138-151 (2018).
25. Wei, X. Settlement archaeology and civilization achievements in the core area of the middle Yangshao culture period. *Huaxia Archaeology*. **5**, 67-89 (2024).