

陇东先秦遗址多光谱遥感与遗址预测模型研究

Archaeological site predictive models based on DEM and multi-spectral remote sensing data on the pre-Qin sites in the Loess Plateau of East Gansu Province
张海;徐艺菁;周静

引用本文:
张海, 徐艺菁, 周静. 2021. 陇东先秦遗址多光谱遥感与遗址预测模型研究. 遥感学报, 25 (12): 2396-2408 DOI: 10.11834/jrs.20211133.
Zhang H, Xu Y J and Zhou J. 2021. Archaeological site predictive models based on DEM and multi-spectral remote sensing data on the pre-Qin sites in the Loess Plateau of East Gansu Province. National Remote Sensing Bulletin, 25 (12): 2396-2408 DOI: 10.11834/jrs.20211133.

在线阅读 View online: <http://www.ygxb.ac.cn/thesisDetails?columnId=19917391>

您可能感兴趣的其他文章

Articles you may be interested in

融合多源时序遥感数据大尺度不透水面覆盖率估算

Estimation of large-scale impervious surface percentage by fusion of multi-source time series remote sensing data
遥感学报. 2020, 24(10):1243-1254.

深度置信网络算法反演Landsat 8 OLI气溶胶光学厚度

Inversion of aerosol optical depth for Landsat 8 OLI data using deep belief network
遥感学报. 2020, 24(10):1180-1192.

WorldView-2影像与OLI影像协同岩性模糊分类

Lithological fuzzy classification by combining WorldView-2 data and OLI data
遥感学报. 2022, 26(6):1247-1259.

多光谱影像NDVI阴影影响去除模型

Research on multispectral-image-based NDVI shadow-effect-eliminating model
遥感学报. 2020, 24(1):53-66.

{{relationArticle5.title}}
{{relationArticle5.enTitle}}

陇东先秦遗址多光谱遥感与遗址预测模型研究

张海¹, 徐艺菁¹, 周静²

1. 北京大学 考古文博学院, 北京 100871;

2. 甘肃省文物考古研究所, 兰州 730000

摘要: 考古遗址预测模型 (SPM) 在考古学研究和文化资源管理中有着广泛的应用, 但传统的构建方法过度依赖数字高程模型 (DEM), 对多光谱遥感数据的应用较少。本研究以陇东黄土高原地区的区域考古调查为例, 利用第三次全国文物普查数据、商业 DEM 数据和 Landsat 8 OLI 多光谱遥感数据, 构建 3 种考古遗址预测模型 (DEM 模型、多光谱遥感模型和混合模型)。研究表明, 黄土高原地区先秦遗址分布对坡向、纵向曲率、地形开阔度等地貌因素具有明显的相关性, 而引入主成分分析法 (PCA) 提取的多光谱因子不仅可以有效提高单纯基于 DEM 派生数据构建的遗址预测模型的精度和预测效率, 而且还能够很好地改善遗址预测模型的空间结构, 增强遗址预测模型的应用效果。基于上述混合模型的分析, 陇东黄土高原地区先秦时期的土地资源开发, 从仰韶文化时期的不足 10%, 迅速提升至龙山文化时期的 43.1%, 先秦时期整体达到了 57.3% 的规模, 土地资源的加速开发奠定了该地区早期文明发展的基础。

关键词: 遥感, 陇东黄土高原, 考古遗址预测模型, Landsat 8 OLI, 先秦时期土地资源开发

引用格式: 张海, 徐艺菁, 周静. 2021. 陇东先秦遗址多光谱遥感与遗址预测模型研究. 遥感学报, 25(12): 2396–2408

Zhang H, Xu Y J and Zhou J. 2021. Archaeological site predictive models based on DEM and multi-spectral remote sensing data on the pre-Qin sites in the Loess Plateau of East Gansu Province. National Remote Sensing Bulletin, 25(12): 2396–2408 [DOI: 10.11834/jrs.20211133]

1 引言

考古遗址预测模型 SPM (Site Predictive Modelling) 是利用地理信息系统 GIS (Geographical Information System) 软件工具实现对考古遗址的赋存环境、生态、文化等变量的空间结构进行量化评估、建模和预测分析的一种重要手段。考古遗址预测模型随着 GIS 技术兴起而出现, 最初由考古学家 Green 等 (1973) 在南美玛雅遗址的空间分析中使用, 后经过 Kvamme (1983) 的改进和应用推广, 逐步成为 GIS 技术在考古学研究和文化资源管理中最重要应用领域之一。21 世纪以来, 国内考古和古环境研究工作者也开始关注和使用考古遗址预测模型开展聚落和环境考古的研究工作。除了理论和方法的综述 (梅启斌和汪诚波, 2006) 之外, 也出现了一些应用方面的研究: 包括利用考古遗址预测模型讨论遗址选址与地貌因素的关

系 (倪金生, 2009; 彭淑贞等, 2010; 乔文文等, 2013; 董振和金石柱, 2015; 郭飞, 2018), 评估特殊文化变量对遗址空间布局的影响 (Zhang 等, 2013), 以及利用遗址预测模型建构古代土地资源利用的 PLUM 模型 (于严严等, 2010) 等。总的来讲, 这些探索性的工作推动了考古遗址预测模型方法的成熟和完善, 比如增加“梯度上升法”提高模型的预测效率 (郭飞, 2018), 但是在具体应用中也存在不可忽视的问题。通常情况下, 构建考古遗址预测模型最基本的地貌因子多源自于数字高程模型 DEM (Digital Elevation Model), 如: 坡度、坡向等 DEM 的一级派生因子, 地形曲率、地貌类型等 DEM 的二级派生因子。由于数据来源单一, 这些地貌变量无法有效表达人类活动的空间异质性 (Spatial heterogeneity) 特征, 而对 DEM 数据的过度依赖几乎是目前所有考古遗址预测模型都存在的问题。实际上, 遥感数据能够提供更

收稿日期: 2021-03-20; 预印本: 2021-09-22

基金项目: 国家重点研发计划 (编号: 2020YFC1521900)

第一作者简介: 张海, 1979 年生, 男, 副教授, 研究方向为 GIS 空间分析。E-mail: haizhang@pku.edu.cn

多的地表分类信息，是除了DEM之外构建考古遗址预测模型的另外一类可选的数据来源。西方文献中很早就有利用遥感数据构建考古遗址预测模型的尝试（Custer等，1986），遗憾的是国内此类研究还十分少见。多光谱遥感数据能够很好表达现代土地利用和地表覆盖状况，在一些情况下同样能够很好反映区域地貌特征。中国黄土高原地区地貌类型多样且差异显著，是开展多光谱遥感构建考古遗址预测模型的理想区域。本项研究利用陇东黄土高原地区考古区域系统调查获取的可靠遗址位置和年代分类数据，综合应用基于DEM和基于多光谱遥感两类数据来源的考古遗址预测模型，并对应用效果进行系统评估，尝试提出多光谱遥感在构建考古遗址预测模型以及区域考古调查和研究中应用的新方法。

陇东地区处于中国黄土高原的中东部边缘，是黄土地貌发育完整的区域。这里的黄土从中更新世晚期开始至全新世持续堆积，最厚可达300 m以上。由于受到河流侵蚀作用的影响，黄土残塬、梁、峁不仅发育典型，而且地形破碎，区域内部的地貌差异较大，与此相关的土地利用类型也有显著差异，适合于开展有关人类活动与地貌、土地利用关系的区域性研究。

本研究所选取的具体区域位于黄河支流泾河上游的支流黑河与达溪河流域（34.92422°N—35.28486°N，107.04597°E—107.78686°E），总面积2373 km²。黑河古称“汭水”属于泾河的一级支流，而达溪河为黑河的最大支流，古称“百里溪”，因商末密须国而得名，现代行政区主要位于甘肃省平凉市的灵台县和泾川县。这里属于亚热带半干旱季风气候区，现代年均降雨量高达620 mm，但集中在夏秋季节，黄土地貌遭水流侵蚀严重，属于黄土丘陵沟壑区。

已有的考古工作表明，该地区自新石器时代以来就是古文化发达的区域，尤其是先秦时期的遗址密集，著名的灵台桥村遗址（甘肃省博物馆考古队，1980）、白草坡西周墓地（魏怀珩和伍德煦，1972；初仕宾，1977）均位于此地（图1）。根据第三次全国文物普查的资料，该区域自仰韶文化中期（6—5 ka BP）以来社会发展呈加速度，至龙山文化晚期（4.3—3.8 ka BP）达到区域社会发展的高峰，不仅聚落数量和规模空前，而且社会分化现象显著。以桥村遗址为代表，中心聚落

面积超过100万 m²，且发现有品类丰富的玉器和目前发现最早的建筑用瓦（包括筒瓦和板瓦）之一。因此，可以利用第三次全国文物普查的资料，把先秦时期当作是一个完整的研究时段，在早期社会复杂化的背景下讨论遗址的空间布局与黄土高原特定环境的关系。

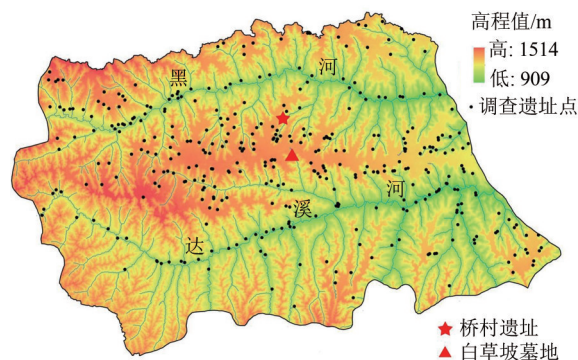


图1 研究区域及其位置

Fig. 1 Research area and its location

李水城和李非（1992）与莫多闻等（1996）在邻近的葫芦河流域的调查表明，仰韶至龙山文化时期气候转向干凉，遗址的空间位置也从河谷阶地向海拔更高的黄土台塬迁移。同样的情况也发生在本区域内，桥村遗址的调查和发掘均显示了龙山文化时期以修建窑洞式房屋为主体的居住形态，遗址倾向于分布在黄土塬面的前缘向谷底延伸的侵蚀面上，较临近河谷的仰韶文化聚落相比，海拔高度和地貌位置都发生了明显变化。修建窑洞需要利用坡面斩崖掏洞，而当窑洞塌陷之后，往往还继续向内掏挖新的窑洞，如此反复的人类活动破坏了黄土台塬坡面的结构稳定，加剧了侵蚀的发生。与此同时，宽阔的黄土塬面则是农业生产的主要区域和古代社群之间交往的通道网络。可以说，该地区自龙山文化晚期开始的以大规模窑洞式居住形态为基础的土地利用方式不仅与特殊地貌位置的关系密切，而且其延续性强，直至今日。因此，该地区地貌类型和土地利用方式的差异都与古代遗址的空间布局存在相关性，可以进一步利用考古遗址预测模型进行分析评估。

2 数据与方法

2.1 数据来源及其处理

本项研究所使用的数据来源主要为

(1) 研究区域的数字高程模型 (DEM) 来自于 ALOS 卫星搭载的 PRISM 传感器利用立体相对获取的 5 m 空间分辨率的数字表面模型 (DSM)。通过对原始数据设置 70° 坡度的阈值, 采用去除地表植被和建筑物的滤波处理方法, 实现从表面模型到地表模型的转换 (Conrad 等, 2015), 并采用均值法重采样至 30 m 空间分辨率以减少误差, 同时降低分辨率以提高后期运算的效率。

(2) 研究区域的遥感数据来源于 Landsat 8 卫星 OLI 陆地成像仪提供的 30 m 空间分辨率的多光谱数据。选择了获取时间为 2015-11-29 的数据, 主要是考虑到冬季落叶后植被相对稀疏, 可有效减少茂密乔木和灌木植被对地表覆盖物判读的影响, 便于更准确地进行土地类型的识别。原始数据通过 ENVI5.3 软件提供的 FLAASH (Fast line-of-sight Atmospheric Analysis of Spectral Hypercubes) 工具纠正大气影响, 为后续研究提供准确的地表覆盖参数。

(3) 考古遗址的基础数据源自第三次全国文物普查 (简称“三普”) 登记在案的原始数据。这些数据详细记录了遗址位置的 GPS 实测经纬度坐标、文化属性和面积等信息。其中, 记录有先秦时期的遗址共有 368 处, 包括仰韶文化遗址 63 处, 龙山文化遗址 277 处。2019 年 7—8 月采用随机抽样的方式对其中 30 处仰韶或龙山文化遗址 (抽样率 10%) 进行了实地复查, 结果表明“三普”资料在遗址位置、文化属性方面的记录比较准确, 但在面积的估算上差异较大。通常情况下, 在利用逻辑回归方程构建遗址预测模型时, 需要有相对准确的遗址地理坐标, “三普”数据基本能满足这一要求。

需要指出的是, 由于黄土丘陵侵蚀区的局部地貌差异很大, 对于 30 m 分辨率的数字高程模型和多光谱遥感数据在一个遗址上存在多个像元值的问题 (单个像元大小 900 m², 而遗址面积多在数千至几十万平方米, 单个遗址涵盖几十至上千个像元), 因此不能直接采用遗址点获取地貌参数或多光谱数值。理论上最佳的解决方案应以准确的遗址文化堆积的实际分布范围对像元进行统计, 但在实践中受到时间、人力、经费等条件制约无法实现。因此, 采用折中方案, 以“三普”记录的遗址面积的均值 4.5 万 m² 为参照, 从每个遗址点出发生成 120 m 半径的圆形缓冲区, 以此为依据统

计各遗址的地貌参数或多光谱数据的均值, 从而最大程度上减少由于遗址内部的差异而带来的统计误差。

2.2 变量分析与选取

构建考古遗址预测模型最关键的因素是选取合适的变量。本项研究的变量主要来自于 DEM 和多光谱遥感, 详见表 1—表 2。

地貌因子包括: (1) 形貌特征, 坡度、坡向、曲率、地表崎岖度、纹理度和凸度; (2) 水文特征, 侵蚀基面高度、河谷深度、河流垂直距离、地形湿度指数和坡长坡度因子; (3) 位置特征, 地形位置指数、相对坡位、地形开阔度、风暴露指数和地貌类型。这些参数的计算均由 DEM 通过 SAGA-GIS7.9 软件提供的地貌分析工具分级派生而来。

将上述反映区域地形地貌特征的因子按考古遗址进行分类统计, 并将研究区域的所有像元作为背景值, 分别观察仰韶和龙山文化的遗址分布频率直方图, 如图 2 所示。

(1) 仰韶文化遗址集中在坡度 10°—25° 区间, 龙山文化遗址的坡度更陡, 集中在 15°—30°, 平均高出 5° 左右; 仰韶文化遗址集中朝向南和西南方向 (150°—270°), 而龙山文化遗址的朝向则更宽, 虽仍以南向为主, 但也出现了一些朝向北的情况。

(2) 仰韶至龙山文化的遗址发生了空间位置上的显著变化。相对坡位指数 (RSP) 增大表明遗址向更高山脊顶部方向移动, 相应的侵蚀基面 (CNBL) 增加, 河谷深度 (VD) 降低。这种变化主要是由于遗址相对海拔位置的升高造成的。

(3) 由于遗址垂直空间位置的变化, 也造成了相应区域环境变量的改变。如更高的地形崎岖度 (TRI) 和地形纹理 (TEX) 特征表明龙山文化遗址多处于沟壑山脊密集交错的复杂地貌, 而更高的风力暴露指数 (WEI) 和地形开阔度 (TO) 代表遗址所处的环境更加开阔, 也更利于彼此之间的交往和社会交流网络的加强。

(4) 上述区域性变量的变化十分显著, 但从地表曲率等局部地貌指标来看, 仰韶至龙山的变化不大, 表明在小尺度地貌选址方面依然保持了连续性。

总之, 无论从区域性还是局部性的尺度看,

仰韶、龙山文化遗址的选址都与地貌存在密切的联系，是可用于构建遗址预测模型的重要因子。

除了地貌因子之外，本项研究中还考虑了土地利用因子，主要是基于多光谱的遥感数据进行地表覆盖类型的区分。Landsat 8的OLI传感器提供了9个波段的数据，其中，第1—7波段覆盖从海岸短波至短波红外波段（波长0.43—2.29 μm），是

可用于区分地表覆盖物类型的主要波段。由于不同波段之间的相关性很强，采用主成因素分析法PCA（Principal Component Analysis）对第1—7波段进行降维。从降维后的特征值占比来看，前3个主成分因子已经能够解释所有波段99%的信息，因此提取前3个主因子作为构建遗址预测模型的变量因子（表2）。

表1 变量来源及描述
Table 1 Source and description of variables

		因子	算法	应用
DEM 衍生	地貌	坡度(Slope)	最大坡降坡度	遗址坡度
		坡向(Aspect)	最大坡降朝向	遗址光照、受风状况
		截面曲率CSC(Cross-Section Curvature)	与坡向正交的地形截面方向曲率	局部范围的遗址横向地形特征
		纵向曲率LC(Longitudinal Curvature)	与坡向平行的地形截面方向曲率	局部范围的遗址纵向地形特征
		地形崎岖度指数TRI(Topographic Ruggedness Index)	中心点高程与邻域点的总差值	局部地形的多样性
		地表纹理TEX(Terrain Surface Texture)	单位面积内山峰与山谷出现的频率	高起伏地形特征: 低值代表平原; 高值代表山脊与山谷密集地区
		地表凸度TSC(Terrain Surface Convexity)	由曲率计算的单位面积地表凸起和凹陷的程度	低起伏地形特征: 低值代表宽谷、山前坡地; 高值代表冲积扇、阶地
	水文	侵蚀基面高度CNBL(Channel Network Base Level)	河网海拔高度的差值面	河网的海拔高度
		谷深VD(Valley Depth)	山脊线差值的基准面减去绝对高程	从源面计算遗址所处的河谷深度
		与河网垂直距离VDCN(Vertical Distance to Channel Network)	河网侵蚀基面高程减去绝对高程	计算遗址与河流的垂直距离
		地形湿度指数TWI(Topographic Wetness Index)	汇流量与坡度比值	模拟静态的土壤含水量指标
		坡长坡度因子LSF(LS-Factor)	通用土壤侵蚀方程的坡长(L)与坡度(S)变量	与地形相关的土壤流失量
	位置	多尺度地形位置指数TPI(Multi-Scale Topographic Position Index)	中心点高程值减去邻域点高程的均值,再按照最大最小值标准化处理	经标准化处理后: 0代表平地, 正值代表山脊, 负值代表山谷, 数值大小代表不同尺度的效果
		相对坡位RSP(Relative Slope Position)	相对于山脊线和山谷线计算的坡度位置	0代表山谷底部, 1代表山脊顶部
		地形开阔度TO(Topographic Openness)	按固定范围计算的中心点与四周最高或最低视线点的高度	有仰视(+)和俯视(-)两种, 可用来衡量遗址的视觉景观
		风暴露指数WEI(Wind Exposition Index)	中心点所在位置受各个方向的风力影响程度	大于1代表受风影响, 小于1代表不受风影响
		地貌类型LFC(TPI Based Landform Classification)	基于TPI的地貌类型划分	共分10类: 高、中、低山, 上、下坡地, 平地, 河谷, 高、中、低河道
Landsat OLI 多光谱	主成分因子	第一主成分因子(FC1)	B1-B7波段主成因素分析	可解释总变量的80%和B1—B7波段
		第二主成分因子(FC2)	B1-B7波段主成因素分析	可解释总变量的16.3%, B5波段
		第三主成分因子(FC3)	B1-B7波段主成因素分析	可解释总变量的2.6%, B1、B2波段

表2 Landsat 8 OLI 多光谱数据PCA 分析的特征值及其占比
Table 2 Eigenvalues and theirs proportion of PCA analysis for Landsat 8 OLI multispectral data

参数	主成分因子						
	FC1	FC2	FC3	FC4	FC5	FC6	FC7
特征值	5409831	1089720	176532.2	18644.29	7445.118533	878.8544	317.3552
百分比/%	80.7	16.26	2.63	0.28	0.11	0.02	0
累计百分比/%	80.7	96.96	99.59	99.87	99.98	100	100

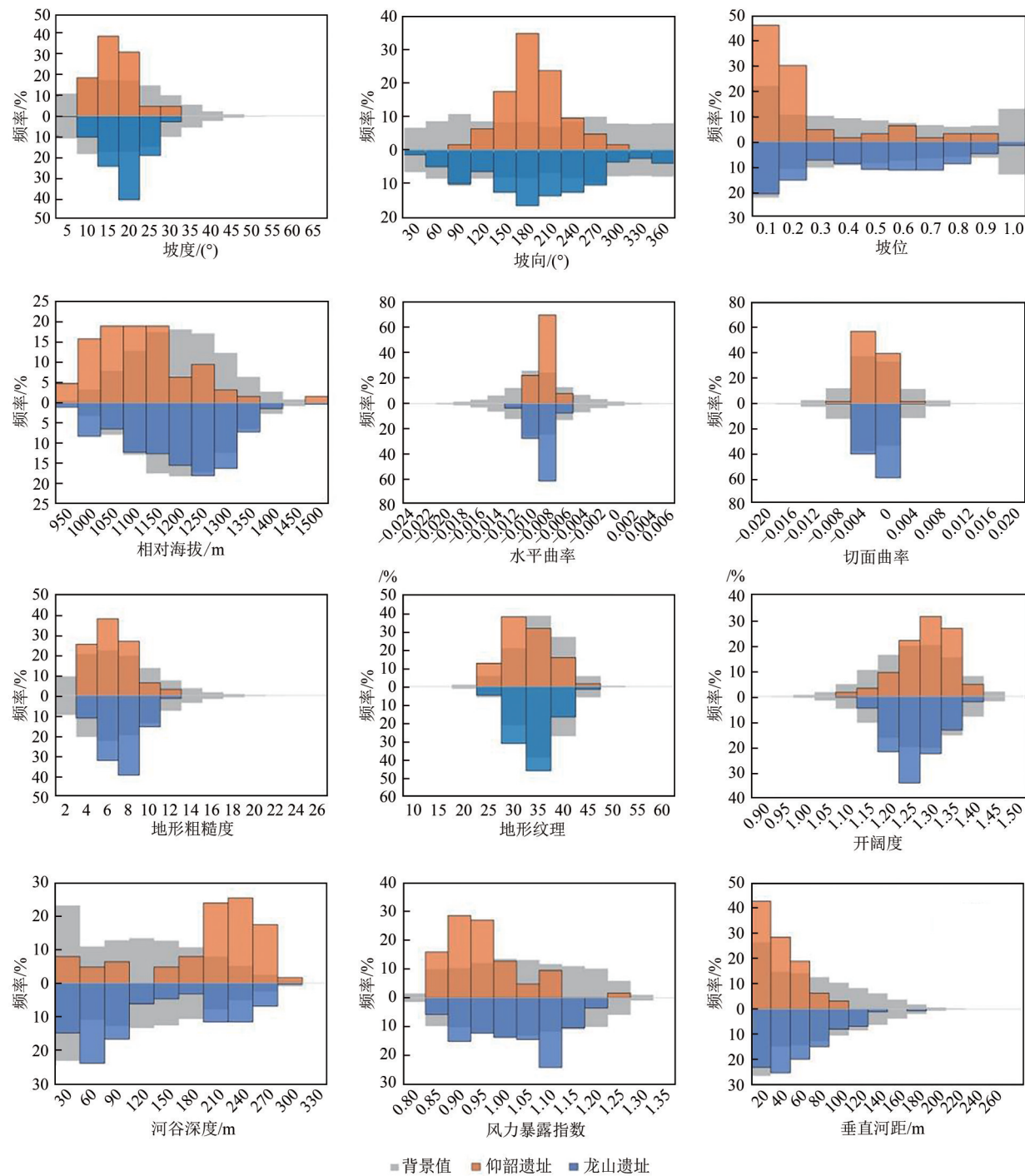


图2 仰韶、龙山文化聚落与各类地貌参数的关系

Fig. 2 The relationship between settlements of Yangshao and Longshan Culture and the geomorphological parameters

结合由特征向量计算的PCA各因子与不同波段之间的载荷矩阵可以发现（表3）：第一主成分因子FC1能够代表所有的7个波段，解释其中80.7%的总信息量，但不可忽视的是第1、2、5波段中有部分独立的信息仍需要用第二（FC2）、三（FC3）主成分因子来解释。

需要注意的是，经过PCA方法降维的主成分因子分量并不具备明确的物理意义，因此也不适合直接讨论它们与考古遗址的关系。利用GIS的栅格统计工具可以对多光谱主成分因子和地貌因子以像元为单位进行相关分析，计算相关系数（ R^2 ），结果如表4所示。

表3 Landsat 8 OLI多光谱数据PCA分析的因子载荷矩阵
Table 3 Component matrix for PCA analysis of Landsat 8 OLI multispectral data

	波段1	波段2	波段3	波段4	波段5	波段6	波段7
FC1	-0.694790	-0.763520	-0.855860	-0.861050	-0.793020563	-0.968980	-0.939340
FC2	-0.054500	-0.031340	-0.154770	0.067721	-0.608054718	0.226916	0.335050
FC3	0.634055	0.605149	0.488309	0.491183	-0.035734178	-0.086180	0.012921
FC4	-0.284460	-0.187600	-0.025270	0.062623	0.009591413	-0.042590	0.063241
FC5	-0.160670	-0.108880	0.005080	0.091919	-0.004018968	0.018586	-0.034770
FC6	-0.036560	-0.018460	0.066516	-0.019570	-0.001280845	0.000641	-0.000410
FC7	0.052709	-0.049690	0.003719	0.001532	-2.50218E-05	-0.000140	0.000306

注：粗体代表高因子载荷。

表4 多光谱因子与地貌因子的空间相关性(R^2)
Table 4 Spatial correlation between PCA factors from OLI multispectral data and terrain factors from DEM (R^2)

	Slope	Aspect	CSC	LC	TRI	TEX	TSC	CNBL	VD	VDCN	TWI	LSF	TPI	RSP	TO+	WEI
FC1	3.88	4.48	13.70	9.29	3.79	2.44	0.72	0.08	1.05	1.73	1.13	7.98	14.20	3.37	15.10	5.35
FC2	5.48	1.75	5.62	2.81	5.00	3.46	0.01	2.03	1.24	0.10	0.10	7.91	5.07	0.01	9.65	0.02
FC3	8.89	2.05	7.10	4.77	8.21	5.51	0.11	0.06	0.01	0.05	0.09	12.20	7.38	1.36	16.70	1.73

注：粗体代表高相关系数。

但总体来讲，多光谱遥感与地貌因子之间的相关程度不高（相关系数均小于20%）。相比而言，地形开阔度指数（正值，TO+）与多光谱因子的相关性最强，说明地形开阔度与现代土地利用程度有一定关系。其中，FC1因子与垂直方向的地表曲率（包括截面曲率CSC和纵向曲率LC）、地形位置（地形位置指数TPI）、地形开阔度有一定的相关性；FC2因子主要与地形开阔度因子相关；FC3因子则与坡度（Slope）、地表崎岖度（TRI）、坡长坡度因子（LSF）和地形开阔度有相关性。因此，多光谱遥感因子只能有限地解释部分地貌因素，更多反映的应该是现代土地的利用状况。

2.3 研究方法

本文采用Kvamme（1983）提供的方法，利用逻辑回归方程（Logistic Regression）构建遗址预测模型。具体的操作流程如图3所示。

（1）选择遗址点（site）与非遗址点（non-site）。首先采用随机的方式对368处调查所知的先

秦遗址点选取其中二分之一（184处）作为遗址点，保留另外二分之一（184处）作为检验点，用作后期对模型预测效果的评估。非遗址点采用随机的方式首先在调查区域内自动生成，再依据其120 m圆形缓冲区范围不与遗址点重合的条件，对个别非遗址点进行调整，保证非遗址点范围内确实没有遗址的存在。

$$g(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p \tag{1}$$

$$P = \frac{1}{1 + e^{-g(x)}} \tag{2}$$

（2）构建最佳模型。利用sigmoid函数得到逻辑回归的线性回归方程。如公式1所示， x 为自变量， $g(x)$ 为因变量， β 为自变量系数。当只有一个自变量时， $g(x)=x$ ，为单变量逻辑回归，据此可分别对单因子变量进行显著性检验，选择强相关的变量（式（2）， $P<0.05$ ）。对于弱相关变量可做进一步归类处理后再进行相关分析，提取其中强相关的部分。然后，将所有提取的强相关因子进

行多变量的逻辑回归分析(式(1)、(2)),并检查方程中各自变量的贡献量,进一步剔除其中弱相关的自变量。最后,考虑到不同地貌变量均来自于同一个DEM数据源,并由此而来的内生相关性,再利用赤池信息量准则AIC(Akaike Information Criterion)(刘璋温,1980)对模型进行进一步的

梯度优化(式(3)): k 是所拟合模型中参数的数量, L 是对数似然值, n 是观测值数目。 k 取值越小, AIC 越小; L 取值越大, AIC 值越小。 k 小意味着模型简洁, L 大意味着模型精确), 选择既简洁又准确的最优预测模型。

$$AIC = (2k - 2L)/n \quad (3)$$

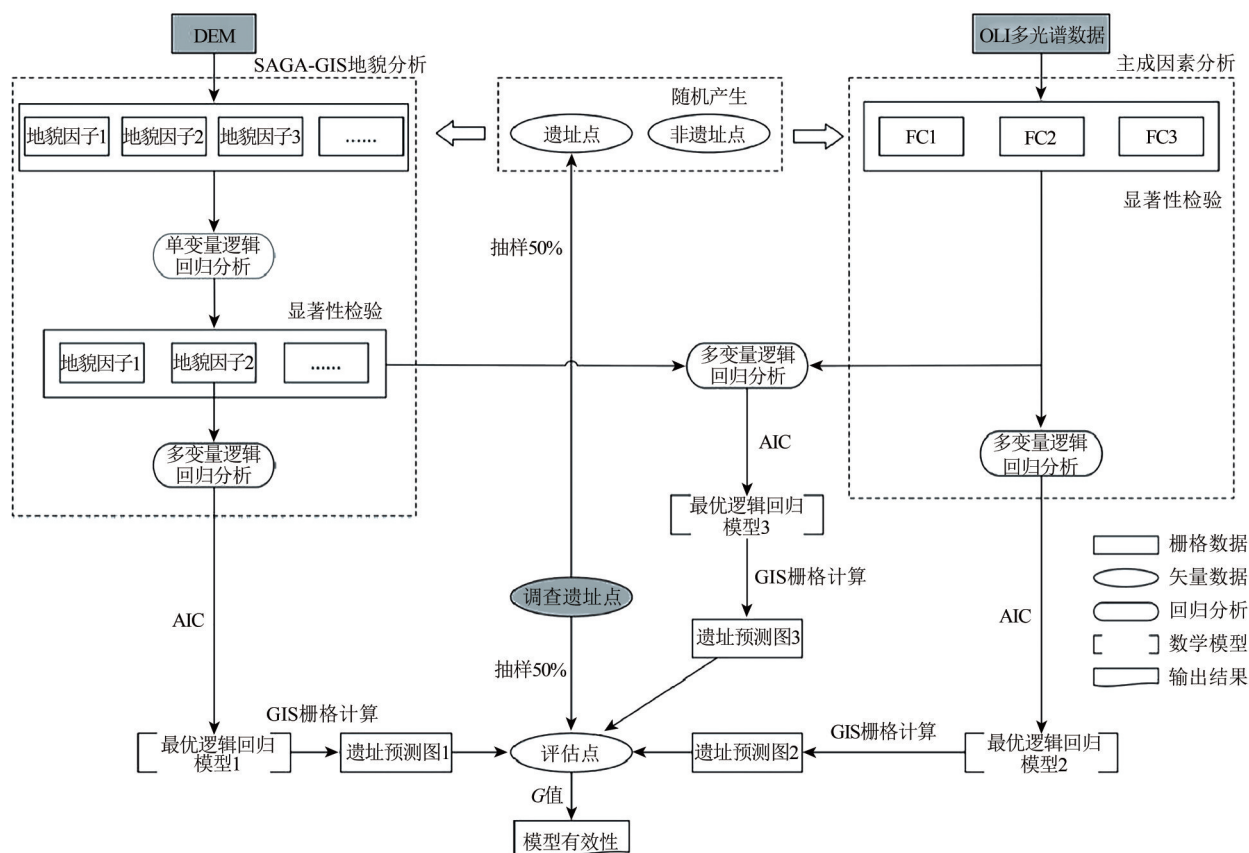


图3 遗址预测模型流程图

Fig. 3 Flow chart of SPM

(3) 基于DEM的地貌因子和基于多光谱的因子分别单独用来构建遗址预测模型,同时构建包含两类因子的混合模型,以便于对两种数据来源的不同模型预测效果进行比对。

模型的有效性评估方面,首先利用GIS的栅格计算器(Raster Calculator)按照上述最佳模型的参数和式(1)、(2)分别构建遗址预测模型,通过GIS空间统计直观计算模型之间的相关性。

然后,采用Kvamme提供的G值计算方法对上述两个预测模型进行实际预测效果的评估。按照式(4)的表述, A_p 为某预测概率的区域面积, A_T 为研究区域总面积, P_i 为此概率下遗址预测的成功率。因此,若研究区域概率区的面积越小,并在此概率区域内发现的遗址数量越多,则增益值 G 越

大;反之 G 值越小。 G 值越趋近于1,则模型的预测效果越好;越趋近于0,则模型预测效果越差。

$$G = 1 - \frac{A_p/A_T}{P_i} \quad (4)$$

为了提高评估的客观性,本文采用预留的184处遗址计算 G 值,并对上述3个预测模型进行综合比对,寻找其中差异的原因。除了 G 值评估的预测效果之外,基于GIS的空间分析方法也可用于不同预测模型的空间结构分析,并以此为基础讨论DEM和多光谱遥感数据综合应用于考古遗址预测模型的优劣。

3 结果分析

本文分别将DEM地貌因子和多光谱因子单独

作为自变量，再将两类自变量混合分别构建遗址预测模型，从而得到以下3个遗址预测模型：

模型1，由DEM地貌因子作自变量。经梯度优化后的最优模型选择的自变量有：坡向（aspect）、纵向曲率（LC）、地形开阔度（TO+）、地形纹理（TEX）、地形位置指数（TPI）、与河流垂直距离（VDCN）、地貌类型（LFC）。这些地貌因子中既有宏观因素，也有局部特征，既包括地形形貌，也包含水文和空间位置特征，表明对遗址进行有效预测的地貌因素较复杂(表5,图4(a))。

模型2，由多光谱遥感因子作自变量。由于在

主成分分析时采用了主轴旋转的方法，做到不同因子内部变异最小，相互之间变异最大的效果。因此，3个主成分因子最终都被选入最优模型，多光谱因子的遗址预测效率较高(表5,图4(b))。

模型3，由模型1和模型2的自变量共同组成。经梯度优化后的最优模型最终选择由多光谱因子的第2、第3主成分因子替代了地貌因子中的坡向（aspect），其他因子均保留，表明混合模型中仍以地貌因子为主，而多光谱数据中第2、3主成分因子具有地貌因子之外的其他因素，且优于坡向因子(表5,图4(c))。

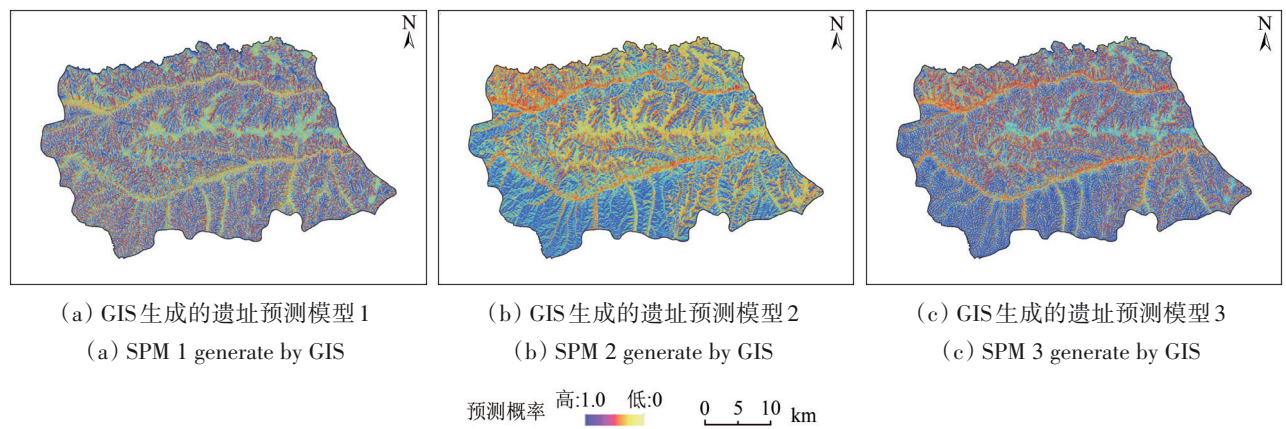


图4 陇东先秦遗址预测模型
Fig. 4 SPMs of pre-Qin sites in East Gansu Province

表5 遗址预测模型的最优二元回归方程因子系数及评估参数

Table 5 Coefficients and evaluation parameters of optimal logistic regression equation of SPMs													
模型	截距	Aspect	LC	TO(+)	TEX	TPI	VDCN	LFC	FC1	FC2	FC3	拟合系数 R^2	AIC
模型1	16.025	-0.001	327.421	-8.433	-0.062	2.007	-0.031	0.536	—	—	—	0.265	444.6
模型2	-0.416	—	—	—	—	—	—	—	-0.005	-0.004	0.005	0.312	420.0
模型3	15.356	—	360.578	-18.314	-0.048	1.515	-0.023	0.875	—	-0.007	0.009	0.444	379.1

从上述3个模型的预测结果的数学评估来看，无论逻辑回归方程的拟合系数 R^2 值还是AIC得分，模型2均优于模型1，如果综合了模型1、2因子的混合模型3，模型优度则显著提高。由于在考古区域系统调查和研究的实际工作中，只有高预测概率才具有实际意义，因此按照大于80%的高预测概率值来评估模型的预测效果。结果表明，无论从预测范围的面积百分比，还是对抽样用于检测的遗址的正确预测率来看，模型2均比模型1的效率高， G 值也明显提升。而混合模型3，其预测效率还能进一步提升（表6）。

表6 G 值检验模型预测的效果($P\geq 0.8$)

Table 6 The testing of effectiveness of SPMs with G value ($P\geq 0.8$)			
模型	参数		
	A_p/A_t (面积百分比)/%	P_t (正确预测率)	G 值
模型1	30.1	0.614	0.51
模型2	27.3	0.707	0.61
模型3	27.2	0.755	0.64

4 讨论

4.1 模型空间结构分析

上述模型评估的情况表明在基于DEM的地貌

因子的基础上引入多光谱遥感数据可以有效地提高考古遗址预测模型的效率。本文尝试对上述模型1、2、3的空间结构进行分析,探讨其中的原因。

原因1,利用GIS的栅格统计工具以像元为单位对模型1、2、3进行交叉统计,并绘制像元密度图,计算线性回归方程及相关系数 R^2 ,结果如图5

所示。模型1与模型2之间没有相关性($R^2=0.073$),表明模型1与模型2的数据空间结构存在很大的差异。然而,混合模型3与模型1、2之间均存在显著的相关性($R^2=0.544, 0.443$),表明混合模型3兼具模型1的地貌因素与模型2的多光谱遥感因素的空间结构特点,尤其是与模型1的相关性更强。

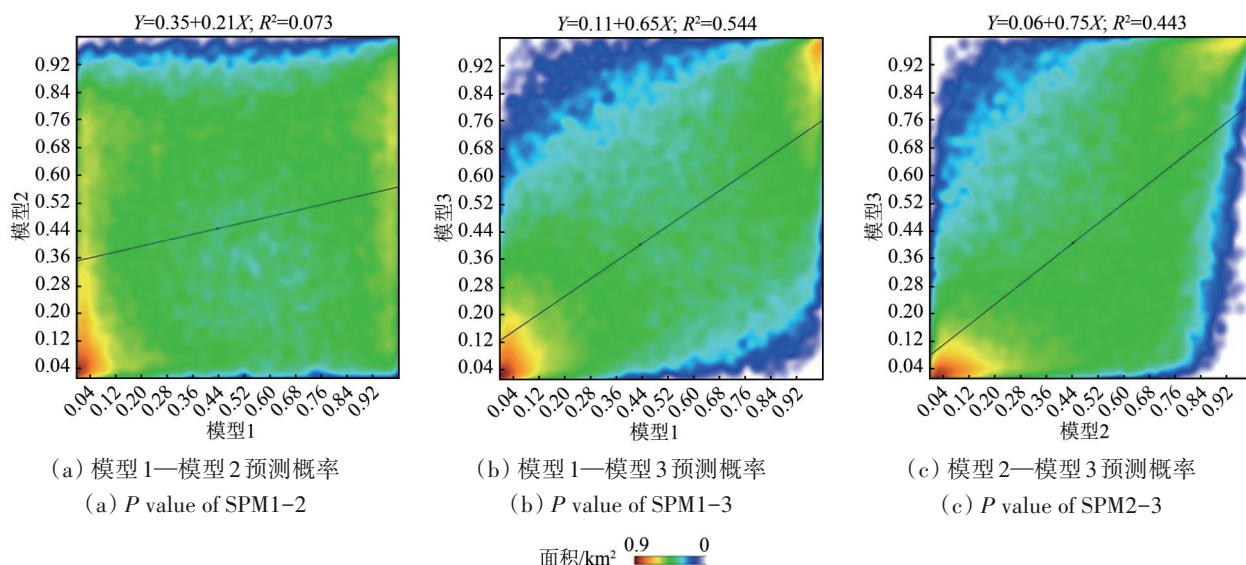


图5 遗址预测模型1、2、3的相关性分析

Fig. 5 The spatial correlations of SPM 1, 2, 3

原因2,本文进一步利用“对关联函数”PCF (Paired Correlation Function) 考察3个预测模型与实际遗址点的空间结构关系。PCF是量子化学中常用的分析方法,用于考察一个多粒子系统中,每个粒子周围的粒子分布状况,它反映了粒子数密度的关联行为。这种研究方法同样适用于普通空间分析,

可以考察不同空间尺度下点的空间分布模式,在考古学聚落形态研究中有较好的应用(Bevan等, 2013)。本文分别将模型1、2、3与调查遗址点进行Cross PCF关联性分析,绘制不同空间距离 r 上的对关联系数 $g(r)$,同时利用蒙特·卡洛随机模拟方法计算显著性水平区间,结果如图6所示。

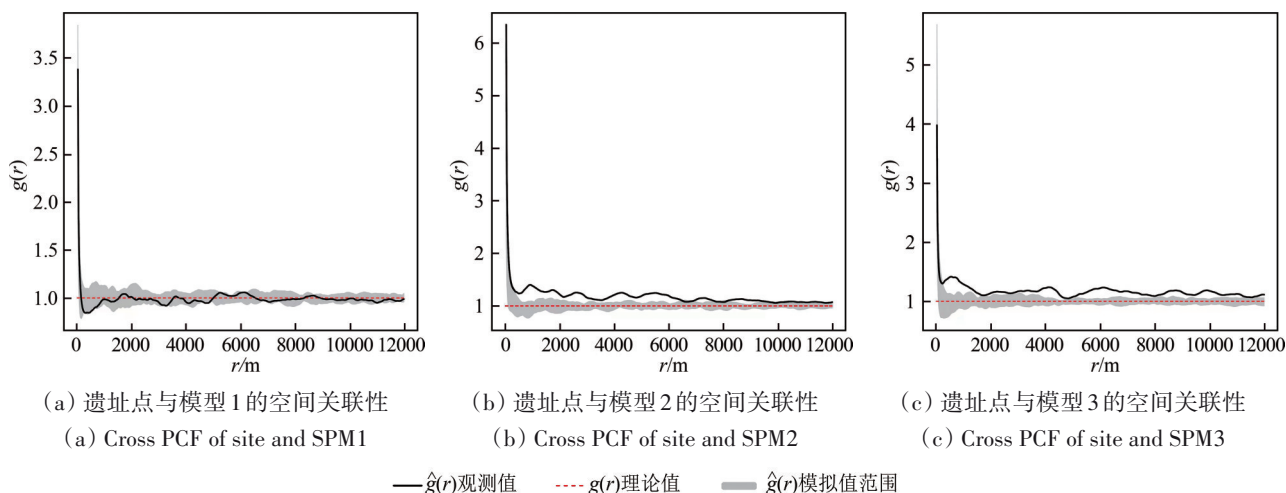


图6 遗址点与预测模型的对关联函数(Cross PCF)分析(蒙特·卡洛随机模拟999次)

Fig. 6 Cross pairwise correlation function (Cross PCF) analysis of sites and SPMs (with 999 Monte Carlo simulations)

模型1与遗址点在各个空间尺度上的相关性均表现为随机状态（未超出蒙特·卡洛模拟的显著性水平区间），因此它们之间不存在关联。相反，模型2、3与遗址点在超出一定的空间尺度上（ $r=380\text{ m}$ ， 250 m ）均存在显著的空间关联特征（超出蒙特·卡洛模拟的显著性水平区间最高值），而且这种空间相关性又具有明显的全局性（global）特点（超出最小空间距离 r 之后，空间相关性始终存在）。这反映出，模型2、3不仅对遗址的预测效率高于模型1，而且更符合遗址点的空间结构特点（点的空间聚合模式），尤其是混合模型3（ $r=250\text{ m}$ ）比模型2（ $r=380\text{ m}$ ）在更小的空间最小尺度即能检测出与考古遗址点的空间相关性，空间结构也更为优化。

上述PCF分析表明，模型2、3之所以较模型1的预测效率更高，主要原因是这两个模型更符合遗址点空间分布的结构特征。模型1是纯粹的地貌因素变量，而模型2则以多光谱遥感数据为因子，更多反映了现代土地利用的特点，即现代人类活动的空间异质性。多光谱遥感数据之所以能更好的表现古代遗址的空间布局特征，主要原因是研究区域内人类活动的延续性——现代人类活动频繁的区域同样也是古代遗址分布密集的地方。综上所述可知，混合模型3将多光谱遥感数据引入基于DEM地貌因素的考古遗址预测模型，可以有效优化考古遗址预测模型的数据空间结构，大大提高模型的预测效率。

4.2 评估可利用土地资源

本文依据上述混合模型3对研究区域先秦时期人类活动状况进行评估，估算可利用土地资源的总量和开发情况。首先，通过对研究区域内遗址间的“最近邻点距离”NND（Nearest Neighbor Distance）进行统计（Boots 和 Getis，1988），可知NND的众数值（mode）为500 m（图7），大致代表了先秦遗址之间的最佳距离模式。如果以此为最小空间度量单位对研究区域2373 km²进行估算，那么最多可随机容纳1715处遗址点。再利用模型3对这1715处备选遗址点进行筛选，计算其中预测概率值大于0.5（模型的最低有效值）的遗址点的数量。最后，重复以上随机过程999次，对最终结果取均值为642，即代表了以遗址预测模型3所计算的研究区域内的最大遗址承载量（即在模型

最低有效预测值内，土地资源的饱和开发程度）。据此可知：仰韶文化时期的聚落相当于只开发了区域总资源的9.8%（≈63/642），龙山文化时期的资源开发率达到43.1%（≈277/642），而整个先秦时期则进一步提升至57.3%（≈368/642）的水平。这个估测数据大体上代表了陇东黄土高原地区先秦时期人口增长与可利用土地资源关系的变化情况（图8）。

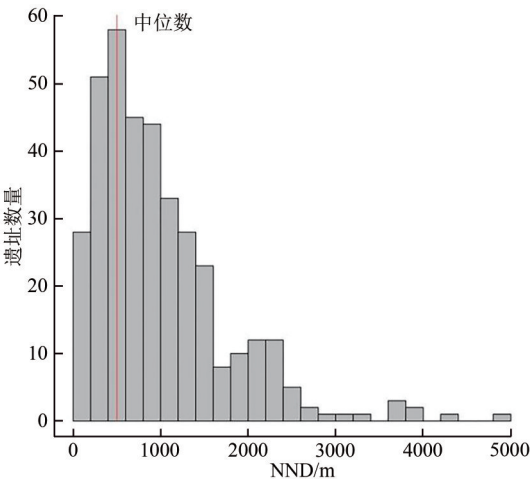


图7 遗址最近邻点距离统计
Fig. 7 Nearest Neighbor Distance of sites

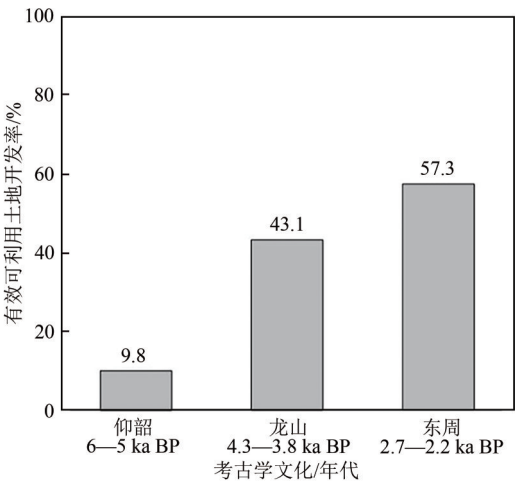


图8 区域土地资源累积开发程度
Fig. 8 Accumulated degree of regional land exploitation

4.3 多光谱遥感与区域考古

本项研究表明，利用多光谱遥感数据构建的遗址预测模型无论从精度还是有效性方面均优于传统的DEM所提供的地貌因子模型。对多光谱数据的主成分分析结果可知，第一主成分因子FC1已经能够解释所有波段80%的信息，且在混合模

型3中属于被剔除的因子,表明多光谱数据本身在很大程度上与地貌信息重合。黄土高原地区土地利用与不同地貌类型高度相关是该区域的一个重要特征,坡地、河谷、塬面在农业生产、居住形态、交通设施等方面分别发挥不同的作用,这种情况自古至今一直延续,也是本项研究中多光谱遥感数据可以有效应用于遗址预测模型的重要原因。

但是,相比单纯的地貌因素而言,多光谱遥感数据能更有效反映出土地利用的状况,尤其是区域土地利用的空间异质性。直观比较模型1、2、3所使用的各类自变量因子,可以看到多光谱因子表现出土地开发集中在黑河、达溪河主河谷和各级黄土塬、梁之上,南部山区则较弱。这种情况与遗址的空间分布是一致的,也表明现代黄土高原地区的土地利用状况是一个长期的历史积累结果,是人类集中开发利用黄土资源的表现,而这种资源的开发模式甚至在先秦时期即已超出了总量的50%以上的规模。

值得注意的是,多光谱的主成分分析中第2、3主成分因子还有一部分变量不能为第1主成分因子所解释,主要表现在第5波段(FC2)和第1、2波段(FC3),表明这3个波段中除了地貌因素之外,还潜在蕴涵着更多与遗址空间分布相关的因素。Landsat 8的OLI数据中:第5波段为调整后的近红外波段($0.845\text{—}0.885\text{ }\mu\text{m}$),相对Landsat 7的NIR近红外波段(第4波段),排除了 $0.825\text{ }\mu\text{m}$ 处水汽吸收特征;第1波段又称为“海蓝波段”或“气溶胶”波段($0.433\text{—}0.453\text{ }\mu\text{m}$)属于新增波段,主要应用于海岸带观测;第2波段,“蓝色波段”($0.45\text{—}0.51\text{ }\mu\text{m}$)相当于Landsat 7的第1波段。相对而言,第5波段的信息量最大,也是第2主成分因子的主要解释部分。曾有研究者通过对新疆阿拉尔市的OLI影像数据的研究,表明从band 1—band 7的各波段的相关系数来看,第1波段的相关性最大,而第5波段与其他波段的相关性最小,可见第5波段不仅包含的地物信息最为丰富,而且这个波段有很大的独立性。通过对所有7个波段中任选3个波段,得到35种组合,计算各种波段组合的OIF值(最佳指数法,即信息量与各波段的均方差之和成正比,与各波段间相关系数之和成反比),其中5-2-1波段组合排序第15位,表明这3个波段组合蕴涵的信息量不容小觑(晁增福

等,2017)。而本项研究也表明,黄土高原地区第5-2-1波段组合对考古学研究具有一定的应用潜力(图9)。

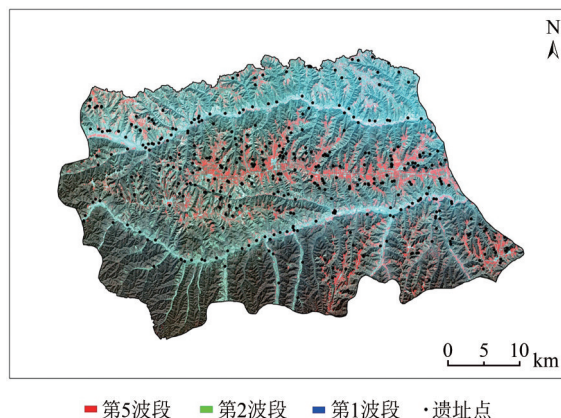


图9 研究区域OLI数据5-2-1波段组合影像

Fig. 9 5-2-1 band combination image of OLI data in the research area

5 结 论

本文以陇东地区区域考古调查记录的数据为基础,从基于DEM和多光谱遥感的基础地貌和土地利用信息中甄选自变量因素,根据Kvamme(1983)提供的方法,利用逻辑回归方程,构建考古遗址预测模型。通过对这两个模型的系统比对,可以发现相比DEM所提供的单纯地貌信息,利用Landsat 8的OLI多光谱遥感数据可以有效地提高考古遗址预测模型的精度和效率,并优化模型的空间结构。在黄土高原地区,由于地貌的差异性大,基于不同地貌的土地利用方式也不同,而多光谱遥感既能有效反映地貌,又能表现人类对不同地貌单元土地利用程度的空间异质性特征,是有效开展区域性考古研究的重要手段。

借助于对混合模型的进一步的研究,我们发现陇东地区人类对黄土地貌资源的开发实际上在先秦时期即已经达到了总体资源的一半以上,尤其是龙山文化时期,聚落数量剧增,人口规模扩大,有效土地资源的开发已经从仰韶文化时期的不足10%而迅速扩大到43.1%。这就为该区域的早期社会复杂化和文明化进程提供了重要的经济基础。这一时期,陇东地区龙山文化的中心聚落中开始出现最早使用瓦的地面式建筑,其性质与同时期的延安芦山峁、榆林石峁相当,与此同时大量的玉器也开始在中心聚落之间流通。至商周时

期, 这里已经发展成与中央王朝关系密切的重要方国。

通过对多光谱遥感数据的主成因素分析并结合不同遗址预测模型的比对, 可以发现 Landsat 8 OLI 数据的 5-2-1 波段的组合能够更好地反映考古遗址的分布状况。尤其是第 5 波段, 不仅包含地物信息丰富, 而且相对其他波段更独立, 在未来黄土高原甚至其他地区的区域考古研究中还值得我们做进一步的深入探索。

参考文献(References)

- Bevan A, Crema E R, Li X Z and Palmisano A. 2013. Intensities, interactions, and uncertainties: some new approaches to archaeological distributions//Bevan A, Lake M, eds. *Computational Approaches to Archaeological Spaces*. Walnut Creek: Left Coast Press: 27-52
- Boots B N and Getis A. 1988. *Point Pattern Analysis*. Newbury Park, CA: SAGE
- Chao Z F, Xing X N and Kang S X. 2017. Study on optional bands combination of OLI image in land use classification. *Hubei Agricultural Sciences*, 56(8): 1567-1569 (晁增福, 邢小宁, 康顺光. 2010. 土地利用分类中 OLI 影像合成最佳波段组合研究. *湖北农业科学*, 56(8): 1567-1569) [DOI: 10.14088/j.cnki.issn0439-8114.2017.08.042]
- Chu S B. 1977. The Western Zhou tombs at Lingtai Baicaoopo, Gansu Province. *Acta Archaeologica Sinica*, 22(2): 99-130 (初仕宾. 1977. 甘肃灵台白草坡西周墓. *考古学报*, 22(2): 99-130)
- Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V and Böhner J. 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7): 1991-2007 [DOI: 10.5194/gmd-8-1991-2015]
- Custer J F, Eveleigh T, Klemas V and Wells I. 1986. Application of landsat data and synoptic remote sensing to predictive models for prehistoric archaeological sites: an example from the delaware coastal plain. *American Antiquity*, 51(3): 572-588 [DOI: 10.2307/281753]
- Dong Z and Jin S Z. 2015. Prediction research on Bohai Kingdom ruins in Yanbian area based on the logic regression model. *Journal of Yanbian University (Natural Science)*, 41(2): 179-184 (董振, 金石柱. 2015. 基于 Logistic 回归模型的延边地区渤海国遗址预测研究. *延边大学学报(自然科学版)*, 41(2): 179-184) [DOI: 10.3969/j.issn.1004-4353.2015.02.018]
- Green E L. 1973. Location analysis of prehistoric Maya sites in northern British Honduras. *American Antiquity*, 38(3): 279-293 [DOI: 10.2307/279715]
- Guo F. 2018. *Study of Archaeology Sites Predictive Distribution based on Logistic Regression Optimization Method—A Case Study of Fenhe River Basin*. Beijing: Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (郭飞. 2018. 基于 Logistic 回归最优化方法的遗址预测分布研究——以汾河流域为例. 北京: 中国科学院遥感与数字地球研究所)
- Kvamme K L. 1983. Computer processing techniques for regional modeling of archaeological site locations. *Advances in Computer Archaeology*, 1(1): 26-52
- Li S C and Li F. 1992. The archaeological survey in the Hulu River area in Gansu Province. *Archaeology*, (11): 971-986 (李水城, 李非. 1992. 甘肃省葫芦河流域考古调查. *考古*, (11): 971-986)
- Liu Z W. 1980. Akanke Information Criterion (AIC) and its significance. *Mathematics in Practice and Theory*, (3): 64-72 (刘璋温. 1980. 赤池信息量准则 AIC 及其意义. *数学的实践与认识*, (3): 64-72)
- Mei Q B and Wang C B. 2004. The research of site predictive models and GIS in archaeology//Collection of Thesis of the Second Symposium on Digitization and Protection of Chinese Cultural Heritage. Hangzhou: [s.n.]: 157-163 (梅启斌, 汪诚波. 2004. 考古 GIS 和遗址预测模型研究//第二届中华文化遗产数字化及保护研讨会. 杭州: [s.n.]: 157-163)
- Mo D W, Li F, Li S C and Kong Z C. 1996. A preliminary study on the paleoenvironment of the middle Holocene in the Hulu River area in Gansu Province and its effects on human activity. *Acta Geographica Sinica*, 51(1): 59-69 (莫多闻, 李非, 李水城, 孔昭宸. 1996. 甘肃葫芦河流域中全新世环境演化及其对人类活动的影响. *地理学报*, 51(1): 59-69) [DOI: 10.11821/xb199601007]
- Ni J S. 2009. Predictive model of archaeological sites in the upper reaches of the Shuhe River in Shandong. *Progress in Geography*, 28(4): 489-493 (倪金生. 2009. 山东沭河上游流域考古遗址预测模型. *地理科学进展*, 28(4): 489-493) [DOI: 10.11820/dlxxjz.2009.04.002]
- Peng S Z, Zhang W and Chen D D. 2010. Model predictability of archaeological sites of the Dawenkou Culture in the Wensi River Basin. *Journal of Taishan University*, 32(6): 34-39 (彭淑贞, 张伟, 陈栋栋. 2010. 汶泗流域大汶口文化考古遗址模型预测. *泰山学院学报*, 32(6): 34-39) [DOI: 10.3969/j.issn.1672-2590.2010.06.008]
- Qiao W W, Bi S B, Wang Q F and Guo Y. 2013. Predictive model of archaeological sites of Longshan Culture in Zhengzhou-Luoyang area. *Science of Surveying and Mapping*, 38(6): 172-174, 181 (乔文文, 毕硕本, 王启富, 郭忆. 2015. 郑洛地区龙山文化遗址预测模型. *测绘科学*, 38(6): 172-174, 181) [DOI: 10.16251/j.cnki.1009-2307.2013.06.023]
- The Archaeology Team of Gansu Museum. 1980. The trial excavation of Qiaocun site during Qijia Culture period in Lingtai County, Gansu Province. *Archaeology and Cultural Relics*, 1(3): 22-24 (甘肃省博物馆考古队. 1980. 甘肃灵台桥村齐家文化遗址试掘简报. *考古与文物*, 1(3): 22-24)
- Wei H H and Wu D X. 1972. A Western Zhou tomb at Lingtai Baicaoopo. *Cultural Relics*, (12): 2-8 (魏怀珩, 伍德熙. 1972. 灵台白草坡西周墓. *文物*, (12): 2-8)
- Yu Y Y, Wu H B and Guo Z T. 2010. A new simulation model for prehistoric land use and carbon storage (PLCM)—an application in Yiluo Valleys. *Quaternary Sciences*, 30(3): 540-549 (于严严, 吴海斌, 郭正堂. 2010. 史前土地利用碳循环模型构建及应用——

以伊洛河流域为例. 第四纪研究, 30(3): 540-549 [DOI: 10.3969/j.issn.1001-7410.2010.03.12]
Zhang H, Bevan A and Guo D S. 2013. The Neolithic ceremonial com-

plex at Niuheliang and wider Hongshan landscapes in northeast-
ern China. Journal of World Prehistory, 26(1): 1-24 [DOI: 10.1007/s10963-013-9062-9]

Archaeological site predictive models based on DEM and multi-spectral remote sensing data on the pre-Qin sites in the Loess Plateau of East Gansu Province

ZHANG Hai¹, XU Yijing¹, ZHOU Jing²

1. School of Archaeology and Museology, Peking University, Beijing 100871, China;

2. Provincial Institute of Cultural Heritage and Archaeology, Gansu Province, Lanzhou 730000, China

Abstract: This paper aims to explore a new method on the traditional archaeological Site Predictive Model (SPM) by integrating Multi-Spectral Remote Sensing (MSRS) resources. The application on the case study from the Loess Plateau of east Gansu Province shows great prospect of the combination of SPM and MSRS on the regional archaeological studies and cultural resource managements.

The SPM is a useful statistic model in archaeological research and cultural heritage management practices. Previous archaeological applications of the SPMs relied heavily on the Digital Elevation Model (DEM) while overlooked the MSRS. This research attempts to integrate the MSRS, mainly the Landsat 8 OLI data, into the SPMs to increase the model's predicting efficiency and to acquire high-quality predictive models, which can be used to evaluate the ancient land use degree in statistics.

This research compares three SPMs that were based on the DEM-derived terrain factors, the factors from MSRS, and the mixed factors. The terrain factors include the following. First, the landform indices, such as slope, aspect, curvature, texture, and convexity. Second, the hydrological indices, such as channel network base level, valley depth, and vertical distance to rivers. Third, the topographic position indices, such as relative slope, topographic openness, and wind exposition. The MSRS factors are derived by the reduction of the PCA method on Landsat 8 OLI's band 1—7. The mixed factors are the combination of terrain and MSRS factors totally. Archaeological data dated to the pre-Qin period from the systematic archaeological survey in east Gansu Province are incorporated into the three SPMs by using the Logistic Regression provided by Kvamme (1983) and AIC gradient optimization. Three assessment indices, including Goodness-of-Fit R^2 , AIC and, G value are calculated to test the three models on their predictive effects. Two spatial analysis methods, including cross raster statistics and cross PCF on sites vs. models are incorporated to evaluate the models on their spatial structures.

The comparison of the three models indicates that the MSRS can significantly improve the validity of the modular prediction with the greater R^2 (model1=0.265, model2=0.312, model3=0.414), G value (model1=0.51, model2=0.61, model3=0.64) and smaller AIC (model1=444.6, model2=420.0, model3=379.1). Meanwhile, models that combined MSRS data can achieve a better spatial structure than the DEM. Further analysis based on the mixed SPM (model3) indicates that the total land-use areas in the eastern part of the Loess Plateau increased from less than 10% during the Yangshao period to up to 43.1% during the Longshan period. The latter laid a solid foundation for the development of social complexity in the Pre-Qin period. The analysis on the MSRS-based model also reveals that the combination of bands 5-2-1 in OLI can provide more information on ancient sites.

The land-use of Loess Plateau in east Gansu Province has demonstrated strong continuity since the prehistoric period. Thus, the modern multi-spectral remote sensing data can be well utilized into traditional archaeological site predictive model not least to improve its spatial structure and raise the model's efficiency. Landsat 8 OLI can provide effective multi-spectral data for site identification and land-use classification. The band 5 of OLI potentially contains information on archaeological sites, which must be studied further.

Key words: remote sensing, Loess Plateau in East Gansu Province, archaeological site predictive model, Landsat 8 OLI, Land-use in Pre-Qin period

Supported by National Key Research and Development Program of China (No.2020YFC1521900)