

DOI: 10.20103/j.stxb.202403310682

刘瑞雪, 李佳轩, 李云. 水库消落带植物多样性空间格局预测模型及环境解释——基于 XGBoost-SHAP 模型框架. 生态学报, 2024, 44(21): 9652-9669.

Liu R X, Li J X, Li Y. Prediction model and environmental interpretation for the spatial pattern of plant diversity in the water-level fluctuation zone of reservoir based on XGBoost and SHAP. Acta Ecologica Sinica, 2024, 44(21): 9652-9669.

水库消落带植物多样性空间格局预测模型及环境解释 ——基于 XGBoost-SHAP 模型框架

刘瑞雪, 李佳轩, 李 云*

深圳大学建筑与城市规划学院, 深圳 518060

摘要:生物多样性的监测与预测对实现生物多样性保护及其可持续管理至关重要。传统方法通过实地调查来构建环境与生物多样性之间的多变量关系模型。空间大数据技术及机器和深度学习算法的发展为探索环境-生物多样性关系和预测生物多样性空间格局提供了新的视角和方法。构建了一种基于 XGBoost 算法的预测模型, 融合实地调查的植物多样性数据和来自多源数据库的环境变量数据, 分别构建了气候、地形、土壤、水文和人类活动 5 类共 34 个环境变量与植物群落物种丰富度、物种多样性和谱系多样性的关系模型, 对丹江口水库消落带的植物多样性空间格局进行预测, 同时结合 SHAP 框架确定关键环境因素; 并进一步预测 2050 年水库消落带的植物多样性空间格局。研究表明, XGBoost 算法在预测水库消落带植物多样性方面表现较好, 3 个多样性指标中谱系多样性的预测模型展现了最优的预测能力, 而物种多样性预测模型的预测能力相对较低。结合 SHAP 分析发现年平均水淹时长、人类足迹与最冷季平均气温是影响消落带植物群落物种丰富度、物种多样性和谱系多样性的关键环境因素, 其中年平均水淹时长的影响最为显著, 随着年平均水淹时长增加, 物种丰富度、物种多样性和谱系多样性降低。本研究构建的可解释预测模型可有效揭示消落带的植物多样性空间格局, 为消落带生物多样性的保护和可持续管理提供科学依据, 为生物多样性的监测和管理提供了新方法, 对评估全球变化对生态系统的影响并促进生物多样性保护有重要意义。

关键词:植物多样性; 空间格局; 预测模型; 机器学习; 消落带

Prediction model and environmental interpretation for the spatial pattern of plant diversity in the water-level fluctuation zone of reservoir based on XGBoost and SHAP

LIU Ruixue, LI Jiaxuan, LI Yun*

School of Architecture & Urban Planning, Shenzhen University, Shenzhen 518060, China

Abstract: Monitoring and predicting biodiversity is crucial for achieving biodiversity conservation and sustainable management. Traditional methods construct multivariate relationship models between the environment and biodiversity through field investigations. The development of spatial big data technologies, along with machine and deep learning algorithms, provides new perspectives and methods for exploring the environment-biodiversity relationship and predicting spatial patterns of biodiversity. In this study, we constructed prediction models based on the XGBoost algorithm, integrating plant diversity data from field surveys and environmental variable data from multiple sources databases. We developed models to predict the spatial pattern of plant diversity in the water-level fluctuation zone of Danjiangkou Reservoir by examining relationships between 34 environmental variables, including climate, topography, soil, hydrology, human

基金项目:深圳市科技计划面上项目(20220807102319001); 2023 年中国工程院战略研究与咨询项目(2023-HZ-03)

收稿日期:2024-03-31; **采用日期:**2024-10-05

* 通讯作者 Corresponding author. E-mail: li_yun@szu.edu.cn

activities, as well as species richness, species diversity, and phylogenetic diversity of plant communities. Additionally, we identified key environmental factors using the SHAP framework. Furthermore, we predicted the future spatial pattern of plant diversity in the water-level fluctuation zone of Danjiangkou Reservoir for 2050. The results showed that the XGBoost algorithm performed well in predicting plant diversity in the water-level fluctuation zones with better predictive ability for the phylogenetic diversity model compared to the species diversity model which had relatively lower predictive performance. Through SHAP analysis, we found that annual average flooding duration, human footprint, and mean temperature during the coldest season significantly influenced species richness, species diversity, and phylogenetic diversity of plant communities in the water-level fluctuation zone. The average annual duration of flooding had the most significant impact, with species richness, species diversity, and phylogenetic diversity decreasing as the average annual flooding duration increased. The interpretable prediction model constructed in this study effectively revealed spatial patterns of plant diversity in the water-level fluctuation zones, providing scientific evidence for conservation and sustainable management of biodiversity in these areas. It also offers a new method for biodiversity monitoring and management, which is of great significance for assessing the impacts of global changes on ecosystems and promoting biodiversity conservation.

Key Words: plant diversity; spatial pattern; prediction model; machine learning; water-level fluctuation zone

生物多样性是维持陆地生态系统功能的基础,在提供重要生态系统服务方面发挥着重要作用^[1-2]。掌握生物多样性的空间格局对其保护和管理至关重要^[3]。绘制生物多样性空间格局具有悠久且丰富的历史。从19世纪开始学者们就有计划、有目的性的对物种数据等资料进行记录和整理^[3]。在过去几十年中,尽管科学家们建立了全球生物多样性信息网络等数据共享平台,在绘制大尺度植物多样性空间格局方面已取得一定进展^[4]。但现存物种分布数据的资料受到如自然条件、采样地点可达性、物种被观察到的概率等多方面限制,绝大多数研究都以粗空间分辨率来评估物种多样性空间格局,多使用专家绘制的范围图^[5]和区域植物清单^[4]的数据,难以支撑生物多样性的精准研究^[6]。较小尺度上的生物多样性信息可通过实地调查获得,但这些信息无法直接扩展到较大的空间尺度^[3]。从那时起,学者们通过构建基于环境变量的生物多样性模型对生物多样性空间格局进行精细化和缩放^[7-8]。这些生物多样性空间格局预测方法的准确性取决于已有的生物多样性数据、环境变量以及预测模型的质量和代表性^[3]。尽管对生物多样性空间格局已有长期且大量的研究,但生物多样性在精细空间分辨率下的大尺度空间格局的相关研究仍然缺乏^[9]。

近年来学者们通过构建模型来估测大尺度区域的植物多样性空间格局^[10]。这些模型通过模拟环境变量与植物多样性的关系对植物多样性空间格局进行调整、精确和缩放,使得可以在一定区域内进行植物多样性空间格局的连续预测^[11]。植物多样性与环境变量的关系往往是复杂、非线性的和具有尺度依赖性的^[8],而且许多环境变量相互作用并表现出较高的共线性,对传统生态统计模型如广义线性模型、典范对应分析模型和非度量多维尺度分析模型等提出了重大挑战^[3]。随着科技发展,数据可用性和模型技术的发展使植物多样性预测模型达到前所未有的分辨率和准确性^[3]。近年来发展起来的机器与深度学习算法是一种强大的建模工具,可有效处理多维数据并且在没有先验规范的情况下揭示变量之间的非线性关系^[12]。因此,机器与深度学习算法已逐渐成为生态学中传统统计方法的替代方案^[7,13],不仅能够有效提高估测准确度和效率,而且同时包含了生态学、地理信息系统、计算机科学的跨学科技术创新,为生物多样性空间格局的研究提供了空间大数据发展背景下的全新视角和解决方案。Večeřa 等人展示了机器学习方法(随机森林模型)在使用来自植被样地数据库的 α 多样性数据进行训练时基于气候和其他环境因子预测固定大小植被样地多样性的潜力^[11]。Chang 比较分析了多种机器学习模型(决策树、随机森林、支持向量机)和深度神经网络模型在估测生物多样性方面的能力,并基于环境因素预测了以色列的生物多样性空间格局^[14]。Cai 等使用机器学习方法(随机森林、极端梯度增强和神经网络模型)和传统统计方法(广义线性模型和广义加性模型)来预测环境驱动的大尺度植物多样性空间格局,并在全球范围内预测植物多样性^[3]。极端梯度增强(Extreme Gradient Boosting,

XGBoost)模型是对梯度提升决策树(Gradient Boosting Decision Tree,GBDT)模型的高效实现与改进,加入正则化规则以减少过拟合的风险,在算法效率及计算精准度方面有明显提升,表现出更精确的预测结果和更高的处理效率^[15],目前已在多个领域得到广泛应用并成果颇丰^[16-17]。尽管机器学习方法相比传统统计方法有优秀的预测能力,但因缺乏可解释性严重限制了其应用,通常被认为是黑箱模型^[18]。基于博弈论和局部解释理论的SHAP(SHapley Additive exPlanations)属于经典的解释框架,可提供Shapley值来估计每个特征变量的贡献^[16-17]。因此,普遍引入SHAP框架对机器学习模型的结果进行解释^[18]。

在已有的植物多样性空间格局预测的研究中,学者们将研究对象集中在热带、亚热带和温带森林等物种丰富、生境条件优越的植物群落上^[3,11,14],尚未有学者关注生态系统脆弱、生境条件复杂且恶劣的植物群落。消落带是由于水位涨落而使水陆交界地带的土地周期性淹没或出露于水面的区域,包括最低水位线和最高水位线之间的区域^[19-20]。消落带作为生态交错带,具有高度生境异质性,为众多物种提供栖息地,是生物多样性热点区域。同时提供如保护水质、稳定堤岸、水土保持、污染阻截和保护水域生态安全等丰富的生态系统服务功能。由于受到多种因素的影响如气候、地形、土壤、河道形态、水文及人类活动等,消落带已成为退化最严重的生态系统之一。因此掌握消落带植物多样性空间格局对评估全球变化对生态系统的影响并支持消落带的生物多样性保护有重要作用。丹江口水库位于长江最大支流汉江和最长支流丹江的交汇处,是我国南水北调中线工程的水源地和亚洲最大的人工湖,其生态安全直接关系到我国华北地区的用水安全^[21]。对丹江口水库消落带植物多样性空间格局进行研究,有助于掌握水库消落带植物多样性空间分布规律,可为水库消落带生物多样性保护提供科学参考,有助于维护水源地的生态安全稳定。前期研究已证实,自1974年丹江口水利枢纽一期工程建成到2014年12月南水北调中线工程正式运行之前,消落带植物群落经过40年的自然演替已达到稳定状态^[22]。综上,本研究以丹江口水库消落带为研究区域,基于2013年消落带植物多样性的实地调查数据和气候、土壤、地形、水文和人类活动等环境特征,使用XGBoost算法构建多源环境变量与消落带植物多样性的关系模型,对消落带植物多样性空间格局进行预测,并结合SHAP框架确定关键环境因素;然后进一步预测南水北调中线工程正式运行后未来情景下(2050年)的消落带植物多样性空间格局,以期为消落带生物多样性保护和可持续生态管理提供科学依据。研究流程图如图1。

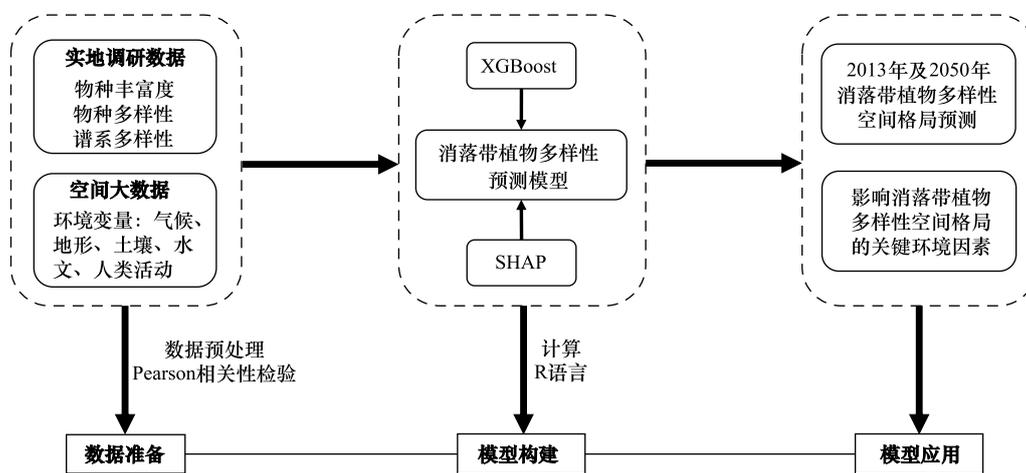


图1 研究流程图

Fig.1 Datasets, metric, and mythological steps used in this study

1 研究方法

1.1 研究区域概况

丹江口水库(32°36'—33°48'N, 110°59'—111°49'E)地处亚热带温暖半湿润季风区(图2),年均温13.7℃,年均降水量873.3mm,年均蒸发量854mm,降雨主要集中在5—10月。地带性植被主要为亚热带落

阔叶林以及部分常绿阔叶林和针阔混交林。水库库岸具有沟谷发育、岸线曲折、地形破碎的特点,库湾数量多,呈现“鸡爪”状。库岸分为石质和土质两种基质类型。石质基质土壤瘠薄,植被稀疏;土质基质土壤条件较好,植被生长相对良好,以草本植物为主^[22]。自 1974 年丹江口水利枢纽一期工程建成到 2014 年 12 月南水北调中线工程运行之前,根据长江水资源保护科学研究所提供的水库水位数据可知,每年 4 月出现最低水位,随后水位逐渐升高,每年 10 月达到正常蓄水位 157 m,库容为 174.5 亿 m³。2014 年 12 月南水北调中线工程运行后,原有的水库消落带被淹没,在更高海拔上形成新的消落带。根据水利部《丹江口水利枢纽调度规程(试行)》等运行调度文件,水库每年 5 月降低到夏季汛期限水位 160 m,10 月后蓄水到正常蓄水位 170 m,库容会达到 339.5 亿 m³^[21]。高差达 10 m 的消落带,使水域生态安全和生物多样性保护面临巨大挑战。

1.2 研究方法

1.2.1 数据获取及处

(1) 样地调查与植物多样性指数计算

于 2013 年 5 至 7 月对丹江口水库消落带植物群落进行实地调查。根据库区地形图及高分辨率遥感影像,综合考虑植被生长情况及可达性,从水库水面以上有植物出现的地方起至海拔 157 m(正常蓄水位)的范围内,选择植物群落进行样方调查,最终调查了 205 个植物群落(图 2)。乔木群落设置 20 m×20 m 的样方;灌木群落设置 5 m×5 m 的样方;草本植物群落设置 2 m×2 m 的样方。记录样方中所有维管植物,测定每个物种的平均高度、盖度和多度。植物种类鉴定依据《中国植物志》和《湖北植物志》。使用手持式 GPS 仪对样方进行空间定位,记录经纬度和海拔。

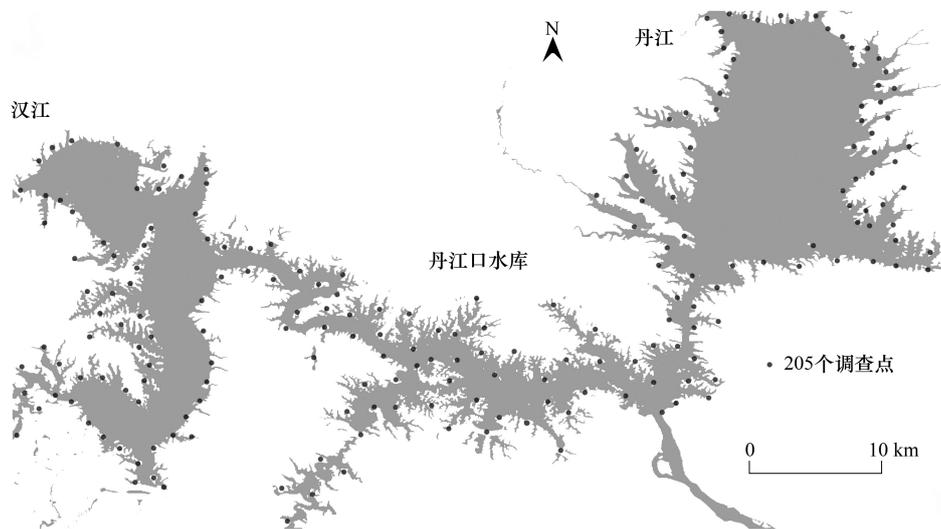


图 2 研究区域及实地调查点分布图

Fig.2 Distribution of 205 sampling sites in the water-level fluctuation zone of Danjiangkou Reservoir

使用 Patrick 丰富度指数 R 来表征植物群落物种丰富度,使用 Shannon-Wiener 多样性指数 H 来表征植物群落物种多样性,计算公式如下^[21]:

$$\text{Patrick 丰富度: } R = S$$

$$\text{Shannon-Wiener 多样性指数: } H = - \sum_{i=1}^S (P_i \ln P_i)$$

其中, S 是样方内所有植物种类的数量。 P_i 为样方内第 i 个种的重要值,即 $P_i = IV_i$, IV_i 是样方内物种 i 的重要值, $IV = (\text{相对高度} + \text{相对盖度} + \text{相对多度}) / 3$ 。

使用样方调查收集到的物种信息在国际植物清单数据库 (<http://www.theplantlist.org/>) 中检索并创建物种目录。样地调查共记录植物种类 183 种属于 51 科 143 属,其中 13 种植物在国际植物清单数据库无法检索

到。因此先根据 Zanne 等人构建的物种进化关系在 *phyloomatic* (<http://phylodiversity.net/phyloomatic/>) 平台将 170 种植物生成具有进化枝长的谱系树,再使用 R 语言的 *V.Phylomaker2* 包将剩余的 13 种植物绑定到各自的科属中^[3],谱系树构建结果如图 3。使用 Faith 提出的 PD 指数来计算植物群落谱系多样性,该指数表示群落中物种在谱系树上进化枝长度总和,使用 R 语言 *picante* 包的 *pd* 函数进行计算^[21]。

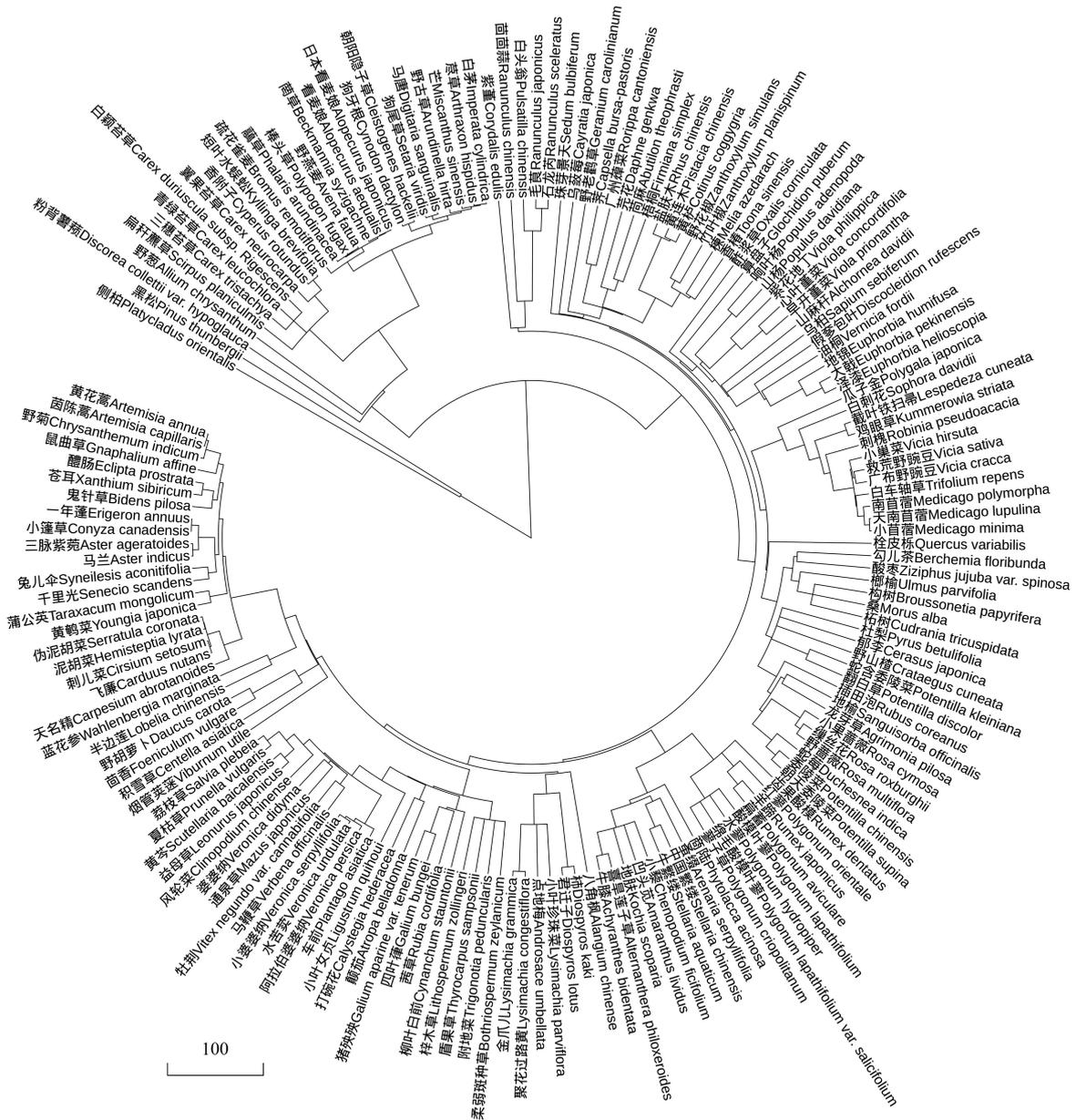


图 3 丹江口水库消落带植物群落物种谱系树

Fig.3 Species phylogenetic tree of plant communities in the water-level fluctuation zone of Danjiangkou Reservoir

(2) 环境因素

本研究建立了包括气候、土壤、地形、水文和人类活动共 5 类 34 个环境变量的数据集(表 1)。气候变量来自世界气候数据库 Worldclim (<https://www.worldclim.org>),选取当前(1970—2000 年)和未来(2040—2060 年)情景的 19 个气候变量,空间分辨率为 1 km。其中 2050 年的气候情景选择 IPCC 第 6 次报告中共享社会经济路径的 SSP245 情景的气候数据。SSP245 代表未来的社会、经济和技术因素与历史趋势并没有明显的变化,是一种类似当前情景下的未来发展情景^[23]。基于高程数据对所有气压下的温度进行线性回归得到递减

率,从而对温度相关的气候因子进行插值降尺度^[24-25],同时结合高程数据使用地理加权回归对降水相关的气候因子进行插值降尺度^[26],从而将 19 个气候变量的空间分辨率统一到 500 m;土壤变量数据来自国家地球系统科学数据中心(<http://www.geodata.cn>)的 2012 年 9 个土壤变量,空间分辨率为 250 m;地形数据来自地理空间数据云(<http://www.gscloud.cn>)发布的 DEM 高程数据,空间分辨率为 30 m,并通过 ArcGIS 空间分析工具和 3D 分析工具提取坡度和坡向;人类活动变量来自开放科学中心(<https://osf.io/5ux86/>)发布的人类足迹指数(2000—2019 年),空间分辨率为 300 m。利用 ArcGIS 对上述所有环境变量统一投影坐标系,并统一空间分辨率为 500 m。水文因素包括年平均水淹时间和水淹频率。根据长江水资源保护科学研究所提供的 1979 年 1 月至 2009 年 12 月的水库水位数据记录和海拔数据,划定 2013 年的消落带空间范围为海拔 140 m 至 157 m;根据水利部的《丹江口水利枢纽调度规程(试行)》等运行调度文件,划定 2050 年的消落带空间范围为海拔 160 m 至 170 m。并计算年平均水淹时间和水淹频率。假定从 2013 年到 2050 年地形、土壤和人类活动变量保持不变。

为降低环境变量之间的共线性对模型预测结果的影响,对 2013 年 34 个环境变量进行 Pearson 相关性检验(图 4),去除相关性较高的影响因子(相关系数绝对值大于 0.8),保留相关性较低的环境变量参与预测模型构建。

表 1 环境变量数据集

Table 1 Environmental variables

类型 Types	变量含义 Meanings of variables	变量名称 Variables names	单位 Unit	类型 Types	变量含义 Meanings of variables	变量名称 Variables names	单位 Unit
生物气候 Bioclimate	年平均气温	BIO1	℃		最暖季降水量	BIO18	mm
	昼夜温差日均值	BIO2	℃		最冷季降水量	BIO19	mm
	等温性	BIO3	-	土壤 Soil	土壤有机碳含量	MS1	(g/kg)
	温度季节性	BIO4	-		土壤质地	MS2	-
	最暖月最高温	BIO5	℃		土壤粘土含量	MS3	(g/kg)
	最冷月最低温度	BIO6	-		土壤淤泥含量	MS4	(g/kg)
	气温年较差	BIO7	℃		土壤阳离子交换能力	MS5	(cmol/kg)
	最湿季平均温度	BIO8	mm		土壤沙含量	MS6	(g/kg)
	最干季平均气温	BIO9	℃		土层厚度	MS7	cm
	最暖季平均气温	BIO10	℃		土壤容重	MS8	(g/dm ³)
	最冷季平均气温	BIO11	℃		土壤 pH 值	MS9	-
	年均降水量	BIO12	mm	地形 Topography	海拔	DEM	m
	最湿月降水量	BIO13	mm		坡向	ASP	(°)
	最干月降水	BIO14	mm		坡度	SA	(°)
	降水季节性变异系数	BIO15	-	水文 Hydrology	年平均水淹时长	FD	(d/a)
	最湿季降水量	BIO16	mm		水淹频率	FF	(次/a)
	最干季降水量	BIO17	mm	人类活动 Human activity	人类足迹	HF	-

BIO1:年平均气温 Annual mean temperature; BIO2:昼夜温差日均值 Mean diurnal range; BIO3:等温性 Isothermality; BIO4:温度季节性 Temperature seasonality; BIO5:最暖月最高温度 Max temperature of warmest month; BIO6:最冷月最低温度 Min temperature of coldest month; BIO7:气温年较差 Temperature annual range; BIO8:最湿季平均温度 Mean temperature of wettest quartre; BIO9:最干季平均气温 Mean temperature of driest quarter; BIO10:最暖季平均气温 Mean temperature of warmest quarter; BIO11:最冷季平均气温 Mean temperature of coldest quarter; BIO12:年均降水量 Annual precipitation; BIO13:最湿月降水量 Precipitation of wettest month; BIO14:最干月降水量 Precipitation of driest month; BIO15:降水季节性变异系数 Precipitation seasonality (coefficient of variation); BIO16:最湿季降水量 Precipitation of wettest quarter; BIO17:最干季降水量 Precipitation of driest quarter; BIO18:最暖季降水量 Precipitation of warmest quarter; BIO19:最冷季降水量 Precipitation of coldest quarter; MS1:土壤有机碳含量 Soil organic carbon fraction; MS2:土壤质地 Soil texture; MS3:土壤粘土含量 Soil clay fraction; MS4:土壤淤泥含量 Soil silt fraction; MS5:土壤阳离子交换能力 Soil cation exchange capacity; MS6:土壤沙含量 Soil sand fraction; MS7:土层厚度 Soil thickness; MS8:土壤容重 Soil bulk density; MS9:土壤 pH 值 pH; DEM:海拔 Elevation; ASP:坡向 Slope aspect; SA:坡度 Slope; FD:年平均水淹时长 Flood duration; FF:水淹

频率 Flood frequency; HF: 人类足迹 Human footprint

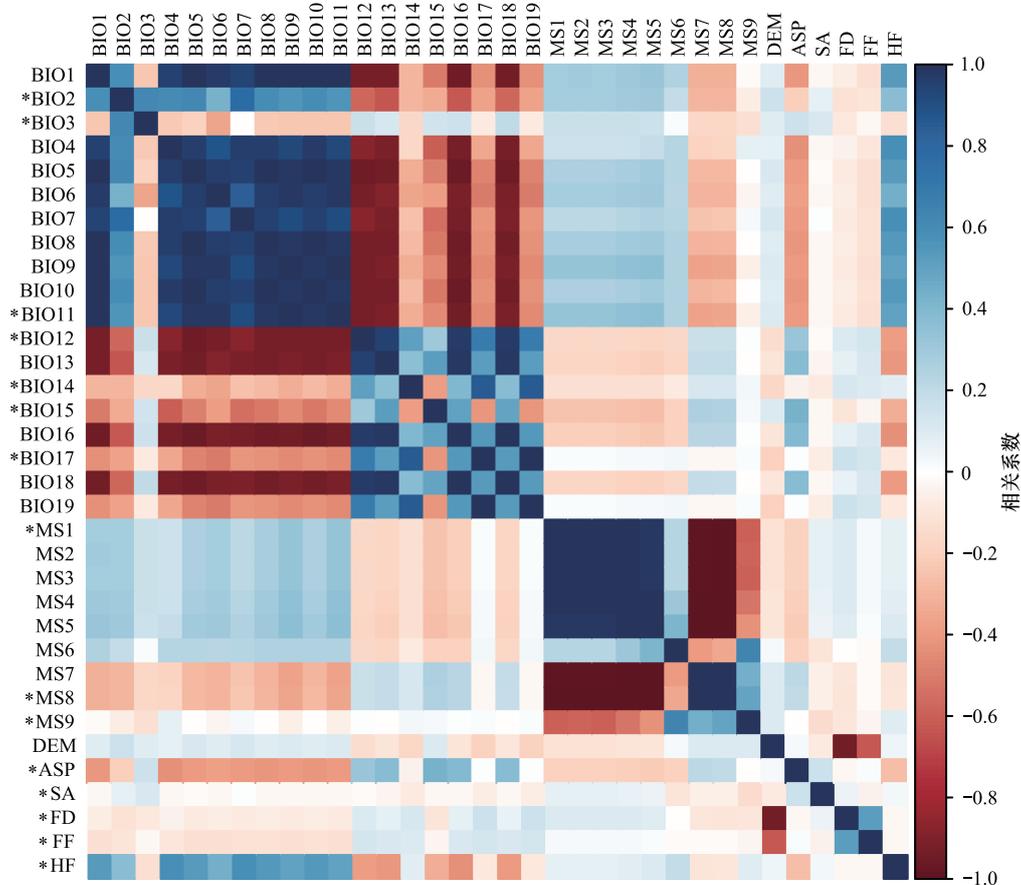


图 4 2013 年环境变量的相关性检验

Fig.4 Pearson correlation test for environmental variables in 2013

*: 经过 Pearson 相关性检验后保留的参与预测模型的环境变量

1.2.2 构建环境变量与植物多样性指标的预测模型

使用 XGBoost 构建多源环境变量与消落带植物多样性的非线性关系,对水库消落带植物多样性进行预测,同时引入 SHAP 归因分析对模型结果进行解释。

(1) XGBoost-SHAP 模型原理及步骤

XGBoost 算法本质上是一个集成算法,通过残差拟合来建立多个评估器,并不断累加最终生成强评估器,每一轮迭代生成的评估器都在上一轮的基础上训练生成,这种训练模式下最终可使模型的预测结果最大程度上接近实际值^[18]。尽管 XGBoost 有更高的预测精度、更快的计算速度和更强的预测能力,但其可解释性低,通常被认为是黑箱模型。统计 XGBoost 模型的特征重要性值只能说明哪个环境变量重要,无法衡量各个特征与预测结果之间的关系。为解决模型解释性差的问题,引入 SHAP 对模型结果进行解释^[16]。由于 SHAP 归因分析采用了博弈论中的 shapley 值作为解释度量,具有对变量全局和局部可解释性强、变量贡献度公平分配及可视化效果出色等优点,弥补了黑箱模型可解释性差的缺陷^[16-17]。因此引入 SHAP 通过 Shapley value 对单一变量与因变量的非线性关系进行解释分析,对各个环境变量进行贡献评估^[27]。分别使用 R 语言 xgboost 包的 xgb 函数构建模型^[28],shapviz 包 shapviz 函数进行 SHAP 值计算,并通过 shapviz 包的 sv_importance 函数对变量进行重要性排序^[29]。

使用 XGBoost 算法构建植物多样性预测模型,设特征数据集为 X (包含气候、土壤、地形、水文和人类活动

等环境变量),对应的植物多样性为 Y , 给定 n 个样本(表示 n 个植物群落), m 个特征数目, 建立训练数据集 $T = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$, 其中 $x_i = \{x_1(1), x_2(2), \dots, x_n(m)\}$, 将训练集输入 XGBoost 模型进行训练。XGBoost 模型利用前向分布算法学习到包含 k 个基评估器的加法模型, 假设 f_k 为第 k 个基评估器, y'_i 代表样地 x_i 对应的植物多样性预测值, 有计算公式如下:

$$y'_i = \sum_{t=1}^k f_t(x_i)$$

此时目标函数为:

$$L = \sum_{i=1}^n l(y_i, y'_i) + \sum_{t=1}^k \Omega(f_t)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|$$

其中: l 为损失函数, 表示预测值与真实值之间的误差; Ω 为正则化函数, 防止模型过拟合。 T 表示每棵树的叶子节点数; ω 表示每棵树叶子的权重; 为了抑制树的生长和防止模型过拟合, 加入 γ 和 λ , λ 为正则化系数, γ 为分裂阈值。经过第 t 轮迭代, 样地 x_i 的植物多样性预测值 y'_i 有表达式:

$$y'_i = y'^{t-1}_i + f_t(x_i)$$

此时目标函数 L 则变化为:

$$L = \sum_{i=1}^n l(y_i, y'^{t-1}_i + f_t(x_i)) + \Omega(f_t)$$

对目标函数 L 进行二阶泰勒展开, 移除常数项 l , 可得:

$$L = \sum_{i=1}^n (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t)$$

其中, g_i 代表损失函数 l 对 $t-1$ 轮迭代得到的植物多样性预测值 y'^{t-1}_i 所求的一阶导数, h_i 为损失函数 l 对 $t-1$ 轮迭代得到的植物多样性预测值 y'^{t-1}_i 所求的二阶导数。XGBoost 模型中引入叶子节点分数 $\omega_{q(x)}$ 代表样本落入的具体叶子节点, 叶子节点分数即为每个样本的预测值, 即可将 $f_t(x)$ 转换为 $\omega_{q(x)}$ 。定义树模型的复杂度, 即为正则化函数 $\Omega(f_t)$ 与叶子节点 j 的样本集合为 I_j , 表达式如下:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$$

$$I_j = \{i \mid q_{(x_i)} = j\}$$

其中, $\sum_{j=1}^T \omega_j^2$ 代表叶子权重。由于需要预测的值 ω 不确定, 对目标函数进行一阶求导并令其数值为 0, 求出叶子节点 j 的数值。最终得到目标函数最小值公式 L 如下:

$$\omega_j = -\frac{G_j}{H_j + \lambda}$$

$$L = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T$$

其中: $G_i = \sum g_i$, $H_i = \sum h_i$ 。

为解决模型解释性较差的问题, 引入 SHAP 框架对模型结果进行解释。SHAP 框架的构建在反映特征重要性的同时, 可以反映各个特征对结果的影响正负性。SHAP 框架将输入的特征当作贡献者, 对每个特征的贡献值进行计算, 最终将贡献值累加得到最终预测^[27], 公式如下:

$$s = f_0 + \sum_{i=1}^M f_i$$

其中, s 为模型预测值, f_i 为每个特征对应的归因值, M 为特征数目, f_0 为所有训练样本的预测均值。

针对 XGBoost 模型, SHAP 提供了 Shapley value 与模型预测值之间的关系转换公式, 即:

$$\ln \frac{s}{1-s} = f_0 + \sum_{i=1}^M f_i$$

每个特征对应的 SHAP 值代表该特征对总体预测结果的贡献度。当归因值 $f_i > 0$ 时, 表明此时该特征对预测值有提升作用, 反之则表明该特征对预测值有降低作用。

(2) 参数调整

XGBoost 模型构建中, 超参数的调整对抑制模型过拟合至关重要。采用网格搜索和交叉验证相结合的方法进行自动化参数调整, 降低时间成本的同时可有效获取最佳参数组合^[30]。经过网格搜索与交叉验证, 本研究中 XGBoost 模型最佳超参数组合为: nrounds = 100, max_depth = 8, eta = 0.1, gamma = 0, colsample_bytree = 0.8, min_child_weight = 1, subsample = 0.7, 其他参数保持默认值。

(3) 模型评估

选择决定系数 (R^2)、平均绝对误差 (MAE) 与均方根误差 (RMSE) 指标用于评估模型效果^[28], 其计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}$$

其中, y_i 为植物多样性指数实际值, y'_i 为植物多样性指数预测值。

决定系数 R^2 的取值区间为 $(-\infty, 1]$, 反映植物多样性指数中的总变异量可通过环境变量在回归分析中的解释在全部样本中的占比, 越接近于 1 说明模型效果越好, 预测结果越接近真实值。MAE 与 RMSE 能够衡量模型预测值与真实值之间的差异, 值越小表明模型结果越精准。

为去除自变量数量对 R^2 值的影响, 引入校正决定系数 (Ra^2) 评估模型准确性。 Ra^2 同时考虑了样本数量与变量个数, 评估效果更加精准^[31]。 Ra^2 通常在 $[0, 1]$ 上取值, 越趋近于 1 代表模型可解释性越高, 模型效果越好。其计算公式如下:

$$Ra^2 = 1 - (1 - R^2) \frac{(n - 1)}{(n - k - 1)}$$

其中, n 为样本数量, k 为自变量数量。

1.2.3 植物多样性空间格局及冷热点分析

基于 XGBoost 得到消落带植物群落物种丰富度、物种多样性和谱系多样性的预测值, 在 ArcGIS 里分别生成消落带植物群落物种丰富度、物种多样性和谱系多样性的空间分布格局。利用 ArcGIS 的热点分析计算 G_i^* 指数, 对消落带植物多样性指标的高值区和低值区进行表征^[32]。计算公式如下:

$$G_i^* = \frac{\sum_j w_{ij} \times x_j - \bar{X} \times \sum_j w_{ij}}{S \times \sqrt{\frac{[n \times \sum_j w_{ij}^2 - (\sum_j w_{ij})^2]}{n - 1}}}$$

其中, w_{ij} 是栅格 i 与 j 之间的权重值矩阵, x_j 是栅格 j 的多样性值, \bar{X} 是所有栅格多样性值的平均值, S 是多样性值的标准差, n 为栅格总数。

如 $G_i^* > 0$ 则该栅格是热点, 反之则是冷点。根据 P 值, $P < 0.05$ 对应的统计置信区间为 95%, $P < 0.1$ 对应

的统计置信区间为 90%,据此划分出极冷点区、冷点区、热点区和极热点区。

2 结果与分析

2.1 2013 年水库消落带植物多样性空间格局及关键环境因素

2.1.1 基于 XGBoost 模型的水库消落带植物多样性空间格局预测

采用 XGBoost 模型分别预测消落带植物群落物种丰富度 R 、物种多样性 H 与谱系多样性 PD (图 5)。对 3 个模型的预测能力进行比较,发现谱系多样性 PD ($R^2 = 0.78, Ra^2 = 0.76, MAE = 98.75, RMSE = 128.31$) 的预测能力最好,物种丰富度 R ($R^2 = 0.71, Ra^2 = 0.69, MAE = 3.13, RMSE = 5.53$) 的预测能力次之,物种多样性 H ($R^2 = 0.62, Ra^2 = 0.59, MAE = 0.25, RMSE = 0.31$) 的预测能力相对较低。2013 年水库消落带植物群落物种丰富度 R 的预测值范围为 $[3, 37]$,物种多样性 H 的预测值范围为 $[0, 3]$,谱系多样性 PD 的预测值范围为 $[536, 2318]$ 。消落带植物群落物种丰富度、物种多样性和谱系多样性的空间格局较为一致,物种丰富度较高的区

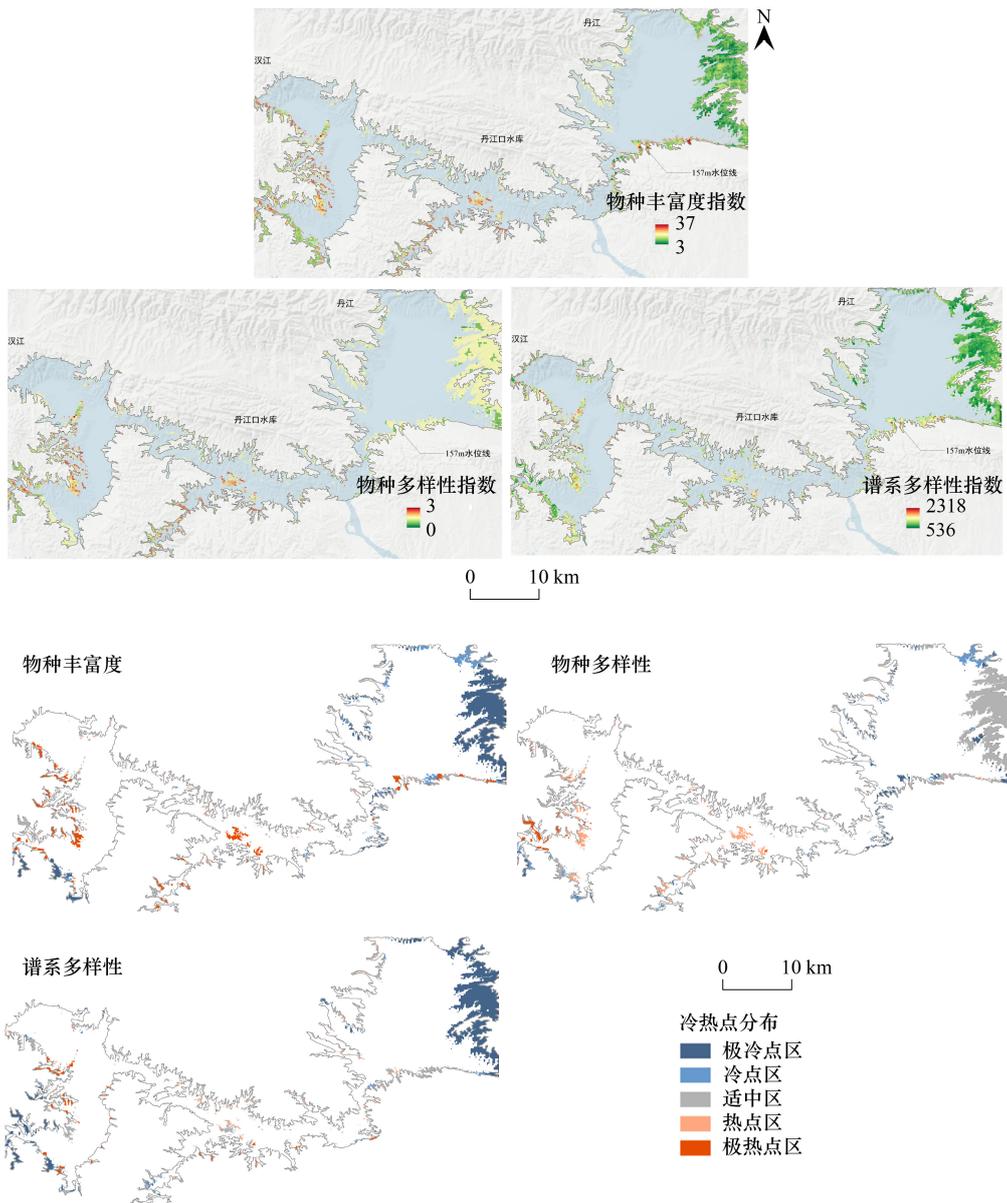


图 5 2013 年丹江口水库消落带植物多样性空间格局预测及冷热点分布

Fig.5 Predicted plant diversity in the water-level fluctuation zone of Danjiangkou Reservoir in 2013

域,一般也具有较高的物种多样性和谱系多样性。植物群落的物种丰富度和谱系多样性的冷点区主要集中在水库东北部和西南角,物种多样性的冷点区零散分布在水库的东南部和西南部,而热点区广泛且零散分布在水库西部和中部(图5)。

2.1.2 基于 SHAP 的解释性分析

基于 SHAP 分析对水库消落带植物多样性预测模型结果进行解释性分析,绘制 SHAP 依赖图来呈现单个

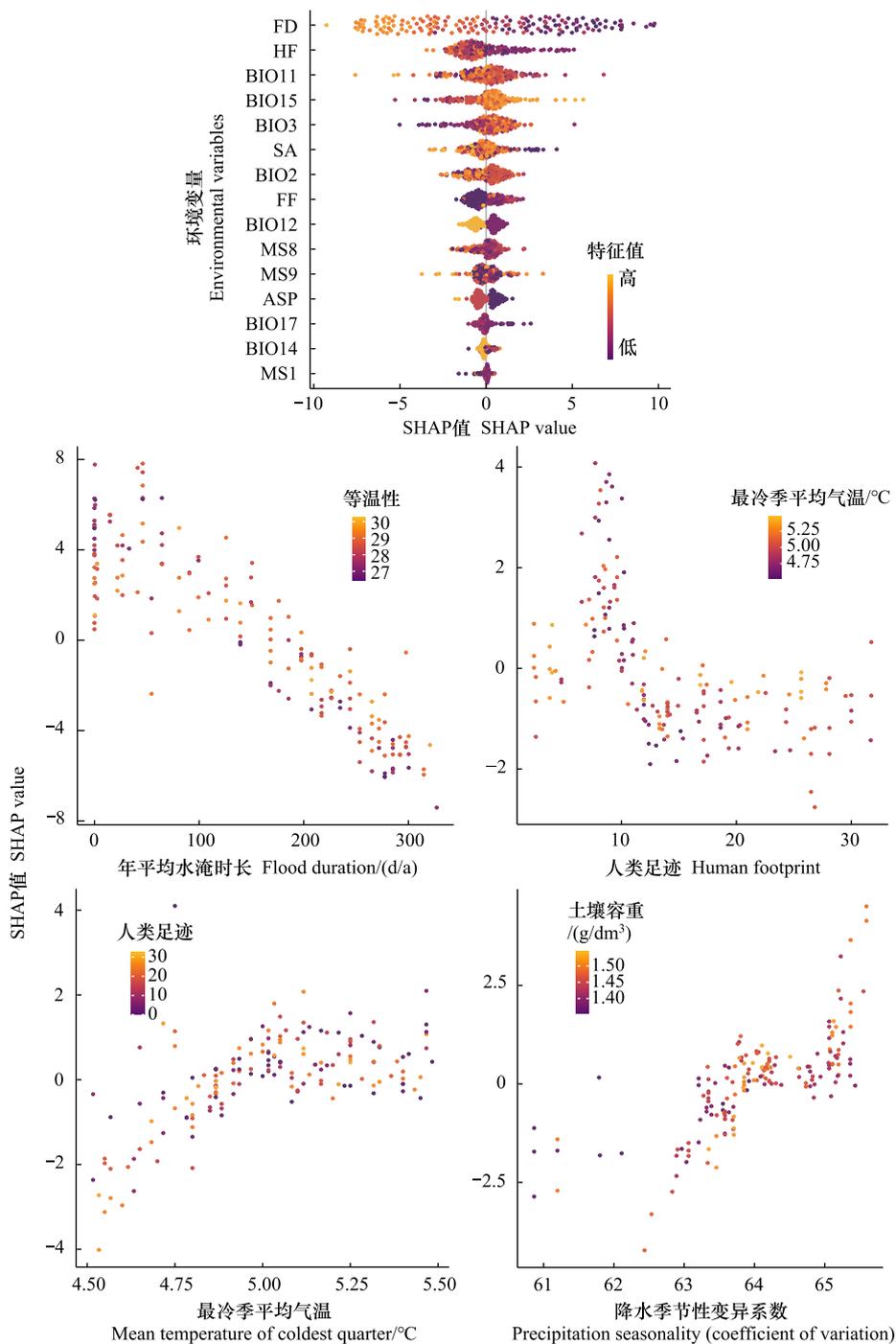


图6 消落带植物群落物种丰富度的 SHAP 特征分析

Fig.6 Summary of SHAP feature analysis of species richness of plant communities in in the water-level fluctuation zone of Danjiangkou Reservoir

环境因素的独立作用及两个关系密切因素之间的共同作用(图 6—8)。年平均水淹时长 FD、人类足迹 HF、最冷季平均气温 BIO11、降水季节性变异系数 BIO15 等因素对消落带植物群落物种丰富度影响较为显著(图 6)。选取影响显著的 FD、HF、BIO11 和 BIO15 绘制 SHAP 依赖图(图 6)。结果表明,影响消落带植物群落物种丰富度的最重要特征为年平均水淹时长,该特征对物种丰富度的影响呈现较明显的反向关系,即随着年平均水淹时长增加物种丰富度降低。而且年平均水淹时长越长,BIO3(等温性)越大,物种丰富度降低的可能性越大;人类足迹对物种丰富度呈现复杂的非线性关系,即随着人类活动强度增加,物种丰富度呈现先上升后降低的趋势。而且人类活动强度较低时,BIO11(最冷季平均气温)越低,物种丰富度下降的可能性越大;降水季节性变异系数和最冷季平均气温对物种丰富度也呈现非线性关系。随着最冷季平均气温的增加物种丰富度先增加,但增加到一定程度后物种丰富度逐渐保持稳定且有回落的趋势。而且最冷季平均气温增加时 HF(人类活动强度)较低,会使物种丰富度保持稳定的可能性增大;当降水季节性变异系数增加,物种丰富度指数先增加,当降水季节性变异系数增加到一定程度,物种丰富度逐渐保持相对稳定的状态。

年平均水淹时长 FD、平均日较差 BIO2、人类足迹 HF、最冷季平均气温 BIO11 等因素对消落带植物群落物种多样性影响较为显著(图 7)。SHAP 依赖分析的结果表明(图 7),影响消落带植物群落物种多样性的最重要特征为年平均水淹时长,该特征对物种多样性的影响呈现较明显的反向关系,即随着年平均水淹时长增加物种多样性降低。而且年平均水淹时长越长,平均日较差越低,会使物种多样性降低的可能性增加;人类足迹对物种多样性有非线性作用,随着人类活动强度增加物种多样性先升高后降低,而且人类活动强度越大,BIO12(年均降水量)越大,物种多样性降低的可能性越大;平均日较差和最冷季平均气温对物种多样性的影响较为一致,随着平均日较差和最冷季平均气温的增加,物种多样性呈现先上升再降低的趋势。而且最冷季平均气温越高,较高的年平均水淹时长会增大物种多样性增加的可能性。随着最冷季平均气温的持续升高,较高的年平均水淹时长反而会增大物种多样性降低的可能性。

年平均水淹时长 FD、平均日较差 BIO2、最冷季平均气温 BIO11 和人类足迹 HF 等因素对消落带植物群落谱系多样性影响较为显著(图 8)。SHAP 依赖分析的结果表明(图 8),影响消落带植物群落谱系多样性的最重要特征为年平均水淹时长,该特征对谱系多样性的影响呈现明显的反向关系,即随着年平均水淹时长增加植物群落谱系多样性降低,而且年平均水淹时长越长,BIO3(等温性)数值越高,则谱系多样性降低的可能性越大;平均日较差和谱系多样性之间的关系呈现出非线性特征。随着平均日较差增加,谱系多样性呈现先增加后降低的趋势,而且 FF(水淹频率)越高,谱系多样性随平均日较差增加所呈现先增加后降低的可能越大;最冷季平均气温在一定范围内的提高能够有效增加谱系多样性,但当最冷季平均气温取值达到约 5.25℃后,最冷季平均气温的增加反而会使谱系多样性降低。而且 BIO17(最干季降水量)较小时,谱系多样性降低的可能性会随着最冷季平均气温的增加而变大;人类足迹对谱系多样性的影响呈现出非线性关系。人类足迹增加过程中存在两个临界值,当人类足迹增加到第一个临界值时,对谱系多样性的促进作用变为抑制作用;当人类足迹继续增加至第二个临界值时,对谱系多样性的抑制作用又变为促进作用。而且在人类足迹到达第一个临界值前,较低的 BIO12(年均降水量)会使谱系多样性增加的可能性增加,而在人类足迹在第一和第二个临界值之间时,较低的年平均降水量会使谱系多样性降低的可能性增加,人类足迹在超过第 2 个临界值之后,较高的年平均降水量会使谱系多样性增加的可能性增加。

通过对比三个指标的影响因素,发现年平均水淹时长、人类足迹与最冷季平均气温对消落带植物群落物种丰富度、物种多样性和谱系多样性都有显著影响,其中年平均水淹时长的影响最大。将 XGBoost 模型输出的特征重要性排名(图 9)与 SHAP 解释结果(图 6—8)对比,发现两种方法的结果在影响最大的环境变量上一致,其他环境变量的影响力排序上有所差异。XGBoost 模型的结果表明影响消落带植物群落物种丰富度较为显著的特征为 FD、BIO2、BIO11 和 HF(图 9),其中 FD、BIO11 和 HF 在 SHAP 结果中也影响较大(图 6),证实了年平均水淹时长、最冷季平均气温与人类活动强度对消落带植物群落物种丰富度的重要影响;FD、BIO2、BIO11 和 BIO3 对物种多样性具有显著影响(图 9),FD、BIO2 和 BIO11 在 SHAP 结果中也是影响力较大的三

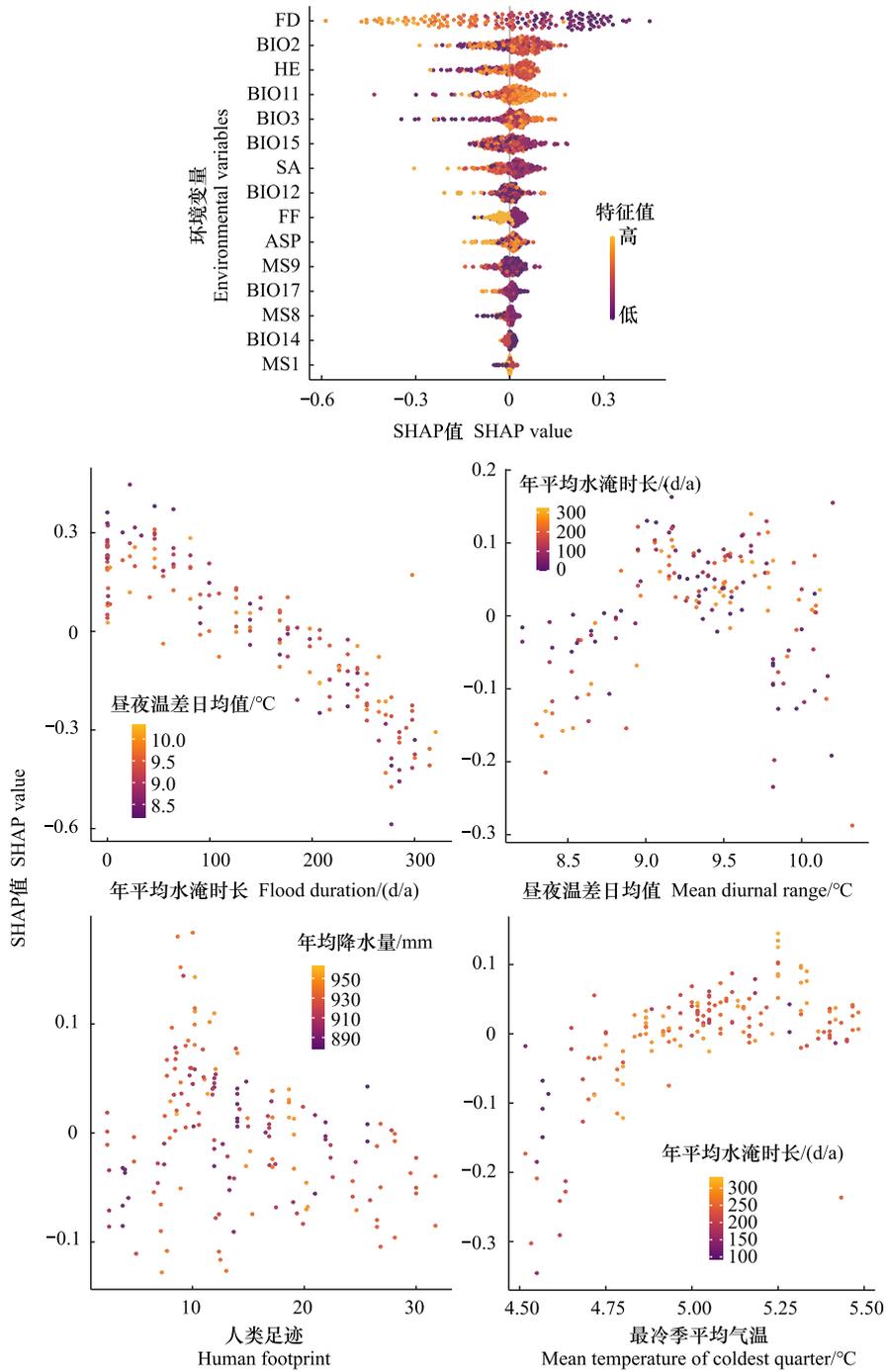


图 7 消落带植物群落物种多样性的 SHAP 特征分析

Fig.7 Summary of SHAP feature analysis of species diversity of plant communities in the water-level fluctuation zone of Danjiangkou Reservoir

个因素(图 7),表明年平均水淹时长、平均日较差以及最冷季平均气温对消落带植物群落物种多样性的重要影响;FD、BIO2、BIO15 和 BIO11 对群落谱系多样性影响显著(图 9),其中 FD、BIO2 和 BIO11 在 SHAP 结果中也是影响较大的因素(图 8),证实年平均水淹时长、平均日较差和最冷季平均气温对消落带植物群落谱系多样性的重要影响。

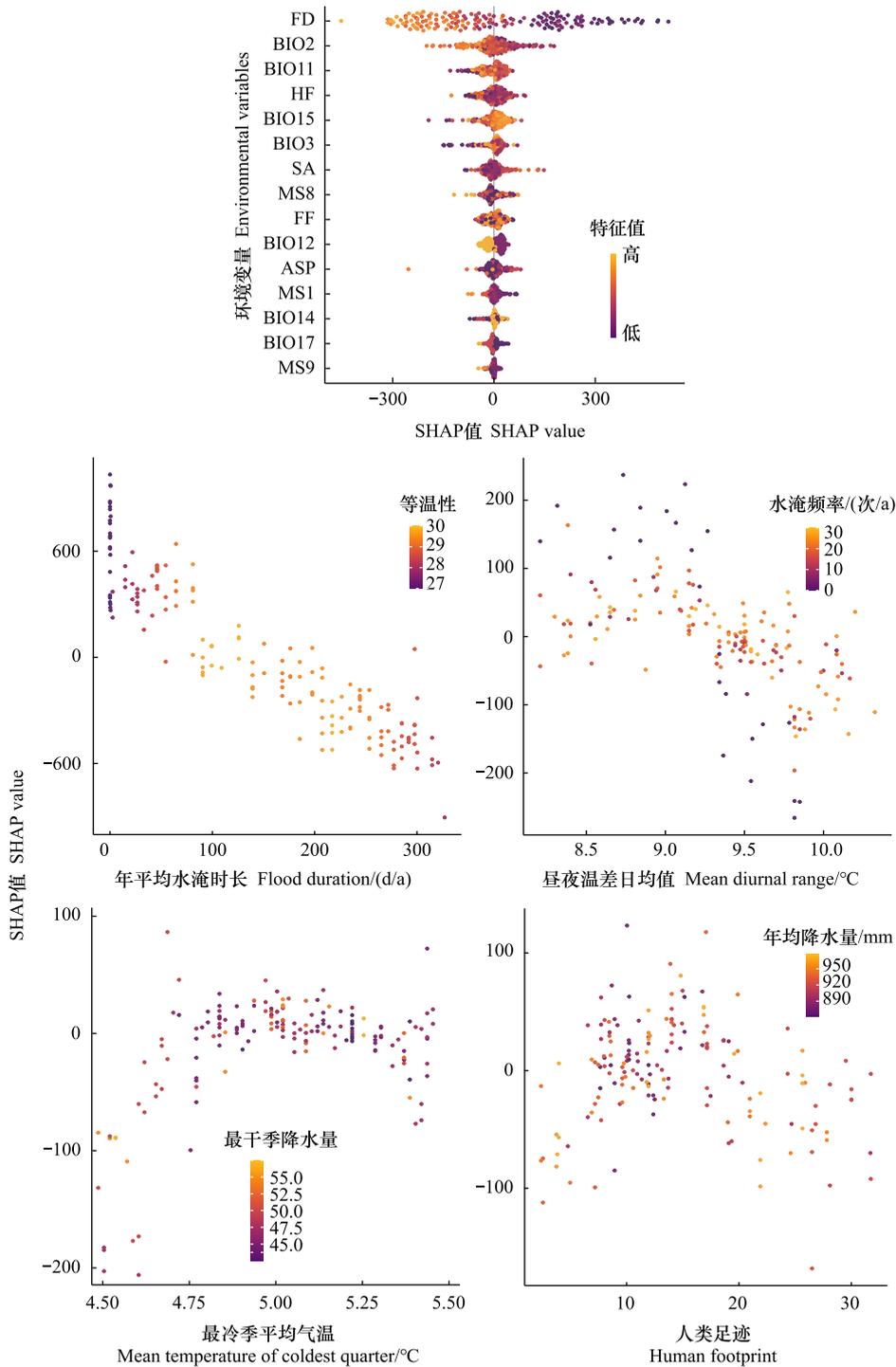


图 8 消落带植物群落谱系多样性的 SHAP 特征分析

Fig.8 Summary of SHAP feature analysis of phylogenetic diversity of plant communities in the water-level fluctuation zone of Danjiangkou Reservoir

2.2 2050 年丹江口水库消落带植物多样性空间格局预测

采用 XGBoost 模型预测 2050 年水库消落带植物多样性空间格局(图 10)。由于南水北调中线工程运行后水位抬升,消落带上升到海拔 160 m 至 170 m 之间的空间范围,2050 年水库消落带面积相比 2013 年明显减少(图 5、图 10)。由于环境因素的变化,2013 年与 2050 年水库消落带植物多样性呈现出明显不同的空间格

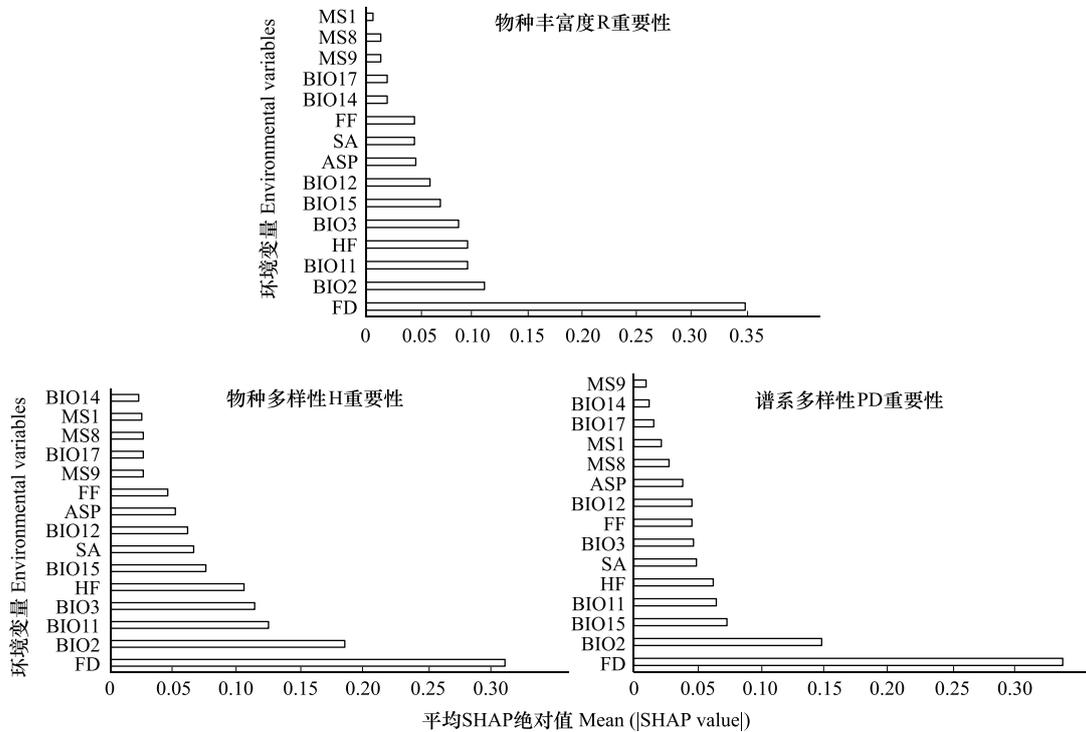


图9 XGBoost 环境变量的重要性
 Fig.9 Importance of environmental variables based on XGBoost models

局。物种多样性的空间分布格局在 3 种指标上都变化明显,2050 年水库消落带植物群落物种多样性水平较 2013 年有明显提升。而且对比分析 2013 年和 2050 年的冷热点图发现物种丰富度、物种多样性和谱系多样性的冷热点空间分布有明显变化。物种丰富度的极冷点区从 2013 年的水库东北部集中分布变成分散分布;物种多样性的极冷点区从 2013 年的零星分散分布转变为集中于水库西南角,而极热点区集中分布在中部且在西部也有零散分布,热点区集中分布在东北部;谱系多样性的极冷点区从 2013 年的水库东北部集中分布转移到西南部和中部,极热点区和热点区的分布区域数量及面积相比 2013 年明显减少(图 10)。

3 结论与讨论

本研究构建了一个基于 XGBoost 算法的预测模型,通过实地调查所得的水库消落带植物多样性数据和多源数据库中提取的环境数据,对 2013 年水库消落带植物多样性空间格局进行预测。由预测结果可知(图 5),XGBoost 模型预测的水库消落带植物群落物种丰富度、物种多样性和谱系多样性的平均值分别为 16.3、2.3 和 1245.6,略低于实地调查的平均值 17.6、2.3 和 1450.6,但与实地调查值非常接近,反映出 XGBoost 模型在消落带植物多样性预测方面具有明显优势且预测结果具有可靠性。2013 年实地调查所获得的消落带植物群落物种丰富度、物种多样性和谱系多样性的真实值范围分别为 [3, 52]、[0, 3]、[537, 3119],而 XGBoost 模型得到的预测范围分别为 [3, 37]、[0, 3]、[536, 2318],通过对比真实值与预测值发现,在低取值范围的多样性指标预测时 XGBoost 模型与真实值较为接近;在高取值范围的多样性指标预测时 XGBoost 模型会略微低估多样性值,在对谱系多样性进行预测时,这一现象较为明显。XGBoost 模型的预测值略低于实地调查的测量值,主要原因是实地调查时多挑选生长情况较好的植物群落进行调查,而生长情况较差的植物群落较少纳入调查,地表无植被覆盖完全不被调查。

尽管 XGBoost 算法相比传统统计方法有优秀的预测能力和计算速率,但因缺乏可解释性严重限制了其应用^[33]。为了解释模型结果,本研究利用 SHAP 对 XGBoost 模型结果进行环境变量的特征分析,发现年平均水

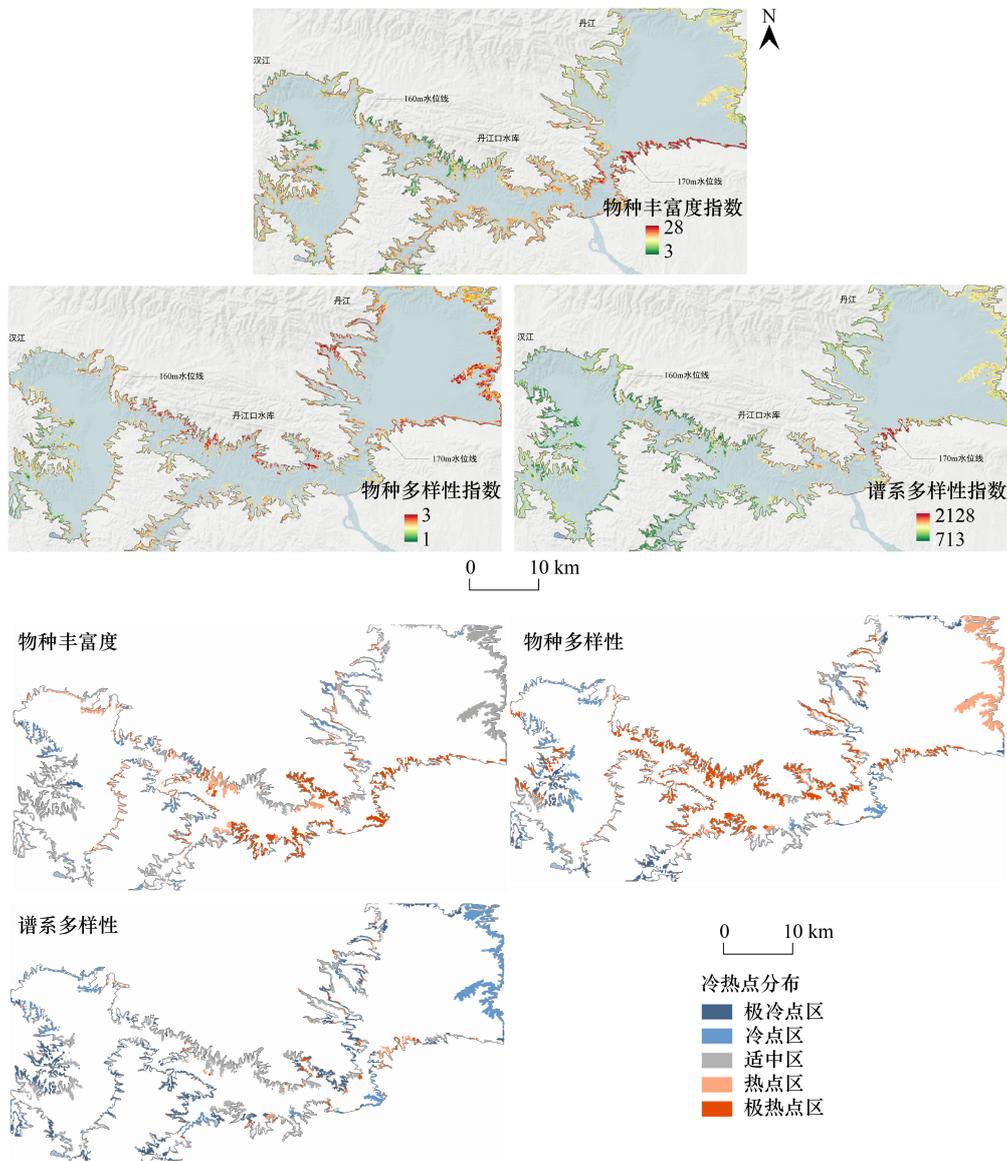


图 10 2050 年丹江口水库消落带植物多样性空间格局预测及冷热点分布

Fig.10 Predicted plant diversity in the water-level fluctuation zone of Danjiangkou Reservoir in 2050

淹时长、人类足迹和最冷季平均气温对消落带植物群落物种丰富度、物种多样性和谱系多样性有显著影响,其中年平均水淹时长的影响最大(图 6—8)。这一结果与 XGBoost 得到的特征重要性排名(图 9)在影响最大的环境变量上保持一致。而且查阅前人对水库消落带植物多样性—环境关系的相关研究,发现虽然前人研究通过实地调查获得植物多样性数据和环境变量数据,采用传统统计方法,研究结果表明水淹影响是决定消落带植物群落空间分布的主导因素^[34—35],对消落带植物群落物种丰富度、物种多样性和谱系多样性起着决定性作用^[21]。基于 XGBoost-SHAP 模型得到的结果与这些基于实地调查的研究结果较为符合,证实了本研究 XGBoost-SHAP 模型具有良好的模型性能和可解释性,在生物多样性—环境关系的相关研究中有较高可行性。

相比前人关于生物多样性空间格局的研究具有一定的优势。基于实地调查所得的植物多样性数据,结合一些公开的、大规模的数据库所获得的空间大数据,实现了对大尺度空间范围精细分辨率的植物多样性空间格局预测,并可以对未来情景的植物多样性空间格局进行前瞻性预测。机器及深度学习算法在生态建模领域的出现为生物多样性监测和管理提供了新方法。这些算法具有复杂的架构,在性能和准确性方面远超传统的

统计方法;采用的方法锚定在 XGboost 机器学习算法的同时引入 SHAP 框架,构建了一个结果可解释的预测模型。此方法促进了对生物多样性和环境之间关系的深入探索,将是未来生物多样性相关研究的一个重要视角,可为复杂的生物多样性形成机制提供更深刻的见解。但也存在一定局限性。构建模型所使用的植物多样性数据是基于实地调查获得,样本量 205 个,虽然对实地调查来说这已是不小的样本量,但对于机器及深度学习算法来说,一般样本量越大预测模型性能越好、预测结果越精准;植物多样性的环境影响机制复杂,尽管已经使用了包括气候、土壤、地形、水文和人类活动 5 类 34 个环境变量,但由于这些环境变量空间大数据的分辨率相比实地调查数据的分辨率过于粗糙,会造成预测结果与实地情况存在一定差异。

参考文献 (References):

- [1] Tilman D, Isbell F, Cowles J M. Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution, and Systematics*, 2014, 45: 471-493.
- [2] Cardinale B J, Duffy J E, Gonzalez A, Hooper D U, Perrings C, Venail, Naeem S. Biodiversity loss and its impact on humanity. *Nature*, 2012, 486: 59-67.
- [3] Cai L, Kreft H, Taylor A, Denelle P, Schrader J, Essl, Weigelt P. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist*, 2023, 237: 1432-1445.
- [4] Ulloa C, Acevedo-Rodríguez P, Beck S, Belgrano M J, Bernal R, Berry P E, Jørgensen P M. An integrated assessment of the vascular plant species of the Americas. *Science*, 2017, 358(6370): 1614-1617.
- [5] Currie D J, Paquin V. Large-scale biogeographical patterns of species richness of trees. *Nature*, 1987, 329: 326-327.
- [6] 乔慧捷, 胡军华. 利用数值模拟重构物种多样性格局的形成过程. *生物多样性*, 2022, 30(10): 355-363.
- [7] Sabatini F M, Jiménez-Alfaro B, Jandt U, Chytrý M, Field R, Kessler M, Bruelheide H. Global patterns of vascular plant alpha diversity. *Nature Communications*, 2022, 13(1): 4683.
- [8] Keil P, Chase J M. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, 2019, 3: 390-399.
- [9] Beck J, Ballesteros Mejia L, Buchmann C M, Dengler J, Fritz S A, Gruber B, Dormann C F. What's on the horizon for macroecology? *Ecography*, 2012, 35(8): 673-683.
- [10] Andermann T, Antonelli A, Barrett R L, Silvestro D. Estimating alpha, beta, and gamma diversity through deep learning. *Frontiers in Plant Science*, 2022, 13: 839407.
- [11] Večeřa M, Divíšek J, Lenoir J, Jiménez Alfaro B, Biurrun I, Knollová I, Chytrý M. Alpha diversity of vascular plants in European forests. *Journal of Biogeography*, 2019, 46: 1919-1935.
- [12] Crisci C, Ghattas B, Perera G. A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, 2012, 240: 113-122.
- [13] Park D S, Willis C G, Xi Z, Kartesz J T, Davis C C, Worthington S. Machine learning predicts large scale declines in native plant phylogenetic diversity. *New Phytologist*, 2020: 1544-1556.
- [14] Chang G. Biodiversity estimation by environment drivers using machine/deep learning for ecological management. *Ecological Informatics*, 2023, 78: 102319.
- [15] 李培坤, 陈旭梅, 鲁文博, 马嘉欣, 刘屹, 王昊. 基于 XGBOOST-SHAP 的地铁建成环境与站点出行距离的非线性关系研究. *铁道科学与工程学报*, 2024, 21(04): 1624-1633.
- [16] 陈小昆, 左航旭, 廖彬, 孙瑞娜. 融合 XGBoost 与 SHAP 的冠心病预测及其特征分析模型. *计算机应用研究*, 2022, 39(6): 1796-1804.
- [17] 廖彬, 王志宁, 李敏, 孙瑞娜. 融合 XGBoost 与 SHAP 模型的足球运动员身价预测及特征分析方法. *计算机科学*, 2022, 49(12): 195-204.
- [18] 罗妍, 王枫, 叶文玲. 基于 XGBoost 和 SHAP 的急性肾损伤可解释预测模型. *电子与信息学报*, 2022, 44(1): 27-38.
- [19] 刘瑞雪, 詹娟, 史志华, 陈龙清. 丹江口水库消落带土壤种子库与地上植被和环境的关系. *应用生态学报*, 2013, 24(3): 801-808.
- [20] 江维薇, 杨楠, 肖衡林. 三峡库区与西南库区消落带植物多样性及群落构建比较. *湖泊科学*, 2023, 35(2): 564-579.
- [21] 刘瑞雪, 林小龙, 陈龙清. 水淹干扰下水库水滨带植物群落谱系结构与物种多样性格局. *生态学报*, 2023, 43(11): 4700-4711.
- [22] 刘瑞雪, 陈龙清. 水淹干扰对水库水滨带植物群落稳定性与种间关系的影响. *生态学报*, 2021, 41(16): 6566-6579.
- [23] O'Neill B C, Tebaldi C, van Vuuren D P, Eyring V, Friedlingstein P, Hurtt, Sanderson B M. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 2016, 9: 3461-3482.

- [24] Patiño J, Collart F, Vanderpoorten A, Martin-Esquivel J L, Naranjo-Cigala A, Mirolo S, Karger D N. Spatial resolution impacts projected plant responses to climate change on topographically complex islands. *Diversity and Distributions*, 2023, 29(10): 1245-1262.
- [25] Karger D N, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza R W, Zimmermann N E, Linder H P, Kessler M. Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 2017, 4: 170122.
- [26] Chauvier Y, Descombes P, Guéguen M, Boulangeat L, Thuiller W, Zimmermann N E. Resolution in species distribution models shapes spatial patterns of plant multifaceted diversity. *Ecography*, 2022(10): e05973.
- [27] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 2014, 41(3): 647-665.
- [28] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016:785-794.
- [29] 徐大星, 纪木强, 涂泽松, 许伟鹏, 徐伟龙, 牛维. 机器学习分析肱骨近端内侧柱不稳定骨折术后失效的风险因素. *中国组织工程研究*, 2024, 28(33): 5295-5301.
- [30] Syarif I, Prugel-Bennett A, Wills G. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *TELKOMNIKA*, 2016, 14(4): 1502.
- [31] Sun H, Luo Q, Xia Z, Li Y, Yu Y. Bottomhole pressure prediction of carbonate reservoirs using XGBoost. *Processes*, 2024, 12(1): 125.
- [32] Sun L, Yu H, Sun M, Wang Y. Coupled impacts of climate and land use changes on regional ecosystem services. *Journal of Environmental Management*, 2023: 116753.
- [33] Rahmanian S, Nasiri V, Amindin A, Karami S, Maleki S, Pouyan S, Borz S A. Prediction of plant diversity using multiseasonal remotely sensed and geodiversity data in a mountainous area. *Remote Sensing*, 2023, 15: 387.
- [34] 刘瑞雪, 陈龙清, 史志华. 丹江口水库水滨带植物群落空间分布及环境解释. *生态学报*, 2015, 35(4): 1208-1216.
- [35] 程莅登, 袁兴中, 孙阔, 唐婷, 袁嘉. 三峡库区消落带植物群落及其功能性状对不同水淹强度的响应. *生态学报*, 2024, 44(11): 4795-4807.