

doi:10.3969/j.issn.2097-0013.2024.04.013

## 基于随机森林赋权信息量的区域滑坡易发性评价 ——以三峡库区秭归至巴东段为例

马 敏<sup>1</sup>, 王江立<sup>1</sup>, 陈 琦<sup>1</sup>, 李景富<sup>2</sup>

MA Min<sup>1</sup>, WANG Jiang-Li<sup>1</sup>, CHEN Qi<sup>1</sup>, LI Jing-Fu<sup>2</sup>

1. 中国地质调查局武汉地质调查中心(中南地质科技创新中心), 湖北 武汉 430205;

2. 广东省有色地质环境中心, 广东 广州 510080

1. Wuhan Center of China Geological Survey (Geosciences Innovation Center of Central South China), Wuhan 430205, Hubei, China;

2. Guangdong Nonferrous Geological Environment Center, Guangzhou 510000, Guangdong, China

**摘要:** 在滑坡易发性评价研究中, 常规的信息量模型通常将不同评价指标的信息量进行简单累加, 而没有重视各评价指标间的权重差异, 导致易发性分区结果不精确。本文以三峡库区秭归至巴东段为例, 提出了一种基于随机森林赋权信息量的滑坡灾害易发性评价方法, 以求提高评价精度。首先利用因子特征分析方法确定滑坡评价因子, 接着运用随机森林模型确定各评价因子的权重, 然后将权重与信息量模型融合, 通过将评价因子信息量加权叠加的方法得到更准确的易发性评价结果。通过统计指标和 ROC 曲线对两个模型进行了分析和评估。结果显示, 传统信息量模型和基于随机森林赋权的信息量模型在测试集的 ROC 曲线下面积(AUC)分别为 0.778 和 0.855, 确定随机森林赋权信息量的方法对传统信息量方法起到了优化作用, 该方法为滑坡灾害风险评估提供了新的思路。

**关键词:** 滑坡; 随机森林模型; 信息量模型; 易发性评价; 三峡库区秭归至巴东段

中图分类号: P642.22

文献标识码: A

文章编号: 2097-0013(2024)-04-0749-15

**Ma M, Wang J L, Chen Q and Li J F. 2024. Regional Landslide Susceptibility Evaluation Based on Random Forest Weighting Information: a Case Study of Zigui to Badong Section in the Three Gorges Reservoir Area. *South China Geology*, 40(4): 749–763.**

**Abstract:** In the study of landslide susceptibility assessment, conventional information models usually simply accumulate the information of different evaluation indicators, without paying attention to the weight differences between each evaluation indicator, resulting in inaccurate susceptibility zoning results. This article takes the Zigui to Badong section of the Three Gorges Reservoir area as an example and proposes a landslide hazard susceptibility evaluation method based on random forest weighting information to improve evaluation accuracy. This method first uses factor feature analysis to determine landslide evaluation factors, and then uses a random forest model to determine the weights of each evaluation factor. Then, the weights are fused with an information model, and a more accurate susceptibility evaluation result is obtained by weighting and superimposing the information of the evaluation factors. Two models were analyzed and evaluated using

收稿日期: 2024-09-25; 修回日期: 2024-10-24

基金项目: 中国地质调查局项目(DD20230706)

第一作者: 马敏(1984—), 女, 工程师, 研究方向为地质数据库建设、地质大数据处理与分析, E-mail: mamin@mail.cgs.gov.cn

statistical indicators and ROC curves. The results showed that the area under the ROC curve (AUC) of the traditional information model and the information model based on random forest weighting were 0.778 and 0.855, respectively, in the test set. The method of determining the weighted information of the random forest optimized the traditional information method, providing a new approach for landslide disaster risk assessment.

**Key words:** landslide; random forest model; information content model; susceptibility assessment; Zigui to Badong section of the Three Gorges Reservoir area

三峡库区是我国地质灾害的重点防治区域,区内共有 5386 处滑坡、崩塌纳入后续工作地质灾害防治,其中大型、特大型滑坡有 1339 处(叶润青等, 2020)。滑坡易发性是指在特定地质和环境条件下,由孕灾地质条件综合决定的某一地区发生滑坡的可能性,评价的重点在于评估滑坡发生的地点和概率(Reichenbach et al., 2018),进行滑坡易发性评价可以为三峡库区的防灾减灾管理提供科学依据,进而推动防灾减灾技术的发展。滑坡易发性评价过程主要涉及评估单元、评价因子、评估模型三个方面。常用的评估单元有栅格单元、地域单元、子流域单元、均一条件单元和斜坡单元(赵晓燕等, 2021; 章昱等, 2023; 赵萍等, 2024)。地域单元往往以行政区域划分,与滑坡的自然发生规律不够吻合;子流域单元需要详细的水文和地形数据;均一条件单元在条件变化较大的区域难以应用;斜坡单元的划分受主观因素影响,且可能因为过于关注单个斜坡而忽视了区域尺度上地质因素间的相互作用(丁丽等, 2023; 李泽芝和王新刚, 2024);而栅格单元易于集成多源数据,保持了空间数据的连续性,划分灵活高效且利于建模和分析,使用最为广泛(陈涛等, 2020; 余淙蔚等, 2022)。

在滑坡易发性研究中,必须考虑历史滑坡事件、地质环境因素和空间位置之间的关系,地质环境因素直接影响易发性评价结果的可靠性和准确性(Pham et al., 2015)。滑坡的发生主要取决于地质、环境条件的一系列动态、静态因素和内在、外在因素。在评价因子的选择方面目前并没有普遍适用的规则(Merghadi et al., 2020)。前人研究表明,地形因素在滑坡敏感性制图中起着主要作用,而这些因素通常来源于数字高程模型(DEM)(Dai F C et al., 2002)。本文利用 DEM 获取了高程、坡度、坡向、曲率、地形起伏度 5 个地形因子。地

层岩性、基础地质结构对斜坡稳定性具有绝对的控制作用,本文选取工程岩组、斜坡结构作为 2 个地质评价因子来进行滑坡易发性评价研究。库水位变化、降雨、土壤含水量都是诱发滑坡发生的重要环境因素,本文选取年均降雨量、土壤含水量、库水缓冲区作为评估研究区水文状况的初步指标。人类工程活动引发的荷载增加、植被破坏、排水系统改变等也直接或间接导致滑坡发生,本文选取植被覆盖度(NDVI)、土地利用类型作为反映人类工程活动的影响指标。以上因子通过调查和收集均可以获取高精度数据,确保滑坡易发性评价的准确性。

评估模型的选择也是大量学者的探究方向,这些模型大致可分为三种:基于物理的模型、基于知识的模型和基于数据的模型。基于物理的模型通常用于小尺度建模(Tofani et al., 2017; Medina et al., 2021);基于知识的模型依赖于专家知识来生成可靠的敏感性地图(Zhu A X et al., 2014);基于数据的模型主要包括三类:统计模型、机器学习模型和深度学习模型,统计模型如信息量模型、权重证据模型等可以计算特定的统计指标,反映滑坡对潜在因素的内部响应(钱紫玲, 2023; 林振等, 2024),机器学习模型如随机森林(杜鹏等, 2024)、决策树(黄发明等, 2021)等被用于滑坡易发性评价及主控因素分析,深度学习模型如卷积神经网络(李明等, 2023)、循环神经网络(李璐等, 2022)等被用于滑坡位移时空预测。然而,各种评价模型都有其局限性,运用组合模型日趋成为滑坡易发性评价的趋势。其中,统计模型与机器学习模型的组合不仅易于实现,而且减少了单一模型的过拟合风险,并结合了统计模型的解释性和机器学习模型的预测能力,使得该组合应用最为广泛(彭亮等, 2021; 王高峰等, 2021; 连志鹏等, 2022)。

信息量模型作为统计模型的代表, 不仅实现简便, 而且评价结果客观, 预测精度高, 因而在实际应用中广泛采用(肖海平等, 2024; 吴兴贵等, 2024), 随机森林模型具有自动特征选择和重要性评估、广泛适用性等多重优势。本文将随机森林模型具有重要性评估的能力应用于信息量模型中, 发挥两个模型的综合优势, 以期提高滑坡易发性评价的精度。

## 1 研究区和数据源

### 1.1 研究区概况

研究区位于长江三峡库区的湖北省秭归至巴东段(图 1), 面积约为 408 km<sup>2</sup>, 为中低山地形和侵蚀峡谷地貌, 山地占 74%, 丘陵占 21.7%, 平地仅占 4.3%; 地势总体海拔为东高西低, 高程为 80 ~

2000 m, 沿长江往南、北方向总体呈变高趋势, 形成以秭归盆地为中心向周围增高的地貌形态。研究区位于大巴山断褶带、川鄂湘黔隆起褶皱带和川东褶皱带三大构造单元的交汇处, 地质构造复杂, 褶皱发育多样。发育有九畹溪、仙女山、牛口、巴东断层等, 以及黄陵背斜、秭归向斜、香炉山背斜、百福坪背斜、韩家河—路口子背斜等(李景富, 2010; 彭令等, 2017)。研究区地层主要为沉积岩, 除部分组、段缺失之外, 志留系下统至三叠系下统均有出露, 主要岩性为灰岩、硅质岩、白云岩, 部分组、段为砂岩、粉砂岩(如罗惹坪组  $S_1lr^1$ ), 分布于黄陵背斜西翼, 链子崖到香溪河河口段; 三叠系巴东组和侏罗系砂岩、页岩、泥岩及灰岩、泥灰岩的地层岩性组合为秭归盆地(秭归向斜)的地层组成, 也是本区灾害发生的重要地层组合(郭培虹, 2010)。

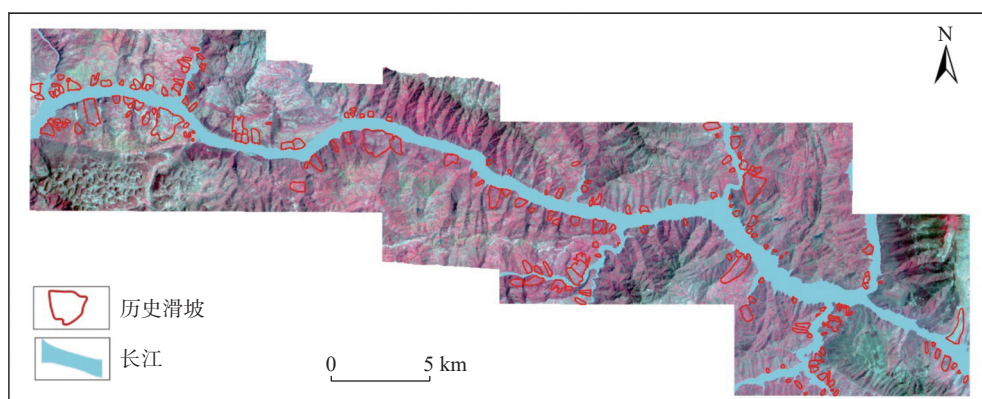


图 1 研究区地理位置及历史滑坡分布图

Fig. 1 Geographical location and historical landslide distribution map of the research area

研究区属北亚热带大陆性季风气候, 四季分明, 雨量充足, 平均年降水量 1000 ~ 1400 mm, 降水因季节不同差异较大, 5~9 月总降水量约占该区全年降水量的 70%。研究区的峡谷段地形险峻, 山高坡陡, 暖湿气流受地形的动力抬升作用易形成局部暴雨, 常引发山体滑坡等地质灾害(张坤等, 2020)。另外, 研究区人口密度高出全国平均人口密度约 3 倍, 人地矛盾突出, 人类的工程活动, 如高边坡、修路、建房、开垦等, 也加剧了区内滑坡灾害的危害。

### 1.2 数据源

考虑到滑坡事件、地质环境因素和空间位置

之间的关系, 结合所能收集的各类相关资料, 本文的数据源包括: (1) 研究区 H49E006010 巴东县幅、H49E006011 泄滩幅、H49E07010 风吹垭幅、H49E007011 秭归县幅、H49E007012 新滩幅 1 : 5 万地质图, 以及 H-49-(09) 巴东县幅 1 : 20 万地质图, 用于获取斜坡结构、工程岩组类型; (2) 空间分辨率为 25 m 的数字高程模型(DEM), 来源于 1997 年版 1 : 5 万地形图, 图幅编号同上, 用于提取高程、坡向、坡度、地表曲率、地形起伏度等地形地貌因子; (3) 滑坡地质灾害详细调查数据来源于原三峡地质灾害防治工作指挥部, 用于获取历史滑坡的位置分布; (4) 空间分辨率为 30 m 的

Landsat-8 遥感影像, 来源于地理数据云 (<https://www.gscloud.cn>), 用于获取土地利用、植被覆盖、土壤相对湿度、库水分布信息; (5) 气象站观测数据, 来源于中国国家地球系统科学数据中心 (<http://www.geodata.cn>), 用于获取年平均降雨信息。

## 2 模型与方法

首先基于因子特征分析方法确定参与本次滑坡易发性评价的评价因子; 然后采用信息量模型计算了各个评价指标的信息量; 接着通过随机森林算法确定了各评价指标的重要性, 并将因子重

要性值作为客观权重, 对各评价因子的信息量进行加权叠加; 最后, 对单一模型和组合模型的评价效果进行了对比分析。技术流程(图 2)主要分为以下六个步骤: (1) 数据获取: 收集研究区内地质环境相关因素, 建立基本环境因子的地理空间数据库; (2) 特征提取: 利用评价因子特征分析方法筛选出合适的评价因子, 并结合各自特征进行分类分级; (3) 样本选择与划分: 确定滑坡样本与非滑坡样本, 确定训练集和测试集; (4) 模型评价: 利用信息量模型和基于随机森林赋权信息量的模型, 分别对研究区进行滑坡灾害易发性评价; (5) 模型评估: 结合统计指数和 ROC 曲线评估两

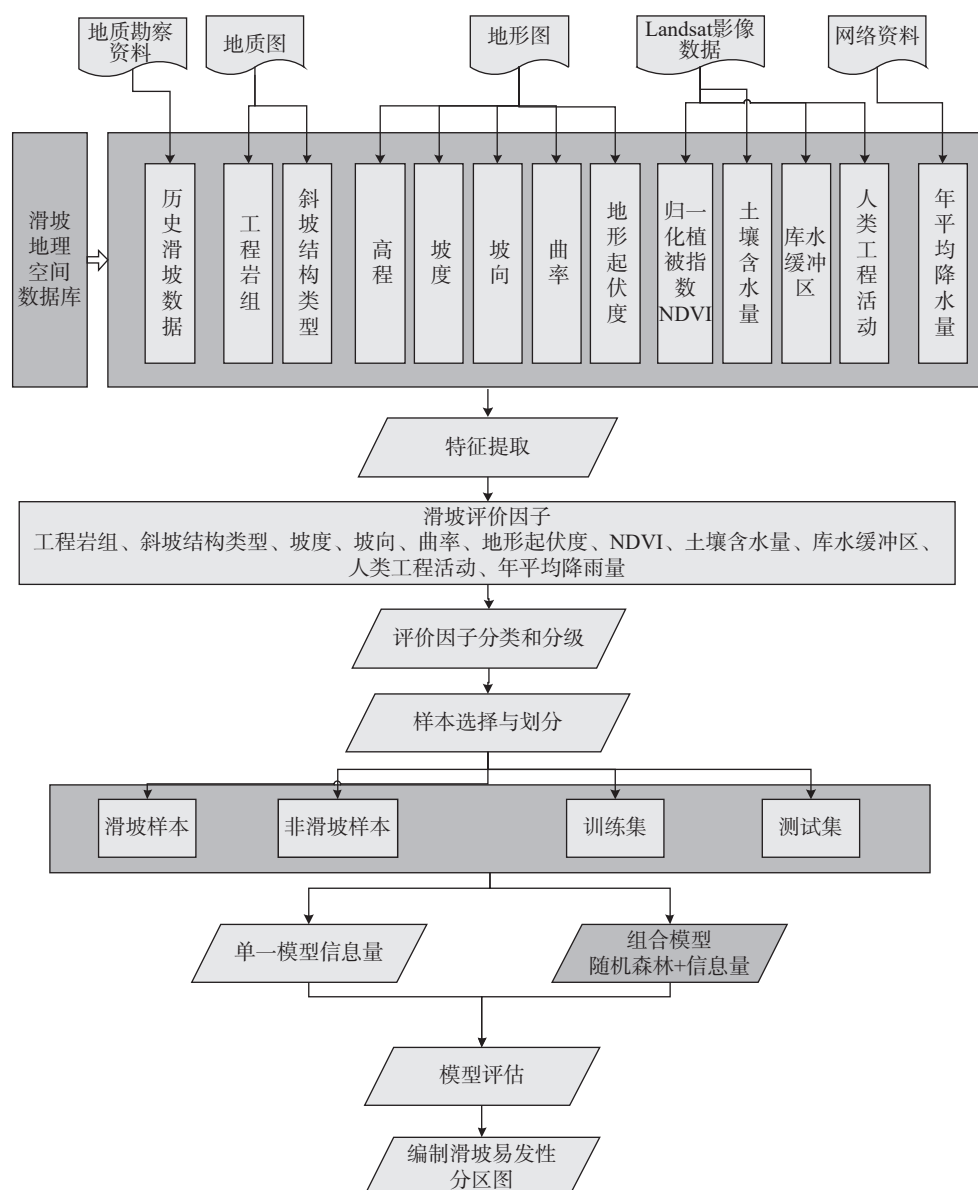


图 2 研究区进行滑坡易发性评价的技术流程图

Fig. 2 Technical flowchart for landslide susceptibility assessment in the research area

个模型的准确性; (6) 编制滑坡易发性分区图并对其进行评估。

## 2.1 评价模型

### 2.1.1 信息量模型

信息量模型是基于滑坡发生的基本环境因素与滑坡之间关联性的信息量计算。该模型的基本思想是通过比较滑坡发生的实际分布与预期分布之间的差异来评估每个环境因素对滑坡发生的贡献(Fell et al., 2008)。基本思路为, 滑坡( $y$ )的发生受众多地质环境因素  $X(x_1x_2\cdots x_n)$  的综合影响, 根据信息预测的理论, 滑坡灾害是否发生与所获取的地质环境因素信息的数量和质量有关, 通过计算每个环境因素的信息量值, 以此来表示该因素对滑坡发生的贡献值。信息量值越大, 表示该因素对滑坡发生的贡献越大(殷坤龙和朱良锋, 2001)。

$$I(y, x_1x_2\cdots x_n) = \log_2 \frac{P(y, x_1x_2\cdots x_n)}{P(y)} \quad (1)$$

式(1)中:  $I(y, x_1x_2\cdots x_n)$  为因素组合  $x_1x_2\cdots x_n$  对滑坡发生所提供的信息量,  $P(y, x_1x_2\cdots x_n)$  为因素  $x_1x_2\cdots x_n$  在组合条件下发生滑坡的概率。

区域滑坡灾害的易发性评价是基于对研究区域进行网格化处理的基础上展开的。假定某研究区域被划分为  $N$  个栅格单元。其中记录有滑坡灾害发生的栅格单元格数量为  $N_0$  个。研究区内具相同因素  $x_1x_2\cdots x_n$  组合的栅格单元共  $S$  个, 其中有  $S_0$  个栅格单元发生了滑坡灾害。依据统计概率作为先验概率的思路, 可以得知在该区域内, 因素  $x_1x_2\cdots x_n$  对滑坡灾害发生所贡献的信息量为:

$$I(y, x_1x_2\cdots x_n) = \log_2 \frac{S_0/S}{N_0/N} \quad (2)$$

转化为样本频率进行计算, 评价因子  $x_i$  对滑坡发生提供的信息量值  $I(x_i)$  可用下式计算:

$$I(x_i) = \ln \frac{N_i/N}{S_i/S} \quad (3)$$

式(3)中  $x_i$  表示评价单元内评价因子所在的等级;  $N_i$  为滑坡落在影响因素  $x_i$  内的单元数;  $N$  为研究区滑坡分布的单元总数;  $S_i$  为研究区内含有影响因素  $x_i$  的单元数,  $S$  为研究区评价单元的总数。

### 2.1.2 随机森林模型

本文涉及到 11 个不同的滑坡评价因子, 包

括: NDVI、年均降雨量、斜坡曲率和坡向等。量化这些滑坡评价因子的重要性对于理解因子之间的潜在关系、识别最具影响力的变量并构建高精度的滑坡易发性评价模型具有重要意义(Reichenbach et al., 2018)。传统随机森林(Classical Random Forest)是一种基于决策树的集成学习方法, 使用有放回抽样(bootstrap)方法对每一个决策树进行训练, 以达到分类或回归的目的。基于随机森林的滑坡评价因子重要性特征值评估(Variable importance measures, VIM)是一种较为常用的机器学习特征重要性计算方法。由于随机森林模型具有高预测率, 对异常值和噪声容忍度高, 高效处理高维数据等优势, 使得随机森林模型特别适用于高维数据研究, 即在进行多维数据分类的同时, 基于 VIM 值量化不同特征的重要性, 筛选出对于研究更重要的特征。

平均不纯度减少是传统随机森林算法中用于评估特征重要性的主要方法之一。在随机森林的构建过程中, 每个特征在决策树节点分裂时都会带来一定的不纯度减少, 通过计算特征在分割节点时减少的不纯度或导致的模型性能下降, 得出特征的重要性。这种不纯度可以通过基尼指数(Gini)来衡量。通过求解某特征在某个节点上分枝前后的基尼指数差值, 确定每个特征在每一棵决策树上所做的贡献量, 最后将某个特征基尼指数变化值除以所有特征基尼指数变化值之和, 求得某个特征归一化后的贡献量, 并根据贡献量大小进行排序, 即为每个特征的重要性(Yu L B et al., 2022)。计算方法如下:

$$Gini(p) = \sum_{k=1}^k p_k \cdot (1 - p_k) = 1 - \sum_{k=1}^k p_k^2 \quad (4)$$

式中,  $Gini(p)$  表示节点  $p$  的 Gini 指数,  $k$  表示有  $k$  个类别, 而  $p_k$  表示第  $k$  个类别的权重。特征  $j$  在节点  $m$  上基尼指数变化量通过计算该特征在该节点分枝前的基尼指数与分枝后基尼指数的差求得。设  $VIM_{jm}$  表示特征  $j$  在节点  $m$  上的基尼指数变化值,  $GI_m$  表示分枝前的基尼指数,  $GI_l$  和  $GI_r$  则为节点  $m$  分枝后产生的两个新节点的基尼指数。可得:

$$VIM_{jm} = GI_m - GI_l - GI_r \quad (5)$$

若特征  $j$  在决策树  $i$  中出现在节点集合  $M$  中, 则特征  $j$  在第  $i$  棵决策树上的基尼指数变化量:

$$VIM_{ij} = \sum_{m \in M} VIM_{jm} \quad (6)$$

若在随机森林中有  $n$  个决策树, 则特征  $j$  的总基尼指数变化量为:

$$VIM_j = \sum_{i=1}^n VIM_{ij} \quad (7)$$

则特征  $j$  贡献量归一化以后的值记为特征  $j$  的贡献量为:

$$VIM_j' = \frac{VIM_j}{\sum_{i=1}^c VIM_i} \quad (8)$$

$\sum_{i=1}^c VIM_i$  表示所有特征的基尼指数差之和,  $VIM_j'$  表示归一化以后得到特征  $j$  的贡献量。

### 2.1.3 加权信息量模型

线性加权法是一种常用的多属性决策分析方法, 它通过为不同的评价指标分配权重, 并将这些权重与各指标的评价值相乘, 从而得到一个综合评价值。该方法计算过程简单、易于理解和操作, 广泛应用于多种类型的多属性决策问题。公式如下:

$$y = \sum_{j=1}^m VIM_j' * I(x_j) \quad (9)$$

式(9)中,  $y$  为线性加权法的综合评价得分, 即综合信息量值;  $m$  为评价指标;  $VIM_j'$  为随机森林计算得到的特征  $j$  的贡献量, 即评价因子  $x_j$  相对应的权重系数,  $I(x_j)$  为  $x_j$  的信息量值。

## 2.2 因子特征分析方法

### 2.2.1 因子相关性分析

基本环境因子与滑坡的发生在一定程度上存在关联, 但各因子之间可能存在的相关性会对预测结果产生不利影响, 并可能增加模型的复杂性和运行时间。因此, 在评价之前需要对基本环境因子进行相关性分析, 以消除高相关性的因子, 提高模型效率和预测准确性。皮尔逊(Pearson)相关系数反映两个因子之间的相关程度, 是两个因素协方差除以标准差的乘积, 相关系数的绝对值为  $0 \sim 1$ , 若绝对值  $> 0.5$ , 则视为强相关变量。不少学者用其进行滑坡因子筛选并取得了不错的结果(尚敏等, 2021; 陈丹璐等, 2024; 何清等, 2024)。本文利用 SPSS Statistic 软件, 采用皮尔逊(Pearson)

相关系数进行相关性分析。

### 2.2.2 因子共线性分析

在滑坡基本环境因素的数据库中, 经常会观察到一种现象, 即某些因素之间具有很强的相关性, 特别是在地形因素中(Yu L et al., 2022)。这种多重共线性通常会导致模型性能下降。为了解决这个问题, 本文使用方差膨胀因子(VIF)和容忍度(Tolerance)来检测和量化基本环境因素中的多重共线性程度。VIF 是一个指数, 用于衡量由于多重共线性导致的估计回归系数方差的膨胀程度(Lin D Y et al., 2011)。容忍度(Tolerance)是 VIF 的倒数, 为多重共线性提供了另一种视角。通常, 当 VIF 大于 5 或容忍度小于 0.2 时, 表示这些因素之间存在多重共线性(Lin D Y et al., 2011)。预测变量  $j$  的 VIF 可以按以下方式计算:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (10)$$

$VIF_j$  是第  $j$  个自变量的方差膨胀因子,  $R_j^2$  是第  $j$  个自变量对所有其他自变量的回归判定系数, 也称为  $R$  平方值, 它表示在所有自变量中, 第  $j$  个自变量能够被其他自变量解释的比例。

### 2.3 模型评估方法

为了验证模型的性能和泛化能力, 本文验证了不同数据集中评估模型的拟合度(Tien et al., 2016)。训练数据集中的评估结果反映了滑坡模型与训练数据的拟合程度。测试数据集中的评估结果则反映了滑坡模型的预测能力(Tien et al., 2016)。本文通过使用基于统计指标和 ROC 曲线的评估方法, 对四种滑坡模型的性能进行了评价和比较。

#### 2.3.1 基于统计指数的评估

在统计学中, 有许多指标可以用来验证模型的性能(Bennett et al., 2013)。在本文中, 我们使用了阳性预测值、阴性预测值、灵敏度(SST)、精度(ACC)、F1 分数(F1 score)(表 1)等统计指标来验证滑坡模型的性能。

TP(True Positive)代表真阳性, 指的是被准确识别为滑坡的栅格数量; TN(True Negative)代表真阴性, 指的是被准确识别为非滑坡点的栅格数量; FN(False Negative)代表假阴性, 指的是实际为

表 1 统计学中的统计指数描述

Table 1 Description of statistical indices in statistics

| 序号 | 统计指数名称         | 公式                              | 描述                                  |
|----|----------------|---------------------------------|-------------------------------------|
| 1  | 灵敏度(SST)       | $SST=TP/(TP+FN)$                | 表示正确归类为“滑坡”的滑坡单元数百分比                |
| 2  | 精度(ACC)        | $ACC=(TP+TN)/(TP+TN+FP+FN)$     | 表示正确划分为“滑坡”和“非滑坡”的单元数占总单元数的比例       |
| 3  | F1分数(F1 score) | $F1\text{分数}=(2TP/(2TP+FP+FN))$ | 灵敏度和精度的调和平均数, 平衡模型在识别“真阳性”和“假阴性”的能力 |

滑坡点但被错误划分为非滑坡点的栅格数量; FP(False Positive)代表假阳性, 指的是实际为非滑坡点但被错误划分为滑坡点的栅格数量。

2.3.2 ROC 曲线

滑坡敏感性建模是一个二分类任务, 其特点是目标变量只有两个可能的取值(滑坡或非滑坡)。ROC 曲线(Receiver Operating Characteristic Curve)是广泛用于评价分类模型性能的工具, 尤其适用于二分类问题。它通过在不同阈值下计算模型的真阳性率(True Positive Rate, TPR)和假阳性率(False Positive Rate, FPR)来全面评估分类器的表现(Merghadi et al., 2020)。真阳性率(TPR)表示所有实际为阳性的样本中被正确识别为阳性的比例; 假阳性率(FPR)表示所有实际为阴性的样本中被错误识别为阳性的比例。ROC 曲线越接近左上角, 模型的性能就越好。AUC 是根据 ROC 曲线计算得出的, 用于量化模型的总体性能。AUC 值的范围从 0 到 1, 其中较高的值表示更好的区分能力和更准确的模型。根据 Sreenivas and Venkataratnam (2005)的研究, AUC 可以分为五个等级: 差 [0.5, 0.6]、一般 (0.6, 0.7]、好 (0.7, 0.8]、非常好 (0.8, 0.9] 和优秀 (0.9, 1]。

3 过程与结果分析

3.1 训练样本的选择

本次研究区总面积为 408 km<sup>2</sup>, 历史滑坡

204 处, 约占研究区面积的 5.9%。根据研究区大小和滑坡规模, 采用 20 m×20 m 的网格作为滑坡易发性评估的基本栅格单元。设定当前已发生滑坡灾害的区域为“危险区”, 尚未发生滑坡灾害的区域为“稳定区”。将 204 处滑坡均作为正样本, 从稳定区随机选取近等量栅格单元作为负样本。并按 7 : 3 的比例将正、负样本划分为训练集和测试集。危险区和稳定区样本栅格单元数量如表 2 所示。

3.2 评价因子的分析

3.2.1 滑坡影响因素分析

影响斜坡稳定性的因素分为内在因素和外在因素两大类。内在因素包括斜坡岩土类型及性质、地质结构等, 这些因素对斜坡变形破坏的形式和规模及稳定性起着控制作用(张泽方等, 2023; 闫举生等, 2024)。斜坡变形破坏多数受结构面控制, 结构面的强度和斜坡结构类型都是影响斜坡稳定性的重要因素。软弱岩层由于强度较低因而更易发生滑动, 不同岩性的接触面也极易成为潜在的滑动面(刘春玲等, 2010; 常宏, 2022)。而当岩层倾向与斜坡倾向一致而岩层倾角小于斜坡倾角(顺向坡)状态时, 最易沿层面形成滑坡(刘中楠等, 2022)。地形地貌条件也是滑坡发生的基本条件之一, 统计结果表明, 滑坡主要发生在 20° ~ 45°的斜坡上, 大于 45°的斜坡多发生崩塌而少滑坡, 低于 20°的斜坡则很少发生滑坡(汪美华等, 2023; 郭飞等, 2024)。

表 2 三峡库区秭归-巴东段滑坡危险区与稳定区样本栅格数

Table 2 Sample grid number of landslide danger and stability zones in the Zigui-Badong section of the Three Gorges Reservoir area

| 危险区样本栅格单元(个) |           |        | 稳定区样本栅格单元(个) |           |        |
|--------------|-----------|--------|--------------|-----------|--------|
| 训练样本(70%)    | 测试样本(30%) | 总样本数   | 训练样本(70%)    | 测试样本(30%) | 总样本数   |
| 41 669       | 17 770    | 59 439 | 42 000       | 18 000    | 60 000 |

外在因素则包括水的作用、风化作用、人为因素等,这些因素所起的破坏作用促进了斜坡变形的发生和发展(宫凤强等, 2007; 彭建兵等, 2020; 廖德武等, 2022)。张婷婷等(2012)、欧光照等(2016)采用因子分析法得出库水位是影响滑坡变形的的主要外在因素、降雨量是次要因素的结论。人类工程活动会改变原有地形地貌、增加地表荷载,从而导致斜坡稳定性下降(张志兼等, 2022; 周洪福等, 2023), 植被对保持土壤结构稳定起着重要作用(杨正荣等, 2023; 赵旭等, 2024)。

3.2.2 评价因子的获取和量化

本文利用研究区 1 : 50000 地形图获取了 25 m×25 m 分辨率的 DEM,通过 ARCGIS 进行分析提取和重采样,生成了高程、坡度、坡向、曲率、地形起伏度 5 个地形因子作为基本评价因子。从研究区 1 : 50000、1 : 200000 地质图中获取了工程岩组、斜坡结构 2 类基本评价因子。利用气象观测数据和 Landsat 影像获取年均降雨量、土壤含水量及库水缓冲区(等级)作为评估研究区水文状况的初步指标。植被覆盖度(NDVI)、土地利用类型作为反映人类工程活动影响的评价指标。

结合各评价因子连续或离散的自身特点,对高程、坡度、坡向、曲率、地形起伏度、NDVI、土

壤含水量、降雨量 8 个连续型评价因子采用自然断点法划分为不同的属性区间。离散型评价因子结合实际进行等级划分:根据岩性特征,将地层单元分为软岩、软硬相间、硬岩三个工程岩组。根据坡度、坡向、下伏地层倾向、倾角的空间位置关系,将斜坡结构类型分为 6 类:飘倾坡、伏倾坡、顺斜坡、横向坡、逆斜坡、逆向坡(陈涛等, 2020)。土地利用类型分为居民地/道路、植被、水体、裸土、裸岩。库水缓冲区因子以斜坡到库岸的垂直距离划分为波动区 [0 m, 50 m]、强影响区 (50 m, 300 m]、中影响区(300 m, 550 m]、弱影响区 (550 m, 2000 m]4 个等级。

3.2.3 评价因子特征分析

利用 SPSS Stastic 软件对 12 个特征因素进行相关性分析和共线性分析,结果见表 3、表 4。由表 3 可见,高程与库水缓冲区的相关性大于 0.5,显示为强相关性,其它因子的相关性值在 -0.397 ~ 0.387 之间,无明显相关关系。由表 4 可见,高程因素的 *VIF* 值为 5.741(大于 5),其余因子 *VIF* 值在 1.009 ~ 3.502 之间,这表明,除高程外的剩余 11 个因子可以独立作为滑坡易发性评价的特征变量。研究区 11 个评价因子的分级分类图见图 3。

3.2.4 评价因子的重要性

本次模型代码基于 python 语言编写,使用

表 3 研究区滑坡初始评价因子相关性表

Table 3 Correlation Table of Initial Evaluation Factors for Landslides in the Research Area

|        | NDVI    | 地形起伏度   | 高程      | 坡度      | 土壤含水量   | 年均降雨量   | 库水缓冲区   | 土地利用类型  | 曲率      | 坡向      | 工程岩组    | 斜坡结构    |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| NDVI   | 1       | -.138** | -.181** | -.038** | .166**  | -.086** | -.068** | 0.000   | -.034** | -.111** | -.067** | .020**  |
| 地形起伏度  | -.138** | 1       | .361**  | .373**  | -.353** | -.025** | .387**  | -.196** | -.008** | .038**  | .093**  | .034**  |
| 高程     | -.181** | .361**  | 1       | .187**  | -.287** | .061**  | .792**  | -.220** | .061**  | .025**  | .246**  | 0.004   |
| 坡度     | -.038** | .373**  | .187**  | 1       | -.098** | -.157** | .144**  | -.250** | .035**  | .035**  | .055**  | .118**  |
| 土壤含水量  | .166**  | -.353** | -.287** | -.098** | 1       | -.033** | -.173** | .236**  | -.009** | -.397** | -.141** | -.031** |
| 年均降雨量  | -.086** | -.025** | .061**  | -.157** | -.033** | 1       | -.186** | .083**  | -.015** | -.055** | .081**  | -.033** |
| 库水缓冲区  | -.068** | .387**  | .792**  | .144**  | -.173** | -.186** | 1       | -.161** | .041**  | .022**  | .144**  | -.015** |
| 土地利用类型 | 0.000   | -.196** | -.220** | -.250** | .236**  | .083**  | -.161** | 1       | -.032** | -.108** | -.117** | .011**  |
| 曲率     | -.034** | -.008** | .061**  | .035**  | -.009** | -.015** | .041**  | -.032** | 1       | .007*   | 0.005   | .012**  |
| 坡向     | -.111** | .038**  | .025**  | .035**  | -.397** | -.055** | .022**  | -.108** | .007*   | 1       | .010**  | .076**  |
| 工程岩组   | -.067** | .093**  | .246**  | .055**  | -.141** | .081**  | .144**  | -.117** | 0.005   | .010**  | 1       | -.168** |
| 斜坡结构   | .020**  | .034**  | 0.004   | .118**  | -.031** | -.033** | -.015** | .011**  | .012**  | .076**  | -.168** | 1       |

\*\* . 在 0.01 级别(双尾), 相关性显著; \* . 在 0.05 级别(双尾), 相关性显著。

表 4 研究区滑坡初始评价因子共线性表

Table 4 Collinearity table of initial evaluation factors for landslides in the research area

| 特征因素  | 容差    | VIF指数 | 特征因素   | 容差    | VIF指数 |
|-------|-------|-------|--------|-------|-------|
| 年均降雨量 | 0.286 | 3.502 | 土地利用类型 | 0.809 | 1.237 |
| 坡度    | 0.797 | 1.255 | 坡向     | 0.767 | 1.304 |
| 库水缓冲区 | 0.319 | 3.137 | 高程     | 0.174 | 5.741 |
| NDVI  | 0.86  | 1.162 | 地形起伏度  | 0.756 | 1.323 |
| 曲率    | 0.991 | 1.009 | 工程岩组   | 0.808 | 1.237 |
| 土壤含水量 | 0.638 | 1.566 | 斜坡结构   | 0.938 | 1.067 |

注: 容差数值与VIF指数数值互为倒数.

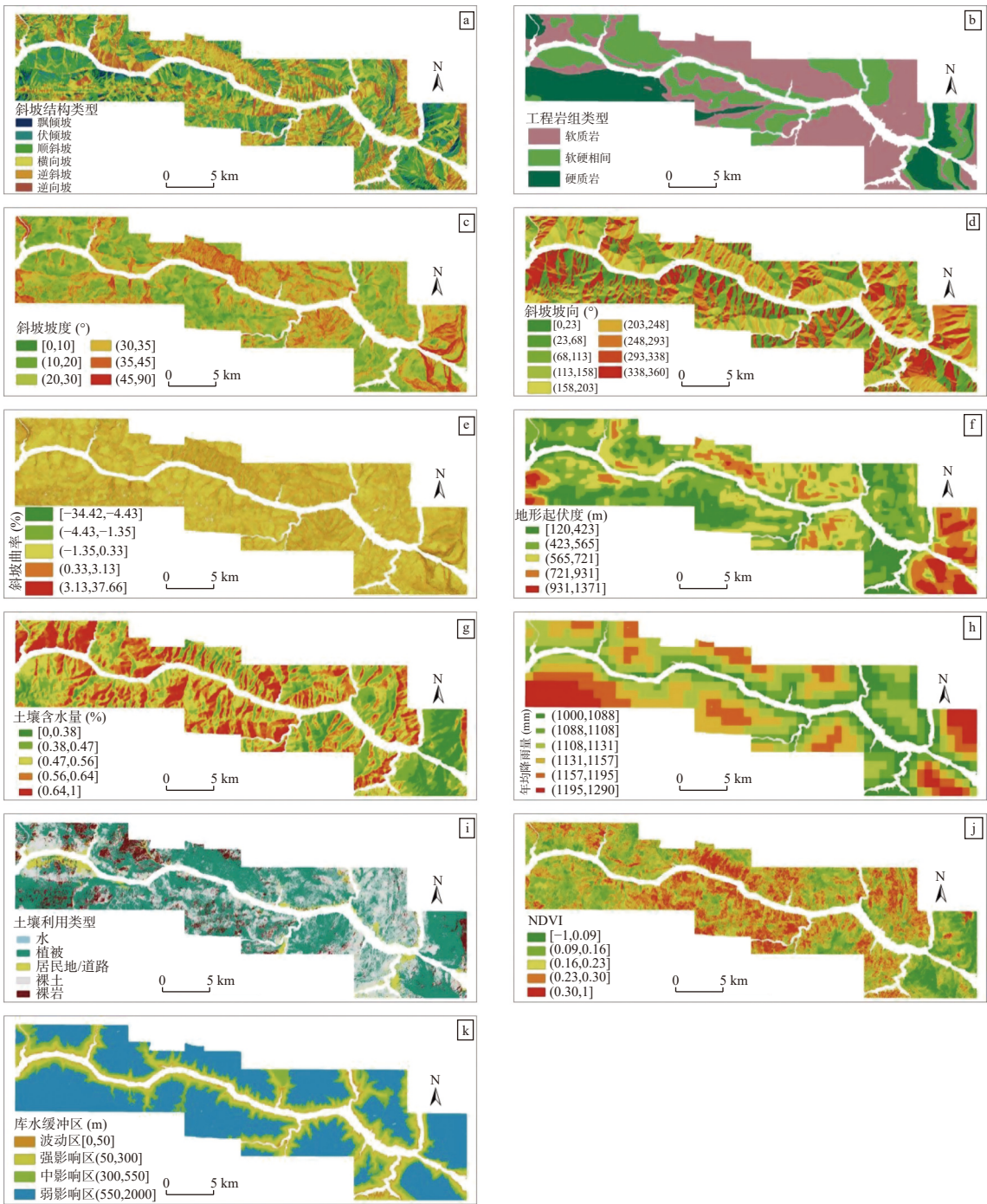


图 3 三峡库区秭归-巴东段滑坡评价因子分级图

Fig. 3 Grading Map of Landslide Evaluation Factors in the Zigui-Badong Section of the Three Gorges Reservoir Area

scikit-learn 库的 RandomForestClassifier 函数,并调用模型的“feature\_importance”属性来实现评价因子的重要性计算。该模型中 n\_estimators 参数设置为 10000。训练过程中训练集和验证集的数据划分为 70% 训练集、30% 验证集。共使用 83 669 条数据,涉及 11 个特征参数。由于随机森林模型中不同的参数设置会对模型的评价精度产生影响,因此在本次研究中基于 Sklearn 库中的 GridSearchCV 方法,通过在指定的参数范围内,按照步长依次调整参数,从而找出精度最高的参数,给出最优化结果。

计算的 11 个滑坡倾向性特征的重要性值如表 5 所示,特征的重要性值越高,表示该因素对滑坡发生的影响越大。可以看出,这 11 个影响因素对研究区内不同类型滑坡的影响各不相同,其中库水缓冲区因子位居首列,这与研究区历史滑坡沿库水分布的特征完全一致。

表 5 研究区滑坡评价因子重要性表  
Table 5 The importance value of landslide causal factors in the research area

| 评价因子  | 重要性   | 评价因子 | 重要性   |
|-------|-------|------|-------|
| 库水缓冲区 | 0.278 | 土地利用 | 0.053 |
| 年均降雨量 | 0.259 | NDVI | 0.036 |
| 地形起伏度 | 0.084 | 工程岩组 | 0.035 |
| 坡度    | 0.074 | 曲率   | 0.028 |
| 坡向    | 0.070 | 斜坡结构 | 0.025 |
| 土壤含水量 | 0.060 |      |       |

3.3 模型的预测和评估

使用训练集构建了信息量模型和加权信息量模型,并分别对这两个模型在训练集和测试集上的性能做了统计分析(表 6)。结果表明,两个模型都表现出了较高的滑坡易发性评价能力,在训练集中,信息量模型的灵敏度(SST)为 0.836,精度(ACC)为 0.767,F1 分数(F1 score)为 0.770,加权信息量的灵敏度(SST)为 0.865,精度(ACC)为 0.791,F1 分数(F1 score)为 0.803。测试集中信息量模型的灵敏度(SST)为 0.808,精度(ACC)为 0.785,F1 分数(F1 score)为 0.787,加权信息量的灵敏度(SST)为 0.874,精度(ACC)为 0.785,F1 分数(F1 score)为 0.800(表 7)。由统计指标可见,加

表 6 研究区滑坡评估参数统计表  
Table 6 Landslide assessment parameters in the research area

| 评估参数  | 训练集    |        | 测试集    |        |
|-------|--------|--------|--------|--------|
|       | 信息量    | 加权信息量  | 信息量    | 加权信息量  |
| TP(个) | 31 548 | 34 361 | 13 807 | 14 934 |
| FP(个) | 12 670 | 11 518 | 4 186  | 5 297  |
| TN(个) | 30 394 | 29 546 | 13 442 | 12 331 |
| FN(个) | 6 183  | 5 370  | 3 285  | 2 158  |

表 7 研究区滑坡预测结果精度表  
Table 7 Accuracy of landslide prediction results in the research area

| 评估参数            | 训练集   |       | 测试集   |       |
|-----------------|-------|-------|-------|-------|
|                 | 信息量   | 加权信息量 | 信息量   | 加权信息量 |
| 灵敏度(SST)        | 0.836 | 0.865 | 0.808 | 0.874 |
| 精度(ACC)         | 0.767 | 0.791 | 0.785 | 0.785 |
| F1 分数(F1 score) | 0.770 | 0.803 | 0.787 | 0.800 |
| AUC             | 0.791 | 0.863 | 0.778 | 0.855 |

权信息量模型均优于信息量模型。

使用训练集和测试集绘制的 ROC 曲线如图 4 所示。传统信息量模型的训练集和测试集 AUC 分别为 0.791 和 0.778,而加权信息量模型训练集和测试集的 AUC 分别为 0.863 和 0.855(表 7)。ROC 曲线分析结果进一步表明,利用加权信息量模型进行滑坡易发性评价的精度得到了明显提高。

3.4 模型验证和比较

利用自然断点法对两类易发性评价模型生成的滑坡敏感性地图进行等级划分,生成滑坡易发性分区图。结合前人的分类经验,将研究区的滑坡易发性划分为五个等级:极高易发区、高易发区、中易发区、低易发区、极低易发区(图 5、图 6)。从图中可以看出,两个模型的评价结果在空间分布上总体相似,历史滑坡大多位于高易发区和极高易发区。通过将评价结果与 11 个滑坡因子(图 3)进行叠加比对,可以发现极高易发区、高易发区主要分布在距库水 300 m 以内、年均降雨量在 1 100 mm 以上、土壤含水量在 50% 以上的软质岩区域,这种情况与滑坡分布的规律基本一致。这表明所建立的两个滑坡模型均能够较好地反映该地区的滑坡易发性分布情况,并都具有一定的预测能力。

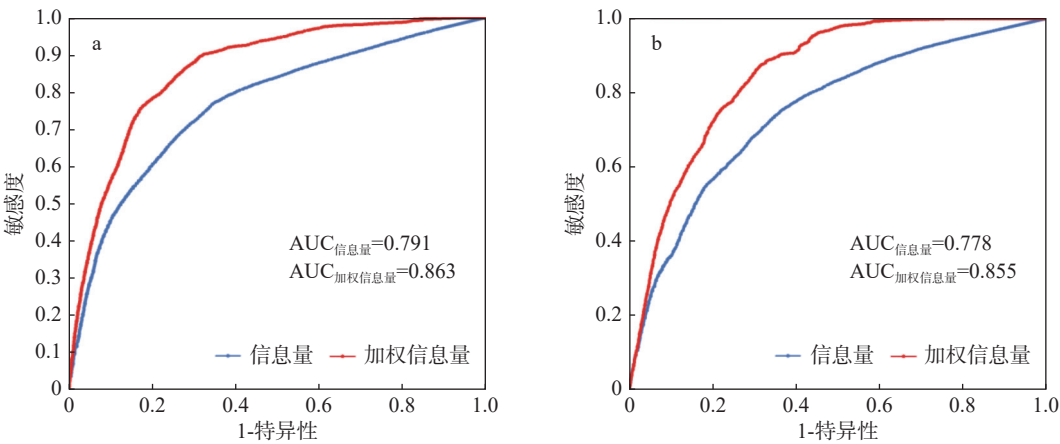


图 4 研究区滑坡训练集 (a) 和测试集 (b) ROC 曲线图

Fig. 4 ROC curves of landslide training set (a) and test set (b) in the research area

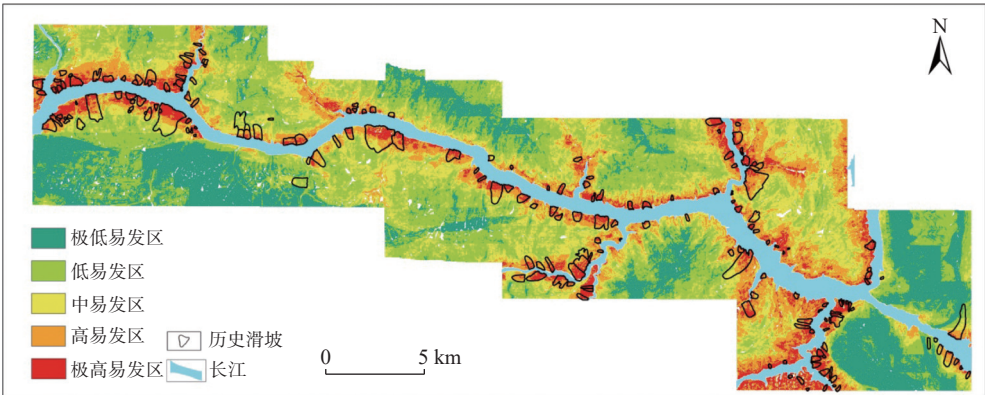


图 5 研究区基于信息量模型的滑坡易发性分区图

Fig. 5 Landslide susceptibility zoning map based on information model in the research area

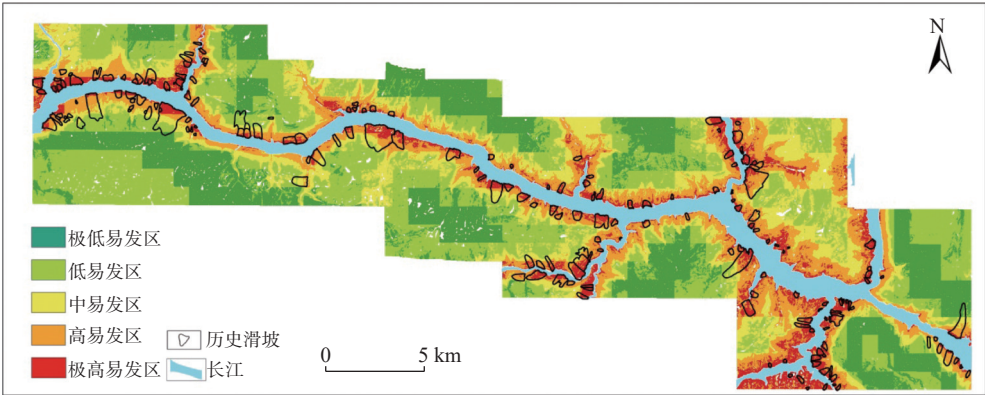


图 6 研究区基于随机森林赋权信息量模型的滑坡易发性分区图

Fig. 6 Landslide susceptibility zoning map based on random forest weighting information model in the research area

表 8 展示了两类模型中极低、低、中、高和极高易发区域的滑坡、栅格单元的数量和百分比。其中,以滑坡中分类等级所占总面积最大者代表该滑坡的分类级别。可以看出研究区的易发性评价结果主要以低级别为主,在两种模型中分别占比 33.8% 和 32.63%,其次是中、极低级别,而极高

级别的易发性占比最少,在两种模型中分别占 6.19% 和 7.26%。可见两类模型的易发性评价呈现出相似的分布规律,且与滑坡发生的实际概率情况相吻合。

从滑坡测试样本中随机选取了 13 个滑坡样本进行检验,分别统计 5 个易发性等级在每个检

验样本中的栅格单元个数,其详细统计数据见表 9 所示。从表 9 中可以看出,基于随机森林加权的信  
息量模型在 13 个检验样本的易发性评价分区结果中,极高易发区、高易发区在加权信息量模型中占比分别为 36.74%、49.34%,而传统信息量模型分别为 34.90% 和 47.26%,可见加权信息

表 8 研究区滑坡评价模型分类统计表

| Table 8 Classification and statistics of landslide evaluation models in the research area |      |         |           |         |           |
|-------------------------------------------------------------------------------------------|------|---------|-----------|---------|-----------|
| 模型                                                                                        | 分类级别 | 滑坡数量(个) | 滑坡数量占比(%) | 栅格数量(个) | 栅格数量占比(%) |
| 传统信息量                                                                                     | 极低   | 0       | 0         | 211 104 | 21.80     |
|                                                                                           | 低    | 2       | 0.98      | 327 357 | 33.80     |
|                                                                                           | 中    | 15      | 7.35      | 207 648 | 21.44     |
|                                                                                           | 高    | 100     | 49.02     | 162 513 | 16.78     |
|                                                                                           | 极高   | 87      | 42.65     | 59 927  | 6.19      |
| 随机森林<br>加权信息量                                                                             | 极低   | 0       | 0         | 223 119 | 23.04     |
|                                                                                           | 低    | 1       | 0.49      | 316 079 | 32.63     |
|                                                                                           | 中    | 12      | 5.88      | 182 465 | 18.84     |
|                                                                                           | 高    | 98      | 48.04     | 176 554 | 18.23     |
|                                                                                           | 极高   | 93      | 45.59     | 70 332  | 7.26      |

表 9 研究区已知滑坡的预测结果可靠性检验

| Table 9 Reliability testing of prediction results for known landslides in the research area |          |       |         |       |      |       |       |       |
|---------------------------------------------------------------------------------------------|----------|-------|---------|-------|------|-------|-------|-------|
| 序号                                                                                          | 滑坡名称     | 模型    | 总栅格数(个) | 极低(个) | 低(个) | 中(个)  | 高(个)  | 极高(个) |
| 1                                                                                           | 黄腊石      | 传统信息量 | 250     | 0     | 2    | 87    | 134   | 27    |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 32    | 158   | 60    |
| 2                                                                                           | 谭家坪      | 传统信息量 | 288     | 0     | 0    | 0     | 77    | 211   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 151   | 137   |
| 3                                                                                           | 谭石爬      | 传统信息量 | 323     | 0     | 0    | 62    | 229   | 32    |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 112   | 174   | 37    |
| 4                                                                                           | 青干河桥头白果树 | 传统信息量 | 233     | 0     | 0    | 42    | 187   | 4     |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 199   | 34    |
| 5                                                                                           | 楚王井      | 传统信息量 | 936     | 0     | 2    | 77    | 480   | 377   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 145   | 326   | 465   |
| 6                                                                                           | 观音阁      | 传统信息量 | 92      | 0     | 0    | 0     | 4     | 88    |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 3     | 89    |
| 7                                                                                           | 张家湾      | 传统信息量 | 249     | 0     | 0    | 2     | 117   | 130   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 99    | 150   |
| 8                                                                                           | 筲箕洼西滑体   | 传统信息量 | 531     | 0     | 41   | 175   | 184   | 131   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 346   | 185   |
| 9                                                                                           | 谭家湾      | 传统信息量 | 183     | 4     | 167  | 12    | 0     | 0     |
|                                                                                             |          | 加权信息量 |         | 0     | 9    | 151   | 23    | 0     |
| 10                                                                                          | 台子湾东     | 传统信息量 | 253     | 0     | 3    | 95    | 155   | 0     |
|                                                                                             |          | 加权信息量 |         | 0     | 10   | 94    | 149   | 0     |
| 11                                                                                          | 白水河      | 传统信息量 | 120     | 0     | 0    | 0     | 53    | 67    |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 50    | 70    |
| 12                                                                                          | 泄滩老镇     | 传统信息量 | 719     | 0     | 0    | 20    | 351   | 348   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 65    | 385   | 269   |
| 13                                                                                          | 老坟园滑坡体   | 传统信息量 | 262     | 0     | 0    | 1     | 127   | 134   |
|                                                                                             |          | 加权信息量 |         | 0     | 0    | 0     | 127   | 135   |
| 合计(个)                                                                                       |          | 传统信息量 | 4 439   | 4     | 215  | 573   | 2 098 | 1 549 |
|                                                                                             |          | 加权信息量 |         | 0     | 19   | 599   | 2 190 | 1 631 |
| 百分比(%)                                                                                      |          | 传统信息量 | 100     | 0.09  | 4.84 | 12.91 | 47.26 | 34.90 |
|                                                                                             |          | 加权信息量 |         | 0.00  | 0.43 | 13.49 | 49.34 | 36.74 |

量模型的易发性评价精度更高。

## 4 结论

本文以三峡库区秭归至巴东段为研究区域, 针对传统信息量模型中各评价因子权重一致的问题, 提出了一种基于随机森林赋权信息量模型的滑坡灾害易发性评价方法。通过特征因子分析筛选了库水缓冲区、降雨量、地形起伏度、坡度、坡向、土壤含水量、土地利用类型、NDVI、工程岩组、曲率、斜坡结构等 11 个评价因子, 利用随机森林模型计算评价因子的重要性并赋权信息量模型, 并分别利用传统信息量模型和加权信息量模型进行了滑坡易发性评价和对比, 得出如下结论:

(1) 研究区内滑坡的统计指标结果表明: 两个模型在滑坡易发性评价方面均表现出良好的性能。ROC 曲线中, 传统信息量模型的训练集和测试集 AUC 分别为 0.791 和 0.778, 而加权信息量模型训练集和测试集的 AUC 分别为 0.863 和 0.855。在训练集中, 信息量模型和加权信息量模型的灵敏度(SST)分别为 0.836、0.865, 精度(ACC)分别为 0.767、0.791, F1 分数(F1 score)分别为 0.770、0.803。测试集中灵敏度(SST)分别为 0.808、0.874, 精度(ACC)分别为 0.785、0.785, F1 分数(F1 score)分别为 0.787、0.800, 可见加权信息量模型在统计指标上均优于传统信息量模型。

(2) 随机选取研究区内 13 个滑坡测试样本对评价结果进行了统计检验, 检验结果显示: 传统信息量模型的评价结果中极高易发区、高易发区占比分别为 34.90% 和 47.26%, 而加权信息量模型中这两个比值分别为 36.74%、49.34%, 可见加权信息量模型的易发性评价精度更高。

(3) 两个模型的滑坡易发性分区图显示了滑坡相似的空间分布特征。滑坡易发性主要以低等级别为主, 其次是中、极低级别, 而极高级别的易发性最低; 历史滑坡大多位于高易发区和极高易发区, 这表明所建立的滑坡易发性评价模型均能够较好地反映该地区的滑坡易发性分布情况, 并都具有一定的预测能力。其中, 极高易发区和高易发区主要分布在研究区库水两岸、降雨量和土

壤含水量均较高的软质岩区域, 这反映了滑坡易发性分布规律与基本地理环境的一致性, 尤其与库水缓冲区分布呈强对应关系, 这与随机森林模型计算的因子重要性结果一致。

## 参考文献:

- 常 宏. 2022. 三峡库区潜在顺层岩质滑坡识别的宏观判据 [J]. 华南地质, 38(2): 265-272.
- 陈丹璐, 孙德亮, 文海家, 辜庆渝. 2024. 基于不同因子筛选方法的 Light GBM-SHAP 滑坡易发性研究 [J]. 北京师范大学学报 (自然科学版), 60(1): 148-158.
- 陈 涛, 钟子颖, 牛瑞卿, 刘 桐, 陈胜云. 2020. 利用深度信念网络进行滑坡易发性评价 [J]. 武汉大学学报 (信息科学版), 45(11): 1809-1817.
- 丁 丽, 郭 凯, 张郝哲. 2023. 基于不同评价单元的地质灾害易发性评价对比 [J]. 云南师范大学学报 (自然科学版), 43(2): 62-68.
- 杜 鹏, 陈宁生, 伍康林, 李 志, 张瀛玉龙. 2024. 基于随机森林模型的藏东南地区滑坡易发性评价及主控因素分析 [J]. 成都理工大学学报 (自然科学版), 51(2): 328-344.
- 官凤强, 李夕兵, 左宇军, 付玉华. 2007. 基于 DDA 方法的滑坡稳定性预测 [J]. 吉首大学学报 (自然科学版), 28(1): 73-76.
- 郭 飞, 赖 鹏, 黄发明, 刘磊磊, 王秀娟, 何政宇. 2024. 基于知识图谱的滑坡易发性评价文献综述及研究进展 [J]. 地球科学, 49(5): 1584-1606.
- 郭培虹. 2010. 基于遥感的岩土质岸坡空间分布研究——以三峡库区秭归—巴东段为例 [D]. 中国地质大学 (武汉) 硕士学位论文.
- 何 清, 李丽琳, 林子安. 2024. 基于 FEEMD-GRU-FC 模型的滑坡位移预测 [J]. 人民长江, 55(7): 108-114.
- 黄发明, 曹中山, 姚 池, 姜清辉, 陈佳武. 2021. 基于决策树和有效降雨强度的滑坡危险性预警 [J]. 浙江大学学报 (工学版), 55(3): 472-482.
- 李景富. 2010. 区域滑坡空间预测的数据挖掘方法探索 [D]. 中国地质大学 (武汉) 硕士学位论文.
- 李 璐, 瞿 伟, 张 勤, 李久元, 王宇豪, 刘祥斌. 2022. 优化循环神经网络在滑坡位移预测中的应用 [J]. 大地测量与地球动力学, 42(6): 594-600.
- 李 明, 蒋委君, 董佳慧, 金少锋, 张宸伟, 牛瑞卿. 2023. 基于机器学习的滑坡灾害易发性评价——以三峡库区为例 [J]. 华南地质, 39(3): 413-427.
- 李泽芝, 王新刚. 2024. 镇域尺度下秦巴山区堆积层滑坡易

- 发性不同单元评价性能对比研究[J]. 西北地质, 57(1):1-11.
- 连志鹏, 厉一宁, 刘磊, 王宁涛. 2022. 基于 ArcGIS 的湖北省远安县地质灾害风险定量评价[J]. 华南地质, 38(4):680-688.
- 廖德武, 郑冰, 杜艳松, 张君恺, 兰中孝, 吴正超. 2022. 兴仁“6·10”彭家洞高速滑坡运动特征与形成机理[J]. 地质科技通报, 41(6):66-76.
- 林振, 卢书强, 梅军. 2024. 基于信息量法的湖北省秭归县滑坡易发性评价[J]. 华南地质, 40(1):152-161.
- 刘春玲, 祁生文, 童立强, 安国庆, 李小慧. 2010. 喜马拉雅山区重大滑坡灾害及其与地层岩性的关系研究[J]. 工程地质学报, 18(5):669-676.
- 刘中楠, 高银梅, 黄勇炜, 姜河, 颜智强, 李策. 2022. 湖南省新宁县斜坡地质灾害形成机理与成因模式[J]. 科技创新导报, 19(30):21-27.
- 欧光照, 吴益平, 张秋霞. 2016. 基于因子分析法的白水河滑坡影响因素分析[J]. 人民黄河, 38(2):99-102.
- 彭建兵, 王启耀, 庄建琦, 冷艳秋, 范仲杰, 王少凯. 2020. 黄土高原滑坡灾害形成动力学机制[J]. 地质力学学报, 26(5):714-730.
- 彭亮, 田浩, 白刚刚, 霍宇翔. 2021. 基于优化组合预测模型的滑坡发展趋势评价[J]. 广西大学学报(自然科学版), 46(5):1228-1235.
- 彭令, 徐素宁, 梅军军, 苏凤环. 2017. 地震滑坡高分辨率遥感影像识别[J]. 遥感学报, 21(4):509-518.
- 钱紫玲. 2023. 基于统计模型的黄土地震滑坡危险性评价[D]. 中国地震局兰州地震研究所硕士学位论文.
- 尚敏, 廖芬, 马锐, 刘昱廷. 2021. 白家包滑坡变形与库水位、降雨相关性定量化分析研究[J]. 工程地质学报, 29(3):742-750.
- 汪美华, 赵慧, 倪天翔, 余洋, 陈红旗. 2023. 近30年滑坡研究文献图谱可视化分析[J]. 中国地质灾害与防治学报, 34(4):75-85.
- 王高峰, 郭宁, 邓兵, 田运涛, 叶振南, 陈宗良, 吕凤兰, 高幼龙. 2021. 不同组合模型区域滑坡易发性及精度分析[J]. 西北地质, 54(2):259-272.
- 吴兴贵, 王宇栋, 王蓝婷, 丁梓逸. 2024. 加权信息量模型在云南澜沧县滑坡危险性评价中的应用[J]. 中国地质灾害与防治学报, 35(3):119-128.
- 肖海平, 万俊辉, 陈兰兰, 范永超, 陈磊. 2024. 加权信息量支持下融合 InSAR 形变特征的滑坡易发性评价[J]. 大地测量与地球动力学, 44(7):718-724.
- 闫举生, 谭建民, 章昱. 2024. 清江流域覃家田滑坡成因机制及稳定性分析[J]. 华南地质, 40(1):162-170.
- 杨正荣, 喜文飞, 史正涛, 杨志全, 肖波. 2023. 复杂植被山区滑坡蠕变与植被覆盖度关系研究[J]. 测绘科学, 48(1):157-165+175.
- 叶润青, 李士垚, 郭飞, 付小林, 牛瑞卿. 2020. 基于 RS 和 GIS 的三峡库区滑坡易发程度与土地利用变化的关系研究[J]. 工程地质学报, 29(3):724-733.
- 殷坤龙, 朱良峰. 2001. 滑坡灾害空间区划及 GIS 应用研究[J]. 地学前缘, 8(2):279-284.
- 余淙蔚, 柳侃, 殷杰, 余斌. 2022. 一种适用于逻辑回归模型评价浅层滑坡易发性的网格尺度划分方法——以2019年福建省三明市群发浅层滑坡为例[J]. 山地学报, 40(1):106-119.
- 张坤, 彭涛, 方家虎, 陈胜华, 罗云洪. 2020. 钟山区马落箐滑坡地质灾害成灾机理分析[J]. 西部资源, (5):88-90.
- 张婷婷, 晏鄂川, 胡显明, 范彬彬. 2012. 基于因子分析法的滑坡变形分析[J]. 长江科学院院报, 29(4):21-25.
- 张泽方, 钱志宽, 魏勇, 朱星, 王林均. 2023. 考虑最优影响因素组合的滑坡易发性评价——以水城区为例[J]. 科学技术与工程, 23(10):4091-4099.
- 张志兼, 黄勋, 蔡雨微, 傅镜羽, 朱悦, 杨锐, 韩超群. 2022. 三峡库区武隆段滑坡灾害驱动因子演变格局与人类活动的影响[J]. 中国地质灾害与防治学报, 33(3):39-50.
- 章昱, 王磊, 伏永朋, 刘亚磊, 黄皓. 2023. 基于斜坡单元与信息量法的丹江口库区典型流域地质灾害易发性评价[J]. 华南地质, 39(3):512-522.
- 赵萍, 赵思逸, 孙雨, 阮旭东, 王宁, 张树衡. 2024. 基于斜坡单元和语义分割的皖南地区滑坡灾害易发性评估[J]. 地质科学, 59(2):562-574.
- 赵晓燕, 谈树成, 李永平. 2021. 基于斜坡单元与组合赋权法的东川区地质灾害危险性评价[J]. 云南大学学报(自然科学版), 43(2):299-305.
- 赵旭, 范豪举, 许敬叔, 杜修力. 2024. 降雨条件下植被根系对非饱和土浅层滑坡的加固作用[J]. 北京工业大学学报, 50(9):1100-1110.
- 周洪福, 方甜, 夏晨皓, 冉涛, 徐如阁, 张景华. 2023. 工程扰动诱发川西杜米滑坡复活变形特征及机理分析[J]. 现代地质, 37(4):1044-1053.
- Bennett N D, Croke B F W, Guariso G, Guillaume J H A, Hamilton S H, Jakeman A J, Marsili-Libelli S, Newham L T H, Norton J P, Perrin C, Pierce S A, Robson B, Seppelt R, Voinov A A, Fath B D, Andreassian V. 2013.

- Characterising performance of environmental models[J]. *Environmental Modelling & Software*, 40: 1-20.
- Dai F C, Lee C F, Ngai Y Y. 2002. Landslide risk assessment and management: an overview[J]. *Engineering Geology*, 64(1): 65-87.
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage W Z. 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning[J]. *Engineering Geology*, 102(3-4): 99-111.
- Lin D Y, Foster D P, Ungar L H. 2011. VIF regression: A fast regression algorithm for large data[J]. *Journal of the American Statistical Association*, 106(493): 232-247.
- Medina V, Hürlimann M, Guo Z Z, Lloret A, Vaunat J. 2021. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale[J]. *Catena*, 201: 105213.
- Merghadi A, Yunus A P, Dou J, Whiteley J, ThaiPham B, Bui D T, Avtar R, Abderrahmane B. 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance[J]. *Earth-Science Reviews*, 207: 103225.
- Pham B T, Bui D T, Prakash I, Dholakia M B. 2015. Landslide Susceptibility Assessment at a Part of Uttarakhand Himalaya, India using GIS-based statistical Approach of Frequency Ratio Method[J]. *International Journal of Engineering Research and Technology*, 4(11): 338-344.
- Reichenbach P, Rossi M, Malamud B D, Mihir M, Guzzetti F. 2018. A review of statistically-based landslide susceptibility models[J]. *Earth-Science Reviews*, 180: 60-91.
- Sreenivas K, Venkataratnam L. 2005. A Spatially Distributed Event-based Model to Predict Sediment Yield[J]. *Journal of Spatial Hydrology*, 5(1): 1-19.
- Tien B D, Tuan T A, Klempe H, Pradhan B, Revhaug I. 2016. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree[J]. *Landslides*, 13(2): 361-378.
- Tofani V, Biscocchi G, Rossi G, Segoni S, D'Ambrosio M, Casagli N, Catani F. 2017. Soil characterization for shallow landslides modeling: a case study in the Northern Apennines (Central Italy)[J]. *Landslides*, 14(2): 755-770.
- Yu L B, Zhou C, Wang Y, Cao Y, Peres D J. 2022. Coupling Data- and Knowledge-Driven Methods for Landslide Susceptibility Mapping in Human-Modified Environments: A Case Study from Wanzhou County, Three Gorges Reservoir Area, China[J]. *Remote Sensing*, 14(3): 774.
- Zhu A X, Wang R X, Qiao J P, Qin C Z, Chen Y B, Liu J, Du F, Lin Y, Zhu T X. 2014. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic[J]. *Geomorphology*, 214: 128-138.