环境因子组合和负样本选取策略对花岗岩区崩岗易发性的影响

郭 飞[1,2]，蒋广辉[1,2]，黄晓虎[1,2※]，王秀娟[1,2]，夏 栋[3]，陈 洋[4]，李小伟[5]

（1. 三峡库区地质灾害教育部重点实验室，宜昌 443002；2. 三峡大学土木与建筑学院，宜昌 443002；3. 三峡大学水利与环境学院，宜昌 443002；4. 广东海洋大学电子与信息工程学院，湛江 524000；5. 中南冶金地质研究所 宜昌 443003）

摘 要：不同环境因子组合和负样本选取策略对崩岗易发性评价结果存在较多不确定性。为探究其对评价结果的影响，该研究以江西省兴国县花岗岩区为例，利用地理探测器探测 17 个环境因子的统计量 q 值，根据累计 q 值百分比大小依次选择 4、7、10 和 17 个环境因子进行组合；利用单随机欠采样、频率比法及改进频率比法等负样本选取策略构建与正样本等量的负样本数据集；采用随机森林模型进行易发性评价，并对评价结果进行对比分析。结果表明：1）3 种负样本选取策略下的模型精度随着因子数量的增加先下降再上升，考虑 4 个环境因子的模型 AUC（area under curve）值分别为 0.729、0.909 和 0.909，较最优环境因子组合仅相差 0.020～0.038，说明考虑主控环境因子，即可得到较为理想的精度；2）通过频率比法选取的负样本数据集更具合理性；3）研究区内高和极高易发区主要分布在兴国县西南部，而极低易发区主要分布在兴国县北部及东部，这与实际情况较吻合。该研究通过探究不同环境因子组合和负样本选取策略对崩岗易发性评价的影响，可为花岗岩区崩岗的防灾减灾提供科学依据。

关键词：易发性；随机森林；崩岗；地理探测器；环境因子组合；负样本选取策略

doi：10.11975/j.issn.1002-6819.202307270

中图分类号：P694 **文献标志码**：A **文章编号**：1002-6819(2024)-01-0199-10

郭飞，蒋广辉，黄晓虎，等. 环境因子组合和负样本选取策略对花岗岩区崩岗易发性的影响[J]. **农业工程学报**，2024，40(1)：199-208. doi: 10.11975/j.issn.1002-6819.202307270 http://www.tcsae.org

GUO Fei, JIANG Guanghui, HUANG Xiaohu, et al. Impact of environmental factor combinations and negative sample selection on Benggang susceptibility in granite areas[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(1): 199-208. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202307270 http://www.tcsae.org

0 引 言

崩岗侵蚀是中国南方花岗岩区最严重和最常见的土壤侵蚀类型之一[1-3]，具有爆发性强、发展速度快、侵蚀量大的特点，易诱发泥石流和山体滑坡等地质灾害，严重威胁地区国土、粮食、生态和公众安全[4]。开展崩岗空间预测对崩岗灾害预警和国土空间规划具有重要科学价值[5]。

崩岗空间预测即崩岗易发性评价，较多文献表明，环境因子评价体系构建和样本集选取策略是崩岗易发性的重要步骤，直接影响崩岗易发性建模结果的精度[6-8]。

目前，环境因子体系多数依靠经验确定，因子过少或过多都会影响最终结果的准确性与可信度[9]。然而，对于环境因子的选取还没有通用的指导原则。KAVZOGLU 等[10] 应用遗传算法在研究区域的 16 个可用因子中寻找最佳因子的组合，结果表明，选取 8 个环境因子，模型就可达到理想的预测精度。MA 等[11] 总结前人研究发现，滑坡易发性评价常使用具有代表性的 6、8、11 个环境因子，使用频率比和证据权重模型进行 3 种因子组合的易发性评价，以获得可靠的滑坡易发性图，预测精度随环境因子数量的增加而降低。JEBUR 等[12] 发现在原有数据集的基础上添加其他环境因子（如地质、土地利用类型等），对模型精度提升有限。PERRIRA 等[13] 为了确定滑坡易发性因子的最佳组合，使用 7 个环境因子进行所有可能的组合，发现使用坡度、反向湿度指数、土地利用类型构建的环境因子体系最优。上述学者基于不同方法对环境因子体系构建进行了探究性研究，绝大多数采用了线性、非线性选择策略或根据经验选择环境因子，尽管有助于避开“维数灾难”而增益模型中变量可解释性，但其忽视了地理要素空间交互影响。崩岗灾害发生是环境因子在特定地理空间上综合作用的结果，基于地理空间异质性角度探求易发性评价的环境因子组合是十分必要的。

另外，数据驱动模型的精度与正负样本的质量密切相关。传统崩岗易发性评价的样本数据集将研究区已发生的崩岗点作为正样本，在未发生崩岗区域采用单随机欠采样（single random undersampling，SRU）法，选取与正样本数量相同的非崩岗点作为负样本。但是 SRU 法可能会导致选择错误的负样本，这是因为对于一个没有

收稿日期：2023-07-28 修订日期：2024-01-08

基金项目：国家自然科学基金项目（42107489）；湖北省自然科学基金项目（2022CFB557）；湖北巴东地质灾害国家野外科学观测研究站开放基金项目（BNORSG202304）；三峡库区地质灾害教育部重点实验室开放基金项目（2022KDZ14）；土木工程防灾减灾湖北省引智创新示范基地项目（2021EJD026）

作者简介：郭飞，博士，硕士生导师，研究方向崩岗灾害风险评估。Email：ybbnui.2008@163.com

※通信作者：黄晓虎，博士，硕士生导师，研究方向地质灾害致灾机理。Email：88569096@qq.com

历史灾害事件的地区并不意味着该地区没有此类灾害[14]。合理选取负样本，建立高效评价模型对于提升易发性评价精度具有重要影响。采用频率比（frequency ratio，FR）法[15]进行易发性评价，在极低、低易发区中随机选取的负样本较 SRU 更为合理，可以提高模型的预测性能。郭衍昊等[16]以山区汶川地震诱发的滑坡为研究区，开展梯度提升决策树、随机森林和耦合模型的精度评价，结果表明频率比法选择滑坡负样本可以明显提高易发性精度。但是其存在需要对连续数据重分类模糊为离散数据的缺点，改进频率比[17]（automatic landslide susceptibility analysis，ALSA）法克服了传统通用方法中频率比值分布的不连续性，提高了各地质灾害影响因子敏感性的区分度，并减小了因子分级的主观性。

易发性评价模型主要包括知识驱动模型、数据驱动模型和机制驱动模型。其中定性评价模型主要依赖专家的主观经验，缺乏定量表达，并且各个专家的评价标准之间存在差异，最终导致无法对评价结果；机制驱动模型对区域物理力学参数要求较高，且这些参数在区域尺度上存在空间变异性并不易获得[18]。与之相比，数据驱动模型在理论基础和实际应用中都具有优势，机器学习是数据驱动模型中的一个重要分支，其在克服过拟合问题、模拟影响因素与敏感性之间的非线性关系以及自动生成最佳特征以实现高预测精度方面表现突出的优势[19]，而成为易发性评价模型的热点研究问题[20-22]。机器学习涵盖的算法有深度神经网络（deep neural networks，DNN）[23]、随机森林（random forest，RF）[24]、支持向量机[25]（support vector machines，SVM）等模型。陈飞等[23]提出信息量（information value，IV）与 DNN 结合的易发性评价模型，结果表明 IV-DNN 模型较信息量模型有更好的精度。牛瑞卿等[26]使用基于粗糙集理论（rough sets，RS）的 SVM 易发性评价，结果表明基于 RS-SVM 的滑坡易发性评价模型具有预测能力强、计算效率高等优点。但这些模型可解性差，泛化能力有限，对噪声数据敏感。RF 是 Bagging 抽样和决策树的集成模型，Bagging 抽样具有准确性高、抗噪声能力强等优势，决策树具有较强的解释性并且易于理解[27-29]，RF 模型集合两者的优点，有优秀的预测性能。因此，本文采用 RF 模型作为易发性评价模型。

综上所述，本文以江西省赣南兴国县花岗岩地区为研究区，根据地理探测器[30]（GeoDetector，GD）探测的 17 个环境因子的统计量 q 值大小进行环境因子组合；采用单随机欠采样、频率比法以及改进频率比法 3 种负样本选取策略，利用 RF 模型探究不同环境因子组合和负样本选取策略对该区域崩岗易发性评价的影响，以期为环境因子体系及样本数据集的构建提供参考，进而提高崩岗易发性评价结果精度 。

1 研究区域和数据

1.1 研究区域

兴国县位于江西省中南部（115°01′～115°51′E，26°03′～26°41′N）。地处南岭东西向复杂构造带东段北侧，地貌以低山、丘陵为主，地势由东北边缘逐渐向中南部倾向。属亚热带季风湿润气候，气候温和，雨量充沛，光照充足，位于赣江支流的平江上游，具有优越的水热条件。成土母岩以花岗岩和砂岩为主，土壤以红壤为主，含沙量大，易侵蚀。值得注意的是，花岗岩区崩岗侵蚀最为严重。土地总面积约 2 853 km^2，土壤与土地利用类型多样。据 2015 年江西省水土保持规划崩岗调查数据显示，兴国县现存崩岗点 2 933 个，面积 1 262.43 hm^2，分布密度为 0.91 个/km^2，其中约 80%（2 460 个）分布于花岗岩区，因此本文以花岗岩区为研究区（图 1）。

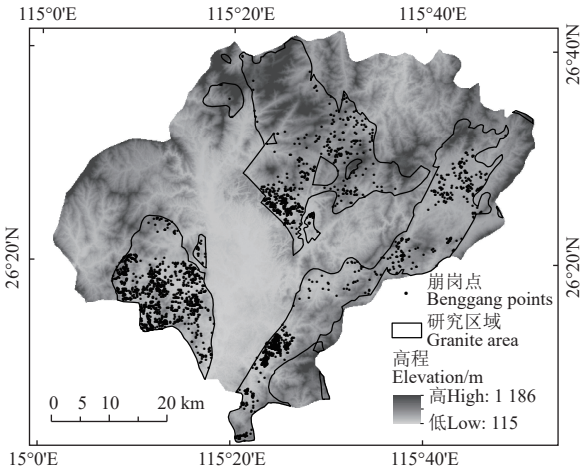


图 1 研究区位置及崩岗分布图

Fig.1 Study area location and Benggang distribution map

1.2 数 据

从《南方崩岗防治规划（2008—2020 年）》获取了兴国县 2 460 处历史崩岗数据，其属性包括崩岗坐标、面积、形态等。

关于崩岗侵蚀机理国内做了大量研究[31-34]，认为崩岗是一种在水力和重力相互作用下，山坡土体受破坏而产生崩塌与冲刷的侵蚀地貌，其发育条件包括：疏松深厚的基岩风化物；降雨径流和重力崩塌的复合作用；地表植被的破坏。据此，从地形地貌、地质条件、气象水文和植被覆盖共 4 个维度选取研究区共 17 个环境因子，数据信息见表 1，各环境因子见图 2。

地形地貌包括坡长因子、坡度、平面曲率、剖面曲率、坡向、地形湿度指数。其中坡长因子是指地表径流源点到坡度减小到可以辨识的沟道之间的水平距离，坡面侵蚀量随着坡长因子变化强弱波状起伏变化；坡度是坡地相对于水平地面的倾斜程度，坡度较大时，山体的稳定性较差，容易发生崩岗；平面曲率是指地面曲率在垂直方向的分量，影响流动的汇聚和分散，进而影响到侵蚀；剖面曲率是指地面曲率在水平方向的分量，影响流动的加速和减速，进而影响到侵蚀；坡向是指坡面法线在水平面上的投影的方向，通过影响土壤水分和日照时数从而对土壤侵蚀产生作用，是引起土壤侵蚀的重要因子；地形湿度指数是区域地形对径流流向和蓄积影响的物理指标，与土壤相对含水率呈线性关系，是影响土壤侵蚀的重要因素。

表 1　数据类型及来源

Table 1　Data types and sources

数据维度 Data dimension	数据名称 Data name	类型 Type	分辨率 Resolution/m	数据来源 Data source
地形地貌 Topography	坡长因子 Slope length factor(LS-factor) 坡度 Slope 平面曲率 Plane curvature (PlanC) 剖面曲率 Profile curvature (ProfC) 坡向 Aspect 地形湿度指数 Topographic wetness index (TWI)	栅格 (.tif)	12.5	ALOS-1 DEM/卫星 (Satellite) (https://search.asf.alaska.edu/#/)
地质条件 Geological conditions	土壤可蚀性 Soil erodibility (GZK) 黏土含量 Clay content (Clay) 砂含量 Sand content (Sand)	栅格 (.tif)	1 000	中国科学院资源环境科学与数据中心 (www.resdc.cn)
气象水文 Meteorology and hydrology	降雨侵蚀力 Average rainfall erosivity (GZR)			
植被覆盖 Vegetation coverage	标准化 VH 通道后向散射系数 Normalized backscatter coefficient of VH channel (VH) 标准化 VV 通道后向散射系数 Normalized backscatter coefficient of VV chanmel (VV)	栅格 (.tif)	10	Sentinel-1
	植被覆盖度 Vegetation coverage (FVCx) 叶面积指数 Leaf area index (LAI) 着色指数 Coloration index (CI) 亮度指数 Brightness index (BI) 修正土壤调整植被指数 Modified soil-adjusted vegetation index (MSAVI)	栅格 (.tif)	10	Sentinel-2 (https://scihub.copernicus.eu/dhus/#/home)

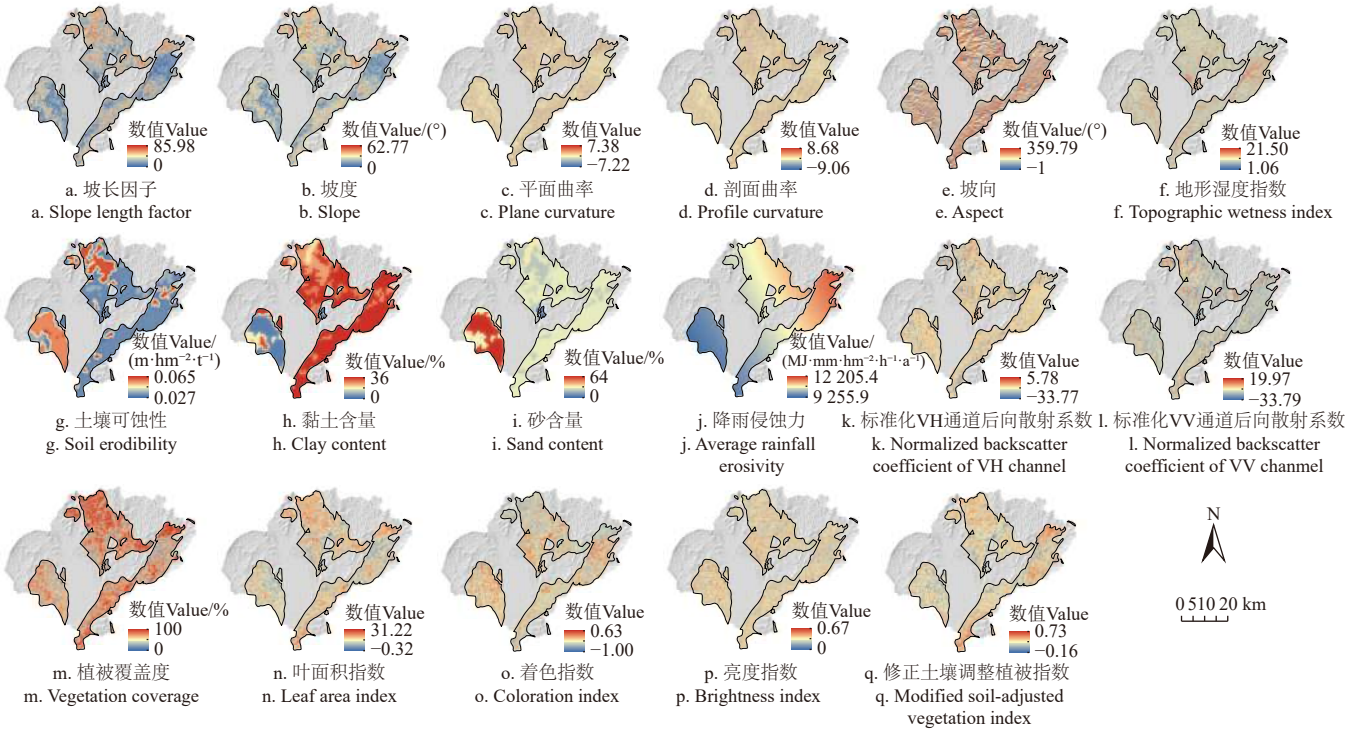


图 2　研究区各环境因子

Fig.2　Environmental factors in the study area

本文的研究区全为花岗岩区，故未考虑地层岩性对崩岗发育的影响，主要选择了土壤可蚀性、黏土含量、砂含量等环境因子。其中土壤可蚀性是指岩石或土壤在水或其他介质的作用下发生化学反应而逐渐溶解或侵蚀的性质，其导致岩石或土壤的物理性质发生一定改变，使岩石或土壤崩塌和侵蚀；黏土含量和砂含量与土壤的物理性质密切相关，而崩岗侵蚀与土壤的岩土力学性质有着密切联系。

气象水文主要选取降雨侵蚀力。降雨侵蚀力是指降雨对土壤表面的冲击力和侵蚀力，较大的降雨侵蚀力可能导致土壤侵蚀等自然灾害。

植被覆盖包括标准化 VH 通道后向散射系数、标准化 VV 通道后向散射系数、植被覆盖度、叶面积指数、着色指数、亮度指数、修正土壤调整植被等，这些环境因子直接或间接的反映植被覆盖情况。其中标准化 VH 通道后向散射系数、标准化 VV 通道后向散射系数以用于提取植被信息、识别植被类型、监测植被动态等；植被覆盖度是指植被（包括叶、茎、枝）在地面的垂直投影面积占统计区总面积的百分比；叶面积指数是指单位土地面积上的总植物叶面积；着色指数是指着色面积占

表面积的百分率，可以用于评估植物的健康状况和生长状况；亮度指数是指在一个平面上，每平方米接收到的光能量在植被覆盖方面，亮度指数，可用于提取植被信息、识别植被类型、监测植被动态等；修正土壤调整植被指数是一种通过减轻土壤对作物监测结果的影响，旨在更准确地监测植被状态的植被指数。

2 研究方法

本研究所用的方法流程主要有 4 个步骤：1）收集历史崩岗信息，根据崩岗发育的影响因素选择相应的环境因子；2）利用地理探测器分析各环境因子统计量 q 值，然后将环境因子 q 值从大到小依次叠加，根据累计 q 值百分比确定环境因子组合；3）采用 SRU、FR 法以及 ALSA 法 3 种负样本选取策略，构建与正样本等量的负样本数据集；4）将样本数据集以 7∶3 的比例分为训练集和测试集，采用训练集数据训练 RF 模型，然后用训练好的 RF 模型对测试集进行计算，通过 ROC 曲线评估模型的预测精度，并计算崩岗易发性结果。

2.1 环境因子相对重要性评估

地理探测器 GD 是无线性假设的，能够探测崩岗的空间分异性，分析不同分层内影响因子对崩岗发生的解释力度。作为空间数据探索和分析十分可靠的工具之一，主要被用来分析各种现象的驱动力和影响因素，可以去除因子之间的相互影响，以定量化的结果筛选出对现象具有重要影响的因子，并提高分析精度。用于崩岗研究的地理探测器的一般假设可以表示为：如果环境因子控制或促成崩岗，则崩岗的空间应与环境因子的空间分布相似。地理探测器由因子探测、交互探测、风险探测、生态探测构成。在探测某因子对属性空间分异性解释程度时，属因子探测，用 q 值[35] 来度量，其表达式为

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} \tag{1}$$

式中 N_h 和 N 分别为分层 h 和全区的样本单元数；L 为崩岗（Y）或影响因子 (X) 的分类或分区；$\sigma^2{}_h$ 和 σ^2 分别是分层 h 和全区的 Y 值的方差。q 表示因子对崩岗发育的解释力，值域为 [0,1]，值越大说明崩岗的空间分异性越明显；如果分层是由自变量 X 生成的，则 q 值越大表示自变量 X 对属性 Y 的解释力越强，反之则越弱。

2.2 负样本选取策略

在建模前，要对研究区正负样本进行选取。负样本是指在研究区域中随机选取的非崩岗点，在进行选取时为保证负样本的合理性，需要在崩岗不易发生的区域进行选取非崩岗点，力求负样本不位于潜在的崩岗区域。因此本研究除了在研究区随机选取负样本外，根据历史崩岗点和环境因子数据之间的空间分布关系，利用 FR 模型和 ALSA 模型计算各崩岗环境因子的频率比值，以环境因子总频率比值为基础，预测并绘制初步的崩岗易发性图，从极低和低易发区中随机选取和崩岗点数量相同的“非崩岗”负样，力求选取的负样本更加合理[36-37]。

2.2.1 频率比法

根据自然历史分析法可知，与过去发生过崩岗相似的地理环境更容易形成新的崩岗。频率比法是地质灾害易发性制图中使用做多的二元统计方法，可用于揭示现有崩岗侵蚀发育特征与每个特定环境因子间的非线性关系。其表达式为

$$F_{ij} = \frac{N_{ij}/N_r}{A_{ij}/A_r} \tag{2}$$

式中 F_{ij} 为为第 i 个因子第 j 个分类的频率比值；N_{ij} 为第 i 个影响因子中第 j 类发生崩岗的面积，hm^2；N_r 为研究区崩岗的总面积，hm^2；A_{ij} 表示第 i 个影响因子中第 j 类所占的面积，hm^2；A_r 表示研究区总面积，hm^2。较大的比值表明分配因子类对崩岗的侵蚀条件贡献更大[38]。其值大于 1 时，表明崩岗与所考虑的环境因子的类别之间有较强的相关性；比值小于 1 时，崩岗与其因子类别之间的关系较小；比值等于 1 时，表明难以判断。

根据频率比计算公式计算各因子各个分类的频率比值，然后利用栅格计算器加权叠加图层得到易发性分区图，研究区崩岗易发性指数 Q 计算如下：

$$Q = \sum q_i \cdot F_{ij} \tag{3}$$

式中 q_i 为第 i 个因子的解释力值。

2.2.2 改进频率比法

ALSA 法是一种改进的频率比法，ALSA 法的优点是无需对连续型数据进行重分类模糊为离散数据。其核心思想是先将连续性数据进行归一化，再以归一化后的单因子值为中心，统计其邻域范围内的崩岗数量和该邻域的面积，通过上述 FR 模型公式计算不同单因子值对于崩岗发生的相对影响程度。

根据改进频率比法计算出各因子频率比值，然后利用栅格计算器加权叠加图层得到易发性分区图，算式如下：

$$Q = \sum q_i \cdot F_i \tag{4}$$

式中 F_i 为第 i 个因子的频率比值。

2.3 随机森林模型

RF 模型[39] 是由多个决策树组成的集成分类模型。在给定自变量 X 下，每个决策树模型都通过投票来选择最优的分类结果。其算法原理是首先利用 bootstrap 抽样从原始训练集中有放回地抽取 K 个样本，且各样本的特征数都与原始训练集相同；再分别对 K 个样本建立决策树模型，得到 K 种分类结果；针对各样本，从总特征 m 中随机选取 n（$n \leqslant m$）个特征作为分裂特征集，从中选择最优特征对节点进行生长，当 $n < m$ 时，每一棵决策树之间又存在差异性。最后，形成随机森林，根据 K 种分类结果的众数决定其最终分类。

2.4 基于 ROC 曲线的精度分析

接受者操作特性（receiver operating characteristic，ROC）曲线广泛应用于模型精度的验证，利用 ROC 曲线能够检验模型对正负类样本的预测能力。通过计算不同阈值（易发性指数）情况下的真阳性率（true positive rate，TPR）和假阳性率（false positive rate，FPR）。

以 FPR 为 x 轴，TPR 为 y 轴，作 ROC 曲线。AUC（area under curve）是 ROC 曲线与 x 轴间围成的面积值，该值多在 0.5～1 之间，越接近 1 表明模型预测性能越好[40]。

3　结果与分析

3.1　环境因子组合

本文在进行环境因子重要性分析时，将未发现崩岗的地方作为负样本，已发现的崩岗点作为正样本，以此计算各环境因子的 q 值。首先将所有因子的 q 值相加，然后按照 q 值从大到小依次叠加，计算累计 q 值百分比，最后为探究最佳环境因子组合，根据累计 q 值百分比选择 56.89%、78.55%、92.88% 以及 100.00% 进行环境因子组合，分别为 4、7、10 和 17 个因子（表 2）。

表 2　环境因子 q 值及环境因子组合累计 q 值百分比

Table 2　Environmental factor q value and environmental factor combination q value cumulative percentage (P_q)

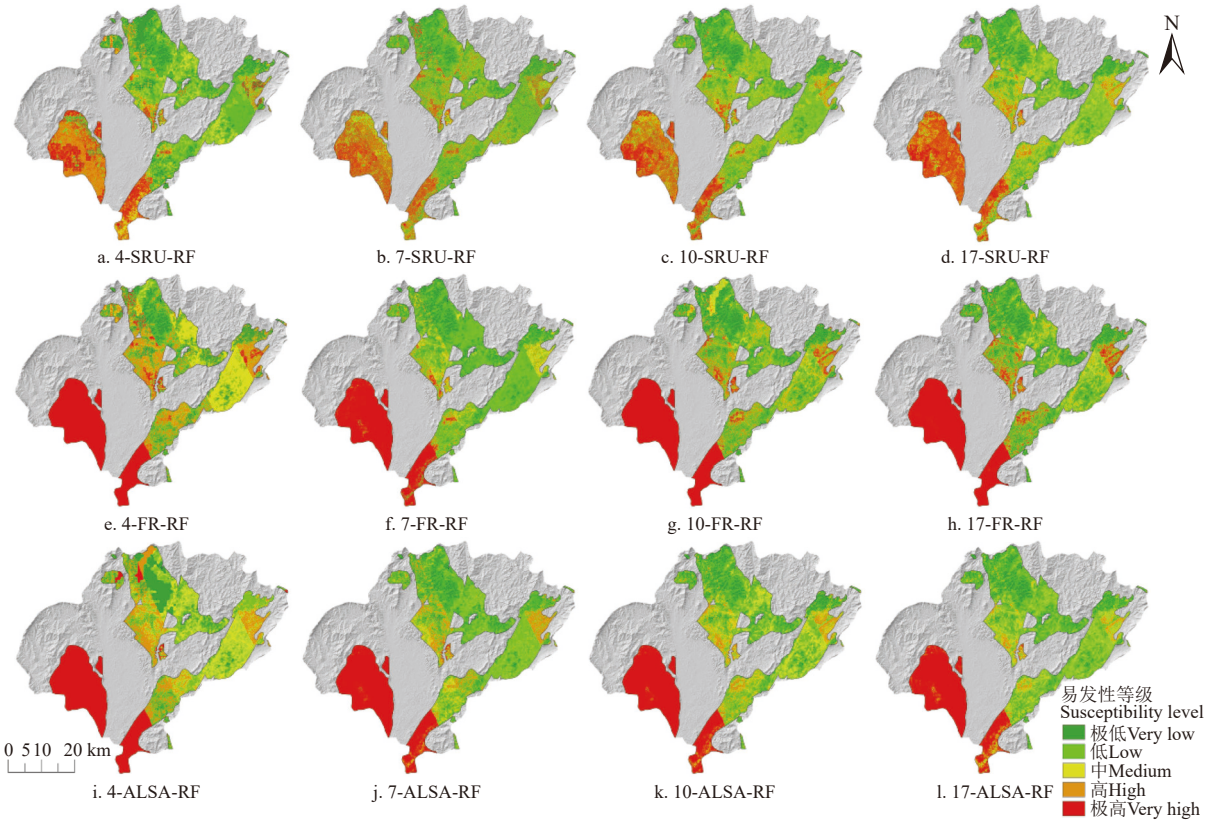
因子 Factor	q 值 q value	P_q/%	因子 Factor	q 值 q value	P_q/%
BI	6.40×10^{-6}	27.30	GZK	2.38×10^{-4}	92.88
Aspect	1.32×10^{-5}	37.32	LS-factor	2.60×10^{-4}	95.01
TWI	2.03×10^{-5}	47.25	LAI	2.94×10^{-4}	97.01
ProfC	2.95×10^{-5}	56.89	Slope	3.44×10^{-4}	98.33
PlanC	5.34×10^{-5}	65.19	Clay	3.99×10^{-4}	99.04
VV	8.39×10^{-5}	72.28	Sand	4.12×10^{-4}	99.53
VH	8.82×10^{-5}	78.55	FVCx	4.15×10^{-4}	99.53
MSAVI	1.25×10^{-4}	84.29	GZR	1.13×10^{-3}	100.00
CI	2.31×10^{-4}	89.87			

3.2　各环境因子组合多重共线性分析

由于环境因子的多样性和复杂性，所选取环境因子间可能存在相关性，各因子之间过高的相关性会降低崩岗易发性评价模型的精度，并增加模型复杂度。因此，本文使用多重共线性理论中的方差膨胀因子（variance inflation factor，VIF）对环境因子进行相关性分析，其中方差膨胀因子 VIF 值大于 5，考虑因子之间的多重共线性问题，反之则不考虑。对各环境因子组合进行多重共线性分析，结果表明，各环境因子组合中环境因子的 VIF 值最大为 4.603，均小于 5，故不考虑环境因子之间的共线性问题，不剔除环境因子。

3.3　崩岗易发性评价

分别以 4、7、10、17 个环境因子构建环境因子体系，然后采用 SRU、FR 法以及 ALSA 法三种负样本选取策略构建原始数据集，将原始数据集划分为两类：从包含 2 460 个崩岗点数据和 2 460 个负样本点数据中随机选取 70% 的崩岗点数据和负样本点数据构建训练集，包含有 3 444 条数据；剩余 30% 的数据作为测试数据集，包含有 1 476 条数据。RF 模型通过训练集训练后，将原始数据集导入到模型中，计算各栅格单元的崩岗易发性指数。为了突出易发区之间的差异性，将易发性指数按自然间断法进行重分类为 5 类，分别对应极低易发区、低易发区、中易发区、高易发区、极高易发区，得到该区域的崩岗易发性评价图（图 3）。



注：小题名中，数字为环境因子数；SRU、FR、ALSA 分别为单随机欠采样法、频率比法、改进频率比法；RF 为随机森林模型。

Note: In the small title, the number represents the number of environmental factors; SRU, FR, ALSA represent the single random undersampling, frequency ratio, automatic landslide susceptibility analysis method, respectively; RF represents the random forest model.

图 3　基于不同环境因子组合和负样本选取策略的崩岗易发性评价

Fig.3　Evaluation of Benggang susceptibility based on different environmental factor combinations and negative samples selection strategies

对比不同条件下崩岗易发性评价结果，分区大致相似，其中环境因子组合为 17 个因子且负样本选取策略为频率比法的结果精度最高，结果显示：兴国县花岗岩区中极高易发区绝大部分分布在兴国县的西南部，部分极高易发区分布在兴国县的中部以及东部；极低易发区绝大部分分布在兴国县的北部以及东部；在兴国县的西南部，因为从极低易发区到极高易发区的变化，在易发区分区图中呈现出一条明显的分界线，这条分界线与 9 981.49 MJ·mm/(hm²·h·a) 的降雨侵蚀力等值线基本吻合（图 4），降雨侵蚀力大小对崩岗的发育有重要影响。

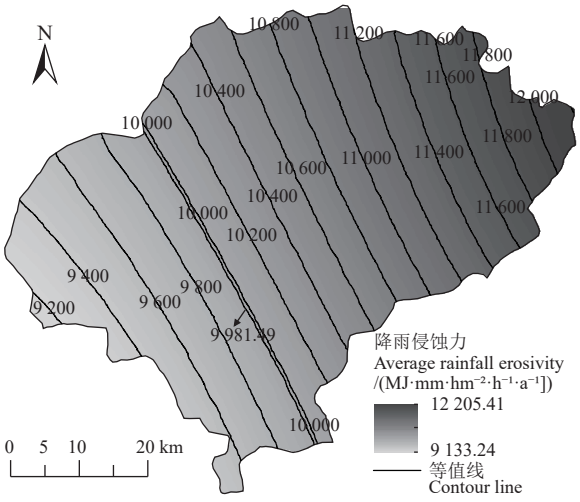


图 4 兴国县降雨侵蚀力等值线

Fig.4 Rainfall erosivity contour line in Xingguo County

3.4 模型精度评价

用训练好的 RF 对测试数据集进行计算，采用 ROC 曲线进行模型精度验证。ROC 曲线下面积（AUC 值）是用于评估模型性能的重要指标，AUC 值越接近 1，表示在模型的可预测性方面性能越好。不用环境因子组合和负样本选取策略下的随机森林模型的 ROC 曲线如图 5。结果显示，负样本选取策略为 SRU 时，AUC 值分别为 0.729、0.711、0.745 和 0.755（环境因子组合分别为 4、7、10 和 17 个因子）；负样本选取策略采用 FR 法选取时，AUC 值分别为 0.909、0.869、0.942 和 0.947（环境因子组合分别为 4、7、10 和 17 个因子）；负样本选取策略采用 ALSA 法选取时，AUC 值分别为 0.909、0.893、0.919 和 0.929（环境因子组合分别为 4、7、10 和 17 个因子）。模型的精度随着因子数量的增大先下降再上升，模型的精度在因子数量为 7 个的时候最小，在因子数量为 17 个的时候最大。

在进行环境因子体系构建时，除结果精度外，还应从计算成本方面综合考虑环境因子的选取。在本研究中，不同的负样本选取策略（SRU、FR 法和 ALSA 法）下，因子数量为 4 个时的模型就达到较高的精度，与因子数量为 17 时的模型精度仅相差 0.026、0.038 和 0.020。说明考虑主控环境因子，即可得到较为理想的精度。在实际工程中，若研究区数据图层获取不易且时间紧迫、任务繁重，可考虑找寻崩岗发育的主控因子开展易发性评价，以节约时间成本；若研究区数据图层多且对模型精度要求较高时，可全面考虑崩岗发育的影响因素构建指标体系，开展评价。

负样本的合理性直接决定了模型的训练集和测试集是否具有合理性，对结果的精度会产生较大影响。对比发现：当负样本采用 FR 法和 ALSA 法选取时，模型精度均比负样本随机选取时高，整体得到了较大提升，AUC 值均高于 0.85，选取的样本更具有合理性，能够有效提高模型的精度；因子数量为 4 个时，负样本采用 FR 法和 ALSA 法选取的模型精度一致；因子数量为 7 个时，负样本采用 FR 法比 ALSA 法选取的模型精度低，ALSA 法选取的负样本更具有合理性，能够有效提高模型的精度；因子数量为 10 个和 17 个，负样本采用 FR 法比 ALSA 法选取的模型精度高，FR 法选取的负样本更具有合理性，能够有效提高模型的精度。

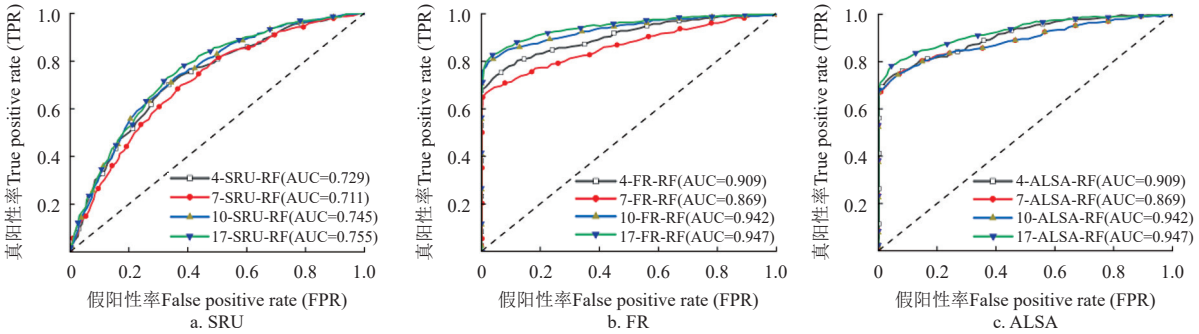


图 5 不同条件下模型的受试者工作特征（ROC）曲线和曲线下面积（AUC）

Fig.5 Receiver operating characteristic (ROC) curve and area under curve (AUC) of the model under different conditions

4 讨 论

本研究通过地理探测器探测崩岗的空间分异性，分析不同环境因子对崩岗发生的解释力，解释力越大，对崩岗发育促进作用越大，在进行环境因子组合时应优先考虑。在分析各环境因子组合中的共线性后构建评价指标体系，并开展了环境因子数量为 4，7，10，17 下的易发性评价，发现随着环境因子数量的增加模型精度呈现先下降再上升的趋势，这与文献 [11, 41-42] 得到的结果较类似。出现这种趋势的原因可能为：1）目前开展崩岗易发性评价所采用的崩岗点多为点数据（即坐标点），

而无反映崩岗范围大小的面数据，造成部分信息统计不全而导致的。2）用于表征环境因子共线性的 VIF 值，一般 VIF 值在 1～5 时，各环境因子间的多重共线性可不考虑，但特殊情况下，自变量间也有可能存在多重共线性[43]。4 个环境因子时，环境因子的 VIF 值都较小；环境因子数量增加到 7 个时，植被覆盖度的 VIF 值出现增大且模型精度下降情况，说明植被覆盖度、坡长因子、叶面积指数及坡度之间可能存在不可忽略的相关性，从而降低了模型精度；环境因子为 10 和 17 时，植被覆盖度、坡长因子、叶面积指数及坡度的 VIF 值仍较大（表 3），但其精度反而提升了，可能是因为新增的环境因子对其精度的提升有促进作用。对于上述 4 个环境因子的相关性，以植被覆盖度和叶面积指数为例，植被覆盖度是指植被（包括叶、茎、枝）在地面的垂直投影面积占统计区总面积的百分比，叶面积指数是指单位土地面积上的总植物叶面积，均反映了植被的覆盖情况，当叶面积增加时，植被覆盖的范围也会相应的有所增加，二者存在正相关关系，这进一步说明植被覆盖度和叶面积指数可能存在着共线性问题。

表 3　不同环境因子数量下因子的 VIF 值

Table 3　Variance inflation factor (VIF) value under different factor numbers

因子 Factors	因子数量 Number of factors			
	17	10	7	4
LS-factor	3.963	3.512	3.463	-
LAI	4.603	4.495	3.443	-
Slope	4.272	3.734	3.687	-
FVCx	4.395	4.189	3.523	1.024

5　结　论

崩岗地质灾害的的空间预测是一个复杂的非线性过程，提高模型的精度对于崩岗预测任务具有重要意义。本文通过探索环境因子的组合和负样本的选取策略，为研究区域提供适当的因子组合和合理的负样本，提高了研究区域的模型精度，为研究区域提供更合理可靠的崩岗预测模型，同时也为其他崩岗频发区域优化崩岗易发性模型提供了思路。本研究结论如下：

1）本文用地理探测器对各环境因子进行重要性分析，并根据重要性大小进行因子组合。精度评价发现，模型的 AUC（area under curve）值随着因子数量的增多均呈现出先减小（7 个因子的时候最小）再增加（17 个因子的时候最大）的规律。但 3 种负样本选取策略下，17 个环境因子较 4 个环境因子模型精度提升范围有限，AUC 值仅增加 0.020～0.038。说明考虑主控环境因子，即可得到较为理想的精度，节约计算成本。

2）采用 3 种负样本选取策略选取的负样本与解译的崩岗点构建样本集。对比 3 种负样本选取策略对随机森林（random forest，RF）模型精度的影响，判断负样本选取策略的合理性。结果显示环境因子为 4 个时，采用频率比法（frequency ratio，FR）和改进频率比法（automatic landslide susceptibility analysis，ALSA）选取的负样本最为合理，AUC 值均为 0.909；环境因子为 7 个时，ALSA 法选取的负样本最为合理，AUC 值为 0.893；环境因子为 10 个和 17 个时，FR 法选取的负样本最为合理，AUC 值分别为 0.942 和 0.947。综上，在本研究区中 FR 法能显著提高模型的预测精度。

3）降雨侵蚀力对崩岗的发育密切相关，崩岗主要发育于降雨侵蚀力为 9 133.24～9 981.49 MJ·mm/(hm²·h·a) 的范围内，高和极高易发区绝大部分分布于研究区的西南部，少部分极高易发区分布于兴国县的中部及东部，极低易发区大部分分布在兴国县北部和东部。

[参　考　文　献]

[1] DUAN X Q, DENG Y S, TAN Y, et al. The soil configuration on granite residuals affects Benggang erosion by altering the soil water regime on the slope[J]. International Soil and Water Conservation Research, 2021, 9(3): 419-432.

[2] DENG Y S, DONG X, CAI C F, et al. Simulation of water characteristic curve in the soil profile of the collapsing gully on granite area of South China based on the fractal theory[J]. Science of Soil and Water Conservation, 2016, 14(2): 1-8.

[3] 李治郡，钟琳婷，黄炎和，等. 基于贴近摄影测量的崩岗侵蚀监测技术[J]. 农业工程学报，2021，37(8)：151-159.
LI Zhijun, ZHONG Linting, HUANG Yanhe, et al. Monitoring technology for collapse erosion based on the nap of the object photograph of UAV[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(8): 151-159. (in Chinese with English abstract)

[4] 文慧，倪世民，冯舒悦，等. 赣南崩岗的发育阶段及部位对土壤水力性质的影响[J]. 农业工程学报，2019，35(24)：136-143.
WEN Hui, NI Shimin, FENG Shuyue, et al. Effects of developmental stages and parts of collapsing gully on soil hydraulic properties in southern Jiangxi[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(24): 136-143. (in Chinese with English abstract)

[5] 林小慧，黄炎和，林金石，等. 基于 DPSIR 模型的崩岗侵蚀风险评价及时空特征[J]. 农业工程学报，2023，39(18)：123-131.
LIN Xiaohui, HUANG Yanhe, LIN Jinshi, et al. Risk assessment and spatial-temporal characteristics of Benggang erosion based on DPSIR model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(18): 123-131. (in Chinese with English abstract)

[6] 张书豪，吴光. 随机森林与 GIS 的泥石流易发性及可靠性[J]. 地球科学，2019，44(9)：3115-3134.
ZHANG Shuhao, WU Guang. Debris flow susceptibility and Its reliability based on random forest and GIS[J]. Earth Science, 2019, 44(9): 3115-3134. (in Chinese with English abstract)

[7] TANG X, HONG H, SHU Y, et al. Urban waterlogging susceptibility assessment based on a PSO-SVM method using a novel repeatedly random sampling idea to select negative

samples[J]. Journal of Hydrology, 2019, 576: 583-595.

[8] HU Q, ZHOU Y, WANG S, et al. Machine learning and fractal theory models for landslide susceptibility mapping: Case study from the Jinsha River Basin[J]. Geomorphology, 2020, 351: 106975.

[9] 罗路广，裴向军，崔圣华，等. 九寨沟地震滑坡易发性评价因子组合选取研究[J]. 岩石力学与工程学报，2021，40(11)：2306-2319.
LUO Luguang, PEI Xiangjun, CUI Shenghua, et al. Combined selection of susceptibility assessment factors for Jiuzhaigou earthquake-induced landslides[J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(11): 2306-2319. (in Chinese with English abstract)

[10] KAVZOGLU T, SAHIN E K, COLKESEN I. Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm[J]. Engineering Geology, 2015, 192: 101-112.

[11] MA S, QIU H, HU S, et al. Quantitative assessment of landslide susceptibility on the Loess Plateau in China[J]. Physical Geography, 2020, 41(6): 489-516.

[12] JEBUR M N, PRADHAN B, TEHRANY M S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale[J]. Remote Sensing of Environment, 2014, 152: 150-165.

[13] PEREIRA S, ZÊZERE J L, BATEIRA C. Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models[J]. Natural Hazards and Earth System Sciences, 2012, 12(4): 979-988.

[14] TANG X, MACHIMURA T, LI J, et al. A novel optimized repeatedly random undersampling for selecting negative samples: A case study in an SVM-based forest fire susceptibility assessment[J]. Journal of Environmental Management, 2020, 271: 111014.

[15] 马啸，王念秦，李晓抗，等. 基于 RF-FR 模型的滑坡易发性评价—以略阳县为例[J]. 西北地质，2022，55(3)：335-344.
MA Xiao, WANG Nianqin, LI Xiaokang, et al. Assessment of landslide susceptibility based on RF-FR model: taking Lueyang county as an example[J]. Notthwestern Geology, 2022, 55(3): 335-344. (in Chinese with English abstract)

[16] 郭衍昊，窦杰，向子林，等. 基于优化负样本采样策略的梯度提升决策树与随机森林的汶川同震滑坡易发性评价[J/OL]. 地质科技通报，1-20
4-01-04]. https: //doi. org/10.19509/ j. cnki. dzkq. tb20230037.
GUO Yanhao, DOU Jie, XIANG Zilin, et al. Optimized negative sampling strategies of gradient boosting decision tree and random forest for evaluating Wenchuan coseismic landslides susceptibility mapping[J/OL]. Bulletin of Geological Science and Technology, 1-20[2024-01-04]. https://doi.org/10.19509/j.cnki.dzkq.tb20230037. (in Chinese with English abstract)

[17] 李郎平，兰恒星，郭长宝，等. 基于改进频率比法的川藏铁路沿线及邻区地质灾害易发性分区评价[J]. 现代地质，2017，31(5)：911-929.
LI Langping, LAN Hengxing, GUO Changbao, et al. Geohazard susceptibility assessment along the Sichuan-Tibet railway and its adjacent area using an improved frequency ratio method[J]. Geoscience, 2017, 31(5): 911-929. (in Chinese with English abstract)

[18] 李坤，赵俊三，林伊琳，等. 基于 SMOTE 和多粒度级联森林的泥石流易发性评价[J]. 农业工程学报，2022，38(6)：113-121.
LI Kun, ZHAO Junsan, LIN Yilin, et al. Assessment of debris flow susceptibility based on SMOTE and multi-Grained Cascade Forest[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(6): 113-121. (in Chinese with English abstract)

[19] 郭飞，赖鹏，黄发明，等. 基于知识图谱的滑坡易发性评价文献综述及研究进展[J/OL]. 地球科学，1-33[2024-01-04]. http://kns.cnki.net/kcms/detail/42.1874.P.20230713.1234.002.html.
GUO Fei, LAI Peng, HUANG Faming, et al. Literature review and research progress of landslide susceptibility mapping based on knowledge graph[J/OL]. Journal of Earth Science, 1-33[2024-01-04]. http://kns.cnki.net/kcms/detail/42.1874.P.20230713.1234.002.html. (in Chinese with English abstract)

[20] ARABAMERI A, CHEN W, LOCHE M, et al. Comparison of machine learning models for gully erosion susceptibility mapping[J]. Geoscience Frontiers, 2020, 11(5): 1609-1620. .

[21] Liu Y, Xu P, Cao C, et al. A comparative evaluation of machine learning algorithms and an improved optimal model for landslide susceptibility: A case study[J]. Geomatics, Natural Hazards and Risk, 2021, 12(1): 1973-2001.

[22] ARABAMERI A, PRADHAN B, LOMBARDO L. Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling[J]. Catena, 2019, 183: 104223.

[23] 陈飞，蔡超，李小双，等. 基于信息量与神经网络模型的滑坡易发性评价[J]. 岩石力学与工程学报，2020，39(S1)：2859-2870.
CHEN Fei, CAI Chao, LI Xiaoshuang, et al. Evaluation of landslide susceptibility based on information volume and neural network model[J]. Chinese Journal of Rock Mechanics and Engineering, 2020, 39(S1): 2859-2870. (in Chinese with English abstract)

[24] GAYEN A, POURGHASEMI H R, SAHA S, et al. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms[J]. Science of the Total Environment, 2019, 668: 124-138.

[25] WEI Y, WU X, WANG J, et al. Identification of geo-environmental factors on Benggang susceptibility and its spatial modelling using comparative data-driven methods[J].

Soil and Tillage Research, 2021, 208: 104857.

[26] 牛瑞卿，彭令，叶润青，等. 基于粗糙集的支持向量机滑坡易发性评价[J]. 吉林大学学报（地球科学版），2012，42(2)：430-439.
NIU Ruiqing, PENG Ling, YE Runqing, et al. Landslide susceptibility assessment based on rough sets and support vector machine[J]. Journal of Jilin University (Earth Science Edition), 2012, 42(2): 430-439. (in Chinese with English abstract)

[27] DOU J, YUNUS A P, BUI D T, et al. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan[J]. Landslides., 2020, 17(3): 641-658.

[28] 李文彬，范宣梅，黄发明，等. 不同环境因子联接和预测模型的滑坡易发性建模不确定性[J]. 地球科学，2021，46(10)：3777-3795.
LI Wenbin, FAN Xuanmei, HUANG Faming, et al. Uncertainties of landslide susceptibility modeling under different environmental factor connections and prediction models[J]. Earth Science, 2021, 46(10): 3777-3795. (in Chinese with English abstract)

[29] 吴润泽，胡旭东，梅红波，等. 基于随机森林的滑坡空间易发性评价：以三峡库区湖北段为例[J]. 地球科学，2021，46(1)：321-330.
WU Runze, HU Xudong, MEI Hongbo, et al. Spatial susceptibility assessment of landslides based on random forest: A case study from Hubei section in the Three Gorges Reservoir Area[J]. Earth Science, 2021, 46(1): 321-330. (in Chinese with English abstract)

[30] 王劲峰，徐成东. 地理探测器：原理与展望. 地理学报，2017，72(1)：116-134.
WANG Jinfeng, XU Chengdong. Geodetector: Principle and prospective[J]. Acta Geographica Sinica, 2017, 72(1): 116-134. (in Chinese with English abstract)

[31] 廖义善，唐常源，袁再健，等. 南方红壤区崩岗侵蚀及其防治研究进展[J]. 土壤学报，2018，55(6)：1297-1312.
LIAO Yishan, TANG Changyuan, YUAN Zaijian, et al. Research progress on Benggang erosion and its prevention measure in red soil region of southern China[J]. Acta Pedologica Sinica, 2018, 55(6): 1297-1312. (in Chinese with English abstract)

[32] 陈晓安，杨洁，熊永，等. 红壤区崩岗侵蚀的土壤特性与影响因素研究[J]. 水利学报，2013，44(10)：1175-1181.
CHEN Xiaoan, YANG Jie, XIONG Yong, et al. Research on the soil characteristics and factors of collapsing erosion in the red soil zone[J]. Journal of Hydraulic Engineering, 2013, 44(10): 1175-1181. (in Chinese with English abstract)

[33] 廖凯涛，刘艳，刘荃，等. 赣州市崩岗侵蚀分布特征与影响因子分析[J]. 水土保持研究，2021，28(6)：126-130.
LIAO Kaitao, LIU Yan, LIU Quan, et al. Distribution characteristics and driving factors of Benggang erosion in Ganzhou city[J]. Research of Soil and Water Conservation, 2021, 28(6): 126-130. (in Chinese with English abstract)

[34] 万赐航，周慧平，王强，等. 植被覆盖度和降雨侵蚀力变化对小流域泥沙连通性的影响[J]. 农业工程学报，2022，38(12)：127-134.
WAN Cihang, ZHOU Huiping, WANG Qiang, et al. Effects of vegetation coverage and rainfall erosivity changes on sediment connectivity in small watersheds[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(12): 127-134. (in Chinese with English abstract)

[35] WANG J F, LI X H, CHRISTAKOS G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China[J]. International Journal of Geographical Information Science, 2010, 24(1): 107-127.

[36] ARABAMERI A, PRADHAN B, REZAEI K, et al. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function–logistic regression algorithm[J]. Land Degradation & Development, 2018, 29(11): 4035-4049.

[37] ARABAMERI A, PRADHAN B, REZAEI K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS[J]. Journal of Environmental Management, 2019, 232(15): 928-942.

[38] RAHMATI O, HAGHIZADEH A, POURGHASEMI H R, et al. Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison[J]. Natural Hazards, 2016, 82(2): 1231-1258.

[39] ANTONIADIS A, LAMBERT-LACROIX S, POGGI J M. Random forests for global sensitivity analysis: A selective review[J]. Reliability Engineering & System Safety, 2021, 206: 107312.

[40] 李建军，陈玉兰，焦菊英，等. 基于多元非线性空间建模的拉萨河流域沟蚀发生风险探测[J]. 农业工程学报，2022，38(17)：73-82.
LI Jianjun, CHEN Yulan, JIAO Juying, et al. Detecting gully occurrence risks using multivariate nonlinear spatial modeling in the Lhasa River Basin of China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 73-82. (in Chinese with English abstract).

[41] 黄发明，刘科技，曾子强，等. 环境因子筛选及其组合方法对滑坡易发性预测的影响规律[J/OL]. 应用基础与工程科学学报，1-20[2024-01-10]. http://kns.cnki.net/kcms/detail/11.3242.TB.20231110.0901.002.html.
HUANG Faming, LIU Keji, ZENG Ziqiang, et al. Influence of environmental factor selection and combination on landslide susceptibility prediction modeling[J/OL]. Journal of Basic Science and Engineering, 1-20[2024-01-10]. http://kns.cnki.net/kcms/detail/11.3242.TB.20231110.0901.002.html. (in

Chinese with English abstract)

[42] GAIDZIK K, RAMÍREZ -HERRERA M T. The importance of input data on landslide susceptibility mapping[J]. Scientific Reports, 2021, 11(1): 19334.

[43] SALMERON R, GARCIA C B, GARCIA J. Variance inflation factor and condition number in multiple linear regression[J]. Journal of Statistical Computation and Simulation, 2018, 88(12): 2365-2384.

Impact of environmental factor combinations and negative sample selection on Benggang susceptibility in granite areas

GUO Fei[1,2], JIANG Guanghui[1,2], HUANG Xiaohu[1,2※], WANG Xiujuan[1,2], XIA Dong[3], CHEN Yang[4], LI Xiaowei[5]

(1. *Key Laboratory of Geological Hazards on Three Gorges Reservoir Area*, *Ministry of Education*, *Yichang* 443002, *China*; 2. *College of Civil Engineering & Architecture*, *China Three Gorges University*, *Yichang* 443002, *China*; 3. *College of Hydraulic & Environmental Engineering*, *China Three Gorges University*, *Yichang* 443002, *China*; 4. *School of Electronic and Information Engineering*, *Guangdong Ocean University*, *Zhanjiang* 524000, *China*; 5. *Central-South Institute of Metallurgical Geology*, *Yichang* 443003, *China*)

Abstract: Benggang is one of the most severe types of soil erosion in the granite areas of southern China, due to the large erosion, strong explosiveness, and fast development speed. Accurate assessment of susceptibility is of great significance for the prevention and control of Benggang damages. In this study, different combinations of environmental factors and negative sample selection strategies were explored the impact on the assessment of Benggang susceptibility. A case study was taken from the granite area of Xingguo County, Ganzhou City, Jiangxi Province, China. A systematic detection was implemented to determine the explanatory power of 17 environmental factors on the development of Benggang using a GeoDetector (GD). According to the cumulative explanatory power percentage, 56.89%, 78.55%, 92.88%, and 100.00% were selected as the environmental factor combinations, corresponding to 4, 7, 10, and 17 environmental factors, respectively. Single random undersampling (SRU) was used to construct a negative sample dataset equal to positive samples using frequency ratio (FR). The susceptibility was calculated in the study area using automatic landslide susceptibility analysis (ALSA). Negative sample data was selected equal to positive samples in the low and extremely low susceptibility areas. The sample dataset was divided into the training and testing datasets in a 7:3 ratio. The training dataset was used to train the random forest (RF) model, and then the trained RF model was to calculate the testing dataset. The prediction accuracy of the model was evaluated to calculate the Benggang susceptibility using the receiver operating characteristic (ROC). The results show that: 1) The model accuracy under the three negative sample selection strategies decreased first and then increased with the increase of the number of factors. The area under curve (AUC) values of the model considering four environmental factors were 0.729, 0.909, and 0.909, respectively. The model accuracy was the lowest at 7 environmental factors, with the AUC values of 0.711, 0.869, and 0.893, respectively. The AUC values of the 10 environmental factors were 0.745, 0.942, and 0.919, respectively. The model accuracy was highest at 17 environmental factors, while the AUC values were 0.755, 0.947, and 0.929, respectively. There was the non-linear correlation between model accuracy and cumulative explanatory power percentage. The difference was only 0.020-0.038, although the accuracy of the model for 4 environmental factors was lower than that of 17 environmental factors. Therefore, the relatively ideal accuracy was achieved when considering the main controlling environmental factors; 2) The improved frequency ratio method was significantly improved the accuracy of the model. When the number of environmental factors was 4, the AUC values of FR and ALSA were both 0.909, and the negative samples selected by FR and ALSA were the most reasonable; When the number of environmental factors was 7, the AUC value of ALSA was 0.893, and the negative sample selected by ALSA was the most reasonable; When the environmental factors were 10 and 17, the AUC values of the FR were 0.942 and 0.947, respectively. In summary, the FR can be expected to select the most reasonable negative samples; 3) The acerage rainfall erosivity was closely related to the development of Benggang. Particularly, the acerage rainfall erosivity was ranged from 9 133.24 to 9 981.49 MJ·mm/(hm²·h·a) within the scope of the study area. The majority of high and extremely high susceptibility areas were distributed in the southwest of the study area, whereas, a small number of extremely high susceptibility areas were distributed in the central and eastern parts of Xingguo County, and the majority of extremely low susceptibility areas were distributed in the northern and eastern. This study has explored the impact of different combinations of environmental factors and negative sample selection strategies on the susceptibility assessment of Benggang. The finding can provide the scientific basis for the disaster prevention and reduction in granite areas.

Keywords: susceptibility; random forest; Benggang; GeoDetector; environmental factor combination; negative sample selection strategy