

引用格式:管家琳,黄炎和,林金石,蒋芳市,姚莹莹,季翔.基于信息量模型与随机森林模型的崩岗风险对比评估[J].山地学报,2021,39(4):539-551.

GUAN Jialin, HUANG Yanhe, LIN Jinshi, JIANG Fangshi, YAO Yingying, JI Xiang. Comparisons between Benggang risk assessments based on Information model and Random Forest model [J]. Mountain Research, 2021, 39(4):539-551.

# 基于信息量模型与随机森林模型的崩岗风险对比评估

管家琳<sup>a</sup>,黄炎和<sup>a</sup>,林金石<sup>a</sup>,蒋芳市<sup>a</sup>,姚莹莹<sup>b</sup>,季翔<sup>b\*</sup>

(福建农林大学 a.资源与环境学院;b.公共管理学院,福州 350002)

**摘要:**崩岗是发生在我国南部地区一种典型的水土流失现象,对其进行风险评估有利于崩岗防控。建立崩岗风险评估体系和崩岗评估模型是崩岗风险评估的基础,而不同的评估模型对崩岗风险的评估结果存在差异。探究不同模型在崩岗风险评估中的应用差异,提高预测精度是目前亟需解决的问题。本研究以福建省安溪县龙门镇小流域为例,根据风险因子与崩岗发生之间的关系选择出主要因子,分别运用信息量模型与随机森林模型对崩岗进行风险评估,对比两种方法的优缺点,进一步探究较优的崩岗风险评估方法。结果表明:(1)信息量模型与随机森林模型均适用于研究区的崩岗发生风险评估,基于随机森林模型对崩岗发生风险的预测精度高于信息量模型(AUC值分别为0.89和0.81),混淆矩阵准确率达到84.09%,且泛化能力较好;(2)高程、地形起伏度、坡度和河网缓冲距离因子是崩岗发生的重要风险因子;(3)2个模型预测的崩岗风险空间分布大致相同,约83%的风险区等级面积趋于一致,且以中、高风险为主。研究结果表明随机森林模型的预测性能总体较信息量模型优,可用于未来崩岗发生的风险评估,为崩岗的防治工作提供参考。

**关键词:**崩岗;信息量模型;随机森林模型;龙门镇;模型对比

**中图分类号:** S157

**文献标志码:** A

崩岗是一种特殊的土壤侵蚀类型,主要发生在广东、福建、江西等地,具有明显的地带性、区域性和突发性特点,对生态系统与人类的生产生活造成了恶劣影响<sup>[1-2]</sup>。崩岗发生风险的预测对及时预防崩岗危害具有重要意义,建立崩岗风险评估体系和崩岗评估模型可以预测崩岗发生风险。不同的风险评估模型对崩岗风险的预测存在差异。现有崩岗风险评估模型包括GIS空间叠加分析的评估模型<sup>[3]</sup>、逻辑回归模型<sup>[4]</sup>、信息量模型<sup>[5]</sup>及生态位适宜度分析

等<sup>[6-7]</sup>。信息量法是从信息论中引出的统计方法,该模型具有高度的可靠性且易实现,同时评价结果具有较高的客观性<sup>[8-9]</sup>,对于崩岗的风险评估应用较为成熟。国内学者使用信息量模型(Information model)对江西省的崩岗发生进行风险评价<sup>[5]</sup>,模型预测结果良好,适用于崩岗风险评价,但是该模型仍存在一定的局限性,比如因子共线性、因子分级或分类对评价精度的影响等。随机森林(Random Forest, RF)是机器学习中基于多个决策树的分类算

收稿日期(Received date):2020-11-02;改回日期(Accepted date):2021-07-20

基金项目(Foundation item):国家自然科学基金(41601557;41571272);福建农林大学杰出青年科研人才计划项目(XJQ201933)。[National Natural Science Foundation of China (41601557;41571272);Funds for Distinguished Young Scholar of Fujian Agriculture and Forestry University (XJQ201933)]

作者简介(Biography):管家琳(1996-),女,江苏盐城人,硕士研究生,主要研究方向:水土保持与风险评估。[GUAN Jialin(1996-),female, born in Yancheng, Jiangsu province, M. Sc. candidate, research on water and soil conservation and risk assessment] E-mail: 1430716868@qq.com

\*通讯作者(Corresponding author):季翔(1984-),女,博士,讲师,主要研究方向:土地可持续利用、土壤侵蚀与评价。[JI Xiang(1984-),female, Ph. D., lecturer, research on landscape ecology and collapsing gully erosion] E-mail: jixiangss@126.com

法,鉴于 RF 可以获取因子变量的重要度以及识别驱动因素,故该方法在医学、生态环境领域<sup>[10-12]</sup>、滑坡等地质灾害方面<sup>[13-15]</sup>有一定的适用性。RF 模型操作简单易于实现,对异常值以及噪声有较高容忍度,由于对多重共线性不敏感可以避免过度拟合的问题,具有较高的预测精度和稳定性<sup>[16-17]</sup>,能够解决崩岗风险评估中因子共线性以及因子分级等问题,可较好地弥补信息量模型的缺点,因此基于该模型对崩岗发生的空间预测值得分析探讨。

厘清崩岗发生与风险因子间的关系,明确崩岗发生的可能性及危害程度,对崩岗预警预报、防灾减灾以及灾后治理有重要意义。本研究以福建省安溪县龙门镇小流域为研究对象,分别运用信息量模型以及随机森林模型两种风险评估方法对崩岗发生风险进行空间预测,并探讨两种模型用于崩岗风险评估时的性能。研究结果对崩岗的预警、防治具有指导作用,可以为我国南方地区的水土保持防控与治理提供参考依据。

## 1 研究区域与数据来源

### 1.1 研究区概况

研究区位于福建省安溪县龙门镇北部小流域( $118^{\circ}4' \sim 118^{\circ}9'E, 24^{\circ}57' \sim 25^{\circ}0'N$ ),属于南亚热带气候,年降水 1700 ~ 2188 mm,年平均气温为  $18^{\circ}C \sim 21^{\circ}C$ 。该区域地形以丘陵山地为主,土壤类型主要为红壤和赤红壤,风化壳厚且疏松。据 2005 年崩岗调查可知,福建省的崩岗数量约占全国的 10.87%<sup>[2]</sup>,受崩岗侵蚀影响剧烈。安溪县龙门镇是福建省崩岗发育的重点区域,分布密集且危害巨大,是崩岗发育的典型示范区,故选择龙门镇小流域为研究区域具有较好的代表性。研究区内崩岗数量为 132 个,具体位置如图 1 所示。

### 1.2 数据来源

基于对崩岗的调查数据以及 ArcGIS 工具创建用于崩岗风险评估的数据库,涉及的相关因子数据来源主要为:(1)福建省地质测绘院提供的安溪县

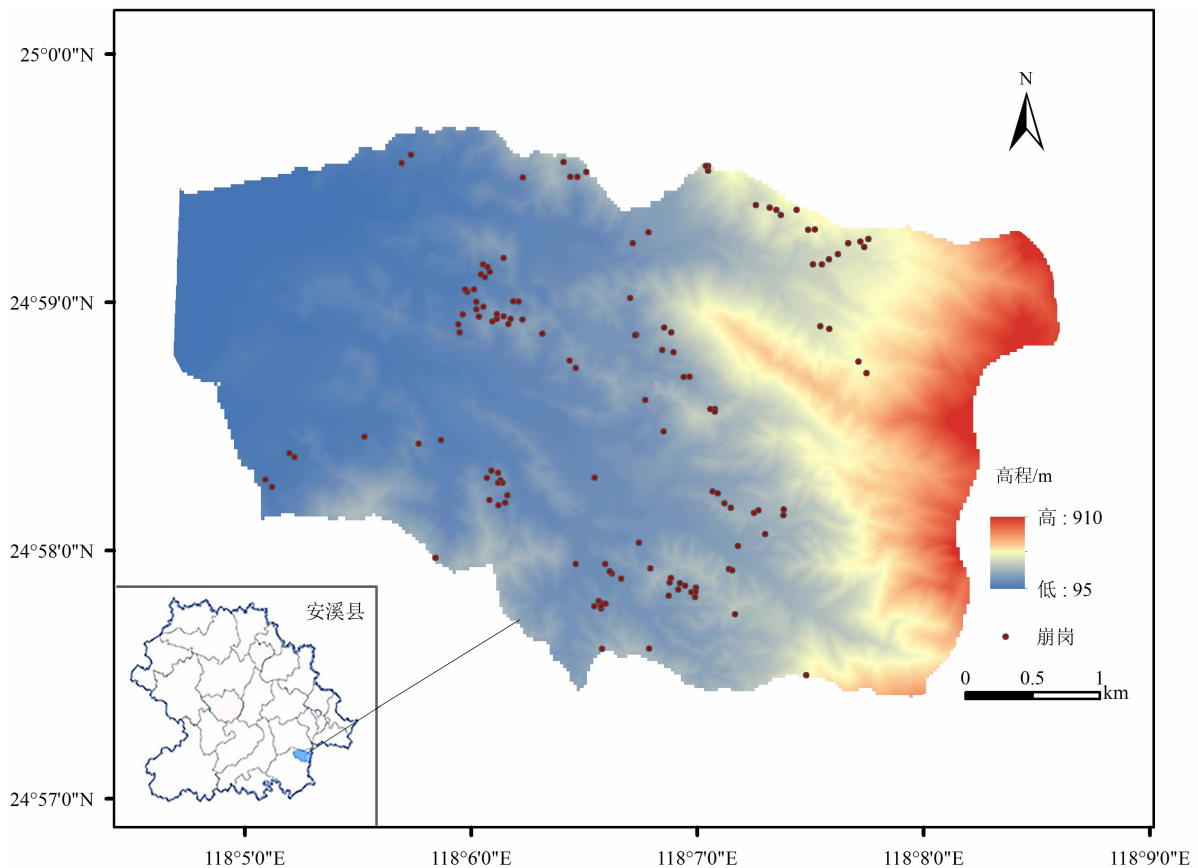


图 1 研究区位置图与崩岗分布图

Fig. 1 Location of study area and the distribution of Benggang

地形图(1:10 000)、高程模型数据(30 m × 30 m 空间分辨率),以获取高程、地形起伏度、坡向、坡度以及曲率等数据信息;(2)龙门镇土地利用现状图(1:10 000)来源于安溪县土壤地形图(1:50 000)和《安溪土壤》;土壤类型分布图(1:50 000)来源于安溪县第二次土壤普查数据;(3)地理空间数据云平台提供的龙门镇 Landsat 植被数据(30 m 分辨率);(4)崩岗数据由 2009 年安溪县崩岗普查成果获得,使用 ArcGIS 10.5 软件将崩岗与特征量数据进行叠加分析,获取研究区高程、地形起伏度、坡度以及土地利用类型等 9 大因子数据。研究区总样本数据以 6:4 比例划分为训练及测试样本数据,其中训练样本数据用以建立模型,测试样本数据用来检验模型。将崩岗数据以点的形式导入 ArcGIS 10.5 软件中以绘制崩岗分布图(图 1)。

## 2 研究方法

### 2.1 崩岗风险因子选择

崩岗的发育受自然与人为因素的双重影响,自然因素是崩岗发育的驱动力,人为因素则是通过后天作用加速或延迟崩岗的发育<sup>[1]</sup>。自然因素包括地形、降雨、土壤以及植被等方面;人为因素主要指人为活动对于自然的改造,包括对土地的利用方式等<sup>[1,18]</sup>。地形起伏度大的一般较容易发生崩岗等地质灾害,随着坡度以及高程的增加会增大崩岗发生的风险<sup>[19]</sup>。崩岗发育通过影响水热条件及植被覆盖具有坡向选择的规律<sup>[20]</sup>;向阳山坡为迎风山坡,降雨带来了额外的水和更大的动能输入,对边坡稳定性产生了相当大的影响<sup>[21-23]</sup>。植被具有固结土壤、保持水土的特性,可以延缓地表径流的形成,减少下切面水分入渗深度<sup>[19]</sup>,是诱使崩岗发生的重要外部因素。不同类型的土壤特性不同,土壤疏松度以及黏粒差异影响坡体的稳定,花岗岩发育的红壤利于崩岗的发育<sup>[20]</sup>。距离水系越近的土壤含水量较大,往往利于崩岗的发育<sup>[24]</sup>,是影响崩岗发生的另一个主要因子。土地利用类型则反映了人类对土体的干预<sup>[4]</sup>可作为崩岗发生的人为因素关键因子。

综上,本次崩岗风险因子的选取是在参阅已有研究成果的基础上,根据研究资料的可获取性与风险因子提取的可操作性原则,以及在难以考虑所有影响因子情况下遵循主导因子原则<sup>[5]</sup>,最终选取高

程、地形起伏度、坡度、坡向、曲率、河网缓冲距离、归一化植被指数(NDVI)、土壤类型以及土地利用类型这 9 个因子作为崩岗发生风险评估的主要因子。

### 2.2 崩岗风险评估模型

#### 2.2.1 信息量模型

信息量模型是通过将数据转换为可量化的信息量值,来衡量研究对象的易发性<sup>[9]</sup>。本研究通过计算因子类相应的信息量值,构建基于信息量模型的崩岗发生风险评估方法,该方法适用于中小尺度的地质灾害危险性预测。崩岗因子类信息量及各因子信息量计算公式分别为<sup>[24]</sup>:

$$I_{ij} = \log_2(N_{ij}/N \times 100) \quad (1)$$

$$I_i = \sum_{j=1}^n I_{ij} \quad (2)$$

式中, $I_{ij}$ 为第  $i$  个因子在第  $j$  等级时的信息量; $N_{ij}$ 为第  $i$  个因子在第  $j$  等级时崩岗的数量(个); $N$ 为研究区崩岗总量(个); $I_i$ 为各因子的信息量值。

每个风险因子对崩岗发育的影响都不同,崩岗的易发性与因子信息量值有关,信息量值越大则表示单元内越易发育成崩岗。基于 ArcGIS 工具,以 30 m × 30 m 大小的栅格单元作为基本单元,运用栅格计算器将信息量值为系数与各因子图层叠加分析,以生成研究区基于信息量模型的崩岗风险值分布图。

#### 2.2.2 随机森林模型

随机森林(Random Forest, RF)是由 BREIMAN 于 2001 年首次提出的一种机器学习算法,通过多棵决策树对样本进行训练和预测<sup>[25]</sup>。该算法将同一个问题分解到不同的决策树中,通过共同学习以提高模型的泛化性能,被誉为当前最好的算法之一。RF 模型的主要思想是:首先使用自助抽样(Bootstrap)技术,从原始数据集  $D$  中随机抽取  $m$  个独立样本,生成  $Dm$  个训练数据集;其次对每个训练集建立对应的决策树模型  $Tm$ ,得到  $Rm$  种分类结果;最后对分类结果进行投票,汇总重复最高的作为模型最优的分类结果<sup>[26]</sup>,具体工作原理流程如图 2。

RF 模型不需要考虑因子间的交互作用,同时是通过随机抽样来构造不同训练集以生成决策树,可以降低变量间的相关性,解决过度拟合的问题<sup>[17,27]</sup>。因此,采用该方法对崩岗发生的可能性进行预测可以忽略风险因子之间的相关关系,并根据

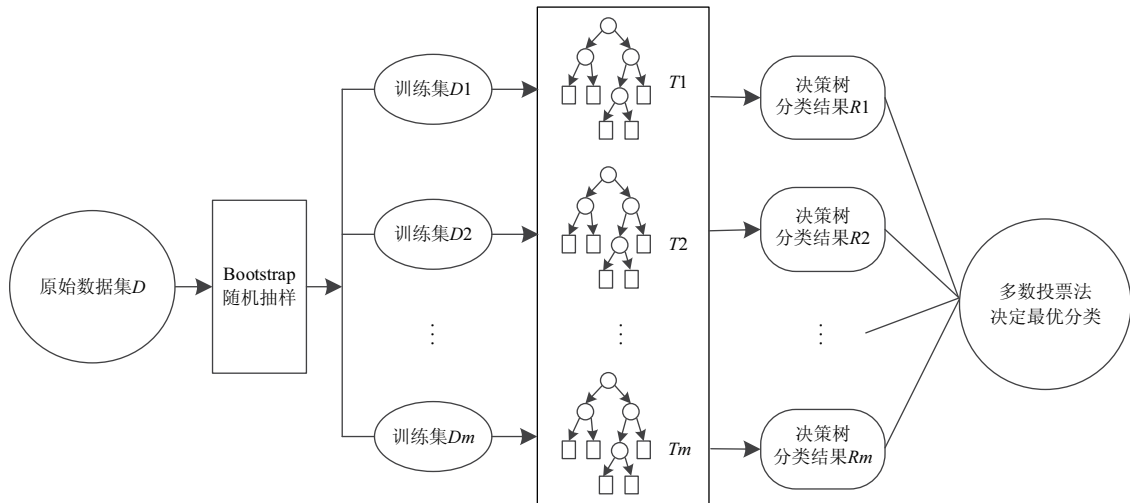


图 2 RF 模型工作原理流程图

Fig. 2 The operating principle flow chart of RF model

各因子在最终分类模型中的重要性  $R_i$ , 来解释对崩岗发育的影响作用。

该模型不仅考虑了崩岗点位数据在崩岗预测中的作用, 还考虑了非崩岗点对于崩岗发生的影响机制。利用 ArcGIS10.5 软件, 基于研究区现存的崩岗点位, 随机生成相同数量非崩岗点数据, 提取对应的因子特征量数据作为模型运算的总样本数据集, 即该模型下的总样本数据共 264 条, 其中崩岗点位与非崩岗点位数据各 132 条。

### 2.2.3 模型精度检验

本研究采用受试者工作特性曲线 (ROC 曲线) 作为信息量模型的精度检验方法。该曲线广泛应用于基于信息量模型地质灾害评价的精度检验<sup>[28-30]</sup>, 表示的是模拟数据与实际数据之间的关系, 利用曲线下面积 (Area Under the Curve, AUC) 来判断模型预测准确度, 取值为 0.5 ~ 1, 值越大表示拟合值与实际值越接近, 则模型预测准确性越高。

混淆矩阵在 RF 模型精度检验的方法中应用较成熟<sup>[31]</sup>, 本研究采用该指标对 RF 模型的预测效果进行验证。采用精确率、召回率以及准确率作为其评价指标, 通过判断 RF 模型的分类预测结果与随机分类结果之间的差异度, 以此来估算模型精度。另外, 为了对比信息量模型和 RF 模型之间的模拟精度, 同时采用 ROC 曲线对 RF 模型进行精度验证。

### 2.3 风险值计算与等级划分

崩岗是由复杂的因子共同作用发生的, 采用公式 (3) 计算崩岗发生的风险值。该值为 0 ~ 1, 值越大表

示该区域越脆弱, 越容易发生崩岗, 其计算公式为:

$$P = \sum_{i=1}^n W_i X_i \quad (3)$$

式中,  $P$  为崩岗发生风险值;  $W_i$  为风险因子  $i$  提供的权重系数;  $X_i$  为第  $i$  个风险因子。

将信息量模型和 RF 模型中得到的  $I_i$  和  $R_i$  代入公式 (3) 中的  $W_i$ , 分别得到基于信息量模型和 RF 模型对应的崩岗发生风险值  $P_i$  和  $P_R$ 。

将崩岗风险值等距分成 5 个类别, 对应的风险等级由小到大依次为极低 (0 ~ 0.2)、低 (0.2 ~ 0.4)、中 (0.4 ~ 0.6)、高 (0.6 ~ 0.8) 以及极高风险 (0.8 ~ 1)。

## 3 结果与分析

### 3.1 风险评估因子数据

将高程、地形起伏度、坡度、坡向、曲率、河网缓冲距离、NDVI、土壤类型以及土地利用类型这 9 个因子以 30 m × 30 m 栅格单元作为崩岗风险评估的基本单元, 建立风险因子数据库。然后对因子进行分级/分类处理, 具体如图 3 所示。

### 3.2 崩岗风险评估

#### 3.2.1 信息量模型

##### (1) 模型预测结果

借助 ArcGIS 10.5 软件分别将 9 个风险因子与崩岗点位分布图叠加, 计算各个因子类对应的崩岗数量, 由公式 (1) 得出风险因子对应的信息量值 (表 1)。



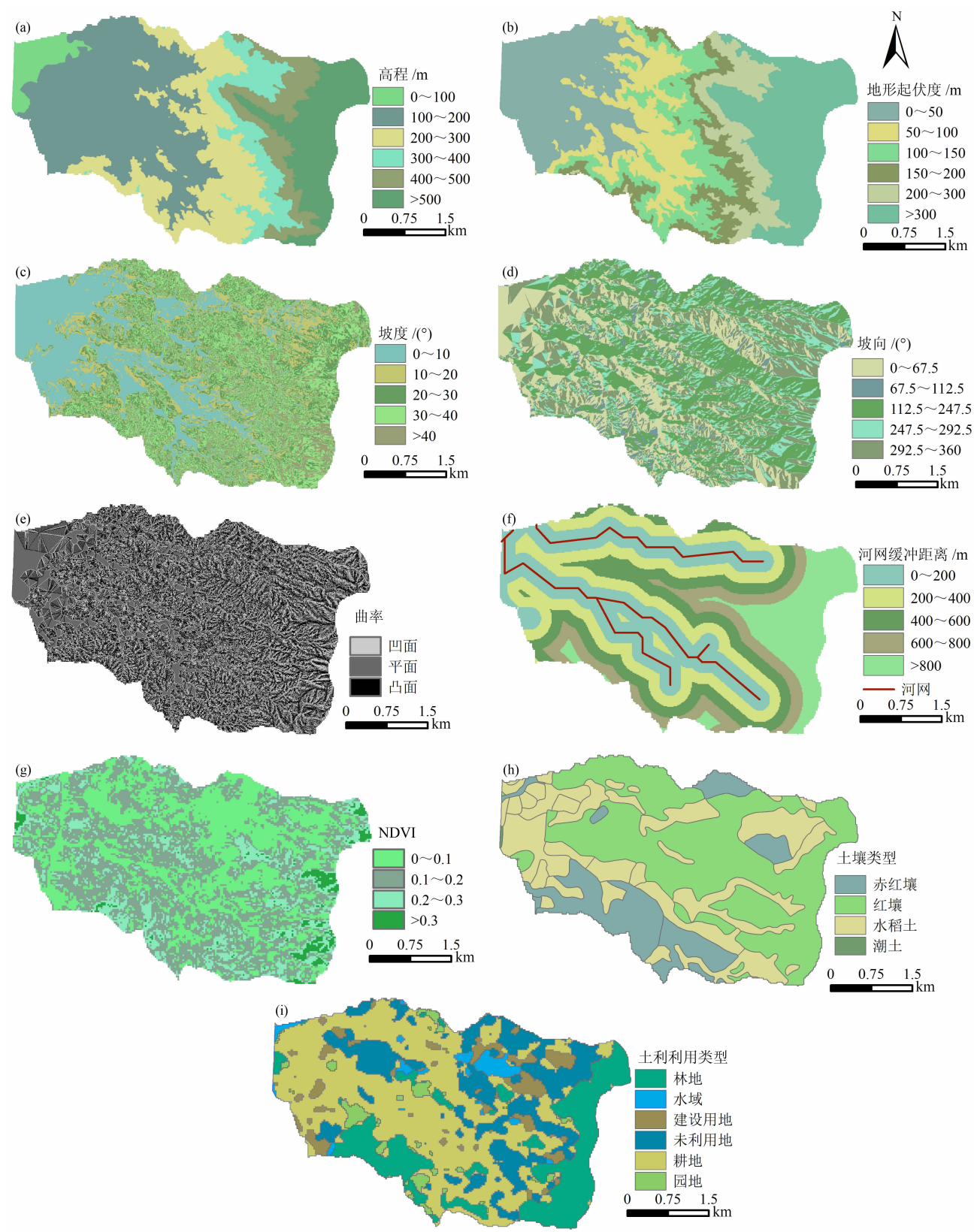


图3 崩岗风险评估因子分级/分类图:(a)高程;(b)地形起伏度;(c)坡度;(d)坡向;(e)曲率;

(f)河网缓冲距离;(g)NDVI;(h)土壤类型;(i)土地利用类型

Fig.3 Grading/ classification of Benggang risk assessment factors: (a) elevation; (b) topographic relief; (c) slope; (d) aspect; (e) curvature; (f) river buffer distance; (g) NDVI; (h) soil type; (i) land-use type

表 1 风险因子信息量值与排序  
Tab. 1 Information value and sequence of risk factors

风险因子	因子类分级	崩岗数量/个	数量密度/%	信息量	排序
高程	A <sub>1</sub>	0	0.00	—	—
	A <sub>2</sub>	52	40.15	5.33	9
	A <sub>3</sub>	52	38.64	5.27	11
	A <sub>4</sub>	16	12.12	3.60	24
	A <sub>5</sub>	12	9.09	3.18	29
	A <sub>6</sub>	0	0.00	—	—
合计	—	132	100.00	17.38	—
地形起伏度	B <sub>1</sub>	15	11.36	3.51	25
	B <sub>2</sub>	33	25.00	4.64	18
	B <sub>3</sub>	37	28.03	4.81	14
	B <sub>4</sub>	19	14.39	3.85	22
	B <sub>5</sub>	15	11.36	3.51	26
	B <sub>6</sub>	13	9.85	3.30	28
合计	—	132	100.00	23.62	—
坡度	C <sub>1</sub>	17	12.88	3.69	23
	C <sub>2</sub>	23	17.42	4.12	21
	C <sub>3</sub>	51	38.64	5.27	12
	C <sub>4</sub>	35	26.52	4.73	16
	C <sub>5</sub>	6	4.55	2.19	34
合计	—	132	100	20.00	—
坡向	D <sub>1</sub>	51	45.45	5.51	5
	D <sub>2</sub>	43	26.52	4.73	17
	D <sub>3</sub>	38	28.03	4.81	15
合计	—	132	100.00	15.04	—
曲率	E <sub>1</sub>	53	40.15	5.33	10
	E <sub>2</sub>	14	10.61	3.41	27
	E <sub>3</sub>	65	49.24	5.62	3
合计	—	132	100.00	14.36	—
河网缓冲距离	F <sub>1</sub>	24	18.18	4.18	20
	F <sub>2</sub>	42	31.82	4.99	13
	F <sub>3</sub>	56	42.42	5.41	6
	F <sub>4</sub>	8	6.06	2.60	33
	F <sub>5</sub>	2	1.52	0.60	35
合计	—	132	100.00	17.79	—
NDVI	G <sub>1</sub>	55	41.67	5.38	8
	G <sub>2</sub>	64	48.48	5.60	4
	G <sub>3</sub>	12	9.09	3.18	30
	G <sub>4</sub>	1	0.76	-0.40	36
合计	—	132	100.00	13.93	—
土壤类型	H <sub>1</sub>	76	57.58	5.85	2
	H <sub>2</sub>	56	42.42	5.41	7
	H <sub>3</sub>	0	0.00	—	—
	H <sub>4</sub>	0	0.00	—	—
合计	—	132	100.00	11.25	—

续表 1

风险因子	因子类分级	崩岗数量/个	数量密度/%	信息量	排序
土地利用类型	I <sub>1</sub>	0	0.00	—	—
	I <sub>2</sub>	12	9.09	3.18	31
	I <sub>3</sub>	33	25.00	4.64	19
	I <sub>4</sub>	78	59.09	5.88	1
	I <sub>5</sub>	9	6.82	2.77	32
	I <sub>6</sub>	0	0.00	—	—
合计	—	132	100.00	16.48	—

注:表中 A1 ~ A6 分别代表高程为:0 ~ 100、100 ~ 200、200 ~ 300、300 ~ 400、400 ~ 500、> 500 m; B1 ~ B6 分别代表地形起伏度为:0 ~ 50、50 ~ 100、100 ~ 150、150 ~ 200、200 ~ 300、> 300 m; C1 ~ C5 分别代表坡度为:0 ~ 10、10 ~ 20、20 ~ 30、30 ~ 40、> 40°; D1 ~ D3 分别代表坡向为:阳坡(112.5° ~ 247.5°)、阴坡(0° ~ 67.5°, 292.5° ~ 360°)、半阳坡(67.5° ~ 112.5°, 247.5° ~ 292.5°); E1 ~ E3 分别代表曲率为:凹面(< 0°)、平面(0°)、凸面(> 0°); F1 ~ F5 分别代表河网缓冲距离:0 ~ 200、200 ~ 400、400 ~ 600、600 ~ 800、> 800 m; G1 ~ G4 分别代表 NDVI:0 ~ 0.1、0.1 ~ 0.2、0.2 ~ 0.3、> 0.3; H1 ~ H4 分别代表土壤类型为:红壤、赤红壤、水稻土、潮土; I1 ~ I6 分别代表土地利用类型为:耕地、园地、林地、未利用地、建设用地和水域。

在信息量模型中,信息量值反映了风险因子对崩岗发生的作用大小,值越大则表示对崩岗发育的贡献程度越大,负值则表示不利于崩岗的发育。由表 1 可知,信息量值较大的因子类分别是:未利用地、红壤、凸面、NDVI 为 0.1 ~ 0.2、阳坡、400 ~ 600 m 的河网缓冲距离。这些条件因素对崩岗的发育起到了促进作用,而因子类为 NDVI > 0.3 的信息量值为负值,则表明该条件下不利于崩岗发生。总体来看,地形起伏度的影响最为显著,贡献程度最大。其次为坡度、河网缓冲距离、高程等,而土壤类型则是对崩岗发生影响最小的因子。

将表 1 中因子类信息量值代入公式(2)得到各因子信息量  $I_i$ 。根据公式(3),借助 ArcGIS 10.5 软件计算得到该模型下的研究区崩岗发生风险值  $p_i$ ,将风险值进行重分类处理,等级划分后生成崩岗发生风险值图(图 4)以及风险等级分布图(图 5)。

(2)模型检验

采用 ROC 曲线对该模型预测结果进行检验,得到的 AUC 值为 0.81,表明信息量模型拟合结果良好,具有一定的准确性,可用于研究区崩岗的风险评估预测。

3.2.2 随机森林模型

采用 Scikit - Learn 库 (<https://scikit-learn.org/stable/>) 的 Random Forest Classifier 和 Python 3.8 实现 RF 模型的编写运算。RF 模型主要设置参数包括决策树的数量、最大样本深度和最小杂质。通过多次实验对比发现,将以上 3 个参数分别设置

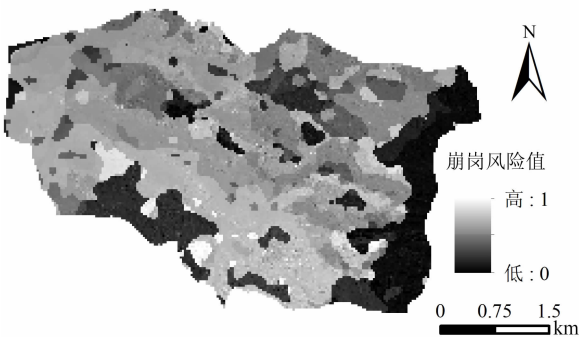


图 4 基于信息量模型的崩岗风险值图

Fig. 4 Benggang risk values predicted by Information model

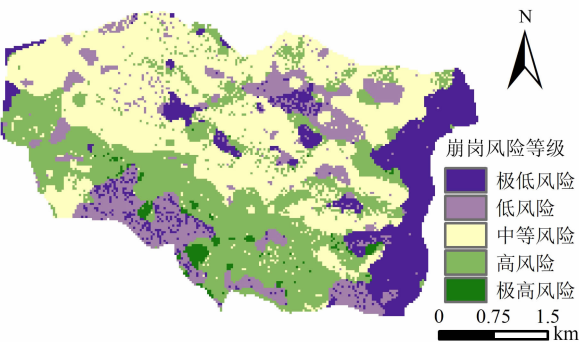


图 5 基于信息量模型的崩岗风险分布图

Fig. 5 Benggang risk levels predicted by Information model

为 80、6 和 0 时,模型预测精度最高。RF 模型中训练样本集为 158 条数据,测试样本集为 106 条数据。

(1)模型预测结果

RF 通过打乱崩岗发生风险因子的取值,分析前

后 RF 模型预测准确性的变化情况,值越大就表示因子的重要性越大<sup>[13]</sup>,即对崩岗发生的贡献就越大。通过对风险因子的重要性分析得出,高程、地形起伏度、坡度、坡向、曲率、河网缓冲距离、NDVI、土壤类型以及土地利用类型 9 个风险因子的重要性  $R_i$  分别为:13.03、14.21、11.69、7.25、10.12、14.79、9.44、8.86 和 10.61。河网缓冲距离是影响崩岗发育最重要的因子,其次为地形起伏度、高程以及坡度等,而坡向的重要性最小。

根据因子的重要性以及公式(3),借助 ArcGIS 10.5 软件计算出基于 RF 模型下的崩岗发生风险值  $P_R$ ,将风险值进行风险分级后,生成崩岗发生风险值图及风险等级分布图,如图 6、图 7 所示。

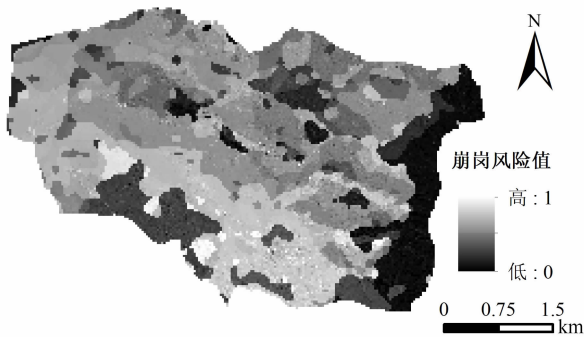


图 6 基于 RF 的崩岗风险值图

Fig. 6 Benggang risk values predicted by RF model

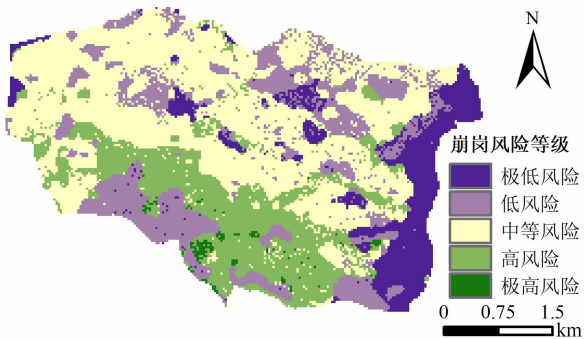


图 7 基于 RF 模型的崩岗风险分布图

Fig. 7 Benggang risk levels predicted by RF model

## (2) 模型检验

采用混淆矩阵对 RF 模型的预测性能进行评价,混淆矩阵以 0.5 为阈值,预测值大于 0.5 则表示发生崩岗,预测值小于 0.5 则表示不发生崩岗。将 106 条测试样本数据代入已经建立好的风险评估模型中进行验证分析。

由混淆矩阵检验结果可知,RF 模型对于崩岗发

生的分类结果的精确率达到 86.89%,召回率为 80.30%,总体准确率为 84.09%,并且该模型的 AUC 值为 0.89,这些指标均表明 RF 模型在预测崩岗发生风险方面具有较高的准确性。

## 3.3 模型结果对比分析

### 3.3.1 精度验证对比

#### (1) ROC 曲线

2 个信息量模型和 RF 模型的 AUC 值均高于 0.80,表明 2 个模型均适用于崩岗的风险评估。但 RF 模型下的 AUC 为 0.89,高于信息量模型的 0.81,表明相比信息量模型,RF 模型拟合结果较好,更适用于研究区的崩岗发生风险的评估。

#### (2) 空间精度分析

为了进一步验证 2 个方法的评估性能,将崩岗点与 2 个模型预测的崩岗发生风险情况进行叠加分析,结果如表 2 所示。

由表 2 可知:(1)区域的崩岗发生风险以中、高风险为主,该风险等级下的面积占研究区面积的 65% 以上;(2)模型的预测结果与实际观察的崩岗点位分布大致吻合,约 70% 的崩岗发育在中、高风险区域,表明信息量模型与 RF 模型对崩岗发生的风险评估结果是可行的;(3)RF 模型预测出的中、高及以上风险区包含的崩岗数量占崩岗总数的 72.73%,高于模型的 69.69%;并且 RF 模型预测出的高、极高风险区面积占比为 21.53%,低于信息量模型的 30.20%,说明 RF 模型的泛化性能较信息量模型好;(4)RF 模型中极低风险区的单位面积崩岗发生数量少于信息量模型,同时高、极高风险区单位面积内崩岗发生数量明显多于信息量模型,这进一步表明 RF 模型对崩岗发生的预测较信息量模型准确。

### 3.3.2 风险因子重要性对比

分别将信息量模型以及 RF 模型计算的风险因子重要性作归一化处理,并制作基于 2 个评估模型的风险因子重要性图,然后计算所有风险因子重要性的均值用以判断该因子是否为影响崩岗发育的重要风险因子(图 8)。

由图 8 可知:(1)在信息量模型中,地形起伏度的贡献值最大,对崩岗的影响极为显著;而在 RF 模型中,河网缓冲距离因子重要性最显著。(2)坡向与河网缓冲距离在 2 个模型间的重要性存在较大差异。(3)综合而言,信息量模型与 RF 模型中高程、



表 2 崩岗发生风险等级及崩岗分布密度的预测结果

Tab. 2 The prediction results of risk levels and distribution density of Benggang

风险等级	风险区面积/km <sup>2</sup>		占总面积比/%		崩岗数量/个		数量密度/%		单位面积崩岗数量/个	
	信息量模型	RF 模型	信息量模型	RF 模型	信息量模型	RF 模型	信息量模型	RF 模型	信息量模型	RF 模型
极低风险	2.84	2.37	15.21	12.67	24	12	18.18	9.09	8.45	5.06
低风险	2.64	3.77	14.11	20.17	16	24	12.12	18.18	6.06	6.37
中风险	7.56	8.53	40.48	45.64	57	51	43.18	38.64	7.54	5.98
高风险	5.37	3.87	28.75	20.69	31	41	23.48	31.06	5.77	10.59
极高风险	0.27	0.16	1.45	0.84	4	4	3.03	3.03	14.81	25.00

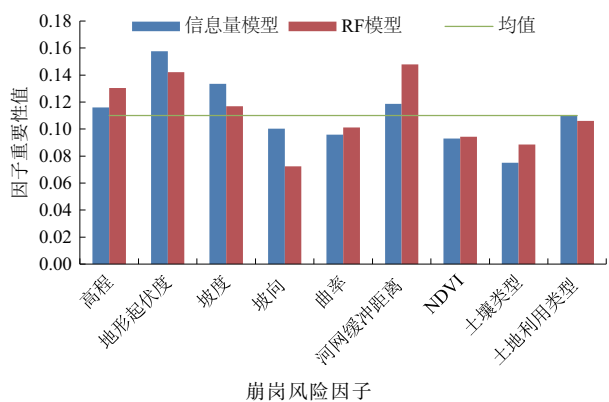


图 8 崩岗风险因子重要性图

Fig. 8 The importance of Benggang risk factors

地形起伏度、坡度以及河网缓冲距离这 4 个风险因子的重要性均高于均值,表明是崩岗发生的重要风险因子。

### 3.3.3 空间预测对比

由图 5、图 7 可知,信息量模型得出的崩岗风险分布趋势与 RF 模型预测结果大致相同,研究区中南部以及西部崩岗发生风险较高。为了进一步分析模型间的空间预测异同,将信息量模型预测结果与 RF 模型结果相减处理,得出模型间的预测风险等级差异(图 9)。

由图 9 分析得出:(1)研究区内约 83% 的风险区面积的等级差为 0,表明信息量模型和 RF 模型对崩岗的预测结果较一致。(2)信息量模型预测风险等级高于 RF 模型的面积占总面积的 14% (风险级差为 1 和 2),而 RF 模型预测风险等级高于信息量模型的面积只有约 3% (风险级差为 -1 和 -3),RF 模型预测结果较信息量模型稳定。

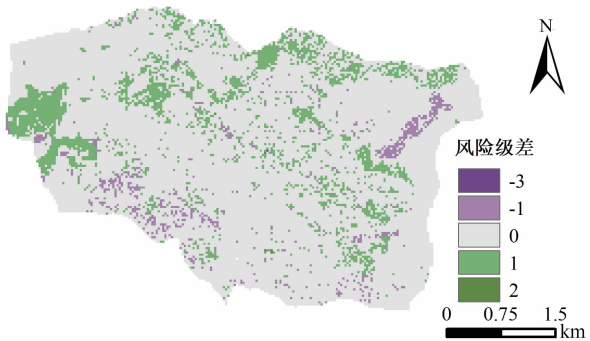


图 9 信息量模型与 RF 模型预测的崩岗风险等级差异

Fig. 9 Differences of Benggang risk levels between results predicted by Information model and RF model

## 4 讨论

有学者在运用 RF 模型对沟蚀的研究中表明到河流的距离对沟蚀有最重要的影响<sup>[32]</sup>,本研究也证实了这一点,在 RF 模型中河网缓冲距离对崩岗发育的影响极为显著。距离河流越近的区域,土体含水量越大,加上降雨所形成的地表径流对花岗岩土体的下切作用<sup>[19]</sup>,这促进了崩岗的发育。一般情况下,崩岗风险因子间存在的相关关系会给模型的预测结果带来信息冗余<sup>[33]</sup>。由于本研究针对信息量模型是将风险因子分级后进行信息量的提取,同时 RF 模型对多重共线性不敏感,因此未考虑风险因子间的相互作用关系。

由研究结果对比发现,信息量模型与 RF 模型在崩岗风险等级上的分布存在差异,可能是因为信息量模型未考虑到非崩岗点对崩岗风险预测的影响,以此为基础建立的崩岗发生风险评估模型无法较好地体现出崩岗的影响机制。RF 模型具有空间预测的高性能以及确定重要性变量的能力<sup>[32]</sup>,但其

缺点在于无法计算事件与影响因素间的空间关系。故在使用该模型对崩岗发生风险进行评估时,风险因子与崩岗间的空间位置关系难以分析,后续可以将双变量统计模型与 RF 模型结合使用以提高模型预测准确性。

后续还可以从以下几方面改进模型的预测结果:首先,通过选取不同的研究区域数据以验证模型预测的准确性,提升模型的科学适用性和推广性。其次,为提升 RF 模型因子变量选取的相对客观性,在模型数据进行随机划分时,可以考虑多次生成不同样本,通过多次出现原则,以确定显著因子变量。再次,可以通过多类型实验验证的方式弥补风险因子以及崩岗点数据偏少的不足。

## 5 结论

对崩岗发生的风险评估有助于及时采取防护措施,减少可能发生的破坏。本研究运用了信息量模型以及 RF 模型对龙门镇小流域崩岗发生风险进行评估,选择了影响崩岗发育的 9 个主要风险因子,并分别绘制研究区崩岗发生风险分布图,探究了适用于研究区域的崩岗发生风险评估方法。主要研究结论为:

(1)信息量模型和 RF 模型的 AUC 值均高于 0.80,说明采用这 2 种方法对研究区的崩岗进行风险评估是切实可行的。通过模型间的空间预测及精度对比分析,RF 模型 AUC 值为 0.89 高于信息量模型的 0.81,且混淆矩阵准确率达到 84.09%,表明 RF 模型具有较高的评估精度。同时 RF 模型较信息量模型有更好的泛化性能,预测出中、高及以上风险区包含的崩岗数量占比为 72.73%,高于信息量模型的 69.69%,因此基于 RF 模型建立的风险评估方法更切合崩岗发生的规律。

(2)对风险因子重要性分析可以得出,在信息量模型以及 RF 模型中,高程、地形起伏度、坡度和河网缓冲距离这 4 个风险因子都是崩岗发生的重要因子。对于信息量模型,地形起伏度是最为显著的因子;而对于 RF 模型,河网缓冲距离是最重要的风险因子。

(3)信息量模型与 RF 模型的崩岗风险分布趋势大致相同(研究区约 83% 风险等级面积预测一致),且崩岗发生风险主要集中分布在研究区域中

南部以及西部,崩岗发生风险以中、高风险为主。

## 参考文献 (References)

- [1] 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] DOI: 10.11766/trxb201807030219
- [2] 冯明汉,廖纯艳,李双喜,等. 我国南方崩岗侵蚀现状调查[J]. 人民长江,2009,40(8):66-68+75. [FENG Minghan, LIAO Chunyan, LI Shuangxi, et al. Investigation on status of hill collapsing and soil erosion in southern China [J]. Yangtze River, 2009, 40(8):66-68+75] DOI: 10.16232/j.cnki.1001-4179.2009.08.025
- [3] 陈嘉林. 福建省典型崩岗区潜在性崩岗风险评估与预测[D]. 福州:福建农林大学,2015:42-54. [CHEN Jialin. Risk evaluation and simulation of the potential collapsing hills in Fujian typical collapsed area [D]. Fuzhou: Fujian Agriculture and Forestry University, 2015:42-54]
- [4] 程冬兵,赵元凌,张平仓,等. 基于 Logistic 模型的江西省崩岗侵蚀风险评估[J]. 中国水土保持科学,2017,15(6):106-116. [CHENG Dongbing, ZHAO Yuanling, ZHANG Pingcang, et al. On the risk assessment of collapse gully erosion in Jiangxi province based on Logistic model [J]. Science of Soil and Water Conservation, 2017, 15(6):106-116] DOI:10.16843/j.sswc.2017.06.013
- [5] 程冬兵,赵元凌,张平仓,等. 基于双变量熵信息法的江西省崩岗侵蚀风险评估[J]. 长江科学院院报,2019,36(2):27-32+38. [CHENG Dongbing, ZHAO Yuanling, ZHANG Pingcang, et al. Risk assessment of collapse gully erosion in Jiangxi province based on bivariate statistical analysis of entropy information [J]. Journal of Yangtze River Scientific Research Institute, 2019, 36(2):27-32+38] DOI: 10.11988/ckyyb.20170984
- [6] 季翔,黄炎和,林金石,等. 崩岗侵蚀沟的时空侵蚀特征及预测[J]. 山地学报,2019,37(1):86-97. [JI Xiang, HUANG Yanhe, LIN Jinshi, et al. Spatio-temporal erosion features and prediction for the erosion gullies on collapsing hills [J]. Mountain Research, 2019, 37(1):86-97] DOI: 10.16089/j.cnki.1008-2786.000402
- [7] 季翔,黄炎和,林金石,等. 基于生态位适宜度的南方花岗岩区崩岗发生敏感性评价方法[J]. 中国农业大学学报,2017,22(10):159-168. [JI Xiang, HUANG Yanhe, LIN Jinshi, et al. Sensitivity assessment method of collapsed gully occurrence in granite region of South China based on niche-fitness [J]. Journal of China Agricultural University, 2017, 22(10):159-168] DOI: 10.11841/j.issn.1007-4333.2017.10.19
- [8] 杜晓晨,陈莉,陈廷芳. 基于 GIS 的凉山州德昌县滑坡危险性评

- 价[J]. 长江流域资源与环境, 2020, **29**(5): 1206 – 1215. [ DU Xiaochen, CHEN Li, CHEN Tingfang. Hazard assessment of landslide in Dechang county of Liangshan state based on GIS [J]. Resources and Environment in the Yangtze Basin, 2020, **29**(5): 1206 – 1215 ] DOI: 10.11870 /cjlyzyhj202005016
- [9] 陈立华, 李立丰, 吴福, 等. 基于 GIS 与信息量法的北流市地质灾害易发性评价[J]. 地球与环境, 2020, **48**(4): 471 – 479. [ CHEN Lihua, LI Lifeng, WU Fu, et al. Evaluation of the geological hazard vulnerability in the Beiliu city based on GIS and information value mode [J]. Earth and Environment, 2020, **48**(4): 471 – 479 ] DOI: 10.14050/j.cnki.1672 – 9250.2020.48.060
- [10] WANG Jing, SHI Leiyu. Prediction of medical expenditures of diagnosed diabetics and the assessment of its related factors using a random forest model, MEPS 2000 – 2015 [J]. International Journal for Quality in Health Care, 2020, **32**(2): 99 – 112. DOI: 10.1093/intqhc/ mzz135
- [11] 李曼, 李园园, 刘焕才. 阿克苏河流域中下游生态系统服务价值对土地利用变化的响应[J]. 内蒙古农业大学学报(自然科学版), 2020, **41**(3): 33 – 37. [ LI Man, LI Yuanyuan, LIU Huancai. The response of ecosystem services value to land use change in the middle and lower reaches of AKSU river basin [J]. Journal of Inner Mongolia Agricultural University (Natural Science Edition), 2020, **41**(3): 33 – 37 ] DOI: 10.16853 /j.cnki.1009 – 3575.2020.03.007
- [12] 王李娟, 孔钰如, 杨小冬, 等. 基于特征优选随机森林算法的农耕地土地利用分类[J]. 农业工程学报, 2020, **36**(4): 244 – 250. [ WANG Lijuan, KONG Yuru, YANG Xiaodong, et al. Classification of land use in farming areas based on feature optimization random forest algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, **36**(4): 244 – 250 ] DOI: 10.11975/j.issn.1002 – 6819.2020.04.029
- [13] 余坤勇, 姚雄, 邱祈荣, 等. 基于随机森林模型的山体滑坡空间预测研究[J]. 农业机械学报, 2016, **47**(10): 338 – 345. [ YU Kunyong, YAO Xiong, QIU Qirong, et al. Landslide spatial prediction based on random forest model [J]. Transactions of the Chinese Society of Agricultural Machinery, 2016, **47**(10): 338 – 345 ] DOI: 10.6041 /j.issn.1000 – 1298.2016.10.043
- [14] SUN Deliang, WEN Haijia, WANG Danzhou, et al. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm [J]. Geomorphology, 2020, **362**: 107201. DOI: 10.1016/j.geomorph.2020.107201
- [15] DOU Jie, YUNUS A P, BUI D T, et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan [J]. Science of the Total Environment, 2019, **662**: 332 – 346. DOI: 10.1016/j.scitotenv.2019.01.221
- [16] 贾南, 陈悦, 康可霖, 等. 基于 RF 的森林火灾风险评价模型及其应用研究[J]. 安全与环境学报, 2020, **20**(4): 1236 – 1240. [ JIA Nan, CHEN Yue, KANG Kelin, et al. Improved forest fire risk assessment model and its application based on the RF algorithm [J]. Journal of Safety and Environment, 2020, **20**(4): 1236 – 1240 ] DOI: 10.13637 /j.issn.1009 – 6094.2019.0899
- [17] 曹正风. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014: 12 – 27. [ CAO Zhengfeng. Study on optimization of random forests algorithm [D]. Beijing: Capital University of Economics and Business, 2014: 12 – 27 ]
- [18] 陈晓安, 杨洁, 肖胜生, 等. 崩岗侵蚀分布特征及其成因[J]. 山地学报, 2013, **31**(6): 716 – 722. [ CHEN Xiaon, YANG Jie, XIAO Shengsheng, et al. Distribution characteristics and causes of collapse erosion [J]. Mountain Research, 2013, **31**(6): 716 – 722 ] DOI: 10.16089/j.cnki.1008 – 2786.2013.06.016
- [19] 李万能, 金平伟, 李岚斌, 等. 南方红壤丘陵区崩岗成因机理的研究进展[J]. 亚热带水土保持, 2014, **26**(3): 30 – 33 + 43. [ LI Wanneng, JIN Pingwei, LI Lanbin, et al. Research progress on the genetic mechanism of collapsing gullies in red soil hilly region of South China [J]. Subtropical Soil and Water Conservation, 2014, **26**(3): 30 – 33 + 43 ]
- [20] 林敬兰, 黄炎和, 林金石, 等. 福建省崩岗侵蚀的地质地貌背景分析[J]. 亚热带水土保持, 2014, **26**(4): 1 – 5. [ LIN Jinglan, HUANG Yanhe, LIN Jinshi, et al. Background analysis on the geology and land form of collapse erosion in Fujian province [J]. Subtropical Soil and Water Conservation, 2014, **26**(4): 1 – 5 ]
- [21] DENG Yusong, DUAN Xiaoqian, DING Shuwen, et al. Effect of joint structure and slope direction on the development of collapsing gully in tuffaceous sandstone area in South China [J]. International Soil and Water Conservation Research, 2020, **8**(2): 131 – 140. DOI: 10.1016/j.iswcr.2020.04.003
- [22] 熊传祥, 王涛, 鲁晓兵. 降雨作用下崩岗形成细观机理模拟[J]. 山地学报, 2013, **31**(6): 710 – 715. [ XIONG Chuanxiang, WANG Tao, LU Xiaobing. Meso-mechanical simulation of slope disintegration erosion under rainfall [J]. Mountain Research, 2013, **31**(6): 710 – 715 ] DOI: 10.16089/j.cnki.1008 – 2786.2013.06.015
- [23] 章智, 陈洁, 林金石, 等. 含水率对安溪县花岗岩崩岗土体胀缩特性的影响[J]. 土壤学报, 2020, **57**(3): 600 – 609. [ ZHANG Zhi, CHEN Jie, LIN Jinshi, et al. Effect of water content on swell-shrink characteristics of collapsed granite soil in Anxi county [J]. Acta Pedologica Sinica, 2020, **57**(3): 600 – 609 ] DOI: 10.11766/trxb201903230072
- [24] 管家琳, 贾秀菊, 季翔. 基于信息量模型的西溪流域崩岗风险评估[J]. 森林与环境学报, 2020, **40**(3): 321 – 328. [ GUAN Jialin, JIA Xiuju, JI Xiang. Risk assessment of collapsing gullies in the Xixi watershed based on information model [J]. Journal of Forest and Environment, 2020, **40**(3): 321 – 328 ] DOI: 10.13324 /j.cnki.jfef.2020.03.013
- [25] BREIMAN L. Random forests [J]. Machine Learning, 2001, **45**

- (1):5-32. DOI: 10.1023/A:1010933404324
- [26] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, **26**(3):32-38. [FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests [J]. Statistics and Information Forum, 2011, **26**(3):32-38]
- [27] 吴孝情, 赖成光, 陈晓宏, 等. 基于随机森林权重的滑坡危险性评价: 以东江流域为例[J]. 自然灾害学报, 2017, **26**(5):119-129. [WU Xiaoping, LAI Chengguang, CHEN Xiaohong, et al. A landslide hazard assessment based on random forest weight: A case study in the Dongjiang River Basin [J]. Journal of Natural Disasters, 2017, **26**(5):119-129] DOI: 10.13577/j.jnd.2017.0514
- [28] 李文彦, 王喜乐. 频率比与信息量模型在黄土沟壑区滑坡易发性评价中的应用与比较[J]. 自然灾害学报, 2020, **29**(4):213-220. [LI Wenyan, WANG Xile. Application and comparison of frequency ratio and information value model for evaluating landslide susceptibility of loess gully region [J]. Journal of Natural Disasters, 2020, **29**(4):213-220] DOI: 10.13577/j.jnd.2020.0422
- [29] 张晓东, 刘湘南, 赵志鹏, 等. 信息量模型、确定性系数模型与逻辑回归模型组合评价地质灾害敏感性的对比研究[J]. 现代地质, 2018, **32**(3):602-610. [ZHANG Xiaodong, LIU Xiangnan, ZHAO Zhipeng, et al. Comparative study of geological hazards susceptibility assessment: Constraints from the information value + logistic regression model and the CF + logistic regression model [J]. Geoscience, 2018, **32**(3):602-610] DOI: 10.19657/j.geoscience.1000-8527.2018.03.18
- [30] 彭珂, 彭红霞, 梁峰, 等. 基于信息量模型的赣州市地质灾害易发性分区[J]. 安全与环境工程, 2018, **25**(5):22-28. [PENG Ke, PENG Hongxia, LIANG Feng, et al. Susceptibility zoning of geo-hazards in Ganzhou city based on the information model [J]. Safety and Environmental Engineering, 2018, **25**(5):22-28] DOI: 10.13578/j.cnki.issn.1671-1556.2018.05.004
- [31] 袁培森, 曹益飞, 马千里, 等. 基于 Random Forest 的水稻细菌性条斑病识别方法研究[J]. 农业机械学报, 2021, **52**(1):139-145+208. [YUAN Peisen, CAO Yifei, MA Qianli, et al. Identification method of rice bacterial leaf streak based on random forest [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, **52**(1):139-145+208] DOI: 10.6041/j.issn.1000-1298.2021.01.015
- [32] 刘坚, 李树林, 陈涛. 基于优化随机森林模型的滑坡易发性评价[J]. 武汉大学学报(信息科学版), 2018, **43**(7):1085-1091. [LIU Jian, LI Shulin, CHEN Tao. Landslide susceptibility assessment based on optimized random forest model [J]. Geomatics and Information Science of Wuhan University, 2018, **43**(7):1085-1091] DOI: 10.13203/j.whugis.20160515
- [33] 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**:928-942. DOI: 10.1016/j.jenvman.2018.11.110

## Comparisons Between Benggang Risk Assessments Based on Information Model and Random Forest Model

GUAN Jialin<sup>a</sup>, HUANG Yanhe<sup>a</sup>, LIN Jinshi<sup>a</sup>, JIANG Fangshi<sup>a</sup>, YAO Yingying<sup>b</sup>, JI Xiang<sup>b\*</sup>

(*a. College of Resources and Environment; b. College of Public Management, Fujian Agriculture and Forestry University, Fuzhou 350002, China*)

**Abstract:** Benggang represents a typical soil erosion phenomenon in the granite region of Southern China, characterized by considerable erosion and extreme landform changes. Different assessment models for Benggang risk exhibited different results, which requires a thorough comparison on their accuracy for the purpose of Benggang control. The establishment of a risk assessment system with a proper model is a prerequisite for Benggang risk assessment. And the differences in the applications of these models should be further determined for promotion of prediction accuracy. In this study, a small watershed in Longmen Town, Anxi county, Fujian province was targeted as case study for Benggang risk assessment. Main factors were selected in the case based on the relationship between risk factors and Benggang events. It compared the advantages and disadvantages between an Information model and a Random Forest model in the process of Benggang risk evaluation, to find out which model had a better

performance. The results show that: (1) Both the Information model and the Random Forest model were applicable to the risk assessment of Benggang in the study area; however, the prediction accuracy of the Random Forest model for the risk of Benggang was higher than that of the Information model (AUC values were 0.89 and 0.81, respectively). The accuracy of the confusion matrix for the Random Forest model reached 84.09% with a better generalization ability. (2) Elevation, topographic relief, slope and river buffer distance mattered for the occurrence of Benggang. (3) The distribution of risk space of Benggang predicted by the two models was equivalent; approximately 83% of the risk areas were of the same level, dominated by medium-risk and high-risk. These study found that the prediction performance of the Random Forest model was generally better than that of the Information model. Random forest model is recommended to be used for Benggang assessment and control in the future.

**Key words:** Benggang; Information model; Random Forest model; Longmen Town; model comparisons

## 中巴喀喇昆仑公路冰湖溃决灾害

(朱颖彦 昆明理工大学/中国科学院、水利部成都山地灾害与环境研究所)

中巴喀喇昆仑公路(简称中巴公路)穿越的西喀喇昆仑-喜马拉雅地区生存着除极地之外最大的陆地冰川群,是世界上著名的跃动型冰川活动区,世界已知的跃动型冰川事件大约90%发生在喀喇昆仑-喜马拉雅地区。冰川跃进,堵塞河道(冰川支谷),形成短暂堰塞冰湖,溃坝泄流,形成洪水或泥石流灾害,历史上造成惨重伤亡与财产损失。中巴公路沿线冰川成因的各类地质灾害频发、类型多样、危害巨大,严重影响中巴经济走廊内重大基础设施的建设与安全运行,是制约中巴经济走廊建设与发展的关键因素。作者统计发现,1780年以来的,大型冰湖溃决事件多起源于印度河上游的新沙勒河谷(Shingshal valley)和希约克河谷上游(Upper Shyok River valley)。近十多年来,中巴公路沿线水系或冰川河谷,并没有发生流域性规模的冰湖溃决灾害,中小型冰湖溃决事件不是因冰川跃进,更多的是因为升温作用造成的冰舌萎缩涌水、冰湖渗透溃堤,或因冰上湖或冰前湖在冰崩、岩崩等激发因素的直接触发下,形成涌浪翻坝,导致冰湖溃决。基于近300年的历史文献,中巴科学家联合开展了10多年的持续野外调查,采用遥感解译、定点观测与案例分析等研究手段,对中巴公路沿线冰湖溃决灾害的河流背景、灾害历史、冰湖的分类、形成与溃决机制等进行时间梳理和汇总分析。

详见本期《中巴喀喇昆仑公路冰湖溃决灾害》一文。