

基于不同因子分级的滑坡易发性区划对比^{*}

——以万州区为例

孙德亮¹, 马祥龙¹, 唐小娅², 文海家³, 密长林⁴

- (1. 重庆师范大学 GIS 应用研究重庆市高校重点实验室, 重庆 401331;
2. 广东药科大学 药学信息工程学院, 广州 510006;
3. 重庆大学 山地城镇建设与新技术教育部重点实验室, 重庆 400045;
4. 山东省临沂市自然资源开发服务中心, 山东 临沂 276000)

摘要:【目的】基于滑坡点属性与研究区全域点属性作为分类基础数据,对位于三峡库区的万州区滑坡易发性区划对比研究。【方法】选取高程、多年平均降雨量、地表切割深度、坡向、距道路距离、坡度、POI 核密度、倾坡类型、岩性、归一化植被指数、曲率、地形起伏度、地形湿度指数等 13 个因子作为影响因子,分别对滑坡点属性与研究区全域点属性使用自然断点法进行因子分类,并通过两种分类结果构建模型预测研究区内滑坡易发程度的空间分布情况。【结果】以研究区全域点属性作为分类数据对应的曲线下面积(Area under curve, AUC)值为 0.79,对应的模型测试集最高精度为 0.755;以滑坡点属性作为分类数据对应的 AUC 值为 0.85,对应的模型测试集最高精度为 0.779;高程、地表切割深度和多年平均降雨量的重要性位于前 3 位,对滑坡的发生有较大影响。【结论】采用滑坡点属性确定因子分类标准建立的随机森林模型精确性和预测能力更优,可对万州区的滑坡灾害危险管理和影响因子分级研究提供参考。

关键词:随机森林模型;因子分类;滑坡易发性区划;万州区;三峡库区

中图分类号:X43

文献标志码:A

文章编号:1672-6693(2021)05-0043-12

滑坡是一种破坏力极强的自然灾害,在世界范围内造成了重大的人员伤亡和经济损失^[1]。据统计,2004—2016 年期间,滑坡在全球共造成了 55 000 余人死亡和高达数百亿美元的经济损失^[2]。因此,对滑坡灾害的有效预测和治理成为了人们重点关注的问题。滑坡易发性区划是结合研究区地形地貌等特征来预测研究区内滑坡易发性程度空间分布情况的方法,是减少滑坡带来的人员伤亡和经济损失的首选措施^[3]。滑坡易发性评价模型一般包括物理力学模型、统计分析模型和机器学习模型^[4]。近些年来,随着 GIS 技术与人工智能的发展,机器学习模型受到了广泛关注。刘睿等人^[5]以三峡库区典型县域巫山县为研究区,基于 GIS 环境与随机森林(Random forest, RF)模型预测研究区内滑坡易发性程度空间分布情况,模拟结果显示 RF 对数据有较好的拟合程度,预测结果客观可靠且精度较高,相较于其他方法具有一定优势。

目前,有关研究多集中于选择新的滑坡致灾因子或改进易发性评价模型,对数据的预处理(包括数据分级、分类等)研究较为缺乏,而对滑坡影响因子的分类是基于机器学习方法进行滑坡易发性评价的基础^[6],不同分类方法对滑坡易发性评价结果的精确度有着较大影响^[3]。在前人的研究中,常用的因子分类方法包括专家经验法^[7]、自然间断分级法^[8]、等间距法^[9]等,其中:专家经验法根据滑坡的分布特征对影响因子进行分类,能够反映各因子与滑坡发生之间的关系,但带有一定的主观性,因为不同的学者可能有不同的分类标准;等间距法分类间隔相同,各分类区域大小相等,与滑坡的分布特征不符;自然间断分级法则使用迭代计算得到分类的断点标准,让每一类因子内数据的差别最小化,从而使因子得到合适的分级,能够更好反映出影响因子在不同类别中的差异。此外,也有学者提出了新的方法对滑坡影响因子进行分类:凌晓等人^[10]提出了对称分级法对致灾因子进行分类,并应用多种分类方法进行滑坡易发性评价,结果表明对称分级法得出的效果最佳;郭建华等人^[11]利用方差分析确定最佳分级数目,指出在分级准确度上方差分析要比常规分级方法更高。闫举生等人^[6]使用频率比法与

* 收稿日期:2021-05-05 修回日期:2021-05-28 网络出版时间:2021-09-18 17:51

资助项目:国家自然科学基金(No. 42071217);重庆市自然科学基金(No. cstc2020jcyj-msxmX0841);国家重点研发计划(No. 2018 YFC1505501);教育部人文社会科学规划项目(No. 20XJAZH002)

第一作者简介:孙德亮,男,副教授,博士, E-mail: sundeliang2002@aliyun.com;通信作者:唐小娅,女, E-mail: xiaoya_tang@126.com

网络出版地址: <https://kns.cnki.net/kcms/detail/50.1165.n.20210916.1201.002.html>

等间距法对因子分类,结果表明基于频率比方法的分类精度较高。以上研究表明,不同的分类方法对滑坡易发性评价结果有着较大的影响。

然而,上述研究都是使用研究区全域点属性作为分类数据,如此虽能较好地反映因子本身的特点,但却无法直观体现出致灾因子对滑坡分布的影响。同时,主观方法可能存在分类冗余、细节缺失的不足,而客观方法又不能完全表现滑坡分布与发生的特点。因此,本研究使用滑坡点属性与研究区全域点属性作为分类基础数据,以位于三峡库区腹地的万州区为研究区,并考虑到该区域与刘睿等人^[5]研究的巫山县地理条件相似,从而采用 RF 模型进行滑坡易发性评价研究以便取得较好的评价效果,期望研究结果为万州区乃至三峡库区的滑坡灾害危险管理和影响因子分级研究提供参考。

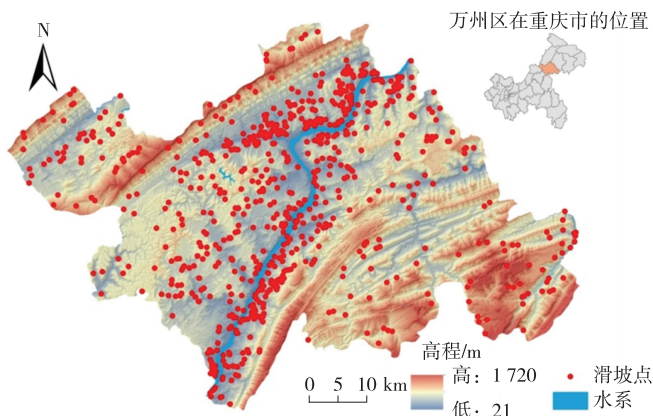
1 研究区概况与数据来源

1.1 研究区概况

万州区位于重庆市东部、三峡库区境内,地处东经 107°55′22″~108°53′25″、北纬 30°24′00″~31°14′58″。万州区沟谷切割深、山高坡陡且暴雨频发,以滑坡为主的地质灾害频发^[12](图 1)。当地地质构造主要由武陵山褶皱带和大巴山弧形褶皱构造组成,主要地层有二叠纪(P)、三叠纪(T)、侏罗纪(J)等;境内多山丘,少平坝和台地。万州区气候为亚热带季风湿润气候,年平均降水量 1 243 mm,且降水主要集中在 5—9 月;境内水系众多,流水切割深,落差较大,流域面积超 100 km² 的河流共计 8 条。多年以来,万州区地质灾害频繁发生,对滑坡灾害进行有效的预测和治理成为了有关部门亟待解决的问题。

1.2 数据来源

本研究所用的 2001—2017 年万州区 773 处历史滑坡资料与各影响因子的数据来源见表 1。



注:底图源自重庆标准地图服务网,审图号:渝 S(2019)055 号,下同

图 1 万州区地理位置及滑坡分布

Fig. 1 The geographical location and landslide distribution of Wanzhou district

表 1 本研究所用数据及数据来源

Tab. 1 The data and their sources in the study

名称	来源	类型	精度
历史滑坡	重庆市地质监测站	数据表	
DEM	Global digital elevation model(GDEM)	栅格	30 m
地质资料	国家地质资料数据中心	栅格	1 : 200 000
行政区划	重庆市国土局	矢量	1 : 100 000
河网	重庆市水利局	矢量	1 : 100 000
卫星影像	地理空间数据云平台	栅格	30 m
多年降雨	重庆市气象局	数据表	30 m
道路	重庆市交委	矢量	1 : 100 000
2016 年重庆市 POI	网络爬虫	数据表	

2 地理空间数据库

2.1 影响因子

滑坡的发生由边坡内部的基本地质条件(如岩性、高程等)与外界环境因素(如人类工程活动、多年平均降雨量等)相互作用而成。本研究结合万州区滑坡的空间规律与内在特征,并参考相关文献^[13-15]后,选取高程、地表

切割深度、倾坡类型等 13 个因子(图 2~3)来构建滑坡影响因子体系,其中:地形起伏度、地表切割深度等因子通过 ArcGIS 对研究区的 DEM 数据进行地形与水文分析得到;多年平均降雨量通过利用各地气象检测站的降雨点数据在 ArcGIS 中进行空间插值运算获取;归一化植被指数(NDVI)通过 ArcGIS 软件对 Landsat-8 OLI 影像的红光和近红外波段使用栅格计算器功能运算获得;对栅格地质图进行矢量化,得到岩性数据;距道路距离通过道路矢量数据在 ArcGIS 中进行邻域分析得到;POI 核密度数据通过 ArcGIS 对兴趣点数据多次试验选择最佳的搜索半径进行密度分析获得。值得一提的是,地表切割深度^[16]与地形起伏度类似,指一定汇水区域中平均高程与最低高程的相对高差,能较好反映研究区内的水流切割强烈程度与沟谷切割深度。三峡库区内多低山,沟谷切割较深,水流侵蚀切割对地表土体的影响程度较大,但在以往研究中较少考虑到沟谷切割深度对于滑坡易发性的影响。因此,引入地表切割深度因子,可以很好地表征沟谷切割深度对于滑坡诱发的影响,对影响因子体系研究起到较大的作用。

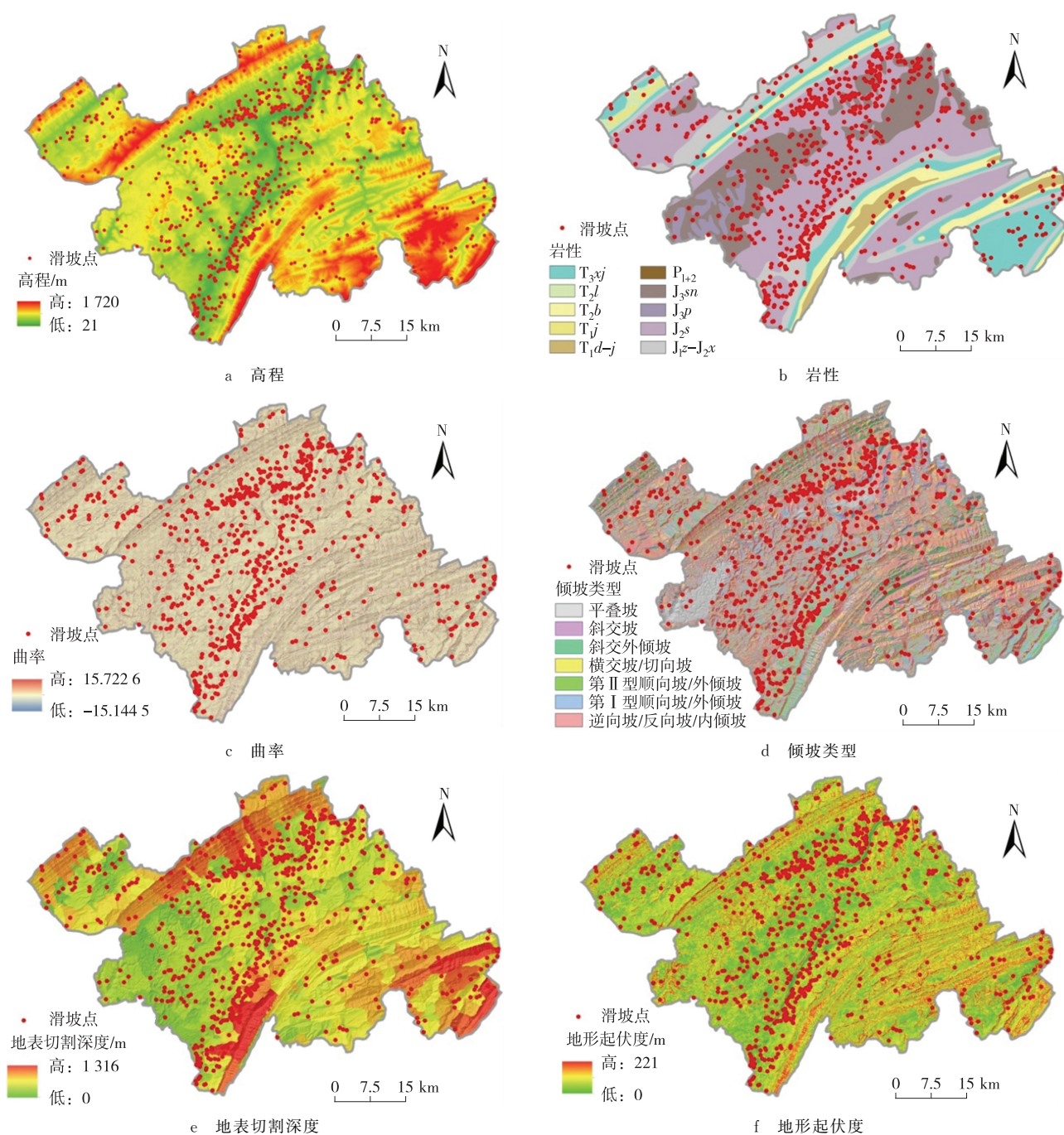


图 2 滑坡影响因子专题图层(I)

Fig. 2 Thematic layer of landslide impact factors (I)

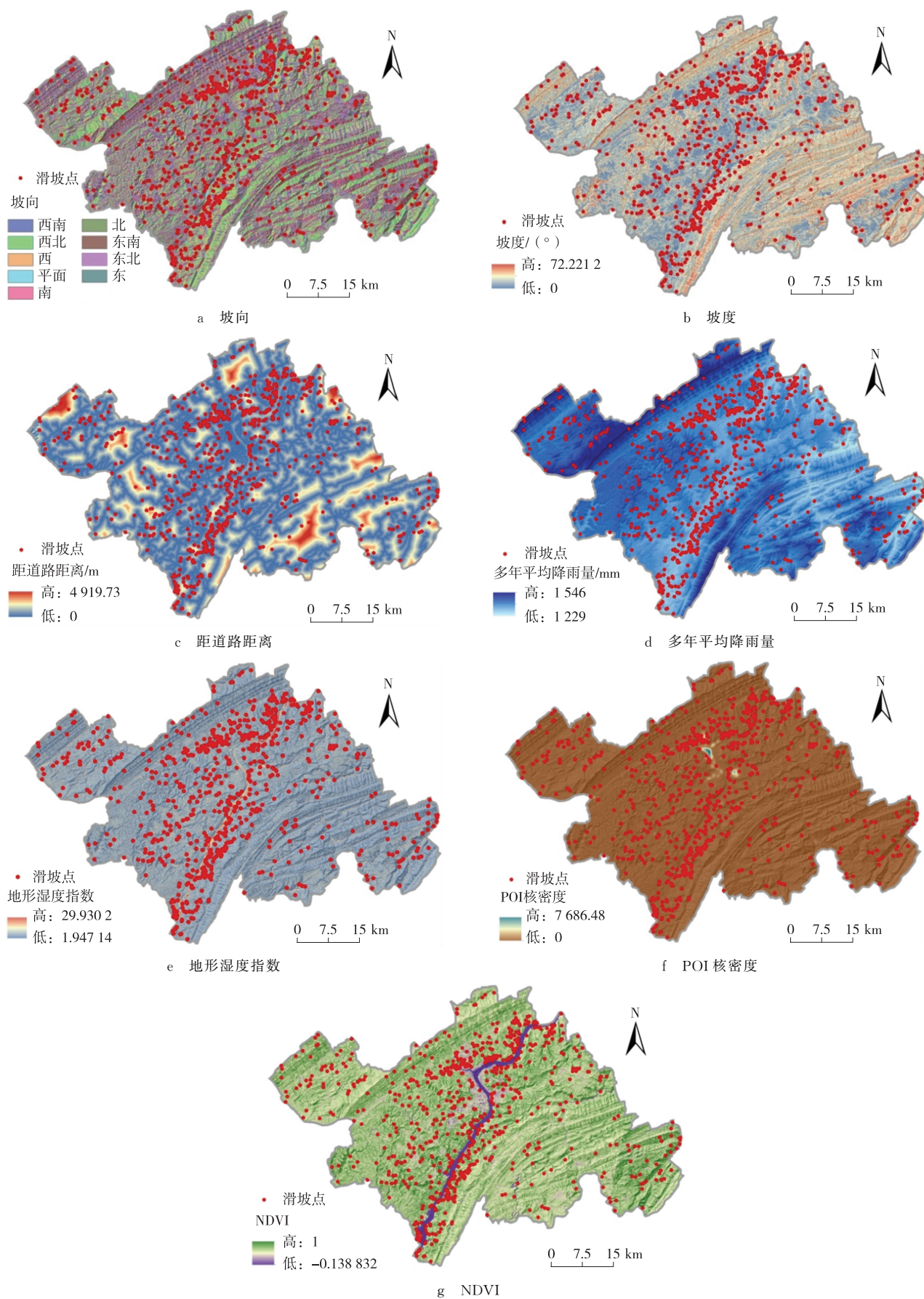


图 3 滑坡影响因子专题图层 (II)

Fig. 3 Thematic layer of landslide impact factors (II)

2.2 影响因子分类

本研究采用 30 m×30 m 的栅格单元作为数据处理与空间分析的基本单元,并在此基础上进行滑坡影响因子的分级与滑坡易发性仿真分析。研究区 13 个因子中除地层岩性、倾坡类型、坡向属于离散型因子可通过它们本身的属性分类之外,其他 10 个因子属于连续型因子,需要选择合适的分级标准。

影响因子的分类主要分为 3 步:首先通过 ArcGIS 进行数据处理与因子提取,得到研究区各因子数据;其次使用研究区全域各因子数据进行自然断点分类,得到原分类标准(表 2);最后将历史滑坡点看作栅格单元,将各因子的属性赋值在历史滑坡点中,并在 ArcMap 中利用自然断点法确定分类间隔,得到改进后的分类标准(表 3)。该分类能更加直观的体现影响因子对滑坡分布的影响。其中需要注意的是,由于道路对附近生态环境的影响距离大致小于 1 000 m,所以本研究中对道路的分类仅考虑 1 000 m 以内的数据。

表 2 全域点属性计算得到的因子分类结果
Tab.2 Factor classification results obtained from global point attribute calculation

影响因子	分级/个	分类标准
高程/m	11	1) <230;2) 230~<345;3) 345~<445;4) 445~<535;5) 535~<620;6) 620~<715; 7) 715~<820;8) 820~<940;9) 940~<1 080;10) 1 080~1 245;11) >1 245
坡度/(°)	7	1) <7.5;2) 7.5~<13;3) 13~<18;4) 18~<24;5) 24~<30;6) 30~38;7) >38
坡向	9	1) 平面;2) 北;3) 东北;4) 东;5) 东南;6) 南;7) 西南;8) 西;9) 西北
曲率	6	1) <-2.5;2) -2.5~<-1.3;3) -1.3~<0.25;4) 0.25~<1.05;5) 1.05~2.4;6) >2.4
距道路距离/m	7	1) 0~<100;2) 100~<200;3) 200~<300;4) 300~<500;5) 500~<710;6) 710~1 000; 7) >1 000;
地形起伏度/m	6	1) <15;2) 15~<25;3) 25~<35;4) 35~<50;5) 50~60;6) >60;
POI 核密度	4	1) <1;2) 1~<2;3) 2~<3;4) 3~<4;5) 4~<5;6) 5~10;7) >10
多年平均降雨量/mm	7	1) <1 308;2) 1 308~<1 337;3) 1 337~<1 363;4) 1 363~<1 389;5) 1 389~<1 421; 6) 1 421~1 464;7) >1 464
地形湿度指数	6	1) <4.6;2) 4.6~<5.9;3) 5.9~<7.5;4) 7.5~<9.6;5) 9.6~<12.4;6) 12.4~16.8; 7) >16.8
倾坡类型	6	1) 顺向坡;2) 斜交外倾坡;3) 斜交坡;4) 横交坡;5) 逆向坡;6) 平叠坡
NDVI	7	1) <0.04;2) 0.04~<0.15;3) 0.15~<0.22;4) 0.22~<0.28;5) 0.28~<0.32;6) 0.32~ 0.38;7) >0.38
岩性	10	1) P ₁₊₂ ;2) J _{3sn} ;3) T _{3xj} ;4) T _{2b} ;5) J _{3p} ;6) J _{1z} -J _{2x} ;7) T _{1d-j} ;8) J _{2s} ;9) T _{1j} ;10) T _{2l}
地表切割深度/m	10	1) <210;210~<305;3) 305~<390;4) 390~<475;5) 475~<545;6) 545~<620; 7) 620~<710;8) 710~<835;9) 835~<1 040;10) 1 040~1 316;

2.3 归一化处理

为了减少不同量纲的影响,对重分类后的数据分别归一化处理,使数值处于[0,1]之间,有关公式如下:

$$x=\frac{x'-x_{\min}}{x_{\max}-x_{\min}}。$$

其中:X 为归一化处理后的结果,x'为各因子的原始数据,x_{min}为每个因子的最小值,x_{max}每个因子的最大值。

3 随机森林模型

RF 模型是应用最广泛的回归算法之一^[17],具有较好的分类结果。RF 原理是组合 n 棵互不关联的决策树 [y(x,θ_k;k=1,2,⋯,n)],基于这些决策树分别判断或预测样本的分类,最后采用多个决策树判断结果的众数作为最终的投票分类决策,具体公式为:

$$Y(x)=\arg\max_z\sum_{i=1}^kI(y_i(x)=Z)。$$

其中: $Y(x)$ 代表 RF 模型判断结果; $y_i(x)$ 表示单棵决策树模型; Z 表示输出变量; $I(\cdot)$ 是一个显式函数。此外,RF 还可以通过全体样本集中剩余的样本作为测试集来判断模型预测的错误率,它也能减少单棵决策树预测的片面性和不准确性,防止因单棵决策树判断结果过拟合而导致的预测精度降低的问题^[18]。RF 的流程图如图 4 所示。

表 3 滑坡点计算得到的因子分类结果

Tab. 3 The factor classification results obtained from landslide point calculation

影响因子	分级/个	分类标准
高程/m	11	1) <195 ; 2) $195 \sim <240$; 3) $240 \sim <290$; 4) $290 \sim <346$; 5) $346 \sim <405$; 6) $405 \sim <471$; 7) $471 \sim <543$; 8) $543 \sim <627$; 9) $627 \sim <754$; 10) $754 \sim 914$; 11) >914
坡度/(°)	7	1) <8.20 ; 2) $8.20 \sim <12.50$; 3) $12.50 \sim <16.56$; 4) $16.56 \sim <20.85$; 5) $20.85 \sim <26.47$; 6) $26.47 \sim 35.23$; 7) >35.23
坡向	9	1) 平面; 2) 北; 3) 东北; 4) 东; 5) 东南; 6) 南; 7) 西南; 8) 西; 9) 西北
曲率	6	1) <-1.73 ; 2) $-1.73 \sim <-0.81$; 3) $-0.81 \sim <-0.12$; 4) $-0.12 \sim <0.58$; 5) $0.58 \sim 1.62$; 6) >1.62
距道路距离/m	7	1) <95 ; 2) $95 \sim <235$; 3) $235 \sim <395$; 4) $395 \sim <570$; 5) $570 \sim <750$; 6) $750 \sim <1\ 000$; 7) $>1\ 000$
地形起伏度/m	6	1) <15 ; 2) $15 \sim <24$; 3) $24 \sim <34$; 4) $34 \sim <46$; 5) $46 \sim 61$; 6) >61
POI 核密度	7	1) <1 ; 2) $1 \sim <2$; 3) $2 \sim <3$; 4) $3 \sim <4$; 5) $4 \sim <5$; 6) $5 \sim 10$; 7) >10
多年平均降雨量/mm	7	1) $<1\ 295$; 2) $1\ 295 \sim <1\ 320$; 3) $1\ 320 \sim <1\ 344$; 4) $1\ 344 \sim <1\ 370$; 5) $1\ 370 \sim <1\ 401$; 6) $1\ 401 \sim 1\ 441$; 7) $>1\ 441$
地形湿度指数	7	1) <4.26 ; 2) $4.26 \sim <5.03$; 3) $5.03 \sim <5.83$; 4) $5.83 \sim <6.82$; 5) $6.82 \sim <8.30$; 6) $8.30 \sim 11.25$; 7) >11.25
倾坡类型	6	1) 顺向坡; 2) 斜交外倾坡; 3) 斜交坡; 4) 横交坡; 5) 逆向坡; 6) 平叠坡
NDVI	7	1) <0.076 ; 2) $0.076 \sim <0.20$; 3) $0.20 \sim <0.265$; 4) $0.265 \sim <0.310$; 5) $0.310 \sim <0.352$; 6) $0.352 \sim 0.398$; 7) >0.398
岩性	10	1) P_{1+2} ; 2) J_3sn ; 3) T_3xj ; 4) T_2b ; 5) J_3p ; 6) J_1z-J_2x ; 7) T_1d-j ; 8) J_2s ; 9) T_1j ; 10) T_2l
地表切割深度/m	10	1) <78 ; 2) $78 \sim <155$; 3) $155 \sim <231$; 4) $231 \sim <307$; 5) $307 \sim <384$; 6) $384 \sim <461$; 7) $461 \sim <537$; 8) $537 \sim <617$; 9) $617 \sim 697$; 10) >697

3.1 模型训练

将去除历史滑坡点附近 500 m 缓冲区以及研究区内河流的区域作为非滑坡区,选取万州区 773 个历史滑坡数据作为正样本,按照 1:1 的比例在非滑坡区内随机选取非滑坡点(773 个)为负样本,构成全体样本集。并采用十折交叉验证法进行 RF 模型的训练与验证。十折交叉验证是将全体样本集(773 个正样本,773 个负样本)平均分成 10 个子集,每次选取 9 个子集训练并生成 n 棵决策树,1 个子集进行验证,轮流选取,并计算错误率,得到每次试验的测试集精度。

3.2 模型检验

模型预测结果的精确度可通过受试者工作特征(Receiver operating characteristic, ROC)曲线来确定。ROC 曲线越靠近左上方,代表模型精度越高,曲线下面积(Area under curve, AUC)值为 ROC 曲线所覆盖的区域面积,可定量判断模型精度大小, AUC 值的大小在[0,1]之间,该值越大代表模型精度越高。

4 滑坡易发性区划结果

将训练完成的 RF 模型应用于研究区数据,预测研究区滑坡发生概率的空间分布情况,将两种分类后的易发性区划图按照相同的分类区间[0~0.35), [0.35~0.5), [0.5~0.7), [0.7~0.85)和[0.85~1)依次分为低、较

低、中、较高和高易发区,得到滑坡易发性分区图(图 5)。由图 5 可知,万州区内的大部分地区都位于低和较低易发区,滑坡高易发区集中于万州区主城和长江与较大的水系附近,这与当地滑坡的实际发育情况一致。

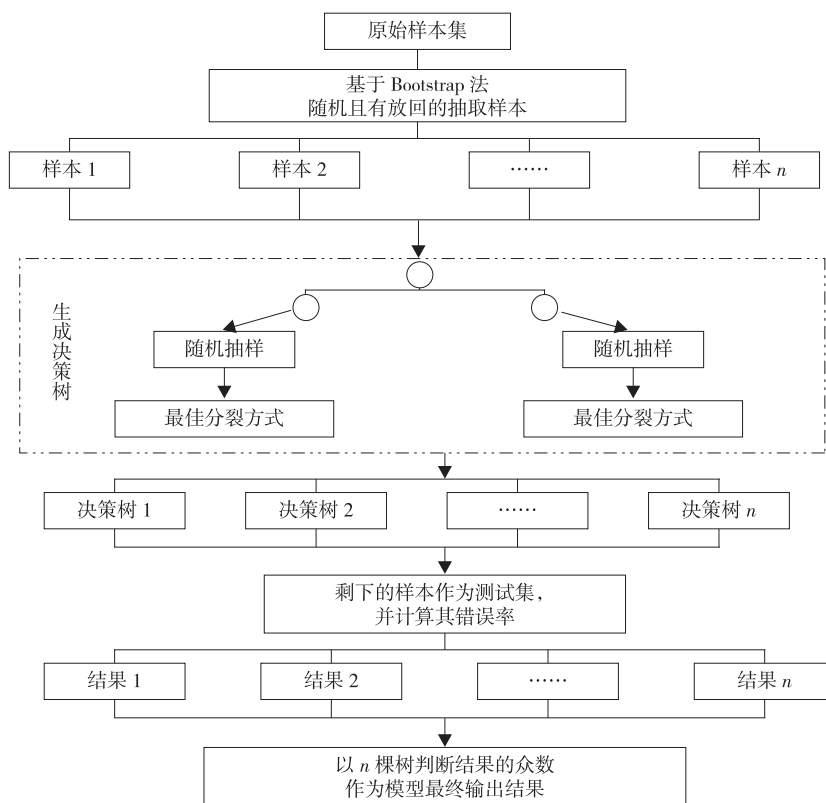


图 4 RF 流程图

Fig. 4 Flow chart of random forest

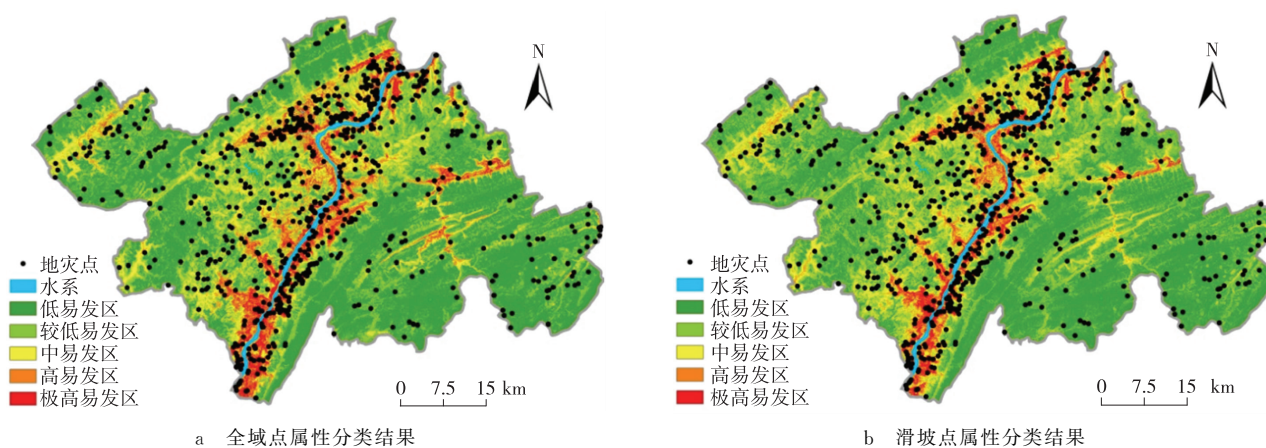


图 5 万州区滑坡易发性区划图

Fig. 5 Mapping of landslide susceptibility in Wanzhou district

由表 4 可知:以滑坡点属性作为分类基础数据得出的高-极高易发区仅占全区 11.45% 的面积,而滑坡数量占研究区历史滑坡总数的 74.65%;以全域点属性作为分类基础数据得出的高-极高易发区占全区 14.02% 的面积,滑坡数量占研究区历史滑坡总数的 56.79%;相比之下,前者得出的高-极高易发区面积更小,包含历史滑坡数据更多。因此,用以上两种因子分类方法得出的分类区间相同的滑坡易发性区划图,以滑坡点属性作为分类基础数据得出的易发性区划图具有更好的预测效果。

4.2 模型精度对比

应用 RF 模型对两种分类结果分别进行滑坡易发性评价(图 6),可以看出,使用滑坡点属性与研究区全域点

属性作为分类基础数据得到的 AUC 值分别为 0.85 和 0.79,两者 AUC 值均高于 0.5,且以滑坡点属性作为因子分类数据的 AUC 值更接近于 1,说明以该分类方法的模型具有良好的精确性和预测能力。

表 4 全域点属性分类的滑坡易发性分级统计
Tab. 4 Statistics of landslide susceptibility classification of all points

易发性等级	全域点属性分类结果					滑坡点属性分类结果				
	栅格数/ 个	面积占比/ %	滑坡数/ 个	滑坡数 占比/%	滑坡密度/ (个·km ⁻²)	栅格数/ 个	面积占比/ %	滑坡数/ 个	滑坡数 占比/%	滑坡密度/ (个·km ⁻²)
低易发区	1 723 398	44.68	31	4.01	0.02	1 616 647	41.92	9	1.16	0.01
较低易发区	963 376	24.98	90	11.64	0.10	1 126 176	29.20	12	1.55	0.01
中易发区	629 122	16.31	213	27.55	0.38	672 283	17.43	175	22.64	0.29
高易发区	356 392	9.24	208	26.91	0.65	286 194	7.42	272	35.19	1.06
极高易发区	184 494	4.78	231	29.88	1.39	155 482	4.03	305	39.46	2.18

RF 模型十折交叉验证的测试集精度结果表明用滑坡点属性与研究区全域点属性作为分类基础数据得到的测试样本最高精度分别为 0.779 与 0.755,选取最高精度样本搭建模型,以滑坡点属性为分类基础数据得到的模型准确度更高。

5 讨论与结论

5.1 基于滑坡点重分类的有效性

利用 ArcGIS 分级统计高程的两种分类结果中各类对应的滑坡点比例、区域所占面积比例与滑坡密度,如表 5 所示。从该表可以看出,在研究区全域点属性作为分类基础数据中,当高程大于 1 245 m 时,滑坡数量为 0,则该分类没有意义,从而造成分类冗余的情况。而且当滑坡过于集中于某一类时,也会导致细节缺失以及易发性评价精度降低的状况。

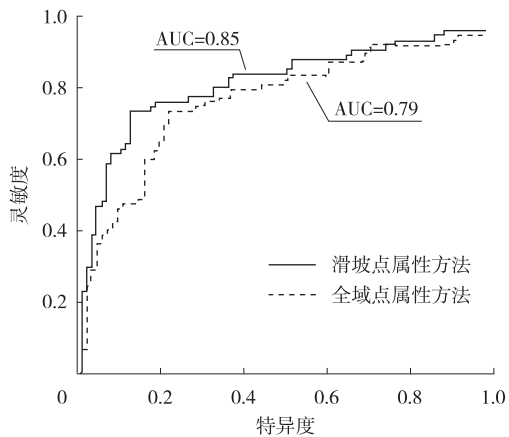


图 6 ROC 曲线
Fig. 6 ROC curve

表 5 高程分类结果对比
Tab. 5 Comparison of elevation classification results

全域点属性分类结果				滑坡点属性分类结果			
高程/m	滑坡点 比例/%	占总面积 比例/%	滑坡密度/ (个·km ⁻²)	高程/m	滑坡点 比例/%	占总面积 比例/%	滑坡密度/ (个·km ⁻²)
<230	0.19	0.05	0.816	<195	0.10	0.04	0.537
230~<345	0.22	0.10	0.473	195~<240	0.12	0.02	1.289
345~<445	0.20	0.13	0.330	240~<290	0.10	0.04	0.537
445~<535	0.12	0.15	0.172	290~<346	0.10	0.06	0.358
535~<620	0.09	0.15	0.129	346~<405	0.12	0.07	0.368
620~<715	0.08	0.13	0.132	405~<471	0.10	0.10	0.215
715~<820	0.04	0.09	0.095	471~<543	0.09	0.12	0.161
820~<940	0.03	0.08	0.081	543~<627	0.09	0.15	0.129
940~<1 080	0.01	0.06	0.036	627~<754	0.09	0.16	0.121
1 080~1 245	0.01	0.04	0.054	754~914	0.04	0.12	0.072
>1 245	0.00	0.02	0.000	>914	0.04	0.13	0.066

以滑坡点属性作为分类基础数据时则很好地避免了上述问题。使用滑坡点自身属性分类,每个分类中都必然存在历史滑坡点,且滑坡在各分类中分布较为均匀。各区域面积差异较大,能直观反映出各区域滑坡密度差异。以滑坡点属性对高程进行分类,结果显示在 $0 \sim < 290$ m 处滑坡密度较高,在 $195 \sim < 240$ m 处滑坡密度达到峰值(1.289 个 $\cdot \text{km}^{-2}$),这一最高值明显高于以全域点属性分类得出的结果,滑坡密度高低差异较为明显。并且全域点属性分类结果最后 5 个级别的滑坡密度均小于 0.01 个 $\cdot \text{km}^{-2}$,分类冗余现象较为严重(表 5)。所以,以滑坡点属性作为分类依据,不仅从主观上避免了分类冗余过多、细节缺失的缺点,而且在客观上能更加完整地表现出滑坡分布与发生的特点。以滑坡点属性对有关数据进行分类更为有效,且更加适合滑坡易发性区划的构建。

5.2 因子重要性排序

滑坡形成机理复杂,对影响因子进行重要性探测可为滑坡灾害风险管理提供参考。用 RF 模型的平均基尼指数减低数^[19]作为判断因子重要性大小的指标,取两种分级结果构建模型后得出指标的均值判断因子重要性,按因子对滑坡发生的影响程度由大到小对 13 个因子进行排序,依次为:高程、多年平均降雨量、切割深度、坡向、距道路距离、坡度、POI 核密度、倾坡、岩性、NDVI、曲率、起伏度、地形湿度指数(图 7)。

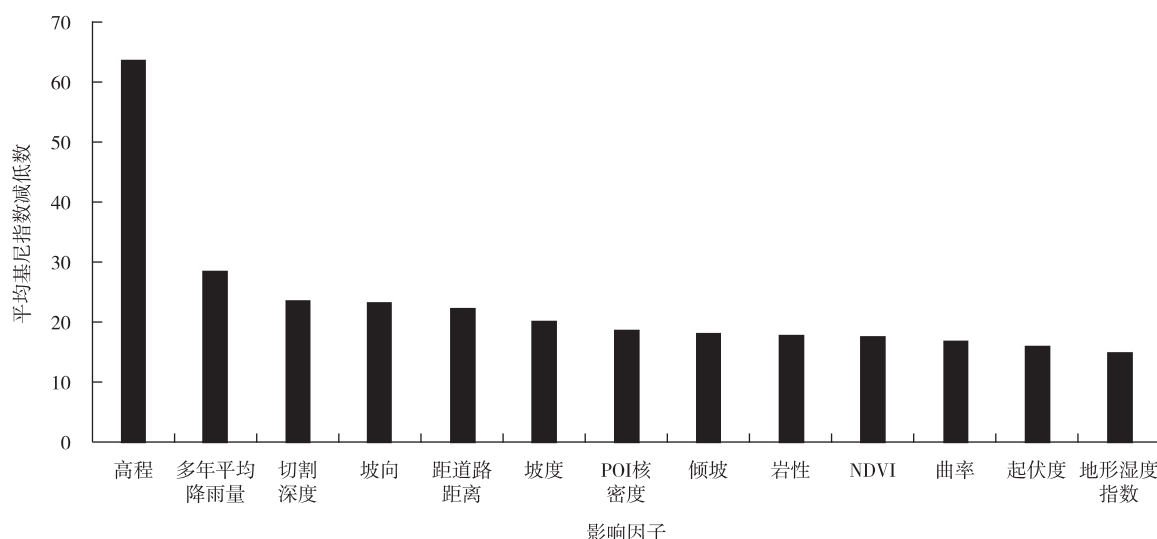


图 7 影响因子重要性排序

Fig. 7 Impact factor importance ranking

万州区境内多中小起伏山地、丘陵,地形复杂,其中:高程较低地区由于开挖坡脚、污水渗透、乱砍滥伐等人类活动较为广泛,使得坡体稳定性严重下降,滑坡的发生概率更高;高程较高地区人类活动现象较少、植被覆盖度高,水土固定能力较强,因此滑坡发生概率更低。已有较多研究表明,高程对滑坡发生的影响较大^[5,20]。降雨会冲刷坡体表面岩土体物质且渗漏的雨水会侵蚀坡体内部,降低坡体稳定性,与滑坡的形成关系紧密,而多年平均降雨量不仅影响岩土体本身的强度,且影响当地地表径流、地下水流量等因素的变化^[21]。地表切割深度对滑坡形成的重要性排在第 3 位,但在以往研究中较少将该因子引入到评价体系,然而三峡库区内沟谷切割深且水流侵蚀切割对地表土体的影响程度较大,故本研究引入该因子对相关研究有一定参考价值。

利用 ArcGIS 将高程、多年平均降雨量、地表切割深度按间隔分级统计万州区历史滑坡的分布情况,如图 8 所示。在本研究中,研究区高程范围为 $21 \sim 1\,720$ m,而当地滑坡集中分布在 $150 \sim 450$ m 的高程范围内,在高程超过 $1\,000$ m 后滑坡密度接近于 0。总体上看,随着高程的增加,滑坡数量与滑坡密度均呈现下降的趋势,高程与滑坡发生的相关性较强,对滑坡形成的影响较大。滑坡数量与滑坡密度均随多年平均降雨量的增加呈先增加后减少趋势,且集中分布于多年平均降雨量为 $1\,300 \sim 1\,380$ mm 之间,滑坡数量分布近似呈正态分布。此外,滑坡数量集中于地表切割深度为 $350 \sim 650$ m 之间,且滑坡密度在地表切割深度为 $350 \sim 650$ m, $1\,000 \sim 1\,050$ m 和 $1\,150 \sim 1\,300$ m 处较高,高低值差异明显,能较好地反映不同区间内该因子对滑坡形成的影响。

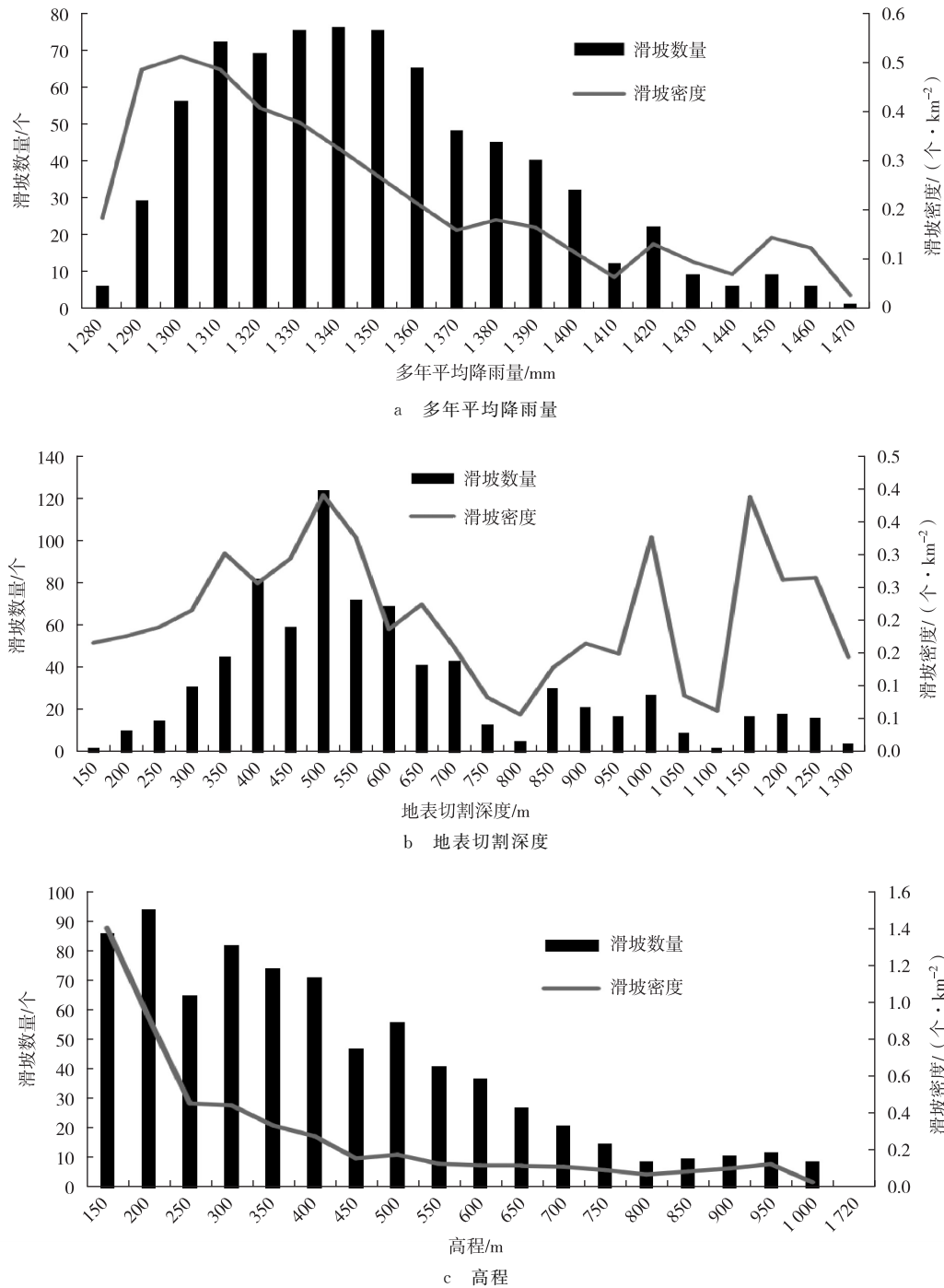


图 8 典型因子与滑坡之间的关系

Fig. 8 The relationship between typical factors and landslides

5.3 结论

本研究以位于三峡库区腹地的万州区为研究区,以历史滑坡点属性与研究区全域点属性作为分类基础数据,基于自然断点法进行分类,得到两种分类结果,并应用 RF 模型对两种分类结果分别进行滑坡易发性区划评价。结果表明:

- 1) 相较于使用研究区全域点属性确定分类标准,采用滑坡点属性确定连续因子分类标准构建的 RF 模型 AUC 值与测试集精度更高,模型的精确性和预测能力更佳。
- 2) 以滑坡点属性确定的分类标准进行因子分类,得到的滑坡易发性区划图高-极高易发区面积更小,包含历史滑坡数据更多,得出的易发性区划图效果更佳,对影响因子分级研究有较好的参考价值。
- 3) 高程、地表切割深度和多年平均降雨量的对滑坡形成的重要性位于前 3 位,以往研究中较少考虑地表切

割深度对于滑坡易发性的影响,引入该因子可定量表征区域内水流切割强烈程度与沟谷切割深度对于滑坡诱发的影响,为滑坡影响因子体系的构建提供了参考。

参考文献:

- [1] PETLEY D. Global patterns of loss of life from landslides[J]. *Geology*, 2012, 40(10): 927-930.
- [2] ILMY H F, DARMINTO M R, WIDODO A. Application of machine learning on Google Earth engine to produce landslide susceptibility mapping (case study: pacitan) [EB/OL]. [2021-05-05]. <https://iopscience.iop.org/article/10.1088/1755-1315/731/1/012028>.
- [3] 孙德亮. 基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D]. 上海: 华东师范大学, 2019.
- SUN D L. Mapping landslide susceptibility based on machine learning and forecast warning of landslide induced by rainfall[D]. Shanghai: East China Normal University, 2019.
- [4] LUO W, LIU C C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods[J]. *Landslides*, 2017, 15: 465-474.
- [5] 刘睿, 施妮娜, 孙德亮, 等. 基于 GIS 与随机森林的巫山县滑坡易发性区划[J]. 重庆师范大学学报(自然科学版), 2020, 37(3): 86-96.
- LIU R, SHI S X, SUN D L, et al. Mapping of landslide proneness in Wushan county based on GIS and random forest[J]. *Journal of Chongqing Normal University (Natural Science)*, 2020, 37(3): 86-96.
- [6] 闫举生, 谭建民. 基于不同因子分级法的滑坡易发性评价: 以湖北远安县为例[J]. 中国地质灾害与防治学报, 2019, 30(1): 52-60.
- YAN J S, TAN J M. Evaluation of landslide susceptibility based on different factor classification methods: taking Yuanan county, Hubei as an example[J]. *The Chinese Journal of Geological Hazard and Control*, 2019, 30(1): 52-60.
- [7] 赵魁. 基于 ArcGIS 的云安区地质灾害易发性分区评价[J]. 地质灾害与环境保护, 2020, 31(4): 38-42.
- ZHAO K. Geological hazard susceptibility mapping evaluation in Yun'an district based on ArcGIS[J]. *Journal of Geological Hazards and Environment Preservation*, 2020, 31(4): 38-42.
- [8] 赖冠中, 陈文音. 基于自然断点法分析的城乡建设用地整理潜力分区研究: 以汕头市濠江区为例[J]. 广西城镇建设, 2019(12): 123-127.
- LAI G Z, CHEN W Y. Research on the regionalization of urban and rural construction land consolidation potential based on natural breakpoint analysis: taking Haojiang district of Shantou city as an example[J]. *Cities and Towns Construction in Guangxi*, 2019(12): 123-127.
- [9] 周婧. 等间距法和均值-标准差法的遥感蚀变分级研究[C]//中国矿物岩石地球化学学会. 中国矿物岩石地球化学学会第九次全国会员代表大会暨第 16 届学术年会文集. 西安: 中国矿物岩石地球化学学会, 2017: 895-896.
- ZHOU J. Research on remote sensing alteration classification based on equal spacing method and mean-standard deviation method[C]//Chinese Society of Mineralogy, Petrology and Geochemistry. The Ninth National Congress of Chinese Society of Mineralogy, Petrology and Geochemistry and Proceedings of the 16th Annual Academic Conference. Xi'an: Chinese Society of Mineralogy, Petrology and Geochemistry, 2017: 895-896.
- [10] 凌晓, 刘甲美, 王涛, 等. 基于致灾因子对称法分级的信息量模型在地震滑坡危险性评价中的应用[J]. 国土资源遥感, 2021, 33(2): 172-181.
- LING X, LIU J M, WANG T, et al. Application of information value model based on symmetrical factors classification method in landslide hazard assessment[J]. *Remote Sensing for Land and Resources*, 2021, 33(2): 172-181.
- [11] 郭建华, 刘初群, 刘翠. 基于遗传算法优化的城市标准循环工况构建[J]. 科学技术与工程, 2017, 17(15): 327-333.
- GUO J H, LIU C Q, LIU C. Construction of urban standard cycle conditions based on genetic algorithm optimization[J]. *Science Technology and Engineering*, 2017, 17(15): 327-333.
- [12] 骆黎. 万州区地质灾害分析与对策研究[D]. 重庆: 西南大学, 2006.
- LUO L. Analysis and countermeasures of geological disasters in Wanzhou district[D]. Chongqing: Southwest University, 2006.
- [13] 许嘉慧, 孙德亮, 王月, 等. 基于 GIS 与改进层次分析法的奉节县滑坡易发性区划[J]. 重庆师范大学学报(自然科学版), 2020, 37(2): 36-44.
- XU J H, SUN D L, WANG Y, et al. Mapping of landslide proneness in Fengjie county based on GIS and improved analytic hierarchy process[J]. *Journal of Chongqing Normal University (Natural Science)*, 2020, 37(2): 36-44.
- [14] WANG Y, SUN D L, WEN H J, et al. Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang county (Chongqing, China) [J]. *International Journal of Environmental Research and Public*

- Health, 2020, 17(12): 1-39.
- [15] 文海家, 张永兴, 柳源. 三峡库区地质灾害防减灾战略关键问题[J]. 中国地质灾害与防治学报, 2004, 15(2): 25-31.
WEN H J, ZHANG Y X, LIU Y. Key issues of geological disaster prevention and mitigation strategy in the Three Gorges reservoir area[J]. The Chinese Journal of Geological Hazard and Control, 2004, 15(2): 25-31.
- [16] 王婷, 潘军, 蒋立军, 等. 基于 DEM 的地形因子分析与岩性分类[J]. 国土资源遥感, 2018, 30(2): 231-237.
WANG T, PAN J, JIANG L J, et al. Topographic factor analysis and lithology classification based on DEM[J]. Remote Sensing for Land & Resources, 2018, 30(2): 231-237.
- [17] 秦璇. 随机森林算法的改进及应用[D]. 兰州: 兰州交通大学, 2020.
QIN X. Improvement and application of random forest algorithm[D]. Lanzhou: Lanzhou Jiaotong University, 2020.
- [18] SUN D L, XU J H, WEN H J, et al. An optimized random forest model and its generalization ability in landslide susceptibility mapping: application in two areas of Three Gorges reservoir, China[J]. Journal of Earth Science, 2020, 31(6): 1068-1086.
- [19] DU G L, ZHANG Y S, IQBAL J, et al. Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu province, China[J]. Journal of Mountain Science, 2017, 14(2): 249-268.
- [20] 郑迎凯, 陈建国, 王成彬, 等. 确定性系数与随机森林模型在云南芒市滑坡易发性评价中的应用[J]. 地质科技通报, 2020, 39(6): 131-144.
ZHENG Y K, CHEN J G, WANG C B, et al. Application of certainty factor and random forests model in landslide susceptibility evaluation in Mangshi city, Yunnan province[J]. Bulletin of Geological Science and Technology, 2020, 39(6): 131-144.
- [21] 陈洪凯, 魏来, 谭玲. 降雨型滑坡经验性降雨阈值研究综述[J]. 重庆交通大学学报(自然科学版), 2012, 31(5): 990-996.
CHEN H K, WEI L, TAN L. Review of research on empirical rainfall threshold of rainfall-induced landslide[J]. Journal of Chongqing Jiaotong University (Natural Science), 2012, 31(5): 990-996.

Resources, Environment and Ecology in Three Gorges Area

Comparison of Landslide Susceptibility Mapping Based on Different Factor Classifications: Taking Wanzhou District as an Example

SUN Deliang¹, MA Xianglong¹, TANG Xiaoya², WEN Haijia³, MI Changlin⁴

(1. The Key Laboratory of GIS Application Research, Chongqing Normal University, Chongqing 401331;

2. Information Engineering College of Pharmacy, Guangdong Pharmaceutical University, Guangzhou 510006;

3. Key Laboratory of New Technology for Construction of Cities in Mountain Area, Ministry of Education, Chongqing 400045;

4. Linyi City Natural Resources Development Service Center, Shandong Province, Linyi Shandong 276000, China)

Abstract: [Purposes] Based on the attributes of the landslide point and the attributes of the entire study area as the basic data for classification, a comparative study of landslide sensitivity in Wanzhou district located in Three Gorges reservoir area is carried out, and a reference for the classification study of the influencing factors of landslide formation is provided. [Methods] Select 13 items including elevation, annual average rainfall, and others are used as landslide hazards as impact factors. The natural breaks classification used to classify the attributes of the landslide point and the whole area of the study area respectively, and the two classification results are used to construct a model to predict the spatial distribution of landslide susceptibility in the study area. [Findings] The area under curve (AUC) value corresponding to the classification data of the study area is 0.79, and the highest accuracy of the corresponding model test set is 0.755; the AUC value corresponding to the landslide point attribute as the classification data is 0.85, the highest accuracy of the corresponding model test set is 0.779; the importance of elevation, surface cutting depth, and annual average rainfall are in the top 3, which has a greater impact on the occurrence of landslides. [Conclusions] The random forest model established by the classification standard of landslide point attribute determination factors has good accuracy and predictive ability, and can provide reference for landslide disaster risk management and impact factor classification research in Wanzhou district.

Keywords: random forest model; factor classification; landslide susceptibility mapping; Wanzhou district; Three Gorges reservoir area

(责任编辑 方 兴)