

DOI:10.16030/j.cnki.issn.1000-3665.2018.02.22

基于子流域特征的泥石流易发性评价

张书豪, 吴光, 张乔, 高飞

(西南交通大学地球科学与环境工程学院, 四川 成都 611756)

摘要: 文章以云南省迪庆藏族自治州金沙江河谷金江镇—林当可为例, 研究了基于子流域单元的泥石流易发性评价模型。首先, 将研究区划分为 1 570 个子流域单元, 其中 70% (1 090 个) 用于模型建立, 30% (480 个) 作为模型测试样本。然后, 采用 9 种子流域单元特征因子 (melton 比率、流域平均径流侵蚀力指数、平均地形湿度指数、流域延伸率、流域平均高程、流域水系密度、流域平均植被覆盖度、流域平均坡度、流域高差率) 的归一化值作为自变量, 泥石流是否发育作为因变量, 运用逻辑回归建立泥石流易发性评价模型, 将易发性划分为 5 个等级。最后, 通过独立的测试样本验证, 易发性评价模型预测性能 ($AUC = 0.821$) 良好; 且随着易发性等级提高, 泥石流绝对和相对发育比例逐级增加, 模型满足合理性要求, 适用于研究区的泥石流易发性评价。

关键词: 泥石流; 易发性评价; 逻辑回归模型; 子流域特征; 金沙江河谷

中图分类号: P642.23

文献标识码: A

文章编号: 1000-3665(2018)02-0142-08

Debris-flow susceptibility assessment using the characteristic factors of a catchment

ZHANG Shuhao, WU Guang, ZHANG Qiao, GAO Fei

(Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan 611756, China)

Abstract: In traditional application of GIS, the susceptibility mapping based on distinct grids usually neglect the organic unity of the natural element. For this reason, this article investigates the susceptibility model based on catchments, using the valleys of the Jinsha River between Jinjiang and Lindangke as a working example. Firstly, the study area is divided into 1 570 catchments, in which 70% (1 090 catchments) are used to build the susceptibility model and 30% (480 catchments) are used as validation samples. Secondly, 9 feature factors, including the melton ratio, average stream power index of catchment, average topographic wetness index of catchment, catchment elongation, average height of catchment, drainage density of catchment, average fractional vegetation cover of catchment, average slope of catchment and catchment relief ratio are normalized and then selected as independent variables. Whether debris-flows occur is the dependent variable. Susceptibility assessment models of debris-flow are then built using the logistic regression and are classified into five degrees. Finally, the model can predict well ($AUC = 0.821$) through the test of independent validation samples. Moreover, both the absolute and relative incidence of debris-flows increase with the susceptibility degree.

Keywords: debris-flow; susceptibility assessment; logistic regression model; characteristic factors of catchments; valley of the Jinsha River

收稿日期: 2017-06-10; 修订日期: 2017-08-21

基金项目: 中国铁路总公司科技开发计划 (结合工程的科研项目) (总公司合同编号: 2010G004 - I; 分合同编号: 滇西指 LX2016 - 1 号 - 1)

第一作者: 张书豪 (1990-), 男, 博士研究生, 主要从事斜坡地质灾害风险评价和稳定性分析。E-mail: zshgostop@126.com

通讯作者: 吴光 (1958-), 男, 博士, 教授, 博士生导师, 研究方向为工程地质学。E-mail: 962444613@qq.com

泥石流易发性分区是指一定环境下泥石流的空间发生概率,主要判断空间中哪些位置容易发生泥石流,而不考虑发生的具体规模和时间^[1]。泥石流易发性分区主要的评价单元类型有栅格单元、地貌单元、行政区单元、均一条件单元。

国内外许多学者依据不同评价单元和特征因子对泥石流展开易发性研究。Wilford 等^[2]在不列颠哥伦比亚地区研究了以 melton 比率和流域长度来划分泥石流与普通洪流流域的方法。Welsh 等^[3]在新西兰山区采用 melton 比率判断泥石流是否发育。Zhou 等^[4]研究了汶川地震灾区的泥石流发育特征,分别建立了各个流域的流域高差、流域面积、流域面积高差比与流域中地震引起的松散物质面积之间的关系,以松散物质面积大小作为该流域是否发育泥石流的判定阈值。Stock 等^[5]研究了世界各地的河谷的侵蚀规律,建立了河道汇水面积和坡度比的双对数关系曲线,并统计了曲线上从泥石流转为普通河流的坡度比范围。黎艳等^[6]以栅格为基本评价单元进行泥石流易发性评价,其结果对相关地区的基础设施规划建设有着重要借鉴意义。

根据上述文献以及其他有关研究,目前泥石流易发性评价多是基于栅格单元,采用如岩性、坡向、坡度、高程等单一离散的栅格单元进行叠加分析。这种评价方式得到的结果不仅忽略了自然单元之间的有机整体性,而且常与真实的环境不符。而流域单元根据一定的地貌特征划分,能反映真实自然环境的演变,以及地貌与地质之间有差别的边界,因此与泥石流的真实发育环境特征更为相符^[7]。另一方面,运用特定流域特征指标建立的经验模型不仅依赖于丰富的专家知识和深入的分析调查,而且对环境特征的变化十分敏感。而机器学习技术基于实际样本,无需人为主观赋权,可以使用高维特征因子快速建立与研究地区相适应的预测模型。

本文以云南迪庆藏族自治州内与丽江市交界的金沙江河谷金江镇—林当可为例,基于子流域评价单元,运用9种不同的流域特征因子,结合GIS环境与机器学习技术(逻辑回归)建立泥石流空间预测模型,并用独立的检验样本集定量评价模型的预测效果。

1 研究区概况及数据来源

1.1 研究区概况

研究区金沙江金江镇—林当可河谷段,通过石鼓大拐弯、虎跳峡大峡谷,位于云南省迪庆藏族自治

州境与丽江市交界(图1)。地形险峻、岭谷相对高差大。研究区为亚热带季风气候,年平均气温在12℃左右,年平均降雨量约为600~800mm,其中5~10月降雨量占全年的80%,干湿季节分明,垂直分带明显。

研究区内外地质营力作用强烈,气候干热昼夜温差大,峡谷重力地质现象和岩层风化剥蚀过程强烈,滑坡、崩塌、错落、卸荷、冰川堆积物等不良地质现象极其发育,为泥石流的形成提供了丰富的物质来源。

1.2 数据来源

本文研究所需的主要基础数据汇总于表1。

表1 数据来源和数据类型
Table 1 Sources and types of data

基础数据	数据来源与制作	数据格式
DEM	地理空间数据云网站(www.gscloud.cn)	30 m × 30 m 栅格数据
水系数据	DEM 提取, Google Earth 卫星影像数字化	矢量
Landsat8 数据	来源美国地质调查局(landsatlook.usgs.gov),用于计算植被覆盖度。	多波段栅格数据
泥石流灾害统计	根据《丽江地质灾害防治规划(2011—2020年)》、《迪庆藏族自治州香格里拉县地质灾害防治规划(2011—2020年)》、卫星影像判释	矢量

2 研究方法

2.1 流域单元划分

本文利用 ArcGis 的 ArcHydro 水文分析插件,依次按照填充洼地—计算流向—计算流量—定义河网—生成子流域栅格划分研究区的子流域。定义河网的最小阈值是划分子流域单元的关键,Lyons 等^[8]认为泥石流能够在汇水面积0.05 km²左右的河道形成,同时考虑到本文研究区的大小和关心的结果尺度,以汇水面积1 km²作为最小阈值生成河网,再生成子流域栅格。接着对部分与卫星影像不符合的流域边界进行适当修正后,得到1570个子流域单元矢量,平均流域面积为2.39 km²,最小流域面积为0.2 km²,最大流域面积为9.32 km²。研究区泥石流子流域分布如图1所示。

2.2 易发性评价因子和分级

本文选取9个评价因子,各因子分级图见图2。

2.2.1 melton 比率

melton 比率是反映一个流域地势的指标,如式(1),1965年由Melton^[9]提出,该值越大则整个流

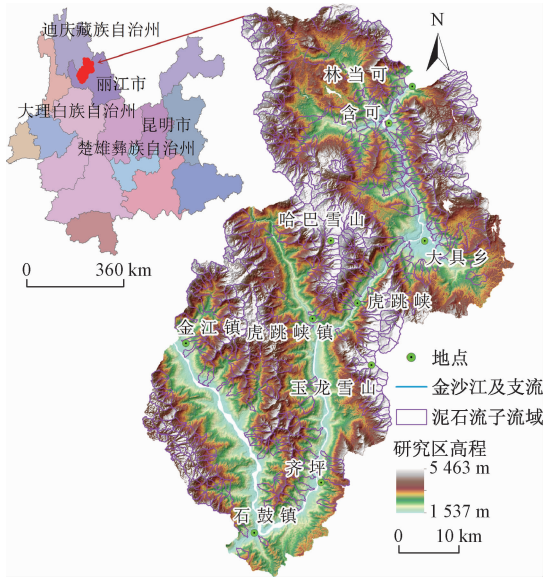


图 1 研究区泥石流子流域分布图

Fig. 1 Distribution of debris-flow catchments

域的地势越陡峻,反之则越平缓。泥石流在运动过程中需要足够的能量携带物质,该指标能体现物质在流域中运动势能大小,从而影响搬运物质的能力。本文将研究区的 $melton$ 比率分为 5 个等级(图 2a)。

$$R_m = \frac{dH}{A} \quad (1)$$

式中: A ——流域面积/ m^2 ;

dH ——流域高差/ m 。

2.2.2 流域平均径流侵蚀力指数

径流侵蚀力指数 SPI 由 Moore 等^[10]提出,反映地表某一位置受到表面径流侵蚀力的强弱,公式如下:

$$SPI = \ln\left(\frac{A_c}{L_c} \tan\beta\right) \quad (2)$$

式中: A_c ——待计算栅格单元上游的汇水面积/ m^2 ;

$\tan\beta$ ——待计算栅格坡度正切值;

L_c ——栅格单元宽度/ m 。

地表 SPI 值较高的部位在地表水作用下受到侵蚀作用较强,反之则较弱,因此能在一定程度上指示水土流失的强弱。本文计算各子流域的平均 SPI 值作为评价因子,并将其分成 4 个等级(图 2b)。

2.2.3 流域平均地形湿度指数

地形湿度指数 TWI 可用来反映地表容易积水的区域,计算公式如下:

$$TWI = \ln\left(\frac{A_c}{L_c \tan\beta}\right) \quad (3)$$

式中: A_c ——待计算栅格单元上游的汇水面积/ m^2 ;

$\tan\beta$ ——待计算位置坡度正切值;

L_c ——栅格单元宽度/ m 。

其他条件一定,坡度增大会导致 TWI 值降低,而上游汇水面积增大则会引起 TWI 值升高,地表 TWI 值越高的部位越容易积水,泥石流的形成需要汇集足够的水源。本文计算各子流域单元的平均 TWI 值,并分成 4 个等级(图 2c)。

2.2.4 流域延伸率

流域延伸率 BE 是将流域面积相等的圆直径与流域长轴作比值,是由 Schumm 等^[11]在 1956 年提出。越接近于 1 则流域形状越接近于圆形,反之,流域形状则越长且窄,其计算公式如下:

$$BE = \frac{2\sqrt{A}}{L/\pi} \quad (4)$$

式中: A ——流域面积/ m^2 ;

L ——流域长轴长度/ m 。

其他条件一定时,接近圆形的流域通常比形状长而窄的流域在出水口的峰值流量更大,因为支流能够在更短的时间差内汇集并到达出水口。本文将研究区子流域延伸率分成 4 个等级(图 2d)。

2.2.5 流域平均植被覆盖度

植被覆盖度 (VFC) 是指植被的叶、枝、茎等的垂直投影面积占所统计区内面积总和的百分比。 VFC 越高,植被覆盖程度越高,反之则越低。 VFC 能够指示一段时间内,流域的水土保持能力。考虑到研究区的植被情况, VFC 的最大值和最小值可以分别近似取 1 和 0 来对计算公式进行简化如下:

$$VFC = \frac{(NDVI - NDVI_{min})}{(NDVI_{max} - NDVI_{min})} \quad (5)$$

式中: $NDVI$ ——植被覆盖指数。

首先计算各栅格 VFC ,再计算各子流域单元的平均 VFC ,并将其分为 5 个等级(图 2e)。

2.2.6 流域平均高程

流域平均高程是子流域单元内所有高程栅格的平均值。研究区属于第二阶梯向第三阶梯的过渡带河谷地区,岩土应力和差异风化垂直分带明显,因此平均高程会影响松散物质的储量。本文将子流域平均高程分成 7 个等级(图 2f)。

2.2.7 流域高差率

流域高差率 R_r 是流域高差与流域长轴长度之比值,计算公式如下:

$$R_r = \frac{dH}{L} \quad (6)$$

式中: dH ——流域高差/m;

L ——流域长轴长度/m。

流域高差率能够简单反映流域地势,该值越大,则地势越陡,反之则越平缓。本文将子流域高差率分成5个等级(图2g)。

2.2.8 流域水系密度

子流域水系密度 D_r 是流域内一定流量阈值生成的水系面积与流域面积之比。本文以100个栅格单元汇水面积为阈值生成水系来计算流域水系密度,流域水系密度指示了流域内受到一定规模水流冲刷的区域

所占的比例,本文将研究区子流域水系密度分为4个等级(图2h),计算公式如下:

$$D_r = \frac{A_g N_{100}}{A} \quad (7)$$

式中: A_g ——栅格面积/ m^2 ;

N_{100} ——子流域内水系的栅格总数/个;

A ——子流域面积/ m^2 。

2.2.9 流域平均坡度

以DEM计算得到的坡度栅格为基础,求取每个子流域内坡度的平均值。流域平均坡度主要体现了流域内各个位置地势的平均水平,本文将子流域平均坡度分为5个等级(图2i)。

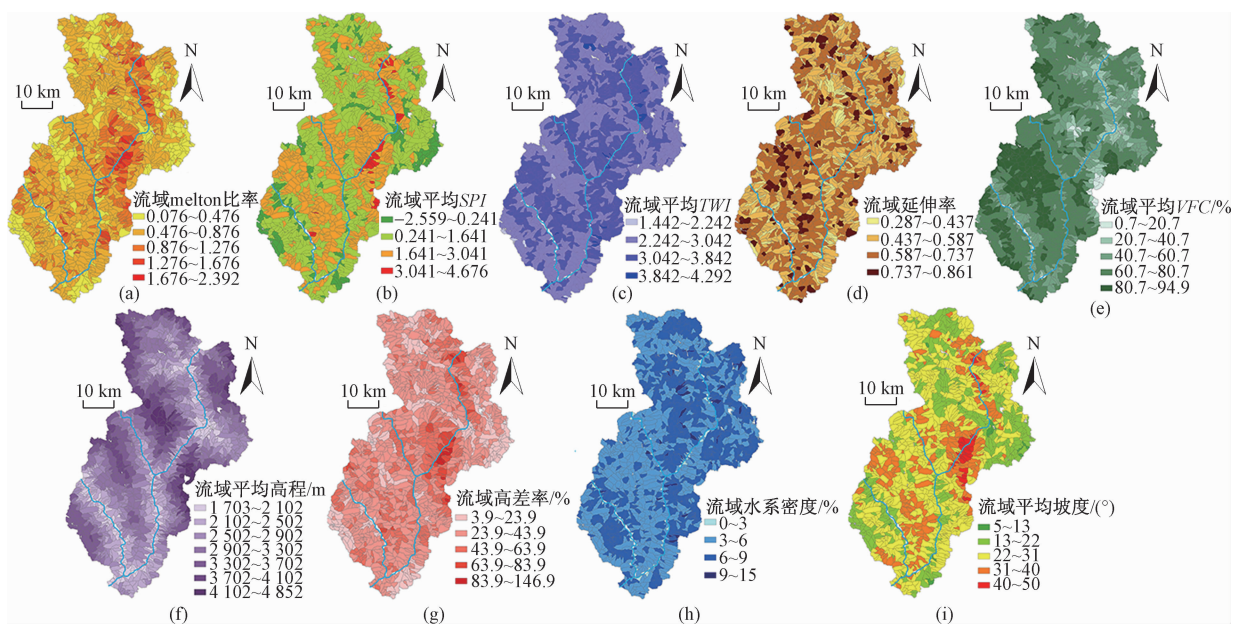


图2 子流域特征因子分级图

Fig. 2 Classification maps showing the characteristic factors of catchments

2.3 子流域特征因子归一化

在模型评价中,各个因子的量纲常常不统一,取值范围也不一致^[12]。逻辑回归本质是线性模型,每个因子内取不同值却对应相同的权重是不合理的。为了解决上述缺点,可以用子流域特征因子归一化后的值进行易发性评价。本文用各因子不同分级下的子流域中发育泥石流的个数比例进行归一化,计算公式如下:

$$d_{ij} = \frac{M_{ij}}{N_{ij}} \quad (8)$$

$$I_{ij} = \frac{d_{ij}}{\sum d_{ij}} \quad (9)$$

式中: d_{ij} ——第 i 个因子下的第 j 个分级内的子流域中发生泥石流的的比例;

M_{ij} ——第 i 个因子下的第 j 个分级中的泥石流子流域个数;

N_{ij} ——第 i 个因子下的第 j 个分级中的子流域总个数;

I_{ij} ——第 i 个因子下的第 j 个分级的归一化值。

通过运用式(8)和(9),并按照2.2节中各子流域特征因子的分级方式进行归一化处理,将计算结果汇总于表2。

2.4 评价模型

逻辑回归模型能揭示一个因变量与多个互不相关的自变量之间的多元回归关系^[13],解决二分类和多分类问题,其通过 sigmoid 函数将多元线性回归方程转化为一个结果输出为0~1之间的逻辑回归函数,具体表达如下式:

表 2 各因子分级泥石流发生率和归一化值

Table 2 Normalized values and debris-flow incidence in each classification of factors

特征因子	分类级别	泥石流子 流域数/个	类别 总数/个	发生 率 d_{ij}	归一 化值 I_{ij}	特征 因子	分类 级别	泥石流子 流域个/个	类别 总数/个	发生 率 d_{ij}	归一 化值 I_{ij}
melton 比率	0.076 ~ 0.476	57	273	0.2088	0.1226	流域平均 高程	1702 ~ 2102 m	24	214	0.1121	0.0373
	0.476 ~ 0.876	265	817	0.3244	0.1904		2102 ~ 2502 m	54	290	0.1862	0.0620
	0.876 ~ 1.276	139	348	0.3994	0.2345		2502 ~ 2902 m	73	334	0.2186	0.0727
	1.276 ~ 1.676	50	96	0.5208	0.3057		2902 ~ 3302 m	95	299	0.3177	0.1057
	1.676 ~ 2.392	9	36	0.2500	0.1468		3302 ~ 3702 m	128	246	0.5203	0.1732
							3702 ~ 4102 m	108	145	0.7448	0.2479
流域平 均径流	-2.559 ~ 0.241	20	184	0.1087	0.0829	流域高 差率	4102 ~ 4852 m	38	42	0.9048	0.3011
	0.241 ~ 1.641	213	717	0.2971	0.2265		3.9% ~ 23.9%	39	238	0.1639	0.0916
侵蚀力 指数	1.641 ~ 3.041	261	615	0.4244	0.3235		23.9% ~ 43.9%	260	814	0.3194	0.1786
	3.041 ~ 4.676	26	54	0.4815	0.3671		43.9% ~ 63.9%	176	428	0.4112	0.2299
							63.9% ~ 83.9%	37	66	0.5606	0.3135
流域平 均地形 湿度 指数	1.442 ~ 2.242	2	24	0.0833	0.0854		83.9% ~ 146.9%	8	24	0.3333	0.1864
	2.242 ~ 3.042	217	745	0.2913	0.2986	流域水系 密度	0% ~ 3%	6	64	0.0938	0.0889
流域平 均地形 湿度 指数	3.042 ~ 3.842	294	769	0.3823	0.3918		3% ~ 6%	238	730	0.3260	0.3091
	3.842 ~ 4.292	7	32	0.2188	0.2242		6% ~ 9%	265	736	0.3601	0.3413
							9% ~ 15%	11	40	0.2750	0.2607
流域延 伸率	0.287 ~ 0.437	34	101	0.3366	0.2478	流域平均 坡度	5° ~ 13°	3	39	0.0769	0.0490
	0.437 ~ 0.587	209	645	0.3240	0.2386		13° ~ 22°	91	481	0.1892	0.1204
流域延 伸率	0.587 ~ 0.737	230	696	0.3305	0.2433		22° ~ 31°	284	771	0.3684	0.2344
	0.737 ~ 0.861	47	128	0.3672	0.2703		31° ~ 40°	124	235	0.5277	0.3358
							40° ~ 50°	18	44	0.4091	0.2604
流域平 均植被 覆盖度	0.7% ~ 20.7%	11	14	0.7857	0.3531						
	20.7% ~ 40.7%	57	126	0.4524	0.2033						
流域平 均植被 覆盖度	40.7% ~ 60.7%	119	382	0.3115	0.1400						
	60.7% ~ 80.7%	225	765	0.2941	0.1322						
	80.7% ~ 94.9%	108	283	0.3816	0.1715						

$$P(y = 1 | x; \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (10)$$

式中: P ——属于泥石流子流域的概率;

x ——各子流域特征因子归一化值;

β ——各子流域特征因子归一化值系数。

$$C = \begin{cases} 1 & P(y = 1 | x; \beta) > t \\ 0 & P(y = 1 | x; \beta) < t \end{cases} \quad (11)$$

式中: C ——分类结果(泥石流或非泥石流);

t ——判别阈值。

本研究中,子流域是否发生泥石流是一个二分类问题,把各个特征因子归一化值输入到多元线性方程中,通过 sigmoid 函数进行转化,并输出结果。将泥石流子流域看作分类 1,则输出结果就是该子流域发育泥石流的概率;通过确定一个判别阈值,将大于该阈值的结果归类为 1(泥石流),而小于该阈值的结果归类为 0(非泥石流)。

3 子流域单元泥石流易发性评价

3.1 样本划分

为了检验模型的预测能力,本文将样本集 70% 划分为训练集,即 1 090 个子流域用于建立模型;30% 的样本划分为验证集,即 480 个子流域,其中含泥石流子

流域 164 个,占比 34.17%,用于验证模型。

3.2 多重共线性诊断

多重共线性指在多元回归中两个或多个变量之间高度线性相关。逻辑回归变量之间的关系采用的是多元线性方程表达,如果各个变量之间存在多重共线性,会导致方程中各个回归系数的估计不准确,因此可能导致很大的预测误差^[14]。本文用 SPSS 17.0 进行变量之间的多重共线性诊断(表 3)。表中各变量容差均大于 0.2,且 VIF (方差膨胀因子)不大于 5。说明这些变量之间没有明显的多重共线性关系,因此均可用于建立预测模型。

3.3 易发性评价模型建立

利用 SPSS17.0 采用训练集中的样本建立逻辑回归模型,以表 2 计算得到的各个子流域特征分级的归一化值 I_{ij} 作为逻辑回归模型的输入变量 x ,以是否是泥石流子流域为输出变量(1 表示泥石流子流域,0 表示非泥石流子流域),得到式(12):

$$\begin{cases} Z = -7.613 + 2.416x_1 + 2.622x_2 + \\ 1.375x_3 - 3.415x_4 + 9.640x_5 + \\ 14.425x_6 + 1.904x_7 + 4.441x_8 + 5.698x_9 \\ P(y = 1 | Z) = \frac{1}{1 + e^{-Z}} \end{cases} \quad (12)$$

其中, $P(y=1|Z)$ 是属于泥石流流子流域的概率, x_1 到 x_9 依次是 melton 比率、流域平均径流侵蚀力指数、流域平均地形湿度指数、流域延伸率、流域平均植被覆盖度、流域平均高程、流域高差率、流域水系密度、流域平均坡度等因子求得的归一化值 I_{ij} 。

表3 多重共线性诊断与因子重要性

Table 3 Collinearity diagnostics and importance of factors

模型	共线性统计		变量重要性
	容差	VIF	
流域平均植被覆盖度归一化值	0.919	1.089	0.07
流域平均径流侵蚀力指数归一化值	0.332	3.013	0.09
流域 melton 比率归一化值	0.480	2.085	0.06
流域平均地形湿度指数归一化值	0.720	1.389	0.04
流域水系密度归一化值	0.885	1.131	0.08
流域平均高程归一化值	0.905	1.105	0.41
流域高差率归一化值	0.354	2.828	0.06
流域延伸率归一化值	0.945	1.058	0.03
流域平均坡度归一化值	0.438	2.282	0.15

4 结果分析与精度评价

4.1 因子重要性分析与泥石流空间易发性分布

因子变量的重要性取决于模型的算法。根据逻辑回归模型输出的变量重要性来看(表3),流域平均高程、平均坡度、平均径流侵蚀力指数、水系密度是9个指标中对泥石流发育影响最显著的,而流域延伸率则是重要性最低的因子。

流域平均高程归一化值的重要性最高(0.41),泥石流发育频率随流域平均高程增加而增加(表2)。由于研究区为第二阶梯向第三阶梯的过渡带的河谷地区,相对高差很大,岩土应力和差异风化垂直带明显;在这些地区,高程越高,岸坡的卸荷作用越充分且向时效变形转化^[15],因此斜坡破坏频发,加上强烈的物理风化作用,高海拔地区有充足的松散物质来源。

流域平均坡度归一化值的变量重要性是0.15,泥石流发育频率在32°~40°之间最大,达到了0.3358,而在小于22°的区间,泥石流的发育比例很低。

流域平均径流侵蚀力指数归一化值的变量重要性是0.09,泥石流发育比例随流域平均径流侵蚀力指数的增加而增加(表2),可见它们之间存在正相关关系。

流域水系密度归一化值的变量重要性是0.08,流域水系密度小于3%的区间泥石流发育频率只有0.09,其他区间的发育频率比较接近,说明该指标也能较好地区分泥石流流域。

melton 比率归一化值重要性是0.06,在区间

1.2764~1.6764 泥石流发育频率最大(0.52),该值与泥石流发育频率没有明显的正负相关关系,也没有遵循部分学者的经验模型判别准则。流域延伸率归一化值重要性只有0.03,各个分级区间的发育频率均在0.25左右(表2),说明这个指标对于本研究区易发性的区分度很差。

图3的泥石流易发性分区图与真实地形边界符合,每个子流域单元对应一个易发性等级,与栅格单元制作的易发性分区图相比,结果更加直观易判释。虎跳峡段是研究区内泥石流沟最为发育的区段,其位于哈巴雪山和玉龙雪山之间。大具一林当可段泥石流也较为发育,尽管其左侧岸坡易发性较低,但其上游有较多的泥石流沟,所以下游也易受到危害。石鼓镇附近区域,石鼓镇至虎跳峡镇左侧段,泥石流沟都不太发育。

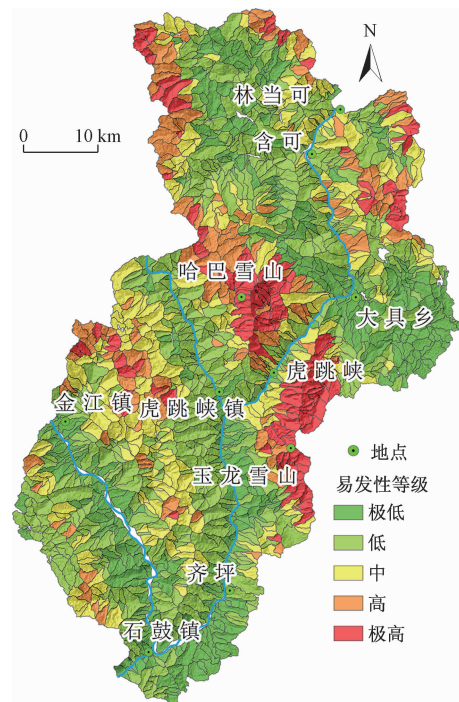


图3 子流域泥石流易发性分级

Fig. 3 Debris-flow susceptibility classification of catchments

4.2 易发性结果预测效果评价

本文分别采用两种方式对泥石流易发性模型预测效果进行评价:一是检验测试集中泥石流流子流域落入各个易发性分区的比例是否合理;二是通过机器学习中常用的 ROC 曲线来对模型的预测效果进行综合评价。由训练集建立的逻辑模型得到的子流域发育泥石流的概率 P 值的范围在 0.009~0.993,将区间划分为5个等级,检验测试集落入各个区间的比例(表4)。

表 4 子流域泥石流易发性检验表

Table 4 Check table of debris-flow susceptibility of catchments

易发等级	易发区间	泥石流流子流域数/个	子流域总数/个	泥石流发育比例	泥石流相对发育比例
极高	0.759 ~ 1	39	45	86.67%	9.24
高	0.559 ~ 0.759	31	49	63.27%	6.20
中	0.359 ~ 0.559	45	91	49.45%	2.61
低	0.159 ~ 0.359	40	173	23.12%	0.64
极低	0.009 ~ 0.159	9	122	7.38%	0.29
合计		164	480		

注:区间泥石流发育比例=区间泥石流流子流域个数/区间子流域总个数,区间泥石流相对发育比例=区间泥石流发育比例/区间子流域个数占比。

合理的易发性分区应该满足:随着易发性等级的提高,泥石流在各个易发性分区内的发育比例与相对发育比例应该逐渐增大。检验结果显示,易发性等级由极低到极高,各分级泥石流发育比例由 7.38% 逐渐升高至 86.67%,同时各分级泥石流相对发育比例也由 0.29 逐渐升高至 9.24,结果满足前述合理性。

ROC 曲线(图 4)是二分类学习中常用的评价分类器表现的曲线,其评价更为全面、客观,不会受到正负样本比例的影响。图 4 纵轴代表正确预测出的泥石流流子流域占有泥石流流子流域个数的比例,横轴代表正确预测出的非泥石流流子流域占有非泥石流流子流域个数的比例。由于 ROC 曲线是逐渐降低泥石流易发性阈值(大于该阈值判断为泥石流,小于该阈值判断为非泥石流)而得到曲线上各个点对应的坐标;如果曲线越靠近左上角,则越多的泥石流集中于一个等级高且范围小的易发性区间内,曲线下的面积(AUC)就越大,显然这样的模型能很好地将泥石流与非泥石流区分开来,因此预测效果就越好。图 4 是随机预测模型与本文逻辑回归模型的 ROC 曲线对比。本文预测模型 AUC 为 0.821,通常易发性评价中 AUC 在 0.7 以上准确性较好^[16]。结合前述的易发性结果合理性检验,本文模型能够客观准确地对研究区内的子流域进行泥石流易发性评价。

5 结论

(1)基于某几种流域特征因子建立的经验模型适用性常常有限。而采用高维流域特征因子,结合可靠的机器学习技术,能够快速建立适用于特定地区的易发性评价模型。

(2)以子流域评价单元得到的泥石流易发性分区制图,相较以栅格单元划分的制图更符合真实地貌单

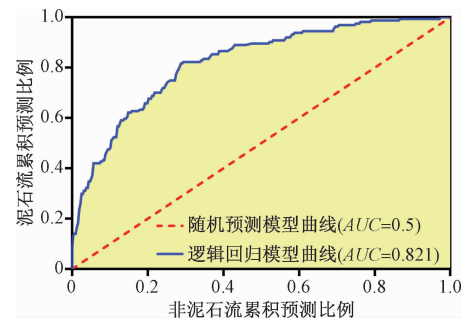


图 4 易发性评价模型 ROC 曲线

Fig. 4 ROC curve of the susceptibility assessment model

元边界,且结果直观易于解释。本研究区内流域平均高程、流域平均坡度、流域平均径流侵蚀力指数、流域水系是 9 个流域特征因子中对泥石流活动影响最显著的因子;研究区内的泥石流高易发区基本沿着 3 000 m 以上高海拔山区分布。

(3)本文通过独立测试集对泥石流易发性评价模型进行检验后,结果显示:随着易发性等级的提高,泥石流在各个易发性分区内的绝对发育比例和相对发育比例均逐渐增大,满足合理性;模型的 ROC 曲线获得的 AUC 值为 0.821,预测精度良好。

参考文献:

- [1] Guzzetti F, Reichenbach P, Cardinali M, *et al.* Probabilistic landslide hazard assessment at the basin scale [J]. *Geomorphology*, 2005, 72 (1/4): 272 - 299.
- [2] Wilford D J, Sakals M E, Innes J L, *et al.* Recognition of debris flow, debris flood and flood hazard through watershed morphometrics [J]. *Landslides*, 2004, 1(1): 61 - 66.
- [3] Welsh A, Davies T. Identification of alluvial fans susceptible to debris-flow hazards [J]. *Landslides*, 2010, 8(2): 183 - 194.
- [4] Zhou W, Tang C, Van Asch T W J, *et al.* A rapid method to identify the potential of debris flow development induced by rainfall in the catchments of the Wenchuan earthquake area [J]. *Landslides*, 2015, 13(5): 1243 - 1259.
- [5] Stock J, Dietrich W E. Valley incision by debris flows: Evidence of a topographic signature [J]. *Water Resources Research*, 2003, 39(39): ESG 1 - 1.
- [6] 黎艳, 陈剑, 许冲, 等. 基于 AHP 的半干旱区泥石流易发性评价: 以金沙江上游奔子栏-昌波河段为例 [J]. *现代地质*, 2015, (4): 975 - 982. [LI Y,

- CHEN J, XU C, et al. Susceptibility assessment of debris flows in the semiarid region based on AHP method: An example from the Benzilan-Changbo segment of the upper Jinsha River [J]. *Journal of Mountain Science*, 2015, (4): 975 – 982. (in Chinese)]
- [7] 吴树仁. 滑坡风险评估理论与技术[M]. 北京: 科学出版社, 2012. [WU S R. Theory and technology of landslide risk assessment [M]. Beijing: Science Press, 2012. (in Chinese)]
- [8] Lyons N J, Mitsova H, Wegmann K W. Improving mass-wasting inventories by incorporating debris flow topographic signatures [J]. *Landslides*, 2013, 11 (3): 385 – 397.
- [9] Melton M A. The Geomorphic and Paleoclimatic Significance of Alluvial Deposits in Southern Arizona: A Reply [J]. *The Journal of Geology*, 1965, 73 (1): 102 – 106.
- [10] Moore I D, Grayson R, Ladson A. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications [J]. *Hydrological processes*, 1991, 5 (1): 3 – 30.
- [11] Schumm S A. Evolution of Drainage Systems and Slopes in Badlands at Perth Amboy, New Jersey [J]. *Geological Society of America Bulletin*, 1956, 67 (5): 597 – 646.
- [12] 王佳佳. 三峡库区万州区滑坡灾害风险评估研究 [D]. 武汉: 中国地质大学, 2015. [WANG J J. Landslide risk assessment in Wangzhou County, Three Gorges Reservoir [D]. Wuhan: China University of Geosciences. 2015. (in Chinese)]
- [13] Ohlmacher G C, Davis J C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA [J]. *Engineering Geology*, 2003, 69 (3 – 4): 331 – 343.
- [14] Chatterjee S, Hadi A S. Regression Analysis by Example [M]. 5th ed. London: John Wiley & Sons Inc, 2012.
- [15] 黄润秋. 中国西南岩石高边坡的主要特征及其演化 [J]. *地球科学进展*, 2005, 20 (3): 292 – 297. [HUANG R Q. Main characteristics of high rock slopes in southwestern china and their dynamic evolution [J]. *Advances in Earth Science*, 2005, 20 (3): 292 – 297. (in Chinese)]
- [16] Youssef A M, Pourghasemi H R, Pourtaghi Z S, et al. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia [J]. *Landslides*, 2015, 13 (5): 839 – 856.

责任编辑: 汪美华

(上接第 135 页)

- [10] 马世国. 强降雨条件下基于 Green-Ampt 入渗模型的无限边坡稳定性研究 [D]. 杭州: 浙江大学, 2014. [MA S G. Study on the stability of infinite slope based on Green-Ampt infiltration model under intense rainfall [D]. Hangzhou: Zhejiang University, 2014. (in Chinese)]
- [11] Chen L, Young M H. Green-Ampt infiltration model for sloping surfaces [J]. *Water Resources Research*, 2016, 42 (7): 887 – 896.
- [12] Ng C W W, Wang B, Tung Y K. Three-dimensional numerical investigations of groundwater responses in an unsaturated slope subjected to various rainfall patterns [J]. *Can Geotech J*, 2011, 38 (5): 1049 – 1062.
- [13] 罗渝, 何思明, 何尽川. 降雨类型对浅层滑坡稳定性的影响 [J]. *地球科学 (中国地质大学学报)*, 2014, 34 (9): 1357 – 1363. [LUO Y, HE S M, HE J C. Effect of rainfall patterns on stability of shallow landslide [J]. *Earth Science (Journal of China University of Geosciences)*, 2014, 34 (9): 1357 – 1363. (in Chinese)]
- [14] 刘果果. 川东红层地区降雨诱发浅层土质滑坡试验及数值研究 [D]. 成都: 成都理工大学, 2016. [LIU G G. Rainfall experiments and numerical analysis of the shallow landslide in the red-bed area in East Sichuan [D] Chengdu: Chengdu University of Technology, 2016. (in Chinese)]

责任编辑: 汪美华