逻辑回归和信息量模型在山区滑坡易发性评价中适宜性的分析

邹凤钗[1],冷洋洋[1,2],陶小郎[1],尹彦迪[1]

(1. 贵州省地质环境监测院,贵州　贵阳　550081;2. 成都理工大学,四川　成都　610059)

[摘　要]以万山区为例,在区域滑坡孕灾条件的基础上,筛选工程地质岩组、斜坡结构、平均坡度、地貌、距构造距离及距河流距离共 6 个易发条件因子,选取逻辑回归模型和信息量模型对山区滑坡进行易发性评价。结果显示逻辑回归模型中中高易发区面积占比分别为 15.78%和 19.70%,82%的地质灾害点落在该区域内;信息量模型中中高易发区面积占比为 12.41%、25.19%,包含了区域 88%的滑坡灾害点。最后通过实际发生的灾害点在各易发区的分布情况进行检验,逻辑回归模型中灾害点落在高易发区的比例远小于信息量模型,且高易发等级中灾害点实际发生的比值较小,说明针对山区区域滑坡地质灾害易发性评价结果预测上,信息量模型的评价结果更为客观准确。

[关键词]滑坡地质灾害;易发性评价;逻辑回归模型;信息量模型;万山区

[中图分类号]P694;P642.22　[文献标识码]A　[文章编号]1000-5943(2022)03-280-07

1　引言

地质灾害易发性评价是对区域地质灾害发生的可能性进行预测,对于区域灾害防治具有现实意义。目前遥感(remote sensing,RS) 和地理信息系统(GIS) 技术与模型相结合的评价方法已广泛应用于灾害评价中。其中赵建粮、Anbalaga 等运用 GIS 分析功能开展地质灾害区划研究(赵建粮 等,2007;Anbalagan,1992);刘艺梁等运用逻辑回归和人工神经网络 2 种模型进行三峡滑坡对比定量分析(刘艺梁 等,2010);魏国灵等运用综合危险性指数法对陆河县的地质灾害易发性进行区划(魏国灵等,2020);贺鹏等利用层次分析法构建评价模型分析西藏札达地区的滑坡灾害易发性(贺鹏 等,2016);田春山等基于 CF 模型和逻辑回归模型对广东省地质灾害易发性进行了评价(田春山 等,2016);高克昌等基于 GIS 和信息量模型对重庆万州的滑坡易发性进行评价(高克昌 等,2006)。上述评价模型各有特色,取得了大量的研究成果。贵州省作为山区省份,地层出露较齐全、工程岩组及地貌类型组合复杂多样,需选择一种适用、易实现的方法进行区域地质灾害易发性评价。

万山区工程地质岩组齐全,基本涵盖了省内致灾工程地质岩组,工程岩组与地貌类型组合多样,有较高的分析研究价值。且区域地质灾害多发区地质环境特征与我省地质灾害发育分布特征有一定相似性,因此研究万山区地质灾害易发性评价方法对于我省防灾减灾具有一定的示范作用。本文结合万山区地质灾害详细调查成果(贵州省地质环境监测院,2021),选取逻辑回归模型和信息量模型研究区域滑坡地质灾害易发性,并对比分析,得出

[收稿日期]2022-02-03　[修回日期]2022-07-06

[基金项目]贵州省地质灾害详细调查及风险评价成果集成(中央自然灾害防治体系建设补助资金 Z195110010003);2021 年度中共贵州省委全面深化改革重大调研课题(贵州山区地质灾害成因机理及防治对策研究)。

[作者简介]邹凤钗(1985—),女,硕士,高级工程师,主要从事矿山地质环境、地质灾害评价工作。E-mail:694554086@qq.com.

[通讯作者]冷洋洋(1987—),男,博士生,高级工程师,主要从事地质灾害调查及评价工作。E-mail:87065420@qq.com.

山区滑坡地质灾害易发性两种评价方法适用程度,有效推动区域地质灾害防控工作。

2　研究区背景

万山区位于贵州省东缘,介于东经 109°11′～109°14′、北纬 27°30′～27°32′之间,面积 842 km^2。地处云贵高原向湘西丘陵过渡的斜坡地带,区内总体地势东部低、西部高、中部隆起,地形切割强烈,区内地貌类型主要有中山、低山侵蚀地貌、中低山、低山溶蚀地貌 4 类。区域属长江流域沅江水系之一级支流锦江及舞阳河汇水范围,锦江、舞阳河分水岭位于万山区南部,境内河流主要靠源区降水补给,雨源特征明显,具有河道狭窄,河床较陡,洪枯变化剧烈的特点。区内出露地层按岩石的工程力学强度不同,可分为坚硬岩类工程地质岩组、软质岩夹坚硬岩类工程地质岩组、软质岩类工程地质岩组、松散岩类工程地质岩组。硬质岩类包括地层单元有:清水江组(Pt_3^1dq),寒武系清虚洞组($Є_2q$)、乌训组($Є_2w$)、寒武系武陵统-芙蓉统($Є_{3-4}$)、奥陶系桐梓组(O_1t)及红花园组(O_1h)地层,其岩性主要为灰岩、白云岩,清水江组主要为变余粉砂岩。软质岩夹坚硬岩类包括地层单元有:南华系富禄组(Pt_3^2bf)、大唐坡组(Pt_3^2bd)、南沱组(Pt_3^2n),震旦系陡山沱组(Pt_3^3d)、灯影组($Pt_3^3Є_1dy$)、老堡组($Pt_3^3Є_1l$),其岩性主要为粘土岩夹灰岩白云岩、粉砂岩等。软质岩类包括地层单元有:寒武系牛蹄塘组($Є_{1-2}n$)、九门冲组($Є_2jm$)、变马冲组($Є_2b$)、杷榔组($Є_2p$),奥陶系大湾组($O_{1-2}d$)。其岩性主要为粘土岩。松散岩类:由第四系(Q)粘土、粉质黏土、砾石层等组成。其结构松散,稳定性差。万山区地质灾害及隐患较严重,其中历史地质灾害共造成 11 人死亡、4 人失踪,损毁房屋 1150 间,直接经济损失 621 万元。现阶段共发育地质灾害 68 处,以滑坡地质灾害为主,滑坡共 47 处,主要集中在万山区东部清水江组(Pt_3^1q)浅变质岩地区以及西部杷榔组粘土岩地区,以小型浅层土质或强风化层滑坡为主。

3　研究方法

3.1　逻辑回归模型

逻辑回归模型主要是在一个因变量和多个自变量之间形成多元回归关系,从而预测任何一块区域某一事件的发生概率。其优势在于进行统计分析时,自变量可以是连续的,也可以是离散的,也没有必要满足正态分布,其性质拟合于地质灾害易发性分析的过程(覃乙根 等,2020)。其步骤如下:

首先确定区域崩塌滑坡易发性评价的因子,利用数学方法进行线性回归分析得到逻辑回归模型的回归系数:

$$\log(P_i/(1-P_i)) = b_0+b_1X_1+b_2X_2+b_3X_3+\cdots+b_pX_p \tag{1}$$

式中,P_i 为评价因子在不同分级条件下发生崩塌滑坡的概率,b_0 为常数项,b_1-b_p 为逻辑回归模型的回归系数,X_1-X_p 为自变量,模型左侧为事件发生几率的自然对数值。

3.2　信息量模型

地质灾害的形成受多种因素影响,信息量模型反映了一定地质环境下最易致灾因素及其细分区间的组合;具体是通过特定评价单元内某种因素作用下地质灾害发生频率与调查区地质灾害发生频率相比较实现的(常亚婷 等,2021;刘蕴坤 等,2020;鲁霞 等,2020;王雷 等,2020;赵帅 等,2019)。

对应某种因素特定状态下的地质灾害信息量为:

$$I_{Aj\to B} = \ln\frac{N_j/N}{S_j/S}(j=1,2,3,\cdots n) \tag{2}$$

式中:$I_{Aj\to B}$—对应因素 A、j 状态下地质灾害 B 发生的信息量;N_j—对应因素 A、j 状态下地质灾害分布的单元数;N —调查区已知有地质灾害分布的单元总数;S_j—因素 A、j 状态分布的单元数;S 为调查区单元总数。

$I_{Aj\to B}$值有正有负,$I_{Aj\to B}$值越大反映了对应因素 A、j 状态下地质灾害发生的可能性较大。

由于每个评价单元受众多因素的综合影响,各因素又存在若干状态,各状态因素组合条件下地质灾害发生的总信息量为:

$$I=\sum_{i=1}^{n}\ln\frac{N_i/N}{S_i/S} \tag{3}$$

式中:I—对应特定单元地质灾害发生的总信息量,指示地质灾害发生的可能性,可作为地质灾害易发性指数。

4　滑坡地质灾害易发性评价

4.1　数据来源

万山区滑坡地质灾害点的基础数据主要来源于 2021 年开展的贵州省万山区地质灾害详细调查及风险评价野外调查成果。基于数据中万山区孕灾地质环境条件与各地质灾害发育之间关系，本文选用 25×25 m 的栅格单元作为模型的计算单元，基于 gis 栅格数据，将研究区划分为 948572 个栅格，比例尺为 1∶5万。

滑坡地质灾害易发性评价影响因子的基础数据有万山区地理国情普查框架要素数据 1∶10000 数字高程模型数据和 2020 年遥感卫星影响数据（影像分辨率优于 1 米）；1∶5万地质地形图。数据来源于贵州省测绘资料档案馆。

4.2　因子选取

选用 IBM SPSS Statistics 25 对滑坡地质灾害进行二元逻辑回归分析。本文共选取了 10 个因子进行初步筛查，包括距河流距离、斜坡结构、构造、海拔高度、平均坡度、曲率、坡型、坡高、地貌、工程地质岩组。经过单因素筛查，排除掉其中影响不显著的因子。最后选取了工程地质岩组、斜坡结构、距河流距离、距构造距离、平均坡度、地貌 6 个对区域滑坡地质灾害影响显著的因子，其中距构造距离主要是指距断层及褶皱的距离。

4.3　逻辑回归模型易发性评价

通过各孕灾因子的逻辑回归分析，各因子对区域滑坡地质灾害的影响程度见表 1。其中工程地质岩组对滑坡地质灾害的形成有较显著影响：软质岩组、清水江组计算比数比（OR）分别是参考变量（硬质岩类）的 6.498 倍和 3.415 倍，说明该岩组分布区易发生滑坡；在距离河流 50～10 m 及 100～200 m 内，滑坡地质灾害易发，数据较为客观的反映出距河流距离与滑坡地质灾害的关系，距离河流距离越近其搬运、堆积及侵蚀作用越强，滑坡地质灾害越易发；距构造距离 0～100 m 滑坡较易发；斜坡结构从逆向→斜逆向→横向→斜顺向→顺向变化时，每提升一级，计算比数比（OR）是上一级的 1.174 倍，反映出区域顺向坡、斜顺向坡滑坡地质灾害相对易发的特征；平均坡度分级中，区域坡度在 25°～38°区间内滑坡较易发；地貌分类中，侵蚀地貌较多发滑坡。

通过 spss 计算各个栅格的概率，运用 gis 根据自然断点法将易发性等级结果划分为四级（图 1），其中极低易发区面积 439.41 km^2，占总面积的 52.19%；低易发区面积 103.91 km^2，占总面积的 12.34%；中易发区及高易发区面积分别为132.86 km^2 和 165.82 km^2，占比分别为 15.78%及 19.70%。

表 1　滑坡易发性各孕灾因子二元逻辑回归分析

Table 1　Binary logistic regression analysis of landslide susceptibility factors

滑坡孕灾因子	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
河流距离（参考变量距离河流大于 200 m）			317.180	3	.000	
河流距离 0～50 m	.673	.131	26.497	1	.000	1.961
河流距离 50～100 m	1.298	.111	136.617	1	.000	3.663
河流距离 100～200 m	1.100	.077	204.803	1	.000	3.005
构造距离（参考变量距离构造大于 1000m）			302.180	4	.000	
构造距离 0～100 m	1.673	.143	56.337	1	.000	3.321
构造距离 100～200 m	1.166	.130	121.533	1	.000	2.115
构造距离 200～500 m	1.039	.081	217.861	1	.000	1.131
构造距离 500～1 000 m	0.887	.053	233.024	1	.000	0.964
斜坡结构（逆向坡、斜逆向坡、横向坡、斜顺向坡、顺向坡梯度值）	.161	.017	86.466	1	.000	1.174
平均坡度分级（参考变量坡度大于 50°）			118.777	5	.000	
坡度<20°	.279	.204	1.867	1	.172	1.322
坡度 20°～25°	.958	.169	32.145	1	.000	2.605
坡度 25°～30°	1.037	.175	35.319	1	.000	2.822
坡度 30°～38°	1.163	.169	47.264	1	.000	3.199
坡度 38°～50°	.643	.174	13.637	1	.000	1.901

续表

滑坡孕灾因子	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
地貌分类(参考变量 构造侵蚀、溶蚀地貌)			643.593	2	.000	
溶蚀地貌	-2.359	.165	203.738	1	.000	.095
侵蚀地貌	1.550	.095	266.425	1	.000	4.713
工程地质岩组(参考变量硬质岩组)			964.133	4	.000	
硬质岩夹软质岩组	.646	.103	39.093	1	.000	1.908
软质岩组	1.872	.090	427.788	1	.000	6.498
松散岩组	1.073	.503	4.552	1	.033	2.925
清水江组(易滑地层)	.347	.094	13.698	1	.000	3.415

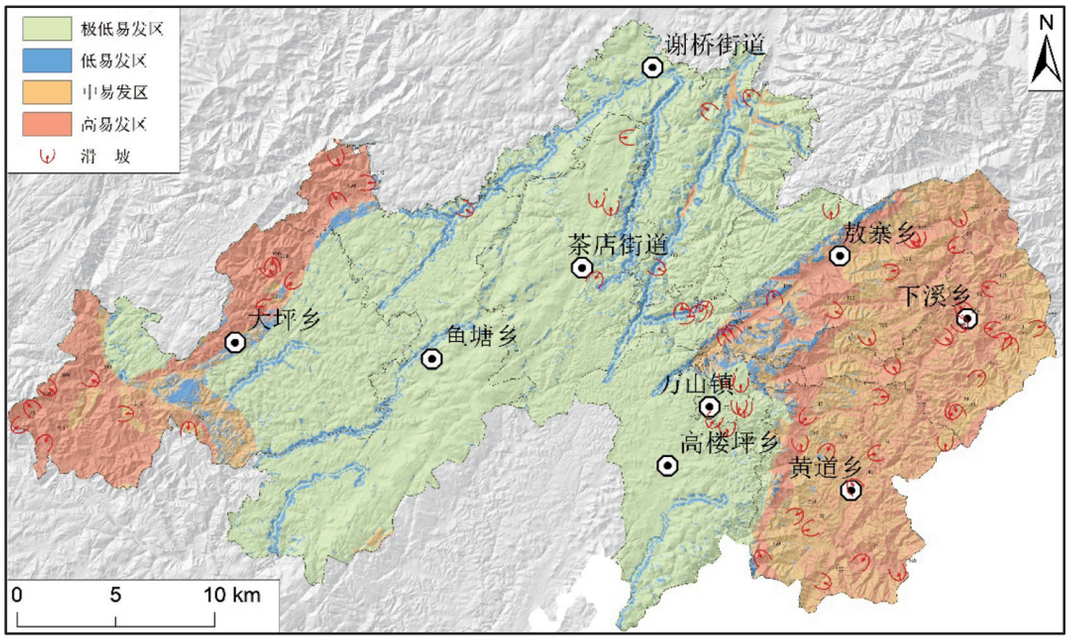

图 1　二元逻辑回归模型滑坡易发性分区图

Fig. 1　Landslide susceptibility partition map in binary logistic regression model

4.4　信息量模型易发性评价

运用 GIS 栅格计算器工具将斜坡结构、工程地质岩组、平均坡度、地貌、距构造距离及距河流距离 6 个不同的因子图层叠加处理,得到总的信息量值(表 2),即区域滑坡地质灾害易发性等级值,结合万山区地质环境条件,根据自然断点法将易发性等级评价结果分为四级:极低易发区(-8.44～-5.67)、低易发区(-5.66～-2.15)、中易发区(-2.14～1.07)、高易发区(1.07～4.70),得到区域滑坡易发性等级评价图(图 2)。其中极低易发区及低易发区面积分别为347.66 km^2 和 177.83 km^2,分别占总面积的 41.29% 及 21.12%;中易发区面积 104.46 km^2,占总面积的 12.41%;高易发区面积为 212.08 km^2,占比为 25.19%。

表 2　滑坡易发性等级信息量计算表

Table 2　Calculation table of landslide susceptibility level information

序号	因子	分级	灾害点密度	栅格数量	栅格比例	信息量
1	工程地质岩组	硬质岩类	0.067	628 632	0.663	-2.286
		软质岩夹硬质岩类	0.097	61 175	0.064	0.403
		软质岩类	0.585	111 127	0.117	1.608
		松散岩类	0.002	1 139	0.001	0.469
		清水江组	0.249	146 499	0.154	0.480

续表

序号	因子	分级	灾害点密度	栅格数量	栅格比例	信息量
2	斜坡结构	顺向	0. 318	195 977	0. 207	0. 432
		顺斜向	0. 143	169 245	0. 178	−0. 223
		横向	0. 279	306 317	0. 323	−0. 145
		逆斜向	0. 142	145 671	0. 154	−0. 080
		逆向	0. 118	131 362	0. 138	−0. 162
3	地貌	低山溶蚀地貌	0. 007	403 019	0. 425	−4. 147
		低山侵蚀地貌	0. 817	311 936	0. 329	0. 910
		中低山溶蚀地貌	0. 036	180 468	0. 190	−1. 668
		中山侵蚀地貌	0. 140	53 149	0. 056	0. 918
4	平均坡度	小于 20°	0. 176	369 065	0. 389	−0. 793
		20°～25°	0. 151	134 867	0. 142	0. 058
		25°～30°	0. 195	127 678	0. 135	0. 372
		30°～38°	0. 275	156 143	0. 165	0. 515
		38°～50°	0. 190	129 966	0. 137	0. 326
		50°～70°	0. 013	307 56	0. 032	−0. 910
		>70°	0. 000	97	0. 000	0. 629
5	距河流距离	0～50 m	0. 056	32 087	0. 034	0. 508
		50～100 m	0. 078	29 760	0. 031	0. 910
		100～200 m	0. 135	55 980	0. 059	0. 830
		>200 m	0. 731	830 745	0. 876	−0. 181
6	距构造距离	主要构造 0～100 m	0. 686	471 685	0. 497	0. 322
		主要构造 100～200 m	0. 144	229 012	0. 241	−0. 520
		主要构造 200～500 m	0. 074	140 535	0. 148	−0. 693
		主要构造 500～1 000 m	0. 062	67 849	0. 072	−0. 146
		主要构造>1 000 m	0. 035	39 491	0. 042	−0. 181

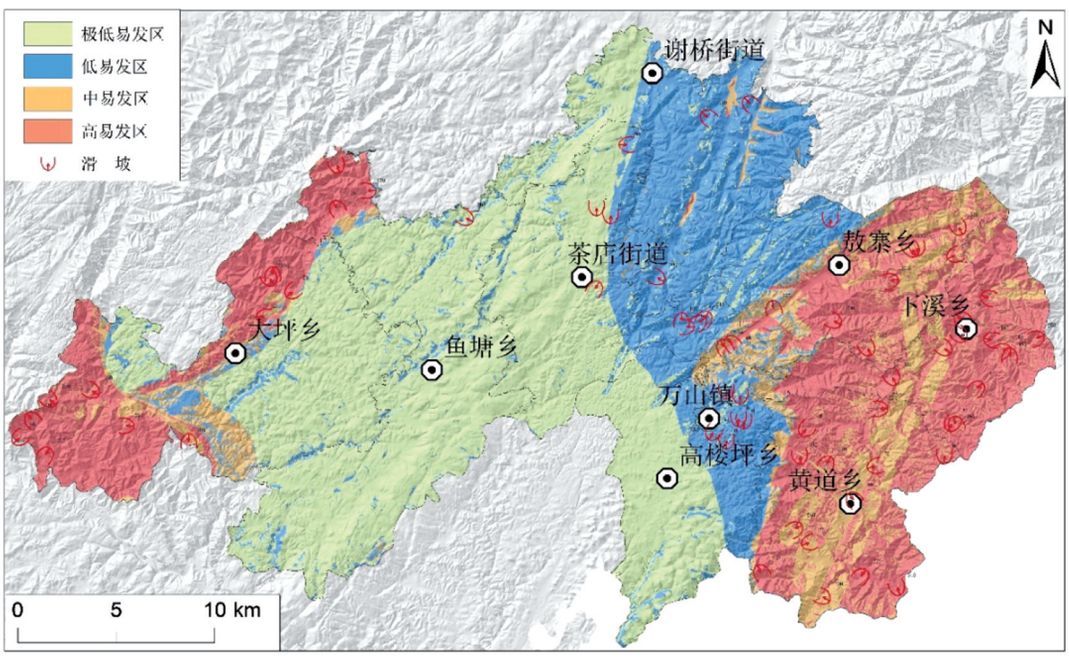

图 2　信息量模型滑坡易发性分区图

Table 2　Landslide susceptibility partition map in information model

5　评价结果分析

基于二元逻辑回归及信息量模型计算得出,万山区滑坡地质灾害中高易发区主要集中在区域东、西部,特别是在斜坡结构为顺向坡、距河流距离 50～200 m、距构造距离 0～100 m、平均坡度在 25°～38°、侵蚀地貌以及工程地质岩组为软质岩、清水江组的分布区是最容易发生滑坡地质灾害的区域。由图 1、图 2 可知,两种模型主要区别是区域东部的高易发区面积差异,信息量模型东部主要以高易发为主,更符合区域的实际工程地质条件。

本文分别对评价结果的合理性和精确性进行检验。对于模型合理性的检验主要通过实际发生的灾害点在各易发区的分布状况(田春山 等,2016),基于以下 2 个检验准则:

$R_i = B_i / A_i (i = Ⅰ, Ⅱ, Ⅲ, Ⅳ)$

(1)灾害点落在高易发区的百分比应该最大。

(2)各易发区灾害点的数量百分比(B_i)和各易发区的面积占整个研究区的总面积的百分比(A_i)的比值(R_i)应该是增大的。

式中:B_i—落在易发区 i 中的滑坡点占区域滑坡总数的百分比;A_i—i 等级易发区的面积占整个研究区面积的百分比。

对合理的地质灾害程度分区 R_i 应该满足 $R_Ⅰ < R_Ⅱ < R_Ⅲ < R_Ⅳ$。检验结果见图 3-4。

通过两种模型计算滑坡易发程度检验结果可知,信息量模型中已知滑坡灾害点落在高易发区的百分比最大,且极低易发区在整个研究区域的面积占比最大。同时,各易发等级中滑坡实际发生的比值 Ri 随之增大。说明这种划分方法得出的易发性等级与实际的滑坡发生情况相吻合,划分结果较理想。且该种模型划分的高易发及中易发中包含了 88%的已知滑坡灾害点。二元逻辑回归模型中亦满足文中验证条件,有 82%的已知滑坡灾害点落在中高易发等级中。但已知滑坡灾害点落在高易发区的比例远小于信息量模型,且高易发等级中滑坡实际发生的比值 Ri 较小,说明针对山区滑坡地质灾害易发性评价,信息量模型的预测结果更为准确。

6　结论

(1)以万山区为例,结合逻辑回归模型和信息量模型对山区滑坡地质灾害易发性进行评价,结果显示工程地质岩组、与河流距离及地形地貌特征三类评价因子对滑坡地质灾害易发程度贡献较大。特别是工程地质岩组为软质岩组、清水江组,与河流距离 50～100 m 且地形坡度在 25°～38°时极易发滑坡地质灾害。

(2)从各易发等级分布状况来看,采用逻辑回归模型评价的中高易发区面积占比分别为 15.78%和 19.70%,82%的地质灾害点落在该区域内;信息量模型评价的中高易发区面积占比为 12.41%、25.19%,包含了区域 88%的滑坡灾害点。

(3)根据实际发生的灾害点在各易发区的分布状况分别对两种模型进行对比分析,逻辑回归模型中灾害点落在高易发区的比例远小于信息量模型,且高易发等级中灾害点实际发生的比值较小,说明针对山区区域滑坡地质灾害易发性评价结果预测上,信息量模型的评价结果更为客观准确。

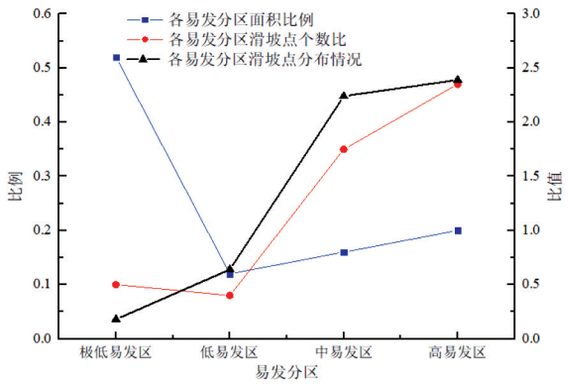

图 3　二元逻辑回归模型评价结果检验图

Fig. 3　Evaluation results in binary logistic regression model

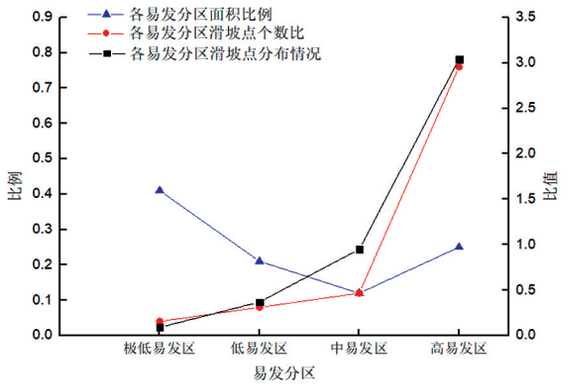

图 4　信息量模型评价结果检验图

Fig. 4　Evaluation results in information model

[参考文献]

常亚婷,刘征宇,向刚,等. 2021. 信息量模型在区域地质灾害危险性评价中的应用[J]. 贵州地质,38(4),430-436.

高克昌,崔鹏,赵纯勇,等. 2006. 基于地理信息系统和信息量模型的滑坡易发性评价——以重庆万州为例[J]. 岩石力学与工程学报,25(5):991-996.

贵州省地质环境监测院. 2021. 贵州省万山区地质灾害详细调查及风险评价报告[R]. 贵州:贵州省地质环境监测院.

贺鹏,童立强,郭兆成,等. 2016. GIS 支持下基于层次分析法的西藏札达地区滑坡灾害易发性评价研究[J]. 科学技术与工程,16(25):193-200.

刘艺梁,殷坤龙,刘斌. 2010. 逻辑回归和人工神经网络模型在滑坡灾害空间预测中的应用[J]. 水文地质工程地质,37(5):92-96.

鲁霞,兰安军,母浩江,等. 2020. 基于信息量模型的盘州市地质灾害易发性评价[J]. 科学技术与工程,20(14):5544-5551.

刘蕴琨,巨能攀,刘秀伟,等. 2020. 贵州省崩滑灾害影响因子易发区间评价[J]. 灾害学,35(3):201-207.

覃乙根,杨根兰,江兴元,等. 2020. 基于确定性系数模型与逻辑回归模型耦合的地质灾害易发性评价——以贵州省开阳县为例[J]. 科学技术与工程,20(1):96 -103.

田春山,刘希林,汪佳. 2016. 基于 CF 和 Logistic 回归模型的广东省地质灾害易发性评价[J]. 水文地质工程地质,43(6):154-170.

魏国灵,金云龙,邱锦安,等. 2020. 粤东陆河县地质灾害易发性评价[J]. 中国地质灾害与防治学报,31(1):52-56.

王雷,吴君平,赵冰雪,等. 2020. 基于 GIS 和信息量模型的安徽池州地质灾害易发性评价[J]. 中国地质灾害与防治学报,31(3):96-102.

赵帅,赵洲. 2019. 基于信息量模型的地质灾害易发性评价[J]. 水力发电,45(3):27-31.

赵建粮,王军见,张晋. 2007. GIS 技术在灵宝市地质灾害易发性分区评价中的应用[J]. 科学技术与工程,7(11):2735-2740.

Anbalagan R. 1992. Landslide hazard evaluation and zonation mapping inmountainous terrain [J]. Engineering Geology, 32 (4): 269-277.

Suitability Analysis for Landslide Susceptibility Assessment in the Mountain Area by Logistic Regression and Information Model

ZOU Feng-chai[1], LENG Yang-yang[1,2], TAO Xiao-lang[1], Yin Yan-di[1]

(1. *Guizhou Geological Environment Monitoring Institute*, *Guiyang* 550081, *Guizhou*, *China*;
2. *Chengdu University of Technology*, *Chengdu* 610059, *Sichuan*, *China*;)

[**Abstract**] Taking Wanshan district as an example, six factors including engineering geological rock group, slope structure, average slope, landform, distance from structure and distance from water system were selected based on the study of landslide hazard condition in Wanshan District. Logistic regression model and information model were used to evaluate the landslide susceptibility, and the results showed that the proportion of medium-high susceptibility area is 15.78% and 19.70% respectively, and 82% of geological disasters fall in this region. In the information model, the area of medium-high susceptibility area accounted for 12.41% and 25.19%, including 88% of the landslide disasters. Finally, the distribution of actual disaster points in each prone area was tested. The proportion of disaster points in the high-prone area in the logistic regression model was much smaller than that in the information model, and the actual ratio of disaster points in the high-prone level was small. It shows that the evaluation result of information quantity model is more objective and accurate for the regional landslide geological hazard susceptibility evaluation.

[**Key Words**] Landslide; Susceptibility evaluation; Logistic regression model; Information model; Wanshan District