

应用 CF 和 Logistic 回归模型编制滑坡危险性区划图

王卫东^{1,2}, 陈燕平¹, 钟 晟¹

- (1. 中南大学 商学院, 湖南 长沙, 410083;
2. 中南大学 土木建筑学院, 湖南 长沙, 410075)

摘 要: 根据地理信息系统(GIS), 以滑坡发生确定性系数(CF)和 Logistic 回归模型编制贵州省滑坡灾害区划图, 其主要步骤为: 应用历史滑坡灾点面积和滑坡致灾因子子集面积计算 CF, 由此确定主要的滑坡致灾因子; 应用 Logistic 回归模型、地理信息系统(GIS)空间分析工具和统计软件 SPSS 寻求最合适的模型描述灾害是否发生(因变量)与致灾因子(自变量)之间的关系; 计算研究区域内各单元的滑坡概率, 划分危险性等级, 绘制基于 GIS 的贵州省滑坡灾害区划图; 最后, 用历史灾害分布数据检验滑坡灾害区划图。研究结果表明: 根据区划图, 贵州省可分为 4 个区域, 即低危险区、中等危险区、高危险区和极高危险区; 对贵州省的危险性分区合理, 结合 CF 与 Logistic 回归模型编制滑坡灾害区划图的方法有效。

关键词: 滑坡; 危险性区划; 确定性因素; Logistic 回归模型; 地理信息系统

中图分类号: P642.22

文献标识码: A

文章编号: 1672-7207(2009)04-1127-06

Landslides susceptibility mapped with CF and Logistic regression model

WANG Wei-dong^{1,2}, CHEN Yan-ping¹, ZHONG Sheng¹

- (1. School of Business, Central South University, Changsha 410083, China;
2. School of Civil and Architectural Engineering, Central South University, Changsha 410075, China)

Abstract: The landslides susceptibility mapped in Guizhou Province was established through integrating a certainty factor (CF) and the Logistic regression model based on geology information system (GIS). The procedures were as follows: Firstly, the key factors were selected according to CF deduced from the area of landslides inventory and the area of factors subclasses. Secondly, the most appropriate model was selected to describe the relationship between hazards (dependents) and the key factors (independents) using the Logistic regression model, GIS spatial analysis tools and SPSS. Thirdly, the landslide probability was calculated, the dangerous levels were divided and landslides susceptibility map was established. Finally, the susceptibility map was verified by the distribution of landslide inventory. The results show that according to susceptibility map, the study area can be classified into four categories of landslide susceptibility, i.e., low dangerous zone, moderate dangerous zone, high dangerous zone and very high dangerous zone. The method of integrating the CF and the Logistic regression model is an effective way to evaluate the landslide susceptibility.

Key words: landslide; dangerous zone; certainty factor; Logistic regression model; geology information system

滑坡是世界范围内发生最多的地质灾害之一。根据中国国土资源部资料^[1], 2007年, 全国各类地质灾害共造成 679 人死亡和失踪, 直接经济损失达 24.75 亿元。在中国西部的贵州省, 2007年, 滑坡导致大量基

础设施被破坏, 150 多人丧生。基于地理信息系统(GIS)的滑坡灾害区划图是历史和潜在滑坡风险的有效评估工具, 可用于土地利用和基础设施规划以及发展早期预警系统^[2]。地理信息系统具有强大的地理信息存储、

收稿日期: 2009-01-12; 修回日期: 2009-03-25

基金项目: 交通部西部项目(200331880201)

通信作者: 王卫东(1971-), 男, 江西上饶人, 博士, 副教授, 从事道路与铁道工程研究; 电话: 0731-82655765; E-mail: 147745@163.com

处理、分析和展示功能,正成为地质灾害分析和风险控制的一个主要平台。作为基本的分析工具,整合了很多分析模块的地理信息系统能有效地管理和操作不同规模(某个局部区域或全国范围)的滑坡灾害空间数据。在过去 20 年中,大量的地理学家和工程技术人员运用GIS识别滑坡,绘制滑坡灾害区划图。为评价滑坡危险性,采用了多种方法,概括为:概率与统计模型^[3-6](如Logistic回归模型^[5]和确定性系数(CF)多元回归模型^[6]),数学模型^[7-8](如人工神经网络模型),确定性模型(安全系数法^[9-10]),经验方法(权重线性组合模型(WLC)^[11-13])等。在此,本文作者基于GIS和SPSS,采用确定性系数(CF)和Logistic回归模型相结合的方法编制贵州省滑坡灾害危险性区划图。用CF提取滑坡的关键致灾因子,将这些因子导入SPSS统计工具,得出Logistic回归模型的各项系数,应用回归模型计算研究区域内滑坡概率,并在ArcGIS下绘制滑坡危险度分区图。

1 滑坡致灾因子

1.1 滑坡致灾因子及其分类

贵州省位于北纬 103°36′~109°35′,东经 24°37′~29°13′,占地面积为 176 167 km²。根据调查,贵州省潜在的滑坡致灾因子有高程、岩性、坡度、坡向、地形地貌、河流切割、地质构造等,滑坡的诱发因子主要是降雨和人类改造自然的活(如公路、铁路的边坡开挖)。由于各致灾因子在不同等级(见表 1)下对滑坡的影响程度不同,因此,有必要将各致灾因子进一步细化等级。据GIS,部分致灾因子及其再分类如图 1~4 所示。其中:

图 1 所示是根据岩性对滑坡的影响程度划分的岩组等级,即稳定岩组、偶滑岩组、较易滑岩组和易滑岩组。

图 2 所示为滑坡和区域地质构造之间的关系。在 ArcMap 环境中,按与地质构造线距离取 20 级缓冲,每级 1 km。然后,统计每级缓冲距离内的历史滑坡灾点数,最后,将 20 级灾点数量进行归类与分析,确定灾害点与构造线的距离关系可分为 3 级,即 0~6 km, 6~16 km 和大于 16 km。

图 3 所示是按地形起伏情况将研究区域划分的等级,共分为 4 级,即大起伏山地、中起伏山地、小起伏山地和平原地区。

图 4 所示为滑坡和公路之间的关系。在 ArcMap 环境中,按与公路距离取 20 级的缓冲,每级 1 km。

再统计每级缓冲距离内的历史滑坡灾点数。最后,将 20 级的灾点数进行归类与分析,确定灾害点与公路的距离关系可分为 3 级,即:0~9 km, 9~15 km 和大于 15 km。

图 5 所示为贵州省历史滑坡分布图,根据国土资源部资料与国家地理信息系统绘制。

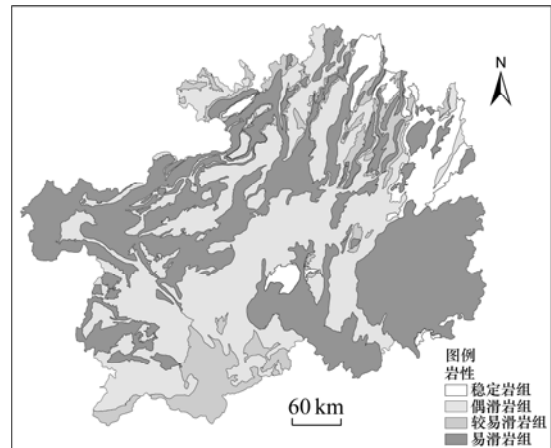


图 1 岩性分组

Fig.1 Groups of slide-prone lithological

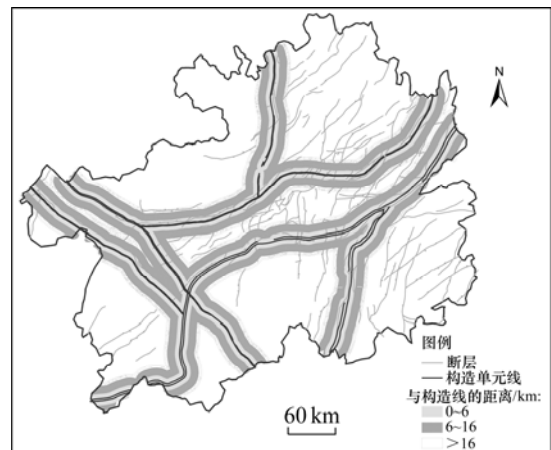


图 2 与构造线的距离分级

Fig.2 Grades of proximity to structural line

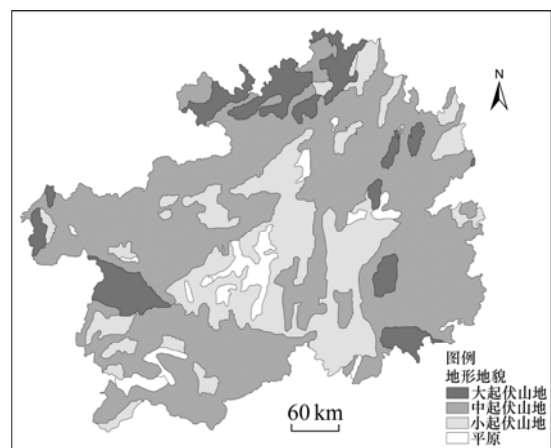


图 3 地形起伏分级

Fig.3 Grades of topography

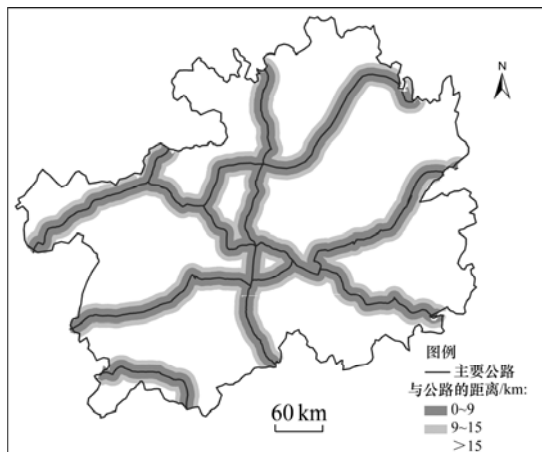


图 4 与公路的距离分级

Fig.4 Grades of proximity to drainage

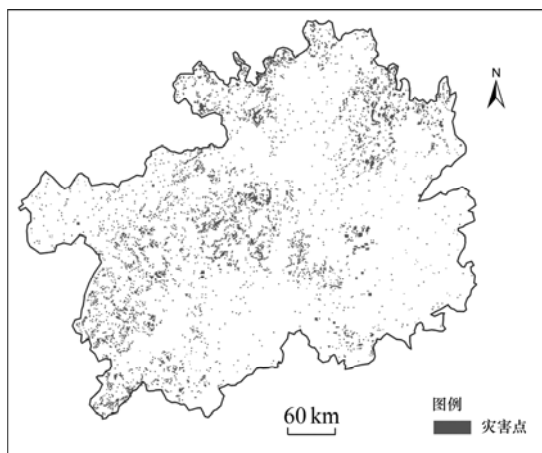


图 5 历史滑坡分布

Fig.5 Distributions of landslides inventory

1.2 滑坡关键因子的选取

确定性系数(CF)是一个概率函数^[14], 其表达式为:

$$K_{CF} = \begin{cases} \frac{P_a - P_s}{P_a(1 - P_s)}, & \text{当 } P_a > P_s \text{ 时;} \\ \frac{P_a - P_s}{P_s(1 - P_a)}, & \text{当 } P_a < P_s \text{ 时。} \end{cases} \quad (1)$$

式中: P_a 为事件在集合 a 中发生的条件概率, 即滑坡在某个因子集中滑坡面积 (S_{ij}^s) 与该因子集面积 (见表 1) 的比值; P_s 为事件在整个集合中发生的先验概率, 即研究区内滑坡面积与研究区总面积之比; K_{CF} 为确定性系数, 其变化区间为 $[-1, 1]$, 当 K_{CF} 为正时, 其值越大, 表明滑坡变形失稳的确定性越高, 即该单元为滑坡易滑区 (特殊地, 当 $P_a=1$ 时, $K_{CF}=1$); 当 K_{CF} 为负时, 其值越小, 表明越不易发生滑坡 (特殊地, 当 $P_a=0$ 时, $K_{CF}=-1$); K_{CF} 接近于 0, 表示先验概率与条件概率接近, 不能确定滑坡是否发生。

据 GIS, 首先将致灾因子划分为不同的因子子集 (见表 1), 然后, 将每个因子图层 (图 1~4) 与历史滑坡图层 (图 5) 进行叠加, 统计每一子集面积和子集中所包含的滑坡面积, 根据式 (1) 计算每一子集的 K_{CF} 。每个因子子集的 K_{CF} 比较分散, 需要 1 个统一值以决定该因子的重要程度, 即确定滑坡发生的关键因子。因此, 将各子集的 K_{CF} 进行合并, 即

$$z = \begin{cases} x + y - xy, & \text{当 } x > y > 0 \text{ 时;} \\ \frac{x + y}{1 - \min(|x|, |y|)}, & \text{当 } x \text{ 和 } y \text{ 异号时;} \\ x + y + xy, & \text{当 } x < y < 0 \text{ 时。} \end{cases} \quad (2)$$

式中: x 和 y 分别指要合并的 2 个 K_{CF} 。

使用式 (2) 时, 1 次仅能合并 2 个 K_{CF} 。本文从最小值开始, 依次合并^[14], 得出 z 。根据 z 确定每一种致灾因子对滑坡的影响程度, 进而确定滑坡发生的关键因子。

首先考虑 10 个致灾因子, 分别为高程、年均降雨量、岩性、地形地貌、与河流的距离、与构造线的距离、与铁路的距离、与公路的距离、坡度和坡向。各因子子集的 K_{CF} 和因子 z 的计算结果如表 1 所示。

K_{CF} 和 z 越大, 表明对应的因子致灾的权重就越大。从表 1 可以看出, 高程、年均降雨量和坡向的 z 出现了负数, 经分析认为这 3 个因子对滑坡的影响不大, 可以不考虑。岩性、与河流的距离、坡度、与构造线的距离、与公路的距离、与铁路的距离、地形地貌这 7 个因子的 z 大于 0, 可以认为它们与滑坡的相关性较大, 是滑坡危险性评价中需要考虑的滑坡致灾因子。

需要说明的是, 雨量是滑坡灾害的重要诱发因子, 一般在滑坡预警预报中考虑的是预报当日之前 1 周的有效雨量 (或累计雨量) 和当天预报雨量。而本文在编制的滑坡危险性区划图过程中, 年均降雨量对滑坡的分布没有显著影响 (见表 1)。

2 Logistic 回归模型及其在滑坡中的应用

所谓 Logistic 回归, 是指应变量为二值 (二项) 分类变量的回归分析。在地质灾害和灾情分析中, 将各致灾因子作为自变量, 灾害的发生与否作为因变量 (其中, 0 代表灾害不发生, 1 代表灾害发生)^[15]。

设 P 为灾害发生的概率, 取值范围为 $[0, 1]$; $1-P$ 为灾害不发生的概率。将 $P/(1-P)$ 取自然对数 ($\ln(P/(1-P))$), 即对 P 进行 Logit 转换, 记为 $\text{Logit } P$, 则 $\text{Logit } P$ 的取值范围为 $(-\infty, +\infty)$ 。以灾害发生的概

表 1 因子子集的 K_{CF} 和因子 z
Table 1 Subclasses' K_{CF} and factors' z

因子	各因子子集	各子集灾点面积/km ²	各子集面积/km ²	K_{CF}	z
高程/m	< 500	617.3	9 674.85	-0.086 310 417	-0.319 2
	500~1 000	4 481.85	74 984.04	-0.147 748 358	
	1 000~1 500	4 984.51	64 195.13	0.113 938 227	
	1 500~2 000	1 849.62	19 731.25	0.278 873 501	
	> 2 000	314.47	7 861.08	-0.441 349 167	
降雨量/(mm·a ⁻¹)	< 900	61.14	2 724.52	-0.692 097 886	-0.772 2
	900~1 100	353.69	8 068.85	-0.385 108 889	
	1 100~1 300	6 859.09	91 493.02	0.080 063 955	
	1 300~1 500	3 610.04	58 223.53	-0.113 390 349	
	> 1 500	1 363.79	16 014.8	0.199 060 537	
岩性	易滑岩组	5006	78 916.46	-0.091 539 219	0.2391
	较易滑岩组	1 807.92	16 986.62	0.374 057 954	
	偶滑岩组	4 679.6	68 958.5	-0.023 524 077	
	稳定岩组	754.23	11 663.47	-0.072 678 028	
与河流的距离/km	0~2	1 750.65	21 667.67	0.151 784 797	0.068 3
	2~8	4 185.54	59 941.31	0.006 839 720	
	8~20	4 928.91	73 861.88	-0.040 946 407	
	20 km 以外	1 382.69	21 053.91	-0.057 215 345	
坡向/(°)	平地	309.58	5 514.56	-0.202 736 645	-0.298 5
	北(0~45, 315~360)	2 781.5	39 862.98	0.005 442 210	
	东(45~135)	3 208.05	47 535.74	-0.029 905 594	
	南(135~225)	2 735.34	41 459.44	-0.053 157 038	
与构造线的距离/km	0~6	2 463.91	28 031	0.226 340 759	0.207 1
	6~16	2 738.33	36 695.33	0.075 436 408	
	> 16	7 045.91	111 799	-0.097 853 660	
地形地貌	大起伏山地	1 458.72	18 025.17	0.153 282 548	0.136 8
	中起伏山地	7 131.88	109 281.37	-0.063 545 179	
	小起伏山地	3 571.45	47 947.28	0.073 633 355	
	平原	85.74	1 270.96	-0.029 705 626	
坡度/(°)	0~1.56	8 124.7	120 997.51	-0.036 042 917	0.165 5
	1.56~12.39	858.35	11 941.34	0.035 831 746	
	12.39~23.22	1 662.89	22 205.15	0.077 545 406	
	23.22~63	1 602.14	21 128.2	0.089 930 721	
与公路的距离/km	0~9	3 110.28	38 223.62	0.158 308 329	0.131 8
	9~15	1 653.1	22 936.25	0.040 119 530	
	> 15	7 484.4	115 365	-0.069 463 825	
与铁路的距离/km	0~6	1 427.5	13 074.52	0.391 699 415	0.591 8
	6~12	1 198.28	12 021.67	0.326 582 301	
	12~20	1 187.77	15 245	0.117 637 732	
	> 20	8 434.23	136 184	-0.114 464 893	

率 P 为因变量, 建立回归方程:

$$\text{Logit } P = B_0 + B_1 I_{1j} + B_2 I_{2j} + \dots + B_n I_{nj} \quad (3)$$

式中: 影响因子 $I_{1j}, I_{2j}, \dots, I_{nj}$ 为自变量; B_1, B_2, \dots, B_n 为逻辑回归系数。令

$$P = \frac{e^z}{1 + e^z} \quad (4)$$

$$\text{则 } z = B_0 + B_1 I_{1j} + B_2 I_{2j} + \dots + B_n I_{nj} \quad (5)$$

Logistic回归模型是普通多元线性回归模型的推广, 它的误差项服从二项分布而非正态分布, 模型中 B_0 为常数, 表示在不接触任何潜在危险或保护因素条件下, 效应指标发生与不发生事件的概率之比的对数。

在式(3)和式(5)中, 若逻辑回归系数 B_1, B_2, \dots, B_n 被确定, 则根据不同的指标 I_{ij} , 可计算某一区域发生灾害的概率 P 。根据 P , 可以划分灾害发生可能性等级。因此, 问题转化为利用各因子指标 I_{ij} , 求算逻辑回归系数 B_1, B_2, \dots, B_n 。

式(3)和式(5)需要 1 个统一量纲指标 I_{ij} , 而表 1 中各因子量纲和取值范围都不同, 因此, 首先需要统一各滑坡致灾因子的量纲。本文应用历史滑坡空间分布, 实现各因子量纲统一。在 ArcMap 中将各因子图层分别与历史灾点图层进行叠加, 应用 GIS 空间分析技术得到致灾因子子集中历史灾害面积 S_{ij}^z , 由下式计算指标 I_{ij} 。

$$I_{ij} = \frac{S_{ij}^z}{\sum_{j=1}^m S_{ij}^z} \quad (6)$$

式中: i 为第 i 个致灾因子; j 为第 i 个致灾因子中第 j 个子集; m 为第 i 个致灾因子的子集个数。

将各因子指标 I_{ij} 作为 Logistic 回归模型的自变量, 灾害是否发生作为因变量。在因子图层和历史灾点图层的叠加过程中, 若某不规则单元内有灾害发生, 则取值为 1, 否则, 取值为 0。利用 ArcGIS 将图像转化为数字矩阵, 将数据导入 SPSS 统计分析软件^[16], 采用 Binary logistic 回归分析, 得到贵州省滑坡危险性因子的 Logistic 回归模型:

$$z = -0.962 - 0.062 \times I_{1j} + 1.211 \times I_{2j} + 0.089 \times I_{3j} + 0.321 \times I_{4j} - 3.673 \times I_{5j} + 1.640 \times I_{6j} + 0.263 \times I_{7j} \quad (7)$$

式中: $I_{1j}, I_{2j}, \dots, I_{7j}$ 依次表示公路、河流、构造线、铁路、坡度、岩性、地形地貌等因子的指标。根据 SPSS 得到的结果, 显著性值接近于 0, 说明差异极显著, 显著性差异 χ^2 为 6 010.809, 模型统计通过检验。

统计贵州省内所有不规则单元的指标 I_{ij} , 运用式(7)和(4), 可计算每个不规则单元的滑坡发生概率 P 。根据所有 P , 用等要素法将贵州省滑坡危险性进行分级, 危险度分区结果如图 6 所示。根据图 6, 分别统计各危险级别覆盖的区域面积和历史滑坡面积, 结果

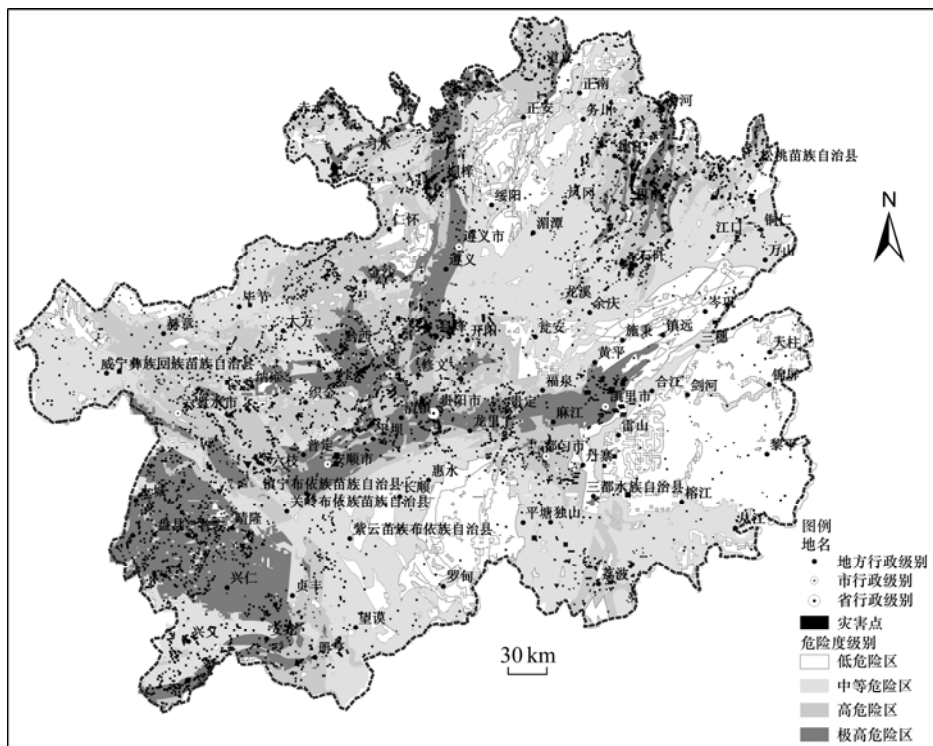


图 6 贵州省危险度分区图

Fig.6 Landslide susceptibility map of Guizhou Province

如表2所示。由表2可见,危险性级别越高的区域,滑坡面积占分级区域面积的比例就越高,即滑坡灾害密度越大。高危险区和极高危险区集中了大量的滑坡灾害,而低危险区和中等危险区的滑坡分布相对稀疏。这说明本文划分的危险性级别区域符合历史滑坡分布现状,编制的贵州省滑坡危险性区划图合理、有效。

表2 危险性级别
Table 2 Dangerous level

危险性级别	滑坡面积/ km ²	分级区域面积/ km ²	滑坡面积百分 比/%
低危险区	610.41	42 512.62	1.44
中等危险区	3 143.20	65 479.97	4.80
高危险区	3 821.77	37 278.87	10.25
极高危险区	4 648.96	30 117.49	15.44

图6表明,在贵州省西部和中部地区滑坡点较多,灾害情况严重,东南部灾点较少。根据表2,贵州省境内低、中、高和极高危险性分级区域面积占贵州省总面积的比例分别为24.2%,37.3%,21.2%和17.3%,说明贵州省约2/3地区属于中、低等地质灾害危险区,约1/3地区属于地质灾害高和极高危险区。

3 结 论

a. 贵州省滑坡主要分布在西部和中部,而东南部灾点较少。低、中、高和极高危险区分别占贵州省面积的24.2%,37.3%,21.2%和17.3%。

b. 根据 K_{CF} 取值的分布情况,在铁路、公路沿线滑坡灾害发生较多,说明贵州省内人为的环境改造是滑坡灾害产生的主要原因之一,在以后的交通基础设施建设过程中,应尽量避免深挖高填。

c. 根据历史灾点分布情况,应用确定性系数(CF)和Logistic多元回归模型编制贵州省滑坡危险性区划图,能客观评价区域滑坡的危险性程度。

参考文献:

[1] 中国国土资源部. 2007年地质灾害防治取得显著成效[R]. 北京: 中国国土资源部, 2008.
Ministry of Land and Resource of China. Remarkable effect achieved in preventing and controlling geological disasters in 2007[R]. Beijing: Ministry of Land and Resource of China, 2008.

[2] 王卫东, 刘武成. 基于GIS的公路地质灾害信息管理和决策支持系统[J]. 中南工业大学学报: 自然科学版, 2003, 34(5): 302-305.
WANG Wei-dong, LIU Wu-cheng. Road's geological hazards information management and decision making support system

research based on GIS[J]. Journal of Central South University of Technology: Natural Science, 2003, 34(5): 302-305.

[3] Baeza C, Corominas J. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques[J]. Earth Surface Processes and Landforms, 2001, 26(12): 1251-1263.

[4] Clerici A, Perego S, Tellini C, et al. A procedure for landslide susceptibility zonation by the conditional analysis method[J]. Geomorphology, 2002, 48(4): 349-364.

[5] 李雪平. 基于GIS的区域斜坡稳定性评价Logistic回归模型研究[D]. 武汉: 中国地质大学工程学院, 2005.
LI Xue-ping. Study on logistic regression model applied to regional slope stability evaluation based on GIS[D]. Wuhan: Engineering Academy, China University of Geosciences, 2005.

[6] 秦晓敏. 基于GIS的滑坡地质灾害危险性评价研究: 以巩留县为例[D]. 乌鲁木齐: 新疆大学资源与环境科学学院, 2007.
QIN Xiao-min. Based on GIS landslide geological disaster hazard evaluation research: Taking Gongliu county as example[D]. Urumqi: College of Resources & Environmental Science, Xinjiang University, 2007.

[7] Pistocchi A, Luzi L, Napolitano P. The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods[J]. Environmental Geology, 2002, 41(7): 765-775.

[8] Lee S, Ryu J H, Won J S, Park H J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network[J]. Engineering Geology, 2004, 71(3/4): 289-302.

[9] Refice A, Capolongo D. Probabilistic modeling of uncertainties in earthquake-induced landslide hazard assessment[J]. Computers & Geosciences, 2002, 28(6): 735-749.

[10] Zhou G, Esaki T, Mitani Y, Xie M, et al. Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach[J]. Engineering Geology, 2003, 68(3/4): 373-386.

[11] WANG Wei-dong, XIE Cui-ming, DU Xiang-gang. Landslides susceptibility mapping based on geographical information system, GuiZhou, south-west China[J]. Environmental Geology, 2009, 58(1): 33-43.

[12] Eastman R. Multi-criteria evaluation and GIS[M]. New York: Wiley, 1999: 493-502.

[13] Jiang H, Eastman J R. Application of fuzzy measures in multicriteria evaluation in GIS[J]. Int J Geog Inf Sci, 2000, 14: 173-184.

[14] 肖桐. 基于GIS的兰州市滑坡空间模拟研究[D]. 兰州: 兰州大学资源环境学院, 2006.
XIAO Tong. GIS-based research on spatial simulation of landslide in Lanzhou[D]. Lanzhou: College of Earth and Environmental Science, Lanzhou University, 2006.

[15] 丛威青, 潘懋, 李铁锋, 等. 基于GIS的滑坡、泥石流灾害危险性区划关键问题研究[J]. 地学前缘, 2006, 13(1): 185-190.
CONG Wei-qing, PAN Mao, LI Tie-feng, et al. Key research on landslide and debris flow hazard zonation based on GIS[J]. Earth Science Frontiers, 2006, 13(1): 185-190.

[16] 张文彤. SPSS统计分析教程(高级篇)[M]. 北京: 北京希望电子出版社, 2002.
ZHANG Wen-tong. SPSS statistic analysis tutorial (Advanced piece)[M]. Beijing: Beijing Hope Electronic Press, 2002.