

# 区域滑坡因子敏感性的 Logistic 回归分析

李雪平<sup>1</sup>, 唐辉明<sup>1</sup>, 周顺平<sup>2</sup>

(1. 中国地质大学 工程学院, 武汉 430074; 2. 中国地质大学 信息工程学院, 武汉 430074)

[摘要] 应用 Logistic 回归模型评价区域滑坡因子敏感性。将巫山县新址西区选作研究区, 选取地形地貌、岩性特征、地质构造等因子, 以 MAPGIS 地理信息系统为技术支撑, 将全区按  $10\text{ m} \times 10\text{ m}$  大小划分为 14 450 个格网单元, 每个格网单元作为一个样本数据, 建立 Logistic 回归模型。对回归模型中取得的各项参数进行分析, 得出坡度为  $15^\circ \sim 30^\circ$ 、坡高 300 m 以内、坡向南、岩性  $T_2b_1$  和  $T_2b_3$ 、顺向坡  $2(\beta > \alpha)$  和切向坡、距有影响的构造线 60 m 内为研究区滑坡的敏感因素。结果表明采用 Logistic 回归方法具有较高可信度, 可为滑坡灾害的防治、预测及土地利用服务。

[关键词] 滑坡; 因子; 敏感性; Logistic; 地理信息系统

[中图分类号] TP319 [文献标识码] A [文章编号] 1672-6561(2005)04-0014-05

[作者简介] 李雪平(1969-), 女, 重庆垫江人, 讲师, 从事环境地质、地理信息系统应用研究。

## 0 引言

影响斜坡稳定性的因素很多, 可归纳为内因和外因。内因包括斜坡岩土性质、结构构造等; 外因包括水文地质条件、地震活动、人类活动等。在滑坡稳定性评价<sup>[1-3]</sup>中, 敏感性分析用来对滑坡稳定性有较大影响的因素进行检查和分析, 确定影响滑坡稳定的敏感因素, 从而使滑坡防治和监测更有针对性, 使治理、设计更安全、可靠和经济。

目前滑坡因子敏感性分析一般采用单因素分析法, 多因素敏感性分析的研究尚处于探讨阶段。Logistic 回归模型是二分类因变量(因变量  $y$  只取两个值)进行回归分析时经常使用的统计分析方法。与线性回归不同, Logistic 回归是一种非线性模型, 普遍采用的参数估计方法是最大似然估计法<sup>[4]</sup>。在随机样本条件下, Logistic 模型的最大似然估计具有一致性、渐进性和渐进正态性<sup>[5]</sup>。由于一个地区历史滑坡的发生用有或无来表示, 可以运用 Logistic 回归模型来比较每种因素对滑坡发生的贡献大小, 即滑坡因子的敏感性分析。如 Ohlmacher Gregory C

等<sup>[6]</sup>在 Kansas 东北地区利用 Logistic 回归模型, 在 ArcView GIS 支持下进行了区域滑坡发生概率的评价制图, 并得出了区域滑坡敏感岩性的排序。Appt Jeremy 等<sup>[7]</sup>在 Oregon 海岸区利用地形变量进行 Logistic 回归来划分滑坡地区与临近地区。Lee S 等<sup>[8]</sup>在韩国 Yongin and Janghung 地区选择了 14 个与滑坡发生有关的因素, 利用 Logistic 回归模型和人工神经网络模型进行滑坡预测, 为灾害防治和土地利用服务。Rowbotham David N<sup>[9]</sup>等在尼泊尔 Phewa Tal 地区在 GIS 技术支持下, 利用 Logistic 模型进行斜坡稳定性分析等。

由于 Logistic 回归模型的非线性, 系数估计采用极大似然估计法, 必须通过迭代计算完成, 不借助于计算机技术, 几乎无法求解。加上地质条件不但与其自身的条件有关, 还与其所处的地理位置有关, 故对地质数据的处理涉及原始资料的输入问题和成果资料的编图问题以及数据多种形式的分类问题, 这是一个棘手的问题。GIS 技术提供了解决问题的途径。故一般进行地质问题的 Logistic 回归基本上采用 GIS 技术作为支撑手段。

## 1 Logistic 回归的概率公式

对一个试验样品在一组自变量作用下所发生的结果用指示变量  $Y$  表示,  $Y$  的赋值规则为

[收稿日期] 2004 12 31

[基金项目] 国家自然科学基金项目(40072085)

$$Y = \begin{cases} 1 & \text{出现成功的结果} \\ 0 & \text{出现失败的结果} \end{cases}$$

记出现成功结果的概率为  $P$ , 出现失败结果的概率为  $Q$ ,  $x_1, x_2, \dots, x_m$  表示对结果  $Y$  的  $m$  个影响因素, 用 Logistic 回归公式表示出现成功结果的概率为

$$P = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}} \quad (1)$$

由于  $P + Q = 1$ , 根据式(1) 得到出现失败结果的概率为

$$Q = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}} \quad (2)$$

由式(2)可以看出, 一个试验样品出现某种结果的概率与有关因素之间呈曲线关系。两个概率的比数定义为发生比 *odds*:

$$odds = P/Q = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{e^{\beta_0} \times e^{\beta_1 x_1} \times \dots \times e^{\beta_m x_m}} \quad (3)$$

取自然对数有

$$\ln(P/Q) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m \quad (4)$$

将  $\beta_0, \beta_1, \beta_m \dots$  称为 Logistic 回归系数。由于式(4) 为非线性方程, 常用 Newton Raphson 方法求解 Logistic 回归系数。

## 2 回归系数的显著性检验

当得到实际问题的经验回归方程后, 不能立即用它去进行分析和预测, 因为回归方程是否真正描述了因变量和自变量之间的统计规律性, 还需用统计方法对回归方程和回归系数进行假设检验。Logistic 回归常提供以下系数进行对比和检验: 回归系数估计量的标准差 ( $S.E.$ ), 回归系数检验的统计量值 ( $Wald$ ),  $Wald$  检验的显著性概率 ( $Sig$ ), 标准化的回归系数 ( $\beta^*$ ), 偏相关系数 ( $R$ )。一般情况下给出  $\exp(\beta)$ , 以便对比。

$$S.E. = \sqrt{\text{var}(\beta)} \quad (5)$$

$$Wald = \left( \frac{\beta}{S.E.} \right)^2 = \frac{\beta^2}{\text{var}(\beta)} \quad (6)$$

$$\beta^* = \frac{\beta S_x}{1.8138} \quad (7)$$

式中:  $S_x$  为  $x$  的标准差

$$R = \sqrt{\frac{Wald - 2df}{N - 2LE_0}} \quad (8)$$

式中,  $df$  为变量的自由度; 分母为  $-2$  倍截距模型

的对数似然值。

大多数统计分析用指明哪些系数在 1%, 5% 或其他适当显著性水平上显著的方式说明统计显著性结果。 $Sig$  值表示计量结果对应的精确显著性水平。检验统计量  $t$  值与  $Sig$  值的关系是

$$Sig(|t| > |t \text{ 值}|) = Sig \text{ 值} \quad (9)$$

要检验假设一般都是某个回归系数等于 0 的原假设。因此在系数为 0 的原假设为真的条件下,  $Sig$  值是得到其系数估计不小于已知估计系数的新样本数据的概率。在原假设为真时,  $Sig$  值越小就越不可能出现这种情况。反过来, 较大的  $Sig$  值意味着样本数据支持原假设。实质上  $Sig$  值就是放弃真错误的真实概率, 即检验的真实显著性。 $Sig$  值的求取一般采用计算机软件, 如 SPSS, SAS, EXCEL 等。

式(5)中  $S.E$  的大小反映了估计量取值波动程度。一般情况下,  $Wald$  值越大或  $Sig$  值越小, 则自变量在回归方程中的重要性越大。单样本  $t$  是检验某个变量的总体均值与指定的检验值之间是否存在显著差异。单样本  $t$  检验的零假设为  $H_0$ : 总体均值与指定检验值之间不存在显著差异。设定显著性水平  $\alpha$  后, 与  $Sig$  值进行比较, 若  $Sig$  小于或等于  $\alpha$ , 则拒绝  $H_0$ , 认为总体均值与检验值之间存在显著差异。标准化的回归系数表示自变量一个标准差的变化所导致的因变量上以其标准差为单位测量的变化。在应用标准化因变量和多个标准化自变量的模型中, 所有自变量与因变量的关系都是以同样的单位(标准差)测量的。所以, 虽然其原始变量是以不同尺度测量的, 但通过标准化后自变量对因变量的作用便具可比性。 $R$  用来检查因变量与每一个自变量之间的偏相关。 $R$  表示在控制其他变量的情况下, 该自变量对因变量的作用。 $R$  的值域为  $(-1, +1)$ , 为正值时表示当自变量增加时事件发生的可能性也增加, 为负值时则相反; 当绝对值很小时表示对应自变量对模型的偏贡献很小, 为负值时则相反。

## 3 研究区概况

### 3.1 地质环境特征

研究区为三峡库区巫山县新城址区。巫山县位于四川盆地东缘, 巫山山脉腹地, 北依大巴山。县城地处三峡水库库段中部, 水库蓄水后, 回水位为 175.4 m, 除北门坡部分居民区外, 绝大部分将被淹没。新址地处淮阳山字型构造西翼反射弧-大巴山

弧形构造、川东褶皱带及川鄂湘黔隆起褶皱带三大构造的结合部位,地质构造体系复杂,褶皱、断裂发育。基岩以三叠系滨海相碳酸盐岩为主,其次为陆相砂泥质岩类。主要包括:三叠系下统嘉陵江组第四段( $T_{1j4}$ )浅灰色中厚层灰岩、白云岩、泥质白云岩。三叠系中统巴东组第一段( $T_{2b1}$ )灰、深灰色泥质灰岩、白云质灰岩,第二段( $T_{2b2}$ )紫红色泥岩夹粉砂质泥岩、粉砂岩,第三段( $T_{2b3}$ )灰白色白云质灰岩、泥质白云岩、深灰色中厚层灰岩。新址西区代表性滑坡有:秀峰寺滑坡、四道桥-邓家屋场滑坡<sup>[7]</sup>。

3.2 影响因子选取

由于原始数据只包含了巫山县城新址西区 6 km<sup>2</sup> 范围,故只选取了影响斜坡稳定的基本因素:地形地貌、工程地质岩组、地质构造等,而降雨、地下水等影响因素由于基本相同没有考虑。具体因子的选取和分类如表 1。

表 1 影响因子及分类

Table 1 Category of the Influence Factors	
滑坡因子	分类
坡度/ $^{\circ}$	$\leq 10$ , 10~15, 15~20, 20~25, 25~30, 30~35, 35~40, $> 40$
高程/m	$\leq 150$ , 150~200, 200~250, 250~300, 300~350, 350~400, 400~450, 450~500, $> 500$
坡向/ $^{\circ}$	315~45, 45~90, 90~135, 135~180, 180~225, 225~270, 270~315
岩性	$T_{1j4}$ , $T_{2b1}$ , $T_{2b2}$ , $T_{2b3}$
坡形	顺向坡 1( $\beta < \alpha$ ), 顺向坡 2( $\beta > \alpha$ ), 逆向坡、切向坡
构造线距离/m	$\leq 61$ , 61~122, 122~183, $> 183$

构造线距离因子分类计算过程:首先将研究区四级以上每一条构造线作单因子 Logistic 回归,根据相关系数  $R$  的大小取舍构造线。构造线取舍的  $R$  限值为 0.1, 所得结果作为有影响构造线保存。计算距最近有影响构造线距离的算法为寻找每一格网单元离得最近的有影响构造线距离。根据对有影响的构造线距离的单因子 Logistic 回归系数聚类,划分为: $\leq 61$  m, 61~122 m, 122~183 m,  $> 183$  m。

4 研究方法及结果分析

4.1 研究方法

研究的技术路线是将新址西区按 10 m $\times$ 10 m 方格划分为 14 450 个格网单元。将每一个单元作为一个样本。其中已发生滑坡地段的单元数为

1 705 个(在计算中,当滑坡面积占格网单元的 40% 时,就认为该格网单元代表滑坡事件发生),记  $Y=1$ , 其余单元  $Y=0$ 。将各因素的连续变量划分为分类变量(见表 1)。在 MAPGIS 系统的支持下,将地形数据、岩性数据、对斜坡稳定有影响的构造线数据输入形成相应的文件。在 MAPGIS 二次开发库的支持下,用 Visual C++ 6.0 进行二次开发,生成斜坡稳定性评价系统。

4.2 结果分析

4.2.1 坡度因子对滑坡敏感性的影响

坡度因子每组  $Sig$  值均达到设计检验水平  $\alpha(0.05)$ 。由标准化回归系数  $\beta^*$  绝对值的大小可知,坡度为  $15^{\circ}\sim 30^{\circ}$  是研究区滑坡的敏感因素,坡度  $15^{\circ}\sim 30^{\circ}$  三类滑坡的  $odds$  分别为参照组  $\leq 10^{\circ}$  的 1.703 7, 1.566 7, 1.801 0 倍。

4.2.2 高程因子对滑坡敏感性的影响

高程因子 200~250 m, 250~300 m,  $> 500$  m 三组  $Sig$  值未达到检验水平。由  $\beta^*$  绝对值的大小可知,高程在 350~500 m 对滑坡发生的影响最大,由于为负值,其意义为降低滑坡发生的因素。高程在 150~300 m 滑坡分类的  $odds$  分别为参照组  $\leq 150$  m 的 1.128 3, 0.973 3, 1.008 8 倍,故可得出高程在 300 m 以内为场址区滑坡易发地段的结论。

4.2.3 坡向因子对滑坡敏感性的影响

坡向因子  $225^{\circ}\sim 270^{\circ}$  组未达到检验水平。 $\beta^*$  最大分别为 1 350~1 800, 900~1 350, 1 800~2 250, 说明南向坡为场址区滑坡发生敏感方向。

4.2.4 岩性因子对滑坡敏感性的影响

$\beta^*$  排列为  $T_{2b1}$ ,  $T_{2b3}$ ,  $T_{2b2}$ ,  $odds$  分别为参照组  $T_{1j4}$  岩性中滑坡发生的 4.950 8, 2.258 2, 1.762 5 倍。

4.2.5 坡形因子对滑坡敏感性的影响

标准化回归系数  $\beta^*$  排队为顺向坡 2( $\beta > \alpha$ )、切向坡、逆向坡。相对于参照组滑坡发生的机会为 1.956 2, 1.527 5, 1.257 9 倍。

4.2.6 距最近构造线距离因子对滑坡敏感性的影响

标准化回归系数  $\beta^*$  排队为  $\leq 61$  m, 61~122 m, 122~183 m, 说明距离越近,滑坡发生可能性越大。

以上几个因素联合作用的影响显著,模型拟合效果较好,模型的预测能力:滑坡发生样本的判对率为 72.08%, 滑坡不发生样本的判对率为 81.44%。根据表 2 结果得到试验区 Logistic 回归模型(10), 滑坡分布预测如图 1。

表 2 研究区滑坡因子敏感性 Logistic 回归分析

Table 2 Result of Logistic Regression on Sensitivity Factor in Research Area

变 量	$\beta$	<i>S. E.</i>	<i>Wald</i>	df	<i>Sig</i>	<i>R</i>	ex p( $\beta$ )	$\beta^*$
坡度对照组 $\leq 10^\circ$			5.4788e008	7	0.0024	-0.0747		
10°~15°	0.227 7	5.787 4 e-005	1.547 6 e007	1	0.021 3	-0.051 1	1.255 7	0.040 14
15°~20°	0.532 8	3.991 5 e-005	1.775 1 e008	1	0.007 6	0.014 0	1.703 7	0.113 0
20°~25°	0.445 0	3.412 3 e-005	1.731 0 e008	1	0.008 3	0.015 1	1.566 7	0.103 1
25°~30°	0.588 4	3.464 0 e-005	2.884 8 e008	1	0.006 4	0.071 7	1.801 0	0.134 1
30°~35°	0.351 6	6.540 1 e-005	2.890 8 e007	1	0.014 6	0.020 3	1.421 4	0.058 3
35°~40°	0.434 2	1.307 0 e-004	1.103 4 e007	1	0.016 8	-0.021 2	1.543 7	0.050 9
> 40°	0.409 6	9.750 7 e-005	1.764 8 e007	1	0.015 3	-0.042 4	1.506 2	0.055 6
高程对照组 $\leq 150$ m			3.295 7 e008	8	0.002 9	0.059 2		
150~200 m	0.120 7	5.162 1 e-005	5.471 2 e006	1	0.037 8	0.097 4	1.128 3	0.022 5
200~250 m	-0.027 0	5.013 8 e-005	2.909 6 e005	1	0.163 0	0.079 2	0.973 3	-0.005 1
250~300 m	0.008 8	4.634 2 e-005	3.598 4 e004	1	0.419 5	0.102 5	1.008 8	0.001 7
300~350 m	-0.173 4	5.466 6 e-005	1.005 6 e007	1	0.027 1	0.038 1	0.840 8	-0.031 4
350~400 m	-0.906 1	4.353 6 e-005	4.331 2 e008	1	0.004 6	-0.132 5	0.404 1	-0.184 2
400~450 m	-0.912 7	6.730 0 e-005	1.839 3 e008	1	0.005 7	-0.118 0	0.401 4	-0.149 2
450~500 m	-0.696 1	7.832 0 e-005	7.900 2 e007	1	0.008 1	-0.109 6	0.498 5	-0.105 5
> 500 m	-0.006 8	8.869 2 e-005	5.863 5 e003	1	0.602 3	-0.028 7	0.993 2	-0.001 0
坡向对照组315°~45°			1.599 2 e007	6	0.011 4	-0.066 0		
45°~90°	-0.065 4	7.150 0 e-005	8.375 6 e005	1	0.081 8	-0.068 1	0.936 7	-0.010 4
90°~135°	0.250 7	2.845 0 e-005	7.767 8 e007	1	0.013 5	0.057 2	1.284 9	0.063 1
135°~180°	0.269 8	2.567 4 e-005	1.104 2 e008	1	0.012 0	0.072 1	1.309 7	0.071 4
180°~225°	0.117 5	5.277 4 e-005	4.958 0 e006	1	0.039 3	-0.006 7	1.124 7	0.021 7
225°~270°	0.062 5	1.038 2 e-004	3.629 7 e005	1	0.102 8	-0.066 0	1.064 5	0.008 2
270°~315°	0.139 4	1.896 5 e-004	5.399 4 e005	1	0.062 7	-0.065 6	1.149 5	0.013 6
岩性对照组 T <sub>1j4</sub>			1.363 2 e009	3	0.001 9	-0.105 2		
T <sub>2b1</sub>	1.599 5	4.153 8 e-005	1.482 9 e009	1	0.002 6	0.263 4	4.950 8	0.332 9
T <sub>2b2</sub>	0.566 8	2.846 9 e-005	3.963 2 e008	1	0.006 0	-0.138 2	1.762 5	0.142 5
T <sub>2b3</sub>	0.814 6	2.388 2 e-005	1.163 3 e009	1	0.003 8	-0.016 0	2.258 2	0.223 6
坡形对照组顺向坡 1			1.280 8 e007	3	0.009 2	-0.036 9		
顺向坡 2	0.671 0	8.359 9 e-005	6.442 0 e007	1	0.008 7	0.051 8	1.956 2	0.098 4
逆 向 坡	0.229 4	9.005 3 e-005	6.490 0 e006	1	0.026 3	-0.030 6	1.257 9	0.032 4
切 向 坡	0.423 6	4.314 0 e-005	9.597 3 e007	1	0.009 9	-0.003 4	1.527 5	0.086 4
构造线对照组> 183 m			3.225 7 e009	3	0.002 1	-0.161 3		
$\leq 61$ m	0.772 5	2.851 6 e-005	7.338 7 e008	1	0.004 4	0.121 4	2.165 2	0.194 0
61~122 m	0.575 2	3.365 2 e-005	2.921 3 e008	1	0.006 4	0.068 4	1.777 4	0.133 0
122~183 m	0.255 0	4.029 3 e-005	4.004 2 e007	1	0.015 8	-0.026 3	1.290 4	0.053 9
截 距	-4.076 7	7.238 2 e-005	3.172 2 e009	1				

$$P = \frac{e^{-4.0767 + 0.2278[x_1=1] + 0.5328[x_1=2] + \dots + 0.255[x_6=3]}}{1 + e^{-4.0767 + 0.2278[x_1=1] + 0.5328[x_1=2] + \dots + 0.255[x_6=3]}}$$

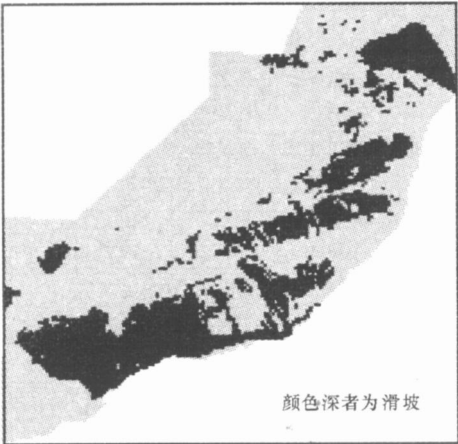


图 1 Logistic 回归模型判断的滑坡发生地段

Fig.1 Landslide Zone Predicted  
by Logistic Regression Model

5 结论

综合以上因素, 可得出坡度为 15° ~ 30°、坡高 300 m 以内、坡向南、岩性 T<sub>2</sub>b<sub>1</sub> 和 T<sub>2</sub>b<sub>3</sub>、顺向坡 2(β>α)和切向坡、距有影响的构造线距离 60 m 之内为研究区滑坡的敏感因素, 这与在工程实践中对试验区斜坡稳定性影响因素的认识是相符的。Logistic 回归模型用于斜坡稳定性分析在影响因素较多的情况下, 可以获得比较客观的结果。

Logistic Regression Analysis on Sensitivity  
of Regional Landslide Factors

LI Xue ping<sup>1</sup>, TANG Hui ming<sup>1</sup>, ZHO U Shun ping<sup>2</sup>

(1. School of Engineering, China University of Geosciences, Wuhan 430074, China;

2. School of Information Engineering, China University of Geosciences, Wuhan 430074, China)

**Abstract** The purpose of this study is to develop sensitivity analysis of regional landslide by Logistic regression model. The west area of Wushan County was chosen as a research area. The Logistic regression model was built that in independent variables were landform, physiognomy, lithological character and geological structure and dependent variable was landslide location data. The research area was divided into 14 450 cells by 10 m×10 m based on MAPGIS. It is concluded that the sensitivity factors of landslide in research area are gradient (between 15° and 30°), height (below 300 m), orientation of slope (south), strata (T<sub>2</sub>b<sub>1</sub> and T<sub>2</sub>b<sub>3</sub>), slope direction (consequent slope and tangential slope), distance to structural line affecting the slope (less than 60 m). The result can be used in prevention of hazard land use planning.

**Key words:** landslide; factor; sensitivity; Logistic; GIS

[ 参 考 文 献 ]

[ 1 ] 朱良锋 吴信才 殷坤龙 等. 基于信息量模型的中国滑坡灾害风险区划研究[ J ]. 地球科学与环境学报, 2004, (26) 3: 52 - 56.

[ 2 ] 杨中宝 彭省临 李朝艳. 基于 GIS 的人工神经网络矿产预测系统设计及应用[ J ]. 地球科学与环境学报, 2005, 27( 1 ): 30 - 33.

[ 3 ] 李亚兰 陈志新 王佳运. 多级模糊模式识别模型在地质环境质量评价中的应用[ J ]. 地球科学与环境学报, 2004, 26( 4 ): 90 - 93.

[ 4 ] 王济川 郭志刚. Logistic 回归模型—方法与应用[ M ]. 北京: 高等教育出版社, 2001.

[ 5 ] 何晓群 刘文卿. 应用回归分析[ M ]. 北京: 中国人民大学出版社, 2001.

[ 6 ] Ohlmacher Gregory C, Davis John C. Using multiple logistic regression and GIS technology to predict landslide hazard in Northeast Kansas, USA[ J ]. Engineering Geology, 2003, 69 ( 03 ): 331 - 343.

[ 7 ] Appt Jeremy, Skaugset Arne, Pyles Marvin. Discriminating between landslide sites and adjacent terrain using topographic variables[ C ] // Geological Society of America, Cordilleran Section, 98th annual meeting, 2002.

[ 8 ] Lee S, Ryu J, Min K, et al. Development and application of landslide susceptibility analysis techniques using geographic information system ( GIS)[ J ]. International Geoscience and Remote Sensing Symposium, 2001, 1: 319 - 321.

[ 9 ] Rowbotham David N, Dudycha Douglas. GIS modelling of slope stability in Phewa Tal watershed, Nepal[ J ]. Geomorphology, 1998, 26( 01 ): 151 - 170.

[ 10 ] 王智济 殷坤龙 伍法权. 长江三峡水利枢纽库区巫山县城镇迁建新址工程地质论证报告[ R ]. 武汉: 中国地质大学, 1995.

[ 英文审定: 苏生瑞 ]