

# 基于 GIS 的滑坡 CF 多元回归模型及其应用

兰恒星<sup>1</sup>, 伍法权<sup>2</sup>, 王思敬<sup>2</sup>

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院地质与地球物理研究所, 北京 100029)

**摘 要:** 目前基于 GIS 滑坡变形失稳危险性评价方法均有各自的局限性。本文在基于数据的数学统计模型的基础上提出 CF 多元回归模型。通过将滑坡确定性稳定系数与回归模型的融合, 在一定程度上解决了滑坡评价过程中影响因子的选择和量化的问题, 有利于建立准确的滑坡分析模型。最后, 将模型应用于云南小江流域, 进行了该区的滑坡空间分布及稳定性分析。

**关键词:** 滑坡发生确定性系数; 多元回归模型; 地理信息系统

**中图分类号:** P642.2

**文献标识码:** A

## 1 引言

滑坡的发生是一种非线性的物理过程, 并可能产生多种滑坡类型<sup>[1,2]</sup>。各种滑坡因子具空间可变性, 因子之间具有空间相关性。如果将各种滑坡类型简单地归为一类, 所产生的预测结果将不具可靠性。目前已有的分析模型均存在一定的局限性和不足, 也没有任何一种模型能正确划分所有的单元。对于同一地区的不同滑坡类型, 应采用不同的预测模型。或者对于同一滑坡类型采用不同的预测模型(采用不同的模型单元)以提高预测的精度<sup>[3]</sup>, 这就存在如何将各种不同预测模型进行合并和融合的问题。

统计分析模型是基于 GIS 进行滑坡变形失稳危险性评价的一种主要方法<sup>[4-6]</sup>。然而, 在实际应用中存在的一个主要问题是如何合并各种不同数据层的信息。在 GIS 中, 数据层存在的可以有多种方式, 可以以连续形式(如 DEM)存在, 也可以以离散形式(如岩性、植被覆盖、土地利用、钻孔数据等)存在。GIS 中不同数据类型的数据结构不同, 描述点线面的三种主要数据类型的数据结构可以是栅格数据结构, 也可以是矢量数据结构。对这些不同类型和结构的数据的合并便成了 GIS 应用中的一个难点<sup>[7,8]</sup>。另外, 所有方法存在的另一个局限是输入

数据具多相性、不均一性及由于数据本身的自然误差或在处理和分析过程产生的不确定性。这些均在一定程度上影响了分析结果。本文采用滑坡发生确定性系数来解决异质数据合并的问题, 并从实用的角度出发提出 CF 多元回归统计模型。

## 2 滑坡发生确定性系数(CF)

采用确定性系数进行滑坡的危险性区划的假定是滑坡的危险性可以根据过去滑坡与确定为诱发因素的数据集(地质、地形等)之间的统计关系进行确定。即未来滑坡在达到与自己或相似条件地区的其它滑坡发生时所处的相似的环境条件时, 将发生滑动。这些类比的滑坡可以是位于研究区, 也可以位于研究者已经掌握了滑坡发生与诱发因素之间相关关系的类似条件地区。其适用的模型单元类型为网格单元和均一条件单元。

定义确定性系数的函数  $f_k$  的一般理论公式为

$$f_k: \begin{cases} A \rightarrow [\min_k, \max_k] \rightarrow [a, b] \\ A \rightarrow \{1, 2, 3, \dots, n_k\} \rightarrow [a, b] \end{cases} \quad (1)$$

其中  $A$  为研究区域;  $\min_k$  和  $\max_k$  是表示连续数据区间;  $1, 2, 3, \dots$  表示非连续数据;  $[a, b]$  是确定性系数区间。通过函数变换, 不同类型的数据将落入相同的区间中  $[a, b]$ , 从而可以进行有效的合并。

任一单元的滑坡确定性系数 CF 定义为单元中

收稿日期: 2002 - 01 - 10; 改回日期: 2002 - 09 - 04。

基金项目: 中国科学院特别资助项目“滑坡、泥石流基础研究”。

作者简介: 兰恒星(1972 - ), 男(汉族), 博士。毕业于中国科学院地质与地球物理研究所工程地质力学开放研究实验室, 现为中国科学院资源与环境信息系统国家重点实验室地图学与地理信息系统专业博士后。电话: 010 - 64889764 e-mail: lanhx@reis.ac.cn

某一假定为真(例如,此单元为滑坡易发区)的确定程度。它根据每一数据层的单元上的事件(如滑坡发生)的先验概率和在特定数据条件下(如岩性为砂岩、坡度在 $30^\circ \sim 40^\circ$ ,坡向北东等)的事件发生(滑坡发生)的条件概率之间的关系确定。

CF作为一个概率函数,最早由 Shortliffe 和 Buchanan(1975)提出,由 Heckerman (1986)进行了改进,表示为下式<sup>[9]</sup>

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)} \text{ if } PP_a \geq PP_s; \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)} \text{ if } PP_a < PP_s, \end{cases} \quad (2)$$

式中  $PP_a$  是事件(滑坡)在数据类  $a$  中发生的条件概率,在实际滑坡应用时可以表示为代表数据类  $a$  的单元中存在的滑坡面积与单元面积的比值; $PP_s$  是事件在整个研究区  $A$  中发生的先验概率,可以表示为整个研究区的滑坡的面积与研究区面积的比值。

通过式(2)函数变换,CF的变化区间为 $[-1, 1]$ 。正值代表事件发生确定性的增长,即滑坡变形失稳的确定性高,此单元为滑坡易发区;负值代表确定性的降低,表示滑坡变形失稳的确定性低,不易发生滑坡;接近于0值代表先验概率与条件概率十分接近,事件发生的确定性不可能进行确定,即此单元不能确定是否为滑坡易发区。

计算出每一数据层的CF后,需要将不同数据层的CF进行合并。假定要合并两个数据层的CF分别为和,合并后的结果为,合并公式如下式

$$z = \begin{cases} x + y - xy, & x, y > = 0 \\ \frac{x + y}{1 - \min(|x|, |y|)}, & x, y \text{ 异号} \\ x + y + xy, & x, y < 0 \end{cases} \quad (3)$$

每个因子数据层的CF计算及合并均在GIS中完成。首先将因子数据层按一定规则进行划分为不同的数据类别,然后在GIS中将每个因子数据层与滑坡层进行叠加,计算因子层中每一数据类中滑坡的面积,滑坡的面积与数据类的面积相比得到滑坡在此数据类中发生的频度。根据公式(2)进行CF的计算,从而确定因子层的每一数据类对于滑坡发生的影响程度,进行因子的敏感性分析。将因子数据层的CF值进行按式(3)两两进行合并,按一定规则对合并后的CF值进行重新划分,得到滑坡变形失稳危险性区划结果。通过与新的滑坡数据的对比,可以确定每一种影响因子对滑坡发生的影响程

度,确定滑坡发生的关键因子。由此可以看出,CF的确定方法实际是属于双变量统计分析的范畴。

### 3 确定性系数与多元回归模型融合

通用的线性多元回归模型具有如下形式

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_mX_m + \varepsilon \quad (4)$$

式中  $Y$  为单元中滑坡发生频率(因变量);

$X$  为滑坡稳定性影响因子;

$B$  为多元回归系数;

$\varepsilon$  不误差。

在采用多元回归模型进行滑坡的危险性评价时,需要解决两个主要问题。一个是自变量,即滑坡影响因子的选择。为了得到较理想的回归效果,需要采用影响滑坡变形失稳的关键性因子,而非所考虑的影响因子越多越好。在所有的滑坡因子中,有一些因子与滑坡的变形失稳及分布之间并不存在明确物理意义的相关关系,在这种情况下,采用逐步的线性统计模型的方法将会使结果产生很大的误差,甚至不合理的结果。要想获得较好的预测结果,需要在模型中仅输入那些对滑坡的发生有显著影响作用的因子。通过滑坡确定性系数的合并及检验,可以非常直观的确定影响滑坡的关键性因子。

另一个主要问题是作为自变量的滑坡因子的量化问题,即异质数据类型的合并问题。如前所述,采用每一个因子的相应数据类别的CF值均位于相同的区间,且其数值的大小反映了相应类别对于滑坡变形失稳的影响程度,因此,将CF值作为多元回归模型中的自变量是一个较理想地解决多元回归模型中自变量量化的方法。

滑坡确定性系数与多元回归模型的融合的一般步骤如图1所示。

### 4 实例

采用CF多元回归模型对小江流域进行滑坡稳定性分析。小江流域位于滇东北,地理坐标是 $25^\circ 32'N \sim 26^\circ 35'N$ ,  $102^\circ 52'E \sim 103^\circ 22'E$ ,面积 $3\,086\text{ km}^2$ ,主要属东川市管辖的小江流域(图2)<sup>[10]</sup>。该区小江深大活动断裂纵贯南北,新构造运动强烈,地震活动频繁,滑坡、泥石流等地质灾害严重,斜坡稳定性差,分布有众多级别不同的滑坡,区域工程地质条件极为复杂。

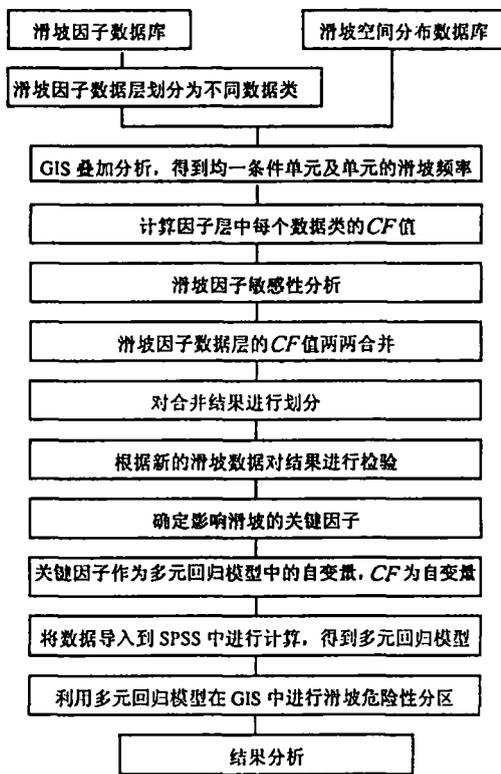


图1 滑坡确定性系数与多元回归模型融合流程  
Fig.1 Integration steps of CF Regression Model

将小江流域所有的因子在 GIS 中进行叠加分析, 得到进行统计计算的近二十万个均一条件单元。对于那些没有实际意义的小多边形、畸形多边形予以删除。每个均一条件单元均包含每一个影响因子的单一条件分组(表 1, 图 3), 从而可以认为单元之间是相互独立的, 有利于进行统计分析。

对小江流域进行影响因子敏感性分析, 得到小江流域滑坡的关键影响因子: 岩性岩组、结构岩组、坡度、高程、断裂。随机选取一组模型单元, 采用各因子数据类的 CF 值作为多元回归模型的自变量 X, 由于岩性岩组与结构岩组有较大的相关性, 采用岩性岩组 CF 与结构岩组 CF 的合并后的值岩性结构

表 1 部分均一条件单元属性

Table 1 Attribute of part of unique condition units

代号	结构类型	坡度范围 (°)	坡向	高程范围 (m)	断裂分段缓冲	凝聚力范围 (kPa)
1	层状碎裂结构岩组	10~20	西	1000~1500	缓冲区以外	22.2~27.6
2	层状碎裂结构岩组	20~30	西	1000~1500	缓冲区以外	22.2~27.6
3	层状碎裂结构岩组	0~10	西	1000~1500	达朵-阿旺 2000~4000m	22.2~27.6

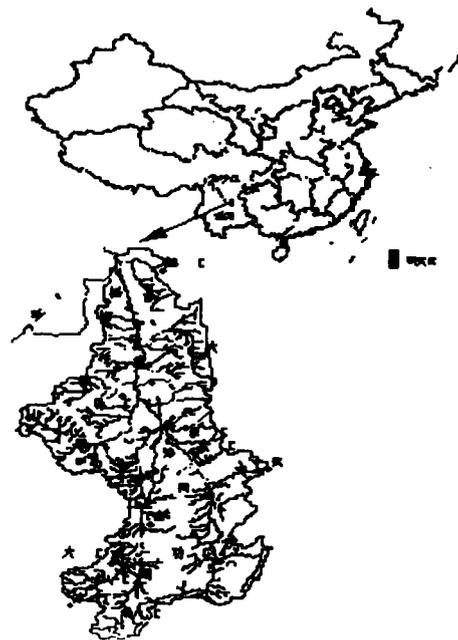


图2 研究区概况  
Fig.2 Study area

CF作为其中一个自变量。将最终叠加图层与滑坡图层进行叠加,确定每个单元的滑坡发生频率作为因变量Y。将数据导入到SPSS统计分析软件中,采用全变量线性多元回归分析得到小江流域的多元回归统计分析模型为

$$Y = -0.863 + 0.136 \times CF_1 + 0.0748 \times CF_2 + 0.123 \times CF_3 + 0.05389 \times CF_4$$

式中 Y为滑坡发生概率,CF<sub>1</sub>为岩性结构的概率函数,CF<sub>2</sub>为坡度的概率函数,CF<sub>3</sub>为高程的概率函数,CF<sub>4</sub>为断层的概率函数。

根据多元回归模型,在GIS中对每一个单元的滑坡发生频率进行计算,最后对小江流域进行稳定性区划,结果如图3。从多元回归分析结果的统计直方图中可以看出(图4),活动滑坡在稳定分区中的比

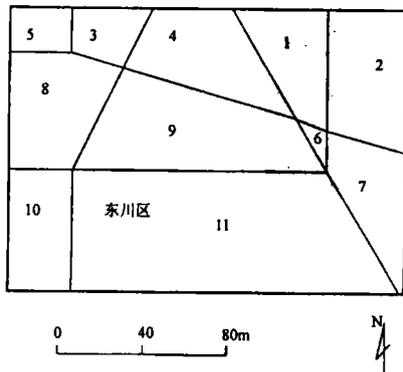


图3 小江流域均一条件单元划分示例  
Fig.3 Unique condition unit classification

例大大减少,而危险性级别较高的5和6类分区中的比例达到将近80%,这证明了多元回归模型对于小江流域的稳定性分析结果具有较高的准确性。

小江流域的不同区域的滑坡稳定性的分布规律有着明显的差异。总体来讲,小江上游地区的稳定性较好,危险性级别较高的不稳定区和极不稳定区分布范围非常小。流域中稳定性较差的地段位于小江流域的中下流地区,尤其汤丹至江口一带较为严重。位于靠近江口的区域,不稳定区域主要沿小江干流分布,左侧(西侧)的支流两侧的亦有不同程度的不稳定区分布,为小江左侧较发育的泥石流沟提供了物质来源。拖布卡至汤丹一带是小江流域的滑坡稳定性最差的区域。不稳定区和极不稳定区广泛分布,特别各支流水系的两则斜坡。

### 5 结论

CF多元回归模型是基于数据的数学统计模型,它的准确程度取决于研究区获得的数据的数量和精度,以及关键因子的选择和量化方法等多个因素。在某个地区获得的预测模型在另一地区不一定适用。通过将滑坡确定性稳定系数与回归模型的融合,可以一定程度以解决滑坡评价过程中影响因子的选择和量化的问题,有利于进行滑坡的因子敏感性分析。根据关键影响因子所建立的回归模型能够准确地进行滑坡空间分布及稳定性分析。

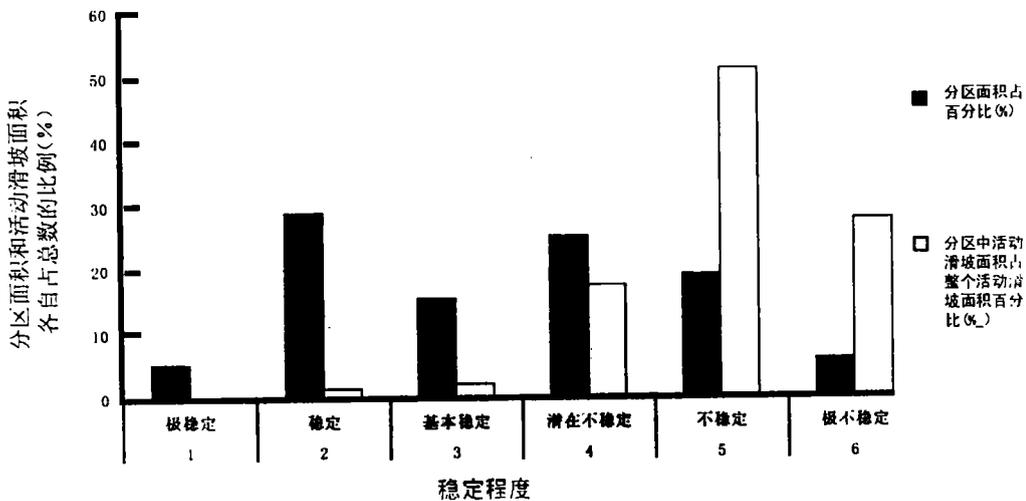


图5 统计分析结果  
Fig.5 Result of statistical analysis

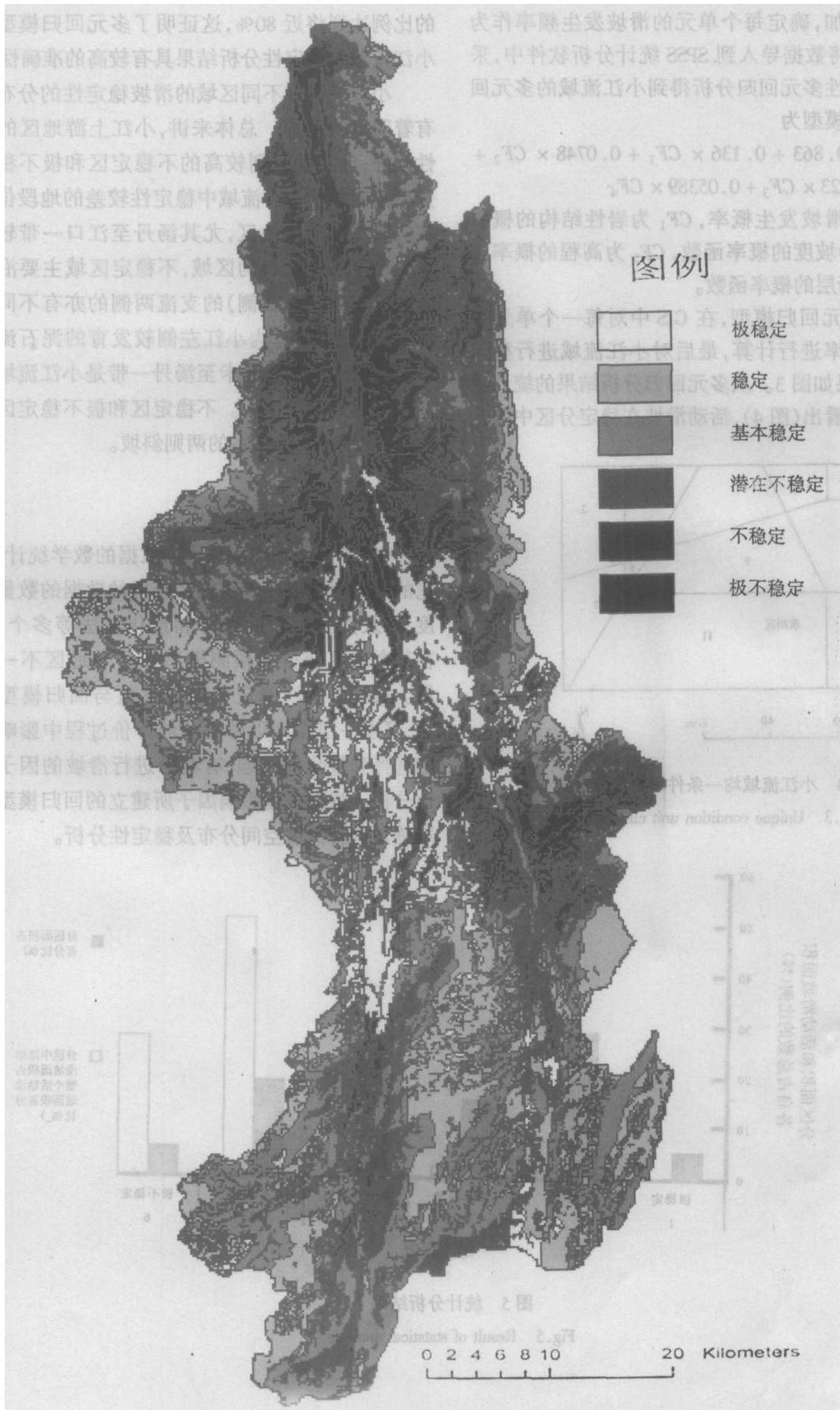


图4 小江流域滑坡稳定性多元回归统计分析划分结果

Fig.4 Hazard zoning by CF-regression model in Xiaojiang

致谢:感谢中国科学院成都山地所提供的小江流域的数字高程模型数据

### 参考文献:

- [1] 秦四清,张倬元,王士天,等.滑坡非线性时间预报理论[A].见:工程地质研究进展[C].成都:西南交通大学出版社,1993,171~181.
- [2] 周萃英,晏同珍.滑坡灾害系统非线性动力学研究[J].长春地质学院学报,1995,25(3):311~316.
- [3] Carrara A., Guzzetti F. Use of GIS technology in the prediction and Monitoring of landslide hazard[J]. *Natural Hazards*, 1999,20: 117~135.
- [4] Baeza C., Corominas J. Assessment of shallow landslide susceptibility by means of statistical techniques[A]. In: Proc VII Int. Symp. on landslides, Trondheim[C], 1996, 1:153~158.
- [5] Carrara A. Multivariate methods for landslide hazard evaluation[J]. *Mathematical Geology*, 1983, 15: 403~426.
- [6] 文宝萍.滑坡预测预报研究现状与发展趋势[J].地学前缘,1996,3(1~2):86~92.
- [7] Fell R. Landslide risk assessment and acceptable risk[J]. *Journal of Canadian Geotechnique*, 1994, 31. 261~272.
- [8] 戴福初,李军.地理信息系统在滑坡灾害研究中的应用[J].地质科技情报,2000,19(1):91~96.
- [9] Heckerman. Probabilistic interpretation of MYCIN's certainty factors [A]. In: L. N. Kanal and J. F. Lemmer(eds). *Uncertainty in Artificial Intelligence*[C]. New York, Elsevier, 1986.298~311.
- [10] 杜榕桓,康志成,陈循谦,等.云南小江泥石流综合考察与防治规划研究[M].重庆:科学技术文献出版社重庆分社.1987,26~50.

## GIS Based Landslide CF Multi-variable Regression Model and Its Application

LAN Heng-xing<sup>1</sup>, WU Fa-quan<sup>2</sup>, and WANG Si-jing<sup>2</sup>

(1. LREIS, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing, 100101 China;

2. Institute of Geology and Geophysics, CAS, Beijing, 100029 China)

**Abstract:** GIS based CF multi-variable regression model is put forward in this paper, since each landslide hazard analysis method has its own disadvantage. By the mean of integration of landslide certainty factor (CF) and multi-variable regression model, some problems can be solved including the selection and quantification of landslide factors. Landslide spatial distribution and stability analysis are carried out in Yunnan Xiaojiang watershed using the CF multi-variable regression model.

**Key words:** landslide certainty factor(CF); multi-variable regression model; geographical information system (GIS)