

GIS 支持下三峡库区秭归县滑坡灾害 空间预测

彭 令, 牛瑞卿*, 陈丽霞

(中国地质大学地球物理与空间信息学院, 武汉 430074)

摘要: 基于 GIS 空间分析和统计模型相结合进行区域评价与空间预测是滑坡灾害研究的重要方向之一。以三峡库区秭归县为研究区, 选择坡度、坡向、边坡结构、工程岩组、排水系统、土地利用和公路开挖作为评价因子。为提高模型的预测精度、可信度和推广能力, 利用窗口采样规则降低训练样本之间的空间相关性。建立 Logistic 回归模型, 对滑坡灾害与评价因子进行定量相关性分析。计算研究区滑坡灾害易发性指数, 对其进行聚类分析, 绘制滑坡易发性分区图, 其中高、中易发区占整个研究区面积的 38.9%, 主要分布在人类工程活动频繁和靠近排水系统的区域。经过验证, 该模型的预测精度达到 77.57%。

关键词: 滑坡; 空间预测; Logistic 回归; 地理信息系统; 三峡库区

文章编号: 1000-0585(2010)10-1889-10

1 引言

滑坡是世界上十大自然灾害之一, 是仅次于地震和洪水的一种严重地质灾害, 滑坡灾害预测预报对于防灾减灾具有十分重要的指导意义^[1,2]。

滑坡灾害空间预测的理论基础是工程地质类比法, 纵观其发展历程, 滑坡灾害空间预测经历了从定性——半定量——定量的过程^[3]。定性研究方法主要依靠专家知识, 通过分析历史滑坡的地质及地貌特征, 进而预测其他具有相似地质及地貌特征的区域。然而, 随着一些定性研究方法与层次、加权等思想的结合, 引入了半定量的研究方法, 意大利学者 Barredol 应用层次分析法 (AHP)^[4]、日本学者 Ayalew 应用权重线性组合法 (WLC)^[5] 进行过滑坡灾害空间评价预测。由于这类方法都依赖于专家经验知识, 具有一定的主观性。因此, 定性及半定量研究方法通常适用于地域性研究^[6,7]。定量方法是在进行定性分析的基础上, 建立预测对象的地质模型, 通过合理假设或简化, 将复杂的研究对象抽象成可以求解的数学模型, 进而选取合理参数, 进行预测计算, 获得评价预测结果。目前的数学模型法可以分为两大类型: 确定性分析方法和非确定性统计分析方法^[3,8], 前者是一种基于物理确定模型和工程地质的方法, 建立在滑坡发生的物理机制基础上, 这种确定性模型需

收稿日期: 2010-02-05; 修订日期: 2010-05-13

基金项目: 国家自然科学基金项目 (40672205); 国家高技术研究发展计划 (2007AA12Z100); 国土资源部三峡库区三期地质灾害防治重大科学研究项目 (SXKY3-2-2)

作者简介: 彭令 (1984), 男, 重庆人, 博士研究生。主要研究方向为地学信息分析、空间知识发现。E-mail: wuhanpl@gmail.com。

* 通讯作者: 牛瑞卿 (1969), 男, 博士, 副教授。主要研究方向为天空地一体化地球观测信息融合与可视化、人类工程活动与岩石体变化遥感监测技术、遥感信息定量化和反演。

要收集工程地质、水文地质等方面的大量数据,因此比较适合特定滑坡的监测、评价和预测,对于大范围的滑坡评价和预测来说很难实现;后者是基于统计分析理论,不需要收集大量有关滑坡物理特征方面的数据,而是通过统计方法分析历史滑坡与滑坡影响因子之间的关系来预测区域上或某段斜坡将来产生滑坡的可能性。统计分析方法中绝大部分利用判别式分析法、神经网络法和逻辑斯谛克(Logistic)回归方法,意大利学者 Carrara 利用逐步式判别法对意大利的斜坡稳定性进行过分类评价^[9];韩国学者 Lee 等利用神经网络法对韩国 Yongin 市的滑坡进行过空间评价预测^[10]。由于判别分析要求变量为连续变量且服从正态分布,而在滑坡空间预测研究中,许多变量为类型变量,如边坡结构等,因此从严格数学意义上判别分析难以满足要求;而神经网络的预测能力与隐含层的层数和各隐含层的节点数有关,因此其性能很大程度上取决于操作者的水平^[10]。为了克服上述问题,Logistic 回归方法被应用到滑坡灾害空间预测研究领域并被广泛接受,如国内外学者 Dai 等^[11]、陈丽霞等^[12]、李雪平等^[13]、Bai 等^[14]、Lee 等^[15]、Ohlmache 等^[16]和 Lulseged 等^[17]利用 Logistic 回归模型进行过滑坡灾害空间评价预测。但是基于统计模型的预测结果均受到训练样本影响,模型预测精度依赖于训练样本数量,而具有空间属性的样本点之间的空间相关性影响模型的可信度及推广能力,研究者面临采集最大数量样本与减少样本空间相关性之间难以达到平衡的问题。因此,利用窗口采用规则,在采集最大数量样本点的同时降低其空间相关性,建立更加稳健的预测模型,并提高模型预测精度,对滑坡灾害空间预测具有重要意义。

本文以三峡库区秭归县境内为研究区,利用窗口采用规则,建立 Logistic 回归模型,分析斜坡稳定性与影响因子之间的定量关系,绘制滑坡灾害空间易发性分区图,为三峡库区滑坡灾害空间预测提供方法支持。

2 Logistic 回归模型

Logistic 回归模型是二分类因变量常用的统计分析方法(因变量 y 只取两个值 1 和 0 或者是和否)。假设用 P 表示滑坡事件出现成功结果(发生滑坡,因变量 y 值为 1)的概率, Q 表示滑坡事件失败(不发生滑坡,因变量 y 值为 0)的概率, x_1 、 x_2 、 \dots 、 x_n 表示对滑坡影响的 n 个因素,用 Logistic 回归公式表示出现成功和失败事件结果的概率分别为^[18]:

$$P = \frac{e^{a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n}}{1 + e^{a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n}} \quad (1)$$

$$Q = \frac{1}{1 + e^{a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n}} \quad (2)$$

通过上式,得到出现某种事件的概率与影响因素之间呈曲线关系。由公式(1)除以公式(2)然后取自然对数有:

$$\ln \left(\frac{P}{Q} \right) = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \quad (3)$$

式中, a 为常数项, b_1 、 b_2 、 \dots 、 b_n 称为 Logistic 回归系数。因此,可以利用 Logistic 回归分析方法建立定量预测的数学模型,概率 P 值作为滑坡预测指数(Landslide Predict Index, LPI)来描述影响因子 x_1 、 x_2 、 \dots 、 x_n 对滑坡的敏感程度或者滑坡发生的可能性大小,从而预测未来滑坡的空间分布特征。

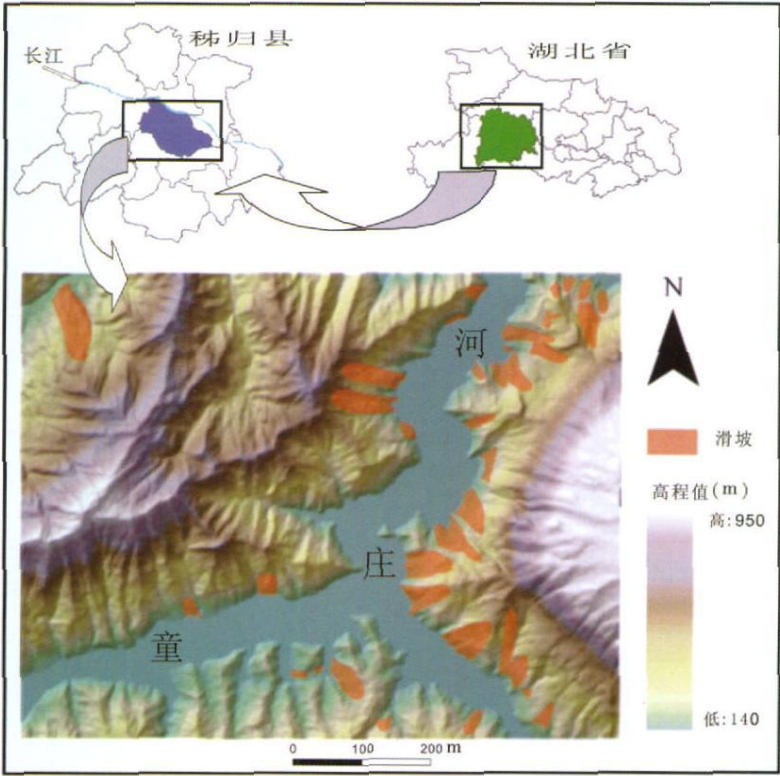


图 1 研究区位置示意图 (叠加历史滑坡分布)
Fig 1 Location of the study area (landslide)

3 研究区概况和数据来源

3.1 研究区

研究区位于三峡库区秭归县境内, 地理坐标为 $110^{\circ}41' \sim 110^{\circ}45' E$ 、 $30^{\circ}55' \sim 30^{\circ}57' N$, 面积约为 $0.27 km^2$ 。图 1 为研究区地理位置示意图 (叠加历史滑坡分布)。该区为鄂西褶皱, 黄陵背斜向秭归向斜转换部位, 为中低山侵蚀地貌。气候属亚热带大陆性季风性气候, 具有四季分明, 雨量充沛, 光照充足, 气候温和等特点^[19]。地质灾害有滑坡、崩塌和塌岸等, 滑坡灾害频繁发生, 造成大量人员伤亡和经济损失, 如近年来区内共发生 30 多起滑坡变形, 危及数百人安全, 经济损失达百万元。

3.2 数据源

本文采用的主要数据源包括: (1) 中巴资源 (CBERS) 卫星数据 1 景 (2004 年 4 月, 轨道号为 04/65), 用于土地利用信息的遥感解译; (2) 1:1 万比例尺地形图, 主要用于提取地形、地貌、公路和水系等相关信息; (3) 1:5 万比例尺地质图, 用于提取地质和工程岩组等信息; (4) 除此之外还有该区历史滑坡存档资料和滑坡野外调查资料及部分航片等, 用于历史滑坡解译和空间定位。研究区内共解译出土质滑坡 37 处, 总面积为 $17132 m^2$, 约占整个研究区面积 6.26%, 其中最小滑坡面积为 $58 m^2$, 最大面积为 $1874 m^2$ 。

4 研究方法

针对滑坡空间预测问题，首先在收集数据资料基础上选择合理的评价预测指标；然后针对预测模型划分适当的模型单元；接着利用采样规则选取训练样本数据集，建立 Logistic 回归预测模型，计算滑坡预测指数，并绘制滑坡易发性分区图；最后对预测模型及预测精度进行检验，其详细技术流程见图 2 所示。

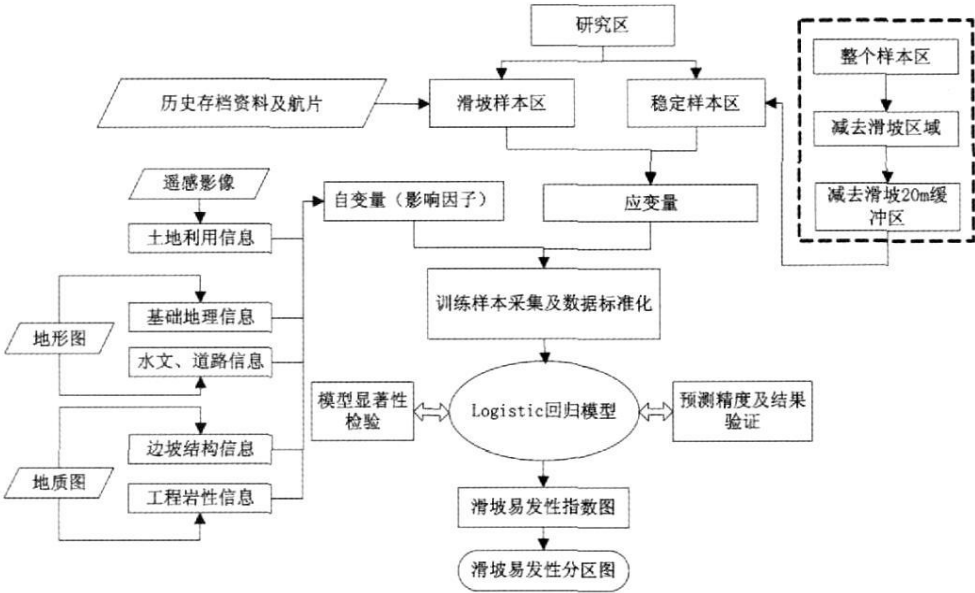


图 2 基于 Logistic 回归模型的滑坡空间预测技术流程图

Fig 2 Procedure of landslide prediction based on Logistic regression model

4 1 评价因子选择

滑坡空间预测需要考虑多种影响因子的综合作用，特别是利用 Logistic 回归模型。理论上，选择的影响因子越多，模型预测精度越高，但是相应增加了模型的复杂度和计算量，同时也会产生回归模型所忌讳的自变量多重共线性。因此，文中只选择对滑坡起重要作用的影响因子，建立既合理又简单有效的回归模型。本文根据研究区的具体特点和前人研究成果^[17]，选取坡度、坡向、边坡结构、工程岩组、排水系统、土地利用和公路开挖作为模型评价因子。滑坡除了受上述因素影响外，还会受降雨等外部因素影响，由于研究区缺乏该类数据，故未纳入评价因子范畴。排水系统和公路开挖对斜坡稳定性都会产生一定影响，一般距离此类因素越近，其对斜坡稳定性影响越大。文中对水系和公路图层作表面距离分析，并把距离值作为自变量。根据统计滑坡体在其他分类因子中的密度分布状况，把分类评价因子划分成不同的状态变量，并将各种变量状态进行编码（表 1）。

4 2 模型单元划分

选取合适的模型单元是滑坡灾害空间预测的基础，对预测结果具有重要影响。本文采用对应于 GIS 中栅格数据结构的网格单元，是滑坡空间预测中的一种常用模型单元，其优点是单元划分简单快捷、模型计算速度快、样本点采集方便，但是同时也存在划分单元与评价因子（如地形）之间的相关性较差的问题^[20]。

表 1 滑坡灾害易发性预测因子变量表

Tab 1 Summary of prediction variables used in the logistic regression model					
因子	变量类型	编码	因子	变量类型	编码
地形坡度	分类变量		地形坡向	分类变量	
0°~5°	哑变量	1	东北	哑变量	1
6°~15°	哑变量	2	正东	哑变量	2
16°~25°	哑变量	3	东南	哑变量	3
26°~35°	哑变量	4	正南	哑变量	4
≥36°	参考变量	5	西南	哑变量	5
边坡结构	分类变量		正西	哑变量	6
顺向坡	哑变量	1	西北	哑变量	7
顺斜坡	哑变量	2	正北	参考变量	8
横向坡	哑变量	3	土地利用	分类变量	
逆斜坡	哑变量	4	建筑用地	哑变量	1
逆向坡	参考变量	5	耕地	哑变量	2
工程岩组	分类变量		林地	参考变量	3
软质岩	哑变量	1	排水系统	连续变量	1
软硬相间质岩	哑变量	2	公路开挖	连续变量	1
硬质岩	参考变量	3			

数据尺度对因子量化具有较大影响，产生的误差在计算过程中逐步传递，并最终影响预测结果，文中将所有评价数据归整为 1：1 万比例尺图层，以保证数据尺度的一致性，减弱数据尺度对预测结果的影响。栅格单元大小对滑坡评价和预测影响很大，直接影响到评价因子的表达。李军等对不同比例尺下栅格单元大小选择进行过研究，提出了计算栅格单元的经验公式^[21]：

$$G_s = 7.49 + 0.0006S - 2.0 \times 10^{-9}S^2 + 2.9 \times 10^{-15}S^3 \tag{4}$$

式中， G_s 为适宜网格大小， S 为原始等高线数据精度的分母。由公式（4）计算出最佳网格单元为 13.29m，文中将所有数据集中到 ArcGIS 软件中，利用 GIS 软件的栅格分析功能，将所有数据转换和重采样为 10m×10m 的网格单元。

4 3 样本集选择

对具有空间属性的数据做统计分析，训练样本需要相互独立，即不存在空间相关性，然而实际上采集的样本并不完全满足这一条件。为提高模型的预测精度、可信度和推广能力，探寻一种采样方法降低样本点之间的空间相关性，对滑坡空间预测具有重要意义^[22,23]。理论上，空间相关性会随着空间距离的增大而减小。本文采用一种窗口采样规则，其思路为：在样本区不重叠的移动采样窗口，保留采样窗口中心的数据点作为样本点，其中窗口的大小分别为 1×1（10m×10m）、3×3（30m×30m）、5×5（50m×50m）和 7×7（70m×70m）

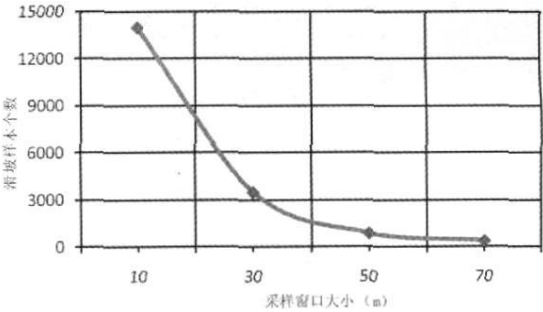


图 3 窗口大小与滑坡样本统计图

Fig. 3 Number of samples in terms of window size

为 1×1（10m×10m）、3×3（30m×30m）、5×5（50m×50m）和 7×7（70m×70m）

等。当采样窗口太小时，具有充足的训练样本点，但是不足以去除样本点之间的空间相关性；当采样窗口太大时，样本点之间的空间相关性降低，但是同时也减少了所能采集的样本点数目，相应丢掉了部分有用信息，相应样本数据集也不满足渐进正态分布。通过统计窗口大小与能采集到的滑坡样本数目（图 3）。最终选择 3×3 大小的采样窗口，即采样的空间距离为 30m，提取样本点作为训练数据集。

对于训练样本数量，一般是采集相同比例的滑坡样本（应变量为 1）与未滑坡样本（应变量为 0）作为训练样本集^[11, 24]。本文利用 3×3 大小的采样窗口在研究区共采集到 3488 个滑坡样本点，约占整个研究区总面积的 1.5%；为消除不同比例训练样本类型对模型的影响，利用相同大小窗口在滑坡 20m 缓冲区范围以外采集相同数量的未滑坡样本点，构成训练样本数据集。

4 4 预测模型分析

在进行 Logistic 回归前，为消除不同属性数据对预测结果的影响，需进行数据的标准化^[17]。文中将连续性因子值标准化到 0~ 1 之间，利用上面选择的训练样本集，在统计软件 SPSS 中进行 Logistic 回归分析，获得回归系数 a 、 b_1 、 b_2 、... b_n ，见表 2。经卡方（Chi-square）检验， P （表中的显著性水平 Sig）值< 0.05（给定的显著性水平），表明各回归系数具有统计学意义。由各评价因子的回归系数值大小不难发现，导致滑坡灾害发生的主要因素依次为土地利用、工程岩组、 $0^\circ \sim 25^\circ$ 的坡度、排水系统，而土地利用的回归系数最大，说明人类工程活动对研究区的垦殖和破坏，急剧加速了滑坡孕灾环境的变异。

表 2 滑坡灾害预测变量的 Logistic 回归系数
Tab 2 The Logistic regression coefficient of each variable

影响因子	回归系数	显著性 (Sig)	影响因子	回归系数	显著性 (Sig)
地形坡度		0.000	地形坡向		0.000
$0^\circ \sim 5^\circ$	1.476	0.000	东北	0.960	0.000
$6^\circ \sim 15^\circ$	2.093	0.000	正东	1.183	0.000
$16^\circ \sim 25^\circ$	1.419	0.000	东南	0.856	0.000
$26^\circ \sim 35^\circ$	0.766	0.000	正南	0.857	0.000
边坡结构		0.000	西南	0.332	0.015
顺向坡	1.289	0.000	正西	0.885	0.000
顺斜坡	-1.571	0.000	西北	0.603	0.000
横向坡	-0.347	0.025	土地利用		0.000
逆斜坡	0.303	0.030	建筑用地	4.227	0.000
工程岩组		0.000	耕地	2.991	0.000
软质岩	2.831	0.000	排水系统	1.868	0.000
软硬相间质岩	2.027	0.000	公路开挖	0.351	0.044
常数项 (a)	-8.885	0.000			

5 结果分析与验证

5 1 预测结果

利用公式（1）计算得到滑坡发生的概率 P 值（LPI），并绘制滑坡灾害空间易发性指数图（图 4A）。理论上，LPI 值为 0~ 1 之间，其值越接近 1，则表明发生滑坡灾害的概率

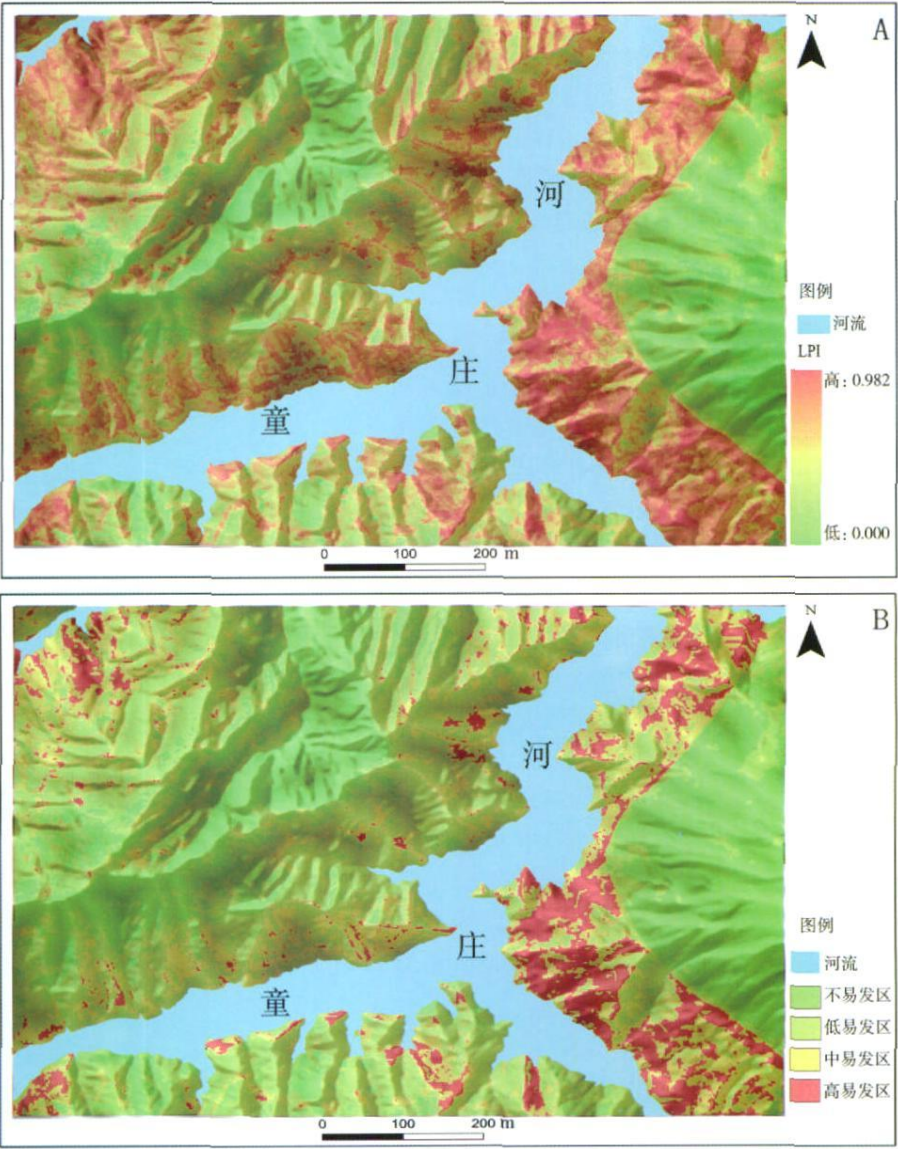


图 4 基于 Logistic 回归模型的滑坡预测图 (A: 易发性指数图; B: 易发性分区图)

Fig 4 The map obtained after the logistic regression analyses (A: probability map; B: susceptibility map)

越大。对 LPI 值进行聚类分析后，得到研究区滑坡易发性分区图 (图 4B)，并分为四个等级：高易发区、中易发区、低易发区和不易发区 (表 3)，其中高、中易发区约占整个研究区面积的 38.9%，主要分布在人类工程活动频繁和靠近排水系统的区域。

5.2 预测模型检验

对于 Logistic 回归模型，似然比模型卡

表 3 滑坡灾害易发性空间预测分类

Tab 3 Spatial prediction degree of landslide hazard

概率 P 值	易发区等级	所占比例 (%)
$P > 0.8$	高易发区	18.5
$0.8 \geq P > 0.6$	中易发区	20.4
$0.6 \geq P > 0.2$	低易发区	40.8
$P \leq 0.2$	不易发区	20.3

方值 (Model chi-square) 越大, 表明所选择的评价因子与预测事件具有越强的相关性。Hosmer-Lemeshow 卡方值 (Goodness of fit) 是评价模型拟合优度的重要指标, 通过计算因变量的观测值与预测值之差而得到该值, 其值越小表示模型拟合越好。Pseudo R^2 值指示模型拟合数据集的优良程度^[25], 通过 $1 - (\ln L / \ln L_0)$ 计算得到, 当值为 1 时表明模型完全拟合数据集, 值为 0 时表明模型完全不能拟合数据集, 当值大于 0.2 时表明模型拟合数据良好^[26]。受试者工作特征曲线下面积 (AUC) 反映预测模型的准确性。理论上, 这一指标取值范围为 0.5~1, 当值为 1 时表示预测完全正确, 值为 0.5 时表示预测完全无价值^[17]。

由表 4 得到, Model chi-square 值较大, 而 Goodness of fit 值较小, Pseudo R^2 为 0.282, 该值大于 0.2, 以及 AUC 为 0.837, 该值接近于 1。综合上述各项指标值, 表明预测模型对研究区数据拟合效果较佳, 预测结果可信。

5.3 预测精度分析

为了定量地评价模型预测精度, 引入成功率验证法^[8, 27, 28], 通过滑坡预测结果和历史滑坡进行比较, 定量描述预测精度。文中将预测指数 LPI 归一化到 0~100 之间, 按照降序进行排列, 然后等分成 100 个单位, 分别计算各个单位内滑坡发生的百分比, 得到滑坡发生累计频率 (%)—预测指数 (%) 曲线图 (图 5)。从图中可以得到, 对于 60%~100% (X 轴 40% 处) 的预测指数, 模型可以预测 77.57% 的滑坡发生在高、中易发区中, 表明模型具有较高的预测精度, 可用于滑坡灾害空间预测分析。

6 结论

利用 GIS 空间分析和统计模型相结合, 定量预测未来滑坡的空间分布, 已经成为滑坡灾害预测研究中的一个重要方向之一。

本研究利用窗口采样规则以降低训练样本之间的空间相关性, 据此而建立的回归模型, 其预测精度、可信度和推广能力都得到提高。在滑坡空间易发性分区图中, 整个研究区面积为 273600m², 其中高、中易发区面积为 106469m², 约占研究区面积的 38.9%, 主要分布在人类工程活动频繁和靠近排水系统的区域; 模型回归系数反映在研究区内土地利用对滑坡灾害影响最为突出。综合表明人类工程活动对研究区域斜坡稳定性具有极大影响。通过模型检验和精度验证, 结果表明预测模型具有较高的可信度和精度, 可用于滑坡灾害空间预测分析。

表 4 模型检验统计表
Tab 4 Summary statistics of the logistic regression model

验证指标	值
$-2\ln L_0$	9670.789
$-2\ln L$	6946.768
Model chi-square	2724.021
Goodness of fit	15.342
Pseudo R^2	0.282
AUC	0.837

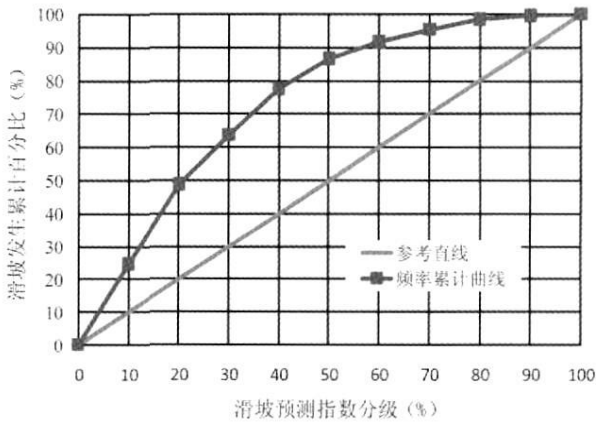


图 5 滑坡发生累计频率与预测指数关系

Fig 5 The relation between landslide accumulative frequency and prediction index

参考文献:

- [1] 陈剑, 杨志法, 刘衢秋. 滑坡的易滑度分区及概率预报模式. 岩石力学与工程学报, 2005, 24 (13): 2392~ 2396
- [2] 戴福初, 李军. 暴雨滑坡泥石流的研究进展与趋向. 地理研究, 1998, 17 (增): 117~ 124.
- [3] 吴益平, 殷坤龙, 陈丽霞. 滑坡空间预测数学模型的对比及其应用. 地质科技情报, 2007, 26 (6): 95~ 100
- [4] Barredol J, Benavidesz A, Herhl J, *et al*. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. International Journal of Applied Earth Observation and Geoinformation, 2000, 2: 9~ 23
- [5] Ayalew L, Yamagishi H, Ugawa N. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides, 2004, 1: 73~ 81.
- [6] Soeters R, Westen C J. Slope instability recognition, analysis, and zonation. In: Turner K A, Schuster R L. Landslides: Investigation and mitigation. Washington, USA: Transport Research Board, 1996. 129~ 177.
- [7] Guzzetti F, Carrara A, Cardinali M, *et al*. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central Italy. Geomorphology, 1999, 31, 181~ 216
- [8] Aleotti P, Chowdhury R. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the Environment, 1999, 58: 21~ 44
- [9] Carrara A, Crosta G, Frattini P. Geomorphological and historical data in assessing landslide hazard. Earth Surface Processes and Landforms, 2003, 28: 1125~ 1142
- [10] Lee S, Ryu J H, Min K, *et al*. Landslide susceptibility analysis using GIS and artificial neural network. Earth Surface Processes and Landforms, 2003, 28: 1361~ 1376
- [11] Dai F C, Lee C F. Landslide characteristics and slope instability modeling using GIS Lantau Island, Hong Kong. Geomorphology, 2002, 42: 213~ 238
- [12] 陈丽霞, 殷坤龙, 张桂荣. 浙江省磐安县滑坡灾害空间预测. 地球与环境, 2005, 33 (增刊): 359~ 363
- [13] 李雪平, 唐辉明, 陈实. 基于 GIS 的 Logistic 回归在区域滑坡空间预测中的应用. 公路交通科技, 2005, 22 (6): 152~ 155
- [14] Bai S B, Wang Jian, Lu G N, *et al*. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. Geomorphology, 2010, 115: 23~ 31.
- [15] Lee S, Min K. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, 2001, 40: 1095~ 1113
- [16] Ohlmacher C G, Davis C J. Using multiple regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering Geology, 2003, 69: 331~ 343
- [17] Lulseged Ayalew, Hiromitsu Yamagishi. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology, 2005, 65: 15~ 31.
- [18] 王济川, 郭志刚. Logistic 回归模型——方法与应用. 北京: 高等教育出版社, 2001.
- [19] 胡立山, 冷坚, 王锡璠, 等. 1: 5 万新滩 (西)、过河口 (西) 幅区域地质调查报告. 湖北: 湖北省地质矿产局, 1995
- [20] 兰恒星, 王苓涓, 周成虎. 地理信息系统支持下的滑坡灾害分析模型研究. 工程地质学报, 2002, 10 (4): 421~ 427
- [21] 李军, 周成虎. 基于栅格 GIS 滑坡风险评价方法中网格大小选取分析. 遥感学报, 2003, 7 (2): 86~ 92
- [22] Bo Huang, Li Zhang, Bo Wu. Spatiotemporal analysis of rural-urban land conversion. International Journal of Geographical Information Science, 2009, 23 (3): 379~ 398
- [23] Getis A, Griffith D A. Comparative spatial filtering in regression analysis. Geographical Analysis, 2002, 34 (2): 130~ 140
- [24] Yesilnacar E, Topal T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Engineering Geology, 2005, 79: 251~ 266
- [25] Menard S. Applied logistic regression analysis. Sage University Paper Series on Quantitative applications in Social Sciences, vol. 106. California: Sage Publications, 1995. 98
- [26] Clark W A, Hosking P L. Statistical Methods for Geographers. New York: John Wiley and Sons, 1986. 518
- [27] 胡德勇, 赵文吉, 李小娟, 等. 不完备样本条件下基于支持向量回归模型的滑坡易发性评价. 地理研究, 2008, 27 (4): 755~ 762

- [28] Chung C F, Fabbri A G. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering & Remote Sensing*, 1999, 65 (12): 1388~ 1399

Landslide hazard spatial prediction in Zigui County of the Three Gorges Reservoir Area based on GIS

PENG Ling¹, NIU Rui-qing¹, CHEN Li-xia¹

(1. Institute of Geophysics and Geomatics, China University of GeoSciences, Wuhan 430074, China)

Abstract: Landslide prediction is very important in disaster prevention and reduction procedures, and it is one of practical research fields to evaluate and predict landslide hazards using statistic analysis model and spatial analysis of GIS. The aim of this study is to analyze landslide susceptibility using Logistic regression model in Zigui County of the Three Gorges Reservoir Area. In this paper, seven evaluation factors are selected, i. e. topographic slope, topographic aspect, bed rock-slope relationship, lithology, land use and distance from road and drainage. In susceptibility mapping, the use of logistic regression is to find the best fitting function to describe the relationship between the presence or absence of landslides (dependent variable) and a set of evaluation factors such as topographic slope and lithology. Here, an inventory map concerning 37 landslides was used to produce a variable, which takes a value of 1 for the presence and 0 for the absence of slope failures. In order to improve the accuracy and credibility of the model prediction, methods to reduce spatial autocorrelation in a logistic regression framework are also discussed. An optimal sampling scheme that can eliminate spatial autocorrelation whilst maintaining enough samples to achieve the accuracy based on the model is developed. The model was tested by the overall model statistics, and the results indicate that the model fits the dataset. The effect of each parameter on landslide occurrence was assessed from the corresponding coefficient that appears in the logistic regression function. The interpretation of the coefficients showed that land use plays a major role in determining landslide occurrence and distribution, although field observations showed that engineering construction exerts great influence on slope stability. With the help of a predicted probability map, the study area was classified into four categories of landslide susceptibility: high, moderate, low and none. The moderate and high susceptibility zones make up 38.9% of the total study area. In comparison to the occurrence of historical landslide hazards, the precision using logistic regression model can be up to 77.57%.

Key words: landslide; spatial prediction; logistic regression; GIS; Three Gorges Reservoir Area