

引用格式:余淙蔚,柳侃,殷杰,余斌.一种适用于逻辑回归模型评价浅层滑坡易发性的网格尺度划分方法——以2019年福建省三明市群发浅层滑坡为例[J].山地学报,2022,40(1):106-119.

YU Congwei, LIU Kan, YIN Jie, YU Bin. A grid-scale division method applicable to logistic regression models for evaluating the susceptibility of shallow landslides——taking the 2019 cluster of shallow landslides in Sanming, Fujian as example [J]. Mountain Research, 2022, 40(1):106-119.

一种适用于逻辑回归模型评价浅层滑坡 易发性的网格尺度划分方法 ——以2019年福建省三明市群发浅层滑坡为例

余淙蔚¹, 柳侃², 殷杰^{1*}, 余斌^{3*}

(1. 华东师范大学 地理科学学院, 上海 200241; 2. 自然资源部丘陵山地地质灾害防治重点实验室, 福州 350002;
3. 成都理工大学 地质灾害防治与地质环境保护国家重点实验室, 成都 610059)

摘 要: 滑坡易发性评价是滑坡防治研究的重要内容之一。在逻辑回归模型中, 常用网格单元作为滑坡易发性评价单元, 具有划分简单, 回归效率高的优点。然而, 在构建模型时直接将研究区的栅格数据转换为矩阵, 每个栅格表示一个评价单元, 存在滑坡样本不平衡, 滑坡地形参数指标不能反映真实的滑坡地形等问题, 造成评价效果统计显著性低。本文以2019年福建省三明市的群发性浅层滑坡为研究对象, 借助DEM数据和航拍影像整理出131个滑坡的面积、高程、岩性和道路建筑邻近性, 采用改进高程矩阵, 获得了更真实的滑坡坡度、剖面曲率和平面曲率等地形参数。根据各影响因子的标准滑坡频率对滑坡的贡献分级, 建立了不同网格尺度的滑坡易发性逻辑回归模型。比较了假设检验中 p 值大小以及 $100\text{ m} \times 100\text{ m}$ 、 $200\text{ m} \times 200\text{ m}$ 、 $500\text{ m} \times 500\text{ m}$ 和 $1000\text{ m} \times 1000\text{ m}$ 等不同网格尺度的滑坡易发性评价结果, 确定以 $500\text{ m} \times 500\text{ m}$ 网格为基本单位的逻辑回归模型评价效果最佳。研究表明改进后的高程矩阵计算的滑坡地形参数能够更好地描述滑坡的地形特征。在评价滑坡易发性时, 逻辑回归模型中的网格尺度大小应确保滑坡发生样本为总样本的30%~70%, 以避免被低估或高估; 总样本数应该在50以上, 以避免模型不显著。本文提出了适用于逻辑回归模型评价滑坡易发性的网格尺度划分方法, 可为滑坡的防灾减灾研究提供一定参考。

关键词: 滑坡; 易发性; 高程矩阵; 逻辑回归; 网格尺度; 福建省

中图分类号:

文献标志码: A

滑坡的易发性研究受到了国内外地质灾害防治领域越来越多的关注。由于滑坡的影响因素较多, 构建一个适用性强的滑坡易发性评价模型是目前研究的难点。随着计算机技术的发展, 数学统计模型越来越多地被引入滑坡易发性研究中, 许多研究通

过岩性、地形、水文条件、植被、人类活动等指标参数与滑坡发生与否之间的关系, 采用线性回归^[1-2]、逻辑回归^[3-5]、信息量模型^[6-7]、人工神经网络^[8-9]、支持向量机^[10]等方法评价滑坡的易发性。这些研究在滑坡的指标选取、模型评价、区划制图等方面已

收稿日期(Received date): 2021-07-28; 改回日期(Accepted date): 2021-12-20

基金项目(Foundation item): 福建省科技防灾课题(闽自然资发[2020]46号)。[Fujian Science and Technology Disaster Prevention Project(Fujian Department of Natural Resources [2020]No. 46)]

作者简介(Biography): 余淙蔚(1999-), 女, 四川成都人, 硕士研究生, 主要研究方向: GIS、地质灾害。[YU Congwei (1999-), female, born in Chengdu, Sichuan province, M. Sc. candidate, research on GIS and geological disasters] E-mail: 18701911323@163.com

* 通讯作者(Corresponding author): 殷杰(1983-), 男, 博士, 教授, 主要研究方向: 自然灾害风险评估与风险管理。[YIN Jie (1983-), male, Ph. D., professor, research on natural disaster risk assessment and risk management] E-mail: jyin@geo.ecnu.edu.cn; 余斌(1966-), 男, 博士, 教授, 主要研究方向: 泥石流灾害与防治、泥石流预报。[YU Bin (1966-), male, Ph. D., professor, research on debris flow hazard and prevention, debris flow forecasting] E-mail: yubin08@cdut.edu.cn

经取得了较好的成果,使得滑坡易发性评价的精度越来越高,但滑坡的指标参数的精度仍有待进一步提高。其中,在滑坡的地形指标参数中,很多易发性模型都用到了滑坡坡度、平面曲率和剖面曲率。获取这些参数时,需要将 3×3 的高程矩阵代入相应公式计算,但研究中均以统一边长的高程矩阵^[1-3],导致在计算小型滑坡时得到比滑坡范围更大的地形指标参数;在计算较大滑坡时,得到的只是滑坡体局部的地形指标参数;两种情况都无法获取代表滑坡的地形指标参数,导致滑坡易发性评价结果的可靠性存疑。

在上述滑坡易发性评价的数学模型中,逻辑回归模型能够反映滑坡事件的二元关系^[11],因而在滑坡易发性研究中得到了广泛的应用^[11-13]。网格单元具有划分简单、回归效率高的优点^[13]而被用作滑坡易发性评价单元。部分研究在构建模型时,直接将研究区的栅格数据转换为矩阵,每个栅格表示一个评价单元^[11-12],使得发生滑坡的样本数远小于总样本数,造成了样本不平衡,难以直接应用于评价中。为了避免这种不平衡,一些研究随机选取与滑

坡样本数相当的非滑坡样本,与滑坡样本共同构成评价样本^[13-14],但这种方法不能充分利用研究区的数据,最终评价效果也不够理想。

多山地丘陵地形的福建省属于我国滑坡最频繁的地区之一。2019 年 5—6 月,福建省三明市受强降雨影响,三明市后山多处发生群发性山体滑坡,造成重大灾害^[15]。本文以 2019 年福建省三明市群发浅层滑坡为例,根据实际滑坡的大小和方向确定高程矩阵的边长和方向,将研究区划分为 $100 \text{ m} \times 100 \text{ m}$, $200 \text{ m} \times 200 \text{ m}$, $500 \text{ m} \times 500 \text{ m}$ 和 $1000 \text{ m} \times 1000 \text{ m}$ 等不同尺度的网格,分别作为逻辑回归模型的基本单元评价滑坡易发性;对比各尺度模型的优劣与适用性,获得适用于逻辑回归模型评价滑坡易发性的网格尺度划分方法,以期为滑坡的防灾减灾研究提供参考。

1 研究数据

1.1 研究区

研究区域位于三明市内(图 1),面积约为

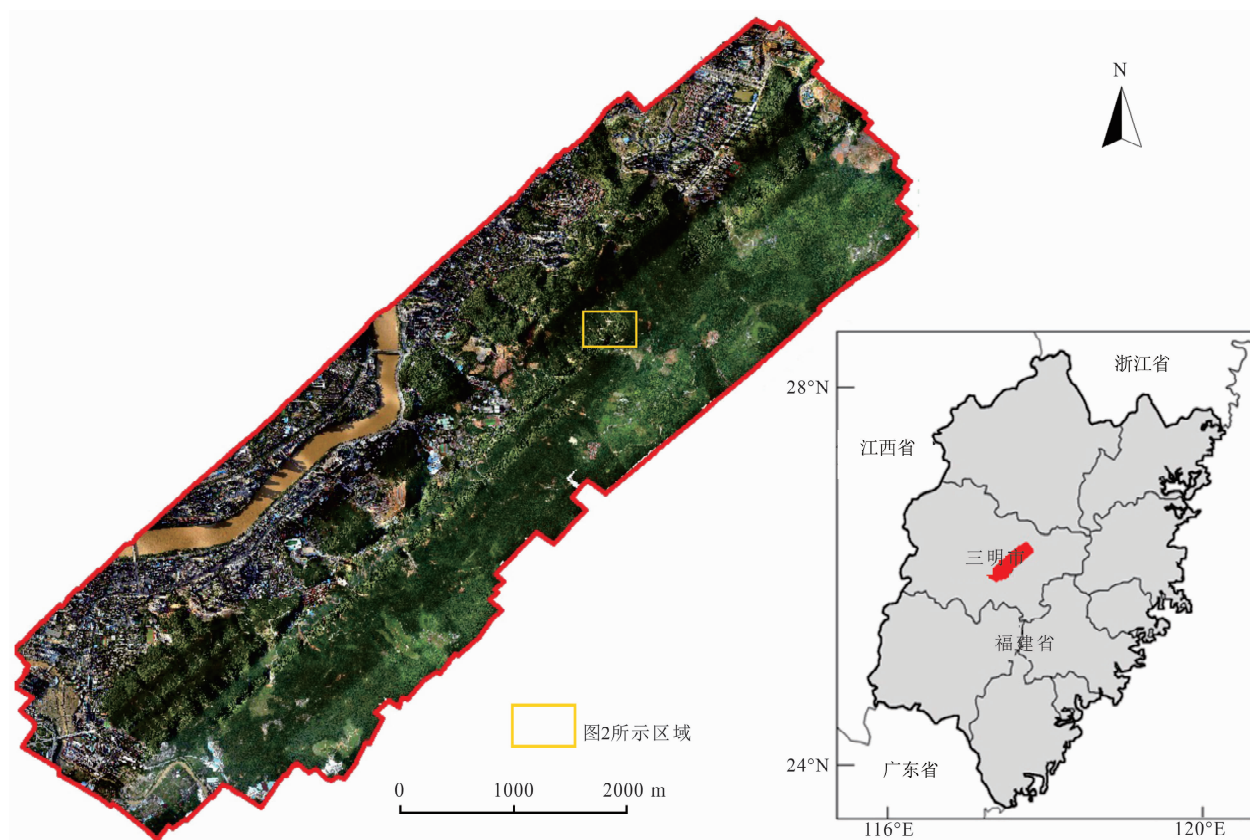


图 1 研究区位置

Fig. 1 Location of study area

27.8 km²。研究范围以三明市区后山为主,位于玳瑁山北麓,以低山及丘陵为主,总体地势东南高,西北低。山顶标高 200 ~ 500 m,最大山顶标高为 1130 m(位于情人谷顶峰),最低标高为 190 m(位于文笔山坡脚),山坡自然坡度为 20° ~ 35°。居民主要沿山脚切坡建房,加上地表水下蚀作用,导致坡脚形成高 10 ~ 30 m,坡度 45° ~ 80°的高陡边坡。高陡边坡发育主要分布于城区北侧和东侧,此处人口稠密,各类建筑物和公路密集,人类工程活动较活跃,属地质灾害高易发地带(图 2) [15]。

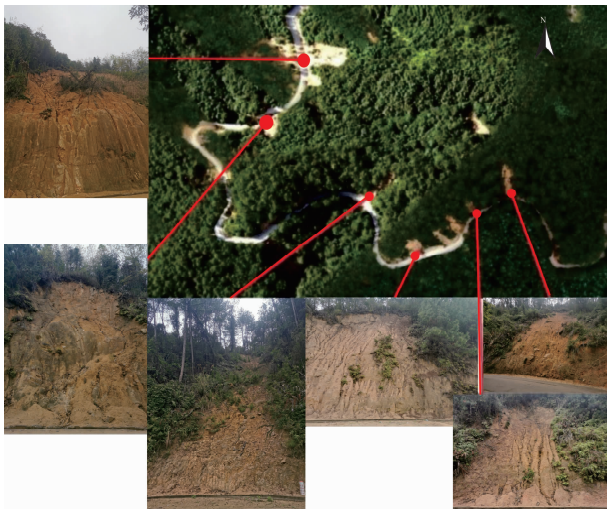


图 2 研究区局部放大图以及部分滑坡影像与实体对照

Fig. 2 Local zoomed-in image of the study area as well comparison between images and realities

三明市降雨量充沛,多年平均降雨量约 1700 mm,主要集中在 3—8 月,平均暴雨日数一般 4 ~ 6 天 [15],暴雨重现期往往也是滑坡、山洪及泥石流的再现期。

研究区岩性分布如图 3 所示 [15],其中绿圈标记的是本文研究范围,粉色标记的是穿城而过的沙溪河。研究区内主要岩性为花岗岩(橙黄色),分布于东南侧;其次在中部分布的粉砂岩(蓝色),以及沿河分布的少量砂卵石和黏土。三明市内主控断裂构造为北东向南平—永安断裂,沿沙溪河方向贯穿全境,走向 40° ~ 50°,破碎带宽度 10 ~ 200 m [15],如图 3 红线所示。

1.2 研究数据

本文所用遥感影像分辨率为 0.5 m,坐标系为 2000 国家大地坐标系;DEM 数据分辨率为 5 m,坐

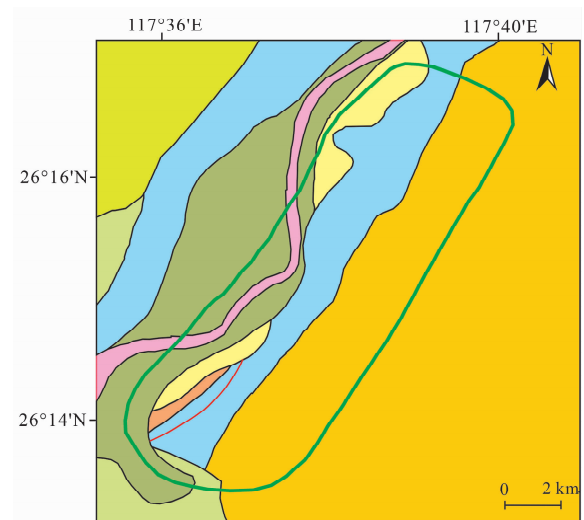


图 3 研究区地质图 [15]

Fig. 3 Geology of study area [15]

标系为 WGS84 坐标系。遥感影像拍摄于 2019 年 6 月下旬,在三明市群发性滑坡发生后不久,完整保留了滑坡发生后的形态。

对三明市 2019 年群发性滑坡后拍摄的遥感影像进行了预处理、配准、裁剪、目视解译,并前往现场对部分滑坡开展了调查,排除了一些难以辨别轮廓和尺寸太小的滑坡后,统计余下的 131 处有效滑坡样本的面积和影响因子数据。在选取滑坡影响因子时,主要参考以往研究中采用的因子(表 1)。

表 1 以往评价因子选取概况

Tab. 1 Selection of causative factors in the previous studies

影响因子	参考文献	本文是否采用
地质构造	[16 - 17]	否
断层	[3, 14, 18 - 19]	否
岩性	[3, 14, 18, 20]	是
高程	[3, 14, 17 - 19, 21]	是
坡度	[3, 14, 17 - 21]	是
曲率	[3, 17, 20 - 21]	是
年均降水量	[16, 20]	否
距河网的距离	[14, 17 - 21]	否
距道路、建筑距离	[3, 14, 18 - 21]	是
植被	[3, 16, 20]	否

由于缺乏研究区域的植被种类分布图,且遥感影像不具备红外波段,难以计算植被指数,因此本文没有选用植被因子;此外,研究区附近仅在西南方向

边缘有一个小型断裂带,研究区西侧距离较远位置有河流,因此研究区域受断裂带和水系影响较小,所以断层、地质构造、水系因子也不纳入影响因子讨论。研究区面积有限,研究范围内的年平均降水差异可以忽略不计,所以本文也未考虑降水因子。

滑动面的重力所产生的总下滑力大于总抗滑力时就会发生滑坡^[23](图 4),在确定滑坡地形时,需要考虑的范围包括上部汇水区(A_U)、潜在滑坡区(A)以及左右两侧凹陷(A_L 和 A_R)部分。当潜在滑坡的坡度在 $20^\circ \sim 45^\circ$,剖面曲率和平面曲率均为负数时,滑坡更易发生^[22]。

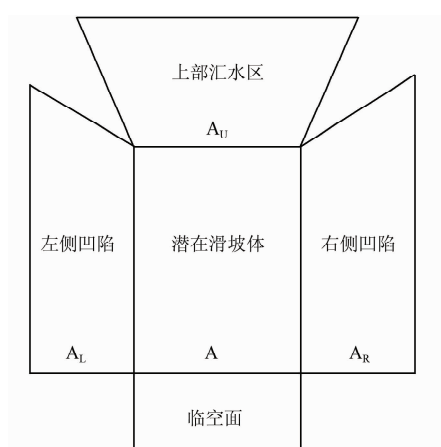


图 4 滑坡结构示意图

Fig. 4 Structure of landslide

在滑坡的易发性研究中,计算滑坡地形参数的高程矩阵边长大多为 30 m^[3,17],研究区内发生的滑坡宽度大多小于 20 m(表 2)。如果采用统一的 30 m 边长的高程矩阵计算滑坡的坡度和曲率等都无法准确地体现滑坡的地形特征。如图 5,边长为 30 m 的高程矩阵计算得到的是滑坡以外更大范围的地形参数。此外,这样计算的坡度和曲率方向是高程变化率最大的方向,只有水平和垂直两种,无法代表滑动方向非水平或垂直的滑坡,同样会影响滑坡地形数据的准确性。

表 2 本文滑坡宽度统计

Tab. 2 Width statistic of landslides in this study

滑坡宽度(w)/m	个数/个
$w \leq 10$	2
$10 < w \leq 20$	88
$w > 20$	13

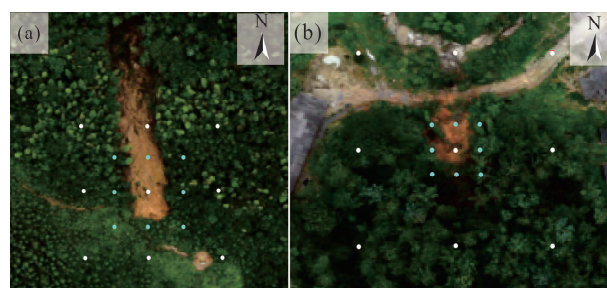


图 5 网格边长为 30 m 的高程矩阵及适应滑坡宽度的网格衡量不同尺度滑坡:

(a) 宽度约 15 m 的滑坡;(b) 宽度约为 10 m 的滑坡

Fig. 5 Elevation matrix with a cell size of 30 m compared with appropriate size evaluating landslides with the width of (a) 15 m and (b) 10 m

注:图中白点表示高程矩阵,蓝点表示适应滑坡宽度的网格。

如采用网格边长为 5 m 的矩阵作为高程矩阵,在计算宽度较大的滑坡地形参数时(图 6),地形数据就只能代表滑坡体局部的特征,而非上游汇水区和两侧所围成的对滑坡有影响的全部区域。高程矩阵的方向也同样可能与滑坡的滑动方向不符。

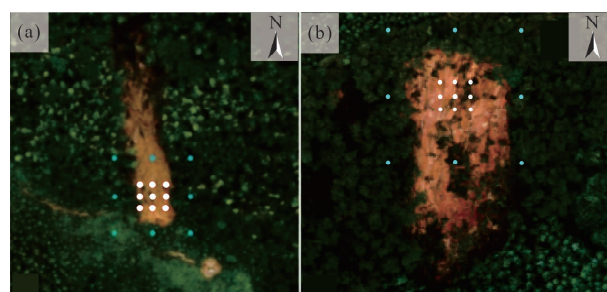


图 6 网格边长为 5 m 的高程矩阵以及适应滑坡宽度的网格衡量不同尺度滑坡:(a) 宽度约 10 m 的滑坡;

(b) 宽度约为 25 m 的滑坡

Fig. 6 Elevation matrix with cell size of 5m compared with appropriate size evaluating landslides with the width of (a) 10 m and (b) 25 m

注:图中白点表示高程矩阵,蓝点表示适应滑坡宽度的网格。

为了获取真实的滑坡地形数据,本文根据滑坡宽度计算滑坡地形参数的高程矩阵边长(图 4、图 7)。矩阵中心点(图 7e 点)需要在潜在滑坡体内(图 4 A 区域),并尽量选择居中位置点;上侧的点(图 7 b 点)需要在滑坡的上部汇水区内(图 5 A_U 区域);两侧的点(图 7d 和 f 点)在滑坡的两侧(图 4 A_R 和 A_L 区域),图 7 中的 b ~ h 的方向代表滑坡方向。如图 5 和图 6 中的蓝色点,为改进后的高程矩阵。

a	b	c
d	e	f
g	h	i

注:图中 a~i 表示高程矩阵中对应位置的高程值。

图 7 用于计算地形参数的高程矩阵

Fig. 7 Elevation matrix for topography calculating

在非滑坡(潜在滑坡)区域的地形参数计算中,因不确定潜在滑坡的边长及方向,本文均采用边长为 5 m 的高程矩阵计算其地形参数。

2 研究方法

2.1 标准滑坡频率

为了确定各滑坡影响因子与滑坡易发性关系,引入了标准滑坡频率(LF_N)方法^[17]。标准滑坡频率通过不同影响因子在单位面积内的不同等级的滑坡比例,判定各影响因子与滑坡发生的关系^[17]。标准滑坡频率的计算公式为:

$$LF_N = \frac{LS_i/C_i}{\sum_{i=1}^n LS_i/C_i} \quad (1)$$

式中, LF_N 为指定因子的标准滑坡频率; LS_i 为该因子第 i 个级别的滑坡总面积(m^2); C_i 为该因子第 i 个级别的总面积(m^2); n 为该因子的分级总数^[17]。本文将每个指定因子按照相等间距分级,统计每个等级的总面积以及该等级内所有滑坡的总面积,计算每个因子的标准滑坡频率。

2.2 逻辑回归方法

逻辑回归一般用于多变量控制的二分类问题,可以用于建立自变量和滑坡发生(因变量取 1)与不发生(因变量取 0)两个结果之间的映射关系。回归结果为 0~1,代表事件发生概率^[3]。虽然逻辑回归仍存在问题,例如对不同情境的适应能力不如决策树算法等,但是回归效率高,便于预测某事件在多个影响因子下发生的概率^[24],能够满足本文需要。逻辑回归模型中,事件发生的概率表示为

$$p(X) = \frac{\exp(\beta X)}{1 + \exp(\beta X)} \quad (2)$$

式中, $p(X)$ 为事件发生概率; X 为自变量矩阵; β 为回归系数,在二项逻辑回归中 $p(X)$ 大于 0.5 表示事件发生,小于 0.5 表示事件不发生^[23]。模型参数通过最大似然估计的方法计算,使模型能够以最大概率再现样本观测数据的估计^[23]。

本研究分别以 DEM 单点和 1000 m × 1000 m、500 m × 500 m、200 m × 200 m 和 100 m × 100 m 的网格为基本单元,利用 MATLAB 软件建立逻辑回归模型评价研究区滑坡易发性。自变量取值为 0~1。单点模型的自变量(影响因子)共分 3 个取值区间,与滑坡关联性最大的区间取 1,最小取 0,中间取 0.5;网格单元模型中,每个网格内单点数为 m ;当单点自变量处于与滑坡关联性最大的范围内时,该单点的自变量值为 $1/m$,否则为 0;网格自变量取值为所有网格内单点自变量值之和。每个网格内滑坡数量由小到大分别为 m_1, m_2, \dots, m_n ,则各网格因变量为 m_i/m_n ($i=1, 2, \dots, n$)。回归结果在 (0,1) 内,所以单点模型的回归结果以 0.5 为阈值,分为低和高 2 个易发性等级;网格模型中,逻辑回归模型计算的评估结果为 0~1,各网格的预测滑坡发生数量由此结果乘以 m_n 获得。

通过 R 语言基于数理统计方法检验评价效果。数理统计方法包括比较 p 值和绘制 ROC 曲线; p 值是假设检验中的弃真概率,当 p 值小于显著性水平(一般为 0.05 或 0.1),说明模型的拟合结果达到预期,大于显著性水平则说明模型评价效果不佳^[22]。ROC (Receiver Operating Characteristic) 曲线若落在坐标轴对角线以上说明检验的正确率高,且越远离对角线可信度越高;反之落在对角线以下说明结果不足够可信。通常用 AUC (area under ROC curve),即 ROC 曲线与横坐标围成的面积大小来定量表示,AUC 越大,说明曲线越靠近上方,模型精度越高^[22]。

3 研究结果

3.1 滑坡和影响因子的关联

各影响因子的滑坡频次和标准滑坡频率统计结果见图 8。坡度是滑坡发生的重要内在条件,较大坡度通常更容易失稳而发生滑坡,但坡度超过 50° 的斜坡土层不易累积,土层厚度太薄,所以也不易发生滑坡^[23]。每个滑坡与非滑坡(即潜在滑坡)的高程矩阵边长不一致,导致二者计算出来的坡度、剖面

和平面曲率范围不一致,标准滑坡频率误差较大,其中剖面曲率和平面曲率的标准滑坡频率出现了大于 1 的极端情况,与理论相悖(图 8 b、c),此时标准滑坡频率参考价值不大。此外,过半数的滑坡发生在道路或建筑旁,说明人类工程活动产生的临空面对滑坡产生影响较大。

综合考虑各因子滑坡频次和标准滑坡频率后,对影响因子的分级如表 3 所示。

3.2 滑坡易发性评价

本文的所有逻辑回归模型中,均以因变量 y 表示模拟结果,影响因子由自变量 $x_1 \sim x_6$ 表示,分别代表坡度、剖面曲率、平面曲率、岩性、道路建筑临近

性和高程。如表 4 中的逻辑回归结果,每个自变量前的系数,表示每个影响因子(自变量)对滑坡易发性的贡献。其中坡度对滑坡的易发性影响最大,岩性和道路建筑临近性对滑坡的易发性也有较大的影响。但是每个模型的剖面曲率、平面曲率和高程因子都存在系数为负数的情况,说明部分影响因子与滑坡的发生存在负相关关系,与事实不符,可能是曲率的标准滑坡频率与发生频次差别太大(图 8),所以曲率的易发范围取值存在误差,将不易引起滑坡的范围包括在内,导致系数出现负数;而高程系数为负可能是因为高程对滑坡贡献相对较小(系数绝对值都最小),所以模型在推导过程中出现了误差。

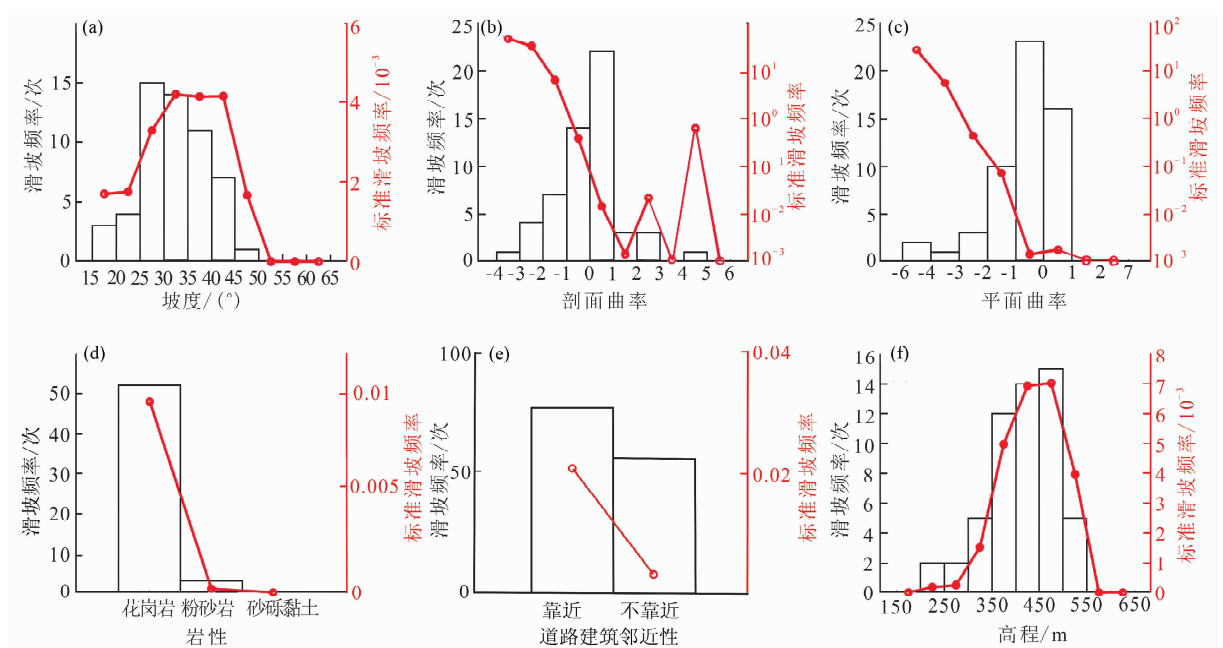


图 8 各影响因子的滑坡频次和标准滑坡频率统计图:

(a)坡度;(b)剖面曲率;(c)平面曲率;(d)岩性;(e)道路建筑邻近性;(f)高程

Fig. 8 Landslide frequency and normalized frequency of causative factors:

(a) Slope; (b) Profile curvature; (c) Plan curvature; (d) Rockiness; (e) Proximity to roads and buildings; (f) Elevation

表 3 影响因子分级表

Tab.3 Classification of causative factors

自变量	影响因子分级					
	坡度/(°) (x_1)	剖面曲率 (x_2)	平面曲率 (x_3)	岩性 (x_4)	道路临近性 (x_5)	高程/m (x_6)
0	$s < 15$ 或 $s > 45$	$pr \geq 3$	$pl \geq 1$	砂卵石和黏土	非道路、建筑旁	$0 \leq e < 300$ 或 $e > 550$
0.5	$15 \leq s < 25$	$0 \leq pr < 3$	$0 \leq pl < 1$	粉砂岩	-	$300 \leq e < 350$ 或 $500 \leq e < 550$
1	$25 \leq s \leq 45$	$pr < 0$	$pl < 0$	花岗岩	建筑、道路旁 60 m 内	$350 \leq e < 500$

表 4 不同尺度的逻辑回归模型
Tab.4 Logistic regression models on different scales

模型	表达式	<i>p</i> 值
单点	$y = \frac{1}{1 + e^{-(-10.80 + 0.74x_1 - 0.15x_2 + 0.01x_3 + 2.02x_4 + 0.76x_5 - 0.01x_6)}}$	(4-1) 2.12×10^{-6}
100 m × 100 m	$y = \frac{1}{1 + e^{-(-9.43 + 2.24x_1 + 9.11x_2 - 2.31x_3 + 0.97x_4 + 0.88x_5 + 0.34x_6)}}$	(4-2) 6.47×10^{-7}
200 m × 200 m	$y = \frac{1}{1 + e^{-(-9.92 + 3.22x_1 + 11.53x_2 - 2.21x_3 + 1.21x_4 + 1.19x_5 + 0.17x_6)}}$	(4-3) 1.07×10^{-6}
500 m × 500 m	$y = \frac{1}{1 + e^{-(-5.16 + 6.33x_1 - 1.16x_2 - 2.93x_3 + 0.85x_4 + 1.46x_5 + 0.84x_6)}}$	(4-4) 0.0326
1000 m × 1000 m	$y = \frac{1}{1 + e^{-(-7.64 + 12.87x_1 - 11.07x_2 + 6.69x_3 + 1.48x_4 + 3.62x_5 - 0.07x_6)}}$	(4-5) 0.43

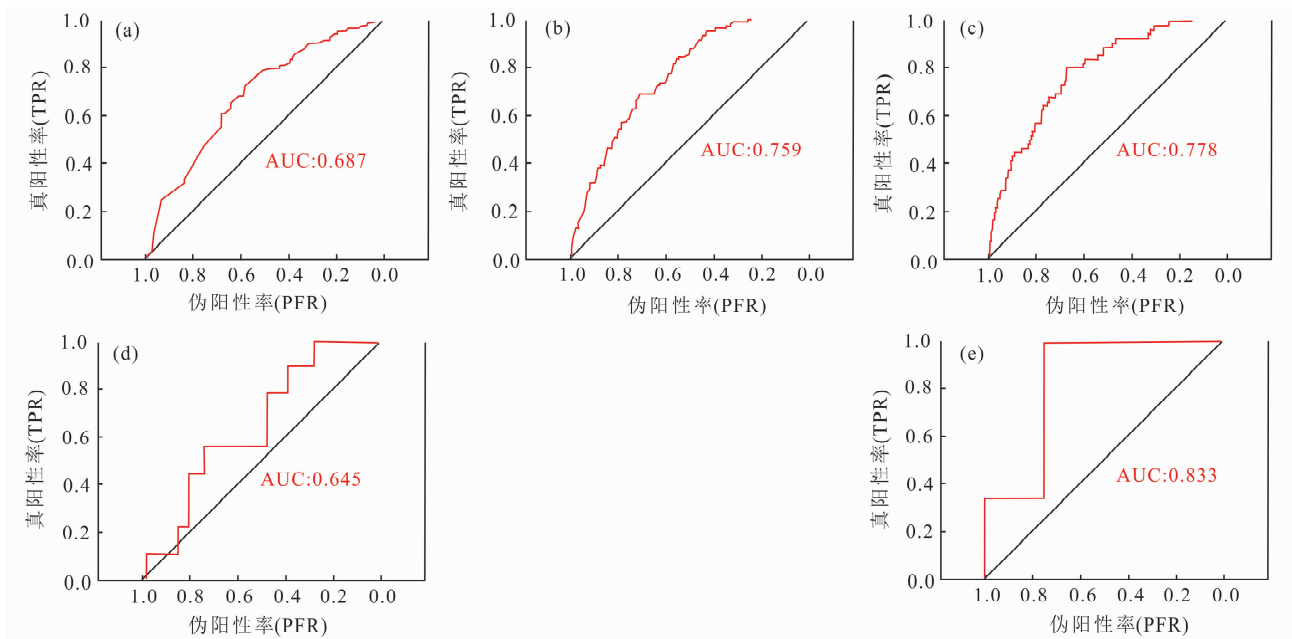


图 9 逻辑回归模型的 ROC 曲线及 AUC 值:

(a)单点模型;(b)100 m × 100 m 模型;(c)200 m × 200 m 模型;(d)500 m × 500 m 模型;(e)1000 m × 1000 m 模型

Fig.9 ROC curves and AUC: (a)Single-point model;

(b)100 m × 100 m model; (c)200 m × 200 m model; (d)500 m × 500 m model and (e)1000 m × 1000 m model

在显著性检验方面(表 4),单点、100 m × 100 m 和 200 m × 200 m 模型的 *p* 值均远小于 0.01,说明通过了 99% 的显著性检验,可以认为模型的评价效果好;500 m × 500 m 模型的 *p* 值小于 0.05,通过了 95% 的显著性检验,评价效果较好,而 1000 m × 1000 m 模型则未通过显著性检验。该模型评价效果不够好的原因可能是样本数量太少,模型无法准确地捕捉到变量间的变化关系。

在 ROC 曲线及其 AUC 检验方面(图 9),曲线基本都是完全位于对角线以上范围,且 AUC 值均在

0.7 左右,模型从统计学角度来看精度已达到预期;只有 500m × 500m 模型的 ROC 曲线有局部位于对角线下方,且 AUC 值为 0.645,比其他模型小,从统计学来看该模型的精度一般。

逻辑回归模型的模拟结果与实际发生情况的可视化对比如图 10 ~ 14。在构建单点模型时,由于研究区域内 DEM 点数量超过 100 万,直接回归效率低下,所以将研究区分为两个大小相近、滑坡数量相似的两部分,分别用于训练样本和测试样本。但该模型全部测试样本的预测结果均为低易发性(图 10),

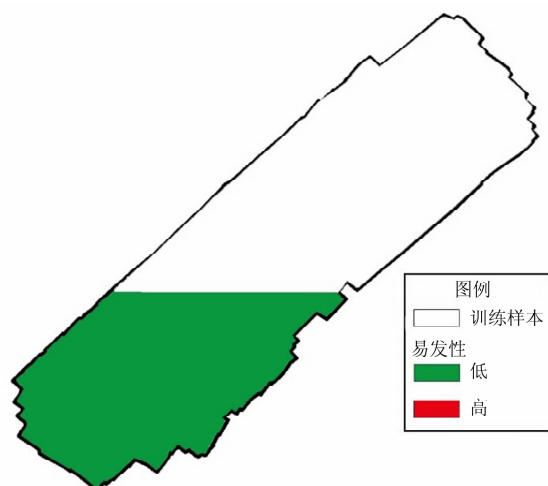


图 10 单点模型的易发性评价结果

Fig. 10 The result of single - point model evaluation

原因在于训练样本中发生滑坡(非零数据)的样本数量只有数十个,只占样本总数的不足万分之一;这种相对于非零数据而言,零(未发生)的个数明显过多的现象被称为零过多现象,也称为样本不平衡现象^[24]。模型计算时即便将所有结果均评估为不发生,该模型仍能通过最严格的显著性检验。这样的零过多现象使得该模型的计算结果没有实际意义。因此,单点逻辑回归模型不适合研究滑坡易发性。

检验网格模型的预测效果采用错位网格方法。

将用于训练的网格分别向北、向西和向西北方向移动半个网格距离后,分别形成新的网格,重新统计新的网格的自变量、因变量,并代入原训练获得的模型中,对比模型的预测结果与实际滑坡数量。这个方法的目的在不改变其他条件的情况下,通过移动网格位置形成新的网格,并在新的网格内分布不同的滑坡数量,从而形成新的样本,检验模型的正确性。

100 m × 100 m 模型没有预测到绝大多数滑坡发生的区域,未发生的区域反而存在个别被高估的情况(图 11)。所以 100 m × 100 m 模型不能预测滑坡易发性。

200 m × 200 m 模型的预测结果较好,整体上存在滑坡的区域都能准确预测(图 12),但高易发区域均没有被识别到,说明该模型在应用中可能会出现相对不安全的情况;同时研究区北部存在被高估的情况,即无滑坡区域被评估存在滑坡风险。

500 m × 500 m 的易发性评价结果最接近实际(表 5)。从空间分布上看,与 200 m × 200 m 模型结果类似,研究区东南角的实际滑坡多发区域并未被预测为较高的易发性,北部实际滑坡较少区域却大多被高估了约 1 个级别,西侧边缘的易发性较低区域,预测结果较好。东南角几乎都只被评估为中等

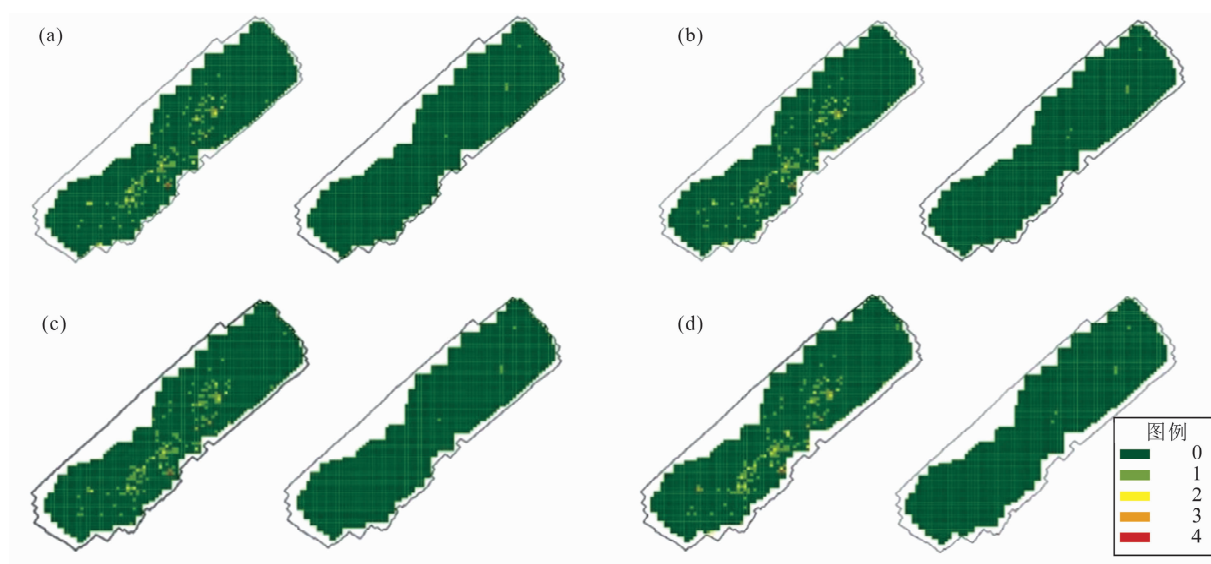


图 11 100 m × 100 m 模型评价易发性的结果与实际在不同网格中的对比:

(a) 原始网格;(b) 西移网格;(c) 北移网格;(d) 西北移动网格

Fig. 11 The results of 100 m × 100 m model evaluation compared with the actual numbers:

(a) Original grid; (b) Westward-offset grid; (c) Northward - offset grid; (d) Northwestward-offset grid

注:4 张子图中,左图表示滑坡实际发生情况,右图表示模型评价的易发性结果(图 12、13、14 同)。

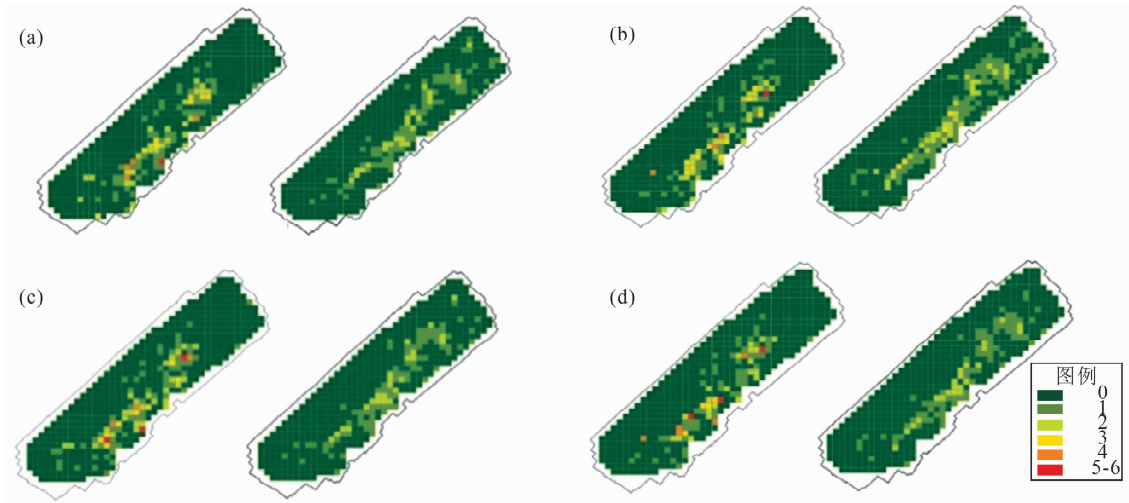


图 12 200 m × 200 m 模型评价易发性的结果与实际在不同网格中的对比:

(a) 原始网格; (b) 西移网格; (c) 北移网格; (d) 西北移动网格

Fig. 12 The results of the 200 m × 200 m model evaluation compared with the actual numbers in:

(a) Original grid; (b) Westward-offset grid; (c) Northward-offset grid; (d) Northwestward-offset grid

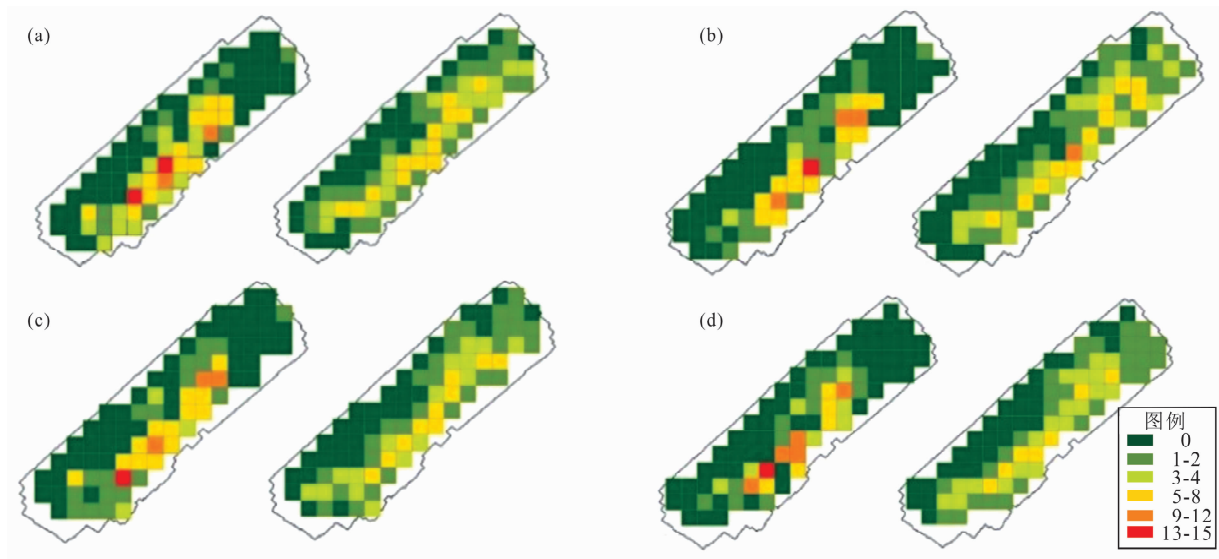


图 13 500 m × 500 m 模型评价易发性的结果与实际在不同网格中的对比:

(a) 原始网格; (b) 西移网格; (c) 北移网格; (d) 西北移动网格

Fig. 13 The results of 500 m × 500 m model evaluation compared with the actual numbers in:

(a) Original grid; (b) Westward-offset grid; (c) Northward-offset grid; (d) Northwestward-offset grid

易发(图 13),但个别网格的实际滑坡数量较大,可能是因为东南角高程较高,山区局地降雨差异,导致评价结果与实际滑坡数量不同。

1000 m × 1000 m 模拟结果和实际相似性不大,误差的分布也没有规律(图 14),不能用于滑坡易发性的预测。

进一步将各个尺度网格模型的预测结果与实际情况比较,统计发生滑坡的样本数量与样本总数的

比例(发生率)、准确率、预测结果与实际相差 1 个级别的样本数量与样本总数的比例(近准确率)、高估率和低估率结果如表 5 所示。

100 m × 100 m 模型的准确率均达到 93%,但是其低估率远大于高估率,且均与发生率完全相等,即所有发生了滑坡的网格均被低估,准确率仍能保证在 90% 以上,可见样本不平衡问题对模型的影响很大; 200 m × 200 m 模型的准确率达到 70% 以上,且

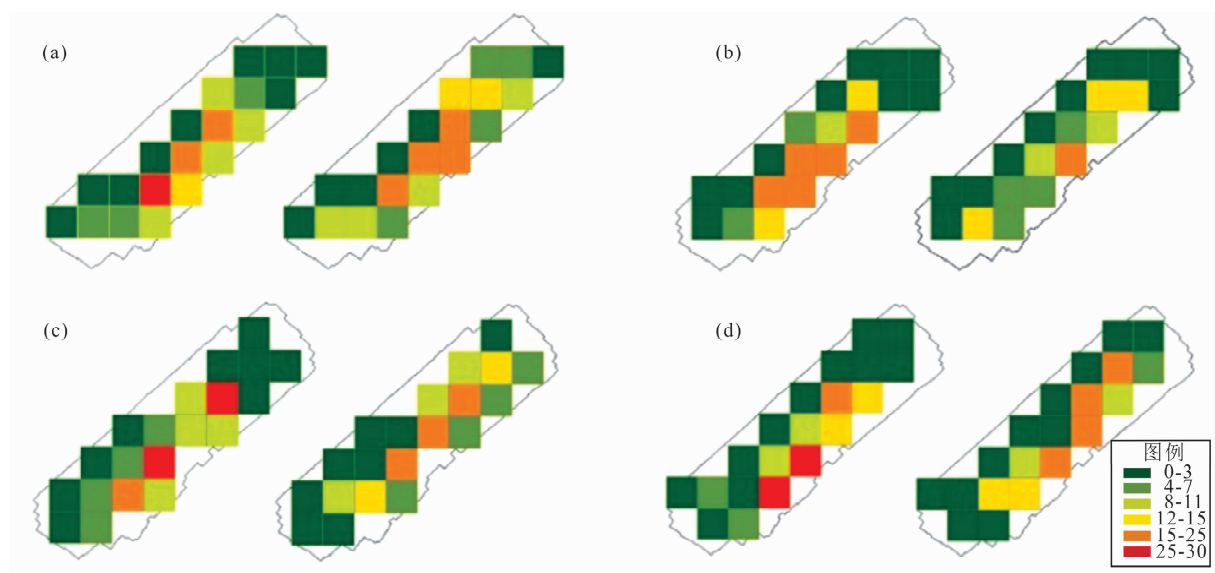


图 14 1000 m × 1000 m 模型评价易发性的结果与实际在不同网格中的对比：
(a) 原始网格；(b) 西移网格；(c) 北移网格；(d) 西北移动网格
Fig. 14 The results of 1000 m × 1000 m model evaluation compared with the actual numbers in：
(a) Original grid；(b) Westward-offset grid；(c) Northward-offset grid；(d) Northwestward-offset grid

表 5 不同尺度的模型评价对比
Tab. 5 Evaluation of different models

模型	网格	样本数量/个	发生率/%	准确率/%	近准确率/%	高估率/%	低估率/%
100 m × 100 m 模型	原始	2141	6.17	93.60	5.46	0.23	6.17
	西移	2141	6.54	93.18	5.98	0.28	6.54
	北移	2141	6.40	93.37	5.74	0.23	6.40
	西北移动	2141	6.59	93.13	5.74	0.28	6.59
200 m × 200 m 模型	原始	552	17.57	75.54	19.38	12.32	12.14
	左移	552	17.03	72.28	21.38	17.03	10.69
	上移	552	18.48	74.64	16.38	11.59	13.77
	西北移动	552	15.76	78.26	15.76	11.23	10.51
500 m × 500 m 模型	原始	85	45.88	40.00	42.35	32.94	27.06
	西移	85	42.35	42.35	36.47	37.65	20.00
	北移	85	44.71	43.53	42.35	31.76	24.71
	西北移动	85	36.47	44.71	43.53	37.65	17.65
1000 m × 1000 m 模型	原始	20	80	40.00	45.00	40.00	20.00
	西移	20	70	55.00	10.00	10.00	35.00
	北移	20	75	50.00	30.00	20.00	30.00
	西北移动	20	60	30.00	55.00	30.00	40.00

相差不大；500 m × 500 m 模型准确率尚可，各个指标的数值大小没有随着网格的移动而出现较大的波动，且几次检验的高估率均高于低估率，在实际预测中相对更安全；1000 m × 1000 m 模型不理想，预测结果不稳定。

综合模型的显著性、评价结果的准确性、实用性

和稳定性,500 m × 500 m 逻辑回归模型评价区域滑坡易发性效果好,评价结果相对安全;虽然该网格单元模型存在误差,但是误差是可控的,可以作为评价滑坡易发性的模型;其次,200 m × 200 m 的逻辑回归模型预测结果较好,能够相对准确地预测结果,但是存在低估率高于高估率的情况,不同网格预测结果可能会出现相对不安全的情况。

4 讨论与结论

4.1 讨论

本文采用改进的高程矩阵计算滑坡的地形参数,为滑坡的易发性评价提供了更可靠的数据。但是在非滑坡(潜在滑坡)区域的地形参数计算中,因不确定潜在滑坡的边长,均采用边长为 5 m 的高程矩阵计算其地形参数,使得滑坡和非滑坡区域统计的坡度、剖面曲率和平面曲率范围大小不一致,统计的标准滑坡频率参考性不够强,最终影响了滑坡的地形指标参数易发范围的准确性,造成建模误差较大。如果滑坡的边长尺度与潜在滑坡的边长尺度相近,统计的标准滑坡频率参考性会更强,滑坡的地形指标参数的易发范围会更准确,建模效果会更好,这个设想是否成立需要在将来的工作中予以验证。

在逻辑回归模型中,如果样本数量太少,模型可能会不显著;如果样本数量太多,可能会因为零过多现象出现大量被低估的结果^[23]。本文中 1000 m × 1000 m、500 m × 500 m、200 m × 200 m 和 100 m × 100 m 网格尺度的逻辑回归模型的样本数分别为 20、85、552 和 2141(表 5),评价效果分别是差、好、较好和较差,说明并非样本数越多越好;100 m × 100 m 网格尺度的逻辑回归模型尽管样本数量很多,但是因为零过多现象(发生率仅 6% 左右),在模型计算参数时因追求最大概率再现观测数据,使得很多样本被低估,造成评价效果较差。但样本太少(如 20 个)评价效果差,而 85 个样本即可达到好的评价效果,因此推测样本数量至少需要 50 才不会影响评价效果。

表 5 中,100 m × 100 m 网格尺度的逻辑回归模型的发生率均为 6% 左右,且与低估率完全相等。因此在评价模型中,即使全部发生样本均被低估,准确率仍然在 94% 左右,且模型显著;200 m × 200 m 的模型发生率在 17% 左右,评价结果较好;500 m ×

500 m 模型的发生率在 40% 左右,评价结果好。因此对比这三种网格尺度与相应的评价结果,推测模型的发生率至少在 30% 以上,才不会出现被低估的情况,不影响评价效果。同理,与零过多现象相反的是零不足现象,即零的数量比预期少^[25],这种情况下,模型的发生率也相应的应该控制在 70% 以下,以避免被高估,影响评价效果。

本文逻辑回归模型评价中,500 m × 500 m 网格尺度单元的滑坡易发性评价效果最好,并不能说明 500 m × 500 m 是最佳网格尺度。对于不同尺度的研究区和不同的滑坡密集程度,逻辑回归评价模型的单元尺度需要根据样本数量和发生率进行调整,确保样本数在 50 以上,并且发生率为 30% ~ 70% 才可以达到较好的评价效果。

4.2 结论

(1) 根据滑坡边长和方向改进后的高程矩阵使计算的滑坡地形数据更准确,可以更好地描述滑坡的地形特征。

(2) 坡度对滑坡的易发性影响最大,岩性和道路建筑可达性对滑坡的易发性也有较大的影响。

(3) 滑坡易发性逻辑回归模型中的网格尺度应确保滑坡发生样本为总样本的 30% ~ 70%,以避免被低估或高估;总样本数应该在 50 以上,以避免模型不显著。

本文在研究过程中发现了一些不足,这也为今后的研究提供了方向,例如,单点模型的滑坡数量相对网格个数较少,为了提高回归效率而将研究区分为训练样本和测试样本,使得滑坡数量更少,代表性较差,模型拟合效果不理想;在模型选择上,逻辑回归模型对于样本数量和发生率的要求较高,如果实验数据难以满足逻辑回归适用条件,可以尝试运用其他的数学模型和方法评估滑坡的易发性。最后,模型的评价结果只在福建三明市得到了检验,在其他地区是否适用还需要进一步研究。

参考文献(References)

- [1] 兰恒星,伍法权,王思敬. 基于 GIS 的滑坡 CF 多元回归模型及其应用[J]. 山地学报, 2002, 20(6): 732 - 737. [LAN Hengxing, WU Faquan, WANG Sijing. GIS based landslide CF multi-variable regression model and its application [J]. Mountain Research, 2002, 20(6): 732 - 737] DOI: 10.16089/j.cnki.1008-2786.2002.06.016
- [2] 赵良军,陈冬花,李虎,等. 基于二元逻辑回归模型的新疆果子

- 沟滑坡风险区划[J]. 山地学报, 2017, **35**(2): 203 - 211.
- [ZHAO Liangjun, CHEN Donghua, LI Hu, et al. A method to assess landslide susceptibility by using logistic regression model for Guozigou region, Xinjiang [J]. Mountain Research, 2017, **35**(2): 203 - 211] DOI: 10.16089/j.cnki.1008-2786.000213
- [3] 冯杭建, 周爱国, 俞剑君, 等. 浙西梅雨滑坡易发性评价模型对比[J]. 地球科学, 2016, **41**(3): 403 - 415. [FENG Hangjian, ZHOU Aiguo, YU Jianjun, et al. A comparative study on plum-rain-triggered landslide susceptibility assessment models in west Zhejiang province [J]. Earth Science, 2016, **41**(3): 403 - 415] DOI: 10.3799/dqkx.2016.032
- [4] AYALEW L, YAMAGISHI H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda - Yahiko Mountains, central Japan [J]. Geomorphology, 2005, **65**(1-2): 15 - 31. DOI: 10.1016/j.geomorph.2004.06.010
- [5] OHLMACHER G C, DAVIS J C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA [J]. Engineering Geology, 2003, **69**(3-4): 331 - 343. DOI: 10.1016/S0013-7952(03)00069-3
- [6] 张桂荣, 殷坤龙, 刘传正, 等. 基于 GIS 的陕西省旬阳地区滑坡灾害危险性区划[J]. 中国地质灾害与防治学报, 2003, **14**(4): 42 - 46. [ZHANG Guirong, YIN Kunlong, LIU Chuazheng, et al. The hazard zoning of landslide supported by GIS in Xunyang region of Shanxi province [J]. The Chinese Journal of Geological Hazard and Control, 2003, **14**(4): 42 - 46] DOI: 10.16031/j.cnki.issn.1003-8035.2003.04.008
- [7] 石菊松, 张永双, 董诚, 等. 基于 GIS 技术的巴东新城区滑坡灾害危险性区划[J]. 地球学报, 2005, **26**(3): 275 - 282. [SHI Jusong, ZHANG Yongshuang, DONG Cheng, et al. GIS-based landslide hazard zonation of the new Badong county site [J]. Acta Geoscientica Sinica, 2005, **26**(3): 275 - 282] DOI: 10.3321/j.issn:1006-3021.2005.03.014
- [8] KAWABATA D, BANDIBAS J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN) [J]. Geomorphology, 2009, **113**(1-2): 97 - 109. DOI: 10.1016/j.geomorph.2009.06.006
- [9] 向喜琼, 黄润秋. 基于 GIS 的人工神经网络模型在地质灾害危险性区划中的应用[J]. 中国地质灾害与防治学报, 2000, **11**(3): 23 - 27. [XIANG Xiqiong, HUANG Runqiu. Application of GIS-based artificial neural networks on assessment of geohazards risk [J]. The Chinese Journal of Geological Hazard and Control, 2000, **11**(3): 23 - 27] DOI: 10.16031/j.cnki.issn.1003-8035.2000.03.006
- [10] 武雪玲, 任福, 牛瑞卿, 等. 斜坡单元支持下的滑坡易发性评价支持向量机模型[J]. 武汉大学学报(信息科学版), 2013, **38**(12): 1499 - 1503. [WU Xueling, REN Fu, NIU Ruiqing, et al. Landslide spatial prediction based on slope units and support vector machines [J]. Geomatics and Information Science of Wuhan University, 2013, **38**(12): 1499 - 1503] DOI: 10.13203/j.whugis.2013.12.006
- [11] 胡涛, 樊鑫, 王硕, 等. 基于逻辑回归模型和 3S 技术的思南县滑坡易发性评价[J]. 地质科技通报, 2020, **39**(2): 113 - 121. [HU Tao, FAN Xin, WANG Shuo, et al. Landslide susceptibility evaluation of Sinan county using logistics regression model and 3S technology [J]. Bulletin of Geological Science and Technology, 2020, **39**(2): 113 - 121] DOI: 10.19509/j.cnki.dzkq.2020.0212
- [12] 邢秋菊, 赵纯勇, 高克昌, 等. 基于 GIS 的滑坡危险性逻辑回归评价研究[J]. 地理与地理信息科学, 2004, **20**(3): 49 - 51. [XING Qiuju, ZHAO Chunyong, GAO Kechang, et al. Logical regression analysis on the hazard of landslide based on GIS [J]. Geography and Geo-Information Science, 2004, **20**(3): 49 - 51] DOI: 10.3969/j.issn.1672-0504.2004.03.011
- [13] 方然可, 刘艳辉, 苏永超, 等. 基于逻辑回归的四川青川县区域滑坡灾害预警模型[J]. 水文地质工程地质, 2021, **48**(1): 181 - 187. [FANG Ranke, LIU Yanhui, SU Yongchao, et al. An early warning model of regional landslide in Qingchuan county, Sichuan province based on logistic regression [J]. Hydrogeology and Engineering Geology, 2021, **48**(1): 181 - 187] DOI: 10.16030/j.cnki.issn.1000-3665.201911034
- [14] 樊芷吟, 苟晓峰, 秦明月, 等. 基于信息量模型与 Logistic 回归模型耦合的地质灾害易发性评价[J]. 工程地质学报, 2018, **26**(2): 340 - 347. [FAN Zhiyin, GOU Xiaofeng, QIN Mingyue, et al. Information and logistic regression models based on coupling analysis for susceptibility of geological hazards [J]. Journal of Engineering Geology, 2018, **26**(2): 340 - 347] DOI: 10.13544/j.cnki.jeg.2017-052
- [15] 柳清文. 福建三明市区后山小流域山洪泥石流形成条件分析及危险性评价研究[D]. 成都: 成都理工大学, 2020: 7 - 13. [LIU Qingwen. Formation condition analysis and risk assessment of mountain torrents and debris flows in Houshan small watershed of Sanming city, Fujian province [D]. Chengdu: Chengdu University of Technology, 2020: 7 - 13] DOI: 10.26986/d.cnki.gedlc.2020.000610
- [16] KEEFER D K, WILSON R C, MARK R K, et al. Real-time landslide warning during heavy rainfall [J]. Science, 1987, **238**(4829): 921 - 925. DOI: 10.1126/science.238.4829.921
- [17] PRADHAN A M S, LEE S R, KIM Y T. A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan, Korea [J]. Landslides, 2019, **16**: 647 - 659. DOI: 10.1007/s10346-018-1112-z
- [18] 许冲, 戴福初, 姚鑫, 等. GIS 支持下基于层次分析法的汶川地震区滑坡易发性评价[J]. 岩石力学与工程学报, 2009, **28**(S2): 3978 - 3985. [XU Chong, DAI Fuchu, YAO Xin, et al. GIS-based landslide susceptibility assessment using analytical hierarchy process in Wenchuan earthquake region [J]. Chinese Journal of Rock Mechanics and Engineering, 2009, **28**(S2): 3978 - 3985] DOI: 10.3321/j.issn:1000-6915.2009.z2.100

- [19] 范强,巨能攀,向喜琼,等. 证据权法在滑坡易发性分区中的应用——以贵州桐梓河流域为例[J]. 灾害学,2015,30(1): 124 – 129. [FAN Qiang, JU Nengpan, XIANG Xiqiong, et al. Landslide susceptibility assessment using weights of evidence in the Tongzi watershed, China [J]. Journal of Catastrophology, 2015, 30(1): 124 – 129] DOI: 10.3969/j.issn.1000-811X.2015.01.024
- [20] 李郎平,兰恒星,郭长宝,等. 基于改进频率比法的川藏铁路沿线及邻区地质灾害易发性分区评价[J]. 现代地质,2017,31(5):911 – 929. [LI Langping, LAN Hengxing, GUO Changbao, et al. Geohazard susceptibility assessment along the Sichuan – Tibet railway and its adjacent area using an improved frequency ratio method [J]. Geoscience, 2017, 31(5):911 – 929] DOI: 10.3969/j.issn.1000-8527.2017.05.004
- [21] 李远远,梅红波,任晓杰,等. 基于确定性系数和支持向量机的地质灾害易发性评价[J]. 地球信息科学学报,2018,20(12):1699 – 1709. [LI Yuanyuan, MEI Hongbo, REN Xiaojie, et al. Geological disaster susceptibility evaluation based on certainty factor and support vector machine [J]. Journal of Geo – Information Science, 2018, 20(12): 1699 – 1709] DOI: 10.12082/dqxxkx.2018.180349
- [22] 潘懋,李铁峰. 灾害地质学[M]. 北京:北京大学出版社, 2012:55 – 58. [PAN Mao, LI Tiefeng. Disaster geology [M]. Beijing: Peking University Press, 2012: 55 – 58]
- [23] 谢宇. 回归分析[M]. 北京:社会科学文献出版社, 2010: 128 – 131. [XIE Yu. Regression analysis [M]. Beijing: Social Science Academic Press, 2010: 128 – 131]
- [24] 谢锋昌,韦博成,林金官. 零过多数据的统计分析及其应用[M]. 北京:科学出版社,2013:1 – 3. [XIE Fengchang, WEI Bocheng, LIN Jinguan. Statistical analysis of zero-inflated data and its application [M]. Beijing: Science Press, 2013:1 – 3]

A Grid-Scale Division Method Applicable to Logistic Regression Models for Evaluating the Susceptibility of Shallow Landslides ——Taking the 2019 Cluster of Shallow Landslides in Sanming, Fujian as Example

YU Congwei¹, LIU Kan², YIN Jie^{1*}, YU Bin^{3*}

(1. Department of Geography, East China Normal University, Shanghai 200241;

2. Key Laboratory of Geohazard Prevention of Hilly Mountains, Ministry of Land and Resources, Fuzhou 350002;

3. State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059)

Abstract: The susceptibility assessment is of great significance in landslides preventing studies, among which grids are often used as assessing units because of simple division and high regression efficiency. However, simply taking each raster as an assessing unit may cause problems, such as sample imbalance and the bias in landslides topography parameter, resulting in insignificant assessment. This study began with examining 131 mass landslides in Sanming recorded in 2019, using DEM data and aerial imagery to establish their extent, as well as their elevation, rockiness and the proximity of roads and building. In addition, the study calculated a range of landslide morphometric measures including slope, profile curvature and plan curvature after improving the elevation matrix according to the topological feature of landslides. Normalized landslides frequency was used to calculate the contribution of each causative factor to landslides occurrence, to establish logistic regression models of landslides susceptibility on different scales. After comparing the p values in hypothesis tests and the result maps of 100 m, 200 m, 500 m and 1000 m regular grids, the model based on the 500 m grid was recognized as the most suitable landslide susceptibility evaluation model, while the others models either underestimated or failed the hypothesis test. It is particularly noticeable that the incidence rates of underestimated models (100 m and 200 m) were lower than 30%, and the sample amount of the failed model (1000 m) was only 20. It was concluded that the improved

elevation matrix represents the real morphometric features of landslides, reducing the error caused by experimental data. Furthermore, the logical regression model was deficient, not only because the number of samples should exceed more than 50 to prevent model from insignificant, but also the incidence rate should be between 30% ~ 70%. These thresholds are considered important to avoid errors caused by the sample imbalance, resulting in too many underestimated or overestimated samples. This paper proposes a grid partitioning method applicable to logistic regression models to evaluate landslide susceptibility, expecting to provide a reference for landslide prevention and mitigation research.

Key words: landslide; susceptibility; elevation matrix; logistic regression; grid scale; Fujian province