



多元自适应回归样条法的滑坡敏感性评价

赵忠国^{1,3,4} 张峰² 郑江华^{1,3,4}

1 新疆大学绿洲生态教育部重点实验室,新疆 乌鲁木齐,830046

2 新疆交通职业技术学院,新疆 乌鲁木齐,831401

3 新疆大学资源与环境科学学院,新疆 乌鲁木齐,830046

4 新疆大学智慧城市与环境建模普通高校重点实验室,新疆 乌鲁木齐,830046

摘要:针对一般滑坡敏感性评价方法不能有效筛选滑坡条件因子的问题,以中国新疆维吾尔自治区新源县为研究区,基于15个滑坡敏感性条件因子,利用多元自适应回归样条法构建了滑坡敏感性指数预测模型,并自动筛选出研究区滑坡敏感性条件因子,在此基础上,实现了新源滑坡敏感性制图。此外,使用逻辑回归方法与多元自适应回归样条法进行精度对比分析。结果显示,采用多元自适应回归样条法构建的滑坡敏感性模型精度优于逻辑回归,其成功率曲线的精度为0.945 4,预测率曲线的精度为0.923 8。同时,模型还筛选出新源县滑坡重要影响条件因子(高程、坡度、降雨量、距断层距离、归一化差分植被指数、平面曲率、岩组)。研究表明,利用多元自适应回归样条构建的新源县滑坡敏感性模型是滑坡预测的有效方法,可为防灾减灾提供决策支持。

关键词:多元自适应回归样条法;逻辑回归;滑坡敏感性指数;敏感性分区

中图分类号:P208;P237

文献标志码:A

滑坡是指大量的岩石、泥土或岩屑物质沿斜坡的运动^[1],被认为是世界最具破坏力的自然灾害之一,它能造成大量财产损失和人员伤亡,尤其在复杂地质环境、极端天气和人为活动影响的山区^[2-4]。为了便于滑坡敏感性的预测,需要利用多种技术和以往的滑坡事件对滑坡敏感性进行准确的识别。

已有研究表明,多种灾害评价模型被应用于地质灾害分析中^[5-7]。模型主要包括统计方法模型和基于机器学习算法的模型。其中,统计方法模型包括双变量统计分析^[8]、统计指数^[9]、频率比^[10-11]、确定因子^[12]、逻辑回归^[10,13]、证据权重^[14]等。然而,统计方法模型需要依赖大量的数据并对数据进行预处理分析。机器学习算法是利用算法对训练数据学习进行建模分析。广泛应用于地质灾害分析的机器学习模型主要包括人工神经网络^[15]、支持向量机^[16]、决策树^[17]、随机森林^[18]等。这些机器学习算法往往运算复杂,对于预测变量与影响因子之间的联系不能合理的解释。一般的统计方法模型需要依赖大量的数据,

仅能分析各个因素类别之间与滑坡的关系。因此,此类方法不能利用较少的变量来分析各因素与滑坡之间的关系。而机器学习方法往往调参复杂,算法运行时间较长,不能自动筛选及合理地解释滑坡空间分布与滑坡条件因子之间的联系。

多元自适应回归样条(multivariate adaptive regression splines, MARS)是泛化能力很强的专门针对高维数据的非线性回归方法^[19]。MARS模型具有建模便利性、模型可解释性、变量可筛选性、运算快捷性等优点^[20],并已应用于积雪覆盖的反演^[21]、地下水潜力^[22]等方面的研究。因此,本文采用MARS模型分析新疆新源县滑坡敏感性,并与逻辑回归模型进行对比分析。

1 研究区及数据源

1.1 研究区概况

研究区位于中国新疆维吾尔自治区伊犁河谷东部的巩乃斯河谷地新源县,如图1所示,经纬度

收稿日期:2019-11-06

项目资助:新疆维吾尔自治区博士后经费(2017(507));新疆青年科技创新人才培养工程(QN2016YX03)。

第一作者:赵忠国,硕士,主要从事地质灾害空间建模与分析。1341785460@qq.com

通讯作者:郑江华,博士,教授。zheng_jianghua@126.com

坐标为 $82^{\circ}28'E \sim 84^{\circ}56'E$, $43^{\circ}03'N \sim 43^{\circ}40'N$ 。根据新源县政府网站,该县总面积为 $7\,581\text{ km}^2$,总人数 32.17 万人。研究区大小河流密布,主要水系为巩乃斯河、恰甫河和特克斯河。新源县具有湿润大陆性中温带气候特征^[23],年降水量在 $270 \sim 880\text{ mm}$,年平均气温 8.5°C 。新源县三面环山,东高西低,海拔范围在 $800 \sim 4\,255\text{ m}$,山地面积占该县总面积的 73.2% 。特殊的地形地貌形成了比较充分的降水,新源县一年中降水主要集中在 4 月至 7 月,该时期的降雨量占全年降雨总量的 $51.5\% \sim 80\%$ ^[23]。全县主要土地覆盖类型为草地和林地,分别占新源县总面积的 68.81% 和 12.85% ^[24]。黄土集中分布在新源县巩乃斯河谷的南北两侧中低山区,厚度不一,不稳定的高陡边坡多。

由于新源县地势地形特殊和气候条件复杂多变,局地强降水事件频发,而广泛分布的黄土容易发生滑坡等地质灾害。多年滑坡数据表明,新源县滑坡主要发生在降雨集中的 5 月至 10 月,且以中小型的黄土状物质浅层滑坡为主。

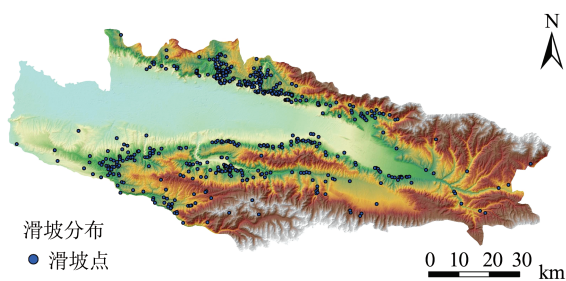


图1 研究区地理位置与滑坡灾害分布

Fig.1 Geographic Location and Disasters Distribution of Study Area

1.2 滑坡条件因子

根据已有文献^[25-26],并结合研究区特有的地形地貌和地质构造,本文共选取了 15 个滑坡条件因子,分别为高程、坡度、坡向、平面曲率、剖面曲率、标准曲率、岩组、距断层距离、归一化差分植被指数(normalized difference vegetation index, NDVI)、距道路距离、距河流距离、降雨量、黄土分布、汇流动力指数(stream power index, SPI)以及地形湿度指数(topographic wetness index, TWI)。

基于地理空间数据云网站获取的数字高程模型(digital elevation model, DEM),分别提取坡向、坡度、平面曲率、剖面曲率、标准曲率、高程、SPI、TWI 条件因子。其中,条件因子 SPI、TWI 为:

$$\begin{cases} S = A_s \times \tan \beta \\ t = \ln \left(\frac{A_s}{\tan \beta} \right) \end{cases} \quad (1)$$

式中, S 为条件因子 SPI 的值; t 为条件因子 TWI 的值; A_s 为单元栅格的汇水面积; β 为局地坡度。

NDVI 是通过地理空间数据云获取的 Landsat 8 遥感影像,经过辐射定标、大气校正、影像镶嵌预处理过程,经过下式计算得到:

$$N = (b_5 - b_4) / (b_5 + b_4) \quad (2)$$

式中, N 为条件因子 NDVI; b_5 和 b_4 分别表示电磁光谱近红外波段和红波段的光谱反射率;NDVI 的值在 $-1 \sim 1$ 之间,数值越大,表示绿色植被越密集,数值越低,表示植被越稀疏^[27]。

岩组、断层数据为 $1:250\,000$ 的数字化地质图;降水数据为年降水等值线的矢量数据;黄土分布数据为伊犁 $1:1\,500\,000$ 数字化黄土分布图。通过欧氏距离计算得到距道路距离、距河流距离和距断层距离。

1.3 滑坡灾害数据

本研究还包括 512 个历史滑坡灾害数据(记为“1”)。同时,为克服滑坡敏感性建模中的非平衡问题,本文使用 Areal Sampling 工具箱中的随机采样工具,每个数据点之间满足最小距离大于 200 m ,在非滑坡区域获取非滑坡灾害数据(512 个,记为“0”)。在此基础上,分别随机按照 7:3 比例在滑坡和非滑坡数据集上划分训练集和测试集^[28],得到 716 个数据的训练集和 308 个数据的测试集。

为便于分析研究,将以上滑坡条件因子、滑坡发生位置数据均转换为空间分辨率为 30 m 的栅格格式。

2 研究方法

2.1 多重共线性分析

多重共线性是指存在与两个或多个变量之间的线性关系。在回归分析中,如果独立变量之间存在多重共线性,会导致回归系数的方差增大,降低模型预测的准确性。因此,本文采用容忍度(tolerance, TOL)和方差膨胀因子(variance inflation factor, VIF)对滑坡条件因子进行多重共线性检验。如果 $VIF > 5$ 或 $TOL < 0.2$,则表明数据之间存在多重共线性问题^[9,29]。

2.2 频率比

在滑坡敏感性制图中,了解滑坡位置分布与

滑坡条件因子的关系是重要的步骤之一^[30]。本文利用频率比(frequency ratio, FR)分析滑坡分布与滑坡条件因子的相关关系, F_{FR} 为FR的值, F_{FR} 被定义为滑坡发生与研究区总面积的比。如果 $F_{FR}=1$,则认为滑坡与条件因子为一般相关; $F_{FR}>1$,则认为高相关; $F_{FR}<1$,则认为低相关^[31]。 F_{FR} 的计算公式如下:

$$F_{FR} = \frac{f_{ij}}{f} = \frac{P_{ij}^*}{P_{ij}} \times \frac{P}{P^*} \quad (3)$$

式中, f_{ij} 是条件因子 j 的分类 i 中滑坡密度; f 是整个研究区中滑坡密度; P_{ij}^* 是条件因子 j 的分类 i 中滑坡数量; P_{ij} 是条件因子 j 的分类 i 中的面积; P 是研究区总面积; P^* 是研究区滑坡发生总数量。

2.3 逻辑回归

逻辑回归(logistic regression, LR)通常用来解释二分因变量或者预测变量之间关系的回归分析方法^[32]。与一般的线性回归模型相比,这些变量可以为连续的变量,也可以是离散的变量。因此,LR是通过预测事件发生概率的二分问题(“0”和“1”)。LR可表示为:

$$\begin{cases} Y = \ln\left(\frac{P_j}{1-P_j}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \\ P_j = \frac{1}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}} \end{cases} \quad (4)$$

式中, Y 表示滑坡事件发生; α 为常量; $(X_1, X_2 \dots X_n)$ 为滑坡的条件因子; P_j 为在 j 栅格处滑坡发生的概率; $P_j/(1-P_j)$ 为可能性; $n(n=0, 1, 2 \dots n)$ 为独立变量的数量; $\beta_n(n=0, 1, 2 \dots n)$ 为LR的回归系数值。

2.4 MARS模型

MARS模型是一种结合样条回归、递归划分等优点的非线性、非参数的回归方法^[33]。同时,MARS模型能自动对连续变量以及二分变量中的非线性关系和交互作用建模,不必对独立变量与因变量之间的关系作假设。

MARS模型的一个重要优势是能对因变量有重要影响的独立变量进行筛选^[22]。模型采用自适应分段线性回归的线性组合和交互作用的基函数(basic function, BF)为基本单元。MARS模型表示为:

$$f(x) = \beta_0 + \sum_{i=1}^n \beta_i h_i(x) \quad (5)$$

式中, β_0 为常量; β_i 是相应基函数的系数; $h_i(x)$ 是第 i 个基函数。基函数表示为:

$$h_i(x) = \max(0, E - x) \text{ 或 } h_i(x) = \max(0, x - E) \quad (6)$$

式中, E 为输入参数的阈值; x 为输入变量。MARS模型构建分为前向逐步过程和后向的剪枝过程两步:第1步,前向逐步增加基函数。即每次选取一对最优基函数加入式(5),找到可能的节点,以此来获得更好的模型性能。然而,在这个过程中,加入过多的基函数会导致模型出现过拟合。第2步,后向剪枝。去除对模型贡献最小的冗余基函数,并找到最佳的模型,该过程中遵循广义交叉验证(generalized cross validation, GCV)原则。其形式如下:

$$G_{GCV} = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - f(x_i)]^2}{\left[1 - \frac{(H+1) + d \times H}{N}\right]^2} \quad (7)$$

式中, H 为基函数的数量; N 是数据的总数; $f(x_i)$ 为MARS模型预测的训练输出值; d 是惩罚系数。Friedman^[33]提出 d 最适值为 $2 \leq d \leq 4$,因此,本文取 $d=3$ 。

2.5 模型的评估

本文中使用真阳性率(true positive rate, TPR)和假阳性率(false positive rate, FPR)来评价滑坡模型的性能。TPR表示滑坡像素被正确分为滑坡的比例, T_{TPR} 为TPR的值;FPR则表示非滑坡像素被正确划分为非滑坡像素的比例^[34], F_{FPR} 为FPR的值。计算公式如下:

$$\begin{cases} T_{TPR} = \frac{T_p}{T_p + F_N} \\ F_{FPR} = \frac{T_N}{F_p + T_N} \end{cases} \quad (8)$$

式中, T_p 和 T_N 为正确分类的像素数量; F_N 和 F_p 为错误分类的像素数量。

接收者操作特征曲线(receiver operating characteristic curve, ROC)常被用来评估滑坡模型的性能^[35]。因此,本文采用ROC曲线对MARS和LR模型性能进行评估。ROC曲线按照数据集分为成功率曲线和预测率曲线,利用训练集构建成功率曲线用来表示模型与被观测的滑坡的拟合程度,使用测试集构建预测率曲线用来表示模型对滑坡预测的效果。此外,为定量地评价模型的性能,还计算了ROC曲线下面积(area under curve, AUC),AUC的值在0.5~1之间。AUC值越高,表示模型的预测能力越好^[29]。如果AUC的值大于0.8,则表示模型的性能良好。

3 结果分析

3.1 多重共线性分析

滑坡条件因子之间的多重共线性分析如表 1 所示,可以看出,VIF 的最大值为 2.012,TOL 的最小值为 0.497。因此,本文中的 15 个滑坡条件因子之间不存在多重共线性。

表 1 滑坡条件因子之间的多重共线性诊断检测

Tab.1 Multicollinearity Diagnostic Indices for Landslide Conditioning Factors

条件因子	容忍度	方差膨胀因子
高程	0.744	1.344
坡向	0.970	1.031
距断层距离	0.906	1.104
标准曲率	0.497	2.012
岩组	0.865	1.156
黄土分布	0.805	1.242
NDVI	0.890	1.124
平面曲率	0.629	1.590
剖面曲率	0.633	1.581
降雨量	0.796	1.257
距河流距离	0.978	1.022
距道路距离	0.840	1.190
坡度	0.741	1.350
SPI	0.829	1.206
TWI	0.784	1.275

3.2 滑坡分布与条件因子之间的关系

本文使用 FR 分析滑坡分布与各条件因子之间的关系,结果如表 2 所示。

表 2 中,高程按照 1 000 m 等间隔分为 8 类, $F_{FR}>1$ 在 1 200~1 800 分类中;在黄土分布的 FR 值中,黄土分类有最大的 $F_{FR}=2.572$;在分类变量岩组中,E 分类的 $F_{FR}=5.626$;距道路距离按 200 m 等间隔划分 5 类,除大于 1 000 m 以外,其他 4 类 $F_{FR}>1$;距断层距离按间隔 1 000 m 划分为 6 类,在 [1 000, 2 000) m 内,FR 值最高;距河流距离按 200 m 等间隔分为 6 类:小于 200 m 范围内,有最高的 $F_{FR}=1.119$;以 20 等间隔将 SPI 分为 5 类,在 [60, 80] 分类中, F_{FR} 最高为 1.952;TWI 使用自然间断点法分为 5 类,在 [1.25, 5.31) 分类中, F_{FR} 有最高值 1.569;按照 100 mm 等间隔将降雨量划分为 5 类,在 [500, 600] mm 中, F_{FR} 最高为 1.510;坡向则被划分为 9 类,东方向有较高的 $F_{FR}=1.203$;标准曲率按自然间断点法分为 5 类, F_{FR} 最高为 1.987,在 [-15.64, -2.19) 内;剖面曲率采用自然

间断点法分为 5 类,在 [1.11, 10.27] 中,剖面曲率有最大的 $F_{FR}=1.711$;平面曲率采用自然间断点法分为 6 类,在 [1.07, 11.65] 中, F_{FR} 有最大值为 1.590;坡度按 10° 间隔划分为 6 类,其中,在 [20°, 30°) 内, $F_{FR}=1.766$;NDVI 按自然间断点法分为 6 类,在 [0.16, 0.39) 内, F_{FR} 较高。

整体上,滑坡主要分布在海拔 1 200~1 800 m、降雨丰沛、水流侵蚀强烈、植被较稀疏的地区。同时,滑坡分布与砂岩、砾岩、泥岩为主的碎屑岩组(岩组(B))有密切关系。此外,滑坡分布与距道路距离、距断层距离以及距河流距离也关系紧密,滑坡分布与三者距离越近,则滑坡与条件因子之间的相关性越高。

3.3 LR 的应用

本文将频率比分类值作为 LR 的输入数据,采用前向 LR 来分析数据。在 SPSS19.0 中计算模型的系数,并应用模型系数生成滑坡敏感指数(landslide susceptibility index, LSI)。计算公式如下: $LSI_{LR} = -7.783 + (\text{坡度} \times 1.214) + (\text{距道路距离} \times 0.887) + (\text{降雨量} \times 1.225) + (\text{剖面曲率} \times 0.769) + (\text{黄土分布} \times 0.288) + (\text{距断层距离} \times 1.157) + (\text{高程} \times 0.830)$ 。

如图 2(a) 所示,本文将滑坡敏感性指数按照自然间断点分为 5 类:极低敏感[0, 0.138)、低敏感[0.138, 0.325)、中敏感[0.325, 0.534)、高敏感[0.534, 0.779)、极高敏感[0.779, 0.993]。

3.4 MARS 模型的应用

本文使用 R 中的“earth”包,建立了 MARS 模型。MARS 模型自动选择需要的变量来构建最佳的模型^[20]。MARS 模型仅筛选了 7 个条件因子(高程、坡度、降雨量、距断层距离、NDVI、平面曲率、岩组(B))。在分类变量中,仅选择分类类别,如岩组(B)类。MARS 模型计算并筛选出变量之间的相对重要性如表 3 所示。由表 3 可知,高程是解释研究区滑坡分布的最重要的条件因子,其次是坡度、降雨量、距断层距离、NDVI 以及岩组。MARS 模型的系数与方程如表 4 所示,根据表 4 生成了 MARS 模型的滑坡敏感性指数(L_{MARS})。

如图 2(b) 所示,本文使用自然间断点法将 MARS 模型运算得到的滑坡敏感性指数分为 5 类:极低敏感[0, 0.102)、低敏感[0.102, 0.314)、中敏感[0.314, 0.565)、高敏感[0.565, 0.804)、极高敏感[0.804, 1]。

3.5 模型验证与比较

在训练集上,LR、MARS 模型成功率曲线的

表 2 滑坡条件因子频率比统计

Tab.2 Frequency Ratio of Landslide Conditioning Factors

条件因子	分类	各级占 比/%	滑坡占 比/%	频率比	条件因子	分类	各级占 比/%	滑坡占 比/%	频率比
高程/m	<1 000	19.55	0.39	0.020	平面曲率	[−9.76, −1.36)	1.34	1.95	1.459
	[1 000,1 200)	8.49	5.27	0.621		[−1.36, −0.44)	9.02	12.30	1.364
	[1 200,1 400)	8.32	20.70	2.489		[−0.44,0.23)	69.95	55.86	0.799
	[1 400,1 600)	6.48	31.05	4.795		[0.23,1.07)	17.11	25.78	1.506
	[1 600,1 800)	7.51	26.95	3.588		[1.07,11.65]	2.58	4.10	1.590
	[1 800,2 000)	9.71	8.98	0.926	坡度/(°)	<10	41.15	11.72	0.285
	[2 000,2 200]	8.58	3.32	0.387		[10,20)	26.05	35.74	1.372
	>2 200	31.36	3.32	0.106		[20,30)	19.25	33.98	1.766
黄土分布	黄土	25.82	66.41	2.572		[30,40)	10.46	14.65	1.400
	非黄土	74.18	33.59	0.453		[40,50]	2.76	3.91	1.414
坡向	平面	0.44	0.00	0.000	标准曲率	>50	0.33	0.00	0.000
	北	13.97	16.6	1.188		[−15.64, −2.19)	3.05	6.05	1.987
	东北	12.08	13.09	1.083		[−2.19, −0.71)	11.58	13.87	1.198
	东	11.20	13.48	1.203		[−0.71,0.36)	61.81	45.31	0.733
	东南	9.57	8.20	0.857		[0.36,1.71)	18.66	28.32	1.518
	南	12.12	9.96	0.822		[1.71,18.65]	4.91	6.45	1.314
	西南	12.69	8.79	0.693	剖面曲率	[−9.86, −1.17)	4.95	6.25	1.262
	西	14.33	14.06	0.981		[−1.17, −0.39)	11.47	18.95	1.651
NDVI	西北	13.59	15.82	1.164		[−0.39,0.24)	60.89	46.09	0.757
	[−0.30,0.08)	5.47	0.78	0.143		[0.24,1.11)	17.21	19.34	1.124
	[0.08,0.16)	12.06	5.66	0.470		[1.11,10.27]	5.48	9.38	1.711
	[0.16,0.24)	19.60	21.68	1.106	岩组	A	49.56	62.50	1.261
	[0.24,0.31)	25.24	29.10	1.153		B	33.56	19.34	0.576
	[0.31,0.39)	23.49	27.93	1.189		C	10.09	8.59	0.852
距道路距离/m	[0.39,0.61]	14.13	14.84	1.050		D	5.85	4.30	0.735
	<400	57.12	66.80	1.169		E	0.94	5.27	5.626
	[400,600)	12.33	13.09	1.061	SPI	<20	47.36	34.77	0.734
	[600,800)	7.29	8.20	1.125		[20,40)	14.90	20.31	1.363
	[800,1 000]	4.92	5.66	1.152		[40,60)	8.32	9.18	1.103
距河流距离/m	>1 000	18.34	6.25	0.341		[60,80]	5.30	10.35	1.952
	<200	9.08	10.16	1.119		>80	24.12	25.39	1.053
	[200,400)	8.46	6.64	0.785	TWI	[1.25,5.31)	28.38	44.53	1.569
	[400,600)	8.10	8.59	1.060		[5.31,6.96)	38.92	42.19	1.084
	[600,800)	7.30	6.05	0.830		[6.96,9.17)	22.49	9.18	0.408
	[800,1 000]	7.02	7.62	1.084		[9.17,12.67)	7.95	2.54	0.319
距断层距离/m	>1 000	60.04	60.94	1.015		[12.67,24.63]	2.26	1.56	0.692
	<1 000	30.20	41.80	1.384	降雨量/mm	<300	1.37	0.00	0.000
	[1 000,2 000)	18.81	26.76	1.423		[300,400)	8.68	0.39	0.045
	[2 000,3 000)	12.36	15.63	1.264		[400,500)	15.89	6.84	0.430
	[3 000,4 000)	8.34	7.81	0.937		[500,600]	42.94	64.84	1.510
	[4 000,5 000]	6.53	2.54	0.389		>600	31.12	27.93	0.897
	>5 000	23.76	5.47	0.230					

AUC 值分别为 0.899 3 和 0.945 4,MARS 模型比 LR 模型高 0.046 1;在测试集上,LR、MARS 模型预测率曲线的 AUC 值分别为 0.905 9 和 0.923 8,MARS 模型比 LR 模型高 0.017 9。

结果表明,LR、MARS 模型的 AUC 值都大于 0.8,因此,两个模型都具有良好的预测能力。但在整体上,MARS 模型的预测能力要优于 LR 模型。

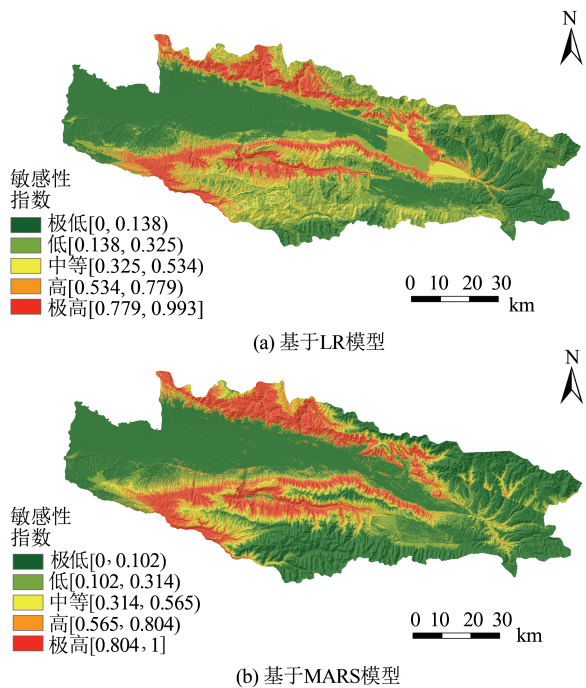


Fig 2 Landslide Susceptibility Map of Xinyuan County

表 3 MARS 模型中各独立变量的贡献度

Tab.3 Contributions of Various Independent Variables in MARS Model

条件因子	模型个数	广义交叉验证	残差平方和
高程	10	100.0	100.0
坡度	9	84.3	84.6
降雨量	9	84.3	84.6
距断层距离	6	31.6	34.7
NDVI	5	23.9	27.6
平面曲率	1	3.8	7.7
岩组(B)	1	3.8	7.7

表 4 MARS 模型的系数与方程

Tab.4 Coefficients and Equations in MARS Model

MARS 方程	系数
常量	+3.647 0
$\max(0, \text{距断层距离} - 1\ 642.35)$	-0.000 1
$\max(0, 1\ 610 - \text{高程})$	-0.004 0
$\max(0, \text{高程} - 1\ 610)$	-0.004 0
$\text{NDVI} \times \max(0, \text{高程} - 1\ 610)$	-0.011 0
$\max(0, \text{NDVI} - 0.282\ 7) \times \text{高程}$	+0.004 0
$\max(0, 17.274 - \text{坡度}) \times \text{高程}$	-0.001 0
$\max(0, 17.274 - \text{坡度}) \times \text{降雨量}$	+0.000 3
$\max(0.002\ 7 - \text{平面曲率}) \times \text{岩组}(B)$	+4.155 0
$\max(0, \text{降雨量} - 550) \times \text{高程}$	-0.000 006

3.6 讨论

滑坡敏感性制图不仅是滑坡危害评估的一个重要步骤,而且有利于旅游发展和可持续土地规划利用管理。因此,需要利用不同分析方法和

以往滑坡事件,对滑坡易发区进行准确识别。现有文献表明,LR 模型在滑坡敏感性评价中已被证明具有可靠性^[36]。同时,MARS 在过去文献中被用来分析土壤特性、地下水潜力、泥石流敏感性制图等不同研究领域均表现良好。Park 等^[22]利用 LR 和 MARS 模型评估了地下水潜力,研究表明,两模型总体精度较好。然而,不管在成功率曲线还是验证率曲线上,MARS 模型的 AUC 值均大于 LR 模型的 AUC 值。也就是说,MARS 模型较 LR 模型具有更高的预测能力,这与本文的结果一致。而 Rotigliano 等^[37]评估了二元 LR 和 MARS 模型的预测性能和鲁棒性,研究表明,MARS 性能稍好一点,但差异较小,无法将 MARS 定义为明显优于二元 LR。

本文分别使用 LR 和 MARS 对研究区滑坡敏感性建模并进行对比分析。结果表明,MARS 可用于研究区滑坡敏感性的建模评估。本研究区的高程、坡度、降雨量、距断层距离、NDVI、平面曲率、岩组对滑坡空间分布有较大的影响。其他相关研究中,高程、坡度、地表粗糙度、坡向、黄土分布、土地覆盖、断层、水系也是重要的条件因子^[23-24]。这些结果与本文结果相似,但坡向、黄土分布、水系在 MARS 中不是重要的因素,而黄土分布在 LR 中则参与了模型构建。这可能与研究区滑坡敏感性建模在方法上的使用差异有关。

此外,MARS 具有较好的预测能力^[20]。MARS 对于回归和分类问题都可建立线性和非线性模型,并能对混合自变量数值型数据进行有效处理,模型可有效地预测数据并便于解释,易于理解。而与其他滑坡敏感性模型比较,如神经网络、支持向量机、LR,MARS 具有良好的解释性并可对变量进行筛选。模型中即使引入其他不必要的条件因子,也可自动筛选变量来构建最佳模型,同时,减少了条件因子筛选和数据预处理的时间。尽管 LR 可根据前向算法对因子进行筛选,但不能自动删除相关性不大的因子。而本文中 MARS 比 LR 有更好预测能力的原因可能是:在不同区域范围,MARS 能够识别依赖变量与自变量之间的不同关系,并能自动删除相关性不大的因子。

4 结 语

本文基于 15 个条件因子,并采用 MARS 模型对新源县滑坡敏感性进行建模。为了分析 MARS 模型的性能,构建了 LR 滑坡敏感性模型

与其进行对比分析。

本文结果表明,MARS模型在本研究区滑坡敏感性建模有较好的性能,MARS模型的整体性能稍优于LR模型,MARS模型具有对变量的自动筛选和更好的解释性。高程、坡度、降雨量、断层距离、NDVI、平面曲率、岩组是研究区滑坡空间分布的主要影响因素。

综上,MARS模型应用于滑坡灾害敏感性预测分析具有可行性和有效性。此外,MARS模型还可用于其他的研究领域,并与其他不同模型进行对比分析,对其有效性进行评估。在本文中,没有涉及地表粗糙度和土地覆盖,这两种滑坡条件因子与滑坡空间分布的关系将在后续研究中进一步探讨。

参 考 文 献

- [1] Cruden D M. A Simple Definition of a Landslide [J]. *Bulletin of Engineering Geology and the Environment*, 1991, 43(1): 27-29
- [2] Shirzadi A, Bui D T, Pham B T, et al. Shallow Landslide Susceptibility Assessment Using a Novel Hybrid Intelligence Approach [J]. *Environmental Earth Sciences*, 2017, 76 (2), DOI: 10.1007/s12665-016-6374-y
- [3] Pham B T, Prakash I, Bui D T. Spatial Prediction of Landslides Using a Hybrid Machine Learning Approach Based on Random Subspace and Classification and Regression Trees [J]. *Geomorphology*, 2018, 303: 256-270
- [4] Wang Y, Wu X, Chen Z, et al. Optimizing the Predictive Ability of Machine Learning Methods for Landslide Susceptibility Mapping Using SMOTE for Lishui City in Zhejiang Province, China [J]. *International Journal of Environmental Research and Public Health*, 2019, 16 (3), DOI: 10.3390/ijerph16030368
- [5] Tao Shu, Hu Deyong, Zhao Wenji, et al. Sensitivity Evaluation of Secondary Landslide Disaster Based on Information and Logistic Regression Model: A Case Study of Northern Wenchuan County [J]. *Geographical Research*, 2010, 29(9): 1 594-1 605(陶舒, 胡德勇, 赵文吉, 等. 基于信息量与逻辑回归模型的次生滑坡灾害敏感性评价——以汶川县北部为例[J]. *地理研究*, 2010, 29(9): 1 594-1 605)
- [6] Bui D T, Tuan T A, Klempe H, et al. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree [J]. *Landslides*, 2016, 13(2): 361-378
- [7] Awawdeh M M, ElMughrabi M A, Atallah M Y. Landslide Susceptibility Mapping Using GIS and Weighted Overlay Method: A Case Study from North Jordan [J]. *Environmental Earth Sciences*, 2018, 77(21), DOI: 10.1007/s12665-018-7910-8
- [8] Bourenane H, Bouhadad Y, Guettouche M S, et al. GIS-Based Landslide Susceptibility Zonation Using Bivariate Statistical and Expert Approaches in the City of Constantine (Northeast Algeria) [J]. *Bulletin of Engineering Geology and the Environment*, 2015, 74(2): 337-355
- [9] Pourghasemi H R, Moradi H R, Aghda S M F. Landslide Susceptibility Mapping by Binary Logistic Regression, Analytical Hierarchy Process, and Statistical Index Models and Assessment of Their Performances [J]. *Natural Hazards*, 2013, 69 (1): 749-779
- [10] Shahabi H, Khezri S, Ahmad B B, et al. Landslide Susceptibility Mapping at Central Zab Basin, Iran: A Comparison Between Analytical Hierarchy Process, Frequency Ratio and Logistic Regression Models [J]. *Catena*, 2014, 115: 55-70
- [11] Chen W, Li W, Hou E, et al. Application of Frequency Ratio, Statistical Index, and Index of Entropy Models and Their Comparison in Landslide Susceptibility Mapping for the Baozhong Region of Baoji, China [J]. *Arabian Journal of Geosciences*, 2015, 8(4): 1 829-1 841
- [12] Dou J, Bui D T, Yunus A P, et al. Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan [J]. *PloS One*, 2015, 10(7), DOI: 10.1371/journal.pone.0133262
- [13] Chen W, Shahabi H, Shirzadi A, et al. A Novel Ensemble Approach of Bivariate Statistical-Based Logistic Model Tree Classifier for Landslide Susceptibility Assessment [J]. *Geocarto International*, 2018, 33(12): 1 398-1 420
- [14] Liu J, Duan Z. Quantitative Assessment of Landslide Susceptibility Comparing Statistical Index, Index of Entropy, and Weights of Evidence in the Shangnan Area, China [J]. *Entropy*, 2018, 20 (11), DOI: 10.3390/e20110868
- [15] Yilmaz I. A Case Study from Koyulhisar (Sivas-Turkey) for Landslide Susceptibility Mapping by Artificial Neural Networks [J]. *Bulletin of Engineering Geology and the Environment*, 2009, 68(3): 297-306
- [16] Tien B D, Shahabi H, Shirzadi A, et al. Landslide Detection and Susceptibility Mapping by Aisars Data

- Using Support Vector Machine and Index of Entropy Models in Cameron Highlands, Malaysia [J]. *Remote Sensing*, 2018, 10 (10), DOI: 10.3390/rs10101527
- [17] Khosravi K, Pham B T, Chapi K, et al. A Comparative Assessment of Decision Trees Algorithms for Flash Flood Susceptibility Modeling at Haraz Watershed, Northern Iran [J]. *Science of the Total Environment*, 2018, 627: 744-755
- [18] Liu Jian, Li Shulin, Chen Tao. Landslide Susceptibility Assessment Based on Optimized Random Forest Model [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(7): 1 085-1 091 (刘坚, 李树林, 陈涛. 基于优化随机森林模型的滑坡易发性评价 [J]. 武汉大学学报·信息科学版, 2018, 43(7): 1 085-1 091)
- [19] Zhang Xiumin, Nan Zhuotong, Wu Jichun, et al. Mapping of Frozen Soils in the Hot Springs of the Qinghai-Tibet Plateau Based on Multiple Adaptive Regression Splines [J]. *Journal of Glaciology and Geocryology*, 2011, 33(5): 1 088-1 097 (张秀敏, 南卓铜, 吴吉春, 等. 基于多元自适应回归样条的青藏高原温泉区域的冻土分布制图 [J]. 冰川冻土, 2011, 33(5): 1 088-1 097)
- [20] Zhang W, Goh A T C, Zhang Y. Multivariate Adaptive Regression Splines Application for Multivariate Geotechnical Problems with Big Data [J]. *Geotechnical and Geological Engineering*, 2016, 34 (1): 193-204
- [21] Kuter S, Akyurek Z, Weber G W. Retrieval of Fractional Snow Covered Area from MODIS Data by Multivariate Adaptive Regression Splines [J]. *Remote Sensing of Environment*, 2018, 205: 236-252
- [22] Park S, Hamm S Y, Jeon H T, et al. Evaluation of Logistic Regression and Multivariate Adaptive Regression Spline Models for Groundwater Potential Mapping Using R and GIS [J]. *Sustainability*, 2017, 9(7), DOI:10.3390/su9071157
- [23] Chang Shunli, Zhang Zhongyue, Sun Zhiquan, et al. GIS-Based Landslide Hazard Analysis and Zoning in Xinyuan County [J]. *Journal of Natural Disasters*, 2011, 20(5): 216-221 (常顺利, 张钟月, 孙志群, 等. 基于 GIS 的新源县滑坡灾害分析与区划 [J]. 自然灾害学报, 2011, 20(5): 216-221)
- [24] Zhou Pinghua, Li Yingcheng, Bai Jie, et al. Evaluation of Landslide Vulnerability in Catchment Unit of Xinyuan County [J]. *Science of Surveying and Mapping*, 2017, 42(9): 82-87 (周平华, 李英成, 白洁, 等. 新源县汇水单元滑坡灾害易发性评价 [J]. 测绘科学, 2017, 42(9): 82-87)
- [25] Zhang Ce, Jie Wenhui, Fu Lihua, et al. Remote Sensing Image and Distribution Characteristics of Landslide Disasters in Xinyuan County, Xinjiang [J]. *Remote Sensing for Land and Resources*, 2017, 29(S1): 81-84 (张策, 揭文辉, 付丽华, 等. 新疆新源县滑坡灾害遥感影像特征及分布规律 [J]. 国土资源遥感, 2017, 29 (S1): 81-84)
- [26] Tao K, Wang L, Qian X. Multi-factor Constrained Analysis Method for Geological Hazard Risk [J]. *International Journal of Engineering and Technology*, 2016, 8(3), DOI: 10.7763/IJET.2016.V6.884
- [27] Pradhan B, Lee S. Landslide Susceptibility Assessment and Factor Effect Analysis: Backpropagation Artificial Neural Networks and Their Comparison with Frequency Ratio and Bivariate Logistic Regression Modelling [J]. *Environmental Modelling and Software*, 2010, 25(6): 747-759
- [28] Chen W, Xie X, Peng J, et al. GIS-Based Landslide Susceptibility Evaluation Using a Novel Hybrid Integration Approach of Bivariate Statistical Based Random Forest Method [J]. *Catena*, 2018, 164: 135-149
- [29] Chen W, Li H, Hou E, et al. GIS-Based Groundwater Potential Analysis Using Novel Ensemble Weights-of-Evidence with Logistic Regression and Functional Tree Models [J]. *Science of the Total Environment*, 2018, 634: 853-867
- [30] Lee S, Hwang J, Park I. Application of Data-Driven Evidential Belief Functions to Landslide Susceptibility Mapping in Jinbu, Korea [J]. *Catena*, 2013, 100: 15-30
- [31] Ozdemir A, Altural T. A Comparative Study of Frequency Ratio, Weights of Evidence and Logistic Regression Methods for Landslide Susceptibility Mapping: Sultan Mountains, SW Turkey [J]. *Journal of Asian Earth Sciences*, 2013, 64: 180-197
- [32] Lachenbruch P A, McCullagh P, Nelder J A. Generalized Linear Models [J]. *Biometrics*, 1990, 46 (4): 291 - 303
- [33] Friedman J H. Multivariate Adaptive Regression Splines [J]. *The Annals of Statistics*, 1991, 19(1), DOI:10.1214/aos/1176347963
- [34] Hong H Y, Liu J Z, Bui D T, et al. Landslide Susceptibility Mapping Using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest Ensembles in the Guangchang Area (China) [J]. *Catena*, 2018, 163: 399-413
- [35] Hussin H Y, Zumpano V, Reichenbach P, et al. Different Landslide Sampling Strategies in a Grid-Based Bivariate Statistical Susceptibility Model [J]. *Geomorphology*, 2016, 253: 508-523

- [36] Pourghasemi H R, Teimoori Yansari Z, Panagos P, et al. Analysis and Evaluation of Landslide Susceptibility: A Review on Articles Published During 2005—2016 (Periods of 2005—2012 and 2013—2016) [J]. *Arabian Journal of Geosciences*, 2018, 11(9), DOI: 10.1007/s12517-018-3531-5
- [37] Rotigliano E, Martinello C, Agnesi V, et al. Evaluation of Debris Flow Susceptibility in El Salvador (CA): A Comparison Between Multivariate Adaptive Regression Splines (MARS) and Binary Logistic Regression (BLR)[J]. *Hungarian Geographical Bulletin*, 2018, 67(4): 361-373

Evaluation of Landslide Susceptibility by Multiple Adaptive Regression Spline Method

ZHAO Zhongguo^{1,3,4} ZHANG Feng² ZHENG Jianghua^{1,3,4}

1 Key Laboratory of Oasis Ecology Education, Xinjiang University, Urumqi 830046, China

2 Xinjiang Vocational and Technical College of Communications, Urumqi 831401, China

3 College of Resources and Environmental Sciences, Xinjiang University, Urumqi 830046, China

4 Key Laboratory of Intelligent City and Environmental Modeling, Xinjiang University, Urumqi 830046, China

Abstract: Objectives: According to general landslide susceptibility evaluation methods, landslide condition factors cannot be effectively selected. **Methods:** The prediction model of landslide susceptibility index (LSI) is constructed by multiple adaptive regression spline (MARS), and the landslide susceptibility condition factors are automatically selected, and the landslide susceptibility map is produced by 15 landslide susceptibility factors. In addition, the accuracy of the model is compared between logistic regression (LR) and MARS. **Results:** The results show that the accuracy of landslide susceptibility model constructed by MARS is better than LR. The accuracy of MARS success curve is 0.945 4, and the accuracy of MARS prediction rate curve is 0.923 8. At the same time, the model also selects the important influencing factors of landslide (elevation, slope angle, rainfall, distance to faults, NDVI, plan curvature, geological petrofabric). **Conclusions:** Research suggests that the MARS is an effective method for landslide prediction in study area and can provide decision support for reducing nature disaster.

Key words: multivariate adaptive regression splines; logistic regression; landslide susceptibility index; susceptibility zoning

First author: ZHAO Zhongguo, master, specializes in spatial modeling and analysis of geological disasters. E-mail: 1341785460@qq.com

Corresponding author: ZHENG Jianghua, PhD, professor. E-mail: zheng_jianghua@126.com

Foundation support: The Xinjiang Uygur Autonomous Region Postdoctoral Funding (2017 (507)); the Xinjiang Youth Science and Technology Innovation Talents Training Project (QN2016YX03).

引文格式: ZHAO Zhongguo, ZHANG Feng, ZHENG Jianghua. Evaluation of Landslide Susceptibility by Multiple Adaptive Regression Spline Method[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(3):442-450. DOI:10.13203/j.whugis20190136(赵忠国, 张峰, 郑江华. 多元自适应回归样条法的滑坡敏感性评价[J]. *武汉大学学报·信息科学版*, 2021, 46(3):442-450. DOI:10.13203/j.whugis20190136)