

文章编号: 1008 - 2786 - (2018)5 - 765 - 09

DOI: 10.16089/j.cnki.1008-2786.000372

基于随机森林模型的泥石流易发性评价

——以汶川地震重灾区为例

刘永垚¹, 第宝锋^{2,1*}, 詹宇¹, Constantine A. Stamatopoulos³

(1. 四川大学 建筑与环境学院, 成都 610065; 2. 四川大学 灾后重建与管理学院, 成都 610207;

3. Stamatopoulos and Associates Co. and Hellenic Open University, 5 Isavron str, 11471 Athens, Greece)

摘 要: 在区域泥石流易发性研究中, 科学确定泥石流易发性主控因子及其贡献率既是关键科学问题, 也是区域泥石流预警预报和风险管理的重要基础。本研究选取汶川地震重灾区, 引入随机森林算法, 以小流域为评价单元, 集合多元因子指标体系, 建立泥石流易发性评价模型, 定量分析了汶川地震重灾区泥石流关键影响因子及贡献率, 并探讨了研究区泥石流易发性的空间分布特征。本文初选了 63 项评价指标, 以模型 AUC 值变化为基础, 筛选出 35 项指标构成易发性评价指标体系, 并用于区域内泥石流易发性主控因子的识别, 结果表明: 流域高差、流域平均坡度、流域内滑坡面积、平均降雨天数是区域内泥石流易发性主控因子, 另外, 沟长比降、大于 10° 积温、年均温、人口密度、村落个数、低覆盖度土地利用方式等在泥石流易发性评价中也发挥着重要作用; 易发性评价结果显示, 极高易发区占比达到了 22.94%, 主要分布于研究区西部, 泥石流易发性较高的小流域主要分布在青藏高原向四川盆地过渡的地形急变带, 同时也与地震带、断裂带、干旱河谷区域密切相关。模型验证结果表明, 平均 AUC 值达 0.84, 模型具有很高的稳定性和准确性, 说明随机森林算法非常适用于区域泥石流易发性评价研究, 机器学习算法结合小流域为单元的方法对区域泥石流易发性评价有良好效果, 可为区域尺度灾害易发性及风险评估提供更为有效的方法参考。

关键词: 泥石流; 易发性; 随机森林算法; 关键因子; 汶川地震

中图分类号: P694

文献标志码: A

泥石流是我国主要的山地灾害之一, 2008 年汶川地震引发的崩塌、滑坡等次生灾害给泥石流活动提供了大量的松散固体物质, 灾区泥石流活动已经逐渐进入了活跃期^[1, 2], 近年来灾区的泥石流灾害对大小聚落、公路桥梁、水电站等造成了严重破坏。通过泥石流易发性研究, 探究泥石流易发性主控因子贡献率、厘清易发性分区对泥石流监测预警和风险管理具有重要意义。

自 20 世纪 70 年代以来, 学者们已经对泥石流

进行了系统的研究, 在泥石流易发性评价方面, 区域泥石流易发性研究通常基于经验和统计方法并结合地理信息系统和遥感展开^[3-5], 主要方法包括: 多因子综合评价法^[6, 7]、人工神经网络评价法^[8]、遥感与 GIS 评价法^[9]、层次分析法^[10]等, 就目前研究成果而言, 评价模型应用程度高, 涉及的区域广阔^[11, 12]。然而, 由于泥石流的复杂非线性特征, 区域泥石流易发性评价研究中很多问题尚未得到系统解决, 评价单元多以栅格或行政区为主, 而以小流域为单元的

收稿日期 (Received date): 2018 - 10 - 24; **改回日期** (Accepted date): 2018 - 10 - 25

基金项目 (Foundation item): 2017 中国和希腊政府间科技合作项目。[2017 Joint China-Greece Intergovernment Science and Technology Cooperation Project]

第一简介 (Biography): 刘永垚 (1993 -), 男, 四川绵阳人, 硕士研究生, 主要研究方向: 水土保持规划与评价。[LIU Yongyao (1993 -), male, born in Mianyang, Sichuan province, M. Sc. candidate, research on soil and water conservation planning] E-mail: liuyongyao93@163.com

*** 通讯作者** (Corresponding author): 第宝锋 (1977 -), 男, 博士, 教授, 主要研究方向: 山地灾害风险、环境遥感。[DI Baofeng (1977 -), male, Ph. D., professor, specialized in mountain disaster risk and remote sensing of environment] E-mail: dibaofeng@scu.edu.cn

研究较少,栅格单元在数据计算方面更具优势,行政单元利于管理决策制定,但是小流域单元更能体现泥石流灾害影响因素的完整性^[13];评价指标选取存在较大的不确定性,一定程度上受人为主观判断影响,缺乏相应的评价标准和技术规范^[12];而目前权重赋值方法可分为主观赋权法、客观赋权法,其中主观赋值法受人为主观影响较大,而客观赋值法受数据影响较大。

对于易发性评价研究中存在的问题而言,随机森林算法是一种非常有前景的非线性方法^[14],该方法是一种基于统计学习理论的机器学习算法,通过集成多棵决策树,然后采用投票方式选出分类结果,与层次分析法等半定量评价方法相比较,主要优势表现在:数据预处理阶段,随机森林模型无需归一化等去量纲规范化处理,可直接进行运算;评价指标选取阶段,处理高维度和大数据量的数据集能力强,不需要做特征选择;易发性评价阶段,模型可获取评价指标贡献率,避免了人为主观赋值的影响,比层次分析法权重赋值规则更简单、准确,模型稳定性与准确性较高^[15, 16]。该方法已经在某些灾害领域成功应用,如滑坡^[17]、森林火灾^[18]、洪水^[19]以及少量的泥石流易发性研究^[20],且具有良好的预测结果。本文以汶川地震重灾区为研究区,以小流域为评价单元,采用随机森林方法,构建汶川地震重灾区易发性评价指标体系,建立评价模型,对评价指标贡献率和易发性分区进行分析从而为灾区泥石流监测预警、风险管理决策的制定提供数据支撑。

1 研究方法

1.1 决策树概述

决策树是随机森林模型的基分类器,是一种广泛应用于分类和预测的单分类器,在机器学习中决

策树属于一种预测模型,它代表的是对象属性和对象值之间的一种映射^[21];现阶段常见的决策树算法包括 CLS、ID3、C4.5、CART 算法等。决策树结构简单、易于理解、分类效率高、能处理高维数据并且分类准确率高;决策树也存在着分类规则复杂、过度拟合与收敛到非全局的局部最优解等缺点^[21]。

1.2 随机森林模型概述

随机森林分类是由多个决策树 $\{h(X, \theta_k), k = 1, \dots\}$ 组成的组合分类模型,参数集 $\{\theta_k\}$ 是独立同分布的随机向量,在给定自变量 X 下,每个决策树分类模型都通过一票投票来选择最优的分类结果。算法如图 1 所示。

首先,利用 bootstrap 抽样从原始训练集 D 中抽取 k 个样本,且每个样本中的特征数(m)都与原始训练集相同;bootstrap 抽样为有放回的抽样,提升了训练集的随机性。

然后对 k 个样本分别建立 k 个决策树模型,得到 k 种分类结果 $\{h_1(X), h_2(X), \dots, h_k(X)\}$;每个样本中随机选取 $n(n \leq m)$ 个特征作为分裂特征集,从中选择最优特征对节点进行生长,当 $n < m$ 时,每一棵决策树之间又存在差异性。

最后,通过上述步骤形成的随机森林,根据 k 种分类结果对每个记录进行投票表决以决定其最终分类,公式如下^[22]。

$$f(x_i) = \text{m_vote}\{h_i(x)\} \quad (i = 1, 2, \dots, k) \quad (1)$$

式中:m_vote 为投票结果。

由上可知,随机森林模型在训练集的随机性与节点分裂最优属性的两处随机性共同作用下,使得随机森林不易产生过拟合,增加了模型的稳定性^[25]。

1.3 ROC 曲线与 AUC 值

ROC (Receiver Operating Characteristics) 曲线与

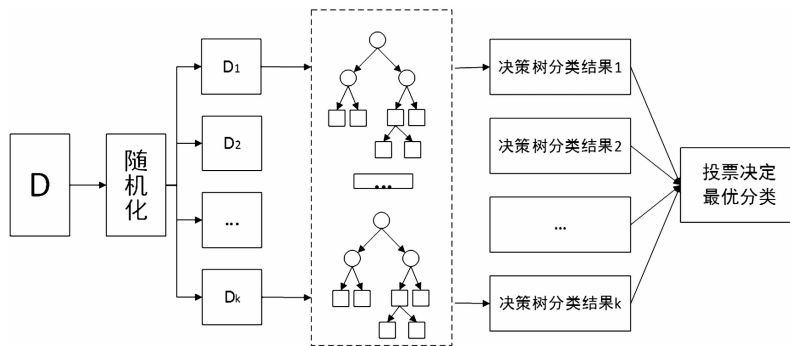


图 1 随机森林示意图

Fig. 1 Schematic illustration of random forest method

AUC(Area Under ROC) 常被用来评价模型性能优劣,为了介绍上述两项指标,下引入二分类数据的混淆矩阵,如表 1 所示。

表 1 二分类数据的混淆矩阵

Tab. 1 Confusion matrix

	Classified positive	Classified negative
Positive	TP	FN
Negative	FP	TN

上表中表示数据集中有正 (Positive) 与负 (Negative) 两种样本,而对这两种的分类包括正确分类 (Classified Positive) 和错误分类 (Classified Negative),分别对应正类样本与负类样本后,得到 4 种分类 TP (True Positive)、FP (False Positive)、FN (False Negative)、TN (True Negative);根据上述 4 种数据,可以计算 TPR (True Positive Rate) 与 FPR (False Positive Rate),公式如下:

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

在分类过程中,阈值 (Threshold) 表示分类器认为某个样本为正样本 (或负样本) 的概率;当调整阈值 (Threshold) 时,每 1 个阈值对应 1 组 TPR 和 FPR,ROC 曲线即为以 TPR 为纵轴、FPR 为横轴的曲线。在曲线下方的区域对应的面积则为 AUC,该

数值介于 0 和 1 之间,一般大于 0.5,越接近 1,表明分类器的效果更好,也就是说模型的准确性越高。

2 研究区与数据

2.1 研究区概况

本文中选取的研究区地处中国西南地区,位于四川、甘肃、陕西三省交界地区,由覆盖汶川地震 41 个重灾县 (区、市) 的 720 小流域构成,总面积达 14.27 万平方千米 (图 2)。

研究区地处中国地形第一、二级阶梯交界过渡地带,西部以高山峡谷区为主,山高坡陡,东部以丘陵、平原为主,海拔范围 450 ~ 5000 m 左右;区域内构造活动极为复杂、地形切割剧烈、地震活动频繁、山地灾害频发^[23];研究区地属山地亚热带湿润季风气候与暖温带大陆性半干旱季风气候交汇区^[24],区域内水系众多,皆属长江左岸支流;河网密布,大小河流超过 200 条,自西向东为岷江、沱江、涪江和嘉陵江流域^[24]。

2.2 数据

小流域是独立的集水地貌单元,可以将其看作一个最小的、独立的“孕灾系统”。本文以小流域为评价单元,数据包括研究区小流域数据、泥石流灾害数据与评价指标数据。

通过遥感解译与 ArcGIS 水文分析,共划分出 718 处小流域 (图 2 所示);此外,为保证汶川地震重灾区的完整性,将区内平原部分与高原部分纳入研

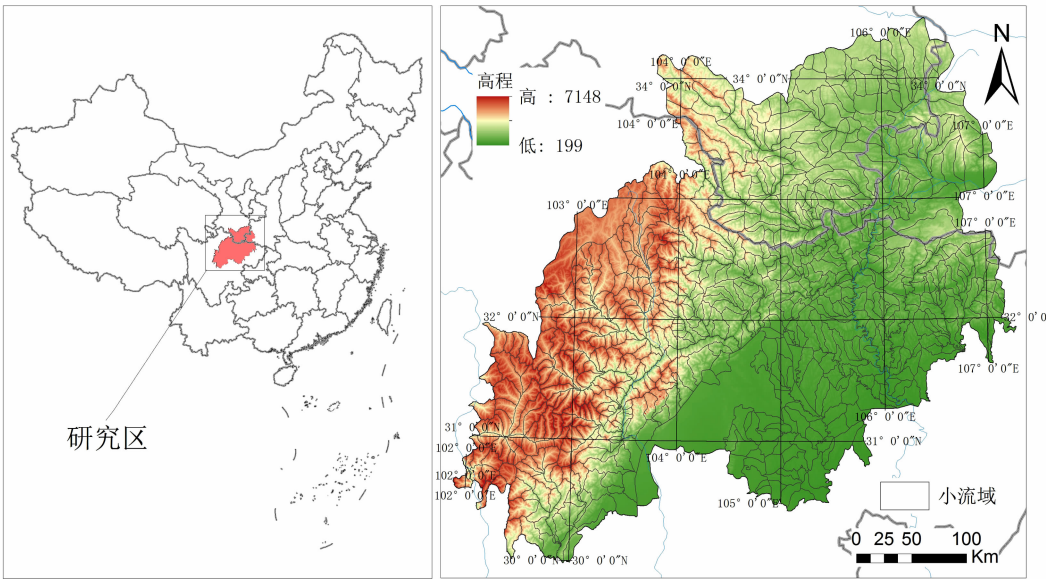


图 2 研究区区位图

Fig. 2 Location map of the study area

究区(编号为 719、720 号小流域),但由于高原与平原地区形成泥石流的条件不充分,不作为本文的评价对象,遂不参与模型运算。

泥石流沟数据共计为 1289 个;在本文中,鉴于随机森里模型分类要求,研究将历史泥石流灾害数据单独提取为一类,作为类别属性值,其中有泥石流灾害事件发生的小流域记为“1”,反之记为“0”,通过二分类,将其分为发生过泥石流与未发生泥石流两种小流域类别。

本文初步选取评价指标时,充分考虑影响泥石流发生的控制性因素和影响性因素,将可能存在的影响因素皆纳入了指标筛选的工作中,再通过模型进行指标筛选,数据主要包括地形、地质、灾害、气象、植被、土壤、土地利用、社会经济等,共计 63 项(表 2)。即研究区内 718 个小流域中,每一个小流域均有 63 组数据与之对应。

与传统的泥石流易发性评价指标比较,本文以小流域为评价单元,可以统计评价单元内滑坡面积、村落个数等以栅格为评价单元无法统计的数据,因此指标涵盖的范围更广、数据量更大。

3 泥石流易发性评价

3.1 模型构建

本研究利用 python 软件实现模型运行,使用相应的 RF 软件包。本文以 718 组数据为基础构建随机森林预测模型,首先将 718 组数据随机分为 D1 - D10 共 10 组子集,分别取其中不同 9 组为训练数据集进行建模,然后取剩余 1 组为测试数据集,重复 10 次,同时采用 10 折交叉验证方法验证(10 - foldcross-validation)模型精度,运算后去除贡献率值最低的评价指标后再次循环运算,并记录每一次的 AUC 值,当 AUC 值达到最高时(0.84),参与运算的指标共计 35 项,据此取 35 项指标构建泥石流易发性评价指标体系,并进行泥石流易发性评价,最后对预测结果进行汇总统计,得到 718 个小流域的易发性数据,同时得到 35 项指标的贡献率。模型中几个重要参数设置如下: max_features(最大特征数) = “sqrt”; n_estimators(决策树数目) = “1000”; min_sample_leaf(最小样本叶片) = 20。

3.2 模型验证

在十折交叉验证中,统计每一次验证准确率及其平均值(表 3)。

表 2 研究数据介绍

Tab. 2 Research data introduction

编号	要素	来源
1	泥石流个数	四川省地质环境监测研究所
2	地形	http://earthexplorer.usgs.gov
3		
4		
5		
6		
7	沟床比降	
8	地质与灾害	http://www.csi.ac.cn
9		
10		
11	气象水文	http://www.resdc.cn
12		
13		
14	植被	(https://vito.be/en)
15		
16		
17	土壤	http://www.resdc.cn
18		
19		
20	社会经济	http://sedac.ciesin.columbia.edu
21		
22		
23	道路长度 * (m)	OpenStreetMap
24		
25		

备注:表中带“*”的指标还有二级分类,如下:
地震烈度:(1) < VI度;(2) VI度;(3) VII度;(4) VIII度;(5) ≥ IX度。
岩石硬度:(1)极硬岩;(2)次硬岩;(3)软硬相间;(4)次软岩;(5)极软岩。
土壤质地:(1)砂土;(2)粉土;(3)粘土。
水力侵蚀:(1)微度;(2)轻度;(3)中度;(4)强度;(5)极强度;(6)剧烈。
土地利用类型:(1)水田;(2)旱地;(3)有林地;(4)灌木林;(5)疏林地;(6)其他林地;(7)高覆盖度草地;(8)中覆盖度草地;(9)低覆盖度草地;(10)河渠;(11)湖泊;(12)水库坑塘;(13)永久性冰川雪地;(14)滩涂;(15)滩地;(16)城镇用地;(17)农村居民点;(18)其他建设用地;(19)沙地;(20)戈壁;(21)盐碱地;(22)沼泽地;(23)裸土地;(24)裸岩石砾地;(25)其他。
道路长度:(1)高速公路;(2)国道;(3)省道;(4)县道;(5)铁路。

由表 3 可知,在交叉验证中,模型总体准确性较高,此外,对于“0”事件(无泥石流发生)的预测准确率(0.85)高于“1”(有泥石流发生)的预测准确率(0.77),这是由于“1”样本数据量较少所导致的。

同时,为了进一步证明评价模型的可靠性,计算交叉验证过程中得到的 ROC 曲线及 AUC 值(图 3)。

表 3 交叉验证准确性

Tab. 3 The accuracy of 10-fold cross-validation

序号	测试样本		预测结果		0 准确率	1 准确率
	0	1	0	1		
1	44	27	56	15	0.73	0.56
2	44	27	54	17	0.77	0.63
3	44	28	44	28	1.00	1.00
4	44	28	43	29	0.98	0.96
5	44	28	56	16	0.73	0.57
6	44	28	46	26	0.95	0.93
7	44	28	50	22	0.86	0.79
8	44	28	54	18	0.77	0.64
9	44	28	54	18	0.77	0.64
10	44	28	45	27	0.98	0.96
平均值	/	/	/	/	0.85	0.77

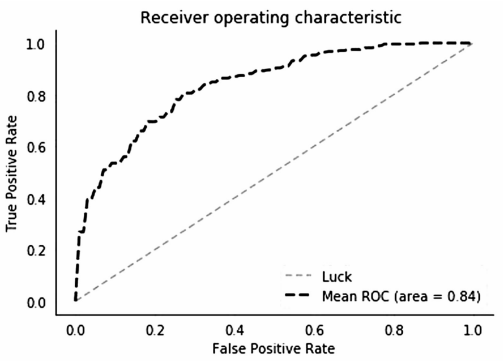


图 3 平均 ROC 曲线图

Fig. 3 Mean ROC curve

由图 3 可知模型十折交叉验证平均 AUC 值达到了 0.84,检验结果十分理想,模型具有很高的准确性。

3.3 易发性评价

模型运行得到 35 项指标的贡献率(图 4),利用 ArcGIS10.2 对易发性评价结果进行处理,通过统计学中常见的自然断点法^[11]划分为 5 个等级,从高到低分别对应极高易发区(0.636 ~ 1)、高易发区(0.454 ~ 0.636)、中易发区(0.275 ~ 0.454)、低易发区(0.123 ~ 0.275)、极低易发区(0 ~ 0.123)5 个等级,并统计贡献率最高的 4 项评价指标在不同易发区内的平均值(图 5)。

3.3.1 评价指标分析

35 项指标贡献率加和等于 1,其中贡献率最高的两项指标为流域高差(0.129)与平均坡度(0.09),极高易发区小流域平均高差达 2841 m,平均坡度 31.69°;而极低易发区小流域平均高差 888 m,平均坡度 17.96°,这两项指标反映了地形对区域泥石流易发性的影响,四川山区泥石流发生的山谷中有 97% 以上的河流的高差在 400 ~ 4000 m 之间^[26],高差大、坡度陡的地形背景下,泥石流启动的势能大,同时还为泥石流提供松散固体物质。滑坡面积贡献率(0.074)位于第三,滑坡个数的贡献率

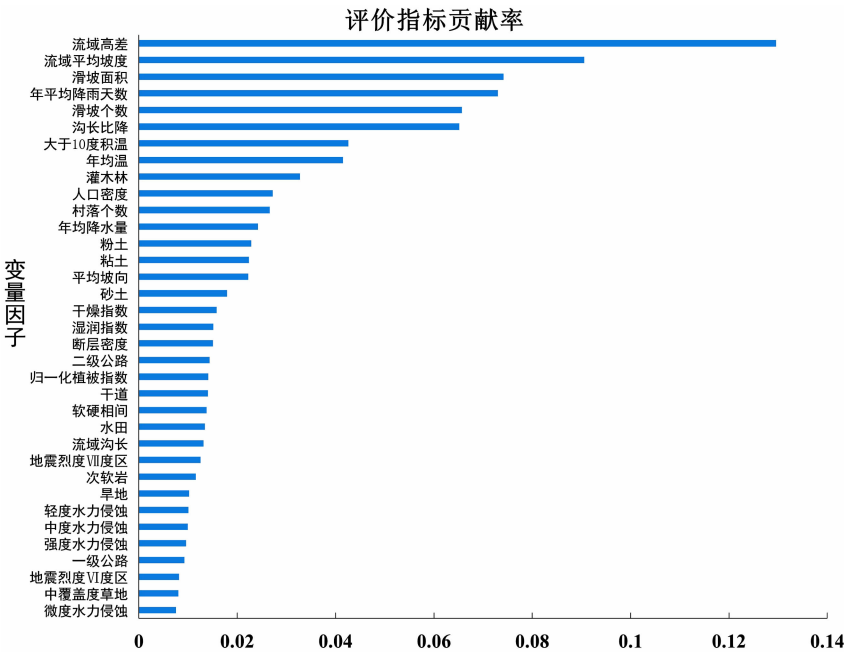


图 4 评价指标贡献率

Fig. 4 Relative importance of the predictors variables

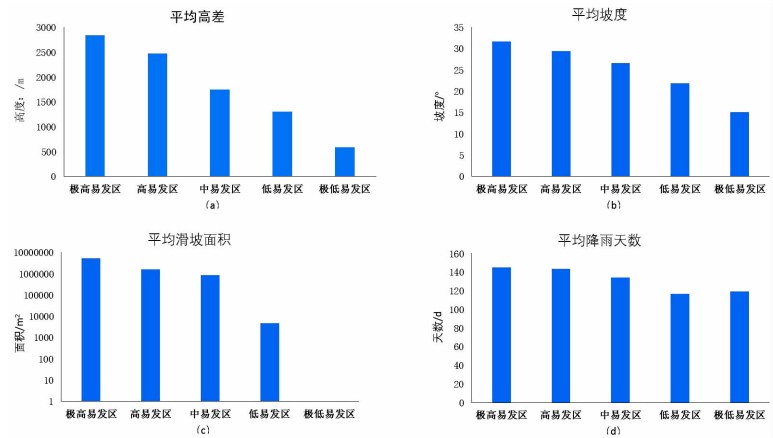


图 5 主要指标不同易发区内平均值

Fig. 5 The mean value of main factors in different susceptibility watersheds

也高达 0.066,这是由于地震后,震区存在大量次生灾害,而以滑坡为主的次生灾害为泥石流的爆发提供了充足的物源。降水是泥石流的触发因子,平均降雨天数(贡献率 0.073)反映了降水对泥石流易发性的影响,由于极低易发区多为平原丘陵,泥石流形成条件不充分,因此降雨天数仍高于低易发区;由于研究区为汶川地震重灾区,地震烈度对泥石流易发性也有一定影响,有地震Ⅶ度区分布的小流域易发性等级较高,评价结果与客观规律相符^[27]。干燥指数(贡献率 0.016)、湿润指数(贡献率 0.015)与大于 10 度积温(贡献率 0.042)反映了气候条件对泥石流易发性的影响,干旱和干湿循环气候条件下,有利于土体开裂,物理风化增强,导致流域内松散物质增多^[30]。除自然环境因子外,研究区内人类活动也不容忽视,流域内的道路密度(贡献率 0.037)、人口密度(贡献率 0.027)与村落个数(贡献率 0.026)几项指标贡献率反映了研究区内人类活动对泥石流易发性存在着较大影响,主要影响方式为不合理的土地利用方式以及道路建设、水电站等工程建设等活动扰动土体、破坏植被,促进了泥石流的发生。

3.3.2 易发性分析

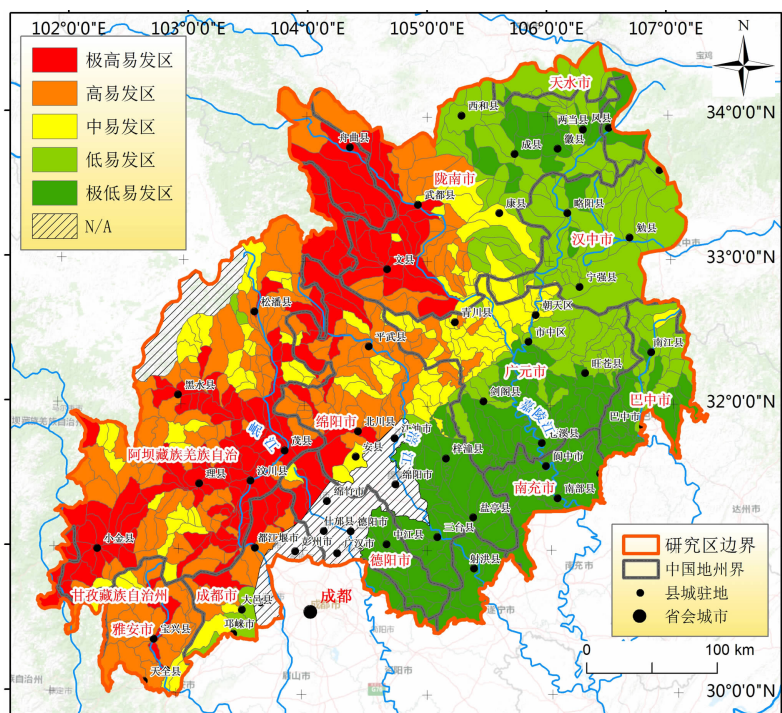
根据评价结果显示(图 6),极低易发区面积最大,共计 163 个小流域,31882.83 km²,占研究区面积 23.86%,易发性较低的区域主要分布于研究区东部,集中于陇南市东部、汉中市、广元市、绵阳市东部、德阳市等丘陵、平原地区;极高易发区占比达到了 22.94%,主要分布于研究区西部,包括陇南市西部、绵阳市西部、阿坝藏族羌族自治州东部、雅安市北部,通过与类似研究成果核对,结果显示,本研究

区四川境内,高、极高易发性区域主要沿岷江流域、大渡河流域分布,与评价结果较为吻合,甘肃南部极高易发区分布面积偏大^[14, 31];高、极高易发性区域地处青藏高原向四川盆地过渡的地形急变带,地形破碎、山高坡陡,泥石流形成条件充分(表 4)。同时,区域内主要有龙门山断裂带分布,因此该区域内地震频发(汶川地震极重灾区分布于该区域),岩层破碎,存在大量次生山地灾害,形成泥石流的物源充足。降水方面,区域内暴雨多发、强降水次数多,容易满足泥石流激发条件。另外区域内岷江及其支流的部分流域都有干旱河谷分布,生态脆弱,水土流失严重。人类活动与泥石流密切相关,而该区域内人口稠密,人类活动强度大,道路、水电站等工程建设密集,一定程度上加速了泥石流发育。上述多因素共同作用下,导致了泥石流高易发性小流域分布于此区域。通过对研究区小流域泥石流易发性空间特征上的定量分析和空间分布上的定性描述,为汶川地震重灾区泥石流监测预警指出了明确的区域重心。

表 4 易发性分级及面积占比

Tab. 4 Classification of the debris flow susceptibility and the ratio of study area

易发性分区	小流域个数	面积/km ²	面积比例/%
极低易发区	163	31 882.83	23.86
低易发区	119	20 866.84	15.61
中易发区	110	19 196.83	14.36
高易发区	173	31 039.51	23.23
极高易发区	153	30 655.84	22.94



N/A:不适用。高原和平原的泥石流形成条件不足,因此这些地区不纳入易发性评价建模范围。

图6 泥石流易发性评价图

Fig.6 Debris flow susceptibility mapping using random forest

4 结论

(1)本文以汶川地震重灾区为研究区,以小流域为评价单元,通过随机森林算法建立了评价模型,模型经交叉验证表现稳定,平均 AUC 值达 0.84,具有很高的准确性,结果表明随机森林模型非常适用于区域泥石流易发性评价研究。

(2)本文初选了 63 项评价指标,以模型 AUC 值变化为基础,筛选 35 项指标构成易发性评价指标体系,并量化了评价指标对泥石流灾害的贡献率,识别出区域内易发性主控因子,结果表明,研究区内流域高差、流域平均坡度、流域内滑坡面积、平均降雨天数对该区域泥石流易发性的影响占主导地位;另外,沟长比降、大于 10° 积温、年均温、人口密度、村落个数、低覆盖度土地利用方式等在泥石流易发性评价中也发挥着重要作用。

(3)易发性评价结果显示,极高易发区占比达到了 22.94%,主要分布于陇南市西部、绵阳市西部、阿坝藏族羌族自治州东部、雅安市北部;研究区内高易发性小流域与地形急变带、断裂带和干旱河谷区域高度重叠。

本文采用机器学习随机森林模型,以小流域为单元,实现了汶川地震重灾区泥石流易发性评价,初步确定了研究区泥石流易发性关键因子构成及其贡献率。研究方法的可行性和成果的可靠性也同时表明机器学习算法可为区域山地灾害相关评价领域研究提供新的方法和思路。

致谢 感谢中国科学院崔鹏院士在论文撰写过程中提供的重要思路与建议,感谢中国地震局地质研究所许冲研究员提供相关数据。

参考文献 (References)

- [1] 崔鹏, 庄建琦, 陈兴长, 等. 汶川地震震后泥石流活动特征与防治对策[J]. 四川大学学报(工程科学版), 2010, 42(5): 10-19 [CUI Peng, ZHUANG Jianqi, CHEN Xingchang. Characteristics and countermeasures of debris flow in Wenchuan area after the earthquake [J]. Journal of Sichuan University (Engineering Science Edition), 2010, 42(5): 10-19]
- [2] BLAHUT J, HORTON P, STERLACCHINI S, et al. Debris flow hazard modelling on medium scale, Valtellina di Tirano, Italy [J]. Natural Hazards and Earth System Science, 2010, 10(11): 2379-2390
- [3] CARRARA A, CROSTA G, FRATTINI P. Comparing models of debris-flow susceptibility in the alpine environment [J]. Geomorphology, 2008, 94(3/4): 353-378

- [4] KAPPES M S, MALET J P, REMAÎTRE A, et al. Assessment of debris-flow susceptibility at medium-scale in the Barcelonnette Basin, France [J]. *Natural Hazards and Earth System Science*, 2011, **11**(2): 627–641
- [5] 阮沈勇, 黄润秋. 基于 GIS 的信息量法模型在地质灾害危险性区划中的应用[J]. 成都理工大学学报(自然科学版), 2001, **28**(1): 89–92 [RUAN Shenyong, HUANG Runqiu. Application of GIS-based information model on assessment of geological hazards risk [J]. *Journal of Chengdu University of Technology*, 2001, **28**(1): 89–92]
- [6] 徐艳琴, 白淑英, 徐永明. 基于两种方法的攀西泥石流易发性评价对比分析[J]. 水土保持研究, 2018, **25**(3): 285–291 [XU Yanqin, BAI Shuying, XU Yongming. Comparative analysis of debris flow susceptibility assessment based on two methods in Panxi district [J]. *Research of Soil and Water Conservation*, 2018, **25**(3): 285–291]
- [7] AUTHOR M K A C, GUPTA A S D, GUPTA R P. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas [J]. *International Journal of Remote Sensing*, 2004, **25**(3): 559–572
- [8] 唐川, 朱大奎. 基于 GIS 技术的泥石流风险评价研究[J]. 地理科学, 2002, **7**(3): 300–304 [TANG Chuan, ZHU Dakui. Assessment of debris flow risk of Yunnan Province by using GIS [J]. *Scientia Geographica Sinica*, 2002, **7**(3): 300–304]
- [9] 王哲, 易发成. 基于层次分析法的绵阳市地质灾害易发性评价[J]. 自然灾害学报, 2009, **18**(1): 14–23 [WANG Zhe, YI Facheng. AHP-based evaluation of occurrence easiness of geological disasters in Mianyang City [J]. *Journal of Natural Disasters*, 2009, **18**(1): 14–23]
- [10] 徐继维, 张茂省, 范文. 地质灾害风险评估综述[J]. 灾害学, 2015, **30**(4): 130–134 [XU Jiwei, ZHANG Maosheng, FAN Wen. An overview of geological disaster risk assessment [J]. *Journal of Catastrophology*, 2015, **30**(4): 130–134]
- [11] 邹强, 崔鹏, 杨伟. G318 川藏公路段泥石流危险性评价[J]. 山地学报, 2013, **31**(3): 342–348 [ZOU Qiang, CUI Peng, YANG Wei. Hazard assessment of debris flows along G318 Sichuan-Tibet highway [J]. *Journal of Mountain Science*, 2013, **31**(3): 342–348]
- [12] 刘希林. 区域泥石流风险评价研究[J]. 自然灾害学报, 2000, **9**(1): 54–61 [LIU Xilin. Regional risk assessment on debris flow [J]. *Journal of Natural Disasters*, 2000, **9**(1): 54–61]
- [13] 邹强, 唐建喜, 李淑松, 等. 基于水文响应单元的泥石流灾害易发性分区方法[J]. 山地学报, 2017, **35**(4): 496–505 [ZOU Qiang, TANG Jianxi, LI Shusong, et al. Susceptibility assessment method of debris flows based on hydrological response unit [J]. *Mountain Research*, 2017, **35**(4): 496–505]
- [14] 熊俊楠, 韦方强, 刘志奇. 四川省小流域泥石流危险性评价[J]. 地球信息科学学报, 2017, **19**(12): 1604–1612 [XIONG Junnan, WEI Fangqiang, LIU Zhiqi. Hazard assessment of debris flow in Sichuan Province [J]. *Journal of Geo-Information Science*, 2017, **19**(12): 1604–1612]
- [15] PHAM B T, PRADHAN B, TIEN B D, et al. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India) [J]. *Environmental Modelling & Software*, 2016, **84**: 240–250
- [16] IVERSON L R, PRASAD A M, MATTHEWS S N, et al. Estimating potential habitat for 134 eastern US tree species under six climate scenarios [J]. *Forest Ecology & Management*, 2008, **254**(3): 390–406
- [17] 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), 2017, **43**(7): 1–7 [LIU Jian, LI Shulin, CHEN Tao. Landslide susceptibility mapping by random forest model [J]. *Geomatics and Information Science of Wuhan University*, 2017, **43**(7): 1–7]
- [18] 李亨, 田原, 邹伦, 等. 基于随机森林方法的滑坡灾害危险性区划[J]. 地理与地理信息科学, 2014, **30**(6): 25–30 [LI Ting, TIAN Yuan, WU Lun. Landslide susceptibility mapping using random forest [J]. *Geography and Geo-Information Science*, 2014, **30**(6): 25–30]
- [19] TRIGILA A, IADANZA C, ESPOSITO C, et al. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampieri (NE Sicily, Italy) [J]. *Geomorphology*, 2015, **249**: 119–136
- [20] KERN A N, ADDISON P, OOMMEN T, et al. Machine Learning Based Predictive Modeling of Debris Flow Probability Following Wildfire in the Intermountain Western United States [J]. *Mathematical Geosciences*, 2017, **49**(6): 717–735
- [21] 赖成光, 陈晓宏, 赵仕威, 等. 基于随机森林的洪灾风险评价模型及其应用[J]. 水利学报, 2015, **46**(1): 58–66 [LAI Chengguang, CHEN Xiaohong, ZHAO Shiwei. A flood risk assessment model based on random forest and its application [J]. *Journal of Hydraulic Engineering*, 2015, **46**(1): 58–66]
- [22] PHAM B T, SHIRZADI A, TIEN Bui D, et al. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India [J]. *International Journal of Sediment Research*, 2017
- [23] 原立峰. 基于 SVM 的泥石流危险度评价研究[J]. 地理科学, 2008, **28**(2): 296–300 [YUAN Lifeng. Debris flow hazard assessment based on SVM [J]. *Scientia Geographica Sinica*, 2008, **28**(2): 296–300]
- [24] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经贸大学 2014 [CAO Zhengfeng. Study on optimization of random forests algorithm [D]. Beijing: Capital University of Economics and Business, 2014]
- [25] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, **26**(3): 32–38 [FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests [J]. *Statistics & Information Forum*, 2011, **26**(3): 32–38]
- [26] 王涛. 汶川地震重灾区地质灾害危险性评估研究[D]. 北京: 中国地质科学院, 2010: 13–14 [WANG Tao. Study on seismic landslide hazard assessment in Wenchuan earthquake severely afflicted area [D]. Beijing: Chinese Academy of Geological

- Sciences, 2010; 13 – 14]
- [27] 张春山. 汶川地震灾区次生灾害隐患排查与工程设计示范 [M]. 北京: 中国大地出版社, 2009; 11 – 25 [ZHANG Chunshan. Investigation and engineering design demonstration of hidden danger of secondary disasters in Wenchuan earthquake stricken area [M]. Beijing; China Land Press, 2009; 11 – 25]
- [28] 许冲. 汶川地震滑坡详细编录及其与全球其他地震滑坡事件对比 [J]. 科技导报, 2012, 30 (25): 18 – 26 [XU Chong. Detailed inventory of landslides triggered by the 2008 Wenchuan earthquake and its comparison with other earthquake events in the world [J]. Science & Technology Report, 2012, 30 (25): 18 – 26]
- [29] 康志成. 中国泥石流研究 [M]. 北京: 科学出版社, 2000; 1 – 4
- [Kang Z, Lee C, Law K, et al. Debris Flow Research in China [M]. Beijing; Science Press, 2004; 1 – 4]
- [30] 陈宁生, Javed Iqbal, 刘丽红. “8·3”鲁甸地震小震大灾背景下次生泥石流的易发性分析 [J]. 成都理工大学学报 (自然科学版), 2016, 43 (1): 102 – 108 [CHEN Ningsheng, JAVED IQBAL, LIU Lihong. Tendency prediction of debris flow triggered by “8·3” Ludian earthquake of Yunnan, China [J]. Journal of Chengdu University of Technology, 2016, 43 (1): 102 – 108]
- [31] 张若琳, 孟晖, 连建发. 基于 GIS 的中国泥石流易发性评价 [J]. 成都理工大学学报 (自然科学版), 2013, 40 (4): 379 – 386 [ZHANG Ruolin, MENG Hui, LIAN Jianfa. Susceptibility evaluation of debris-flow in China based on GIS [J]. Journal of Chengdu University of Technology, 2013, 40 (4): 379 – 386]

Debris Flows Susceptibility Assessment in Wenchuan Earthquake Areas Based on Random Forest Algorithm Model

LIU Yongyao¹, DI Baofeng^{2,1*}, ZHAN Yu¹, Stamatopoulos C. A.³

(1. College of Architecture & Environment, Sichuan University, Chengdu 610065, China;

2. Institute for Disaster Management and Reconstruction, Sichuan University, Chengdu 610207, China;

3. Stamatopoulos and Associates Co. and Hellenic Open University, 5 Isavron Str, 11471 Athens, Greece)

Abstract: As debris flow is one of the major mountain disasters in China, the research on debris flow susceptibility is of great significance for monitoring and risk management. With respect to debris flow susceptibility, identifying the key variables and their relative importance are critical for developing warning systems and managing risk. This paper focused on the most important variables and discussed the spatial characteristics of debris flow susceptibility in the Wenchuan earthquake-hit area by using a machine learning algorithm (i. e., random forest). Watershed was selected as the basic assessment unit, accompanied with a multiple-factor index system, to model debris flow susceptibility. The results of model evaluation showed adequate stability and accuracy, with the 10-fold cross-validation average area-under-curve (AUC) values of 0.84, indicating the random forest model was suitable to evaluate regional debris flow susceptibility. In this paper, on the basis of the change of AUC value, 35 of 63 evaluation indicators were preferred to form a susceptibility evaluation index system, and used to identify the main control factors of debris flow susceptibility in the region. We found that the elevation difference, average slope, landslide, and average rainy days played the most important role in determining the regional debris flow susceptibility. The results of susceptibility assessment indicated that the highest susceptibility watersheds reached 22.94%, was mainly distributed in the western part of the study area. The spatial distributions of the debris flow susceptibility displayed the high-susceptibility watersheds were highly coupled with the locations of the topographical extreme belt, fault zone, seismic belt, and dry valleys. The regional debris flow susceptibility evaluated by the random forest model developed in this study could be applied as a reference for regional-scale disaster susceptibility assessment, and the results also reflected that machine learning algorithm could provide a new method and idea for regional debris flows susceptibility assessment.

Key words: debris flow; susceptibility; random forest algorithm; key variables; Wenchuan earthquake