

任涛, 田国亮, 宁志杰, 等. 基于地理探测器和随机森林的岩溶塌陷易发性评价[J]. 灾害学, 2023, 38(3): 227–234. [REN Tao, TIAN Guoliang, NING Zhijie, et al. Evaluation of Karst Collapse Susceptibility Based on Geodetector and Random Forests [J]. Journal of Catastrophology, 2023, 38(3): 227–234. doi: 10.3969/j.issn.1000-811X.2023.03.035.]

基于地理探测器和随机森林的岩溶塌陷易发性评价*

任涛^{1,2,3}, 田国亮², 宁志杰⁴, 周爱红³, 李宽², 陈石²

(1. 河北省煤田地质局环境地质调查院, 河北 石家庄 050085; 2. 山东省地质矿产勘查开发局第六地质大队, 山东 威海 264209; 3. 河北地质大学 城市地质与工程学院, 河北 石家庄 050031; 4. 中国地质大学(北京) 中国地质科学院, 北京 100083)

摘要: 岩溶塌陷受多种因素影响, 其发生具有突然性, 分布具有空间分异性。为合理的对怀化地区岩溶塌陷进行易发性评价, 采用地理探测器和随机森林模型对怀化地区开展岩溶塌陷易发性评价, 分析了岩溶塌陷的主要影响因素和影响因素之间的相互作用, 将 Geodetector-IV-RF、IV-RF 和 IV 模型的易发性评价结果进行对比分析。结果表明: 岩溶发育程度、岩溶水位与基岩面关系和岩溶塌陷密度这 3 个影响因素对岩溶塌陷的解释程度较高; 岩溶塌陷影响因素两两之间相互作用类型为双因子增强或非线性增强, 都对岩溶塌陷发生起增强作用; Geodetector-IV-RF 模型的评价结果精度最高, 分区结果可以为研究区岩溶塌陷的防治工作提供辅助预警作用。

关键词: 岩溶塌陷; 地理探测器; 随机森林; 易发性评价

中图分类号: P642.26; X43; X915.5 **文献标志码:** A **文章编号:** 1000-811X(2023)03-0227-08

doi: 10.3969/j.issn.1000-811X.2023.03.035

岩溶塌陷是岩溶发育地区典型的地质灾害, 其发育具有隐蔽性, 发生具有突然性^[1-2]。近年来, 随着城市建设不断发展, 岩溶塌陷频发, 对地表建筑物和交通道路等造成破坏, 严重威胁人们的生命财产安全。城镇规划和工程建设防灾减灾的需要对岩溶塌陷的研究提出了新的要求, 岩溶塌陷易发性评价已逐渐成为岩溶塌陷防治研究的热点^[3-4]。

影响岩溶塌陷的因素较多, 并且影响因素与岩溶塌陷之间存在复杂的非线性映射关系, 很难用简单的数学模型或定性方法来描述。随着计算机技术的发展, 神经网络法^[5]、灰色理论^[6]、模糊综合评价^[7]和层次分析法^[8]等方法广泛应用于岩溶塌陷易发性评价, 这些方法在一定范围内获得的研究效果显著, 但也存在着一些不足。神经网络法受样本数量影响较大, 且训练时间长, 易陷入局部极值问题; 灰色理论受数据波动影响较大, 预测精度有限; 模糊综合评价法在选择隶属度函数方面缺乏依据, 计算较复杂; 当影响因素数量较多时, 层次分析法难以确定各因素权重, 并且该方法主观性较强。岩溶塌陷具有空间分异性, 成因复杂, 影响因素繁多, 各个影响因素对岩溶塌陷的解释程度不同, 目前在岩溶塌陷指标赋权中应用较多的层次分析法受主观判断影响较多, 未能基于岩溶塌陷的空间分异性来分析各种影响因素的权重, 并且相当一部分影响因素是定性数据, 对其进行合理的赋值存在一定难度。因此, 有必要根据现有的计算模型, 探索预测岩溶

塌陷易发性评价的新方法。

地理探测器 (Geographical detector, Geodetector)^[9]是探测空间分异性的主要方法之一, 能够度量自变量对因变量空间分异性的解释力, 并且适用于类型量和数值数据, 该方法得到了较多应用。王桂林等^[10]利用地理探测器对选取的 13 个岩溶塌陷影响因素进行因子探测, 筛选出对岩溶塌陷的空间分异性解释力较大的 5 个影响因素来进行岩溶塌陷易发性评价; 支泽民等^[11]运用地理探测器对西藏江达县的 85 处滑坡和影响滑坡稳定性的 8 个影响因素进行了探测, 分析了江达县滑坡稳定性的主要影响因素以及影响因素两两之间的相互作用对滑坡稳定性的影响; 刘婷等^[12]利用地理探测器初步确定了人工切坡造成滑坡的主要影响因素和权重值。近年来, 随着机器学习方法在地质灾害易发性预测方面的不断应用和发展, 基于频率比和信息量 (information value model, IV) 等方法的机器学习耦合模型的研究成为了易发性预测建模领域发展的一个新方向。随机森林模型 (random forests model, RF) 由于其预测效果好, 具有较强抗干扰能力和很强的建模适应能力等优点^[13-14], 成为了一种具有较高准确性和可靠性的地质灾害易发性预测模型。孙德亮^[15]论证了随机森林在重庆市奉节县滑坡易发性评价中优于逻辑回归和人工神经网络模型; 邓念东等^[16]将信息量-随机森林耦合模型用于陕西省神木市崩塌易发性评价, 相比单一模型, 耦合模型的预测精度更高。

* 收稿日期: 2022-09-03 修回日期: 2023-02-09

基金项目: 国家自然科学基金资助项目 (41807231); 河北省自然科学基金项目 (D2019403182)

第一作者简介: 任涛 (1996-), 男, 汉族, 河北邢台人, 硕士, 主要从事工程地质与地质灾害方向的研究。E-mail: 2082108929@qq.com

通信作者: 田国亮 (1987-), 男, 汉族, 河北沧州人, 工程师, 主要从事岩土工程勘察研究方向的研究。E-mail: 634348643@qq.com

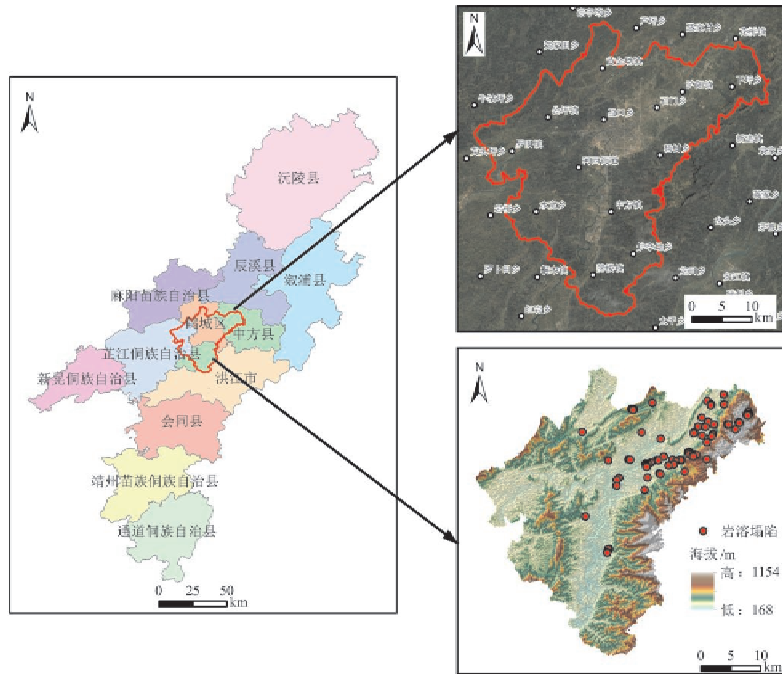


图1 研究区地理位置和岩溶塌陷点分布示意图

(注:该图基于国家测绘地理信息局标准地图服务网站下载的标准地图(审图号:湘S(2022)034号)制作,底图无修改,下同)

本文以湖南省岩溶发育的怀化地区作为研究区,结合地理探测器探测空间分异性的能力,并基于随机森林模型,建立了基于地理探测器和信息量-随机森林耦合(Geodetector-IV-RF)的岩溶塌陷易发性预测模型。同时,利用地理探测器来对岩溶发育程度、到断层的距离、土层结构、土层厚度、岩溶水位波动幅度、岩溶水位与基岩面关系、地下水开采强度、岩溶塌陷密度等8个岩溶塌陷影响因素进行因子探测和交互作用探测,筛选出岩溶塌陷的主要影响因素,通过对比分析IV模型、IV-RF模型和Geodetector-IV-RF模型这三种模型的预测精度,选择最合理的岩溶塌陷易发性评价分区结果以期对怀化地区的岩溶塌陷防治提供科学参考。

1 研究区概况

研究区位于怀化市南部,范围主要包括怀化市鹤城区、中方县和芷江侗族自治县的部分乡镇,区域东西长约49.61 km,南北宽约44.36 km,总面积约1 122 km²。地势呈东南和西北部高,中部低,区内海拔最高为1 154 m,最低为168 m。怀化地区属温湿亚热带季风湿润气候,雨量充足,年均降雨量为1 326.3 mm。研究区位于沅麻盆地和雪峰冲断带交汇地区,地质构造复杂,断层发育较多。区内地层出露较齐全,碳酸盐岩地层分布较广,岩溶发育,岩溶水主要接受大气降雨和外围碎屑岩山区的外缘水补给。研究区共发生岩溶塌陷群49处,岩溶塌陷坑111个,为研究区内主要类型地质灾害^[17](图1)。

2 研究方法

地理探测器^[18-19]是探测空间分异性,以及揭示其背后驱动力的一种统计学方法。该方法无线

性假设,对多变量共线性免疫,地理探测器的因子探测器和交互作用探测器可以用来分别探测影响因素的解释力以及影响因素两两之间的交互作用。随机森林^[20]是一种包含多个决策树的集成学习算法。随机森林算法利用Bootstrap法重采样生成多个训练集,每一个训练集对应一个决策树,在构建决策树时随机选取训练集数据属性,各个决策树之间相互独立,将全部决策树输出类型的众数或平均值作为最终输出。

利用地理探测器对岩溶塌陷易发性评价指标进行空间分异性探测,剔除对岩溶塌陷空间分布解释力不强的评价指标,建立新的岩溶塌陷易发性评价指标体系。对比多组不同灾害点与非灾害点比例样本训练下的模型性能,来选择合适比例的训练样本。利用训练好的Geodetector-IV-RF模型对研究区进行岩溶塌陷易发性评价,将其评价分区结果与IV模型和IV-RF模型的评价分区结果进行对比分析,选择最佳岩溶塌陷易发性预测结果。

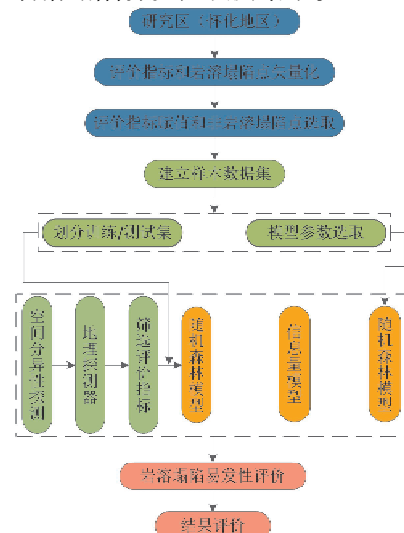


图2 随机森林模型预测岩溶塌陷易发性的流程图

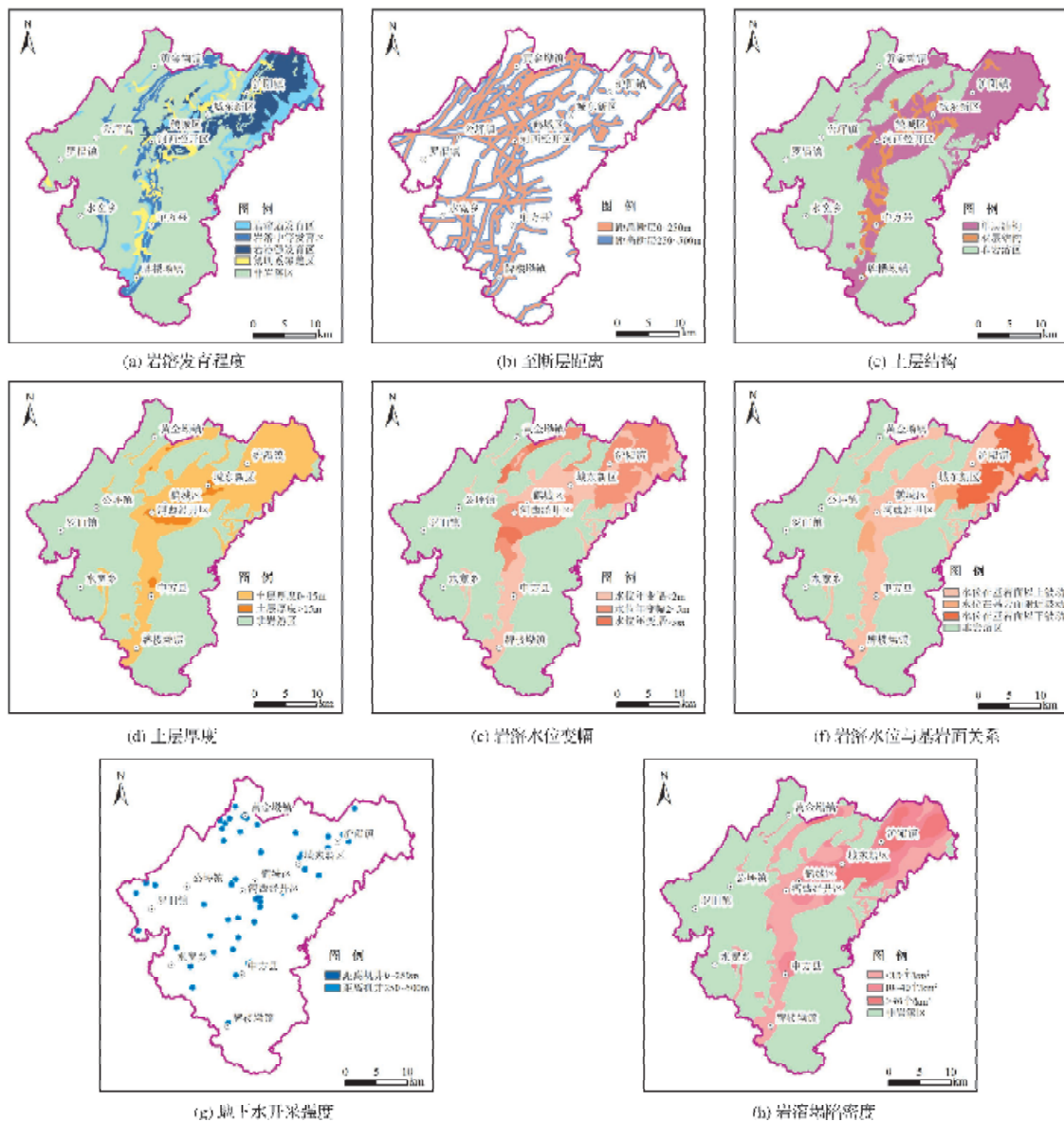


图3 研究区岩溶塌陷各影响因素分区图

3 岩溶塌陷影响因素敏感性分析

3.1 主要影响因素的确定

研究区岩溶塌陷的发育和发生主要受下伏碳酸盐岩岩溶发育程度、地质构造条件、第四系覆盖层特征、地下水动力条件等因素的影响,同时,岩溶塌陷发育强度是以上因素对岩溶塌陷影响的综合体现。因此,本次岩溶塌陷易发性评价将以下8个影响因素作为易发性评价指标:①岩溶发育条件,包括岩溶发育程度(X_1)和至断层距离(X_2);②第四系覆盖层条件,包括土层结构(X_3)和土层厚度(X_4);③地下水动力条件,包括岩溶水位变幅(X_5)、地下水位与基岩面的关系(X_6)以及地下水开采强度(X_7);④岩溶塌陷发育强度以岩溶塌陷密度(X_8)表示^[21]。将各岩溶塌陷影响因素矢量图转为25 m×25 m分辨率的栅格图(图3),将1 122 km²的研究区划分为1 774行1 984列,共1 795 200个栅格。选取85处典型岩溶塌陷点作为灾害点样本。

为了更为合理的选取非岩溶塌陷点,采用信息量法对研究区进行岩溶塌陷易发性评价,将评价结果分为高、较高、中、较低和低5个等级,在较低和低易发区中随机生成非岩溶塌陷点(为节省篇幅,在此不详细说明评价过程),建立包含灾害点和非灾害点的样本点库。利用地理探测器对岩溶塌陷进行因子探测来筛选出影响岩溶塌陷分布特征的主导因素,考虑到样本的不均衡性可能会对探测结果造成影响,随机生成灾害点数量0.5倍、1倍、2倍、…、5倍的非灾害点,将灾害点分别和不同倍数的非灾害点组成6组样本点库,将样本点和岩溶塌陷影响因素相连接,建立6组样本数据库。

基于建立的6组样本数据,利用地理探测器进行因子探测来度量岩溶塌陷与影响因素之间空间分布的相似性,计算得到各个影响因素的统计量 q 值(图4),由图4可知6组样本数据的探测结果不尽相同,但各组探测结果中 q 值大于0.1的影响因素是相同的,分别是岩溶发育程度(X_1)、岩溶水位与基岩面关系(X_6)和岩溶塌陷密度(X_8),这三个影响因素的平均 q 值分别为0.352、0.126和

0.854, 其平均 q 值之和占全部影响因素平均 q 值和的 85.01%。至断层距离(X_2)、土层结构(X_3)、土层厚度(X_4)、岩溶水位变幅(X_5)和地下水开采强度(X_7)这 5 个影响因素的 q 值均小于 0.1, 对怀

化地区岩溶塌陷的影响程度较小。根据地理探测器的因子探测结果, 岩溶发育程度(X_1)、岩溶水位与基岩面关系(X_6)和岩溶塌陷密度(X_8)这三个影响因素是怀化地区岩溶塌陷主要影响因素。

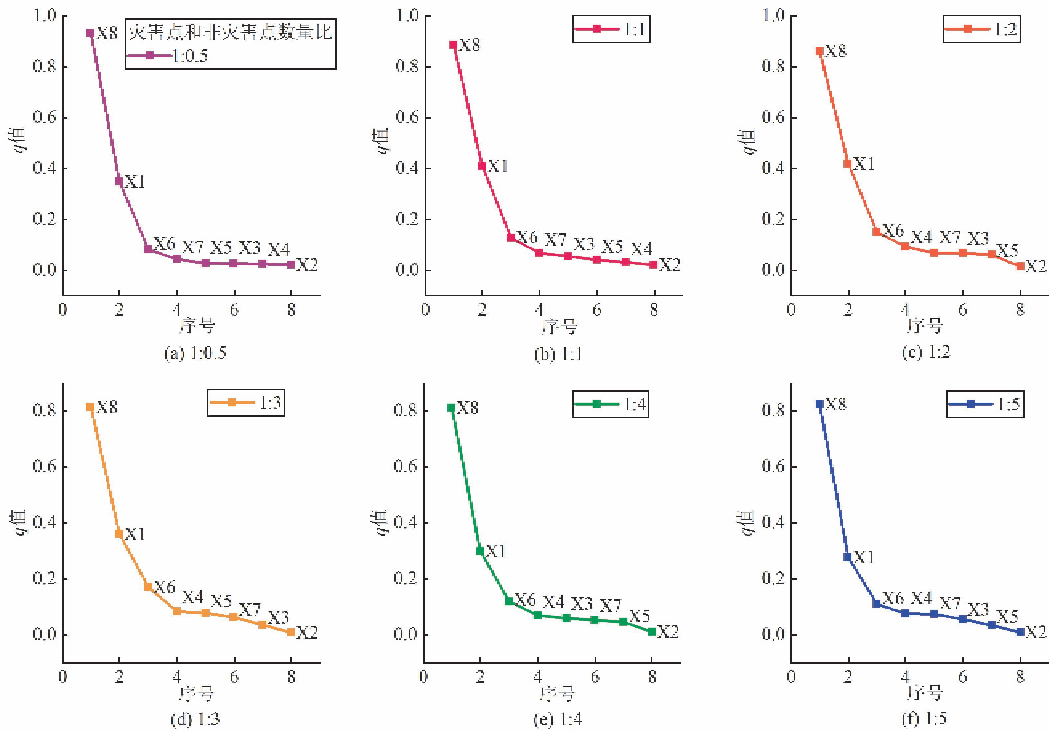


图4 岩溶塌陷影响因素 q 值

表1 不均衡样本下怀化地区岩溶塌陷8个影响因素的交互作用

交互因子	灾害点/非灾害点 1:1		灾害点/非灾害点 1:3		灾害点/非灾害点 1:5	
	交互 q 值	交互作用	交互 q 值	交互作用	交互 q 值	交互作用
构造 \cap 地下水开采强度	0.121	非线性增强	0.093	非线性增强	0.096	非线性增强
构造 \cap 土层厚度	0.067	非线性增强	0.101	非线性增强	0.100	非线性增强
构造 \cap 土层结构	0.107	非线性增强	0.050	非线性增强	0.068	双因子增强
构造 \cap 岩溶发育程度	0.437	双因子增强	0.398	非线性增强	0.304	非线性增强
构造 \cap 岩溶水位变幅	0.070	双因子增强	0.097	非线性增强	0.050	非线性增强
构造 \cap 岩溶塌陷密度	0.891	双因子增强	0.822	双因子增强	0.824	双因子增强
构造 \cap 岩溶水位与基岩面关系	0.161	非线性增强	0.200	非线性增强	0.136	非线性增强
地下水开采强度 \cap 土层厚度	0.086	双因子增强	0.118	双因子增强	0.121	双因子增强
地下水开采强度 \cap 土层结构	0.096	双因子增强	0.099	双因子增强	0.144	非线性增强
地下水开采强度 \cap 岩溶发育程度	0.453	双因子增强	0.416	双因子增强	0.331	双因子增强
地下水开采强度 \cap 岩溶水位变幅	0.136	非线性增强	0.174	非线性增强	0.127	非线性增强
地下水开采强度 \cap 岩溶塌陷密度	0.889	双因子增强	0.834	双因子增强	0.842	双因子增强
地下水开采强度 \cap 岩溶水位与基岩面关系	0.231	非线性增强	0.257	非线性增强	0.199	非线性增强
土层厚度 \cap 土层结构	0.117	非线性增强	0.113	双因子增强	0.145	非线性增强
土层厚度 \cap 岩溶发育程度	0.413	双因子增强	0.383	双因子增强	0.304	双因子增强
土层厚度 \cap 岩溶水位变幅	0.113	非线性增强	0.174	非线性增强	0.134	非线性增强
土层厚度 \cap 岩溶塌陷密度	0.887	双因子增强	0.819	双因子增强	0.830	双因子增强
土层厚度 \cap 岩溶水位与基岩面关系	0.182	非线性增强	0.264	非线性增强	0.198	非线性增强
土层结构 \cap 岩溶发育程度	0.421	双因子增强	0.384	双因子增强	0.302	双因子增强
土层结构 \cap 岩溶水位变幅	0.149	非线性增强	0.156	非线性增强	0.120	非线性增强
土层结构 \cap 岩溶塌陷密度	0.891	双因子增强	0.839	双因子增强	0.849	双因子增强
土层结构 \cap 岩溶水位与基岩面关系	0.240	非线性增强	0.263	非线性增强	0.223	非线性增强
岩溶发育程度 \cap 岩溶水位变幅	0.463	非线性增强	0.393	双因子增强	0.291	双因子增强
岩溶发育程度 \cap 岩溶塌陷密度	0.958	双因子增强	0.970	双因子增强	0.972	双因子增强
岩溶发育程度 \cap 岩溶水位与基岩面关系	0.431	双因子增强	0.383	双因子增强	0.331	双因子增强
岩溶水位变幅 \cap 岩溶塌陷密度	0.899	双因子增强	0.842	双因子增强	0.834	双因子增强
岩溶水位变幅 \cap 岩溶水位与基岩面关系	0.229	非线性增强	0.284	非线性增强	0.222	非线性增强
岩溶塌陷密度 \cap 岩溶水位与基岩面关系	0.899	双因子增强	0.869	双因子增强	0.879	双因子增强

3.2 影响因素的交互作用

地理探测器的交互作用探测功能可以用来分析影响因子两两相互作用时对岩溶塌陷分异性的解释程度是增强还减弱,线性还是非线性等。借助交互作用探测器来得到不同比例样本的8个岩溶塌陷影响因素的交互作用(表1),由于篇幅限制在此列出部分比例样本的交互作用分析结果。由表1可以看出,岩溶塌陷各影响因素的交互作用对岩溶塌陷的解释力高于单一因子的解释力,影响因素交互作用对岩溶塌陷的影响是增强的。相同2个影响因素不同灾害点和非灾害点比例样本得出的交互作用类型不尽相同,但都是增强关系。8个影响因素的平均交互综合作用从高到低为:岩溶塌陷密度(6.178) > 岩溶发育程度(3.279) > 岩溶水位与基岩面关系(2.329) > 岩溶水位变幅(1.983) > 土层结构(1.926) > 地下水开采强度(1.926) > 土层厚度(1.881) > 构造(1.759)。其中,不同比例样本的岩溶塌陷密度和其他影响因素的交互作用 q 值均大于0.80,交互综合作用最强,交互类型均为双因子增强;其次,岩溶发育程度与其他因素的交互作用 q 值均在0.29之上,岩溶发育程度的交互综合作用居于第二,交互作用既有双因子增强也有非线性增强;岩溶水位与基岩面关系与其他因素的交互作用 q 值均大于0.10,岩溶水位与基岩面关系和岩溶塌陷密度、岩溶发育程度的交互作用为双因子增强,和其他因素的交互作用为非线性增强,交互综合作用居第三。为了减小弱解释力影响因素对岩溶塌陷易发性预测模型性能的影响,综合因子探测和交互作用探测的分析结果,将岩溶发育程度、岩溶水位与基岩面关系和岩溶塌陷密度这三个影响因素作为岩溶塌陷易发性评价指标。

4 岩溶塌陷易发性评价与分析

4.1 易发性建模过程

训练样本中灾害点和非灾害点比例的不同可能会影响岩溶塌陷易发性预测模型的性能,为了合理选择该比例,将构建好的不同比例样本按照7:3随机划分为训练样本和测试样本,其中测试样本保持相同均为50组。采用MATLAB语言编写RF模型分别对6组样本进行训练和测试,以1:1比例样本为例进行模型参数寻优说明。RF模型的参数较少,其中 $ntree$ (决策树的数目)和 $mtry$ (构建决策树分支随机选取的变量个数)这两个参数是影响RF模型性能最重要的参数。 $mtry$ 一般取全部变

量个数的平方根,将 $mtry$ 的值设置为1或2,将 $ntree$ 的取值范围设置为0~500,在此建立100个随机森林模型,取所有模型决定系数和均方误差平均值作为当前决策树数量下的决定系数和均方误差,逐次改变 $mtry$ 的值来训练模型,根据模型对测试样本的预测结果(图5)来选择最合适的参数值组合。由图7可知, $ntree$ 值在50~500范围时,模型预测结果比较平稳,由图5a和图5b可知 $mtry$ 值为2时,模型的拟合程度比 $mtry$ 为1时高。选取 $ntree$ 为54, $mtry$ 为2作为最佳参数组合,此时 R^2 为0.942, MSE 为0.015,模型预测精度最高,可以满足实际工程需求。将训练好的岩溶塌陷易发性预测模型对测试样本进行预测,模型预测结果的 R^2 为0.959, MSE 为0.012。

对其余5个训练好的RF模型进行测试,得到预测结果(表2),从测试样本预测结果整体来看(表2),不同比例样本的模型预测结果不同,1:0.5比例样本的模型测试效果最差,如果以预测值0.5为是否判别灾害点的界限,该组样本训练的模型误判3组非灾害点;1:1至1:4比例样本模型的测试结果随着灾害点和非灾害点比例的增大,测试样本预测结果的 R^2 和 MSE 也逐渐增大,但1:2、1:3和1:4比例样本的模型测试效果整体很接近,且准确率均为100%,差距不明显;1:5比例样本的模型预测效果低于1:2比例样本的模型测试效果。从灾害点和非灾害点样本预测结果分别来看,1:0.5比例样本的灾害点样本的 MSE 小于其他四组样本,但非灾害点样本的 MSE 又远大于其他四组样本;而其他五组比例样本训练模型的灾害点和非灾害点样本预测的准确率均为100%。从地质灾害的防御角度来看,模型对灾害点的预测值越接近1越好,预测模型在不对非灾害点进行误判的前提下,模型对非灾害点易发性的预测值稍大会适当增大易发性,从结果看是偏安全的。综合以上分析,怀化地区岩溶塌陷易发性预测模型的训练样本灾害点与非灾害点比例取1:1比较合适。

4.2 易发性预测

基于GIS平台将研究区栅格转点,提取研究区所有栅格对应的岩溶发育程度、岩溶水位与基岩面关系和岩溶塌陷密度这三个影响因素的信息量值构建预测样本集作为模型输入,利用测试好的RF模型对研究区进行岩溶塌陷易发性预测。将岩溶塌陷易发性预测结果链接到对应的栅格上,采用自然断点法将研究区分为高、较高、中、较低和低易发5个等级,得到怀化地区岩溶塌陷易发性分区图(图6a)。

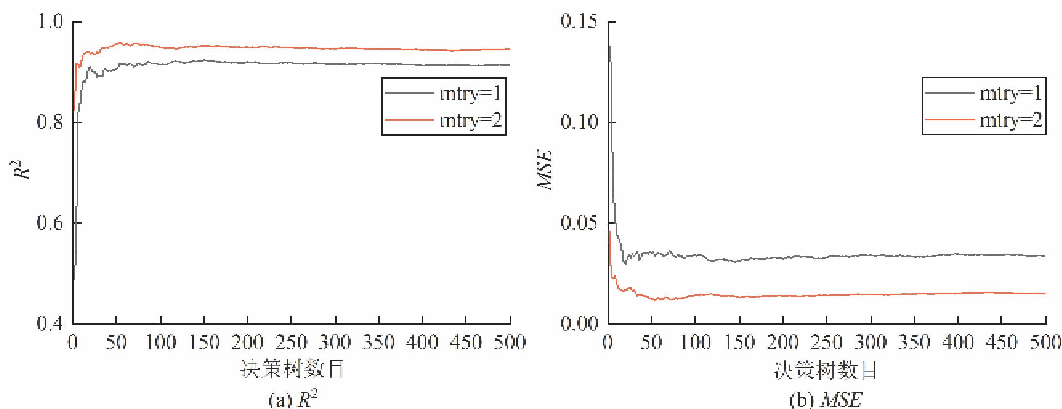


图5 RF模型参数对性能的影响

表 2 不同灾害点与非灾害点比例的岩溶塌陷样本的预测结果

灾害点与非灾害点比例		1:0.5	1:1	1:2	1:3	1:4	1:5
测试样本	R^2	0.790	0.959	0.985	0.986	0.987	0.976
	MSE	0.059	0.012	0.004	0.004	0.004	0.007
	准确率/%	94	100	100	100	100	100
灾害点	MSE	2.27×10^{-5}	0.001	0.002	0.008	0.008	0.014
	准确率/%	100	100	100	100	100	100
非灾害点	MSE	0.118	0.023	0.007	0.001	0.001	0.001
	准确率/%	88	100	100	100	100	100

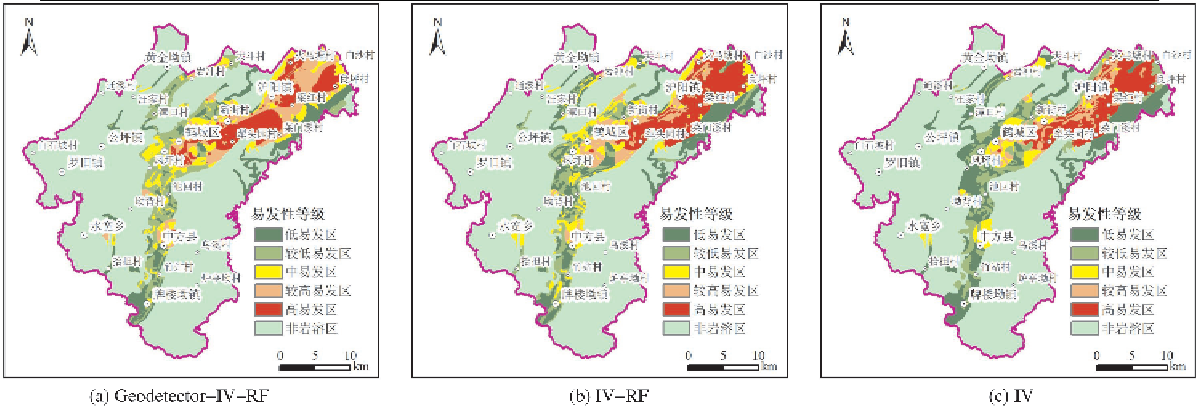


图 6 不同模型下岩溶塌陷易发性分级图

为了进一步分析 Geodetector-IV-RF 模型的评价效果, 将其与 IV-RF 模型和 IV 模型的评价结果进行对比。选择全部 8 个影响因素作为模型输入, 训练样本点和测试样本点均与上述 1:1 比例样本点一致, 对 IV-RF 模型进行训练和测试, 利用测试好的 IV-RF 模型对研究区进行岩溶塌陷易发性评价, 采用自然断点法对评价结果进行分区, 得到基于 IV-RF 模型的岩溶塌陷易发性分区图(图 6b), 基于 IV 模型的岩溶塌陷易发性分区图见图 6c。

4.3 易发性结果分析

Geodetector-IV-RF 模型、IV-RF 模型和 IV 模型预测的怀化地区岩溶塌陷易发性结果如表 3 所示。由表 3 可见随着岩溶塌陷易发性等级由低到高, 三个模型的频率比值也随着增大。其中 Geodetector-IV-RF 模型易发性预测结果从低到高的频率比值分别为 0.000、0.056、0.238、1.365 和 3.442; IV-RF 模型预测结果从低到高的频率比值分别为 0.000、0.000、0.513、1.418 和 3.069; IV 模型的易发性评价结果从低到高的频率比值分别

是 0.000、0.167、1.055、2.109 和 3.011。将高和较高易发性等级的频率比之和除以所有等级易发性频率比之和, 就可以得到模型对岩溶塌陷易发性预测结果的频率比精度。计算可得 Geodetector-IV-RF 模型、IV-RF 模型和 IV 模型的频率比精度分别为 0.942、0.897 和 0.807, Geodetector-IV-RF 模型的频率比精度高于 IV-RF 模型和 IV 模型, 能够很好的反映岩溶塌陷的分布规律。结合图 6 可知, 三种岩溶塌陷易发性预测模型对怀化地区岩溶塌陷预测的易发性图大体上是一致的, 高和较高易发区主要分布于中方镇、城东新区-梁红村一带、鹤城区、犁头园村-菜门溪村一带等地区, 高和较高岩溶塌陷易发区岩溶发育强, 多数岩溶水位在基岩面以下和附近波动, 岩溶塌陷发育较多。高和较高岩溶塌陷易发区包括的怀化主城区、城东新区和中方县城等地人口密集、交通发达、经济发展程度较高, 需要加强对岩溶塌陷的监测和预警, 在高和较高易发区施工要注意避免人为诱发岩溶塌陷。

表 3 岩溶塌陷易发性预测模型的频率比精度分析

模型	易发性等级	栅格数	栅格占比/%	灾害点数	灾害点占比/%	频率比
Geodetector-IV-RF 模型	高	117 556	19.48	57	67.06	3.442
	较高	119 575	19.82	23	27.06	1.365
	中	119 222	19.76	4	4.71	0.238
	较低	126 221	20.92	1	1.18	0.056
	低	120 849	20.03	0	0.00	0.000
IV-RF 模型	高	134 179	22.24	58	68.24	3.069
	较高	95 137	15.77	19	22.35	1.418
	中	110 704	18.35	8	9.41	0.513
	较低	107 339	17.79	0	0.00	0.000
	低	154 177	25.55	0	0.00	0.000
IV 模型	高	115 546	19.15	49	57.65	3.011
	较高	80 790	13.39	24	28.24	2.109
	中	60 564	10.04	9	10.59	1.055
	较低	127 563	21.14	3	3.53	0.167
	低	217 073	35.97	0	0.00	0.000

在地质灾害易发性评价结果检验中, 常采用 ROC 曲线的验证方法。采用 ROC 曲线来对 Geodetector-IV-RF 模型、IV-RF 模型和 IV 模型的评价结果的精度进行检验, 得到三种模型的 AUC 值分别为 0.948、0.873 和 0.851 (图 7), 三种岩溶塌陷易发性预测模型都具有较高的评价精度, 但 Geodetector-IV-RF 模型的 AUC 值最大, 具有更好的评价效果。Geodetector-IV-RF 模型、IV-RF 模型和 IV 模型的 ROC 曲线验证结果和频率比精度的分析结果是一致的。由模型预测结果得频率比精度分析和 ROC 曲线验证可知, Geodetector-IV-RF 模型预测效果最好。

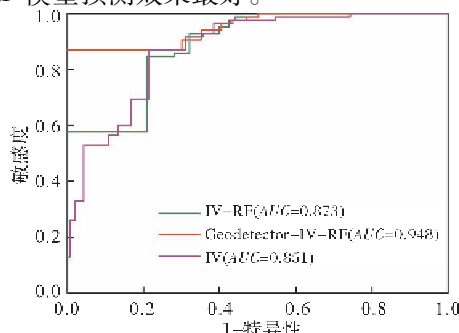


图 7 岩溶塌陷易发性预测模型 ROC 曲线

5 结论

(1) 采用地理探测器对怀化地区岩溶塌陷影响因素进行空间分异性探测可知, 岩溶发育程度、岩溶水位与基岩面关系和岩溶塌陷密度这三个影响因素对岩溶塌陷的影响程度较大, q 值分别为 0.352、0.126 和 0.854, 影响因素两两相互作用对岩溶塌陷的解释起增强作用。

(2) 对不同比例样本训练的模型对测试样本的预测效果对比可知, 在考虑地质灾害防御安全性上, 基于 1:1 比例样本训练的 RF 模型可以更合理对怀化地区岩溶塌陷易发性进行预测。

(3) 利用频率比精度和 ROC 曲线方法对 Geodetector-IV-RF 模型、IV-RF 模型和 IV 模型的评价效果进行对比分析, Geodetector-IV-RF 模型的频率比精度和 AUC 值均为最高, 该模型的预测精度较高, 能够很好的反映怀化地区岩溶塌陷易发程度。

(4) 研究区岩溶塌陷高和较高易发区面积占比分别为 19.48% 和 19.82%, 灾害点占比分别为 67.06% 和 27.06%。高和较高易发区主要分布于中方镇、城东新区-梁红村一带、鹤城区、犁头园村-菜门溪村一带等地区, 其中怀化主城区、城东新区和中方县城等地人口密集、交通发达、经

济发展程度较高, 需要加强对岩溶塌陷的监测和预警。

参考文献:

- [1] 雷明堂, 蒋小珍. 岩溶塌陷研究现状、发展趋势及其支撑技术方法[J]. 中国地质灾害与防治学报, 1998, 9(3): 1-6.
- [2] 韩庆定, 罗锡宜. 广东佛山市高明区李家村岩溶塌陷群成因机理分析[J]. 中国地质灾害与防治学报, 2021, 32(4): 56-64.
- [3] 蒙彦, 雷明堂. 岩溶塌陷研究现状及趋势分析[J]. 中国岩溶, 2019, 38(3): 411-417.
- [4] 涂婧, 魏瑞均, 杨戈欣, 等. 湖北武汉岩溶塌陷时空分布规律及其影响因素分析[J]. 中国地质灾害与防治学报, 2019, 30(6): 68-73, 93.
- [5] 毛世格, 管振德, 阎春恒. 基于小波包分形和神经网络的地震与岩溶塌陷识别[J]. 地震学报, 2018, 40(2): 195-204.
- [6] 李光辉, 潘懋, 杨志双, 等. 基于灰色聚类法的鞍山城区岩溶塌陷危险性区划[J]. 吉林大学学报(地球科学版), 2006, 36(S1): 75-79.
- [7] 武鑫, 黄敬军, 缪世贤. 基于层次分析-模糊综合评价法的徐州市岩溶塌陷易发性评价[J]. 中国岩溶, 2017, 36(6): 836-841.
- [8] 王恒恒, 张发旺, 郭纯青, 等. 基于层次分析法的城市岩溶塌陷危险性评价: 以武汉市南部为例[J]. 中国岩溶, 2016, 35(6): 667-673.
- [9] WANG J F, ZHANG T L, FU B J. A measure of spatial stratified heterogeneity[J]. Ecological Indicators, 2016, 67: 250-256.
- [10] 王桂林, 强壮, 曹聪, 等. 基于地理探测器与层次分析法的岩溶地面塌陷易发性评价: 以重庆中梁山地区为例[J]. 中国岩溶, 2022, 41(1): 79-87.
- [11] 支泽民, 陈琼, 张强, 等. 地理探测器在判别滑坡稳定性影响因素中的应用: 以西藏江达县为例[J]. 中国地质灾害与防治学报, 2021, 32(2): 19-26.
- [12] 刘婷, 谭建民, 郭飞, 等. 人工切坡下滑坡易发性评价中权重修正方法研究: 以赣州市沙地镇为例[J]. 自然灾害学报, 2021, 30(5): 217-225.
- [13] MERGHADI A, A YUNUS, J DOU, et al. Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance [J]. Earth - Science Reviews, 2020, 207: 103225.
- [14] 黄发明, 胡松雁, 闫学涯, 等. 基于机器学习的滑坡易发性预测建模及其主控因子识别[J]. 地质科技通报, 2022, 41(2): 79-90.
- [15] 孙德亮. 基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D]. 上海: 华东师范大学, 2019.
- [16] 邓念东, 石辉, 文强, 等. 信息量支持下的随机森林模型的崩塌易发性评价[J]. 科学技术与工程, 2021, 21(6): 2210-2217.
- [17] 雷明堂, 戴建玲, 吴远斌, 等. 湘西鄂东皖北地区岩溶塌陷 1:5 万环境地质调查成果报告[R]. 桂林: 中国地质科学院岩溶地质研究所, 2020.
- [18] WANG J F, LI X H, CHRISTAKOS G, et al. Geographical detectors - based health risk assessment and its application in the neural tube defects study of the Heshun region, China[J]. International Journal of Geographical Information Science, 2010, 24(1): 107-127.
- [19] 王劲峰, 徐成东. 地理探测器: 原理与展望[J]. 地理学报, 2017, 72(1): 116-134.
- [20] 王小川, 史峰, 郁磊. MATLAB 神经网络 43 个案例分析[M]. 北京: 北京航空航天大学出版社, 2013.
- [21] 吴远斌, 刘之葵, 殷仁朝, 等. 基于 AHP 和 GIS 技术的湖南怀化地区岩溶塌陷易发性评价[J]. 中国岩溶, 2022, 41(1): 21-33.

Evaluation of Karst Collapse Susceptibility Based on Geodetector and Random Forests

REN Tao^{1,2,3}, TIAN Guoliang², NING Zhijie⁴, ZHOU Aihong³, LI Kuan² and CHEN Shi²

(1. Institute of Environmental Geology, Hebei Coal Geology Bureau, Shijiazhuang 050085, China;

2. NO.6 Institute of Geology and Mineral Resources Exploration of Shandong Province, Weihai 264209, China;

3. School of Urban Geology and Engineering, Hebei GEO University, Shijiazhuang 050031, China;

4. Chinese Academy of Geological Sciences, China University of Geosciences (Beijing), Beijing 100083, China)

Abstract: Karst collapse is affected by many factors, its occurrence is sudden, and its distribution has spatial differentiation. In order to reasonably evaluate the susceptibility of karst collapse in Huaihua area, geographic

detectors and random forest model are used to evaluate the susceptibility of karst collapse in Huaihua area, and the main influencing factors and the interaction between the influencing factors are analyzed. The susceptibility evaluation results of Geodetector - IV - RF, IV - RF and IV models are compared and analyzed. The results show that: The degree of karst development, the relationship between the karst water level and the bedrock surface, and the density of karst collapse have a high degree of explanation for karst collapse; The interaction type between the two influencing factors of karst collapse is two factors Enhancement or nonlinear enhancement can enhance the occurrence of karst collapse; The evaluation results of the Geodetector - IV - RF model have the highest accuracy, and the partition results can provide auxiliary early warning for the prevention and control of karst collapse in the study area.

Keywords: karst collapse; geodetector; random forests; susceptibility evaluation

(上接第 226 页)

参考文献:

- [1] 王伟奇. 中国泥石流现状与浅析[J]. 科技信息, 2009(29): 597, 745.
- [2] 唐邦兴, 杜榕桓, 康志成, 等. 我国泥石流研究[J]. 地理学报, 1980, 35(3): 259 - 264.
- [3] 康志成, 韦方强, 丁明涛. 中国泥石流重大灾害与治理点分布[C]//第五届海峡两岸山地灾害与环境保育学术研讨会暨第六届全国泥石流学术会议论文集, 2006: 76 - 92.
- [4] 张书豪, 吴光. 随机森林与 GIS 的泥石流易发性及可靠性[J]. 地球科学, 2019, 44(9): 3115 - 3134.
- [5] 赵源, 刘希林. 人工神经网络在泥石流风险评价中的应用[J]. 地质灾害与环境保护, 2005, 16(2): 135 - 138.
- [6] 王常明, 田书文, 王虹虹, 等. 泥石流危险性评价: 模糊 c 均值聚类 - 支持向量机法[J]. 吉林大学学报(地球科学版), 2016, 46(4): 1168 - 1175.
- [7] 周仲礼, 张乾荣, 曹赛男. 基于决策树理论的泥石流危险性评价研究[J]. 重庆交通大学学报(自然科学版), 2017, 36(4): 77 - 81.
- [8] 刘永奎, 第宝锋, 詹宇, 等. 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[J]. 山地学报, 2018, 36(5): 765 - 773.
- [9] 李永威, 徐林荣, 谷丰宇, 等. 孕灾环境对泥石流危险性影响[J/OL]. 地球科学, 2023. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C45S0n9f12suRadTyEV12pW9UrhTDCdPD66r-NJ9scI9qlomumccvBDJQjchacqK4JZYBEfUrNGs34wWggGyQk1&uniplatform=NZKPT>.
- [10] 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.
- [11] 刘林通, 孟兴民, 郭鹏, 等. 基于流域单元和信息量法的白龙江流域泥石流危险性评价[J]. 兰州大学学报(自然科学版), 2017, 53(3): 292 - 298, 308.
- [12] 王晓朋, 潘懋, 徐岳仁. 基于流域单元的泥石流区域危险性评价[J]. 山地学报, 2006, 24(2): 177 - 180.
- [13] 牛岑岑. 泥石流危险度评价指标的提取与等级划分[D]. 长春: 吉林大学, 2013.
- [14] 姚振国, 刘建周, 牛贝贝, 等. 流域面积对沟道泥石流发育的影响分析[J]. 资源环境与工程, 2019, 33(2): 217 - 219.
- [15] MELTON M A. The geomorphic and paleoclimatic significance of alluvial deposits in southern Arizona: a reply[J]. The Journal of Geology, 1966, 74(1): 102 - 106.
- [16] 向灵芝, 崔鹏, 张建强, 等. 汶川县地震诱发崩塌灾害影响因素的敏感性分析[J]. 四川大学学报(工程科学版), 2010, 42(5): 105 - 112.
- [17] 毛佳睿, 李铁锋, 田运涛, 等. 基于物源特征的白龙江流域泥石流易发性评价[J]. 科学技术与工程, 2020, 20(28): 11479 - 11490.
- [18] 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.
- [19] 吕红燕, 冯倩. 随机森林算法研究综述[J]. 河北省科学院学报, 2019, 36(3): 37 - 41.
- [20] 李坤, 赵俊三, 林伊琳, 等. 基于 RF 和 SVM 模型的东川泥石流易发性评价研究[J]. 云南大学学报(自然科学版), 2022, 44(1): 107 - 115.
- [21] PANDEY V K, SHARMA K K, POURCHASEMI H R, et al. Sedimentological characteristics and application of machine learning techniques for landslide susceptibility modelling along the highway corridor Nahan to Rajgarh (Himachal Pradesh), India[J]. CATENA, 2019, 182: 104150.
- [22] RAHMATI O, TAHMASEBPOUR N, HACHIZADEH A, et al. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion[J]. Geomorphology, 2017, 298: 118 - 137.
- [23] 扈秀宇, 秦胜伍, 窦强, 等. 基于 GIS 和随机森林模型的泥石流敏感性分析: 以吉林省洮南市北部山区为例[J]. 水土保持通报, 2019, 39(5): 204 - 210, 217, 2.
- [24] 张书豪, 吴光, 张乔, 等. 基于子流域特征的泥石流易发性评价[J]. 水文地质工程地质, 2018, 45(2): 142 - 149.

Assessment of Debris Flow Vulnerability in Xiangcheng County, Ganzi Tibetan Autonomous Prefecture Based on RF model

WANG Xiangqian, DING Mingtao, HE Yufeng

(School of Earth Science and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: In this paper, the debris flow disaster in Xiangcheng County, Ganzi Tibetan autonomous Prefecture is selected as the research object, and the watershed unit is selected as the evaluation unit based on the machine learning algorithm of machine forest (RF), combining 9 index factors (watershed area, average elevation of the watershed, average plane curvature of the watershed, average profile curvature of the watershed, average slope of the watershed, elevation difference of the watershed, melton ratio, main rock formations of the watershed, and gully density of the watershed) to build the debris flow vulnerability evaluation model, determine the weight of indicator factors, select the optimal combination of factors, and use the ROC, AUC and ACC to compare the accuracy and stability of the model. The results show that the ACC value and AUC value of the prediction model are 0.8596 and 0.9342 respectively, and the basin unit as the evaluation unit can achieve good prediction accuracy.

Keywords: debris flow; susceptibility; random forest; Xiangcheng county of Sichuan