

基于Logistic回归和随机森林的清江流域长阳库岸段堆积层滑坡易发性评价

曾 斌, 吕权儒, 寇 磊, 艾 东, 许汇源, 袁晶晶

Susceptibility assessment of colluvium landslides along the Changyang section of Qingjiang River using Logistic regression and random forest methods

ZENG Bin, LYU Quanru, KOU Lei, AI Dong, XU Huiyuan, and YUAN Jingjing

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.202205044>

您可能感兴趣的其他文章

Articles you may be interested in

基于RBF神经网络信息量耦合模型的滑坡易发性评价

Landslide susceptibility assessment by the coupling method of RBF neural network and information value: A case study in Min Xian, Gansu Province

黄立鑫, 郝君明, 李旺平, 周兆叶, 贾佩钱 中国地质灾害与防治学报. 2021, 32(6): 116–126

北京山区突发性地质灾害易发性评价

Assessment on the susceptibility of sudden geological hazards in mountainous areas of Beijing

罗守敬, 王珊珊, 付德荃 中国地质灾害与防治学报. 2021, 32(4): 126–133

基于不同因子分级法的滑坡易发性评价——以湖北远安县为例

§{suggestArticle.titleEn}

闫举生, 谭建民 中国地质灾害与防治学报. 2019, 30(1): 52–60

基于遥感影像多尺度分割与地质因子评价的滑坡易发性区划

Landslide susceptibility assessment based on multi-scale segmentation of remote sensing and geological factor evaluation

李文娟, 邵海 中国地质灾害与防治学报. 2021, 32(2): 94–99

基于快速聚类-信息量模型的汶川及周边两县滑坡易发性评价

Landslide susceptibility assessment based on K-means cluster information model in Wenchuan and two neighboring counties, China

周天伦, 曾超, 范晨, 毕鸿基, 龚恩慧, 刘晓 中国地质灾害与防治学报. 2021, 32(5): 137–150

考虑滑带强度参数分区取值的堆积层滑坡稳定性分析方法

§{suggestArticle.titleEn}

闫玉平, 肖世国 中国地质灾害与防治学报. 2020, 31(2): 44–49



关注微信公众号, 获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.202205044

曾斌, 吕权儒, 寇磊, 等. 基于 Logistic 回归和随机森林的清江流域长阳库岸段堆积层滑坡易发性评价 [J]. 中国地质灾害与防治学报, 2023, 34(4): 105-113.

ZENG Bin, LYU Quanru, KOU Lei, et al. Susceptibility assessment of colluvium landslides along the Changyang section of Qingjiang River using Logistic regression and random forest methods[J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(4): 105-113.

基于 Logistic 回归和随机森林的清江流域长阳库岸段堆积层滑坡易发性评价

曾斌¹, 吕权儒², 寇磊³, 艾东³, 许汇源³, 袁晶晶³

(1. 中国地质大学(武汉)环境学院, 湖北 武汉 430000; 2. 湖北省电力勘测设计院有限公司, 湖北 武汉 430040; 3. 湖北省地质局第七地质大队, 湖北 宜昌 443000)

摘要: 清江流域地质环境背景条件复杂, 特别是沿清江库岸地质灾害频发; 而过往的地质灾害易发性评价多针对行政区域范围, 鲜有针对库岸带的专门评价, 并且所用评价指标体系的针对性以及评价方法的可靠性仍有进一步提升的空间。为构建一套更加符合清江流域库岸带地质灾害发育特征的易发性评价指标体系, 获得更加准确、适用性强的易发性区划成果, 以清江流域渔峡口—资丘段为研究区, 以两岸涉水斜坡体为研究对象, 以斜坡单元为评价单元, 构建坡度、坡向、高差、坡型、归一化植被指数、地形湿度指数、斜坡结构类型、工程地质岩组、堆积层厚度、河谷演化类型等 10 个指标组成的易发性评价体系; 采用基于归一化确定性系数的 Logistic 回归和随机森林方法构建评价模型并得到不同易发性区划成果。评价结果显示: 高易发区主要分布于清江干流渔峡口东—资丘东段左岸顺向斜坡体的中—下部涉水区域, 且 Logistic 回归模型在地形地貌复杂的库岸段的适用性要优于随机森林模型。研究表明: 所选堆积层厚度及河谷演化类型指标很好地代表了清江库岸段的独特地质背景条件特点; 在非行政区划范围的特定研究区且当历史滑坡样本数量有限的情况下, Logistic 回归模型能够较好地学习灾害发育规律并具备可靠的易发性预测能力。

关键词: 堆积层滑坡; Logistic 回归; 随机森林; 易发性评价

中图分类号: P642.22

文献标志码: A

文章编号: 1003-8035(2023)04-0105-09

Susceptibility assessment of colluvium landslides along the Changyang section of Qingjiang River using Logistic regression and random forest methods

ZENG Bin¹, LYU Quanru², KOU Lei³, AI Dong³, XU Huiyuan³, YUAN Jingjing³

(1. School of Environmental Studies, China University of Geosciences, Wuhan, Hubei 430000, China; 2. Power China Hubei Electric Engineering CO. Ltd., Wuhan, Hubei 430040, China; 3. The Seventh Geological Brigade of Hubei Geological Bureau, Yichang, Hubei 443000, China)

Abstract: The geological and environmental background conditions of the Qingjiang River Basin are highly complex, particularly with frequent geological disasters along the Qingjiang reservoir bank. Previous susceptibility assessment for

收稿日期: 2022-05-27; 修订日期: 2022-09-14

投稿网址: <https://www.zgdzzhyfzxb.com/>

第一作者: 曾斌(1980-), 男, 湖北宜昌人, 博士, 副教授, 主要从事地质灾害调查、评价、监测预警、防控治理等方面的研究工作。

E-mail: zengbin@cug.edu.cn

通讯作者: 袁晶晶(1989-), 男, 湖北宜昌人, 水工环工程师, 主要从事地质灾害防治、监测预警方面的工作。E-mail: 937295811@qq.com

geological disasters was mostly focused on administrative areas and seldom specialized evaluations for the reservoir bank zone. Furthermore, there is still room for improvement in the evaluation index system, as well as in the pertinence and reliability of the evaluation method. To address these shortcomings, a more suitable susceptibility evaluation index system was constructed to obtain accurate and applicable susceptibility zoning results. The Yuxiakou to Ziqiu section of the Qingjiang River Basin was chosen as the research area, with the wading slope body on both sides of the river selected as the research object and the slope unit chosen as the evaluation unit. A susceptibility evaluation system composed of ten indicators, including slope, aspect, elevation range, slope type, *NDVI*, *TWI*, slope structure type, engineering geological rock formation, accumulation thickness, and valley evolution, was constructed. The logistic regression and random forest methods were used to construct the evaluation model based on the normalized certain factors, and different susceptibility zoning results were obtained. According to the evaluation results, the high-prone areas were mainly distributed in the middle to lower water wading areas of the left bank from the east of Yuxiakou to the east of Ziqiu, along the main stream of the Qingjiang River. The logistic regression model showed better applicability in the reservoir-bank section with complex topography and landforms. The research revealed that the accumulation thickness and valley evolution indicators were effective in representing the unique geological background conditions of the Qingjiang reservoir bank. The logistic regression model was able to learn the developmental law of disasters and has a reliable susceptibility prediction ability.

Keywords: accumulation thickness; Logistic regression; random forest; susceptibility assessment

0 引言

清江流域长阳库岸段地质背景条件复杂,堆积层滑坡是区内主要滑坡类型,近年来区内老滑坡复活和新滑坡发生的频率不断上升,对河谷两岸斜坡体上居民的人身财产安全产生了严重的威胁,因此开展以斜坡体为基本单元的滑坡易发性评价工作、提供更加准确的区划成果是一件事关民生的重要工作,具有较高的科研和社会价值。前人在清江流域区域滑坡易发性评价方面已经开展了部分研究工作:如雷深涵^[1]总结了长阳县地质灾害发育特征和规律,肖尚德^[2]开展了长阳境内清江流域滑坡分布规律和防灾对策研究工作。

易发性评价中的三个关键步骤分别是:划分具有适用性的评价单元、构建具有针对性的评价指标体系、采用结果可靠的评价方法。从评价单元划分上看,斜坡单元和栅格单元是常见的两种评价单元,目前的划分方法有水文分析法、曲率分水岭法、*r.slopeunits*法、*MIA-HSU*法^[3-5]。常用的评价方法则有数理统计法和机器学习法:数理统计法有信息量模型、频率比模型、证据权模型和熵指数模型等^[6-9];机器学习模型有支持向量机(SVM)、人工神经网络(ANN)和随机森林(RF)等^[10-12];这些方法的基础数学原理和模型构建方法不同,根据研究区的地质背景条件不同,方法的适用性也有很大差异。

前人在清江流域长阳库岸段易发性评价中侧重于区域研究,评价结果具体运用到单个斜坡体上的精度还

有待进一步提高,且较少考虑干流河谷区这一独特的地质背景条件。针对这些不足,本文以清江流域渔峡口—资丘段为研究区,以斜坡单元作为评价单元,构建基于归一化确定性系数的 Logistic 回归和随机森林易发性评价模型并完成不同等级易发性分区预测,经 ROC 曲线验证得到更优的区划成果。研究成果不仅能够为清江库岸段堆积层滑坡易发性评价及政府部门的防灾减灾工作提供依据,同时也能够为类似山区库岸带地质灾害易发性评价研究提供有益参考。

1 研究区概况

1.1 研究区位置

研究区位于湖北省宜昌市长阳土家族自治县西段清江干流流域,地处鄂西南清江中下游、隔河岩水库上游 10 km 位置,沿清江干流呈条带状分布,东西长 42 km,南北宽 12 km,总面积 504 km²,途径渔峡口镇和资丘镇。研究区地理位置如图 1 所示。

1.2 地质背景条件

研究区地形南北高、东西低,清江自西向东从研究区中部穿过汇入长江,最高海拔 1 824 m,最低为清江水位 180 m。断裂带主要位于清江干流左岸,褶皱轴线总体呈 NWW—NEE 分布,导致干流两岸基岩产状有显著差异。地层跨越寒武系—第四系,岩性以滨海—浅海相的碳酸盐岩为主,80% 以上的斜坡体出露基岩为三叠系大冶组(*T₁d*)灰岩,二叠系大隆组(*P₃d*)硅质岩及炭质泥岩、吴家坪组(*P₃w*)灰岩、孤峰组硅质岩夹泥岩(*P₂g*)、



图 1 研究区地理位置图

Fig. 1 Geographical location map of the study area

茅口组(P_2m)灰岩、栖霞组(P_2q)灰岩,少量出露石炭系(C_{1-2})和泥盆系(D)灰岩、砂岩、粉砂岩。第四系主要为崩坡积物(Qh^{dl+col}),由砾石、砂土和黏性土夹杂碎石组成,坡体下部厚、中上部薄,结构松散、富水性和透水性好。

1.3 库岸带堆积层滑坡发育特征

研究区堆积层滑坡物质来源包含两个方面:一是第四系崩坡积物;二是表层基岩由于风化、构造破碎作用与历史滑坡崩塌体堆积、人工斜坡体开挖等作用共同形成的松散堆积物。滑坡滑移面或滑带则多沿堆积层与中强风化基岩形成,以直线或折线形态为主,倾向清江或临近沟谷。研究区有完整编录资料的堆积层滑坡 41 处,平均滑体厚度 11.44 m,滑坡规模小-中型为主。历史滑坡分布如图 2 所示。

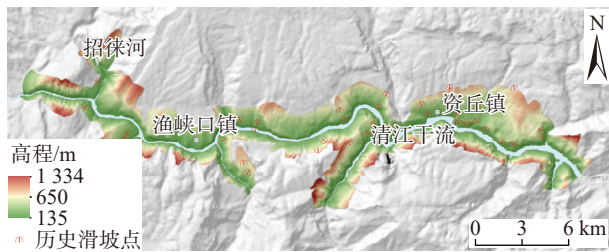


图 2 研究区历史滑坡分布图

Fig. 2 Distribution map of historical landslides in the study area

1.4 库岸带堆积层滑坡影响因素

影响库岸带堆积层滑坡发生的各类因素中:斜坡体的岩性条件、构造条件和临空条件是较为重要的影响因素。从岩性条件上看,含有软弱夹层的“二元结构”是堆积层滑坡形成的重要岩性组合基础,当斜坡体上部具有软弱岩夹层(如 P_3d 的薄层炭质泥页岩)时,该类岩层更容易转变为潜在滑面。从构造条件上看,斜坡体结

构为顺向坡,当坡向在 $ESS \sim WSS$ 之间时,软弱岩层能够成为潜在滑移面。从临空条件上看,清江干流的下切作用使左岸的斜坡体顺层潜入清江,坡脚处于水位以下且前缘临空,随着江水对坡脚的不断冲蚀掏空,前缘岩土体不断滑入江中,为后缘滑动提供临空条件。

2 研究方法

本文研究技术路线如图 3 所示。

2.1 评价指标提取方法

从影响堆积层滑坡发生的各类因素来看,斜坡结构和岩性组合条件是影响滑坡发生的主要本体因素,地形地貌、水文、地质构造为次要本体因素,河谷演化类型是重要区域历史演化背景影响因素。本文初选评价指标体系共包含 5 个类别、共计 11 个评价指标(图 3)。

其中:坡度、坡向、高差、坡型可由 DEM 直接计算得到;工程地质岩组和距断层距离由野外实测与项目资料得到;归一化植被指数由 Landsat8 数据在 ENVI(美国 Exelis Visual Information Solutions 公司研发遥感图像处理平台)中计算得到;地形湿度指数由 DEM 根据水文学公式计算得到;河谷演化类型根据常宏等^[13-14]对清江中下游河道变迁历史和河流演化规律总结得来;斜坡结构类型根据野外实测的产状点利用经验贝叶斯克里金插值法(EBK)拟合得到研究区基岩产状分布情况,然后经过与坡向的差值,从而将斜坡结构类型分为顺向坡、斜向坡、横向坡和逆向坡;堆积层厚度不能采用简单的类比或者转化为某单一地形指标的函数,论文利用已有滑坡点堆积层厚度数据,采用分区拟合方法计算全区堆积层厚度。

2.2 水文分析法

水文分析法是常用的斜坡单元划分方法。它是基于 ArcGIS 中的“Hydrology”模块,以 DEM 为原始数据,通过填挖、提取流向和流量、生成河网栅格、提取集水区等步骤,提取出山脊线和山谷线,并以此为边界划出斜坡单元。斜坡单元划分合理程度用形状指数(morphology index, MI)来评判^[15]。

2.3 指标体系构建及易发性评价方法

2.3.1 K-means 聚类法

基于多元统计分析学的 K-means 聚类法可以用于连续型评价指标子类划分中,它可以将样本数据按照统计数据特性的相似程度进行划分。该算法通过多次迭代来寻找最优聚类中心,使每个样本的特征值点与其所属类均值的距离和最小^[16]。该方法最大优势是消除了分类者主观的影响,可以根据相似性判断准则对所参与

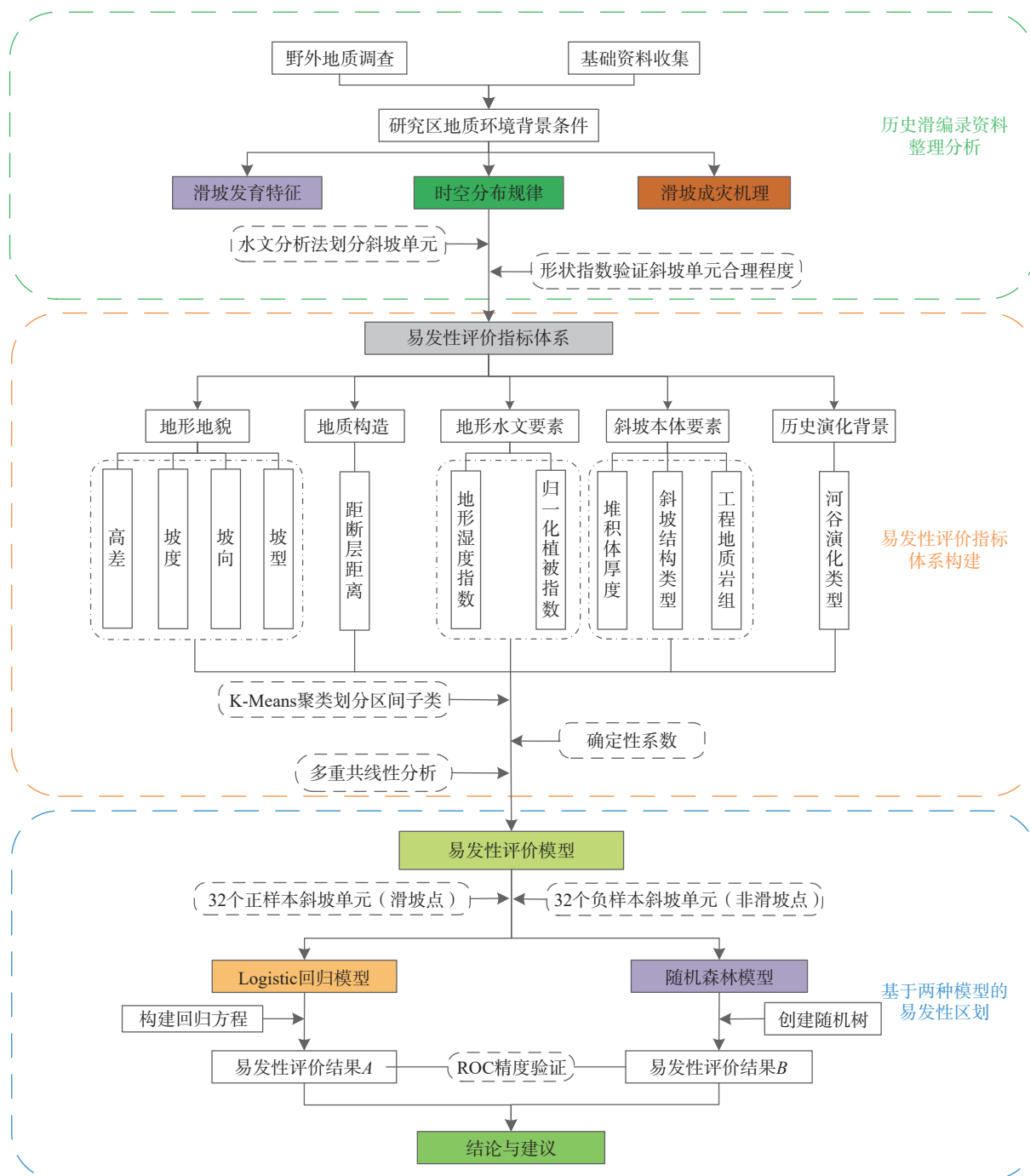


图3 技术路线图

Fig. 3 Technology roadmap

子类划分的全部数据进行归类。

2.3.2 确定性系数法

确定性系数(certain factors, CF)是二元统计学中的一种概率函数,可以较为精确地分析某种因素对事件发生的贡献程度。CF值为 $[-1, 1]$ 之间的数值,CF值越大,说明该子类对滑坡发生的贡献程度越大;CF值越小,说明该子类对滑坡发生的贡献程度越小^[17]。该方法

结果清晰,对各子类与滑坡发生的关联性定量化程度较高,且可以用于后续评价指标赋值。

2.3.3 Logistic 回归原理

Logistic 回归模型的基础理论是多元统计学,是一种非线性评价模型,该模型对数据类型和数据总体分布情况的要求较低,既可以选用连续型变量也可以选择分类型变量。很好地解决了滑坡易发性评价模型建立中

变量类型不统一的问题^[18-19]。

2.3.4 随机森林原理

随机森林采用 Bootstrap 重抽样方法从样本中抽取多个样本, 并对每个样本构建决策树(random forest trees), 然后将这些决策树组合在一起, 通过投票得出最终分类结果^[20]。在构建决策树时, 每一棵决策树都有一个训练集, 要构建 K 棵决策树, 需要产生 K 个训练集。在训练集形成的过程中, 采用随机且放回抽样方法从所

有样本中随机抽取一定量样本, 共产生 S 个训练子集。对最终的模型输出结果进行投票, 所得票数最多的分类结果作为算法最终的输出结果^[21]。

3 易发性评价结果

3.1 基础数据源

初选评价指标的来源、数据类型及子类划分方法如表 1 所示。

表 1 评价指标来源及处理方式
Table 1 Sources and processing methods of the evaluation indicators

评价指标	来源	数据类型	子类划分方式说明
坡度	5mDEM(栅格)	连续型	(1)归一化植被指数、地形湿度指数: 用K-means聚类法将指标分为5个子类区间 (2)堆积层厚度: 由野外实测堆积层点数据通过经验贝叶斯克里金法插值为面数据, 用K-means聚类法分为5个子类区间 (3)工程地质岩组分为6个子类区间: II-1(坚硬-较坚硬碎屑岩)、II-4(较坚硬碎屑岩)、III-1(坚硬碳酸盐岩)、III-2(坚硬碳酸盐岩夹碎屑岩)、III-3(较坚硬碳酸盐岩)、III-4(较坚硬碳酸盐岩夹碎屑岩) (4)斜坡结构类型分为4个: 顺向坡、斜向坡、横向坡、逆向坡 (5)河谷演化类型分为2个: 迁移性河谷、持续性河谷 (6)坡向分为8个: 北-东北-东-东南-南-西南-西-西北 (7)坡型分为3个: 凸形坡、平直坡、凹形坡
高差	5mDEM(栅格)	连续型	
归一化植被指数	Landsat8-OLT(栅格)	连续型	
地形湿度指数	5mDEM(栅格)	连续型	
距断层距离	线数据(矢量)	连续型	
堆积层厚度	点数据(矢量)	连续型	
工程地质岩组	面数据(矢量)	分类型	
斜坡结构类型	面数据(矢量)	分类型	
河谷演化类型	面数据(矢量)	分类型	
坡向	5mDEM(栅格)	分类型	
坡型	5mDEM(栅格)	分类型	

3.2 斜坡单元划分及评价指标处理结果

本文利用水文分析法, 在 ArcGIS 中将研究区分为 200 个斜坡单元, 平均 $MI=1.95$, 正态偏差为 $-0.1 \sim 0.15$, 符合正态分布规律, 整体形状基本规则。在此基础上, 将 6 个连续型指标利用 K-means 聚类完成子类区间划分, 其余 5 个分类型指标根据规范推荐划分子类区间。子类区间划分结果及在斜坡单元内的对应分类结果如图 4 所示。

在将全部评价指标划分子类区间后, 即可计算得到每个子类区间的确定性系数。根据计算结果: “距断层距离”指标的确定性系数与该指标与滑坡发生关联性的既有规律偏差较大, 说明研究区滑坡发生受该指标的影响的程度很小, 因此从后续评价指标体系中去除。将剩余的 10 个评价指标构建多重共线性诊断模型, 模型 Durbin-Watson 检验值为 $1.353 < 2$, 显著性值 $Sig=0.01$, 说明多重共线性诊断模型通过一阶自相关检验且具有显著性。

3.3 基于归一化确定性系数值的 Logistic 回归滑坡易发性评价结果

Logistic 回归模型采用 SPSS 软件构建。选取 32 处滑坡点斜坡单元作为 “ $Y=1$ ” 阳性样本、随机选取 32 处

经野外查证未发生滑坡的斜坡单元作为 “ $Y=0$ ” 阴性样本共同组成模型的全体样本, 抽取其中 70% 用于模型构建, 剩余 30% 用于模型质量检验。对于坡向、坡型、工程地质岩组、斜坡结构类型、河谷演化类型这 5 个分类型变量, 将它们的子类设置为虚拟变量并赋予其对应的编码纳入模型; 对于坡度、高差、 $NDVI$ 、 TWI 、堆积层厚度这 5 个连续型变量, 直接采用原始数值纳入模型。

构建 Logistic 回归模型后, 需对其模型的拟合优度和预测质量进行验证。本文所构建的模型 Hosmer-Lemeshow 检验的 $Sig=0.85 > 0.05$, 表明模型拟合较优, 且能充分利用已有样本信息最大化拟合模型。模型总体预测正确精度达 84.4%, 在滑坡点上的预测精度达 87.5%, 模型总体预测正确率较高, 模型可以用于后续回归方程的构建。

接下来用模型来对样本进行回归分析, 并计算各个指标的回归结果。单纯从全部指标的显著性水平来看, 并不是每个评价指标都满足 $Sig < 0.05$ 的判别标准, 个别连续型指标和个别分类型变量的哑变量出现了 $Sig > 0.05$ 的情况。从严格数理统计意义上讲, 这些变量与滑坡发生并不具备统计学意义, 但是由于评价指标的选

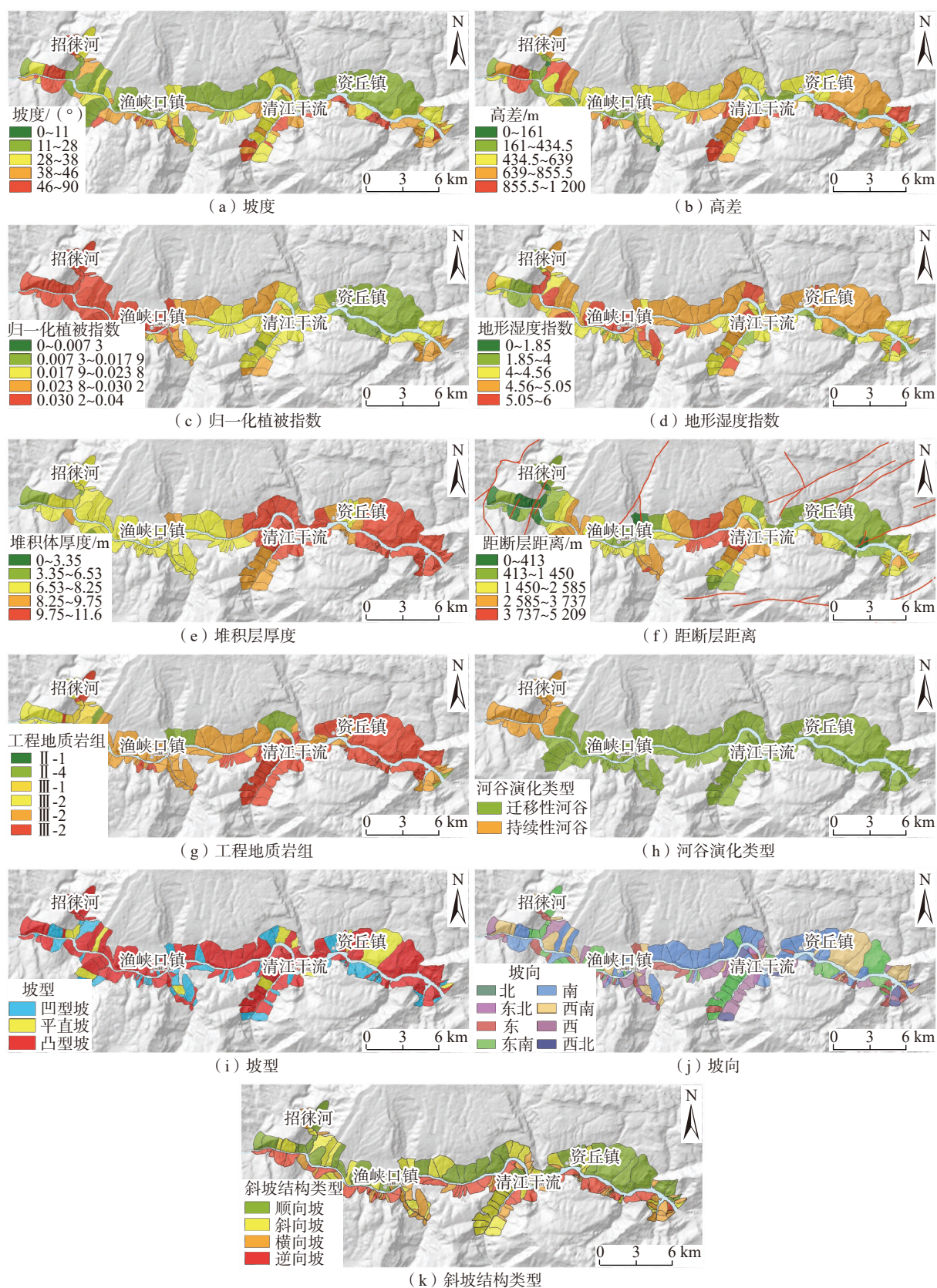


图 4 评价指标分类结果图

Fig. 4 Classification results of all evaluation index

取、分类标准的差异和滑坡发生的特殊性与复杂性,多数因素是作为间接影响因素参与到滑坡发展、变形过程中,并不是直接的诱发因素,因此使每个指标都与滑坡发生具备直接的回归关系是比较困难的,但其代表的地质意义不能被忽视,因此在显著性水平差异与标准值相差不是很大的情况下,仍需要将其纳入到回归方程中。最终的回归方程如式所示。

$$\begin{aligned} \text{Logit}P = & 0.366x_1 + 0.001x_2 + 311.267x_3 + 0.905x_4 \\ & + 0.189x_5 - 1.206x_{61} + 0.548x_{62} + 21.924x_{71} \\ & + 19.925x_{72} + 25.595x_{73} + 23.090x_{74} + 23.585x_{75} \\ & + 27.093x_{81} + 28.184x_{82} + 4.972x_{91} \\ & + 7.087x_{92} + 23.610x_{101} \end{aligned} \quad (1)$$

将根据式(1)计算所得的易发性指数在 ArcGIS 中利用自然断点法分为四个等级:不易发区 [51, 62)、低易发区 [62, 143)、中易发区 [143, 230)、高易发区 [230, 263]。研究区基于归一化确定性系数的 Logistic 回归易发性评价结果如图 5 所示。

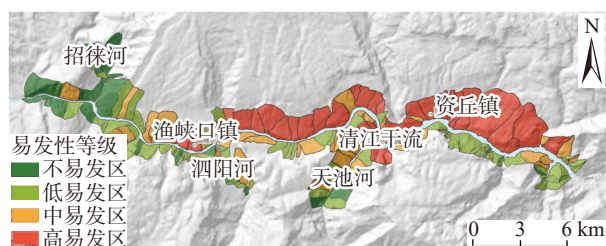


图 5 基于归一化 CF 值 Logistic 回归滑坡易发性分区图

Fig. 5 Landslide susceptibility zoning map based on normalized CF-valued logistic regression

3.4 基于归一化确定性系数值的随机森林滑坡易发性评价结果

本文随机森林模型构建采用 SPSS Modeler 中的 Random Forest Trees 函数来创建数据流文件。为了保持前后样本一致性,将 Logistic 回归中的训练样本集也作为随机森林的训练样本,从正负样本中随机选取 70% 的斜坡单元作为训练样本,30% 的斜坡单元作为测试样本。

在模型参数中,预测评价指标个数(m_{try})和随机树的个数(n_{tree})是两个重要的参数。其中 m_{try} 为构建随机树时从随机抽取的评价指标个数,一般为样本全部评价指标个数的根号值; n_{tree} 是模型随机树的构建数量。本次选取 $m_{\text{try}}=3$, $n_{\text{tree}}=2\ 000$, 选取“自动处理不平衡数据”的方法来修正不平衡数据。

运行数据流文件后,可得随机森林模型。根据模型置信度报告,其预测可信度概率处于 0.556~1 的平均

正确性达 88.2%, 2 折交叉验证的折叠正确性为 0.571, 基尼系数 $gini=0.281$, 说明不论是随机树分支还是总体模型的准确性均较好。从模型纳入的评价指标来看,原始训练样本评价指标中的“河谷演化类型”指标对应的频率权重在四舍五入后小于 1, 不具备纳入最终模型的要求,因此该指标从最终随机森林模型的评价指标中去除。

用构建的随机森林模型进行预测,再根据预测的二分类变量结果值和对应的可信度概率换算为易发性指数,将易发性指数在 ArcGIS 中利用自然断点法分为四个等级:不易发区 ($Y=0, P=1$)、低易发区 ($Y=0, P<1$)、中易发区 ($Y=1, P<0.889$)、高易发区 ($Y=1, P>0.889$)。研究区基于归一化确定性系数的随机森林易发性评价结果如图 6 所示。

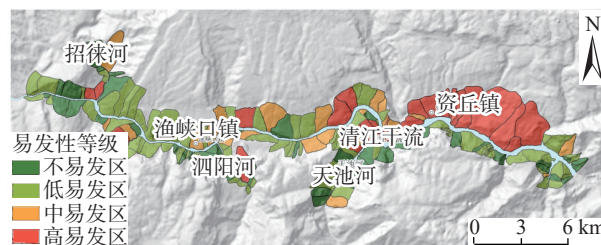


图 6 基于归一化 CF 值的随机森林滑坡易发性分区图

Fig. 6 Landslide susceptibility zoning map based on normalized CF-valued random forest results

4 讨论

4.1 基于不同模型的易发性评价结果对比

将历史滑坡点分别与两种易发性结果(图 5、图 6)进行对应可以发现:随着易发性等级的提高,两种模型对应区域内历史滑坡点的分布比例和面积也提高。中易发区和高易发区包含了 90% 以上的历史滑坡点,也说明两种模型评价结果与历史滑坡的实际情况较为吻合,符合研究区滑坡分布规律。

两种评价模型的 ROC(receiver operating characteristic curve)曲线如图 7 所示。其中 Logistic 回归模型 AUC (曲线下面积)=0.821, 随机森林模型 $AUC=0.769$, 两种模型对于高易发区均有较好的预测,基本符合历史滑坡点的分布规律,但两种模型在各个易发性等级上的预测结果和准确性有一定差异。

Logistic 回归模型预测得到的中高易发区要多于随机森林。由于 Logistic 回归模型是以历史滑坡数据为参考采用数理统计方法完成预测,具有一定的统计学规律,因此在将历史滑坡点所在斜坡单元预测为高易发的基础上,将具有相似条件的斜坡单元均预测为了中、高

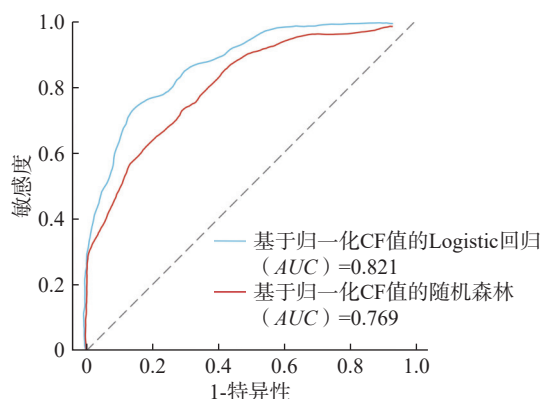


图 7 两种模型易发性结果 ROC 曲线

Fig. 7 ROC curves of susceptibility results for two models

易发区;但在右岸的少量逆向、横向坡的地段预测为了中易发性,右岸预测结果准确性低于随机森林模型。

从分布情况来看,随机森林基本将历史滑坡点所在的斜坡单元预测为了高易发区,其余多数被预测为了低易发区和不易发区,整体预测结果偏向于保守。从工程地质类比的角度来看,具有相似地质环境背景条件的易发性应具有一定相似性,而在左岸高易发区附近的顺向坡预测为了中、低易发区;而在右岸的逆向、横向坡地段预测结果较好。

4.2 不同易发性评价模型方法的适用性

从方法上看,Logistic 回归属于数理统计方法,它将全部样本一次性纳入到回归方程构建中,是一种对样本数据的拟合过程。随机森林是通过将样本进行大量随机抽样来构建学习规则的一种方法,是一种对数据内部之间关联性和影响作用的一种学习过程。

滑坡是一个复杂的地质体演化过程及结果,由于地质体强烈非均质性及各向异性,滑坡易发性评价实际上是一个受到众多因素影响的复杂非线性过程,任何一种模型只能是从某个方面得到更优预测结果,而无法涵盖全部影响要素;并且在要素的提取学习过程中,也会存在部分滑坡信息识别的精度问题。目前的众多研究中关于数理统计模型和机器学习模型易发性预测性能对比结果中,绝大多数都认为机器学习模型无论是从泛化适用能力还是预测结果准确性方面都优于数理统计模型。而在本文研究区中,预测性能更好的模型是数理统计模型,说明模型适用性在不同研究区具有很大的差异性,并不能一概而论说哪种模型具有普适性,而要根据研究区实际条件来确定。针对类似本文这样,研究区面积小、地质背景条件较为复杂,且掌握滑坡样本数目不是很多的情况,Logistic 回归方法的适用性要更强。

5 结论

(1)K-means 聚类法可以更加准确地划分指标子类区间,克服了传统划分方法中主观因素强的缺陷。确定性系数法能够定量化实现指标子类与滑坡的关联性强弱并给出关联系数,具有较强的客观性,也能为后续易发性模型构建提供更加准确的参考依据。

(2)基于研究区地质环境条件特色,选择“河谷演化类型”和“堆积层厚度”两个特色指标,有针对性的构建了易发性评价指标体系。根据库岸段堆积层滑坡成灾机理,评价指标中的“斜坡结构类型”和“工程地质岩组”是两个较为关键的指标,“距断层距离”在研究区与滑坡发生关联性不大。

(3)构建了基于归一化确定性系数的 Logistic 回归和随机森林两种易发性评价模型,经 ROC 曲线验证认为 Logistic 回归模型的预测结果准确性高于随机森林,成果适用性更好;而随机森林预测结果偏保守,适用性相对较差。因此建议在面积较小、灾点资料有限的研究区优先考虑如 Logistic 回归等数理统计方法;但若评价面积较大、评价单元选择栅格时,无论是从样本个数还是训练集个数都会呈几何倍增加,使用随机森林模型将会得到更好的结果。

参考文献(References):

- [1] 雷深涵,华骐,郭峰,等.长阳县地质灾害发育特征[J].地质灾害与环境保护,2013,24(1):30-35. [LEI Shenhan, HUA Qi, GUO Feng, et al. Characteristics of the geological hazards in Changyang [J]. Journal of Geological Hazards and Environment Preservation, 2013, 24(1): 30-35. (in Chinese with English abstract)]
- [2] 肖尚德,李智民,刘云彪,等.湖北省清江流域滑坡分布规律与减灾对策研究[J].工程地质学报,2012,20(4):514-521. [XIAO Shangde, LI Zhimin, LIU Yunbiao, et al. Distribution law and hazard mitigation strategies of landslides in Qingjing River valley, Hubei [J]. Journal of Engineering Geology, 2012, 20(4): 514-521. (in Chinese with English abstract)]
- [3] 王凯,张少杰,韦方强.斜坡单元提取方法研究进展和展望[J].长江科学院院报,2020,37(6):85-93. [WANG Kai, ZHANG Shaojie, WEI Fangqiang. Slope unit extraction methods: advances and prospects [J]. Journal of Yangtze River Scientific Research Institute, 2020, 37(6): 85-93. (in Chinese with English abstract)]
- [4] ALVIOLI M, MARCHESINI I, REICHENBACH P, et al. Automatic delineation of geomorphological slope units with r. slopeunits v1.0 and their optimization for landslide susceptibility

- modeling [J]. *Geoscientific Model Development*, 2016, 9(11): 3975–3991.
- [5] SUN Xiaohui, CHEN Jianping, HAN Xudong, et al. Landslide susceptibility mapping along the upper Jinsha River, southwestern China: A comparison of hydrological and curvature watershed methods for slope unit classification [J]. *Bulletin of Engineering Geology and the Environment*, 2020, 79(9): 4657–4670.
- [6] 黄立鑫, 郝君明, 李旺平, 等. 基于 RBF 神经网络-信息量耦合模型的滑坡易发性评价——以甘肃岷县为例 [J]. *中国地质灾害与防治学报*, 2021, 32(6): 116–126. [HUANG Lixin, HAO Junming, LI Wangping, et al. Landslide susceptibility assessment by the coupling method of RBF neural network and information value: A case study in Min Xian, Gansu Province [J]. *The Chinese Journal of Geological Hazard and Control*, 2021, 32(6): 116–126. (in Chinese with English abstract)]
- [7] 白光顺, 杨雪梅, 朱杰勇, 等. 基于证据权法的昆明五华区地质灾害易发性评价 [J]. *中国地质灾害与防治学报*, 2022, 33(5): 128–138. [BAI Guangshun, YANG Xuemei, ZHU Jieyong, et al. Susceptibility assessment of geological hazards in Wuhua District of Kunming, China using the weight evidence method [J]. *The Chinese Journal of Geological Hazard and Control*, 2022, 33(5): 128–138. (in Chinese with English abstract)]
- [8] 李光辉, 铁永波, 白永建, 等. 则木河断裂带(普格段)地质灾害发育规律及易发性评价 [J]. *中国地质灾害与防治学报*, 2022, 33(3): 51. [LI Guanghui, TIE Yongbo, BAI Yongjian, et al. Distribution and susceptibility assessment of geological hazards in Zemuhe fault zone(Puge section) [J]. *The Chinese Journal of Geological Hazard and Control*, 2022, 33(3): 51. (in Chinese with English abstract)]
- [9] 王涛, 刘甲美, 栗泽桐, 等. 中国地震滑坡危险性评估及其对国土空间规划的影响研究 [J]. *中国地质*, 2021, 48(1): 21–39. [WANG Tao, LIU Jiamei, LI Zetong, et al. Seismic landslide hazard assessment of China and its impact on national territory spatial planning [J]. *Geology in China*, 2021, 48(1): 21–39. (in Chinese with English abstract)]
- [10] ZHAO Shuai, ZHAO Zhou. A comparative study of landslide susceptibility mapping using SVM and PSO-SVM models based on grid and slope units [J]. *Mathematical Problems in Engineering*, 2021, 2021: 1–15.
- [11] XIAO Ting, YIN Kunlong, YAO Tianlu, et al. Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China [J]. *Acta Geochimica*, 2019, 38(5): 654–669.
- [12] YU Chenglong, CHEN Jianping. Application of a GIS-based slope unit method for landslide susceptibility mapping in Helong city: comparative assessment of ICM, AHP, and RF model [J]. *Symmetry*, 2020, 12(11): 1848.
- [13] 常宏, 金维群, 王世昌, 等. 清江中下游岸坡稳定性对河流地貌过程的响应 [J]. *工程地质学报*, 2011, 19(5): 756–763. [CHANG Hong, JIN Weiqun, WANG Shichang, et al. Slope stability response to river geomorphic process in middle and lower reaches of Qingjiang River [J]. *Journal of Engineering Geology*, 2011, 19(5): 756–763. (in Chinese with English abstract)]
- [14] 常宏, 谭建民, 李明. 清江河道变迁与滑坡崩塌 [M]. 武汉: 中国地质大学出版社, 2018. [CHANG Hong, TAN Jianmin, LI Ming. Changes of Qingjiang River and landslide collapse [M]. Wuhan: China University of Geosciences Press, 2018. (in Chinese)]
- [15] 陈伟. 山区村镇滑坡灾害风险评估研究 [D]. 武汉: 武汉大学. [CHEN Wei. Study on risk assessment of landslide disaster in mountainous villages and towns [D]. Wuhan: Wuhan University. (in Chinese with English abstract)]
- [16] GUO Zizheng. Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management [J]. *Geoscience Frontiers*, 2021, 12(6): 101249.
- [17] 王峰. 四川省南江县地质灾害易发性区划研究 [D]. 成都: 成都理工大学. [WANG Feng. Study on regionalization of geological disaster susceptibility in Nanjiang County, Sichuan Province [D]. Chengdu: Chengdu University of Technology. (in Chinese with English abstract)]
- [18] 罗路广, 裴向军, 崔圣华, 等. 九寨沟地震滑坡易发性评价因子组合选取研究 [J]. *岩石力学与工程学报*, 2021, 40(11): 2306–2319. [LUO Luguang, PEI Xiangjun, CUI Shenghua, et al. Combined selection of susceptibility assessment factors for Jiuzhaigou earthquake-induced landslides [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2021, 40(11): 2306–2319. (in Chinese with English abstract).]
- [19] 王锐. 基于 GIS 和 Logistic 回归模型的降雨型滑坡易发性研究 [D]. 杭州: 浙江大学. [WANG Rui. Study on the susceptibility of rainfall landslide based on GIS and Logistic regression model [D]. Hangzhou: Zhejiang University. (in Chinese with English abstract)]
- [20] 吴润泽, 胡旭东, 梅红波, 等. 基于随机森林的滑坡空间易发性评价——以三峡库区湖北段为例 [J]. *地球科学*, 2021(1): 321–330. [WU Runze, HU Xudong, MEI Hongbo, et al. Spatial susceptibility assessment of landslides based on random forest: A case study from Hubei section in the Three Gorges Reservoir area [J]. *Earth Science*, 2021(1): 321–330. (in Chinese with English abstract)]
- [21] 郝国栋. 基于随机森林模型的商南县滑坡易发性评价 [D]. 西安: 西安科技大学, 2020. [HAO Guodong. Evaluation of landslide susceptibility in Shangnan County based on random forest model [D]. Xi'an: Xi'an University of Science and Technology, 2020. (in Chinese with English abstract)]