

文章编号: 1001-1749(2023)03-0403-07

基于地理加权回归模型的区域滑坡易发性评价

栗 解, 陈建华

(成都理工大学 地球物理学院, 成都 610059)

摘要: 滑坡对人员和经济造成了巨大威胁, 因此对区域滑坡进行易发性评价至关重要。为了最低化滑坡带来的损失, 以 2013 年四川省芦山县滑坡集中区域滑坡点为源数据, 选取 6 种评价因子, 采用多重共线性分析确定因子可用性, 使用地理加权回归模型对研究区进行滑坡易发性评价, 采用线性回归模型与其对比分析, 使用自然断点法生成滑坡易发性评价图。结果表明: 地理加权回归模型评价结果优于线性回归模型, 更适合区域性滑坡评估; 芦山滑坡点集中区域的滑坡发生受植被覆盖和断层影响较大; 在评价结果划分的等级中, 极高—高等级区域内的滑坡点占比和点密度最大, 极低—低等级区域最小, 说明评价结果与实际情况一致; 地理加权回归模型评价结果图可直观地分辨出各区域滑坡程度的分布范围, 可为该区域防灾减灾、土地规划和经济发展提供较好的参考依据。

关键词: 区域滑坡; 地理加权回归; 线性回归; 易发性

中图分类号: P 642.22 **文献标志码:** A **DOI:** 10.3969/j.issn.1001-1749.2023.03.16

0 引言

滑坡已成为世界上发生频次最多的地质灾害之一, 给人员和经济造成了巨大的损失, 破坏了大量基础设施^[1]。同时滑坡时刻威胁着社会、环境和经济的安全, 因此对某一区域滑坡进行合理的易发性评价有助于滑坡灾害防治、治理和决策。

滑坡是一种地理空间分布事件, 其评价存在错综复杂的空间问题。回归模型分析是解决空间问题的重要手段和方法, 一般将回归分析模型分为线性回归模型 (Ordinary Linear Regression, OLR) 和地理加权回归模型 (Geographically Weighted Regression Model, GWR)^[2]。OLR 模型就是假设回归系数随着位置的改变不发生变, 而 GWR 模型的系数是随着空间位置的改变而改变的。GWR 是近几年

新的空间分析方法之一, 它是将空间结构嵌入到线性回归方程模型当中, 以此用来探测空间关系的非平稳性^[3-4]。

回归分析模型在多领域得到应用分析, Kang 等^[5]利用地理加权回归模型获得降水与 NDVI 空间相关性, 为农业管理提供帮助; 祝新明等^[6]利用地理加权回归对地表温度进行降尺度研究, 获取地表参数与温度之间的度量关系; Zhang 等^[7]用地理加权模型结合多尺度高速公路交通流来解释空间相互作用的异质性; Fu 等^[8]利用地理加权回归预测土壤重金属空间分布, 获得空间金属分布格局更合理, 精度更高; Lesch 等^[9]使用线性回归为基础, 推导出的空间预测模型在表征土壤和作物空间变化方面表现良好; 周茂袁等^[10]使用线性回归对武汉疫情解封时间进行预测, 结果与实际预期高度吻合, 准确性高。

收稿日期: 2022-02-11

基金项目: 四川省科技计划项目 (2019YFG0187)

第一作者: 栗解 (1996—), 男, 硕士, 研究方向为空间分析模型与方法, E-mail: jax942583379@163.com。

通信作者: 陈建华 (1976—), 男, 副教授, 硕士生导师, 主要从事空间模型分析研究, E-mail: chjh3@163.com。

以往研究表明,回归分析模型应用相对普遍,其解决领域问题的结果也较理想,但在滑坡评价方面应用相对较少。因此选择芦山县滑坡集中区域为研究区,应用地理加权回归模型进行实验研究,采用线性回归模型与其对比分析,以获得与区域真实情况更接近的滑坡易发性结果图,为未来的区域滑坡灾害防范和经济发展工作提供可靠参考。

1 研究方法

地理加权回归模型是典型的统计学方法之一,它是用回归原理研究具有空间或区域分布特征的两个或多个变量之间数量关系的方法。线性回归模型是利用数理统计中回归分析,来确定多种变量间相互依赖的定量关系的一种统计分析方法。

1.1 地理加权回归模型

地理加权算法的核心是地理加权方程的建立,其具体建立方式如下。

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i) x_{i1} + \dots + \beta_p(u_i, v_i) x_{ip} + \epsilon_i$$

$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad (1)$$

其中: y_i 与 $x_{i1}, x_{i2}, \dots, x_{ip}$ 是因变量 y 与自变量 $x_{i1}, x_{i2}, \dots, x_{ip}$ 在位置 (u_i, v_i) 的观测值; (u_i, v_i) 是第 i 个采样点的坐标; $\beta_k(u_i, v_i)$ 是第 i 个采样点上的第 k 个回归参数;点 i 被当作回归点。 $\epsilon_i \sim N(0, \sigma^2)$,当 $i \neq j$ 时, ϵ_i 与 ϵ_j 的协方差 $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ 。那么方程就简单记为式(2)。

$$y_i = \beta_{i0} + \sum_{k=1}^p \beta_{ik} x_{ik} + \epsilon_i$$

$$i = 1, 2, \dots, n \quad (2)$$

$$f(\beta) = \sum_{i=1}^n [y_i - \beta_0(u_0, v_0) \dots - \beta_p(u_0, v_0)]^2 \omega_i(u_0, v_0) \quad (3)$$

式(3)中: (u_0, v_0) 为研究区的任意一点; y_i 为因变量; $\beta_0 \dots \beta_p$ 为该点回归系数; ω_i 为该点权重。

$$\begin{cases} \hat{\beta} = (X^T W(i) X)^{-1} X^T W(i) Y \\ W(i) = \begin{bmatrix} \omega_{i1} & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \omega_{in} \end{bmatrix} \\ \omega_{ik} = \exp(-(d_{ij}/b)^2) \\ b = c - (1 - \delta) | c - a | \\ \delta = 0.618 \end{cases} \quad (4)$$

式(4)中: $\hat{\beta}$ 为系数 β 的加权最小二乘估计量,而不是具有恒定权重的矩阵; ω_{ik} 为选择的高斯权函数;

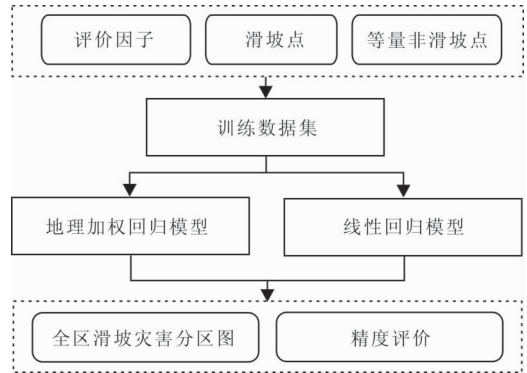


图1 技术路线

Fig. 1 Technical route

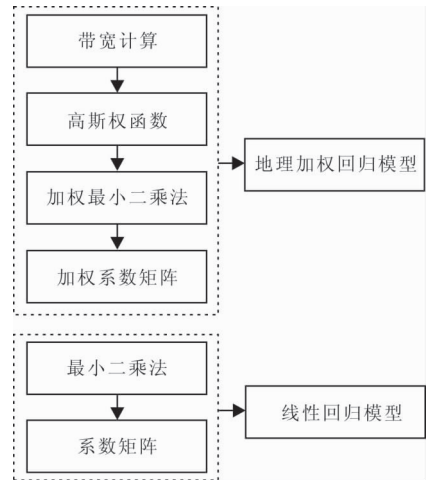


图2 模型构建

Fig. 2 Model building

d_{ij} 为两点之间的距离; b 为选择的最佳带宽。最优带宽的计算采用黄金分割法、点之间最大距离 a 和最小距离 c 。地理加权的结果对内核函数的选择相对不敏感,但它敏感的是选择的特定的内核函数带宽。在上述过程中,空间权重系数选取 $\hat{\beta}$ 每一列的最小值。

1.2 线性回归模型

线性回归模型算法核心是线性回归方程的建立,建立方式如下:

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i + \epsilon$$

$$i = 1, 2, \dots, n \quad (5)$$

式(5)中: y_i 与 x_1, x_2, \dots, x_i 构成因变量和自变量的关系; β_i 是第 i 个因素的回归参数; ϵ 为回归方程的误差项。

1.3 技术路线

实验技术路线如图1所示。

1) 构建滑坡灾害易发性评价数据集,选取合理

的评价因子、已知滑坡点和等量的非滑坡点构成该研究区滑坡灾害易发性评价数据集,并将其按照随机抽样 8:2 的比例分为训练和验证数据集。

2) 训练回归分析模型,将训练数据集分别应用于回归模型中。结合带宽计算,其最佳带宽为 948,使用高斯函数确认权重、加权最小二乘法确定加权系数矩阵,以此确立地理加权回归模型。使用最小二乘法计算系数矩阵,确定线性回归模型。模型构建流程见图 2。

3) 全区滑坡灾害分级图。利用自然断点法将全区预测值等级分为极低、低、中、高和极高 5 个灾害易发性等级。

4) 精度评价。评价模型预测能力的采用验证数据集,判断其优劣采用受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC)和在坐标轴中曲线下的面积值(Area under the Curve, AUC)来进行分析。

2 实验与分析

2.1 研究区概况

芦山县隶属于四川省雅安市,位于四川盆地周山区西缘,雅安东北部,青衣江上游,北与汶川县连界,东与成都毗邻,西接甘孜。县内地质构造复杂,年降雨量大,其相对海拔差较大在 557 m~5 289 m 之间,地势北高南低,呈南北狭长山岳河谷地貌。在其南部,地势较平坦,人群较密集,发生的滑坡也更多(图 3)。受 2013 年地震影响,该县域内滑坡灾害频发,对人民的生命财产安全造成了巨大的损失,同时也时刻威胁着该区域内的群众。对芦山县高危区域进行滑坡易发性评价和分级,能有效地减少往后芦山滑坡造成的损失,也能对该县域的防灾和发展提供有效的辅助决策支持。

2.2 数据处理

影响滑坡的评价因子众多,合理地选取评价因子对模型的构建至关重要。为了使实验正常开展,在多数数据源头获取基础数据,比如:在地理空间数据云获取 30 m 数字高程影像,在欧空局获取 2013 年哨兵影像等。

滑坡因子影响滑坡发生的方式有差异,如归一化植被指数反映区域植被覆盖程度,能体现基本地质条件的稳定性。地震动峰值加速度反映区域地表地质点在 earthquake 影响下的加速度值,体现区域地质稳

定受地震影响程度,其值与滑坡影响力呈正相关性。土地类型反映不同土地的利用情况,表现了该区域的土地土壤含水量及地表结构稳定程度,含水量多的土壤以及地表结构的改变都会增加滑坡发生概率。断层影响区域地质破碎程度,距其远近反映岩石承力强度,其较低承力强度会加剧滑坡的发生。坡度和起伏度反映斜坡的型体特征,在一定程度上影响滑坡物质方向。所以依靠选取评价因子的客观存在性、显著性和继承性三原则^[11],同时通过对研究区的地域环境分析,选取了归一化植被指数、地震动峰值加速度、土地类型、断层、坡度和起伏度共计 6 种滑坡评价因子,在 6 种评价因子数据中,包含了描述性、连续性和离散性数据,然而实验的开展需要离散数据,故需使用 GIS 软件对其数据进行分级处理。归一化植被指数和土地类型分类按影响滑坡的大小分为 9 类;地震动峰值加速度的提取是先将面转栅格,再按照值的大小分为 9 类;断层是先将断层多层缓冲,再缓冲面数据转为栅格数据,根据点到断层的距离分为 9 类,距离越近,值越大;坡度和起伏度分类按照坡度大小和起伏度大小分为 9 类。为将其进行离散化,所有的 9 类数据按照数值 1-9 赋值。

在 GIS 软件中,处理 602 个滑坡点,生成与滑坡点等量的随机点。将滑坡点赋值为“1”,随机点赋值为“0”,将其合并组成一个 1 204 个预样本点数据,将其中随机抽样的 963 个点作为训练样本,剩下的 241 个点作为验证样本。按位置匹配每个滑坡评价因子的值,并将其值提取至样本点中,将其数据作为 GWR 和 OLR 模型构建和验证数据集。为便于模型预测输入,采用研究区的 590 571 个 30 m×30 m 网格单元的中心点组成全区预测集。

2.3 多重共线性分析

滑坡受到多因子影响,各因子之间存在相互制约、相互促进或互不影响的关系。在选取评价因子作为输入的时候,有必要对多个因子进行多重共线性分析来分析因子的可用性。多重共线性分析结果主要包含方差膨胀因子(Variance Inflation factor, VIF)及容忍度(Tolerance, TOL),一般以 5 和 0.1 作为 VIF 和 TOL 的临界值。在分析结果中同时满足 TOL 大于 0.1 且 VIF 小于 5,则因子间不存在共线性问题,因子可使用;否则因子间存在共线性问题,该因子不可使用。以此,基于 SPSS 软件对各因子进行多重共线性分析,其分析结果如表 1 所示。

表 1 多重共线性分析结果

Tab.1 Multicollinearity analysis results

自变量因子	TOL	VIF
归一化植被指数	0.888	1.126
地震动峰值加速度	0.851	1.175
土地类型	0.984	1.016
距断层距离	0.911	1.098
坡度	0.323	3.095
起伏度	0.323	3.095

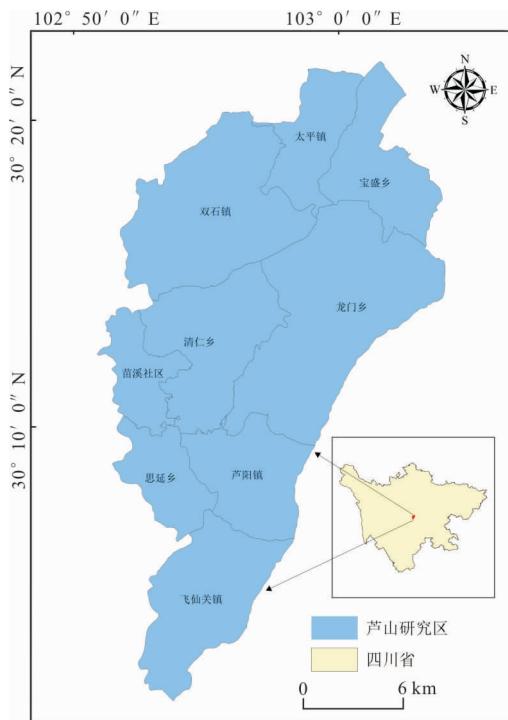


图 3 研究区概况图

Fig. 3 Location of study area

由表 1 可知, TOL 和 VIF 都没有超出规定界限, 表明所选取的评价因子不存在多重共线性问题, 在该实验中使用。

2.4 结果分析

研究采用 GWR 和 OLR 模型对滑坡易发性进行评价, 通过其模型对全区的 590 571 个网格进行滑坡易发性评价预测。为了避免人为干扰, 将预测平面按照自然断点法分类并将其风险等级分为极低、低、中、高、极高五个级别, 并将其制作成芦山县滑坡高风险地区易发性评价分级图(图 4)。

从图 4(a) 可知, 滑坡主要发生在上部区域, 大部分滑坡点发生的区域都与该图层的高危区域分布比较接近。特别是在滑坡点越密集的区域, 预测值

就越高, 发生滑坡的概率就越大。在非滑坡点区域内, 预测的值大多数都偏向于中低等级。其极高—高等级区域较多分布在断层及植被覆盖较少的区域, 这是由于少植被的区域土壤疏松及断层周围的岩体破碎, 岩土位置易发生变化。从图 4(b) 中可知, 滑坡点所在的区域与高等级及以上区域大部分重合, 其周围区域也都偏向于高等级及以上, 呈面状分布。统计结果如表 2 所示, GWR 评价结果的极高—高等级区域面积占总面积的比例适中, 落在其中的滑坡点数目最多, 相对滑坡点密度最大, 在其区域内的滑坡分布更集中; 落在其极低—低等级区域的滑坡点数目最小, 相对滑坡点密度最小, 在其区域内发生滑坡概率较小, 符合实际情况。而 OLR 评价结果的极高—高等级区域面积占比与极低—低等级区域面积占比接近, 其滑坡点密度在极高—高等级区域最大, 极低—低等级区域内最小, 但其差距不够明显。从图 4 可知, 在 OLR 模型的评价结果中, 高等级及以上区域片状分布更广和密集, 不符合正常生活环境, 而 GWR 模型高等级及以上区域分布更稀疏, 中及以下等级区域分布符合真实情况; GWR 模型在高等级及以上区域滑坡点密度大于 OLR, 在低等级及以下区域滑坡点密度小于 OLR。从图 4(c) 和图 4(d) 可知, GWR 模型评价结果局部区域内呈现高中依次分布, 符合实际情况, OLR 模型结果在区域内就呈现高中低跨度分布, 与实际情况相斥。上述表明, GWR 模型的评价结果优于 OLR 模型。

为了进一步检验模型预测结果的可靠性和准确性, 模型的精度检验是必须的。常用的模型检验方法是受试者工作特征曲线, 即为 ROC 曲线, 其原理是把灵敏度和特异度结合起来综合评价结果的准确度或判别效果的一种方法。AUC 来表示模型的预测能力即模型的优劣。AUC 值介于 0.5~1 之间, AUC 值越接近于 1, 说明模型的预测能力越好。AUC 值介于 0.5~0.7 之间时, 模型预测具有较低的准确性; AUC 值介于 0.7~0.9 时, 模型预测具有一定的准确性; AUC 值在 0.9 以上时, 模型具有较高准确性; AUC 值低于 0.5 时, 模型预测能力较差, 此时的预测结果不予使用。

利用 ROC 曲线对地理加权模型的预测结果评价如图 5 所示, 其 GWR 和 OLR 模型 AUC 值分别为 0.78、0.76, 证明其预测具有一定应用意义。根据上述指标, GWR 模型的 AUC 值大于 OLR, 结果表明, GWR 预测能力优于 OLR。

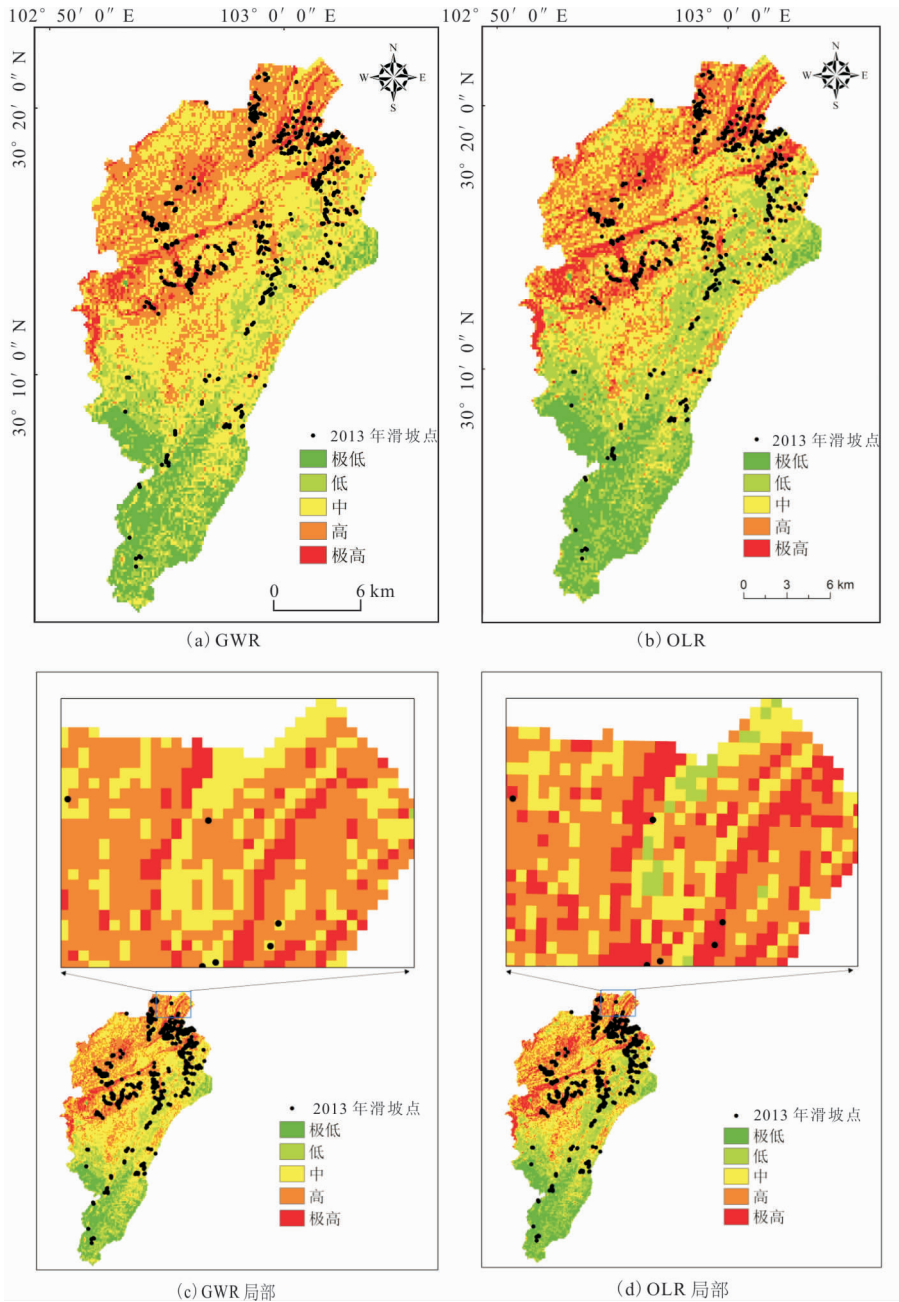


图 4 芦山滑坡集中地区易发性评价结果图

Fig. 4 Map of susceptibility evaluation results of Lushan landslide concentrated area

表 2 评价统计结果

Tab. 2 Evaluation statistical results

统计项	模型	极高	高	中	低	极低
面积占比/%	GWR	4.9	29.7	42.2	21.3	2.9
	OLR	9.1	26.4	28.9	23	13.6
滑坡点数目	GWR	77	268	207	48	2
	OLR	157	184	170	81	10
滑坡点占比/%	GWR	12.8	44.5	34.4	8	0.3
	OLR	26.1	30.6	28.2	13.4	1.7
相对滑坡点密度	GWR	2.612	1.498	0.815	0.376	0.103
	OLR	2.868	1.159	0.975	0.583	0.125

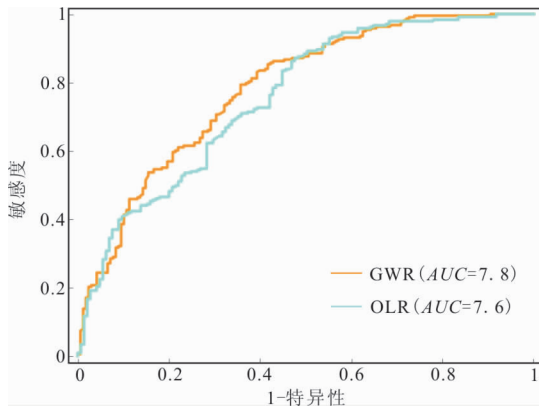


图 5 ROC 曲线检验结果

Fig. 5 ROC curve test results

3 结论

笔者以芦山滑坡集中区域为研究区,选取 6 种评价因子并对其进行分级处理及多重共线性分析,结合地理加权回归模型对该区域进行滑坡易发性评价,使用线性回归模型与其对比分析,采用 ROC 曲线和 AUC 值对模型的预测精度进行评价。实验获得相关结论如下:

1) 地理加权回归模型评价结果优于线性回归模型,更适合区域性滑坡评估。

2) 在研究区评价结果划分的极低、低、中、高和极高五个等级中,高等级及以上区域内的滑坡点占比和点密度最大,低等级及以下区域最小,说明评价结果与实际情况一致。

3) 地理加权回归模型评价结果图,可直观地分辨出各区域滑坡程度的分布范围,由其可知,芦山滑坡点集中区域的滑坡发生受植被覆盖和断层影响较大,在后期滑坡防治上,可以在植被覆盖较少的区域和距离断层较近的区域进行初步危险隐患排查,必要时对其添加抗滑桩或者挂网喷浆,同时评价结果图能为决策者在未来区域的经济发展规划上提供可靠的科学依据和支撑。

参考文献:

[1] 陈燕平. 基于 GIS 的贵州省滑坡地质灾害易发性多模型综合评价 [D]. 长沙: 中南大学, 2010.
CHEN Y P. Multi-model comprehensive evaluation of landslide geological hazard susceptibility in Guizhou Province based on GIS [D]. Changsha: Central South University, 2010. (In Chinese)

[2] 覃文忠, 王建梅, 刘妙龙. 混合地理加权回归算法研究 [J]. 武汉大学学报, 2007, 32(2): 115-116.
QIN W Z, WANG J M, LIU M L. Research on hybrid geo-weighted regression algorithm [J]. Journal of Wuhan University, 2007, 32(2): 115-116. (In Chinese)

[3] 覃文忠. 地理加权回归基本理论与应用研究 [D]. 上海: 同济大学, 2017.
QIN W Z. Research on the basic theory and application of geographically weighted regression [D]. Shanghai: Tongji University, 2017. (In Chinese)

[4] CHRIS BRUNSDON, A. STEWART FOTHERINGHAM, MARTIN E. CHARLTON. Geographically weighted Regression: a method for exploring spatial nonstationarity [J]. Geographical Analysis, 1996, 28(4): 281-298.

[5] KANG L, DI L, DENG M, et al. Use of geographically weighted regression model for exploring spatial patterns and local factors behind NDVI-Precipitation correlation [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote sensing, 2014, 7(11): 4530-4538.

[6] 祝新明, 宋小宁, 冷佩, 等. 多尺度地理加权回归的地表温度降尺度研究 [J]. 遥感学报, 2021, 25(08): 1749-1766.
ZHU X M, SONG X N, LENG P, et al. Downscaling of land surface temperature based on multi-scale geo-weighted regression [J]. Journal of Remote Sensing, 2021, 25(08): 1749-1766. (In Chinese)

[7] ZHANG L, CHENG J, JIN C, et al. A multiscale Flow-Focused geographically weighted regression modeling approach and its application for transport flows on expressways [J]. Applied Sciences, 2019, 9(21): 4673.

[8] FU P, YANG Y, ZOU Y S. Prediction of soil heavy metal distribution using geographically weighted regression kriging [J]. Bulletin of Environmental Contamination and Toxicology, 2022, 108(2): 344-350.

[9] S. M. LESCH, D. L. CORWIN. Prediction of spatial soil property information from ancillary sensor data using ordinary linear regression: Model derivations, residual assumptions and model validation tests [J]. Geoderma, 2008, 148(2): 130-140.

[10] 周茂袁, 崔宁, 王秀丽, 等. 基于加权线性回归模型的武汉解封时间的预测 [J]. 应用概率统计, 2021, 37(02): 192-200.
ZHOU M Y, CUI N, WANG X L, et al. Prediction of unsealing time in Wuhan based on weighted linear regression Model [J]. Applied Probability and Statistics,

2021,37(02):192-200. (In Chinese)

- [11] 李利峰, 张晓虎, 邓慧琳, 等. 基于 SVM-LR 融合模型的滑坡灾害易发性评价——以山阳县为例[J]. 科学与技术工程, 2020, 20(26):10618-10625.

LI L F, ZHANG X H, DENG H L, et al. Landslide

susceptibility assessment based on SVM-LR fusion model: A case study of Shanyang County [J]. Science Technology and Engineering, 2020, 20(26):10618-10625. (In Chinese)

Evaluation of regional landslide susceptibility based on geographically weighted regression model

SU Jie, CHEN Jianhua

(Chengdu University of Technology, College of Geophysics, Chengdu 610059, China)

Abstract: Landslides pose a significant threat to people and the economy. Therefore, it is imperative to evaluate the susceptibility of regional landslides. For the loss minimization of landslide, this paper takes the 2013 fourth county of Sichuan province area landslide point as the source data, selecting six kinds of evaluation factors, the multicollinearity analysis to determine availability factor, using a geographically weighted regression model to evaluate landslide liability in the study area, using a linear regression model with the comparative analysis, Natural breakpoint method is used to generate landslide susceptibility evaluation map. The results show that the evaluation result of a geographically weighted regression model is better than that of a linear regression model, and it is more suitable for regional landslide evaluation. The landslide occurrence in Lushan's concentrated area is greatly influenced by vegetation cover and fault. Among the grades of the evaluation results, the proportion and density of landslide points are the largest in the extremely high-high grade area and the smallest in the extremely low-low grade area, indicating that the evaluation results are consistent with the actual situation. Furthermore, the geographically weighted regression model can intuitively distinguish the distribution range of landslide degrees in each region, providing a good reference for disaster prevention and mitigation, land planning, and economic development in this region.

Keywords: regional geological landslide; geographically weighted regression; general linear regression; susceptibility