

• 滑坡堰塞湖灾害机理与防控 •

DOI:10.12454/j.jsuese.202400283



本刊网刊

## 考虑空间异质性的降雨滑坡易发性预测研究

张幸福<sup>1,2</sup>, 姜元俊<sup>3\*</sup>, 阿比尔的<sup>1,2</sup>

(1. 重庆交通大学 河海学院, 重庆 400074; 2. 重庆交通大学 山区公路水运交通地质灾害重庆市教委重点实验室, 重庆 400074;  
3. 中国科学院 水利部 成都山地灾害与环境研究所, 四川 成都 610041)

**摘要:** 现有滑坡易发性预测方法未能充分考虑地形、土壤和植被等环境因素的空间异质性, 也无法准确反映极端降雨对滑坡易发性的影响。为了克服这些限制, 引入了一种结合深度嵌入聚类(DEC)的动态雨量阈值分区方法, 通过深度学习技术, 根据环境因素将研究区域划分为具有相似特征子区域, 实现了滑坡预测模型的精细化空间异质性分析; 在此基础上, 提出基于混合分布的动态雨量阈值模型以区分非极端降雨与极端降雨, 并采用贝叶斯方法动态更新模型参数, 提高了模型对不同降雨类型的适应性和预测的时效性。以通江县为案例, 采用多任务学习自适应神经树模型(MLANT), 结合深度嵌入 DEC 模型与混合分布阈值模型, 对滑坡易发性进行预测。结果表明, 本文方法在精确度、F1 分数及受试者工作特征曲线下面积 AUC 值等关键性能指标上显著优于传统依赖统一阈值的模型。特别是与传统的基于前期有效降雨量方法相比, 预测效果提升显著, 预测滑坡密度和数量由 0.038 事件/km<sup>2</sup> 和 44 个滑坡事件提升至 0.044 事件/km<sup>2</sup> 和 59 个滑坡事件, 充分证实了在滑坡易发性预测中使用深度嵌入聚类(DEC)的动态雨量阈值分区考虑空间异质性和区分不同降雨事件的重要性和有效性。

**关键词:** 滑坡易发性; 深度嵌入聚类(DEC); 空间异质性; 混合分布降雨阈值; 多任务学习自适应神经树模型

中图分类号: P642.22

文献标志码: A

文章编号: 2096-3246(2025)04-0012-17

滑坡是一种严重威胁人类生命和财产安全的频发自然灾害。根据国际灾害数据库(EM-DAT), 2014 年至 2018 年, 全球因滑坡造成的死亡人数达到 4 914 人, 27 110 人流离失所, 经济损失高达 21 亿美元。这种灾害的频发性和破坏力, 迫切需要科学有效的预测模型来评估其易发性, 以减轻其对社会的影响。

在滑坡易发性模型研究中, 物理模型、启发式模型、统计模型及机器学习模型被广泛应用。物理模型如 Whiteley 等<sup>[1]</sup>所展示, 尽管其预测精度较高, 但因复杂的数据需求不适合大规模应用。启发式模型依赖专家知识, 如 Basu 等<sup>[2]</sup>所述, 这类模型因主观性强而难以客观评估。统计模型, 如逻辑回归、频率比率和证据权重法被广泛采用, 但 Aditian<sup>[3]</sup>和 Chen<sup>[4]</sup>等的研究显示, 这些模型在处理滑坡条件因素的复杂非线性关系时存在局限。机器学习方法, 如 Huang 等<sup>[5]</sup>提出的支持向量机(SVM)、Merghadi 等<sup>[6]</sup>提出的随机森林(LR)以

及 Youssef 等<sup>[7]</sup>提出的人工神经网络, 在提取非线性关系和大规模数据处理上表现出色, 然而这些模型在处理动态环境变化时常显不足, 特别是在需要同时适应多种预测任务和快速变化的环境条件时, 其适应性和灵活性有限。

空间异质性是滑坡易发性研究中的核心问题, 涉及滑坡影响因素与滑坡事件之间的关系在不同地理位置的显著变化。传统模型, 如地理加权回归(GWR)被广泛用于捕捉这种空间变异性, 如 Hong 等<sup>[8]</sup>在将 GWR 与 SVM 和 LR 进行对比时发现, GWR 在捕捉局部空间关系方面更具优势。然而, GWR 模型假定地质现象的线性关系, 并依赖简单的空间核函数生成权重矩阵, 这些预定义的距离衰减函数难以准确捕捉复杂的非线性关系而影响预测精度<sup>[9]</sup>。此外, GWR 采用的最佳平均尺度未能充分考虑多变量空间数据特征的不同尺度变化, 导致在复杂滑坡预测任务中的性能

收稿日期: 2024-04-26 修回日期: 2024-08-09 网络出版日期: 2025-02-25

基金项目: 国家自然科学基金项目(42172320)

作者简介: 张幸福(1998—), 男, 硕士。研究方向: 滑坡易发性。E-mail: zhangxingfu001@163.com

\* 通信作者: 姜元俊, 研究员, E-mail: yjjiang@imde.ac.cn

不足。基于此,尽管有研究尝试解决空间异质性问题,但大多未能充分实现。如:Yang等<sup>[10]</sup>提出的局部尺度滑坡易发性模型仅考虑了控制因素的局部空间异质性,未涉及滑坡本身的空间异质性。Deng等<sup>[11]</sup>通过在各空间聚类中建立不同的局部模型,在预测细颗粒物方面表现良好,但该方法未能在整个研究区域内建立一个有效的全局滑坡模型,因而在处理广泛分布的地理数据时表现不足。Wang等<sup>[12]</sup>采用了k-means方法将网格单元进行分组,在一定程度上考虑了地理环境的影响,但由于k-means方法主要依赖距离度量,并未被设计用来直接处理数据的空间或地理特性,故在处理地形、土壤和植被等具有显著空间异质性的环境因素时,往往无法充分捕捉这些因素的复杂非线性特征。

极端降雨作为滑坡的重要触发因素,已经在全球范围内受到广泛关注。随着气候变化加剧,极端降雨事件的频率和强度显著增加,这直接导致了滑坡灾害增多。在过去的几十年中,研究人员开展了多种方法评估和预测由极端降雨引发的滑坡风险。其中,物理模型是一种常见的方法,这类模型结合坡面稳定性分析和水文模型来评估降雨如何影响坡面的稳定性。这些模型通常需要详细的地质数据和精确的水文条件输入,能够在理论上为滑坡预测提供坚实的科学基础<sup>[13]</sup>。然而,物理模型在应用中存在显著的局限性。首先,它们对数据的需求极高,包括高分辨率的数字高程模型(DEM)和详尽的降雨数据,这使得模型在广阔地区的应用变得计算成本高昂且操作复杂<sup>[14]</sup>。

与物础模型相对的是历史统计分析方法,这类方法通过分析历史滑坡记录来估计未来滑坡的概率。这种方法的优点在于操作简便且易于实现,可以快速评估一个地区的滑坡风险<sup>[15]</sup>。然而,统计方法的有效性严重依赖于历史数据的完整性和准确性。在很多情况下,尤其是在缺资料地区,长期、完整的滑坡记录难以获取,这限制了该方法的应用范围;更为关键的是,统计方法基于历史事件的重现性进行预测,对于非周期性或由罕见极端降雨事件引发的独特滑坡,其预测能力往往不足<sup>[16]</sup>。

综上,现有模型在适应多种预测任务和应对快速变化的环境条件时常表现出适应性和灵活性不足的问题。特别是在处理地形、土壤和植被等具有显著空间异质性的环境因素时,这些模型往往无法有效捕捉其复杂非线性特征。此外,对于非周期性或由罕见极端降雨事件引发的独特滑坡,现有方法的预测能力通常不足。为了解决这些问题,本文提出以下创新方法:

1)针对k-means方法在处理空间或地理特性时的

局限,提出深度嵌入聚类分析(DEC)方法,有效捕捉地形、土壤和植被等环境因素的复杂非线性关系,解决传统聚类技术在空间异质性处理上的不足。

2)针对传统模型在预测由极端降雨引发的非周期性滑坡事件方面能力不足的问题,提出基于混合分布模型的动态雨量阈值模型,有效区分了极端降雨与非极端降雨对滑坡易发性预测影响,考虑了由极端降雨引发的滑坡风险,提高了极端气候条件下非周期性滑坡事件的预测能力。

3)针对传统物理、启发式、统计和机器学习模型在适应多种预测任务和快速变化的环境条件时的局限性,提出多任务学习自适应神经树模型(MLANT),动态调整其处理策略以适应不同的预测任务和环境变化,提升模型在处理空间异质性和极端气候事件预测任务中的灵活性和准确性,尤其在面对复杂的非线性环境因素时,比传统单一任务模型表现出更高的适应性和预测能力。

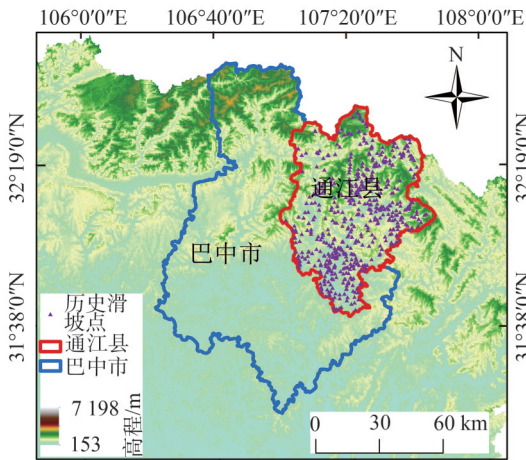
## 1 研究区域概况及滑因子选取

### 1.1 研究区域概况

本文以中国四川省巴中市通江县的滑坡易发性为研究对象,该区域以复杂的地质结构和多变的气候条件而闻名。研究区面积约为4 123 km<sup>2</sup>,位于四川省东北部,横跨云贵高原向四川盆地的过渡区,如图1所示,其地形主要由山地和丘陵组成。该县所处的地理位置及其独特的地质条件,使得该地区成为降雨型滑坡的高发区域。通江县属于亚热带季风气候,年均降雨量介于1 000至1 500 mm之间,特别是在5至9月的雨季,降雨量占全年总量的60%以上。

季节性和短时强降雨导致地表径流量迅速增加,使土壤迅速饱和,尤其在地形陡峭的区域。饱和状态增加了土体重量,降低了抗剪强度,显著提高了滑坡风险。在通江县,疏松的砂质土壤和高可塑性的黏土在饱和后特别容易失去内聚力和摩擦力,是滑坡发生的主要原因。多山地形加剧了地表水的积聚和下渗,使土壤在雨季时更易饱和<sup>[17]</sup>。浅层土壤因接触面积大,受降雨直接影响显著,水分迅速聚集,增加负荷,导致浅层土质边坡滑坡高发。因此,本文特别关注通江县浅层土质边坡的滑坡易发性<sup>[18]</sup>。

从2010年至今,通江县记录的滑坡事件主要发生在降雨集中的月份,显示出明显的季节性和降雨关联性,这些滑坡事件不仅对人民的生命和财产安全构成严重威胁,还对基础设施造成了破坏,进而影响该地区的交通、经济和社会稳定<sup>[19]</sup>。



(a) 通江县地理位置及历史滑坡分布



(b) 典型滑坡遥感图



(c) 现场调查图



图 1 研究区概况及典型滑坡现场调查与感影像

Fig. 1 Overview of the study area and on-site investigation and imagery of typical landslides

## 1.2 研究方法与滑坡数据处理

### 1.2.1 研究方法

1) 数据准备。获取 2010—2022 年历史滑坡数据和对应的逐日降雨量数据, 以及其他滑坡本底因子。

2) 深度嵌入聚类分析 (deep embedding clustering, DEC) 方法。利用深度学习技术, 将地形、土壤、植被等空间数据嵌入低维空间, 在低维空间中对数据进行聚类, 识别相似地理单元, 捕捉复杂的非线性关系和空间异质性。

3) 基于混合分布模型的动态雨量阈值模型考虑不同类型降雨事件对滑坡易发性的影响, 提高预测精准性。

a. 分别对极端降雨部分和非极端降雨部分建模, 使用 Gamma 分布对降雨量不超过阈值  $u$  的数据进行建模, 使用广义帕累托分布 (GPD) 对超过阈值  $u$  的数据进行建模。

b. 使用贝叶斯参数估计结合先验知识和观测数据, 计算模型参数的后验分布。

c. 滑坡发生的时间概率计算。整合非极端降雨和极端降雨部分的预测结果, 计算在降雨量超过阈值部分时, 滑坡发生的时间概率。也就是计算当降雨量超过设定的阈值时, 模型预测滑坡发生的概率值。

4) 通过幂律函数拟合的有效降雨量衰减系数得到前期有效降雨量。

5) 集成多任务学习模型。

a. 计算其他静态滑坡因子的滑坡空间易发性值, 得到滑坡空间概率并与分区动态降雨量阈值模型得到的滑坡时间概率相乘, 即可以得到滑坡预测易发性值。

b. 将前期降雨量衰减系数与 MLANT 模型结合, 得到动态前期有效降雨量, 并与其他静态因子一起参与参数计算。

6) 模型验证。使用历史数据验证模型, 通过 AUC 和 F1 分数评估性能, 对比两种处理降雨量方法的模型。

7) 预测滑坡易发性。按照相同策略构建滑坡易发

性预测数据集并使用训练验证后的模型进行预测分析。

综上所述,本文滑坡易发性预测的技术路线如图2所示。

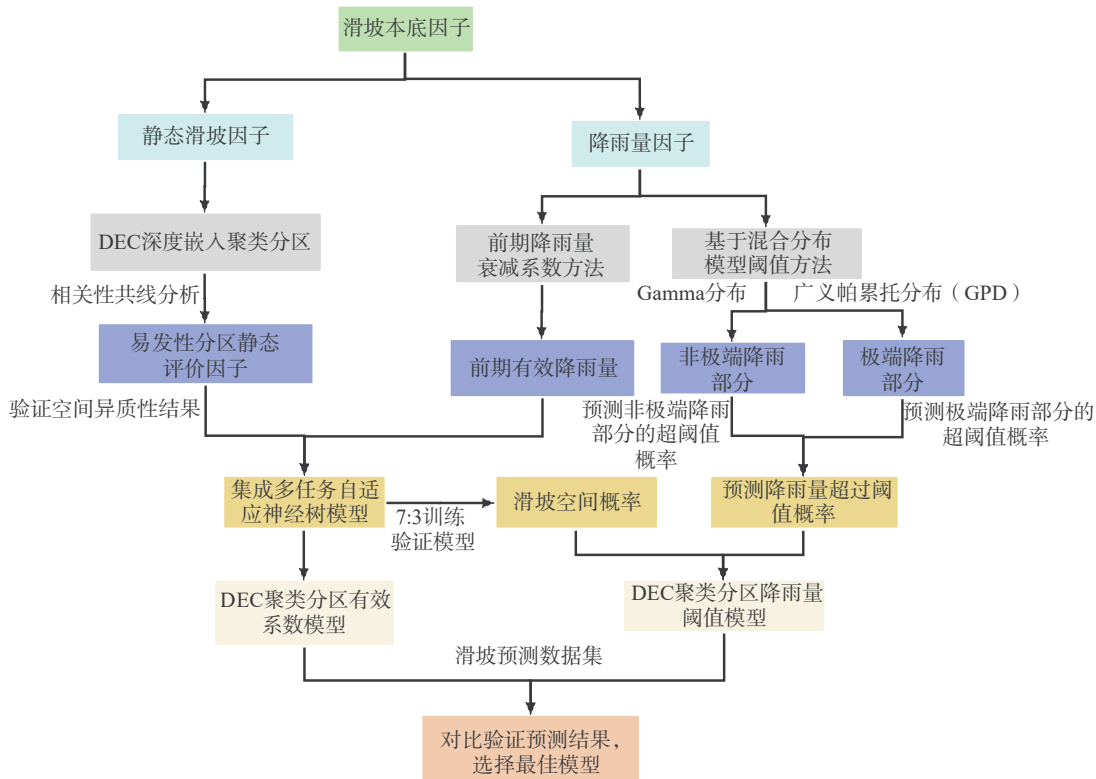


图2 滑坡易发性预测技术路线

Fig. 2 Landslide susceptibility prediction technology road

1.2.2 数据来源和处理方法

历史滑坡数据来源于巴中市自然资源局,覆盖2010至2022年间的滑坡灾害记录,提供了历史滑坡事件的详细资料。通过91卫图平台下载研究区1:200 000地质图,并进行矢量化处理以获取地层岩性信息。从地理空间数据云 (<https://www.gscloud.cn/>) 下载ASTERGDEM 30 m分辨率数字高程模型(DEM),用于提取高程、坡度、坡向和地形起伏度等信息。土地利用数据来自GlobeLand30项目 (<http://www.globallandcover.com/>),分辨率为30 m,提供研究区的土地利用类型信息。逐日降雨量数据(2000至2022年),来源于中国科学院资源环境科学与数据中心 (<https://www.resdc.cn/>)。LandsatTM8遥感影像用于提取归一化植被指数(NDVI),增强对植被覆盖变化的分析能力。高分一号遥感数据来源于星图地球今日影像 (<https://daily.geovisearth.com/dataBuy>),用于滑坡点的验证分析。共收集到历史滑坡点592处,主要以中小型规模的浅层土质滑坡为主。

1.3 滑坡因子分析

通过调查和参考文献,考虑研究区地形地貌、地质、水文环境、降雨4个方面,初步选取15个影响因子(高程、坡度、坡向、平面曲率、剖面曲率、地形起伏度、

地表粗糙度、土地利用、地层岩性、土壤湿度、归一化植被指数(NDVI)、地形湿度指数(TWI)、距河流距离、距道路距离、日降雨量因子),建立滑坡易发性影响因子体系,如图3所示<sup>[20]</sup>。

通过计算滑坡密度(LC)、信息值(IV)和频率比(FR)等指标,可以评估不同因子对滑坡发生概率的影响<sup>[21]</sup>。滑坡密度(LC)衡量滑坡点的分布密度,信息值(IV)量化了特定因子对滑坡事件的贡献度,频率比(FR)用于比较不同因子分区内滑坡事件的发生频率。通过分析这些指标,能够识别出对滑坡发生具有显著影响的因子,如附录A表A1所示。分析结果显示,滑坡易发程度在某些地形和植被条件下显著增加。特定的环境因素,如较低的植被覆盖度、适中的坡度、靠近河流的距离和适度的地形起伏度,均与较高的滑坡风险相关<sup>[22]</sup>;反之,高植被覆盖度、极端坡度和较大的地形起伏度区域的滑坡风险较低。

1)高滑坡风险因素

当土壤湿度为346~358时,滑坡密度(LC)为0.278,信息值(IV)为0.089,频率比(FR)为1.543。土壤湿度较高的区域容易导致土壤饱和,从而增加滑坡发生的可能性。坡度为13°~18°时,滑坡密度(LC)为0.24,信息值(IV)为0.080,频率比(FR)为1.360,表明

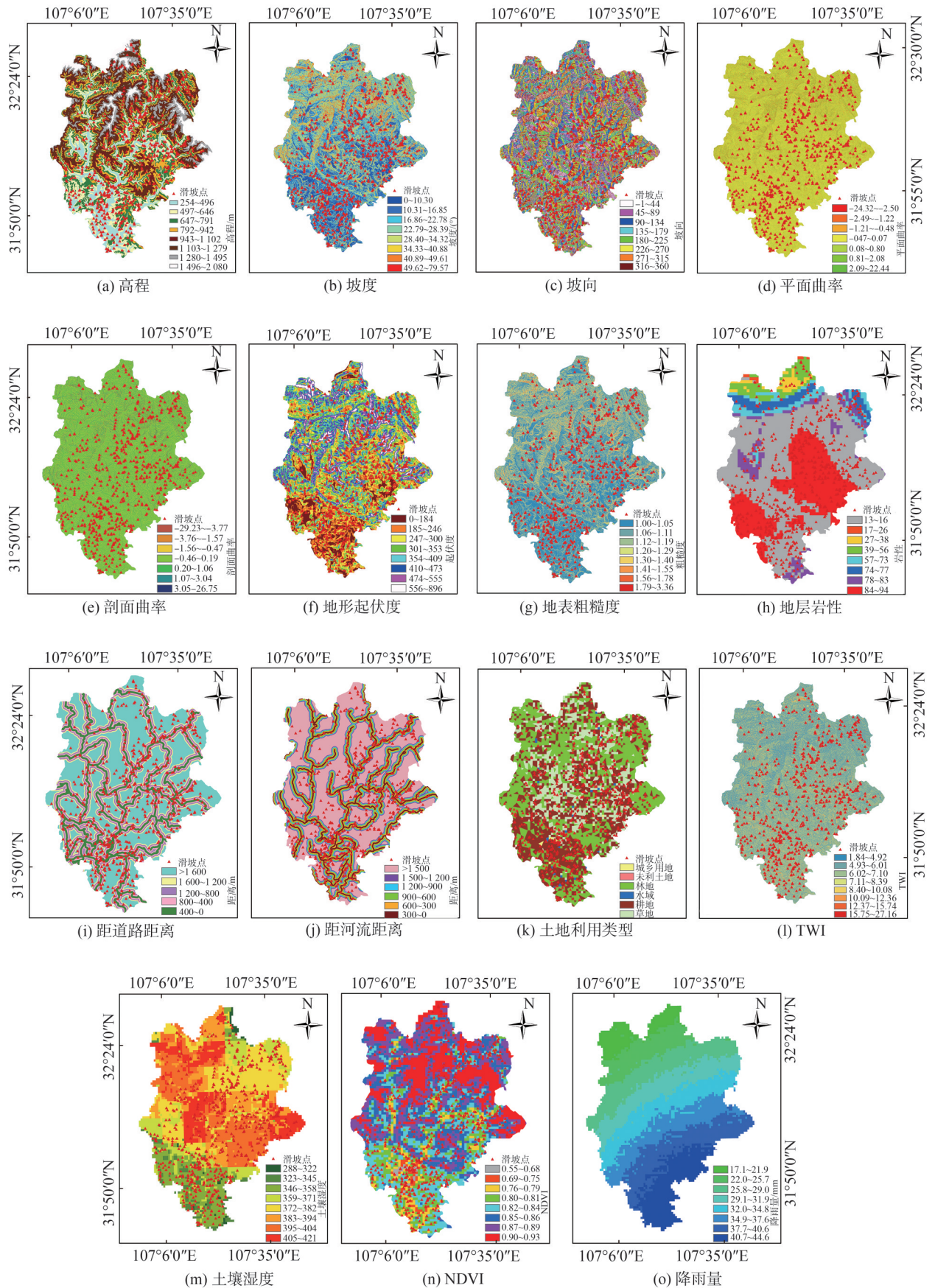


图3 滑坡易发性因子分级

Fig. 3 Landslide susceptibility factor classification

中等坡度区域的滑坡风险较高。这可能是因为这些区域水分容易积聚,导致土壤饱和,增加了滑坡的可能性。当距河流的距离为0~300 m时,是最接近河流的区域,其滑坡密度(LC)为0.29,信息值(IV)为0.090,频率比(FR)为1.630,说明河流的侵蚀作用和边坡的不稳定性导致滑坡风险较高。

### 2) 低滑坡风险因素

当剖面曲率为-20.15~3.69时,表现出较低的滑坡风险,滑坡密度(LC)为0.09,信息值(IV)为-0.002,频率比(FR)为0.499。较低的剖面曲率可能表明地形较为平坦,较少的地形变化有助于减少土壤运动。NDVI为0.9~0.93时,滑坡风险较低,LC为0.016,IV为-0.039,FR为0.091。高植被密度通常增强了地表稳定性,有助于固定土壤,防止侵蚀。地形起伏度为556~896 m时,滑坡风险较低,LC为0.072,IV为-0.010,FR为0.401。较大的地形起伏可能有助于水流排出,从而降低土壤饱和与滑坡风险。

### 3) 其他滑坡风险因素

在进行滑坡易发性评估时,由于直接获取大面积岩土体的力学参数存在困难,因此在滑坡易发性模型中可以考虑将岩性与力学参数的映射关系作为一个重要的输入变量<sup>[23]</sup>。通过统计分析不同岩性区域的滑坡密度(LC)、信息值(IV)和频率比(FR),建立岩性与滑坡风险之间的映射关系。例如,灰岩和白云岩的地层单位(如大冶组、雷口坡组和吴家坪组)通常具有较高的内聚力和内摩擦角,在干燥条件下表现出较高的稳定性。因此,这些岩性区域滑坡发生的风险相对较低。具体数据如附录A表A1所示,灰岩和白云岩的滑坡密度介于0.035至0.171之间,信息值介于-0.007至-0.001之间,频率比在0.197至0.949之间。其中,吴家坪组的频率比为0.949,表明其滑坡发生风险较低。相对而言,页岩、泥灰岩和泥岩的地层单位(如飞仙关组、白田坝组和千佛岩组)通常吸水性强,湿润后强度显著降低,容易在降雨后引发滑坡。数据显示,这些岩性的滑坡密度从0.074到0.780,信息值从-0.007到0.002,频率比从0.413到4.339不等。特别是白田坝组的频率比为4.339,显示其滑坡风险显著增高。通过上述分析,可以将岩性与力学参数的映射关系纳入滑坡易发性预测模型中,间接考虑力学参数对滑坡易发性的影响,以提高模型的准确性和可靠性。

地下水位对滑坡易发性评估至关重要,但由于直接获取大范围数据困难,通常采用地形湿度指数(TWI)和土壤湿度作为间接参数参考岩性的处理方法<sup>[24]</sup>。TWI值较高的区域往往地下水位高,滑坡风险增加。例如,当TWI值在7.11至8.39之间时,滑坡密度(LC)为0.249,频率比(FR)为1.386,显示较高的滑坡风险。土壤湿度高的区域也意味着较高滑坡风险,如当土壤湿度在346至358之间时,频率比达到1.543。通过历史滑坡事件验证和校准模型,可以确保在没有直接地下水位测量数据的情况下,通过间接测量TWI和土壤湿度等来有效预测滑坡风险。

## 1.4 滑坡因子选取

通过分析附录A表A1的数据,可以初步使用IV来识别可能对滑坡预测具有显著影响的因子。信息值作为一个统计指标,当其值高于0.02时,表明相关因子具有一定的预测能力。筛选出的因子包括土地利用、剖面曲率、坡向、坡度、平面曲率、高程、距河流距离、距道路距离、地形起伏度、地层岩性、地形湿度指数(TWI)以及土壤湿度。这些因子在各自的特定分类区间内表现出潜在的预测效果,因此被选为后续深入分析的候选因子。为进一步验证这些滑坡因子的选择是否会存在互相干扰的情况,并确定最佳滑坡因子组合,计算这些因子与滑坡事件之间的相关性,利用Pearson或Spearman相关系数评估它们之间的线性或非线性关系<sup>[25]</sup>。此外,通过计算方差膨胀因子(VIF)来检查这些因子之间的多重共线性<sup>[26]</sup>。通常地,VIF值大于10表明存在潜在的重大共线性问题。表1与图4分别为方差膨胀因子和候选滑坡因子相关性图。

根据Pearson相关性系数的标准,系数小于0.2表示因子之间几乎不相关,0.2到0.4表示低度相关,0.4到0.7表示中度相关,超过0.7则表示存在高度共线性。从图4的分析结果来看,除了距河流和距道路距离之外,其他各滑坡因子的相关性系数均低于0.4,表明它们之间的相关性不强。进一步地,由表1可知,大多数滑坡因子的VIF值都小于10,符合一般的统计要求,显示它们之间不存在严重的共线性问题。然而,距河流距离的VIF值接近10,并且考虑到其与道路距离显示出中度共线性,这可能影响模型的稳定性和解释力,基于这些分析,决定从滑坡易发性评价模型中剔除距河流距离这一因子。

表1 滑坡因子方差膨胀因子

Tab. 1 Landslide factors variance inflation factors

评价因子	土地利用	剖面曲率	坡向	坡度	平面曲率	高程	距河流距	距道路距	地形起伏	地层岩性	TWI	土壤湿度
VIF	5.09	1.66	4.12	6.93	1.67	6.28	9.74	5.17	5.33	3.08	5.63	3.91

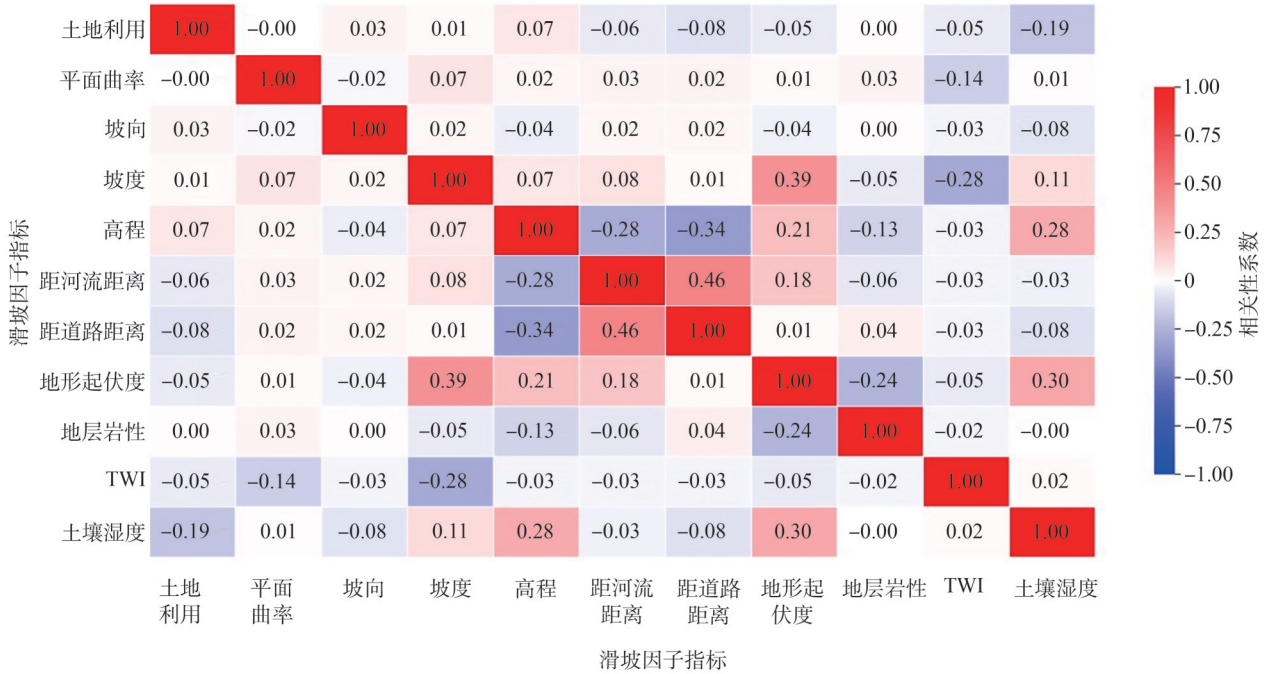


图4 滑坡因子相关性图

Fig. 4 Correlation diagram of landslide factors

## 2 考虑空间异质性的深度嵌入聚类分区与降雨阈值分析

为了有效地解决现有研究在处理空间异质性中的不足,提出了深度嵌入聚类分析(DEC)方法。DEC是一种结合深度学习和聚类分析的无监督学习方法<sup>[27]</sup>,通过自编码器预训练,学习数据的低维表示,在

该表示空间中执行聚类。通过结合深度嵌入聚类(DEC)技术,动态地根据地形、土壤、植被等环境因素将研究区域划分为具有相似地质特性的子区域,从而有效考虑滑坡的空间异质性,提高滑坡易发性预测精度。

### 2.1 考虑空间异质性的深度嵌入聚类分区

DEC算法步骤如图5所示<sup>[28]</sup>。

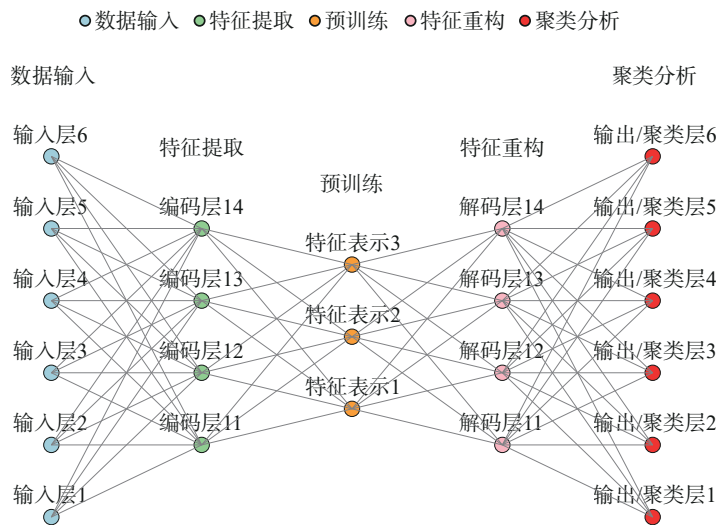


图5 DEC算法原理

Fig. 5 DEC algorithm principle

#### 2.1.1 自编码器的预训练

DEC算法首先通过自编码(autoencoder, AE)预训练,学习数据的低维表示。自编码器是一种特殊的神

经网络,目标是将输入 $x$ 编码成一个低维表示 $z$ ,然后再从 $z$ 重构回 $x$ 。自编码器由编码器 $f_\phi$ 和解码器 $g_\phi$ 组成,其训练目标是 minimized 重构误差,通常使用均方误

差(MSE)作为损失函数:

$$L_{AE}(\theta, \phi) = \frac{1}{N} \sum_{i=1}^N \left\| x_i - g_{\phi}(f_{\theta}(x_i)) \right\|_2^2 \quad (1)$$

式中,  $L_{AE}$  为自编码器重构损失函数,  $\theta$  和  $\phi$  分别为编码器和解码器的参数,  $N$  为样本数量,  $x_i$  为第  $i$  个原始输入样本。

### 2.1.2 聚类层初始化

在自编码器训练完成后, 利用编码器学到的特征  $z_i = f_{\theta}(x_i)$  来初始化聚类中心。通常通过在特征  $z$  上应用 k-means 算法完成。

### 2.1.3 联合优化

DEC算法的核心是通过最小化聚类分配和特定分布之间的 Kullback-Leibler (KL) 散度来联合优化聚类中心和网络参数<sup>[29]</sup>。首先, 计算样本点到聚类中心的软分配概率  $q_{ij}$ :

$$q_{ij} = \frac{\left(1 + \left\| z_i - \mu_j \right\|_2^2\right)^{-1}}{\sum_k \left(1 + \left\| z_i - \mu_k \right\|_2^2\right)^{-1}} \quad (2)$$

式中,  $\mu_j$  为聚类中心  $j$  的位置,  $z_i$  为样本  $i$  的低维表示。然后, 定义目标分布  $p_{ij}$ , 以增强高概率的分配:

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_k \left( q_{ik}^2 / \sum_i q_{ik} \right)} \quad (3)$$

式中,  $q_{ik}$  表示样本  $i$  属于任意中心  $k$  的软分配概率, 通过最小化  $q_{ij}$  和  $p_{ij}$  之间的 KL 散度来进行优化:

$$L_{DEC} = KL(\mathbf{P} \parallel \mathbf{Q}) = \sum_i \sum_j p_{ij} \lg \frac{p_{ij}}{q_{ij}} \quad (4)$$

式中,  $L_{DEC}$  为 DEC 联合优化损失,  $\mathbf{P}$  为目标分布矩阵,  $\mathbf{Q}$  为当前软分配分布矩阵。通过迭代优化过程, DEC 调整聚类中心的位置和低维特征表示, 以获得更好的聚类效果。

### 2.1.4 可视化聚类分区

应用深度嵌入聚类(DEC)算法对通江县滑坡相关因子进行聚类分区, 首先对包括高程、坡度、土地利用等在内的多个滑坡相关因子进行归一化处理, 以统一不同因子的量纲。随后, 利用自编码器网络进行预训练, 通过编码器部分学习了数据的压缩表示, 而解码器部分则负责重构原始输入, 以最小化输入和输出之间的均方误差。自编码器训练完成后, 将其编码输出用于初始化聚类中心, 进而通过最小化 Kullback-Leibler 散度的方式, 联合优化聚类层和特征表示, 结果如图 6 所示<sup>[30]</sup>。

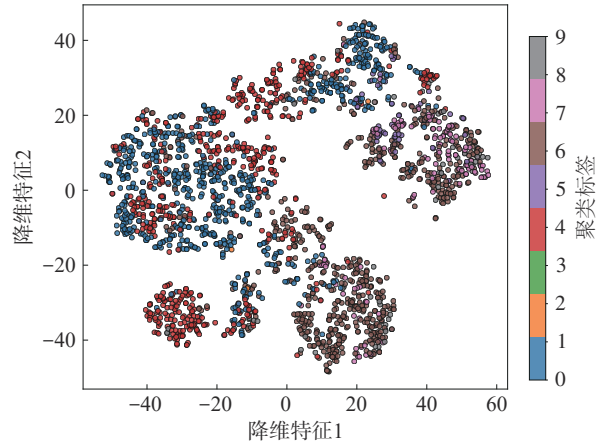


图6 通江县深度嵌入聚类滑坡因子分区

Fig. 6 Deep embedded clustering landslide factor zoning map of Tongjiang County

图6展示了基于DEC算法的聚类结果, 它将通江县的滑坡因子数据根据易发性分区进行了聚类。由图6可知, 不同的颜色代表了不同的聚类分区, 每个分区内的点表现出了相似的特征。通江县聚类分区图如图7所示。

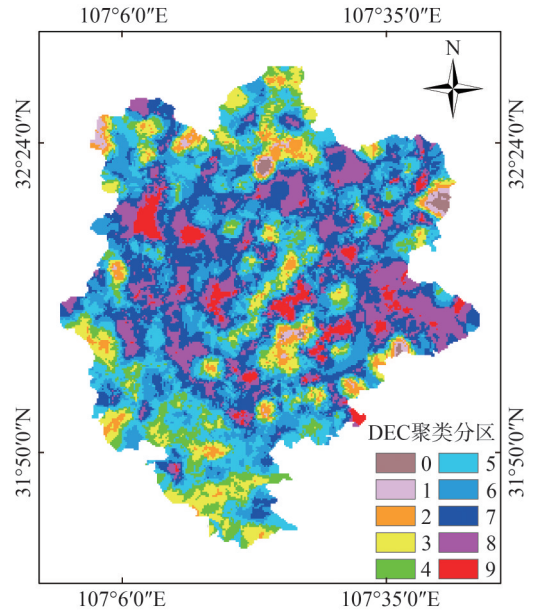


图7 通江县聚类分区图

Fig. 7 Cluster zoning map of Tongjiang County

通过结合深度嵌入聚类(DEC)技术, 动态地根据地形、土壤、植被等环境因素将研究区域划分为具有相似地质特性的子区域, 能够有效考虑滑坡的空间异质性。

## 2.2 考虑空间异质性的降雨阈值分析

为有效解决现有研究在处理降雨量因子中的不足, 提出基于混合分布模型的动态雨量阈值模型。使用该模型分别对非极端降雨(使用 Gamma 分布)和极端降雨(使用广义帕累托 GPD 分布)进行建模, 并使用

贝叶斯方法更新参数。通过这种方法,能够更准确地反映不同类型降雨对滑坡的影响,并动态适应不同环境和气候条件,从而提高预测的精确度和及时性,并使用前期有效降雨量系数法对比验证。

### 2.2.1 基于混合分布模型的动态阈值法

1)在模型中分别对非极端降雨部分和极端降雨部分进行建模,并使用贝叶斯方法更新参数。对于非极端降雨部分降雨量不超过某个阈值 $u$ 的数据,选择Gamma分布进行建模。对于极端降雨部分超过阈值 $u$ 的数据,使用广义帕累托分布(GPD)进行建模。图8所示为通江县DEC聚类分区历史滑坡对应降雨量阈值和整体降雨量阈值分布CFD图,采用95%分位作为阈值 $u^{[31]}$ 。

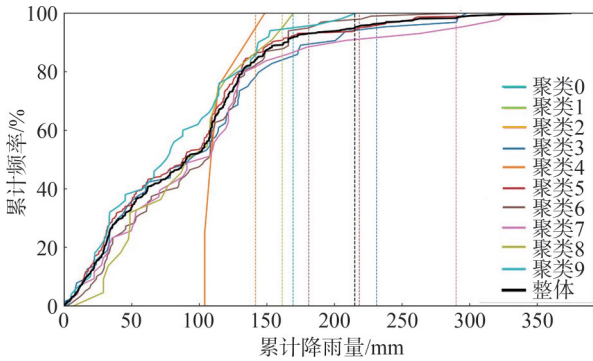


图8 聚类分区与整体降雨量阈值CFD图

Fig. 8 Cluster partition and overall rainfall threshold CFD diagram

2)贝叶斯参数估计。结合先验知识和观测数据,计算模型参数的后验分布。假设非极端降雨部分的参数为 $\theta_{\text{non-extreme}} = (\alpha_{\text{non-extreme}}, \beta_{\text{non-extreme}})$ ,其先验分布为 $p(\alpha_{\text{non-extreme}}, \beta_{\text{non-extreme}})$ ;极端降雨部分的形状参数为 $\zeta$ ,尺度参数为 $\sigma$ ,其先验分布分别为 $p(\zeta)$ 和 $p(\sigma)$ 。

非极端降雨部分的似然函数:

$$L(\theta_{\text{non-extreme}}|\text{data}) = \prod_{i: Y_i \leq u} \text{Gamma}(Y_i | \alpha_{\text{non-extreme}}, \beta_{\text{non-extreme}}) \quad (5)$$

式中, $Y_i$ 表示第 $i$ 条观测降雨量, $u$ 为降雨量阈值, $\theta_{\text{non-extreme}}$ 为非极端降雨部分的Gamma分布参数向量, $\text{Gamma}(\cdot)$ 为Gamma分布的概率密度函数, $L(\theta_{\text{non-extreme}}|\text{data})$ 为非极端段似然函数。

极端降雨部分的似然函数:

$$L(\zeta, \sigma|\text{data}) = \text{GPD}(Y_i - u | \zeta, \sigma) \quad (6)$$

式中, $L(\zeta, \sigma|\text{data})$ 为极端段似然函数,GPD为概率密度函数。

Gamma的概率密度函数为:

$$\text{Gamma}(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad (7)$$

式中: $\alpha$ 为形状参数; $\beta$ 为尺度参数; $\Gamma(\alpha)$ 代表Gamma函数,用于对Gamma分布的概率密度函数进行归一化处理; $x = Y_i - u$ 为极端降雨量部分。

GPD的概率密度函数为:

$$\text{GPD}(x|\zeta, \sigma, \mu) = \frac{1}{\sigma} \left(1 + \zeta \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\zeta} - 1} \quad (8)$$

式中, $\mu$ 为位置参数。根据贝叶斯公式,计算参数的后验分布:

$$p(\theta_{\text{non-extreme}}|\text{data}) \propto L(\theta_{\text{non-extreme}}|\text{data}) \cdot p(\theta_{\text{non-extreme}}) \quad (9)$$

$$p(\zeta, \sigma|\text{data}) \propto L(\zeta, \sigma|\text{data}) \cdot p(\zeta) \cdot p(\sigma) \quad (10)$$

3)预测与超阈值概率计算。使用从后验分布中抽取的参数样本,模拟未来的降雨量数据。对非极端降雨部分的数据进行Gamma分布拟合,计算非极端降雨数据的形状参数 $\hat{\alpha}$ 和尺度参数 $\hat{\beta}$ 。

$$\hat{\alpha} = \frac{\left(\frac{1}{n} \sum_{q=1}^n Y_q\right)^2}{\frac{1}{n-1} \sum_{q=1}^n (Y_q - \hat{\mu})^2} \quad (11)$$

$$\hat{\beta} = \frac{\frac{1}{n-1} \sum_{q=1}^n (Y_q - \hat{\mu})^2}{\frac{1}{n} \sum_{q=1}^n Y_q} \quad (12)$$

式中, $Y_q$ 为非极端降雨数据点, $n$ 为非极端降雨数据的数量, $\hat{\mu} = \frac{1}{n} \sum_{q=1}^n Y_q$ 为样本均值。使用先验分布和似然函数计算后验分布的参数,进行贝叶斯更新。先验分布为Gamma $(\alpha_0, \beta_0)$ ,观测数据的似然函数为Gamma $(\hat{\alpha}, \hat{\beta})$ ,则后验分布的形状参数和尺度参数为:

$$\alpha_{\text{posterior}} = \alpha_0 + n \cdot \hat{\alpha} \quad (13)$$

$$\beta_{\text{posterior}} = \beta_0 + n \cdot \hat{\beta} \quad (14)$$

式中, $\alpha_{\text{posterior}}$ 为更新后的后验参数, $\beta_{\text{posterior}}$ 为更新后的后验尺度。

图9为前期通过聚类得到的聚类分区6(图7)的20 a累计10 d降雨量数据,使用前述后验形状参数和后验尺度参数生成极端和非极端降雨数据。

4)计算超阈值概率。根据模拟的降雨量数据,计算其超过滑坡发生阈值 $u$ 的概率。使用Gamma分布的累积分布函数计算降雨量超过阈值 $u$ 的概率:

$$P(\text{降雨量} > u | \text{非极端降雨量}) = 1 - F\left(u | \alpha_{\text{posterior}}, \beta_{\text{posterior}}\right) \quad (15)$$

式中, $F$ 为Gamma分布的累积分布函数,使用广义帕累托分布的累积分布函数GPD计算降雨量超过阈值 $u$

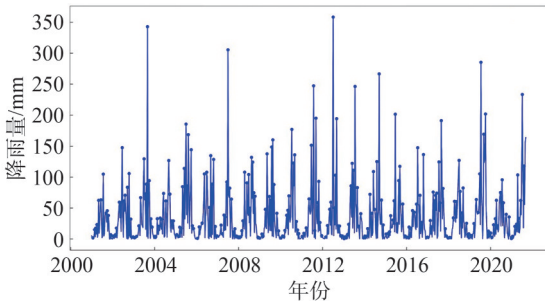


图9 通江县聚类分区6的20 a累计10 d降雨量数据  
Fig. 9 Tongjiang County Cluster 6, 20 a cumulative 10-day rainfall data

的概率:

$$P(\text{降雨量} > u | \text{极端降雨量}) = \left( 1 + \frac{\hat{\xi}(u - \hat{\mu})}{\hat{\sigma}} \right)^{-\frac{1}{\hat{\xi}}} \quad (16)$$

式中,  $\hat{\mu}$  为观测阶段的非极端样本均值,  $\hat{\xi}$ 、 $\hat{\sigma}$  分别为 GPD 后验形状与尺度参数。

表2 通江县2021年9月5日混合分布模型拟合结果

Tab. 2 Fitting results of the mixed distribution model in Tongjiang County on September 5, 2021

阈值/mm	非极端部分均值 $m_{\text{Gamma}}$	非极端部分 标准差 $\sigma_{\text{Gamma}}$	GPD 形状参数 $\xi$	GPD 位置参数 $\mu$	GPD 尺度 参数 $\sigma_{\text{GPD}}$	超阈值概率	K-S 统计量
215.02	82.24	50.79	-0.28	0.85	61.86	0.002 55	0.011
148.22	84.11	42.60	-1.81	-0.18	29.62	0.037 80	0.020
135.06	81.22	43.90	-1.78	0	5.22	0.090 90	0.027
228.28	85.32	55.96	-1.15	-0.23	95.47	0.003 10	0.022
141.57	107.63	4.14	-2.54	-0.04	17.59	0.000 00	0.027
218.32	78.13	50.45	-0.12	0.16	54.01	0.002 85	0.019
180.90	87.68	48.27	0.17	6.05	30.23	0.023 50	0.020
290.00	89.30	57.97	-2.12	-0.35	81.82	0.000 20	0.023
161.26	80.96	40.84	-3.71	-0.09	31.20	0.018 15	0.019
169.41	71.60	46.24	-2.43	0.19	112.54	0.009 70	0.013

### 2.2.2 有效降雨量衰减系数方法

考虑到不同时间段内的降雨对滑坡触发潜力的贡献程度不同,本文采用幂律函数定量描述累积降雨量与观测时段之间的动态关系。通过对历史降雨量数据和滑坡事件进行幂律拟合,可以计算得到有效降雨量衰减系数,以此反映降雨量在不同天数的影响权重。

累积降雨量  $R_{\text{cumulative}}$  的计算是基于前期每天降雨量的加权求和。权重,也即衰减系数的计算公式如下<sup>[33]</sup>:

$$\gamma = \sum_{t=1}^m \left( 1 - \left( \frac{t}{t+1} \right)^\delta \right) \quad (18)$$

式中,  $\gamma$  为累积的降雨量衰减值,  $t$  为时间或离散时刻的索引,  $\delta$  为衰减系数。 $\delta$  用来控制降雨量对滑坡影响随

将非极端降雨部分和极端降雨部分的超阈值概率相加,得到总的超阈值概率:

$$P(\text{降雨量} > u) = P(\text{降雨量} > u | \text{非极端降雨量}) + P(\text{降雨量} > u | \text{极端降雨量}) \quad (17)$$

采用最近 10 d 的降雨数据作为分析窗口,随数据更新持续滚动,确保即时性。根据最新气候调整极端降雨阈值  $u$ , 以适应气候与季节变化,使模型能预测未来某天的滑坡易发性。

表2为通江县2021年9月5日的拟合分布结果。为验证 Gamma 分布和广义帕累托分布(GPD)对历史降雨数据的拟合程度,采用 K-S (Kolmogorov-Smirnov) 统计量<sup>[32]</sup>。K-S 统计量是一种非参数检验方法,通常, K-S 值小于 0.05 被认为数据与分布拟合良好。表2结果显示,所有的值都明显小于 0.05 的常用阈值,因此使用 Gamma 分布和广义帕累托分布(GPD)对非极端和极端降雨数据进行拟合的效果良好。

时间衰减的速度,具体来说,  $\delta$  越大,降雨对滑坡的影响衰减得越快;  $\delta$  越小,降雨的影响持续得更久。累积降雨  $R_{\text{threshold}}$  是通过分形曲线确定的,具体如式(19)所示:

$$R_t = C(t+1)^\delta \quad (19)$$

式中,  $R_t$  为在第  $t$  时刻的降雨量值(通过分形曲线计算的),系数  $C$  为常数,其与衰减系数  $\delta$ , 可通过拟合实际滑坡事件与降雨数据得到,具体结果如图 10 所示。确定完研究区分区后对分区阈值和整体阈值进行拟合<sup>[34]</sup>。

图 11 为通过拟合通江县系数  $C$  和衰减系数  $\delta$  得到的通江县 DEC 聚类分区的整体降雨量衰减系数变化。将得到的降雨量衰减系数与降雨量相乘即为滑坡有效降雨量。

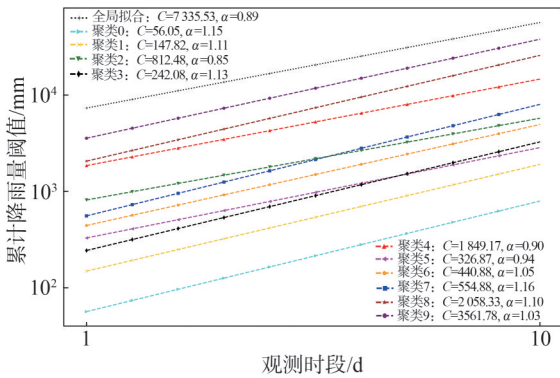


图 10 通江县各聚类分区与整体的分形曲线降雨量拟合  
Fig. 10 Fractal curve fitting of rainfall for different cluster partitions and overall model in Tongjiang County

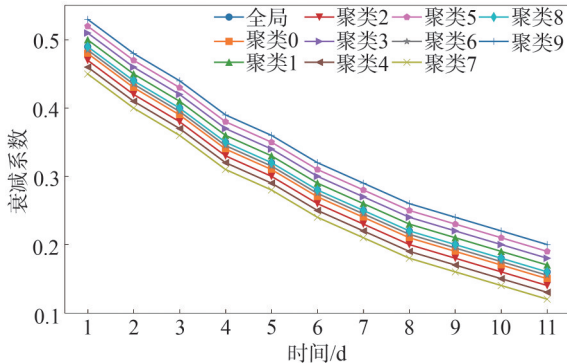


图 11 通江县各聚类分区与整体的降雨量衰减系数变化  
Fig. 11 Variation of rainfall attenuation coefficients for different cluster partitions and overall model in Tongjiang County

### 3 滑坡易发性建模方法及精度分析

#### 3.1 多任务学习自适应神经树(MLANT)模型

在探究复杂数据结构的预测模型中,多任务学习自适应神经树(MLANT)模型展现了独特的优势。MLANT模型是一种机器学习框架,旨在将决策树的解释性与神经网络的表示能力相结合<sup>[35]</sup>。在MLANT模型中,决策节点被构建为包含神经网络层的数据处理单元。每个决策节点的操作可以用以下公式描述<sup>[36]</sup>:

$$f_{\text{decision}}(x; \Theta_d) = \psi(W_d \cdot x + b_d) \quad (20)$$

式中, $x$ 为输入特征, $\Theta_d$ 包括权重 $W_d$ 和偏置 $b_d$ , $\psi$ 为激活函数。对于叶节点,其功能是终结数据的传播并输出预测结果,可表示为:

$$f_{\text{leaf}}(x; \Theta_l) = \omega(W_l \cdot x + b_l) \quad (21)$$

式中, $\Theta_l$ 为叶节点的参数集, $W_l$ 和 $b_l$ 为权重和偏置。

模型的自适应机制涉及以下关键步骤:1)节点分裂。基于性能指标,如减少误差,一个节点可以分裂成两个子节点以期提高模型的表达能力。2)节点剪枝。如果某节点的移除不会对模型的预测准确性产生负面影响

响,则该节点可以被剪枝,从而简化模型结构。3)节点合并。相似功能的节点可以合并为一个,以进一步优化模型的效率和性能。该自适应机制确保了模型能够在提供精确预测的同时,保持其结构的最优化。

#### 3.2 MLANT模型结合降雨量阈值与衰减系数

##### 3.2.1 混合分布模型的动态阈值与MLANT模型集成

结合多任务学习自适应神经树模型(MLANT)和混合分布模型,首先,利用地形、土壤、植被等静态因子计算空间性的滑坡易发性概率。然后,将此空间概率与基于混合分布模型的动态阈值模型得到的时间概率相乘,得到最终的滑坡易发性预测概率。MLANT模型设定的动态分析窗口随每日更新的降雨数据持续滚动,确保了预测的实时性。此外,模型根据最新的气候数据动态调整极端降雨阈值 $u$ ,增强了对环境变化的适应性。

##### 3.2.2 有效降雨量与MLANT模型集成

在集成有效降雨量与多任务学习自适应神经树模型(MLANT)的策略中,关键步骤包括数据的预处理和动态特征融合。所有相关的地质和气候因子,包括有效降雨量,在输入模型之前均进行归一化和缩放处理,这一过程统一了不同数据源的尺度和分布,减少了原始数据差异对模型输出的影响。在特征融合阶段,将有效降雨量等动态变量与土壤类型、植被覆盖率等静态环境因子结合,并通过主成分分析(PCA)或其他特征选择方法,提取对模型预测最有影响力的特征组合。同时,时间序列分析被用于挖掘降雨数据中的时间模式,提升模型对时序变化的响应能力。MLANT模型通过其自适应决策树架构训练这些集成的特征,每个决策节点根据输入的特征动态调整决策边界和输出,允许模型实时且精确地评估滑坡风险。

#### 3.3 滑坡易发性模型精度检验

样本数据集被分为训练集和测试集,比例为7:3,采用深度嵌入聚类(DEC)技术对研究区域进行分区,明确区分了地形与气候特征的独特子集。通过将聚类分区确定的特征与多任务学习自适应神经树(MLANT)模型相结合,显著提升了模型在各评估指标上的表现<sup>[37]</sup>,具体如图12和表3所示。

各模型的性能评估结果显示:引入深度嵌入聚类(DEC)技术,明显改善了传统滑坡易发性预测方法的局限;DEC聚类分区方法通过识别和应对研究区域内的空间异质性,显著提高了滑坡预测的精度和适应性。对比分区方法与整体模型的表现,数据显示,DEC聚类分区模型在关键性能指标上优于整体模型,尤其是受试者工作特征曲线下面积AUC值的提升,从整体模型的0.86提升至分区模型的0.94,这表明分区模型

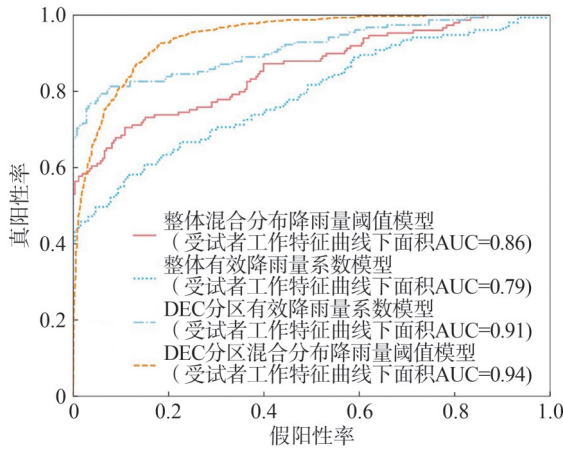


图12 各模型的ROC精度曲线与性能比较

Fig. 12 ROC accuracy curves of each model and performance comparison

表3 各模型性能对比

Tab. 3 Comparison of the performance of each model

模型	精确度	召回率	F1得分	AUC值
DEC分区混合分布降雨量阈值	0.77	0.90	0.82	0.94
DEC分区有效降雨量系数	0.88	0.73	0.8	0.91
整体混合分布降雨量阈值	0.82	0.63	0.74	0.86
整体有效降雨量系数	0.71	0.56	0.62	0.79

在区分滑坡与非滑坡事件方面具有更高的准确性和可靠性。进一步地,在DEC聚类分区的基础上,提出

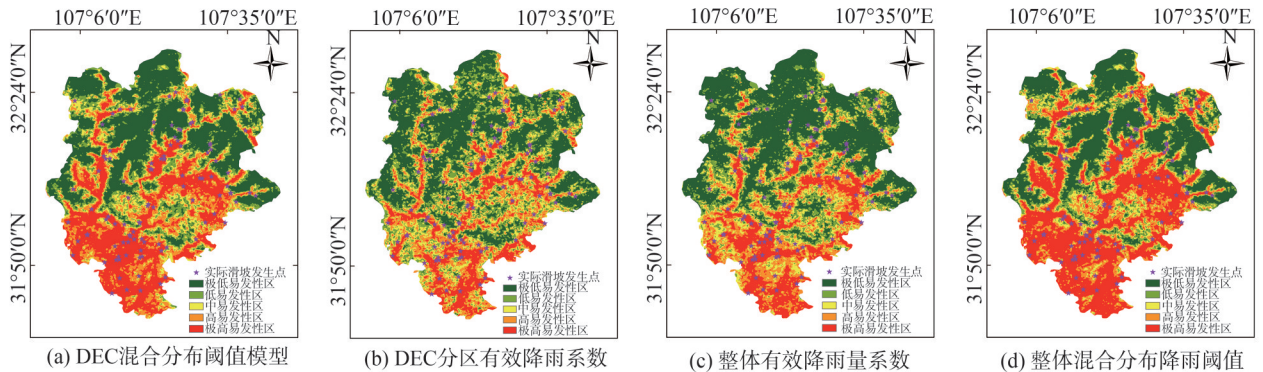


图13 DEC聚类分区与整体区域降雨易发性模型结果对比

Fig. 13 Comparison of the results of the DEC clustering and overall regional rainfall susceptibility models

2021年9月5日,通江县发生了93例滑坡事件,采用滑坡密度和滑坡数量评价滑坡易发性预测结果<sup>[38]</sup>。通过对聚类模型与整体模型进行系统比较,发现聚类模型在高风险区域的表现明显优于整体模型,如图14所示,DEC聚类阈值分区模型在高易发性区与极高易发性区的滑坡密度分别达到了0.036 92和0.046 92事件/km<sup>2</sup>,滑坡事件数量分别为24和35,均高于其他模型。这一结果强调了聚类阈值分区模型在识别和预测高风险滑坡事件方面的优势。

相对而言,DEC聚类有效降雨系数模型虽然在同

基于混合分布的动态降雨量阈值模型,以更精确地区分非极端与极端降雨事件。此模型采用贝叶斯方法动态更新模型参数,增强了模型对环境变化的响应能力。性能比较中,对比混合降雨量阈值模型与有效降雨量系数模型可知,尽管有效降雨量系数模型在精确度上略高,但混合降雨量阈值模型在召回率和AUC值上表现更优,反映了该模型区分滑坡与非滑坡事件的能力更强。综上,采用基于混合分布的动态降雨量阈值模型能有效提高滑坡易发性预测的准确性和实时性,尤其是在处理环境条件变化和极端降雨事件时。故DEC聚类分区结合动态降雨量阈值模型提供了一种有效的滑坡预测策略。

### 4 滑坡易发性动态预测与结果验证

为验证各预测模型的有效性,采用4种不同的模型,预测2021年9月5日四川省通江县滑坡易发性。以2021年9月4日及以前处理的滑坡因子数据和9月5日的降雨量数据作为输入,构建深度嵌入聚类(DEC)分区混合分布降雨阈值模型(图13(a))、多任务学习自适应神经树(MLANT)模型集成的DEC分区模型(有效系数版本,图13(b))、整体有效系数模型(图13(c))和整体混合分布降雨阈值模型(图13(d)),以探究包括地形、土壤类型、植被覆盖率和近期降雨量在内的环境因素如何影响滑坡易发性预测的准确性与实时性。

样的风险区域内表现稍逊,其滑坡密度分别为0.035 33和0.044 59事件/km<sup>2</sup>,成功预测的滑坡事件数量为23和21,但仍优于整体模型。此外,整体模型在低风险区域表现出了较高的误报率。特别是整体有效降雨系数模型在低易发性区的滑坡事件数量显著,达到21事件,其滑坡密度为0.02 937事件/km<sup>2</sup>,这可能导致资源的非有效分配和不必要的预警。这一发现表明,虽然整体模型在处理大范围数据时具备一定的便利性,但在精确性和误报率控制方面还有待优化。

图15为2021年9月5日发生在四川巴中市通江

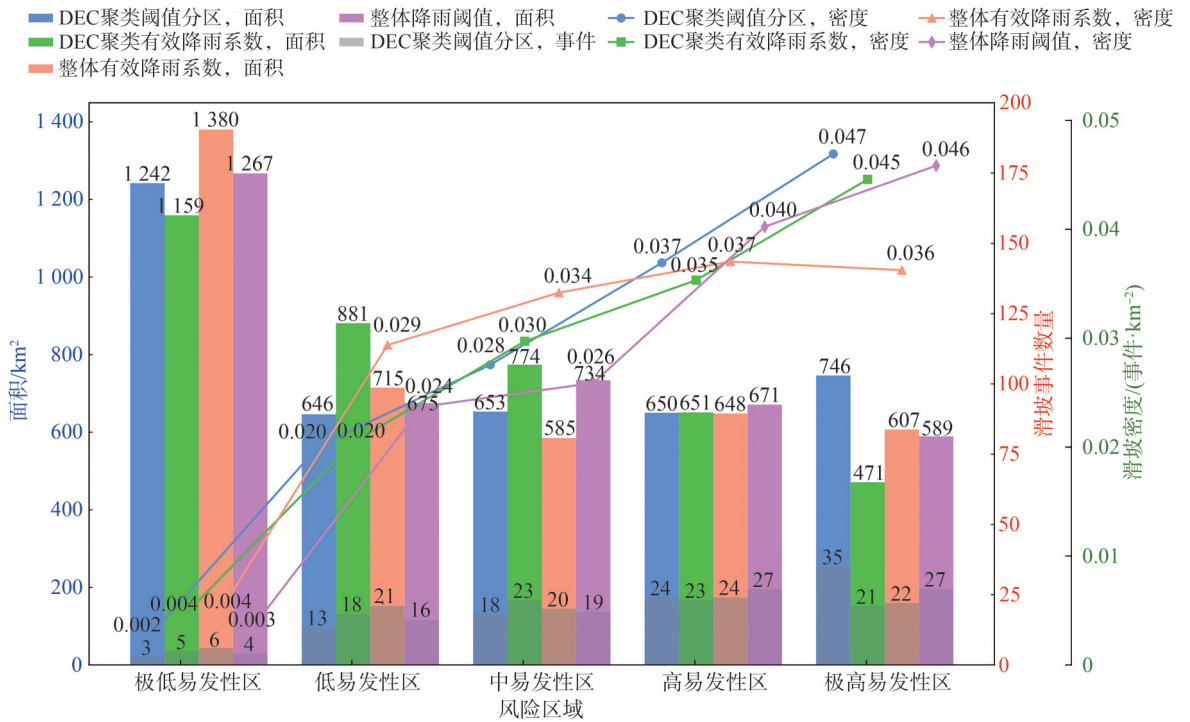


图 14 通江县 9 月 5 日各模型预测的滑坡易发性分区面积与滑坡数量统计

Fig. 14 Area and count statistics for each model's predicted landslide susceptibility zones in Tongjiang County on 5th, Sept.

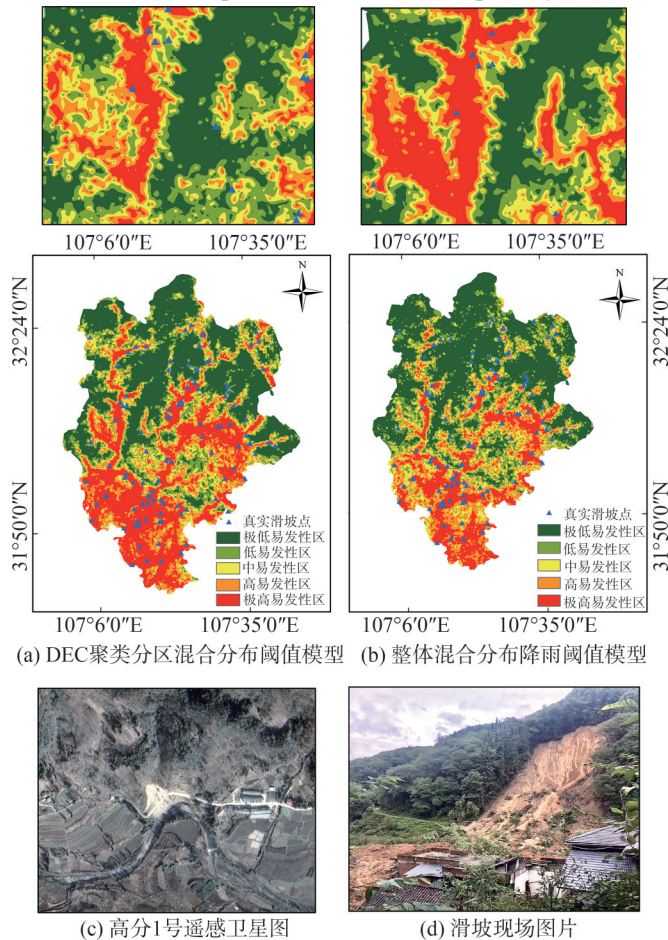


图 15 遥感影像与现场调查验证DEC聚类分区和整体阈值图

Fig. 15 Verification of DEC cluster partition and overall threshold map using remote sensing images and field surveys

县空山镇五福村二社柒家坝的滑坡事件。对比DEC聚类分区阈值模型与整体阈值模型的预测结果发现,DEC聚类分区阈值模型与整体阈值模型均成功预测了事件发生,但在风险区域预测上展现出显著差异,DEC聚类分区阈值模型通过聚类分析精确地界定了高风险区域,使预测更集中于真正的滑坡潜在地点。整体阈值模型预测了过宽的高风险区域,可能导致资源浪费和不必要的社区紧张。

## 5 结 论

本文通过整合深度嵌入聚类(DEC)、多任务学习自适应神经树(MLANT)模型与混合分布雨量阈值方法,系统提升了滑坡空间异质性与极端降雨下的滑坡易发性预测能力,主要结论如下。

1) DEC聚类分区模型显著提升了预测性能。在高易发性区和极高易发性区,DEC模型的滑坡密度分别达到0.036 92和0.046 92事件/km<sup>2</sup>,成功预测59个滑坡事件。与未考虑空间异质性的整体模型相比,成功预测滑坡事件总数分别提升28.3%(相较于有效降雨系数模型)和9.3%(相较于整体降雨阈值模型)。这说明考虑空间异质性有助于提升滑坡风险识别能力。

2) 混合分布动态阈值模型增强极端降雨响应。基于DEC分区,混合分布动态阈值模型在高易发性和极高易发性区的滑坡预测密度分别提升4.5%和5.2%,成功预测滑坡事件数量提升34.1%。与仅使用聚类有效降雨量系数的模型相比,该方法在极端降雨情景下的预测精度和及时性均明显提高,尤其对高风险区域极端事件的响应更加敏感。

综上,本文提出的分区+混合分布模型有效提升了滑坡预测的准确率和实用性。

附录见本刊网络版,扫描标题旁的二维码可阅读网络全文。

### 参考文献:

- [1] Whiteley J S, Chambers J E, Uhlemann S, et al. Geophysical monitoring of moisture-induced landslides: A review[J]. *Reviews of Geophysics*, 2019, 57(1): 106–145.
- [2] Basu T, Pal S. A GIS-based factor clustering and landslide susceptibility analysis using AHP for Gish River Basin, India[J]. *Environment, Development and Sustainability*, 2020, 22(5): 4787–4819.
- [3] Aditian A, Kubota T, Shinohara Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia[J]. *Geomorphology*, 2018, 318: 101–111.
- [4] Chen Wenwu, Zhang Shuai. GIS-based comparative study of Bayes network, Hoeffding tree and logistic model tree for landslide susceptibility modeling[J]. *CATENA*, 2021, 203: 105344.
- [5] Huang Yu, Zhao Lu. Review on landslide susceptibility mapping using support vector machines[J]. *CATENA*, 2018, 165: 520–529.
- [6] Merghadi A, Abderrahmane B, Tien Bui D. Landslide susceptibility assessment at Mila Basin (Algeria): A comparative assessment of prediction capability of advanced machine learning methods[J]. *ISPRS International Journal of Geo-Information*, 2018, 7(7): 268.
- [7] Youssef A M, Pourghasemi H R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia[J]. *Geoscience Frontiers*, 2021, 12(2): 639–655.
- [8] Hong Haoyuan, Pradhan B, Sameen M I, et al. Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China) [J]. *Geomatics, Natural Hazards and Risk*, 2017, 8(2): 1997–2022.
- [9] Fang Zhice, Wang Yi, Duan Gonghao, et al. Landslide susceptibility mapping using rotation forest ensemble technique with different decision trees in the Three Gorges Reservoir area, China[J]. *Remote Sensing*, 2021, 13(2): 238.
- [10] Yang Yang, Yang Jintao, Xu Chengdong, et al. Local-scale landslide susceptibility mapping using the B-GeoSVC model[J]. *Landslides*, 2019, 16(7): 1301–1312.
- [11] Deng Min, Yang Wentao, Liu Qiliang, et al. Heterogeneous space-time artificial neural networks for space-time series prediction[J]. *Transactions in GIS*, 2018, 22(1): 183–201.
- [12] Wang Qian, Wang Yi, Niu Ruiqing, et al. Integration of information theory, K-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges area, China[J]. *Remote Sensing*, 2017, 9(9): 938.
- [13] Godt J W, Baum R L, Savage W Z, et al. Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework[J]. *Engineering Geology*, 2008, 102(3/4): 214–226.
- [14] Nedumpallile Vasu N, Lee S R, Pradhan A M S, et al. A new approach to temporal modelling for landslide hazard assessment using an extreme rainfall induced-landslide index[J]. *Engineering Geology*, 2016, 215: 36–49.
- [15] Guzzetti F, Galli M, Reichenbach P, et al. Landslide hazard assessment in the collazzone area, Umbria, central Italy[J]. *Natural Hazards and Earth System Sciences*, 2006, 6(1): 115–131.
- [16] Cascini L, Bonnard C, Corominas J, et al. Landslide hazard and risk zoning for urban planning and development

- [C]//Hung R, Fell R, Couture R, et al. Landslide risk management, 2005: 209–246.
- [17] Yi Jingsong, Cheng Yingjian, Xu Yi, et al. Meteorological warning threshold of regional landslide disaster based on process rainfall[J]. *Bulletin of Soil and Water Conservation*, 2024, 44(2): 1–8. [易靖松, 程英建, 徐乙, 等. 基于过程降雨的红层丘陵低山区滑坡灾害气象预警阈值研究[J]. *水土保持通报*, 2024, 44(2): 1–8.]
- [18] Glade T, Crozier M, Smith P. Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “antecedent daily rainfall model”[J]. *Pure and Applied Geophysics*, 2000, 157(6): 1059–1079.
- [19] Cai Haojie, Han Haihui, Zhang Yulian, et al. Convolutional neural network landslide recognition based on terrain feature fusion[J]. *Journal of Earth Sciences and Environmental Studies*, 2022, 44(3): 568–579. [蔡浩杰, 韩海辉, 张雨莲, 等. 基于地形特征融合的卷积神经网络滑坡识别[J]. *地球科学与环境学报*, 2022, 44(3): 568–579.]
- [20] Huang Faming, Cao Yu, Fan Xuanmei, et al. Influence of different landslide boundaries and their spatial shapes on the uncertainty of landslide susceptibility prediction[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2021, 40 (Supp2): 3227–3240. [黄发明, 曹昱, 范宣梅, 等. 不同滑坡边界及其空间形状对滑坡易发性预测不确定性的影响规律[J]. *岩石力学与工程学报*, 2021, 40(增刊 2): 3227–3240.]
- [21] Liu Haizhi, Xu Hui, Bao Hongjun, et al. Optimization experiment of early identification of landslides susceptibility areas in medium and small mountainous catchment based on ensemble learning[J]. *Engineering Science and Technology*, 2022, 54(6): 12–20. [刘海知, 徐辉, 包红军, 等. 基于集成学习的山区中小流域滑坡易发区早期识别优化试验[J]. *工程科学与技术*, 2022, 54(6): 12–20.]
- [22] Li Muiyang. Study on shallow landslides distribution and rainfall threshold considering vegetation factors[D]. Beijing: Beijing Forestry University, 2021. [李牧阳. 考虑植被因素的浅表层滑坡分布及雨量阈值研究[D]. 北京: 北京林业大学, 2021.]
- [23] Guo Zizheng, Yin Kunlong, Fu Sheng, et al. Evaluation of landslide susceptibility based on GIS and WOE-BP model[J]. *Earth Science*, 2019, 44(12): 4299–4312. [郭子正, 殷坤龙, 付圣, 等. 基于 GIS 与 WOE-BP 模型的滑坡易发性评价[J]. *地球科学*, 2019, 44(12): 4299–4312.]
- [24] Huang Faming, Zhang Yinlang, Guo Zizheng, et al. Influence of different classification methods on regional landslide susceptibility zoning[J/OL]. *China Industrial Economics*, 2023: 1–13. [2023–11–23]. <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=SCLH20231122002&dbname=CJFD&dbcode=CJFQ>.
- [25] Mu Ke, Xie Wanli, Liu Qiqi, et al. Research on landslide susceptibility evaluation based on logistic regression and LR coupling model[J]. *Disaster Science*, 2022, 37(3): 212–218. [穆柯, 谢婉丽, 刘琦琦, 等. 基于 LR-RF 模型的滑坡易发性评价——以铜川市耀州区为例[J]. *灾害学*, 2022, 37(3): 212–218.]
- [26] Chen Xinyu, Shi Yun, Wen Yongxiao, et al. Landslide susceptibility evaluation based on certainty factor and support vector machines[J]. *Science Technology and Engineering*, 2023, 23(2): 518–527. [陈芯宇, 师芸, 温永啸, 等. 基于确定性系数与支持向量机的滑坡易发性评价[J]. *科学技术与工程*, 2023, 23(2): 518–527.]
- [27] Li Kang, Ni Tongguang, Xue Jing, et al. Deep soft clustering: Simultaneous deep embedding and soft-partition clustering[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2023, 14(5): 5581–5593.
- [28] Xie J, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis[C]//*Proceedings of the International Conference on Machine Learning*. PMLR, 2016.
- [29] Li Junxia, Qian Yuhua, Ma Guoshuai, et al. Depth clustering algorithm for adaptive edge samples recognition[J]. *Journal of Southwest University(Natural Science Edition)*, 2023, 45(3): 34–46. [李俊霞, 钱宇华, 马国帅, 等. 自适应边缘样本识别的深度聚类算法[J]. *西南大学学报(自然科学版)*, 2023, 45(3): 34–46.]
- [30] Zhang Zijing, Zhang Jianxun, Quan Wenjun, et al. Deep autoencoder clustering algorithm for multiple time sequences[J]. *Computer Applications Research*, 2023, 40(8): 2387–2392. [张梓靖, 张建勋, 全文君, 等. 多元时序的深度自编码器聚类算法[J]. *计算机应用研究*, 2023, 40(8): 2387–2392.]
- [31] Vessia G, Pisano L, Vennari C, et al. Mimic expert judgement through automated procedure for selecting rainfall events responsible for shallow landslide: A statistical approach to validation[J]. *Computers & Geosciences*, 2016, 86: 146–153.
- [32] Lyu Qi. Statistical inference studies for Gompertz distribution under complex censored schemes[D]. Beijing: Beijing Jiaotong University, 2020. [吕琪. 复杂删失方案下 Gompertz 分布的统计推断研究[D]. 北京: 北京交通大学, 2020.]
- [33] Li Changjiang, Ma Tuhua, Sun Leling, et al. A new approach to calibrate antecedent precipitation for rainfall-triggering landslide forecast[J]. *Journal of Mountain Sci-*

- ence,2011,29(1):81-86.[李长江,麻土华,孙乐玲,等.降雨型滑坡预报中计算前期有效降雨量的一种新方法[J].山地学报,2011,29(1):81-86.]
- [34] Chinkulkijniwat A, Salee R, Horpibulsuk S, et al. Landslide rainfall threshold for landslide warning in Northern Thailand[J]. *Geomatics, Natural Hazards and Risk*, 2022, 13(1): 2425-2441.
- [35] Tanno R. Adaptive neural trees[C]//Proceedings of the International Conference on Machine Learning. PMLR, 2019.
- [36] Chen Yu. Application of artificial neural networks in the design of biomedical materials[J]. *E3S Web of Conferences*, 2021, 233: 02003.
- [37] Gao Binghai, He Yi, Zhang Lifeng, et al. Dynamic evaluation of landslide susceptibility by CNN considering InSAR deformation: A case study of Liujiaxia Reservoir[J]. *Journal of Rock Mechanics and Engineering*, 2023, 42(2): 450-465.[高秉海,何毅,张立峰,等.顾及InSAR形变的CNN滑坡易发性动态评估——以刘家峡水库区域为例[J].岩石力学与工程学报,2023,42(2):450-465.]
- [38] Huang Faming, Zeng Shiyi, Yao Chi, et al. Uncertainties of landslide susceptibility prediction modeling: Influence of different selection methods of “non-landslide samples”[J]. *Advanced Engineering Sciences*, 2024, 56(1): 169-182.[黄发明,曾诗怡,姚池,等.滑坡易发性预测建模的不确定性:不同“非滑坡样本”选择方式的影响[J].工程科学与技术, 2024, 56(1): 169-182.]

## Rainfall-induced Landslide Susceptibility Prediction Considering Spatial Heterogeneity

ZHANG Xingfu<sup>1,2</sup>, JIANG Yuanjun<sup>3\*</sup>, ABI Erdi<sup>1,2</sup>

(1.School of River and Ocean, Chongqing Jiaotong University, Chongqing 400074, China;

2.Key Laboratory of Geological Disaster Reduction for Highway and Waterway Transportation in Mountainous Areas,

Chongqing Municipal Education Commission, Chongqing Jiaotong University, Chongqing 400074, China

3.Institute of Mountain Hazards and Environment, CAS, Chengdu 610041, China;)

### Abstract:

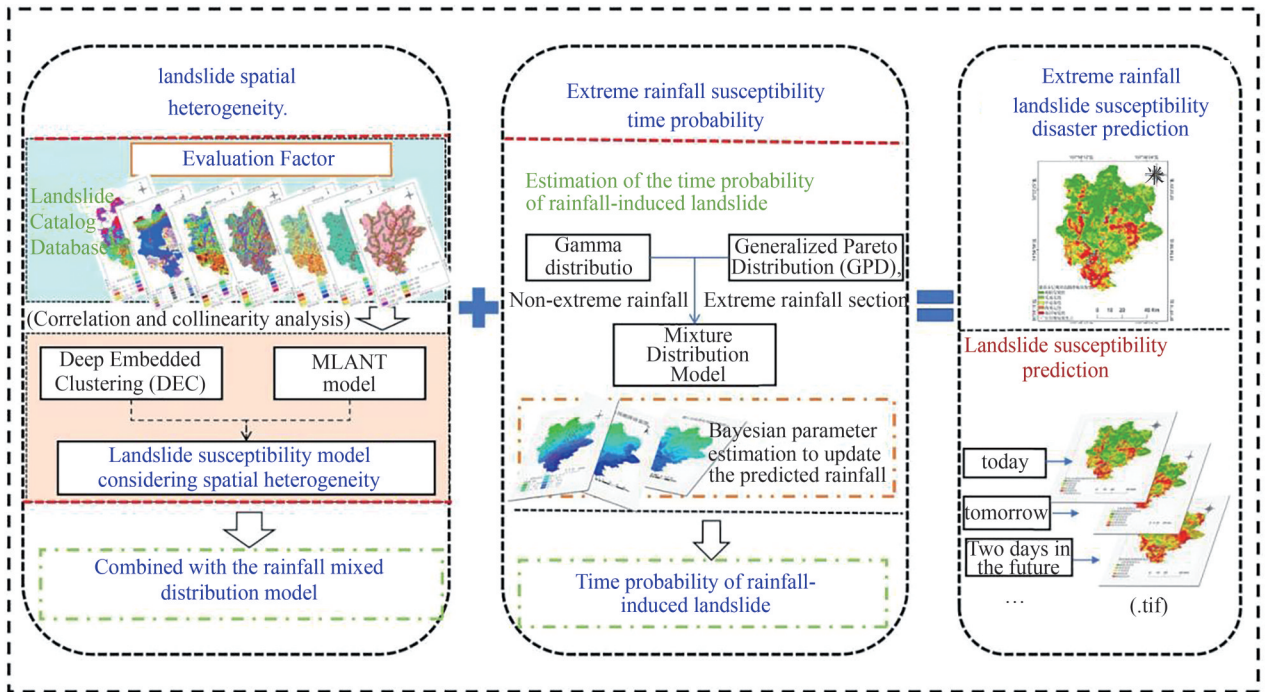
**Objective** Landslides represent a severe and frequent natural hazard, posing significant threats to human life and property. Current models for predicting landslide susceptibility exhibit two primary limitations: the inability to fully capture the spatial heterogeneity of environmental factors such as terrain, soil, and vegetation, and the failure to accurately distinguish between landslides induced by extreme and non-extreme rainfall events. These shortcomings hinder accurate forecasting and reduce the models' adaptability to diverse environmental conditions and rapidly changing climatic patterns. Therefore, this study introduces an innovative approach that combines Deep Embedded Clustering (DEC) with a Dynamic Rainfall Threshold (DRT) model based on a mixed distribution. In addition, a Multi-Task Learning Adaptive Neural Tree (MLANT) model has been developed to enhance model flexibility and prediction accuracy, particularly in varying environmental conditions and during extreme weather events.

**Methods** This research applied three key methodologies to address the limitations of existing landslide susceptibility models. Deep Embedded Clustering (DEC): DEC was utilized to resolve spatial heterogeneity issues. Using deep learning techniques, environmental variables such as terrain, soil, and vegetation were embedded into a low-dimensional space to capture their complex, nonlinear relationships. The model clustered these representations to identify sub-regions with similar geological and environmental features. This clustering-based zoning enhanced the model's capability to analyze landslide susceptibility by managing spatial heterogeneity more effectively than traditional methods. Dynamic Rainfall Threshold (DRT) model: The DRT model was introduced to improve forecasting accuracy by distinguishing between the effects of extreme and non-extreme rainfall on landslide initiation. It integrated Gamma and Generalized Pareto Distributions (GPD) to represent rainfall events of varying intensities. Bayesian methods were employed to dynamically update model parameters, enabling real-time adaptation to changing rainfall conditions. This allowed for greater prediction precision and timeliness under both extreme and non-extreme rainfall events. Multi-Task Learning Adaptive Neural Tree (MLANT): The MLANT model was developed to address the inflexibility and limited adaptability of traditional prediction models. Unlike conventional models, which struggled to adjust to diverse prediction tasks and shifting environmental conditions, MLANT dynamically modifies its processing strategies to match specific tasks and environmental variations. This multi-task learning approach enhances the model's accuracy and flexibility in predicting landslides, particularly in nonlinear environments with complex terrain and rapidly changing weather conditions. The models were validated using real-world data from Tongjiang County, a region prone to landslides. Inputs such as terrain features, soil characteristics, vegetation indices, and rainfall patterns were utilized to train and evaluate the models. The DEC, DRT, and MLANT models were compared to traditional susceptibility models to assess improvements in prediction accuracy and adaptability.

**Results and Discussions** The results demonstrated significant improvements in prediction accuracy when using the proposed models. DEC-Based Clustering: The DEC model predicted higher landslide densities in high- and very-high-susceptibility zones by capturing spatial heterogeneity. Specifically, it predicted landslide densities of 0.036 92 and 0.046 92 events per square kilometer in high and very high-risk zones, respectively, identifying a total of 59 landslide events. This marked an improvement over traditional models that did not consider spatial heterogeneity, such as the overall effective rainfall coefficient model, which predicted fewer events and lower densities. These findings highlighted the impor-

tance of integrating spatial heterogeneity in susceptibility modeling. Dynamic Rainfall Threshold Model (DRT): The DRT model further enhanced accuracy by effectively distinguishing the effects of different rainfall intensities. Its mixed-distribution approach, incorporating both Gamma and GPD distributions, improved the accuracy of landslide predictions under dynamic climatic conditions. In the practical application to Tongjiang County, the DRT model also predicted 59 landslide events in high and very high susceptibility zones, with densities of 0.036 92 and 0.046 92 events per square kilometer, respectively. This outperformed the DEC-based effective rainfall coefficient model, which exhibited lower accuracy and fewer correctly identified events. MLANT Performance: The MLANT model significantly improved flexibility and prediction performance across varied tasks. MLANT efficiently addressed various environmental conditions by dynamically adjusting its internal strategies, particularly in response to landslides triggered by extreme rainfall. Evaluation metrics, such as precision, F1 score, and ROC-AUC, demonstrated that MLANT outperformed traditional models that rely on static thresholds. In Tongjiang County, MLANT increased predicted landslide density from 0.038 events/km<sup>2</sup> and 44 events (using traditional methods) to 0.044 events/km<sup>2</sup> and 59 events, demonstrating superior performance in both frequent and rare landslide scenarios.

**Conclusions** The models developed in this research effectively overcome the limitations of traditional landslide susceptibility prediction approaches by integrating Deep Embedded Clustering (DEC), a Dynamic Rainfall Threshold (DRT) model based on mixed distributions, and a Multi-Task Learning Adaptive Neural Tree (MLANT). Spatial heterogeneity is addressed through DEC clustering, which dynamically partitions regions based on environmental characteristics, improving prediction accuracy. Rainfall differentiation is achieved through the DRT model, which significantly improves the model's capability to predict landslides triggered by both extreme and non-extreme rainfall events by adapting to varying climatic conditions. The MLANT model provides flexibility and adaptability, delivering improved accuracy across multiple prediction tasks and in rapidly changing environmental conditions, outperforming traditional single-task models. These findings demonstrate the critical importance and effectiveness of integrating DEC, DRT, and MLANT models in enhancing landslide susceptibility prediction, particularly in areas characterized by complex environmental conditions and frequent extreme weather events. This research contributes both theoretical and practical advancements, providing valuable tools for enhancing landslide risk management and mitigation strategies.



**Key words:** landslide susceptibility; deep embedded clustering (DEC); spatial heterogeneity; mixed distribution rainfall threshold; multi-task learning adaptive neural tree model

(编辑 李轶楠)

引用格式: Zhang Xingfu, Jiang Yuanjun, Abi Erdi. Rainfall-induced landslide susceptibility prediction considering spatial heterogeneity[J]. Advanced Engineering Sciences, 2025, 57(4): 12–28. [张幸福, 姜元俊, 阿比尔的. 考虑空间异质性的降雨滑坡易发性预测研究[J]. 工程科学与技术, 2025, 57(4): 12–28.]