

基于 GIS 和 Logistic 回归模型的兰州市滑坡灾害敏感性区划研究

方苗^{1,2}, 张金龙¹, 徐瑛¹

(1. 中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000)

(2. 中国科学院研究生院, 北京 100049)

摘要: 针对兰州市脆弱的地质环境和频繁发生的滑坡灾害, 以兰州市为研究区, 采用 Logistic 回归模型, 以 ArcMap 和 SPSS 软件为工具, 选取地层岩性、地貌、坡度、断层构造、植被覆盖度、7-9 月平均降水、道路等作为滑坡灾害影响因子。首先对每个影响因子分级并计算每个因子指标值, 然后在 Arcmap 中对影响因子图层进行叠加操作, 最后在 SPSS 软件中运用 Logistic 回归方法, 计算出每个影响因子的系数值并建立 Logistic 回归模型。根据 Logistic 回归模型在 Arcmap 中绘制兰州市滑坡灾害敏感性区划图, 区划图和实际的滑坡分布情况基本吻合。模型的 Kappa 系数值和 ROC 曲线下面积值 AUC 值分别为 0.623 和 0.709, 两种方法的检验结果均表明模型模拟效果较好, 能应用到兰州市滑坡灾害敏感性区划研究中。

关键词: 滑坡灾害; 敏感性; Logistic 回归; 影响因子; Kappa 值; ROC 曲线

1. 引言

中国是一个自然灾害频发的国家, 每年受灾人口高达 2 亿以上, 因灾造成的直接经济损失逐年增加^[1]。地质灾害是自然灾害的一种重要类型, 而中国是世界上地质灾害最严重的国家之一。仅 2010 年 1 月至 7 月^[2], 全国共发生地质灾害 26009 起, 其中滑坡 19101 起、崩塌 4756 起、泥石流 911 起、地面塌陷 332 起、地裂缝 161 起、地面沉降 36 起, 造成 843 人死亡、失踪, 直接经济损失 33.44 亿元。相关部门的统计结果表明, 滑坡灾害是所有地质灾害中出现次数最多、发生频率最高、造成人员和财产损失最严重的一种地质灾害^[3]。因此, 科学、有效地防治地质灾害尤其是滑坡灾害, 是我国各级政府必须面对的一个重大课题。

滑坡灾害敏感性区划是关于滑坡灾害的发生、发展和分布规律的综合研究^[4]。其划分结果是科学、正确、及时的预测预报滑坡灾害的基础。滑坡灾害敏感性区划方法主要分为两大类, 一种是定性的方法, 另外一种是定量的方法。定性方法是借助于对事物的经验、知识、观察及对发展变化规律的了解, 科学地进行分析、判断的一类方法。定量方法是根据统计数据、检测数据、同类和类似系统的数据资料, 按有关标准, 应用科学的方法构造数学模型进行量化评价的一类方法。

定性方法主要包括专家经验法、层次分析法和加权线性组合法。张文等^[5]以专家经验值为基础, 进行四川平武高平铺库区滑坡易发性评价; 武立新等^[6]基于层次分析法 (AHP) 对山体滑坡危险度评价进行研究; Ouri 等^[7]用层次分析法和 GIS, 完成了 Langan 流域滑坡危险性区划; 许湘华^[8]利用加权线性组合法建立贵州省滑坡危险性分区图; Aykut 等^[9]使用加权线性组合法分析黑海地区的滑坡灾害敏感性并制作区划图。定性方法会根据专家的经验、认识、地域不同而产生不同的结果, 因此定性方法一般只用在区域性的研究当中。

定量方法主要包括确定性方法、人工智能法和多元统计法。确定性方法是通过研究滑坡发生的物理机制的基础上建立数学模型, 计算时需要收集灾害点的各项物理参数, 这种方法适用于局部范围内的单点滑坡评价。Gokceoglu 等^[10]用确定性方法完成了土耳其 Mengen 地

区的滑坡敏感性制图, Baldelli 等^[11]用确定性方法分析了意大利 Messina 地区滑坡敏感性, 两者均取得了较好的研究成果。人工智能方法主要包括专家系统、人工神经网络和模糊逻辑法。Wislocki 等^[12,13]开发了滑坡灾害规划的知识库系统, 对可能影响拟开发场地的滑坡灾害进行评价; Lee 等^[12,14]通过建立空间数据库, 选取与滑坡灾害相关的因素输入到人工神经网络中来分析滑坡敏感性; Ercanoglu 等^[12,15]利用模糊逻辑和 GIS 对土耳其 Yenice 北部的滑坡敏感性进行了评价; 人工智能方法评价过程相对复杂, 而且刘艺梁等^[16]的研究案例也证明人工神经网络方法评价效果不如 Logistic 回归方法好, 在实际应用中出现的频率依然较低。多元统计法主要包括多元线性回归分析、判别分析和 Logistic 回归分析。马还援等^[17]利用多元线性回归分析对安徽省广德县 30 个滑坡的稳定性进行分析, 但是忽略滑坡的发生发展过程是一个复杂的、非线性过程, 因此多元线性回归分析缺少一定合理性; 陆新等^[18]把判别分析法用在重庆市奉节中小型滑坡稳定性评估中, 结果证明了这种方法的有效性; 王卫东等^[4]用 Logistic 回归模型评价贵州省地质灾害敏感性, 并对地质灾害敏感性进行区域划分; Peter 等^[19]用 Logistic 回归模型预测 CWNF(Clearwater National Forest)地区的滑坡灾害敏感性; Lulseged 等^[20]用 Logistic 回归模型评价 Kakuda 和 Yahiko 山区的泥石流灾害敏感性, 并制作灾害敏感性图; 以上三个 Logistic 回归模型应用案例都已证明了此模型的预测精度相对其他预测模型来说较高。Logistic 回归的优势在于进行统计分析时, 自变量可以是连续变量, 也可以是离散变量, 也不要求自变量呈正态分布, 对自变量的分布未作任何要求, 因此增强了模型的应用范围和灵活性。

本文以兰州市为研究区, 根据兰州市滑坡灾害实际情况选取影响兰州市滑坡灾害的因子, 采用 Logistic 回归模型, 借助 ArcGIS 和 SPSS 软件完成研究区的滑坡灾害敏感性区划工作。区划结果能为兰州市高效、科学、及时地开展滑坡灾害防治工作提供依据, 为辖区内的经济建设和社会发展做贡献。

2. 研究区概况

兰州市地处青藏高原与黄土高原交汇部位的黄河谷地, 黄河从西向东穿城而过, 具有带状盆地城市的特征, 市区位于著名的天水—兰州地震带上^[21]。市区南北两边高山环绕, 根据兰州市国土局提供的地貌数据, 兰州市由石质山地、黄土梁区、黄土丘陵区、河谷地和断陷盆地五大类地貌单元组成^[22], 每个大类又包括多个子类。

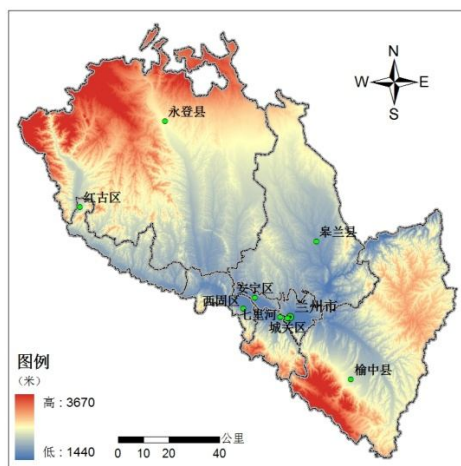


图 1 兰州市数字高程图
Fig.1 Picture of Lanzhou DEM

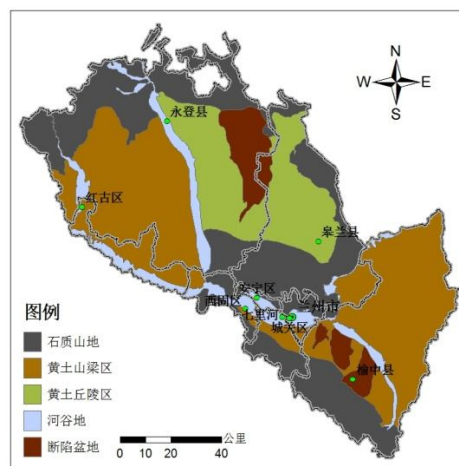


图 2 兰州市地貌图
Fig.2 Picture of Lanzhou Landforms

本地区经历了多个时期的造山运动，地质构造复杂、地形起伏强烈^[22]。构成兰州市南北两山主体的是上覆第四系黄土的新生界红色碎屑岩类，地质分类属于松散岩土，其结构非常疏松，加之高差大，形成的垂直截立面比较多，比较典型的是兰州火车站附近的皋兰山，是由松散的碎屑岩土和疏松的黄土构成，皋兰山北坡坡面极其陡峭，起伏度很大，存在多处垂直立面，相关部门野外调查和历年统计结果显示这一地区是兰州市滑坡灾害发生较多的地方。气候上，兰州地区属于温带半干旱气候，年均降水量 327mm，但是降水集中于 7 至 9 月份、且多以暴雨形式出现^[23]。由此造成黄河两岸山区土质结构疏松，沟壑纵横，地表裸露，多数地方植被覆盖度较低。兰州市是西北地区重要的工业城市，近年来人类工程活动（比如：开挖坡脚、修建道路、采矿、渣土堆积等）对区内地质环境的影响日益强烈，使得境内滑坡地质灾害频繁发生。

3. 研究方法

根据滑坡灾害评价原理和方法以及统计学知识，基于 GIS 和 Logistic 回归模型的兰州市滑坡灾害敏感性区划的流程如图 3 所示，步骤如下：第一步，选择兰州市滑坡灾害的影响因子；第二步，把影响因子根据分类标准划分为二级分类因子，同时根据计算规则计算各类因子的指标值；第三步，对所有影响因子图层进行矢量叠加操作，生成一幅综合评价图层；第四步，找出综合评价图层中哪些多边形中包含有滑坡点，包含了则赋值 1，同时在滑坡点一定半径范围之外选择未包含滑坡点的多边形，赋值为 0，以此构建统计分析的样本数据；第五步，把样本数据导入到 SPSS 软件中进行回归分析，根据分析结果构建回归模型；第六步，根据回归模型计算全区滑坡敏感性并制作敏感性图，最后验证回归模型的模拟效果。

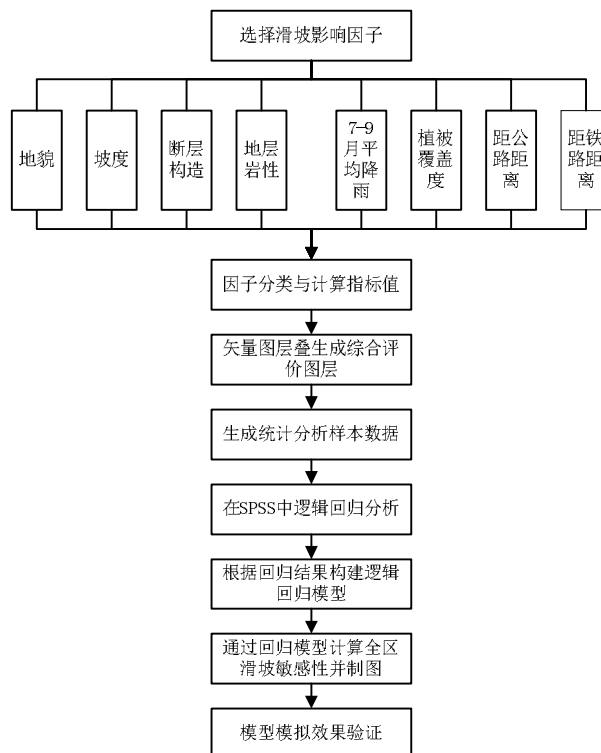


图 3 滑坡灾害敏感性 Logistic 回归评价流程图

Fig.3 Flowchart of Logistic Regression Assessment of Landslide Hazard Susceptibility

4. 数据处理

4.1 影响因子选择

控制地质灾害活动程度的条件，从总体上分为地质条件、地形地貌条件、气候条件、水文条件、植被条件和人类活动等。滑坡灾害的发生、发展、演化过程受到孕育因子和诱发因子的双重控制。各种因子在滑坡形成的过程中有着不同的作用，因子的影响程度也随着区域变化而不同，针对兰州市滑坡灾害发生的实际情况，选取以下 8 个影响因子对兰州市滑坡敏感性进行分析，即：地层岩性、断层构造、高程、坡度、植被覆盖、7~9 月月平均降雨、道路（公路、铁路）。

4.2 因子量化方法

用统计学方法评价地质灾害的过程中，各个影响因子也就是自变量必须得量化，自变量量化的方法有多种。确定性系数 CF 值能反映了灾点的聚集程度，兰恒星等^[24]用 CF 值法确定滑坡发生的关键因子，研究云南小江流域滑坡敏感性；王卫东等^[25]用 CF 值作自变量，研究贵州省滑坡危险性区划。许湘华^[26]用每个影响因子的各个二级分类因子中滑坡灾害面积除以各个二级影响因子的面积，得到一个面积比，然后对各个面积比进行规范化处理，用规范化处理的值作为各个二级分类因子的指标值。以上方法，都考虑到了灾点的分布情况，在滑坡敏感性评价中都是可取的方法。但是，作者考虑用每个影响因子的各个二级分类因子中滑坡灾害点的个数，除以各个二级分类因子的面积，得到一个比值，这个比值不但可以反映单位面积内滑坡灾害点的分布情况，而且不同值之间的比较，能反映两个不同区域内滑坡灾害的发生频率与活动程度。所采用的计算公式如下：

$$x_{ij} = \frac{N_{ij}}{S_{ij}} \quad (1)$$

式中： $i=1,2,\dots,n$ ，是影响因子序号； $j=1,2,\dots,m$ ，是各个影响因子的二级分类因子序号。 x_{ij} 是各类二级因子的指标值； S_{ij} 是各类二级因子的面积； N_{ij} 是各类二级因子中滑坡灾害点的个数；按照（1）式计算出一个初始的 x_{ij} ，然后再把 x_{ij} 归一化处理。

4.3 因子分析与处理

地层岩性：地层岩性是孕育一切地质灾害的物质基础，已有岩体的岩性特征、下伏岩体的岩性特征及其之间的关系显著影响到滑坡的类型、演化趋势等多方面的特征。对于土石性质比较软弱，结构比较松散破碎而蓄水条件较好排水不易的地层，就比较容易产生滑坡。根据岩层类型把地层岩性稳定性划分为以下五个类别，见表 1。

表 1 地层岩性稳定性划分

Table 1 Classes of Geological and Lithological Stability

稳定性级别	地层岩性
稳定岩组	岩浆岩类，花岗岩，闪长岩，橄榄岩，玄武岩（火山岩），辉绿岩等
偶滑岩组	厚层砂岩，厚层砾岩，页岩，凝灰岩，白云岩，大理岩，石英岩，石英砂岩
较易滑岩组	中层粉砂岩，薄层页岩，泥灰岩，灰岩，泥岩，板岩，千枚岩夹碳酸盐，硅质岩，煤系地层
易滑岩组	粘性土，卵石土，碎石土为主的松散堆积层，泥、砂、砾等混合组成物，人工填挖堆积物

断层构造: 断层构造对滑坡的影响主要表现在, 无论是正断层或者是逆断层, 在断层两侧均产生一定宽度的断层破碎带。在断层破碎带或节理裂隙密集处, 由于岩体破碎, 抗剪强度显著降低, 构成软弱带。

坡度: 坡度制约着重力和流水侵蚀作用的强度, 控制着潜在水土流失和崩塌出现的概率。只有处于一定地貌部位、具备一定坡度的斜坡才可能发生滑坡。一般江、河、湖(水库)、海、沟的岸坡, 前缘开阔的山坡、铁路、公路和工程建筑物边坡等都是易发生滑坡的部位。

地貌: 地貌是形成滑坡的基础条件^[22]。从区域地貌条件看, 滑坡主要形成于山地、山间谷底、江河两岸等地, 通常情况下, 地形起伏较大、切割越剧烈的地方都是滑坡高发区。从局部地形看, 要有适宜的斜坡坡度、高度和形态, 以便形成滑动的临空面, 这对滑坡的形成具有最直接的作用。

植被覆盖度: 植被覆盖度的作用体现在植物根系的固结作用, 根系可以固持土体, 稳定山坡土层, 减轻降水对土体的冲击。植被覆盖率的高低反映出植被对地表保护程度的高低。

7~9 月降雨: 国内外大量滑坡实例说明, 降雨对滑坡的产生起着第一位的影响作用。一般要有以下两个条件: 足够的降雨强度; 要有较长的连续降雨时间。兰州地区具有明显的温带大陆性季风气候的特征, 降水量偏少, 各地及各月分配也不均匀, 连阴雨、暴雨, 大雨多集中于 7~9 月。这与兰州市的滑坡灾害多发生在 7~9 月的事实相符合。

人类活动: 工程活动可以破坏坡体, 在坡体下开挖坡脚与边坡挖土使坡脚下部失去支撑, 当斜坡支撑不了斜坡上部质量时就会失去平衡。另外, 修筑铁路、公路、修建农田水利、采矿过程中乱弃土石、废渣等, 这些都可提供大量松散固体物质。兰州市作为西北地区重要的交通枢纽和工业城市, 公路和铁路网密集, 道路修建工程密度大、强度高, 因此是兰州市滑坡灾害发生的一个重要影响因素。

影响因子在不同等级下对滑坡的影响程度不同, 因此, 需要对每个影响因子进一步细化等级。同时, 在 ArcMap 中计算影响因子二级分类因子的指标值, 二级分类因子及指标值计算结果如表 2 所示。表中 x_{ij}^0 表示初始化指标值, x_{ij} 代表归一化后指标值。

表 2 影响因子分级与指标值
Table 2 Classes and Index Value of Impacted Factors

因子	分级序号	二级因子	二级因子中滑坡数	二级因子面积 (km ²)	x_{ij}^0	x_{ij}
地层岩性	1	稳定岩组	5	1059.02	0.0047	0.0502
	2	偶滑岩组	48	2232.48	0.0215	0.2297
	3	较易滑岩组	32	1507.35	0.0212	0.2265
	4	易滑岩组	121	4032.86	0.0300	0.3205
	5	极易滑岩组	72	4439.29	0.0162	0.1731
距断层距离(m)	1	>10000	106	4789.38	0.0221	0.2077
	2	6000—10000	48	2407.95	0.0194	0.1823
	3	2000—6000	62	3327.00	0.0186	0.1748
	4	1000—2000	28	1222.33	0.0229	0.2152
	5	<1000	34	1451.41	0.0234	0.2200
坡度	1	<10	124	6614.47	0.0187	0.1698

(°)	2	10—25	128	5648.04	0.0227	0.2062
	3	25—35	20	877.85	0.0228	0.2071
	4	>35	6	130.64	0.0459	0.4169
	<hr/>					
地貌	1	石质山地	94	4301.76	0.0219	0.1894
	2	黄土山梁区	102	5046.73	0.0202	0.1751
	3	黄土丘陵区	22	2065.19	0.0107	0.0923
	4	河谷地	58	1025.04	0.0566	0.4911
	5	断陷盆地	5	831.96	0.0060	0.0521
<hr/>						
植被覆盖度	1	<0.2	22	1149.76	0.0191	0.1598
	2	0.2—0.35	113	7122.96	0.0158	0.1322
	3	0.35—0.55	87	2945.63	0.0295	0.2469
	4	0.55—0.75	44	1008.87	0.0436	0.3649
	5	>0.75	12	1043.77	0.0115	0.0962
<hr/>						
7~9月降雨(mm)	1	<160	14	2211.59	0.0063	0.0779
	2	160—180	55	3030.78	0.0181	0.2237
	3	180—205	149	6134.91	0.0243	0.3004
	4	>205	61	1893.71	0.0322	0.3980
<hr/>						
距公路距离(m)	1	>1000	162	10849.10	0.0149	0.0617
	2	600—1000	32	930.39	0.0344	0.1424
	3	300—600	47	973.36	0.0483	0.2
	4	100—200	20	250.08	0.0800	0.3313
	5	<100	17	266.06	0.0639	0.2646
<hr/>						
距铁路距离(m)	1	>2000	214	11867.34	0.0180	0.0849
	2	1000—2000	24	664.47	0.0361	0.1702
	3	500—1000	24	362.14	0.0663	0.3126
	4	200—500	6	222.66	0.0269	0.1268
	5	<200	10	154.38	0.0648	0.3055

5. Logistic 回归模型及其在滑坡中的应用

5.1 Logistic 回归模型

在地质灾害分析中,影响因子可以作为自变量,而灾害的发生与否可以作为二值变量(0代表灾害不发生,1代表灾害发生)^[27]。由于影响因子变量中存在不连续变量(如:地层岩性等级图),所以多元线性回归法将不适用于推导此类自变量和因变量之间的关系。而采用 Logistic 回归模型可以解决这种类型的问题^[4,28]。Logistic 回归模型的具体公式如下:

$$P = \frac{e^z}{1 + e^z} \quad (2)$$

$$z = b_0 + b_1 \cdot x_{1j} + b_2 \cdot x_{2j} + \dots + x_{nj} \quad (3)$$

式中 P 为灾害发生的概率,取值范围为 (0,1); b_1, b_2, \dots, b_n 为逻辑回归系数, b_0 表示某一因素改变一个单位时,效应指标发生与不发生事件的概率之比的对数变化值。如果逻辑回归系数确定,则根据不同的指标值 x_{ij} ,即可计算某一区域发生灾害的概率 P 值。根据 P 值大小,可以划分灾害敏感性等级。因此,需要利用各因子指标值求算逻辑回归系数。

5.2 模型应用

在 ArcMap 中把分级的因子图层进行叠加,生成综合评价图层。综合评价图层中的多边形被称作“均一条件单元”,每个均一条件单元都包含每一个影响因子的单一条件分组,从而可以认为均一条件单元之间是相互独立的,这样便于进行统计分析^[4,24]。把滑坡灾害点与综合评价图层进行叠置,判断哪些均一条件单元中包含有滑坡灾害点,如果包含有滑坡灾害点,这个均一条件单元取值为 1,表明这个单元内发生了滑坡灾害;因为滑坡灾害点附近一定范围内的自然条件与滑坡灾害点所在的自然条件极其类似,虽然没有表现出明显的滑坡活动,但是其孕育滑坡灾害的可能性非常大,仅仅在当前未明显表现出来。在选择未发生滑坡灾害的区域作为统计分析的样本时,为了使样本更具代表性和说服力,需要在滑坡灾害点一定距离之外选择没有发生滑坡灾害的区域。鉴于此,作者所采用的方案是:对滑坡灾害点做半径为 100 米的缓冲区,得到滑坡灾害点缓冲区图层,图层中包含与滑坡灾害点数量相等的圆,把滑坡灾害点缓冲区图层与综合评价图层进行相交分析,判断综合评价图层中哪些均一条件单元没有与任何滑坡灾害点缓冲圆相交,没有相交的均一条件单元取值为 0,表明这个单元内没有发生滑坡灾害。

把从综合评价图层中获取的统计样本导入 SPSS 软件中进行 Binary Logistic Regression 分析,根据分析结果构建 Logistic 回归模型。Logistic 回归分析结果中,坡度和 7~9 月平均降雨量的系数所对应的显著水平 Sig 值略大于 0.05 外,其余均小于 0.05,可以认为回归结果通过 5%显著性水平检验。因此,根据回归分析结果中的各因子系数值,构建回归模型的式子如下所示:

$$z = -0.958 + 5.965x_{1j} + 1.339x_{2j} + 3.441x_{3j} + 1.066x_{4j} - 2.143x_{5j} - 3.971x_{6j} - 4.079x_{7j} + 6.102x_{8j} \quad (4)$$

式中: $x_{1j}, x_{2j}, \dots, x_{mj}$ 分别代表地层岩性、地貌、7~9 月平均降雨、植被覆盖度、坡度、距公路距离、距铁路距离和断层构造等因子的指标值。

根据上面构建的 Logistic 回归模型,在 ArcMap 中计算综合评价图层中每个均一条件单元的滑坡灾害概率值 P,最后对整个图层的概率值进行分类,总共分为五种类型。Arcmap 自动提供了几种分类方法,虽然分类方法各有利弊,但不影响灾点的整体分布^[26]。最后,在 ArcMap 中制作兰州市滑坡灾害敏感性区划图,如图 4 所示。

滑坡灾害敏感性区划结果如表 3 所示,从表中数据可以看出,滑坡灾害落在高度敏感区和极度敏感区的占滑坡总数的 64.39%,落在中度敏感以上区域的占滑坡总数的 82.38%,说明区划结果与实际滑坡灾害分布情况基本吻合,区划结果可以用来指导滑坡灾害的防治工作。根据区划结果图,兰州市滑坡灾害防治重点区域在图 4 中高度敏感区和极度敏感区,中度敏感区可以有选择的参考,而不敏感区和轻度敏感区不作为防治重点。

表 3 兰州市滑坡灾害敏感性区划结果

Table 3 Result of Landslide Susceptibility Zoning Study in Lanzhou City

敏感性分区	分区面积 (Km ²)	面积比例 (%)	分区内滑坡点个数	滑坡比例 (%)
不敏感	1903.19	14.34	28	10.07
轻度敏感	4018.26	30.28	31	11.15
中度敏感	3861.30	29.09	50	17.99
高度敏感	2570.22	19.37	76	27.34
极敏感	918.03	6.92	103	37.05

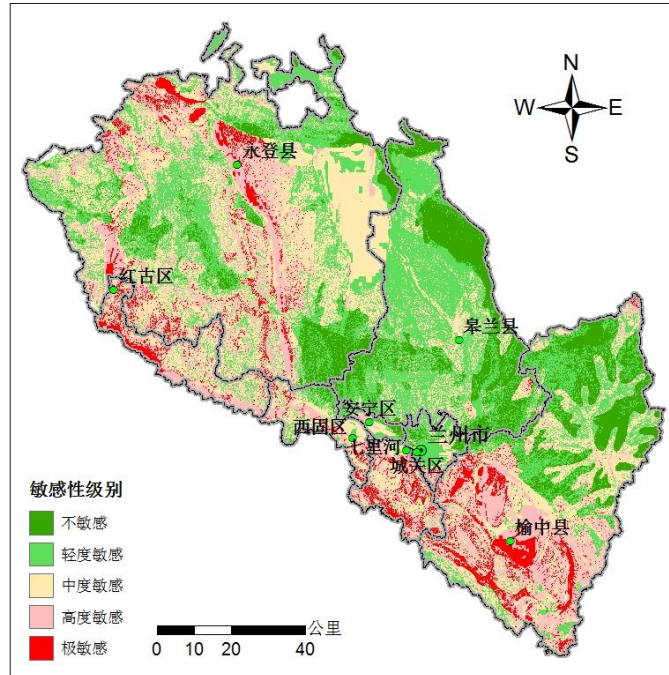


图 4 兰州市滑坡灾害敏感性区划图

Fig. 4 Zoning Map of Lanzhou City Landslide Hazard Susceptibility

5.3 结果验证

根据 Logistic 回归模型完成兰州市滑坡灾害敏感性区划，为了验证区划结果的合理性和模型的准确性，需要对模型进行检验。选取 Kappa 值和 ROC 曲线这两种指标对模型效果进行检验，两种方法都可以用来比较模型模拟精度。

对于检验样本，以 1 代表滑坡发生，0 代表滑坡未发生，则观测值和模型模拟值之间的关系可由下面的二分类表来表示：

表 4 观测值和模拟值的二分类表

Table 4 Two Classification Table of the Observed Value and the Simulated Value

观测值	模拟值	样本数
1	1	N_1
1	0	N_2
0	1	N_3
0	0	N_4

N_1 表示观测值和模拟值一致且都是滑坡发生的数量； N_2 表示观测值是滑坡发生而模拟值是滑坡未发生的数量； N_3 表示观测值是滑坡未发生而模拟值是滑坡发生的数量； N_4 表示观测值和模拟值一致且都是滑坡为发生的数量。

5.3.1 Kappa 值验证

Kappa 系数可反映观测值和模拟结果之间的一致性。其计算过程如下所示：

$$kappa = \frac{p_a - p_e}{1 - p_e} \quad (5)$$

$$p_a = \frac{N_1 + N_4}{N_1 + N_2 + N_3 + N_4} \quad (6)$$

$$p_e = \frac{(N_1 + N_2)(N_1 + N_3) + (N_4 + N_2)(N_4 + N_3)}{(N_1 + N_2 + N_3 + N_4)^2} \quad (7)$$

Kappa 值法是通过模拟结果取不同的阈值,得到不同的 Kappa 值,其中最大的 Kappa 值被用来评估模型^[29,30],评估标准为^[31]: Kappa 值<0.4, 失败; 0.4~0.55, 一般; 0.55~0.7, 好; 0.7~0.85, 很好; >0.85, 非常好。

将观测样本值和模型模拟值通过(5)、(6)、(7)式进行计算,得到 Kappa 值为 0.623,说明模型模拟结果好,回归模型可以应用到兰州市滑坡灾害敏感性区划研究中。

5.3.2 ROC 曲线验证

受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)是根据一系列不同的二分类方式,以真阳性率(灵敏度)为纵坐标,假阳性率(1-特异度)为横坐标绘制的曲线,即 ROC 曲线来评估两类分类算法的性能。其描绘的是诊断中敏感性和特殊性之间的制约关系,能更全面客观地对评估系统进行分析^[32]。ROC 曲线下的面积值在 1 和 0.5 之间,面积越接近 1,说明模型的判别效果越好。其评判标准为:在 AUC>0.5 的情况下,AUC 越接近于 1,说明诊断效果越好。AUC 在 0.5~0.7 时有较低准确性,AUC 在 0.7~0.9 时有较高准确性,AUC 在 0.9 以上时有很高准确性。把 ROC 曲线应用到本文的回归模型检验中,其中以下三个概念需要特别说明:

① 灵敏度(Sensitivity):就是把实际为真值的判断为真值的概率,即实际上是滑坡且模型判断为滑坡的概率。

② 特异度(Specificity):就是把实际为假值的判断为假值的概率,即实际上不是滑坡且模型判断不是滑坡的概率。

③ 误判率:就是把实际为假值的判断为真值的概率,其值等于 1-特异度(1-Specificity),即实际上不是滑坡但模型却判断成滑坡的概率。

将观测样本值和模型模拟值导入 SPSS 软件中进行 ROC 分析,得到模型的 ROC 曲线和 AUC 值,曲线如图所示,AUC 值为 0.709,表明模型的模拟结果具有较高的准确性,回归模型可以应用到兰州市滑坡灾害敏感性区划研究中。

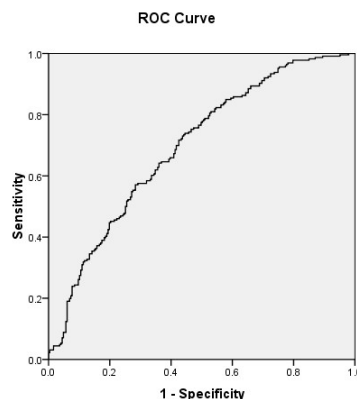


图5 Logistic回归模型的ROC曲线

Fig.5 ROC Curve of Logistic Regression Model

通过 Kappa 值法和 ROC 曲线法验证,说明本次研究所得的兰州市滑坡灾害敏感性的

Logistic 回归模型可以用来计算滑坡敏感性, 以完成兰州市滑坡灾害敏感性区划工作。

6. 结论与讨论

6.1 结论

(1) 针对兰州市地质构造复杂、地貌类型多样、滑坡灾害事件频繁发生的特点, 本文选择 Logistic 回归模型, 以 ArcGIS 和 SPSS 为工具, 选取 8 种滑坡灾害影响因子, 建立滑坡灾害敏感性评价的 Logistic 回归模型, 完成了兰州市滑坡灾害敏感性区划。研究结果与兰州市滑坡灾害的实际分布情况具有很高的符合度, 可以用于兰州市滑坡灾害防治工作中。

(2) 文章对滑坡灾害敏感性评价与区划的方法进行了回顾, 同时对应用 Logistic 回归模型进行滑坡灾害敏感性评价与区划的流程进行了详细的阐述。对滑坡灾害影响因子选择、影响因子的分级、分级指标值的计算、统计样本的建立、样本分析、模型建立、模型应用、模型验证等过程进行了讨论与分析, 使整套评价方法具有一定的借鉴意义。

(3) 通过 Kappa 值和 ROC 曲线对回归模型的模拟效果进行验证, 证明了模型的有效性, 同时也说明整套评价方法适合于兰州市滑坡灾害的敏感性评价与区划。Arcmap 具有强大的数据分析与处理功能, SPSS 软件具有完善的数据统计分析功能, 两者的有效结合进一步提高了空间数据分析与处理的效率。

6.2 讨论

(1) 文章运用 GIS 方法和 Logistic 回归模型完成兰州市滑坡灾害敏感性评价, 且取得较好的研究结果, 但是并不意味着 GIS 方法和统计学方法能完全取代其他的敏感性评价方法, 各种评价方法只有结合具体的实际情况, 辅以必要的野外调查和室内分析, 才能在具体的研究区域内得到较理想的评价结果。

(2) 本文中在构建统计样本的过程中, 用滑坡灾害点做半径为 100 米得缓冲区, 然后判断哪些多边形不与缓冲区相交, 以此作为未发生滑坡的候选样本。对滑坡灾害点做缓冲区在从理论上来说具有一定的合理性, 但是缓冲半径的大小如何设定目前尚没有一个统一的标准。点缓冲分析半径的大小, 影响着缓冲操作之后生成的圆的大小, 继而影响样本选择的区域范围, 当半径较小, 与之相交的多边形较多, 样本的选择余地较大; 当半径较大, 与之相交的多边形较少, 样本的选择余地较小。而样本选择是否具有代表性直接影响着回归分析结果和模型模拟效果。但是, 本文没有针对缓冲半径大小与模型评价结果好坏之间的关系进行分析与讨论, 只能在下一步的工作中继续深入研究。

参考文献:

- [1] Li Baojun. Current Situation and Development of Disaster Information Management System in China[J]. Disaster Reduction in China, 2004, (1):26-30. [李保俊. 中国灾害信息管理系统现状与发展[J]. 中国减灾, 2004, (1): 26-30.]
- [2] The Author of Science and Technology Daily Difficulties of Geological Disaster Prediction[N]. Science and Technology Daily, 2010-08-26 (5). [科技日报综合. 地质灾害预报到底难在哪[N]. 科技日报, 2010-08-26 (5).]
- [3] Li Yuan, Meng Hui, Dong Jing, *et al.* Main Types and Characteristics of Geohazard in China—Based on the results of Geohazard Survey in 290 Counties[J]. The Chinese Journal of Geological Hazard and Control, 2004, 15 (2):29-34. [李媛, 孟晖, 董颖等. 中国地质灾害类型及其特征——基于全国县市地质灾害调查成果分析[J]. 中国地质灾害与防治学报, 2004, 15 (2):29-34.]

- [4] Wang Weidong, Zhong Cheng. The Application of GIS-based Logistic Regression for Geological Hazards Zonation[J]. *Geotechnical Investigation&Surveying*, 2009, (11): 5-10. [王卫东, 钟晟. 基于 GIS 和 Logistic 回归模型在地质灾害危险性区划中的应用[J]. *工程勘察*, 2009, (11): 5-10.]
- [5] Zhang Wen, Bai Shibiao, Wang Jian. Regional Landslide Susceptibility Assessment Based on Expert Experience—— A Case Study of Gaopingpu Reservoir Area, Pingwu County, Sichuan Province[J]. *Journal of Geological Hazards and Environment Preservation*, 2010, 24(4):20-24. [张文, 白世彪, 王建. 基于专家经验值的滑坡易发性评价——以四川平武高平铺库区为例[J]. *地质灾害与环境保护*, 2010, 24(4):20-24.]
- [6] Wu Lixin, Wang Jianguo, Yang Jizhun. Application of Analytic Hierarchy Process in Assessment of Landslide Danger Degree[J]. *YOUSE JINSHU*, 2009, 61(4): 66-69. [武立新, 王建国, 杨计准等. 层次分析法在山体滑坡危险度评价中的应用[J]. *有色金属(矿山部分)*, 2009, 61(4): 66-69.]
- [7] Ouri A E, Amirian S. Landslide Hazard Zonation Using MR and AHP Methods and GIS Techniques in Langan Watershed, Ardabil, Iran[J]. In: *International conference on ACSRS 2009*, 19-22, Oct, Beijing, China.
- [8] Xu Xianghua. Establishing Map of Landslide Disaster Zonings in Guizhou[J]. *Journal of Railway Science and Engineering*, 2009, 6(4):59-63. [许湘华. 贵州省滑坡灾害危险性分区图编制研究[J]. *铁道科学与工程学报*, 2009, 6(4):59-63.]
- [9] Aykut A, Serhat D, Fikri B. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models[J]. *Environ Geol*, 2008, (54):1127 - 1143.
- [10] Gokceoglu C, Aksoy H. Landslide Susceptibility Mapping of the Slopes in the Residual Soils of the Mengen Region(Turkey) by Deterministic Stability Analyses and Image Processing Technique[J]. *Engineering Geology*, 1996, (44):147-161.
- [11] Baldelli P, Aleotti P, Polloni G. Landsliding-susceptibility Numerical Mapping at the Messina Strait crossing site, Italy[J]. In: *Proceedings of the 7th International Symposium on Landslides*, Trondheim, June, 1996, 1:153-158.
- [12] Neaupane K M, Piantanakulchai M, Li Ye (Translate). Analytic Network Process Model for Landslide Hazard Zonation[J/OL]. Available From: http://www.cigem.gov.cn/qingbao_ReadNews.asp. [Neaupane K M, Piantanakulchai M, 李焯(译). 滑坡灾害分区的网络分析法模型[J/OL]. 获取地址: http://www.cigem.gov.cn/qingbao_ReadNews.asp.]
- [13] Wislocki A P, Bentley S P. An Expert System for Landslide Hazard and Risk Assessment [J]. *Computers and Structures*, 1991, 40(1):169-172.
- [14] Lee S, Ryu J H, Lee M J, *et al.* Use of an Artificial Neural Network for Analysis of Susceptibility to Landslides at Boun, Korea[J]. *Environmental Geology*, 2003, 34(1):59-69.
- [15] Ercanoglu M, Gokceoglu C. Assessment of Landslide Susceptibility for a Landslide-prone area (north of Yenice, NW Turkey) by Fuzzy Approach[J]. *Environmental Geology*, 2002, 41(6):720-730.
- [16] Liu Yiliang, Yin Kunlong, Liu Bin. Application of Logistic Regression and Artificial Neural Network in Spatial Assessment of Landslide Hazards[J]. *Hydrogeology&Engineering Geology*, 2010, 37(5):92-96. [刘艺梁, 殷坤龙, 刘斌. 逻辑回归和人工神经网络模型在滑坡灾害空间预测中的应用[J]. *水文地质工程地质*, 2010, 37(5):92-96.]
- [17] Ma Haiyuan, Song Chuanzhong, Xiang Fan, *et al.* The Multiple Regression Analysis Applied to the Slope Stability, Analysis of Guangde County, in Anhui Province[J]. *City Geology*, 2006, 1(2):42-45. [马还援, 宋传中, 向钊等. 多元回归分析在安徽省广德县区域边坡稳定性分析中的应用[J]. *城市地质*, 2006, 1(2):42-45.]

- [18] Lu Xin, Wu Yue, Zhu Jiankai. Application of Classification Method in Stability Assessment of Smalland Medium Landslides [J]. Chinese Journal of Underground Space and Engineering, 2010, 6(6):1270-1275. [陆新, 吴越, 朱建凯. 判别分析法在中小型滑坡稳定性评估中的应用[J]. 地下空间与工程学报, 2010, 6(6):1270-1275.]
- [19] Peter V, Paul G, Randy B. Spatial Prediction of Landslide Hazard Using Logistic Regression and GIS[C]//4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4):Problems, Prospects and Research Needs. Banff, Alberta, Canada, September, 2000.
- [20] Lulseged A, Hiromitsu Y. The Application of GIS-based Logistic Regression for Landslide Susceptibility Mapping in Kakuda-Yahiko Mountains, Central Japan[J]. Geomorphology, 2005, 65(1-2):15-31.
- [21] Ma Jinhui, Nian Yanyun, Cai Dihua, *et al.* Factors of Regional Landslide Risk and Correlation between Landslide and Geology Structure in Lanzhou Area[J]. Journal of Natural Disasters, 2006, 15(3):14-17. [马金辉, 年雁云, 蔡迪花, 等. 兰州地区滑坡风险因素及其与区域构造的关系[J]. 自然灾害学报, 2006, 15(3):14-17.]
- [22] Lanzhou Bureau of Land and Resources, Gansu College of Geological Environment Monitoring. Geological Disaster Prevention Plan in Lanzhou City[M]. Lanzhou: Gansu Science and Technology Press, 2009. [兰州市国土资源局, 甘肃省地质环境监测院. 兰州市地质灾害与防治 [M]. 兰州: 甘肃省学出版社, 2009.]
- [23] Chang Junjie. Formative Causes of Landslide and Debris Flow in Lanzhou City and Preventives[J]. Research of Soil and Water Conservation, 2003, 10(4):250-252. [畅俊杰. 兰州市市区滑坡泥石流危害、成因及其防治对策[J]. 水土保持研究, 2003, 10(4):250-252.]
- [24] Lan Hengxing, Wu Faquan, Zhou Chenghu, *et al.* Analysis on Susceptibility of GIS-based Landslide Triggering Factors in Yunnan Xiaojiang Watershed[J]. Chinese Journal of Rock Mechanics and Engineering, 2002, 21(10):1500-1506. [兰恒星, 伍法权, 周成虎等. 基于 GIS 的云南小江流域滑坡因子敏感性分析[J]. 岩土力学与工程学报, 2002, 21(10):1500-1506.]
- [25] Wang Weidong, Chen Yanping, Zhong Cheng. Landslides Susceptibility Mapped with CF and Logistic Regression Model[J]. Journal of Central South University(Science and Technology), 2009, 40(4):1127-1132. [王卫东, 陈燕平, 钟晟. 应用 CF 和 Logistic 回归模型编制滑坡危险性区划图[J]. 中南大学学报(自然科学版), 2009, 40(4):1127-1132.]
- [26] Xu Xianghua. Approach Study of Landslides Susceptibility Mapping Using Logistic Regression Model[J]. Journal of Railway Science and Engineering, 2010, 7(5):87-91. [许湘华. 用 Logistic 回归模型编制滑坡灾害敏感性区划图的方法研究[J]. 铁道科学与工程学报, 2010, 7(5):87-91.]
- [27] Cong Weiqing, Pan Mao, Li Teifeng, *et al.* Key Research on Landslide and Debris flow Hazard Zonation based on GIS[J]. Earth Science Frontiers, 2006, 13(1):185-190. [从威青, 潘懋, 李铁峰等. 基于 GIS 的滑坡、泥石流灾害危险性区划关键技术研究[J]. 地学前缘, 2006, 13(1):185-190.]
- [28] Wang Jichuan, Guo Zhigang. Method and Application of Logistic Regression Model [M]. Beijing: Higher Education Press, 2001. [王济川, 郭志刚. Logistic 回归模型——方法与应用[M]. 北京: 高等教育出版社, 2001.]
- [29] Cao Mingchang, Zhou Guangsheng, Weng Ensheng. Application and Comparison of Generalized Models and Classification and Regression Tree in Simulation Tree Species Distribution[J]. Acta Ecologica Sinica, 2005, 25(8):2031-2040. [曹铭昌, 周广胜, 翁恩生. 广义模型及分类回归树在物种分布模拟中的应用与比较[J]. 生态学报, 2005, 25(8):2031-2040.]
- [30] Fieldings A H, Bell J F. A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models[J]. Environmental Conservation, 1997, (24):38-49.

- [31] Monsrud R A, Leemans R. Comparing Global Vegetation Maps with the Kappa Statistic[J]. Ecological Modelling, 1992, 62:275-293.
- [32] Wei Xiuxi, Zhou Yongquan. A New Performance Categories Evaluation Method Based on ROC Curve[J]. Computer Technology and Development, 2010, 20(11):47-50. [韦修喜, 周永权. 基于 ROC 曲线的两类分类问题性能评估方法[J]. 计算机技术与发展, 2010, 20(11):47-50.]

Landslide Susceptibility Zoning Study in Lanzhou City Based on GIS and Logistic Regression Model

Fang Miao^{1, 2}, Zhang Jinlong¹, Xu Zhen¹

(1. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: For fragile geological environment and the frequent occurrence of landslides in Lanzhou city, in this paper, based on Logistic regression model and adopted ArcGIS and SPSS softwares, selected elevation, slope, July-September average rainfall, vegetation cover, roads, railways, lithology, faults as impacted factors of Lanzhou landslide disaster. First of all, impacted factors were classified and calculated index values of every class at the same time. and then overlapped all impacted factors in Arcmap of ArcGIS software. The end, use of Logistic multiple non-linear statistical methods to calculate the coefficient of each impacted factor in SPSS software, according to the statistical results, Logistic regression model was gained. Drawed the Lanzhou landslide susceptibility zoning map based on Logistic regression model in Arcmap, zoning map of landslides accord with the actual distribution of landslides. Kappa value and AUC value of the model were 0.623 and 0.709 respectively, the tested results of the two methods showed that the model is good and can be applied to the Lanzhou landslide susceptibility zoning studies.

Key words: Landslide hazard; Susceptibility; Logistic regression; Impacted factor; Kappa value; ROC curve

附属信息:

基金项目: 中国科学院“西部之光”人才计划资助项目“兰州市地质灾害脆弱性及其防治的信息化研究”(Y028A11001)资助。

作者简介:

方苗(1984—),男,陕西安康人,硕士研究生,主要从事GIS应用开发与地质灾害防治研究。

Email: fangmiaomiaomiao@126.com Tel: 13893257155

单 位: 中国科学院旱区寒区环境与工程研究所遥感与地理信息科学研究室