

# 基于 Logistic 回归的区域地质灾害 综合气象预警模型

徐 晶<sup>1,2</sup> 张国平<sup>1</sup> 张芳华<sup>1</sup> 齐 丹<sup>1</sup>

(1. 国家气象中心, 北京 100081; 2. 中国气象科学研究院)

**提 要:** 应用 Logistic 回归拟合地质灾害发生概率, 建立了区域地质灾害综合气象预警模型, 实现了降雨引发地质灾害发生概率的动态预报。降雨观测和数值降雨预报作为模型的动态输入, 利用信息量模型集成地学因子得到的总信息量作为模型的静态预报因子。预警模型在栅格点上预报降雨引发地质灾害的发生概率, 并以等级形式发布预警。用 2004 年实际观测对预报进行检验, 结果表明当年 80% 的灾害都达到了预警发布的标准。2006 年汛期的业务运行对台风碧利斯引发的地质灾害做出了较准确的预报预警。说明模型能满足预警业务的要求, 是提高地质灾害气象预警水平的有效途径。

**关键词:** Logistic 回归 地质灾害 降雨 空间分析 信息量模型

## Regional Integrated Meteorological Forecasting and Warning Model for Geological Hazards Based on Logistic Regression

Xu Jing<sup>1,2</sup> Zhang Guoping<sup>1</sup> Zhang Fanghua<sup>1</sup> Qi Dan<sup>1</sup>

(1. National Meteorological Centre, Beijing 100081; 2. Chinese Academy of Meteorological Sciences)

**Abstract:** Regional Integrated Meteorological Forecasting and Warning Model for Geological Hazards is constructed based on Logistic regression of the probabilities of geological hazard occurrences. Dynamic forecasting of the probability of rainfall-induced geological hazard occurrence is thus realized. Rainfall observations and numerical rainfall forecasts are used as dynamical inputs of the model. Total information produced by information model that integrates factors of geosciences is used as the static predictor of the model. The model forecasts probabilities of geological hazard occurrences on a grid, and releases warning messages in the form of a 5-level hierarchy. Validation of the model with observational data for the year 2004 shows that 80% of the geolog-

ical hazards of the year have been identified as Warning Level 3 or above. The operation of the model during the rainy season of the year 2006 has made forecasts and warnings with high accuracy for geological hazards triggered by Typhoon Bilis. All these show that the model can satisfy the requirements of warning operations, and is an effective way to improve the meteorological warnings for geological hazards.

**Key Words:** geological hazard information model logistic regression rainfall spatial analysis

## 引 言

滑坡、泥石流、崩塌等地质灾害的发生受多种地理、地质和环境因素的影响,如坡度、高程、岩性、植被状况等,但降雨是最重要的触发条件。国土资源部全国县(市)地质灾害调查结果显示,降雨不仅是全部的泥石流,也是 90% 的滑坡和 81% 的崩塌灾害的引发因素<sup>[1]</sup>。

近十余年来,降雨引发地质灾害的机理及预报预警技术的研究取得了长足进展<sup>[2-13]</sup>。其中,基于统计学的 Logistic 回归,利用高分辨率的地学数据来拟合地质灾害的发生概率是预报预警技术的发展方向之一<sup>[14-17]</sup>。由于引发地质灾害的降雨过程和灾害发生的地学背景因子在物理性质、时空尺度和观测手段等方面有很大的差异,已有的工作未能有效地将动态降雨因子引入发生概率模型,只能进行灾害易发性评估,不能对地质灾害的发生进行动态预报。

本文利用遥感技术、地理信息技术以及数值天气预报技术的发展成果,应用 Logistic 回归,建立了具有动态预报能力的区域地质灾害综合气象预警模型。在建立模型过程中,着重解决了如下两个问题:

(1) 灾害触发雨量的空间分析。准确还原灾害发生点的历史触发雨量是建立模型的关键之一。以往的研究只是简单采用灾害点邻近测站的历史记录来代替,这样做往往会低估触发雨量。采用薄板平滑样条函数空间

分析方法则能够客观地还原历史灾害触发雨量。

(2) 地学因子的集成与表现形式。将影响地质灾害发生的各地学因子以恰当的形式表示出来,是建立模型所需解决的另一个关键问题。信息量模型能够有效地集成对灾害发生有显著驱动作用的地学因子,并以总信息量的形式作为模型的预报因子。

## 1 数据来源及处理

### 1.1 历史灾情数据

地质灾害数据由国土资源部中国环境监测院提供,时间跨度为 1980—2004 年。降雨引发的滑坡、泥石流和崩塌等地质灾害统一作为历史灾情考虑。假设地质灾害点在当年发生地质灾害前的时期为发育期,这段时间内的降雨作为未触发灾害的降雨事件;发生灾害时的降雨则作为触发灾害的降雨事件。两类事件同时作为样本来拟合模型。

### 1.2 雨量数据

雨量数据为 1990—2004 年全国各气象站观测的每日到 20 时为止的前 24 小时累计降雨量序列,该数据来自国家气象信息中心,经空间分析后用于建模。

为了恢复灾害发生点的触发雨量以及前期历史降雨,重建相关的降雨指标,运用了薄板平滑样条函数方法<sup>[18-19]</sup>进行雨量空间分析。该方法具有良好的局地特性,除了站点坐标之外,还可以将高程、地形等影响降雨分

布的因素考虑进来,有可能还原出较周围测站记录更高的局地强降雨,提高了插值的准确性。同时,该方法的实现无需过多的人工干预,能够满足预警业务自动运行的需要。

### 1.3 地质、地理和环境数据

根据预报区域的地质、地理和环境特征,运用地质学、力学等原理从物理机制上分析地质因子与地质灾害发生的内在联系,以此来确定地质灾害发生的背景条件。这部分内容另文详述<sup>[20]</sup>。根据分析结果,采用高程、坡度、高差、断层密度、地震烈度、工程地质岩土分类和土地利用地质灾害发生的地质背景因子。

地质因子数据由国家气象中心提供,并反演成为栅格形式。数据属性及处理的详细情况见表 1。

表 1 建立模型所采用的地质、地理和环境数据

数据	属性
坡度 a	0~90°
高程 b/m	0~8841
高差 c/m	0~2510
工程地质 岩土类型 d	1:岩浆岩建造,2:碎屑岩建造,3:碳酸盐岩建造,4:变质岩建造,5:土体工程地质分类
断层密度 e /km·100km <sup>-2</sup>	0~200,200~400,400~600,600~800,800~1000,>1000
地震烈度 f	1~9
土地利用类型 g	1:耕地,2:林地,3:草地,4:水域,5:城镇,6:未利用土地(裸土地和沼泽等)

a~c 图像的分辨率为 1:25 万;d,f 图像的分辨率为 1:400 万;e 图像的分辨率为 1:50 万;g 图像的分辨率为 1:10 万。

## 2 气象预警模型

### 2.1 模型的设计

地质灾害综合气象预警模型是中央气象台地质灾害气象预警业务系统的核心部分。地质灾害气象预警业务根据每天 24 小时数值降雨预报和前期降雨实况,在覆盖全国的

栅格点上进行地质灾害气象预警的等级预报。模型预报因子分成动态的降雨输入和静态背景两部分,通过 Logistic 回归,预报降雨引发地质灾害的发生概率。

拟合模型时,通过空间分析而还原的灾害发生点的历史逐日降雨构成模型的动态输入。由于目前的地质灾害气象预警业务只在我国汛期(5—9 月)运行,因此灾害记录和历史降雨也只取汛期发生的。灾害发生点的地质因子通过信息量模型集成为灾害发生的静态条件。进行预报时,在栅格点上提取各地学因子,通过信息量模型集成为预报区域的静态条件。栅格点上的实况降雨通过空间分析方法来恢复,结合栅格点上的数值降雨预报一起构成模型的动态输入。预警模型在预报区域的栅格点上预报降雨触发地质灾害的发生概率,并通过释用的方式发布预警等级。

图 1 给出了地质灾害综合气象预警模型的框架图。

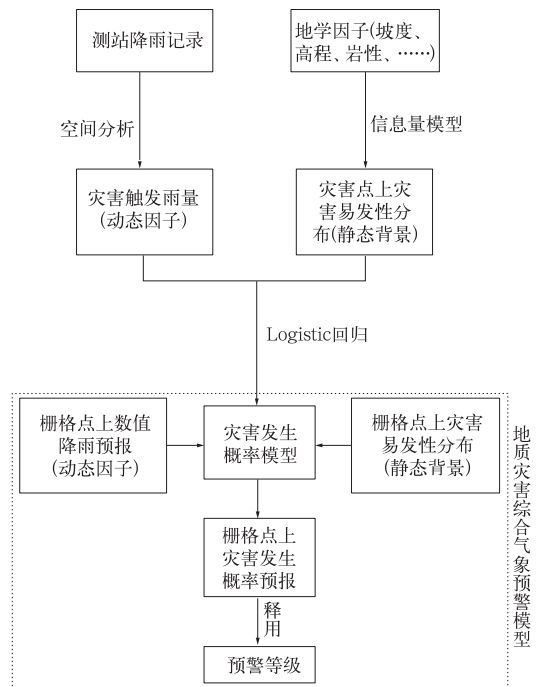


图 1 地质灾害综合气象预警模型框架图

## 2.2 模型预报因子

### 2.2.1 降雨因子

大量的研究工作表明,短时的强降雨和多日的连续降雨都会引发不同类型的地质灾害。短时强降雨可用当日 24 小时雨量来表征。对于多日的连续降雨而言,由于地表径流、水分蒸发等过程使得进入岩土体的雨水量小于观测到的日降雨量之和,因此通常用前期有效雨水量来表征对灾害发生有影响的前期降雨作用<sup>[9]</sup>。前期有效降雨的计算方法为

$$R_w = \sum_{i=1}^n k^{i-1} R_i \quad (1)$$

式中  $R_w$  为前期有效雨量,  $R_i$  为前第  $i$  天降雨量;  $k$  为衰减系数,这里取为 0.8;  $n$  为前期天数,这里取为 14。由于我国各地气候差异较大,触发灾害的降雨量也因地而异,因此雨量都以当地多年年平均雨量为基准,取为相对雨量。

此外,有时在当日降雨量和前期有效雨量都不很大的情况下,前期降雨的连续性也是灾害的触发因素。因此,前 14 天内的降雨天数也作为降雨触发条件加入到模型中。

### 2.2.2 地质灾害易发性的总信息量

采用信息量模型综合考虑与地质灾害易发性相关的地质背景因素,得到各灾害发生点的易发性评估,作为预警模型的背景条件因子。信息量模型在地质灾害易发评估中已得到广泛应用<sup>[21-23]</sup>。信息量模型从整个预报区域的角度,考虑各灾害分布单元中地质因子的具体状态对灾害发生的影响。设  $S$  为预报区域单元总数,  $S_{i,A_i}$  为第  $i$  个因子状态为  $A_i$  的单元数,  $N$  为地质灾害分布的单元数,  $N_{i,A_i}$  为第  $i$  个因子状态为  $A_i$  的灾害分布单元数,则对于每个第  $i$  个因子状态为  $A_i$  的单元的信息量,即对该单元灾害易发性的贡献为

$$I_{i,A_i} = \ln\left(\frac{N_{i,A_i}/N}{S_{i,A_i}/S}\right) \quad (i = 1, 2, \dots, n) \quad (2)$$

各因子对该单元灾害易发性的贡献之和,即该单元的总信息量为

$$I = \sum_{i=1}^n I_{i,A_i} \quad (3)$$

总信息量  $I$  越大,灾害发生的可能性越大。

对预报区域的每个单元都按式(3)计算该单元的总信息量。取灾害发生点所在单元的总信息量,作为预警模型中表征灾害易发性的静态预报因子。拟合模型时,平原、沙漠、水面等降雨引发地质灾害可能性极小或没有的地区以及缺乏数据的地区事先予以消空,以免在这些地区产生虚假的预报结果。结果表明,我国区域地质灾害易发性的总信息量变化范围为  $-13.29 \sim 6.31$ 。

## 2.3 Logistic 回归

地质灾害的发生概率模型基于广义线性模型中的 Logistic 回归来建立<sup>[24]</sup>。对于地质灾害预报而言,所要回归的随机变量为灾害发生的概率  $p_i$ ,其取值范围为  $(0, 1)$ ,而预报因子则为引发地质灾害的气象和地质环境因素。这些因子的取值范围一般在  $(-\infty, +\infty)$ 。为了能建立回归关系,需要引入连接函数,将  $p_i$  的取值范围映射到  $(-\infty, +\infty)$ 。具体形式为:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^m \beta_j x_{i,j} \quad (i = 1, \dots, n) \quad (4)$$

式(4)称为 Logistic 回归。其中,  $x_{i,j}$  ( $j = 1, \dots, m$ ) 为  $m$  个预报因子对应于第  $i$  个观测样本的取值,  $p_i$  为第  $i$  个观测样本的灾害发生概率,  $\beta_j$  ( $j = 0, \dots, m$ ) 为需要拟合的模型参数。模型的观测样本均为汛期触发或未触发地质灾害的降雨事件,模型也只在天气预报为有雨时预报地质灾害的发生概率。

## 2.4 预报结果的释用与预警发布

在网格点上进行预报时,输入表征网格

点上灾害易发性的总信息量、过去 14 天的降雨观测以及未来 24 小时降雨预报,即可得到网格点的地质灾害发生概率。相对于降雨事件总体而言,地质灾害的发生频率毕竟是很小的,加之数据统计的不完备,模型预报出来的灾害发生概率在量级上是非常小的,不能直接用来指导防灾预警。因此,预报结果应通过释用的方式来指导预警。按照中央气象台业务运行的规范,降雨引发的地质灾害分为 5 个预报警报等级<sup>[12]</sup>,分别是(1)可能性很小、(2)可能性较小、(3)可能性较大、(4)可能性大和(5)可能性很大。其中 3 级以上的预警通过新闻媒体公布。据此,根据灾害发生的观测实况,按照百分位 0.25、0.45、0.65 和 0.85 将模型拟合的全部灾害发生概率分为 5 级,3 级以上可基本覆盖实际发生的灾情。业务运行时,根据模型预报的发生概率所在的百分位区间确定其预警等级。

## 2.5 模型拟合结果及检验

采用 1990—2003 年间汛期降雨日的降雨量和灾情作为观测数据,来拟合灾害发生点上的灾害发生概率,结果见表 2。

表 2 地质灾害综合气象预警模型参数拟合结果

预报因子	拟合值	标准误差
常数	-10.088515	0.088031
当天降雨量	6.561416	0.105088
前期有效降雨量	2.228535	0.147125
前期降雨天数	0.300200	0.006149
总信息量	0.022872	0.003212

方差分析结果表明,上述因子均通过  $\alpha = 0.001$  的显著性检验。从误差分析结果看,总信息量的参数拟合误差最大,但也仅为拟合值的 1/7。这表明这些因子确实与灾害发生概率有显著关系,选为预报因子在统计意义上是合理的。

用 2004 年的历史灾情数据和降雨观测对模型的预警进行检验,结果表明 80% 以上的灾害发生均给出了 3 级以上的预警,即需要对公众发布信息。模型性能是较为理想

的。但也存在 35% 的空报率,原因在于预报因子往往具有大尺度的空间连续性,目前的预报还难以精细到地质灾害的空间尺度上。

2006 年第 4 号热带风暴碧利斯于 7 月 14 日中午 12 时 50 分在福建霞浦一带沿海登陆,登陆后减弱的低压一直维持到 18 日,造成南方出现大范围连续 3~5 天强降雨,局部地区累计降雨 400~500mm。湖南、江西、福建、广东、广西和云南多处山地灾害频发,灾情严重。

7 月 13—18 日,中央气象台与国土资源部联合发布了准确的地质灾害预报预警。图 2 显示了“碧利斯”期间的地质灾害气象等级预报与所收集到的 85 例地质灾害。其中 15、16 日发生的灾害分别有 74.9% 和 66.7% 落在 3 级(可能性较大)或 3 级以上预警区域。

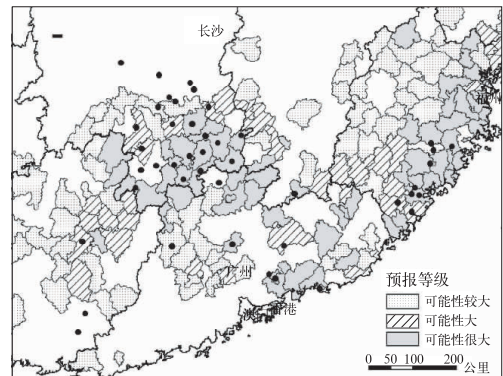


图 2 2006 年 7 月台风碧利斯期间气象灾害地质预报与实况(圆点)对比

## 3 结 论

基于 Logistic 回归,将动态的降雨条件与相对静态的地学因素结合起来,拟合降雨触发地质灾害的发生概率,建立了具有动态预报能力的区域地质灾害综合气象预警模型。统计检验以及实例验证结果都表明,该模型对降雨诱发地质灾害的发生概率具有较好的预报能力,适合于地质灾害气象预警系统的业务运行。主要结论有:

(1) 薄板样条平滑函数是对历史灾害的触发雨量进行空间分析的有效方法。该方法可考虑高程对降雨分布的影响,显著改善降雨插值效果,并能够满足预警业务自动运行的需要。

(2) 信息量模型是集成地学因子进行灾害易发性评估的有效方法,总信息量可用作预警模型的预报因子。进一步提高遥感数据的空间分辨率,改善遥感数据的提取和分类是改进信息量模型的关键。

(3) 模型预报出的灾害发生概率应以释用的方式指导防灾预警。预警等级可根据实际情况,以灾害发生概率的百分位来划分制定。随着灾情数据的不断丰富,预警等级也应及时加以调整。

(4) 由于模型中的预报因子往往具有较大尺度的空间连续性,因此模型存在空报的问题。提高地学数据的空间分辨率,以及数值降雨预报的精细化程度,将有助于从根本上降低空报率,提高地质灾害预报的准确性。

## 参考文献

- [1] 李媛,孟晖,董颖,等. 中国地质灾害类型及其特征—基于全国市县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.
- [2] 张延治. 辽东半岛山洪泥石流与诱发暴雨[J]. 气象,1991(1):43-47.
- [3] 郁淑华. 诱发泥石流灾害的四川盆地大暴雨过程分析[J]. 气象,2002(8):30-33.
- [4] 单九生,刘修奋,魏丽,等. 诱发江西滑坡的降水特征分析[J]. 气象,2004(1):13-15.
- [5] 王锡稳,陶建红,冯军,等. 陇南“5·31”特大泥石流灾害成因分析[J]. 气象,2004(10):43-46.
- [6] 郁淑华,徐会明,何光碧,等. 基于 $\eta$ 数值预报模式的四川盆地泥石流滑坡预报系统[J]. 气象,2005(6):47-50.
- [7] 魏丽,郑有飞,单九生. 暴雨型滑坡灾害预报预警方法研究评述[J]. 气象,2005(10):3-6.
- [8] 薛建军,徐晶,张芳华,等. 区域性地质灾害气象预报方法研究[J]. 气象,2005(10):24-27.
- [9] 崔鹏,杨坤,陈杰. 前期降雨对泥石流形成的贡献——以云南蒋家沟为例[J]. 中国水土保持科学,2003,1(1):11-15.
- [10] 刘传正,温铭生,唐灿. 中国地质灾害气象预警初步研究[J]. 地质通报,2004,23(4):303-309.
- [11] 姚学祥,徐晶. 跨学科的地质灾害预警工程[J]. 中国工程科学,2004,6(6):9-14.
- [12] Yao Xuexiang and Xu Jing. The forecasting and warning system of geological disasters in China[J]. Engineering Science, 2004, 2(3):30-35.
- [13] 姚学祥,徐晶,薛建军,等. 全国地质灾害气象预报预警方法研究[J]. 中国地质灾害与防治学报,2005,16(4):97-102.
- [14] Atkinson P. M. and Massari R. Generalized Linear Modelling of Susceptibility to landsliding in the central Apennines, Italy[J]. Computers and Geosciences, 1998, 24(4):373-385.
- [15] Dai F. C. and Lee C. F. Landslide Characteristics and Slope Instability Modeling Using GIS, Lantau Island, Hong Kong[J]. Geomorphology, 2002, 42: 213-228.
- [16] Ohlmacher G. C. and Davis J. C. Using Multiple Logistic Regression and GIS Technology to Predict Landslide Hazard in Northeast Kansas, USA[J]. Engineering Geology, 2003, 69:331-343.
- [17] Dai F. C. and Lee C. F. A Spatiotemporal Probabilistic Modeling of Storm-induced Shallow Landsliding Using Aerial Photographs and Logistic Regression[J]. Earth Surface Processes and Landforms, 2003, 28:527-545.
- [18] Hutchinson M. F. Interpolation of Rainfall Data with Thin Plate Smoothing Splines-Part I: Two Dimensional Smoothing of Data with Short Range Correlation[J]. Journal of Geographic Information and Decision Analysis, 1998, 2:139-151.
- [19] Hutchinson M. F. Interpolation of Rainfall Data with Thin Plate Smoothing Splines-Part II: Analysis of Topographic Dependence[J]. Journal of Geographic Information and Decision Analysis, 1998, 2: 152-167.
- [20] 张国平,徐晶,毕宝贵. 滑坡泥石流灾害与若干自然地理因子的关系分析[R]. 中国地理学会 2006 年学术年会. 兰州:中国地理学会,2006.
- [21] 柳源. 中国地质灾害(以崩、滑、流为主)危险性分析与区划[J]. 中国地质灾害与防治学报,2003,14(1): 95-99.
- [22] 吴树仁,胡道功,谭成轩,等. 长江三峡库首区滑坡灾害空间预报[J]. 中国地质灾害与防治学报,1998,9(增刊):141-147.
- [23] 朱良峰,吴信才,殷坤龙,等. 基于信息量模型的中国滑坡灾害风险区划研究[J]. 地球科学与环境学报,2004,26(3):52-56.
- [24] McCullagh P. and Nelder J. A. Generalized Linear Models[M], 2nd ed., New York: Chapman and Hall, 1989.