

基于 Fisher 判别分析法的泥石流预报模型研究

王文川¹, 吴海波¹, 赵晓慎¹, 邱林²

(1. 华北水利水电学院 水利学院, 河南 郑州 450011; 2. 华北水利水电学院 环境与市政工程学院, 河南 郑州 450011)

摘要: 针对泥石流形成原因复杂, 形成机理认识不足, 预测预报困难等特点。根据泥石流形成的直接成因降雨因素, 运用判别分析理论, 提出了泥石流预报的 Fisher 判别分析模型。该模型基本思想是将一系列与泥石流发生相关的因子数据投影到某一个方向, 使得投影后数据尽可能地分组, 借助组内方差尽量小, 组间方差尽量大的一元方差分析思想确定判别函数, 依据判别准则, 判定新样本的归属类别。该方法具有操作简单易懂及计算速度快的特点, 可及时对泥石流状况进行判别。实例研究结果表明, 提出的 Fisher 判别分析泥石流预报模型能够有效提高预报精度。

关键词: Fisher 判别分析法; 判别函数; 判别准则; 泥石流

文献标识码: B

文章编号: 1000-288X(2011)06-0104-04

中图分类号: P642.23

Forecasting Model of Debris Flow Based on Fisher Discrimination Analysis Method

WANG Wen-chuan¹, WU Hai-po¹, ZHAO Xiao-shen¹, QIU Lin²

(1. Faculty of Water Conservancy Engineering, North China Institute of Water Conservancy and

Hydroelectric Power, Zhengzhou, He'nan 450011, China; 2. Faculty of Environmental and Municipal

Engineering, North China Institute of Water Conservancy and Hydroelectric Power, Zhengzhou, He'nan 450011, China)

Abstract: Due to the complicated causes, the poor understanding to debris flow and difficulties in prediction, a Fisher discrimination analysis model for forecasting the debris flow was developed according to the theory of discrimination analysis. Its basic idea is that a series of factor data related to the debris flow occurrence was projected to some direction and the data projected can be grouped as far as possible. The discriminant function was determined according to the univariate variance analysis, which requires smaller intraclass variance and larger interclass variance. And then new samples were sorted into some class according to this criterion. The proposed model is simple and run fast, and can predict the condition of debris flow in time. The case study showed that Fisher discrimination analysis model can improve the prediction accuracy of debris flow.

Keywords: Fisher discrimination analysis method; discriminant function; criterion; debris flow

近年来,受特殊的自然地理环境、极端灾害性天气以及经济社会活动等多种因素的共同影响,各地山丘区洪水、泥石流、滑坡等灾害频发,造成的人员伤亡、财产损失、基础设施损毁及生态环境破坏十分严重,已成为我国防洪减灾工作中亟待解决的突出问题^[1]。泥石流作为一种由降雨引起的山洪灾害,具有发生突然、历时短暂、来势凶猛和破坏力强等特点^[2]。其形成原因复杂,而且目前对其形成的机理及其影响因素认识不足。但其预测预报作为一项重要减灾手段,受到了国内外学者的广泛关注,是减灾研究的热点和关键问题^[3]。国内外对泥石流预报模型的研究^[3]集中在泥石流形成背景,降雨强度与雨量关系,

泥石流形成机理及泥石流物理特性等方面。随着现代数理统计分析理论和计算机技术的快速发展, Bayes 判别分析法、逐步判别分析法、遥感技术等方法和技术已应用于地学领域的研究^[1-5]。本研究根据判别分析理论,从暴发是由降雨直接导致的成因着手,提出泥石流预报的 Fisher 判别分析模型,实例研究结果表明, Fisher 判别分析法预报效果更佳。

1 Fisher 判别分析预报模型^[6-14]

1.1 基本原理

Fisher 判别分析法是 Fisher 提出的最优判别准则以及判别函数的求解法。其基本思想是将 k 组 m

维样本数据投影到某一个方向,使得投影后各组之间尽可能地分开,借助组内方差尽量小,组间方差尽量大的一元方差分析思想确定判别函数,依据一定的判别准则,判别新样本属于哪个组。

1.2 判别函数的建立

设有 k 个总体 $G_t (t=1, \dots, k)$, 从 G_t 中分别取 n_t

个 m 维的样本, 有 $n = \sum_{t=1}^k n_t$:

$$\begin{cases} X_i^{(t)} = [x_{i1}^{(t)}, x_{i2}^{(t)}, \dots, x_{im}^{(t)}]^T \\ (i=1, 2, \dots, n_t; t=1, 2, \dots, k) \end{cases} \quad (1)$$

将 $X_i^{(t)}$ 的数据投影到某一 m 维常数向量 C 上, 得到投影点之间的一元线性组合为:

$$y(X) = C^T X \quad (2)$$

采用 $\bar{y}^{(t)}$ 表示总体样本 G_t 的均值, \bar{y} 表示 k 组总均值。

$$\begin{cases} \bar{y}^{(t)} = C^T \bar{X}^{(t)} \\ \bar{y} = C^T \bar{X} \end{cases} \quad (t=1, 2, \dots, m) \quad (3)$$

式中: $\bar{X}^{(t)} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_i^{(t)}$, $\bar{X} = \frac{1}{n} \sum_{t=1}^k \sum_{i=1}^{n_t} X_i^{(t)}$, 其余符号意义同前。

于是得到组内差 E 和组间差 B 。

$$\begin{cases} E = \sum_{t=1}^k \sum_{i=1}^{n_t} [C^T X_i^{(t)} - C^T \bar{X}^{(t)}]^2 = C^T W C \\ B = \sum_{t=1}^k n_t [C^T \bar{X}^{(t)} - C^T \bar{X}]^2 = C^T U C \end{cases} \quad (4)$$

式中: $W = \sum_{t=1}^k \sum_{i=1}^{n_t} [X_i^{(t)} - \bar{X}^{(t)}][X_i^{(t)} - \bar{X}^{(t)}]^T$;

$$U = \sum_{t=1}^k n_t [\bar{X}^{(t)} - \bar{X}][\bar{X}^{(t)} - \bar{X}]^T.$$

E 和 B 分别具有自由度为 $n-k$ 和 $k-1$, 则一元方差分析中检验统计量 F 为:

$$F = \frac{B/(k-1)}{E/(n-k)} = \frac{C^T U C / (k-1)}{C^T W C / (n-k)} \quad (5)$$

公式(5)中 F 值越大, 表明总体样本 G_t 之间的均值有显著差异。 F 值越大, 也使得对系数 C 的目标函数值 $\phi(C)$ 达到极大值:

$$\varphi(C) = \frac{B}{E} = \frac{C^T U C}{C^T W C} \quad (6)$$

由特征值极值性质可知, 公式(6)中 $\phi(C)$ 的极大值转化为求 $W^{-1} \cdot U$ 的最大特征值和相应特征向量问题, 其极大值求解方程:

$$|U - \lambda W| = 0 \quad (7)$$

公式(7)的最大特征根, 记 $W^{-1} \cdot U$ 的非零特征根为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$, ($r \leq m$), 取出最大特征值 λ_1 及其对应的特征向量 \hat{C}_1 , 则得到了判别函数:

$$y(X) = \hat{C}_1^T X \quad (8)$$

1.3 判别准则

建立的判别函数需要通过一定的判别准则来对检验对象进行归类。采用 Mahalanobis 距离法进行判别。对任一个需要判别归属的样品 \tilde{X} , 与总体 G_t 的 Mahalanobis 距离为:

$$d_t^2 = d_t^2(\tilde{X}, G_t) = [\hat{C}_1^T \tilde{X} - \hat{C}_1^T \bar{X}^{(t)}]^T [\hat{C}_1^T \frac{W^{(t)}}{n_t - 1} \hat{C}_1^T]^{-1} [\hat{C}_1^T \tilde{X} - \hat{C}_1^T \bar{X}^{(t)}] \quad (t=1, 2, \dots, k) \quad (9)$$

比较样品 \tilde{X} 与 G_t 的距离 d_t^2 的大小, 依据判别规则 $d_t^2 = \min_{1 \leq j \leq k} \{d_j^2\}$, 则可判断样品 $\tilde{X} \in G_t$ 。

1.4 Fisher 判别分析预报模型检验

设 $G_t \sim N_m(\bar{y}^{(t)}, \frac{W}{n-k})$; $t=1, 2, \dots, k$ 。当假设检验 $H_{01}: \bar{y}^{(1)} = \bar{y}^{(2)} = \dots = \bar{y}^{(k)}$ 被接受, 说明建立判别分析预报模型没有必要; 若 H_{01} 被拒绝, 就需要检验每两个总体间差异的显著性, 即 $H_{02}: \bar{y}^{(i)} = \bar{y}^{(j)}$; $i, j = 1, 2, \dots, k$; $i \neq j$, 其统计量 F_{ij} 为

$$F_{ij} = \frac{(n-m-k-1)n_i n_j}{m(n-k)(n_i + n_j)} d_{ij}^2 \sim F(m, n-m-k-1) \quad (10)$$

式中: $d_{ij}^2 = [\bar{X}^{(i)} - \bar{X}^{(j)}]^T C (C^T \frac{W}{n-k} C)^{-1} C^T [\bar{X}^{(i)} - \bar{X}^{(j)}]$; 其它符号同前。

经检验后, 若某两个总体样本差异不显著, 则将检验的两个总体合并成一个总体, 再与剩下的总体样本重新建立判别函数。

2 Fisher 判别分析预报实现

根据 Fisher 判别分析理论, 对泥石流预报发生与否的相关因子, 建立 Fisher 判别函数, 及对新样本数据进行判别是否发生泥石流, 即可知 $k=2$, 其相关步骤为:

(1) 根据样本得到各组的平均值 $\bar{X}^{(t)}$ 和协方差矩阵 S_t 。

$$\begin{cases} \bar{X}^{(t)} = [\frac{1}{n_t} \sum_{i=1}^{n_t} x_{i1}^{(t)}, \dots, \frac{1}{n_t} \sum_{i=1}^{n_t} x_{im}^{(t)}] \\ S_t = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} [X_i^{(t)} - \bar{X}^{(t)}][X_i^{(t)} - \bar{X}^{(t)}]^T \end{cases} \quad (t=1, 2) \quad (11)$$

(2) 计算总样本的协方差 S 和判别函数的待估参数 C^T , 总样本的协方差 S 为各组样本协方差 S_t 之和的无偏估计。则公式(2)中的判别函数可表示为^[11-14]:

$$y(X) = C^T X = S^{-1} [\bar{X}^{(1)} - \bar{X}^{(2)}]^T X \quad (12)$$

根据公式(12)可知, 泥石流判别函数的待估参数 $C^T = S^{-1} [\bar{X}^{(1)} - \bar{X}^{(2)}]^T$ 。

(3) 计算判别临界值 y_0 。根据样本数据计算泥

石流发生 G_1 组和未发生 G_2 组的代表判别函数值 $\bar{y}^{(1)}$ 和 $\bar{y}^{(2)}$:

$$\begin{cases} \bar{y}^{(1)} = \sum_{i=1}^m C_i^T \bar{X}_i^{(1)} \\ \bar{y}^{(2)} = \sum_{i=1}^m C_i^T \bar{X}_i^{(2)} \end{cases} \quad (13)$$

则临界值 y_0 为

$$y_0 = \frac{n_1 \bar{y}^{(1)} + n_2 \bar{y}^{(2)}}{n_1 + n_2} \quad (14)$$

(4) 对新样本做出判别。任意的一个 m 维的新样本 \hat{X} 代入判别函数中所得的值记为 \hat{y} , 则出现两种情况:

$$\text{当 } \bar{y}^{(1)} > y_0 \text{ 时, } \begin{cases} \hat{y} > y_0 & (\hat{X} \in G_1) \\ \hat{y} < y_0 & (\hat{X} \in G_2) \\ \hat{y} = y_0 & (\text{待判}) \end{cases} \quad (15)$$

$$\text{当 } \bar{y}^{(2)} > y_0 \text{ 时, } \begin{cases} \hat{y} > y_0 & (\hat{X} \in G_2) \\ \hat{y} < y_0 & (\hat{X} \in G_1) \\ \hat{y} = y_0 & (\text{待判}) \end{cases} \quad (16)$$

(5) 判别函数的检验。由公式(9)—(10)对判别函数进行 F 检验, 对于给定的信度 α , 从 F 分布表查出 $F(m, n_1 + n_2 - m - k - 1)$ 的值, 若 $F > F_\alpha$, 在信度 α 检验水平下, 判别函数有效。

3 实例应用

3.1 数据源及泥石流影响因子选取

泥石流的发生, 不仅与次降雨作用有关, 其前期降雨在一定程度上对泥石流的发生产生间接地影响, 选择当日降雨量和前期降雨的实效雨量作为泥石流发生的影响因子^[4,15]。

以文献[4]和文献[15]中对密云县境内所采集的泥石流与降雨数据进行分析(表 1)。以 34 个样本数据为研究对象, 建立 Fisher 判别分析预报模型, 再用样本数据对模型本身进行检验。

3.2 结果分析

按照 Fisher 判别分析法流程图经运算可得密云县泥石流 Fisher 判别分析预报模型为:

$$y(X) = 0.0236x_1 + 0.054x_2 \quad (17)$$

式中: x_1 ——前 15 d 实效雨量; x_2 ——当日降雨量。

根据公式(11)计算得到 $d_{12}^2 = 4.949$, $F = 7.808 > F_{0.01}(2, 29) = 5.42$, 故判别函数在信度 $\alpha = 0.01$ 条件下有效, 可以对其样本数据进行预报。

由公式(16)可得其判别临界值为 7.916。根据公式(19), 代入 x_1 和 x_2 , 如果其值大于 7.916 则发生泥石流, 如果小于 7.916 则不发生泥石流, 得判别结果详见表 1。

从表 1 中模型预报结果可知, 前 16 个样本数据

出现 4 个预报错误, 预报正确率为 75%(表 2)。后 18 个样本数据出现 1 个预报错误, 预报正确率 94.4%(表 2)。密云县泥石流 Fisher 判别分析预测模型整体预报正确率为 84.7%, 与文献[4]Bayes 判别分析预报模型进行比较, 减少了 1 次预报错误, 预报正确率提高了 3.7%。

表 1 降雨因子对泥石流影响分析

样本序号	是否发生泥石流	前 15 d 实效雨量/mm	当日降雨量/mm	判别
1	是	21.53	108.3	否
2	是	19.52	123.8	否
3	是	80.80	153.3	是
4	是	3.80	179.0	是
5	是	80.37	158.3	是
6	是	33.76	179.0	是
7	是	25.98	226.3	是
8	是	32.14	236.5	是
9	是	49.04	116.5	否
10	是	34.16	130.0	是
11	是	62.91	193.5	是
12	是	30.89	237.0	是
13	是	33.25	128.8	否
14	是	88.03	277.2	是
15	是	93.57	169.7	是
16	是	68.51	133.9	是
17	否	17.10	74.1	否
18	否	60.72	114.0	否
19	否	11.20	119.8	否
20	否	74.21	86.8	否
21	否	5.64	86.8	否
22	否	10.22	105.4	否
23	否	14.69	105.5	否
24	否	28.78	141.7	是
25	否	16.92	79.5	否
26	否	16.18	73.4	否
27	否	16.77	71.0	否
28	否	29.17	49.0	否
29	否	39.64	88.8	否
30	否	5.10	82.6	否
31	否	3.12	82.8	否
32	否	20.59	78.6	否
33	否	58.15	96.4	否
34	否	24.38	104.9	否

表 2 Fisher 判别分析预报结果

项目	样本检验	发生泥石流现象	未发生泥石流现象
样本预报	发生泥石流现象	12	4
	未发生泥石流现象	1	17
预报率	发生泥石流现象	75%	25%
	未发生泥石流现象	5.6%	94.4%

4 结语

作为世界上受泥石流灾害威胁最为严重的国家之一,研究其预报模型,提高预报精度,对提高我国山洪灾害防御能力具有重要作用。本研究通过 Fisher 判别分析法对泥石流产生的当日降雨量和 15 d 实效雨量进行了分析,提出了 Fisher 判别分析泥石流预报模型。该模型具有操作简单易懂和计算速度快等特点,并将其运用于密云县泥石流的预报中,结果正确率高达 84.7%,值得在泥石流预报中推广应用。

[参 考 文 献]

- [1] 陈雷. 陈雷在山洪灾害防治县级非工程措施建设会议讲话[EB/OL]. (2010-04-24)[2010-09-09]. http://www.gov.cn/gzdt/2010-11/25/content_1753454.htm.
- [2] 师哲,张平仓,舒平安. 泥石流监测预报预警系统研究[J]. 长江科学院院报,2010,27(11):115-119.
- [3] 崔鹏,高克昌,韦方强. 泥石流预测预报研究进展[J]. 中国科学院院刊,2005,20(5):363-369.
- [4] 郑国强,张洪江,刘涛,等. 基于 Bayes 判别分析法的密云县山洪泥石流预报模型[J]. 水土保持通报,2009,29(1):83-87.
- [5] 孟凡奇,李广杰,李明,等. 逐步判别分析法在筛选泥石流评价因子中的应用[J]. 岩土力学,2010,31(9):2925-2929.
- [6] 白云飞,江露英. 边坡稳定性预测的 Fisher 判别分析模型[J]. 有色金属:矿山部分,2010,62(2):49-52.
- [7] 李秀珍,王成华,宋刚. 基于 Fisher 判别分析法的潜在滑坡判识模型及其应用[J]. 中国地质灾害与防治学报,2009,20(4):23-26.
- [8] 董陇军,李夕兵,白云飞. 急倾斜煤层顶煤可放性分类预测的 Fisher 判别分析模型及应用[J]. 煤炭学报,2009,34(1):58-63.
- [9] 赵选民,徐伟,师义民,等. 数理统计[M]. 北京:科学出版社,1999:303-308.
- [10] 朱永生. 实验数据多元统计分析[M]. 北京:科学出版社,2009:48-54.
- [11] 袁志发,宋世德. 多元统计分析[M]. 北京:科学出版社,2009:255-266.
- [12] 张尧庭,方开泰. 多元统计分析引论[M]. 北京:科学出版社,1982:172-179.
- [13] 王学民. 应用多元分析[M]. 上海:上海财经大学出版社,2009:178-190.
- [14] 暴奉贤,陈宏立. 经济预测与决策方法[M]. 广州:暨南大学出版社,2007:251-255.
- [15] 王礼先,于志民. 山洪及泥石流灾害预报[M]. 北京:中国林业出版社,2001:121-175.
- [5] 张金池,李海东,林杰,等. 基于小流域尺度的土壤可蚀性 K 值空间变异[J]. 生态学报,2008,28(5):2199-2206.
- [6] Wischmeier W H. Cropping-management factor evaluations for a universal soil-loss equation[J]. Soil Sci. Soc. Am. Proc., 1960, 24:322-326.
- [7] 刘宝元,张科利,谢云. 土壤侵蚀模型[M]. 北京:中国科学技术出版社,2001.
- [8] Wischmeier W H. Use and misuse of the universal soil loss equation[J]. J. Soil and Water Conserv., 1976,31:5-9.
- [9] Williams J R, Jones C A, Dyke P T. The EPIC model and its application[C]//Proc. ICRISAT-IBSNAT-SYSS Symp. on Minimum Data Sets for Agro-technology Transfer, 1983:111-121.
- [10] Williams J R, Sharply A N. EPIC erosion productivity impact calculator I: Model documentation [J]. US Department of Agriculture Technical Bulletin, 1990, 1768:105-124.

(上接第 98 页)