

# 模糊聚类法在大坝监测资料分析中的应用

廖 铨<sup>1</sup>, 蔡德所<sup>1</sup>, 李 苗<sup>2</sup>, 张吉燕<sup>1</sup>, 王 立<sup>3</sup>

(1. 三峡大学 水利与环境学院, 湖北 宜昌 443002; 2. 华北水利水电大学 资源与环境学院, 河南 郑州 450011; 3. 三峡大学 土木与建筑学院, 湖北 宜昌 443002)

**摘要:** 为了解模糊聚类法在大坝监测资料分析中应用的适宜性, 利用该法对水布垭面板堆石坝 2013 年面板挠度变形监测数据开展了研究。首先将监测管道划分为若干监测点, 运用模糊聚类分析法对监测点进行分类, 并采用 F 统计量评价聚类效果以确定最佳分类, 对分类结果进行分析比较, 可确定关键的面板变形监控点。随后建立面板挠度变形统计模型, 考虑上游水深、温度、时效等因素, 利用多元逐步回归分析法对关键点监测数据进行预报拟合。结果表明, 利用模糊聚类分析法处理大坝监测数据具有便捷、高效的优点, 模型预报结果拟合度较高。

**关 键 词:** 面板变形; 模糊聚类; 统计模型; 多元逐步回归

中图法分类号: TV698

文献标志码: A

DOI:10.16232/j.cnki.1001-4179.2015.13.022

大坝安全监测的主要目的是掌握大坝的运行性状及各监测量的变化趋势。随着现代监测技术的快速发展, 大坝监测测点更加全面, 监测数据量巨大, 若对监测数据逐一处理分析, 必定需要大量人力和时间, 不利于及时掌握大坝运行状况。为提高监测资料分析效率, 诸多学者将模糊数学、时间序列、神经网络、灰色理论、小波分析、卡曼滤波等理论和方法引入监测数据分析中, 取得了显著的效果<sup>[1]</sup>。为了更加充分和及时地利用丰富的监测信息, 本文在模糊数学的基础上, 尝试利用模糊聚类分析法, 对水布垭大坝面板变形监测数据进行模糊聚类分析, 然后通过最佳分类确定关键点。对关键点进行详细的分析不仅能减少工作量, 还可以及时掌握大坝的真实运行状况<sup>[2]</sup>。

## 1 模糊聚类分析基本原理

假设论域  $U = \{x_1, x_2, \dots, x_n\}$  为被研究对象, 每个对象又由  $m$  个不同的值来刻画, 即

$$x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}, i = 1, 2, \dots, n \quad (1)$$

用  $r_{ij} = R(x_i, x_j)$  表示  $x_i$  与  $x_j$  之间的相似程度, 确

定模糊相似矩阵主要借助传统聚类分析法(相似系数法、距离法、主观评分法)。对一个被研究对象集  $U = \{x_1, x_2, \dots, x_n\}$ , 其中模糊相似矩阵可表示为

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1j} \\ r_{21} & r_{22} & \cdots & r_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nj} \end{bmatrix} \quad (2)$$

给定不同的阈值  $\lambda$ , 如果  $r_{ij} \geq \lambda$ , 则  $x_i$  与  $x_j$  被聚为一类。但是, 若以此为分类标准, 则有可能导致同一类对象被划分到不同的类别, 为了避免这种情况影响分类效果, 需将模糊相似矩阵转换成模糊等价矩阵(具有传递性)。一般采用实用简捷的二次法求传递包闭  $t(R)$ , 即以模糊相似矩阵  $R$  为基础, 依次求二次方, 当第一次出现  $R^k \cdot R^k = R^k$  时,  $R^k$  即为传递包闭  $t(R)$ 。然后, 再给定不同的阈值  $\lambda$ , 即可得到相应的动态聚类结果<sup>[3]</sup>。

## 2 水布垭大坝监测资料

水布垭大坝是目前世界上已建和在建的最高混凝

收稿日期: 2015-02-03

作者简介: 廖 铨, 男, 硕士研究生, 研究方向为水工建筑物安全监测。E-mail: chengliaomiao@163.com

通讯作者: 蔡德所, 男, 教授, 博士生导师, 主要从事水利工程安全监测的光纤传感技术研究及工作。E-mail: caidesuo@vip.

163.com

土面板堆石坝<sup>[4]</sup>。面板作为最主要的防渗体系,对其变形展开监测显得尤为重要,水布垭面板堆石坝同时应用电解液倾斜仪法和光纤陀螺仪(FOG)法两种方法监测面板变形,其中,FOG 监测管道铺设在面板上最长的 0+212 断面,监测仪器布置见图 1。FOG 是一种基于 sagnac 效应的敏感角速度传感器<sup>[5]</sup>,将 FOG 安装在运载小车中,小车在监测管道中运行,管道发生变形时,FOG 能测量出角速度,对角速度进行二次积分得出相应的线位移,最终得到 FOG 运行轨迹曲线,即大坝面板挠度曲线<sup>[6-7]</sup>。

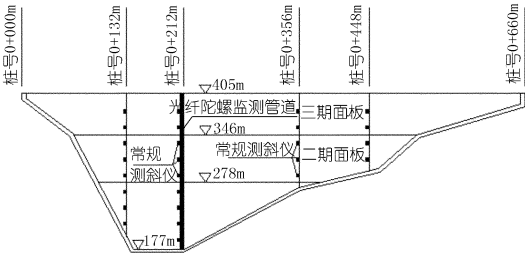


图 1 水布垭面板堆石坝面板挠度变形监测仪器布置

光纤陀螺监测具有精度高、测点连续等优点,但是其获取的数据量庞大,增加了相关研究人员的工作量,不利于及时掌握大坝的运行状态。本文主要针对类似光纤陀螺监测数据庞大的监测类别,首先运用模糊聚类分析法对监测点进行分类,采用 F 统计量评价聚类效果确定最优分类,然后对分类结果进行分析比较,确定关键的面板变形监控点。

3 建立模糊相似矩阵

将  $n$  个监测点定义为论域  $X = \{x_1, x_2, \dots, x_n\}$ , 每个监测点有 12 期不同的监测数据  $x_i = \{x_{i1}, x_{i2}, \dots, x_{i12}\}$ 。  $r_{ij} = R(x_i, x_j)$  表示  $x_i$  与  $x_j$  两监测点之间的相似程度。对面板变形监测点来说,相似程度表示两测点测量值的逼近程度和挠度变形趋势的一致性。计算  $r_{ij}$  主要是通过传统的聚类分析方法,如相似系数法、距离法、主观评分法。本文根据监测点的具体情况,选用相似系数法中的相关系数法来确定相似关系  $r_{ij}$ 。

$$r_{ij} = \frac{\sum_{k=1}^m |x_{ik} - \bar{x}_i| |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \tag{3}$$

式中,  $m$  为期数,  $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$ ,  $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$ 。

根据公式(3)计算水布垭面板坝 2013 年各监测点的相关系数,构成模糊相似矩阵  $R$ 。

$$R = \begin{bmatrix} 1 & 0.50 & 0.41 & 0.51 & 0.58 & 0.54 & 0.70 \\ 0.50 & 1 & 0.83 & 0.97 & 0.85 & 0.91 & 0.70 \\ 0.41 & 0.83 & 1 & 0.81 & 0.71 & 0.76 & 0.59 \\ 0.51 & 0.97 & 0.81 & 1 & 0.87 & 0.94 & 0.73 \\ 0.58 & 0.85 & 0.71 & 0.87 & 1 & 0.93 & 0.83 \\ 0.54 & 0.91 & 0.76 & 0.94 & 0.93 & 1 & 0.77 \\ 0.70 & 0.70 & 0.59 & 0.73 & 0.83 & 0.77 & 1 \end{bmatrix} \tag{4}$$

4 聚类分析

模糊相似矩阵  $R$  只满足自反性和对称性,不具备传递性,若用模糊相似矩阵做聚类分析,会导致同一监测点被划分到不同类,这样结果显然不准确。所以要建立模糊等价关系,需将模糊相似矩阵  $R$  转换成模糊等价矩阵  $R^*$ ,使模糊矩阵  $R^*$  满足自反性和对称性的同时满足传递性,再进行动态聚类分析,能有效避免同一监测点被划分到不同类。采用二次方法求模糊相似矩阵  $R$  的传递闭包  $t(R)$ ,依次求  $R^2, R^4, \dots, R^{2^i}$ ,当第一次出现  $R^k \cdot R^k = R^k$  时,  $R^k$  为传递闭包  $t(R)$ ,  $t(R)$  即为所求的模糊等价矩阵  $R^*$ 。最终求得模糊等价矩阵  $t(R) = R^4 = R^*$ 。

$$R^* = \begin{bmatrix} 1 & 0.70 & 0.70 & 0.70 & 0.70 & 0.70 & 0.70 \\ 0.70 & 1 & 0.83 & 0.97 & 0.93 & 0.94 & 0.83 \\ 0.70 & 0.83 & 1 & 0.83 & 0.83 & 0.83 & 0.83 \\ 0.70 & 0.97 & 0.83 & 1 & 0.93 & 0.94 & 0.83 \\ 0.70 & 0.93 & 0.83 & 0.93 & 1 & 0.93 & 0.83 \\ 0.70 & 0.94 & 0.83 & 0.94 & 0.93 & 1 & 0.83 \\ 0.70 & 0.83 & 0.83 & 0.83 & 0.83 & 0.83 & 1 \end{bmatrix} \tag{5}$$

阈值  $\lambda \in [0, 1]$ , 由大到小依次取模糊等价矩阵  $R^*$  中的每个元素,  $\lambda$  截集  $R_\lambda^* = [r_{ij}^{(\lambda)}]$ 。

$$r_{ij}^{(\lambda)} = \begin{cases} 1, & r_{ij} \geq \lambda \\ 0, & r_{ij} < \lambda \end{cases} \tag{6}$$

由此可求得各监测点的动态聚类,可有效避免同一测点被聚类到不同类别中。水布垭大坝 2013 年面板变形监测点动态模糊聚类结果见表 1。

表 1 2013 年水布垭大坝面板变形监测点动态模糊聚类结果

阈值 $\lambda$	分类数	所属类别						
		S01	S02	S03	S04	S05	S06	S07
1	7	1	2	3	4	5	6	7
0.97	6	1	2	3	2	4	5	6
0.94	5	1	2	3	2	4	2	5
0.93	4	1	2	3	2	2	2	4
0.83	2	1	2	2	2	2	2	2
0.7	1	1	1	1	1	1	1	1

5 确定最佳分类

选定不同的阈值  $\lambda \in [0,1]$  ,通过模糊聚类分析,可了解不同的分类情况,从而获得一个动态的聚类效果,能够非常形象直观地了解总体样本的分类情况。但是,为了确定最佳阈值  $\lambda$  ,需对聚类效果进行综合评价分析。选用 F 统计量来确定最佳阈值  $\lambda$  ,进而得到最佳分类<sup>[8]</sup>。所有监测点记为  $X = \{x_1, x_2, \cdots, x_n\}$  样本空间,每个监测点  $x_i$  有  $m$  期不同的监测值,  $x_i = \{x_{i1}, x_{i2}, \cdots, x_{im}\}$  ,  $(i = 1, 2, \cdots, 7)$  ,  $x_{ij}$  表示第  $i$  个监测点第  $j$  期的变形监测值,原始监测数据矩阵记为  $X_{n \times m}$  ,  $\overrightarrow{x} = (\overline{x_1}, \overline{x_2}, \cdots, \overline{x_m})$  为总体样本中心向量,其中  $\overline{x_k} = \frac{1}{n} \sum_{i=1}^n x_{ik} (k = 1, 2, \cdots, 12)$  。

设阈值  $\lambda$  对应的分类数为  $v$  ,第  $r$  类的监测点数为  $n_r$  ,相应的监测点记为  $\overrightarrow{x_1^{(r)}}, \overrightarrow{x_2^{(r)}}, \cdots, \overrightarrow{x_{n_r}^{(r)}}$  ,第  $r$  类监测点的聚类中心向量为  $\overrightarrow{x^{(r)}} = (\overline{x_1^{(r)}}, \overline{x_2^{(r)}}, \cdots, \overline{x_m^{(r)}})$  ,其中  $\overline{x_k^{(r)}} = \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ik}^{(r)} (k = 1, 2, \cdots, m)$  ,表示第  $k$  期变形的平均值。依次求第  $r$  类的聚类中心向量到总体样本中心向量的距离  $d_r$  ,以及第  $r$  类中第  $i$  个样本  $\overrightarrow{x_i^{(r)}}$  到其中心向量  $\overrightarrow{x^{(r)}}$  的距离  $d_r^i$  ,其中  $d_r = \| \overrightarrow{x^{(r)}} - \overrightarrow{x} \|$  ,  $d_r^i = \| \overrightarrow{x_i^{(r)}} - \overrightarrow{x^{(r)}} \|$  。 F 统计量可表示为

$$F = \frac{\sum_{r=1}^v n_r \| \overrightarrow{x^{(r)}} - \overrightarrow{x} \|^2 / (v - 1)}{\sum_{r=1}^v \sum_{i=1}^{n_r} \| \overrightarrow{x_i^{(r)}} - \overrightarrow{x^{(r)}} \|^2 / (n - v)}$$

(7)

F 统计量服从自由度为  $v - 1, n - v$  的 F 分布,分母表示各类内部的距离,分子表示类与类之间的距离。所以 F 越大说明类与类之间的距离越大,各类内部的距离越小,相应分类也就更合理。若  $F > F_{0.05}(v - 1, n - v)$  ,说明类与类之间的差异非常显著,分类效果更好。如果在分类中,  $F > F_{0.05}(v - 1, n - v)$  的 F 值有多个,则通过比较  $\frac{[F - F_{0.05}(v - 1, n - v)]}{F_{0.05}}$  来选择最大的 F 值,即为最佳分类。利用原始监测数据  $X_{n \times m}$  及不同阈值  $\lambda$  计算的计量 F 值结果见表 2。

从表 2 的统计结果可知,当  $\lambda = 0.93$  和  $\lambda = 0.94$  时,  $F > F_{0.05}$  ;但是  $\lambda = 0.93$  时,  $\frac{(F - F_{0.05})}{F_{0.05}} = 0.8 > 0.24$  ,所以取  $\lambda = 0.93$  ,共分为 4 类,此分类效果最佳。最佳分类表示在此 4 类监测数据中,各类别之间的测值差别较大或测点之间相距较远,而同类中的测

值较接近。结合原始数据分析, S01 和 S07 挠度变形相对较小,但是 S01 处于面板的底部, S07 处于面板顶部。 S03 位于运行管道 150 m 处(坝高 1/3 附近)面板挠度变形最大位置。根据原始数据资料分析,水布垭面板堆石坝面板挠度曲线如图 2 所示,呈“W”双谷形<sup>[9]</sup> ,若在“W”双谷曲线上作一条 650 mm 的水平线,可发现 S02、S04、S05、S06 处于 4 个交点附近,即面板挠度变形曲线方程的 4 个不同解。因此,可判断 S03 处于面板挠度变形最大位置,为关键监测点,面板易产生破坏,需重点考虑对其进行预报拟合。 S02、S04、S05、S06 处于 4 个交点附近,选择一个典型监测点分析即可反映其它 3 个点的情况,本文选择 S05 进行预报拟合。

表 2 不同阈值  $\lambda$  对应统计量 F 值

阈值 $\lambda$	分类数	F	$F_{0.05}$	$(F - F_{0.05}) / F_{0.05}$
1	7	-	-	-
0.97	6	64.8	230	-
0.94	5	23.8	19.2	0.24
0.93	4	16.67	9.28	0.80
0.83	2	5.28	6.61	-
0.7	1	-	-	-

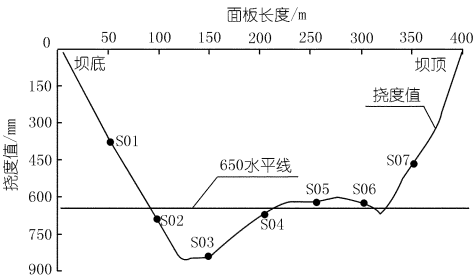


图 2 2013 年水布垭面板堆石坝面板平均相对挠度

6 关键点逐步回归分析

在大坝实际运行中,影响面板挠度变形的因变量因素众多,利用多元逐步回归分析可以很好地解决各因素之间的相关关系。近年来,在大坝安全监测资料分析中引入多元逐步回归分析方法,取得了非常显著的效果。多元逐步回归分析的基本思想是一个有进有出的过程。一方面是,开始选取一个预报因子,然后按预报因子对因变量的作用显著程度大小,依次将其引入到回归方程中;另一方面,对引入的每个预报因子进行检验,若原先引入的因子由于后面因子的引入而变得不再显著时,就要将先引入的因子剔除。因此,逐步回归是一个有因子引入也有因子剔除的进出过程,对每个过程都必须进行 F 检验,以保证每次引入新的因子以前,回归方程中只含有显著因子。重复上述步骤,直到回归方程中引入的因子不再显著,同时也没有不

显著的因子从回归方程中剔除为止,最终得到回归方程<sup>[10]</sup>。

对面板挠度变形最大处 S03 和典型监测点 S05 的数据进行逐步回归分析,考虑上游水深、温度、时效 3 个因素进行预报拟合。建立面板挠度变形统计模型

$$D = a_0 + a_1H + a_2H^2 + a_3H^3 + bT + c\ln\theta \quad (8)$$

式中,  $H$  为上游水深;  $T$  为监测当天温度;  $\theta$  为时间, d。

借助 SPSS 软件对 S03、S05 两点监测数据进行多元逐步回归分析,计算得到对应的回归方程为

$$D_{S03} = -507.55 + 9.83H - 7.627 \times 10^{-5}H^3 - 0.1427T + 0.768\ln\theta \quad (9)$$

$$D_{S05} = -351.17 + 6.81H - 5.284 \times 10^{-5}H^3 - 0.098T + 0.533\ln\theta \quad (10)$$

经逐步回归分析,求得 S03 点监测数据预测值与实测值的复相关系  $R$  为 0.91,标准偏差为 2.05,  $F$  检验值为 26.38。S05 点监测数据预测值与实测值复相关系  $R$  为 0.88,标准偏差为 2.12,  $F$  检验值为 21.62。说明两监测点的预报拟合程度均较好,两者挠度拟合曲线如图 3、4 所示。

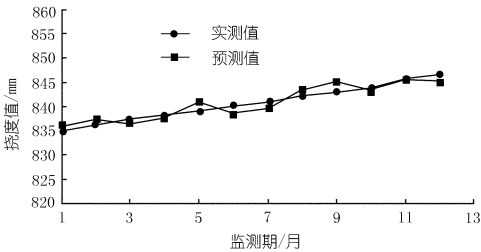


图 3 S03 监测点拟合曲线

7 结 语

通过分析水布垭面板堆石坝面板挠度变形资料,发现模糊聚类的差异性符合面板挠度变形一般规律,将模糊聚类分析法应用于大坝监测资料分析是可行的。采用多元逐步回归分析对关键监测点数据进行预

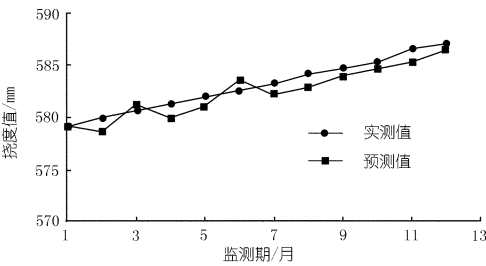


图 4 S05 监测点拟合曲线

报模拟,其拟合程度较高,可有效地预报面板挠度变形。将模糊聚类分析和多元逐步回归分析结合起来处理大坝监测资料,能够提高工作效率,随时掌握大坝的运行状况。

参考文献:

[1] 张平. 狮子滩大坝监测资料分析与参数反演[D]. 成都: 四川大学, 2004.

[2] 周洪波, 薛桂玉, 付成华. 应用 Matlab 聚类方法分析大坝监测资料[J]. 人民长江, 2003, (11): 46-48.

[3] 谢季坚, 刘承平. 模糊数学方法及其应用[M]. 武汉: 华中科技大学出版社, 2006.

[4] 熊泽斌, 杨启贵, 张运建. 水布垭高面板堆石坝设计[J]. 人民长江, 2007, (7): 19-21.

[5] 张桂才. 光纤陀螺原理与技术[M]. 北京: 国防工业出版社, 2008.

[6] 蔡德所, 李昌彩, 卫炎, 等. 光纤陀螺用于混凝土面板堆石坝面板挠度测量的试验研究[J]. 水力发电学报, 2006, (8): 79-82.

[7] 孙东亚, 姜德生, 王立新, 等. 光纤陀螺用于混凝土面板堆石坝面板挠度测量的试验研究[J]. 水利学报, 2000, (4): 14-18.

[8] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版社, 2004.

[9] 黎佛林, 蔡德所, 秦鹏, 等. 水布垭水电站面板挠度监测方法比较[J]. 水力发电, 2013, (1): 82-84, 88.

[10] 吴中如. 水工建筑物安全监控理论及其应用[M]. 北京: 高等教育出版社, 2003.

(编辑: 胡旭东)

(下转第 99 页)



帮助移民脱贫致富<sup>[4]</sup>。同时加强移民思想引导和心理疏导,增强移民的机遇意识、自强意识和开拓开放意识,努力营造文明和谐的社会环境,以助力解决移民矛盾。

参考文献:

[1] 黎爱华,张鹤,张春艳. 水利水电工程移民稳定问题对策研究[J]. 人民长江,2010,(23).

[2] 湖北省移民局. 湖北省 2012 年度大中型水库移民后期扶持政策实施情况监测评估报告[R]. 湖北省移民局,2012.

[3] 陈绍军,程军,史明宇. 水库移民社会风险研究现状及前沿问题[J]. 河海大学学报:哲学社会科学版,2014,(2).

[4] 梁福庆. 水库移民社会稳定管理的思考及对策[J]. 三峡大学学报:人文社会科学版,2014,(1).

(编辑:邓玲)

Study on reservoir resettlement conflicts and settlement mechanism of Hubei Province

CHEN Beibei<sup>1</sup>, PENG Daiwu<sup>2</sup>, ZHANG Shu<sup>3</sup>

(1. Resettlement Project Consulting Center, Hubei University of Economics, Wuhan 430074, China; 2. Business Administration College of Hubei University of Economics, Wuhan 430074, China; 3. Petition Office of Hubei Immigration Bureau, Wuhan 430074, China)

**Abstract:** As the major indicator of resettlement conflicts, the number and content of immigrants’ petitions from 2010 to 2012 recorded by local immigration administrations in 30 resettlement cities, counties and districts of Hubei Province are collected and sorted out. The characteristics of reservoir resettlement conflicts are summarized, and three types of resettlement conflicts including social, natural and psychological conflicts and their manifestations are analyzed. Measures like improving the efficiency of the policy system, strengthening the inter – departmental cooperation, gradually playing down the identities of immigrants, and offering assistance to immigrants for eliminating poverty and seeking prosperity, are put forward to resolve resettlement conflicts.

**Key words:** reservoir resettlement; resettlement conflicts; types of conflicts

(上接第 89 页)

Application of fuzzy clustering method in dam monitoring data analysis

LIAO Cheng<sup>1</sup>, CAI Desuo<sup>1</sup>, LI Miao<sup>2</sup>, ZHANG Jiyan<sup>1</sup>, WANG Yili<sup>3</sup>

(1. College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China; 2. College of Resources and Environment, North China University of Water Resources and Electric Power, Zhengzhou 450011, China; 3. College of Civil Engineering and Architecture, China Three Gorges University, Yichang 443002, China)

**Abstract:** In order to understand the applicability of fuzzy clustering method in analysis of dam monitoring data, it is applied to analyze the slab deflection deformation series 2013 of Shuibuya concrete face rockfill dam. Firstly, the monitoring pipeline was divided into several sections to set up monitoring points. The fuzzy clustering method is used to classify the deformation monitoring point, and the best classification is determined with F statistic evaluation. The classification results are analyzed and compared, so the key slab deformation monitoring points were determined. Then a slab deformation statistical model was established by consideration of the upstream water depth, temperature and aging factors, and the monitoring data of key points were predicted and fitted with multiple stepwise regression analysis. The analysis results show that using fuzzy clustering method to deal with dam monitoring data possesses advantages of convenient, efficient, and the fitting degree of the model prediction results is relatively high.

**Key words:** slab deformation; fuzzy clustering; statistical model; multiple stepwise regression