

Fisher 最优分割法在汛期分期中的应用

刘克林¹, 王银堂¹, 胡四一¹, 高 波²

(1. 南京水利科学研究院水文水资源与水利工程科学国家重点实验室, 江苏 南京 210029;

2. 水利部国际合作与科技司, 北京 100053)

摘要 : 针对传统汛期分期多采用定性、统计分析方法, 其结果往往带有不确定性的缺陷, 介绍了 Fisher 最优分割法的基本原理和分割步骤。以海河流域密云水库为例, 选取反映水库流域暴雨洪水季节性规律的 5 个指标, 根据专家评判法给出各指标的权重系数, 计算目标函数, 进而进行汛期的分期计算。综合分析和合理性验证表明, 该方法具有多指标聚类、满足时序性划分且能判断分几期较优等特点, 较适用于汛期的定量分期研究。

关键词 : Fisher 最优分割法; 汛期分期; 密云水库

中图分类号 : P333.9

文献标识码 : A

文章编号 : 1006-764X(2007)03-0014-03

Application of Fisher optimal dissection method to flood season division // LIU Ke-lin¹, WANG Yin-tang¹, HU Si-yi¹, GAO Bo² (1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China; 2. International Cooperation and Science and Technology Department of Ministry of Water Resources, Beijing 100053, China)

Abstract : Flood season division by conventional methods is often realized by qualitative statistical analysis, and the results are of high uncertainty. To overcome the disadvantage, the Fisher optimal dissection method as well as its basic principle and steps was introduced. With Miyun Reservoir in Haihe River Basin taken as an example, 5 indexes reflecting the seasonal change of rainstorm and flood of the reservoir region were selected, and the weight coefficients of each index were given based on specialists' judgment. Then, the objective function was derived, and the flood season was divided for the reservoir region. Synthetic analysis and rationality validation show that the method is of multi-factor clustering characteristic, and it can realize time sequence division and confirm the optimal stages, therefore, the method is suitable for quantitative division of flood season.

Key words : Fisher optimal dissection method; division of flood season; Miyun Reservoir

近年来, 随着我国社会经济的快速发展, 用水需求不断加大, 水库作为重要的供水水源地, 人们对其蓄水量和供水保证率的要求也在逐步提高。利用水库分期汛限水位调控洪水资源, 在保障防洪安全的前提下不失时机地多蓄水, 是当前优化水库运行管理机制、缓解水资源短缺矛盾的一个重要途径^[1]。确定和调整分期汛限水位的一个重要前提就是基于水库流域暴雨洪水的季节性变化规律对水库汛期进行科学合理的分期。

目前常用的分期方法主要有成因分析法、数理统计法、模糊集法^[2-3]、分形法^[4]、模糊系统聚类法^[5]等。成因分析法从成因背景出发, 物理概念明确, 但由于缺乏精准的量化指标, 分期较粗略。数理统计法以统计特征因子(降雨、径流等)在汛期内的频率分布作为划分汛期的标准, 但对于如何分期具有较

大的主观性。模糊集法以汛期隶属度来定量描述非汛期到汛期、汛期到非汛期的演变规律, 但由于对指标阈值的选取主观性较大, 使得分期结果带有不确定性。

汛期分期在数学上可以定义为一个时间序列的聚类问题。另外它还具有一些基本特性: 一是影响因子众多, 流域的暴雨洪水受天气系统、环流形势以及下垫面条件等多种因素的综合影响, 所以应综合多个影响因子进行分期; 二是水文系列具有较强的时序性, 汛期分期不同于对散点样本的聚类分析, 分期不能破坏时序性; 三是汛期分期除了要解决如何分期, 还需要确定分几期最优或较优。目前常用的定量分析方法如分形法、模糊系统聚类法等, 处理上就存在上述一些问题, 或只能考虑单个影响因子, 或不能保证时序性, 没有确定分几期较优的定量标准。

因此,针对汛期分期的基本特性,需要研究更为适用的分期方法。

Fisher 最优分割法^[6-8]是对有序样本进行分类的一种统计方法,它具有多指标聚类、不破坏样本原有顺序等特点,并能够根据定义的目标函数确定分期较优,因此能较好地应用于汛期分期中。事实上,该方法作为一种传统的线性分类方法,已经在农业区划、气象统计预报、地震周期预报、工业产品检测、医学分析等许多方面得到了成功的应用。

1 Fisher 最优分割法的原理和分割步骤

Fisher 最优分割法的分类依据是样本的总离差平方和最小,进行分割的原则是使得各类内部样本之间差异最小,而各类之间的差异最大。

设有 n 个按一定顺序排列的样本,测得 m 项指标,样本和指标之间通过指标特征值 x_{ij} ($i = 1 \sim n, j = 1 \sim m$) 构建关系矩阵 X :

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

若各指标间的物理量纲不同,可将指标特征值进行一致无量纲化处理:

$$x_{ij}' = \frac{x_{ij}}{x_{\max j}} \tag{1}$$

式中: x_{ij}' 为一致无量纲化后的指标特征值; $x_{\max j}$ 为对应于第 j 个指标 x_{ij} 中的最大值。

按照各指标对样本分类的重要程度赋以不同的权重系数 $\omega_1, \omega_2, \dots, \omega_m$, 加权平均后可将多指标特征值矩阵转化为一维特征值向量 Y :

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11}' & \cdots & x_{1m}' \\ \vdots & \ddots & \vdots \\ x_{n1}' & \cdots & x_{nm}' \end{pmatrix} \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_m \end{pmatrix} \tag{2}$$

以向量 Y 作为初始分类的样本特征值,就可以对样本序列进行分割,最优分割的步骤如下:

a. 定义类直径。类内部样本之间的差异程度用类直径来表示,直径越小则差异越小。设某一类 $G_{i,j} = \{y_i, y_{i+1}, \dots, y_j\}$, 其中 $j > i$, 定义其直径 $D(i,j)$ 为样本离差平方和,则

$$D(i,j) = \sum_{r=i}^j (y_r - \bar{y}_{ij})^2 \tag{3}$$

$$\bar{y}_{ij} = \frac{1}{j-i+1} \sum_{r=i}^j y_r$$

式中: y_r 为样本特征值; $\bar{y}_{ij}$ 为均值。

b. 定义目标函数。若将 n 个样本分成 k 类: $\{y_{j_1}, y_{j_1+1}, \dots, y_{j_2-1}\}, \{y_{j_2}, y_{j_2+1}, \dots, y_{j_3-1}\}, \dots, \{y_{j_k},$

$y_{j_k+1}, \dots, y_{j_{k+1}-1}\}$ 其中 $j_1, j_2, \dots, j_k$ 为 k 个分点,其下标满足 $1 = j_1 < j_2 < \dots < j_k \leq j_{k+1} - 1 = n$ 。最优分割的实质就是寻找某一组分点,使得所有分类的直径总和最小,定义这种分类的目标函数为

$$B(n,k) = \min \sum_{r=1}^k D(j_r, j_{r+1} - 1) \tag{4}$$

式中 $n = j_{k+1} - 1$ 。

c. 求解最优分类。有如下定理:有序样本的最优 k 类分割一定是在其某一个截尾子段最优 $k-1$ 类分割之后再添加一类形成的。容易验证有如下递推公式:

$$B(n,k) = \min_{k \leq j \leq n} [B(j-1,k-1) + D(j,n)] \tag{5}$$

假设要分 k 类,首先找分点 j_k 使式(5)达到极小,即

$$B(n,k) = B(j_k-1,k-1) + D(j_k,n) \tag{6}$$

得到第 k 类 $G_k^* = \{y_{j_k}, y_{j_k+1}, \dots, y_n\}$, 然后找到第 $k-1$ 个分点 j_{k-1} , 使它满足:

$$B(j_k-1,k-1) = B(j_{k-1}-1,k-2) + D(j_{k-1}, j_k - 1) \tag{7}$$

得到第 $k-1$ 个分类 $G_{k-1}^* = \{y_{j_{k-1}}, y_{j_{k-1}+1}, \dots, y_{j_k-1}\}$, 类似的方法得到所有类 $G_1^*, \dots, G_k^*$, 这就是分类数为 k 的最优分类结果。

d. 确定较优分类数目。确定较优分类数有两种方法:①绘制目标函数 $B(n,k)$ 随分类数 k 的变化曲线,取该曲线拐弯处的 k 值为较优分类数;②计算非负斜率 $\beta(k) = \left| \frac{B(n,k) - B(n,k-1)}{k - (k-1)} \right|$, 当 $\beta(k)$ 较大时表示分 k 类优于分 $k-1$ 类,当 $\beta(k)$ 接近 0 时表示没有继续细分的必要,一般取 $\beta(k)$ 最大时的 k 值为较优分类数。

2 实例分析

密云水库地处我国华北地区,属潮白河流域,水库以上控制流域面积为 15 788 km²。流域属中纬度大陆性季风气候,夏季较短,气候炎热而多雨,冬季较长,气候寒冷而干燥。汛期降雨多以暴雨形式集中出现,多年平均降雨量约 550 mm,降雨量年内分配极为不均,汛期 6~9 月份约占全年的 80%,有必要研究汛期分期问题。

2.1 基本资料及指标选择

水库流域汛期分期的主要影响因子为暴雨和洪水在汛期内的时程分布特性和变化规律,其中暴雨反映了气象因素的影响,而洪水则是气象与下垫面等多种因素综合作用的反映。本文基于水库控制流域的降雨资料和入库流量资料来分析,降雨采用密

云水库控制流域的 10 个代表性雨量站 1963 ~ 2000 年逐日雨量资料 ,利用泰森多边形法得到各年逐日流域面平均雨量 ;入库流量采用密云水库 1962 ~ 2000 年逐日平均入库流量资料。

确定研究的论域为密云水库的汛期 (6 月 15 日至 9 月 15 日) ,以侯 (5 ~ 6d 为 1 个侯) 为基本分期单元 ,论域内共有 18 个侯。共选取 5 个指标 :侯面平均雨量、侯最大 1 日面平均雨量、侯最大 3 日面平均雨量、侯内暴雨天数 (日雨量超过 50 mm 的降雨天数) 和侯平均入库流量。指标特征值利用基本资料统计得出 ,见表 1。

表 1 指标特征值

侯序号	月(侯)	侯面平均雨量/mm	侯最大1日面平均雨量/mm	侯最大3日面平均雨量/mm	侯内暴雨天数/d	侯平均入库流量/(m ³ ·s ⁻¹)
1	㉙(4)	12.7	7.5	11.2	1	13.38
2	㉙(5)	13.1	8.5	12.4	1	12.90
3	㉙(6)	15.1	9.1	13.8	12	15.70
4	㉙(1)	24.8	18.3	22.8	29	29.82
5	㉙(2)	21.5	12.8	18.7	9	42.81
6	㉙(3)	25.0	15.3	21.9	18	44.28
7	㉙(4)	21.9	13.0	18.9	15	43.80
8	㉙(5)	28.5	16.1	25.3	37	59.18
9	㉙(6)	29.5	18.5	24.6	29	90.04
10	㉙(1)	28.9	18.5	25.2	34	109.00
11	㉙(2)	24.9	14.2	21.6	17	89.58
12	㉙(3)	27.0	14.4	24.1	30	112.40
13	㉙(4)	15.0	10.5	13.9	9	90.89
14	㉙(5)	12.1	8.2	11.0	5	69.60
15	㉙(6)	17.0	10.9	14.0	6	57.69
16	㉙(1)	13.9	9.4	12.2	3	50.79
17	㉙(2)	8.1	5.6	7.5	1	40.00
18	㉙(3)	7.5	5.8	6.9	0	36.16

注 ㉙(4)表示 6 月份的 第 4 个侯 ,即 6 月 15 ~ 20 日。

2.2 分期计算

由于各指标的物理量纲不同 ,首先应用式 (1) 将指标特征值进行一致无量纲化处理 ,再由专家评判法得出 5 个指标相应的权重系数 , $\omega=(0.25\ 0.125\ 0.125\ 0.25\ 0.25)^T$,代入式 (2) 计算得出向量 Y 。

由式 (3) ~ (7) 计算得出不同截断样本长度的目标函数 $B(i,k)$,其中 $i=k+1,k+2,\dots,m;k=2,3,\dots,m$ 。这里 n 取 18 ,计算结果见表 2。

若将 18 个样本分 3 类 ,由表 2 可知 $B(n,k)=B(18,3)=0.29$,相应分点 $j_k=j_3=13$,得出第 3 个分类 $G_3^*=\{y_{13},y_{14},\dots,y_{18}\}$,再把前 12 个样本分成 2 类 ,查表 2 得 $B(12,2)=0.21$, $j_2=4$,即第 2 个分类 $G_2^*=\{y_4,y_5,\dots,y_{12}\}$,从而第 1 个分类 $G_1^*=\{y_1,y_2,y_3\}$,依据同样方法可以得到所有可能分类的最优分类结果。

绘制目标函数 $B(n,k)$ 及其非负斜率 $\beta(k)$ 随

表 2 目标函数 $B(i,k)$ 计算结果

<i>i</i>	<i>k</i> = 2	<i>k</i> = 3	<i>k</i> = 4	<i>k</i> = 5	<i>k</i> = 6	<i>k</i> = 7	...	<i>k</i> = 18
3	0.00(3)						...	
4	0.00(4)	0.00(4)						
5	0.00(4)	0.00(5)	0.00(5)					
6	0.00(4)	0.00(5)	0.00(5)	0.00(6)				
7	0.00(4)	0.00(5)	0.00(5)	0.00(6)	0.00(7)			
8	0.00(4)	0.00(8)	0.00(8)	0.00(8)	0.00(8)	0.00(8)		
9	0.10(4)	0.00(8)	0.00(8)	0.00(8)	0.00(8)	0.00(8)		
10	0.10(4)	0.00(8)	0.00(8)	0.00(8)	0.00(10)	0.00(10)		
11	0.10(4)	0.00(8)	0.00(11)	0.00(11)	0.00(11)	0.00(11)		
12	0.20(4)	0.00(8)	0.00(8)	0.00(12)	0.00(12)	0.00(12)		
13	0.20(4)	0.10(8)	0.00(13)	0.00(13)	0.00(13)	0.00(13)		
14	0.30(4)	0.20(13)	0.00(13)	0.00(13)	0.00(13)	0.00(13)		
15	0.40(4)	0.20(13)	0.00(13)	0.00(13)	0.00(13)	0.00(13)		
16	0.50(4)	0.20(13)	0.00(13)	0.00(13)	0.00(13)	0.00(13)		
17	0.60(4)	0.20(13)	0.10(13)	0.00(17)	0.00(17)	0.00(17)		
18	0.70(14)	0.20(13)	0.10(13)	0.00(17)	0.00(17)	0.00(17)	...	0.00(18)

注 括号内的数字为目标函数所对应的分点 j_k ,以各侯的序号表示。

分类数 k 的变化曲线 ,如图 1 所示。

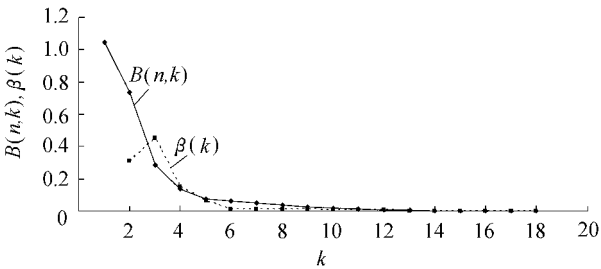


图 1 $B(n,k)$ 和 $\beta(k)$ 随分类数 k 的变化曲线

从图 1 中可以看出 ,目标函数 $B(n,k)$ 在 $k=3$ 处明显拐弯 ,非负斜率 $\beta(k)$ 在 $k=3$ 处取得最大值 ,因此分析认为 ,密云水库汛期以分 3 期较优 ,分期结果见表 3。

表 3 Fisher 最优分割法分期结果与现行分期方案比较

分期方法	前汛期	主汛期	后汛期
Fisher 最优分割法	06-15 ~ 06-30	07-01 ~ 08-15	08-16 ~ 09-15
现行方案		06-15 ~ 08-10	08-11 ~ 09-15

由表 3 可知 ,本次分期计算结果是在现行方案原有主汛期内划分出一个前汛期 ,表明密云水库流域暴雨洪水主要集中在主汛期 ,主汛期可以从 7 月 1 日开始 ,主、后汛期的分界点与现行方案相差 5 d ,这与本次分期划分时段以侯为单元有关。总体来看 ,Fisher 最优分割法的分期结果能较好地反映暴雨洪水在汛期的变化规律 ,且与现行方案基本一致 ,验证了该方法的实用性及其结果的合理性。

3 结 语

a. 本文总结了汛期分期的数学本质和基本特性 ,针对目前分期方法在适用性等方面的不足 ,将 Fisher 最优分割法应用于汛期分期中。以密云水库为例 ,在确定出分 3 期较优的基础上 , (下转第 37 页)

变曲线越来越趋近于双曲线,产生这种情形的根源是不同 σ_3 下的应力-应变曲线叠加了相同的应力-应变直线;②当 σ_3 相同时,随着胶凝材料含量的增加,拟合良好段对应的应变越来越大,拟合效果越来越好。

2 模型参数的确定

双曲线模型共 8 个参数,即 $k, n, R_f, c, \varphi, F, G, D$ 。在本次试验中,由于未能测出材料的侧向应变,因此不能确定 F, G, D ,只能确定 k, n, R_f, c, φ 这 5 个参数。在本次计算中对泊松比不进行非线性迭代。

2.1 c 和 φ 的确定

从超贫胶结材料三轴试验曲线可以看出,C50 超贫胶结材料应力-应变曲线的上升段基本为双曲线,破坏时的应变值基本在 2% 左右,考虑到试验曲线还可以延长,所以建议破坏值取 2% ~ 5%。在本次计算中,考虑到了双曲线的拟合情况,所以破坏时的应变值取为 4%。

根据拟合的双曲线,取轴向应变为 4% 时对应的 σ_3 与 $\sigma_1 - \sigma_3$,可得出不同胶凝材料含量的超贫胶结材料的性质参数:黏聚力 c 和内摩擦角 φ ,见表 2。

表 2 不同胶凝材料含量的超贫胶结材料参数

胶凝材料含量/ ($\text{kg} \cdot \text{m}^{-3}$)	c/kPa	$\varphi/(\text{^\circ})$
30	451.29	41.30
50	505.06	47.15
60	1128.11	47.42

这里需要说明的是表 2 中的 c 和 φ 是对应叠加后虚拟曲线的取值,而不是超贫胶结材料真实的黏聚力和内摩擦角。

2.2 R_f 的确定

由本构模型切线弹性模量的推导过程可知:

$$(\sigma_1 - \sigma_3)_f = \frac{2c \cos \varphi + 2\sigma_3 \sin \varphi}{1 - \sin \varphi} \quad (3)$$

根据 $R_f = \frac{(\sigma_1 - \sigma_3)_f}{(\sigma_1 - \sigma_3)_u}$ (式中 $(\sigma_1 - \sigma_3)_u$ 为主应力差渐近值)可以算出不同胶凝材料含量的超贫胶结材料在不同围压下的值,取各自的平均值,就得到不同胶凝材料含量相应的 R_f 。计算得到胶凝材料含量为 30 kg/m^3 , 50 kg/m^3 , 60 kg/m^3 时 R_f 值分别为 0.86 0.88 0.80。

2.3 k 和 n 的确定

由 $\lg(E_i/p_a)$ 与 $\lg(\sigma_3/p_a)$ 关系曲线及拟合方程,可以计算 k 和 n 。计算得到胶凝材料用量为 30 kg/m^3 , 50 kg/m^3 , 60 kg/m^3 时,对应的 k 和 n 值分别 3277.93 0.384 6 4897.79 0.441 5 5754.4 0.350 5。

3 结 论

本文在三轴试验结果基础上,针对试验曲线存在明显的软化特性,引入虚加刚性弹簧法建立了超贫胶结材料的非线性本构模型,把实际材料的切线模量表达为虚拟邓肯-张模型的切线模量与虚加弹簧的弹性模量之差,并给出了不同胶凝材料含量虚拟邓肯-张模型参数和虚加弹簧的弹性模量,成功地建立了具有软化特性的超贫胶结材料的本构模型,为超贫胶结材料的非线性分析奠定了基础。

参考文献:

[1] 张镜剑,孙明权.超贫胶结材料土石面板坝研究阶段报告[R].郑州:华北水利水电学院,2002.
[2] 钱家欢,殷宗泽.土工原理与计算[M].北京:中国水利水电出版社,1996:55-57.
[3] 郭诚谦,陈慧远.土石坝[M].北京:中国水利水电出版社,1981:584-589. (收稿日期 2007-05-08 编辑 骆超)

(上接第 16 页)将汛期划分为前汛期(6 月 15 日 ~ 6 月 30 日)、主汛期(7 月 1 日 ~ 8 月 15 日)和后汛期(8 月 16 日 ~ 9 月 15 日)。分期结果与现行方案大体一致,验证了本文方法的合理性。

b. Fisher 最优分割法具有多指标聚类、时序性划分以及能确定分几期较优等特点,基本满足汛期分期问题的特性,因此适用于汛期的定量分期研究。通过实例分析,本文认为综合分析汛期分期的影响因素很重要,即指标的合理选取及权重系数的确定是研究的关键,有待进一步探讨。

参考文献:

[1] 胡四一,高波,王忠静.海河流域洪水资源安全利用:水库汛限水位的确定与运用[J].中国水利,2002(10):105-

108.
[2] 陈守煜.从研究汛期描述水文系统模糊集分析的方法[J].水科学进展,1995,16(2):133-138.
[3] 陈守煜.水文水资源系统模糊识别理论[M].大连:大连理工大学出版社,1992:152-160.
[4] 侯玉,吴伯贤,郑国权.分形理论用于洪水分期的初步探讨[J].水科学进展,1999,19(2):140-143.
[5] 高波,刘克林,王银堂,等.系统聚类法在水库汛期分期中的应用[J].水利水电技术,2005,36(6):1-5.
[6] 张尧庭,方开泰.多元统计分析引论[M].北京:科学出版社,1982:447-449.
[7] 和宏伟,张爱玲.Fisher 最优分割法在云南地震分期中的应用[J].地震研究,1994,17(3):231-239.
[8] 郭玉文.Fisher 法在空气环境监测点位优化中的应用[J].环境工程,2002,20(4):64-66.

(收稿日期 2006-09-07 编辑 骆超)