

K 均值聚类和模糊 C 聚类对黄前水库流域水文要素测报分析

徐珊

临沂市水文局

DOI:10.32629/hwr.v4i8.3246

[摘要] 在黄前水库观测资料的基础上,运用K均值聚类和模糊C聚类的方法,对黄前水库的降雨量和水文要素进行分析,由于流域降水是影响黄前水库来水量的最重要的影响因素,建立了水库来水量与流域降水量之间的线性关系。这些分析结果对黄前水库的合理开发与利用有指导作用。

[关键词] K均值聚类; 模糊C聚类; 黄前水库流域; 水文要素测报分析

中图分类号: TV75 **文献标识码:** A

1 黄前水库概况

地理位置: 黄前水库位于泰安市岱岳区泰山东麓,东经 $116^{\circ} 20' \sim 117^{\circ} 59'$,北纬 $35^{\circ} 38' \sim 36^{\circ} 28'$,水源由泰山、长清岭和大青山的三条溪流汇聚而来,流域为扇形,控制流域面积为292K m²,多年平均来水量6000万m³。

气候条件: 库区属暖温带大陆性半湿润季风气候,多年平均气温 18.5°C ,年平均降雨量758mm,75%的降雨量集中于6-9月份,最大年降雨量1303mm(1964年),连续3日暴雨量238.6mm,最小年降雨量334.4mm(1989年),多年平均径流深254mm。

工程概况: 黄前水库于1958年动工兴建,1960年建成蓄水,水库位于泰安市岱岳区北部黄前镇,坝址以上控制流域面积292km²,总库容 $8248 \times 104 \text{m}^3$,兴利库容 $5913 \times 104 \text{m}^3$ 。黄前水库防洪标准为1983年水利工程三查三定洪水复核结果:百年一遇设计,三日净雨317.1mm,最高库水位209.0m,千年一遇校核,三日净雨486mm,最高库水位210.65m,可能最大降雨校核,三日净雨1014.5m,最高库水位212.17m,兴利水位209.0m,死水位190.60m。

黄前水库位于大汶河流域上游,黄前水库流域降雨及水库状况又与流域水文地质、水体自净能力等方面密切相关,同时,工农业生产的快速发展使流域水资源开发利用程度过高,水文情势改变较大,增加了水环境变化的不确定因素。因此,水库流域要素研究十分复杂。

目前,黄前水库流域降雨径流等水文要素及水库调度基于水库建设以来的整合信息路径不够清晰,缺乏理论与实证研究支撑,没有研究水库流域降雨关联要素、水库变化因子整合预测模型,无法形成针对性的整合行动目标;流域降雨的整体趋势把握不够,且主要采用传统的分析方法解决问题等。因此,无论是把握降雨

径流变化规律,还是科学解决水资源开发利用与流域生态水利建设环境之间的矛盾,实现水库及流域可持续开发利用都具有重要的理论和实践意义。

2 流域降雨变化特征

黄前水库流域有五个雨量监测站:独路、彭家峪、下港、西麻塔和黄前水库,1964年至2012年的年降雨量如图2-1。

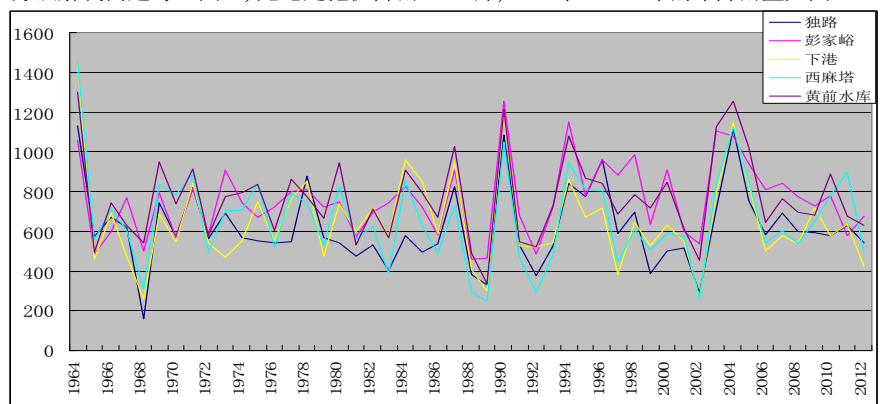


图 2-1 五站年降雨量对比

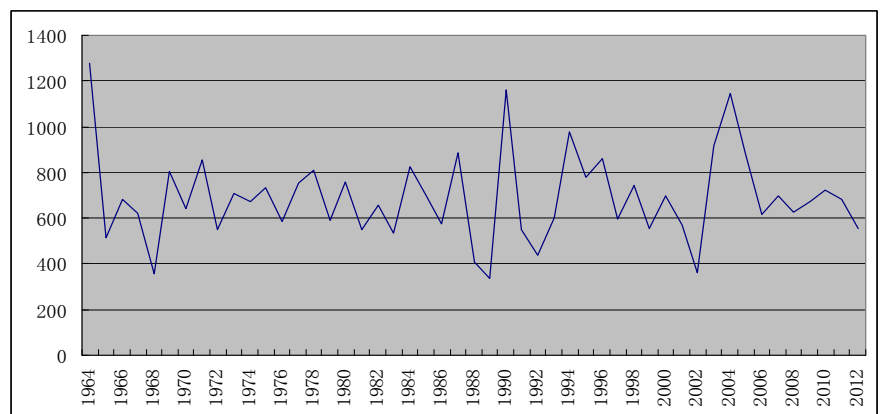


图 2-2 五站年降雨量均值

从图2-1至图2-2年降雨量变化情况
况分析得出,独路、彭家峪、下港、西
麻塔和黄前水库五站监测降雨量具有
如下特点:

- (1)降雨量具有同步性。
- (2)各站降雨量大小没有确切的排列顺序。
- (3)多年降雨量变化幅度区间基本稳定。
- (4)没有出现连续多年的高峰、低谷。
- (5)五站均值变化代表性与各站一致。

3 聚类分析黄前水库流域水文要素

聚类分析的目的是把分类对象按照一定规则分成若干类,这些类不是事先给定的,而是根据数据的特征确定的,对类的数目和类的结构不必作任何假定。在同一类里的这些对象在某种意义上倾向于彼此相似,而在不同类里的对象倾向于不相似。聚类属于无监督学习,同回归、朴素贝叶斯、SVM等都是有类别标签的,而聚类的样本中却没有给定类别,只有表示特征的数据列。因此,通过对于黄前水库流域水文特征要素进行聚类分析,对水文要素进行测报,是一种科学的发现潜在规律的方法。

3.1 K均值聚类黄前水库流域水文要素测报分析

将独路、彭家峪、下港、西麻塔和黄前水库五站的年降雨作为输入参数进行K聚类,也就是将黄前水库流域降雨五站监测的降雨量统筹考虑进行聚类分析,评判历年和每年的降雨情况。

1964-2012年五站年降雨量K聚类分析设置最终分类从3类开始,在程序设计时自动分类到6类,以后可以根据水文监测的需要调整分类,模型运行后每一分类情况如下:

六类分法K均值聚类的轮廓图

通过以上K聚类分析得出结论如下:

(1)使用K均值聚类分析降雨等水文要素,适用于选择事先约定的标准降雨量作为初始凝聚点,这样根据约定的分类数目和初始凝聚点,然后运行聚类模型,得到十分满意的效果。

(2)分类不宜过粗或过细,采用5类

①设置最终分类个数为3类时:

第一类	1964 1971 1978 1984 1987 1990 1994 1996 2003 2004 2005
第二类	1965 1968 1988 1989 1992 2002
第三类	1966 1967 1969 1970 1972 1973 1974 1975 1976 1977 1979 1980 1981 1982 1983 1985 1986 1991 1993 1995 1997 1998 1999 2000 2001 2006 2007 2008 2009 2010 2011 2012

②设置最终分类个数为4类时:

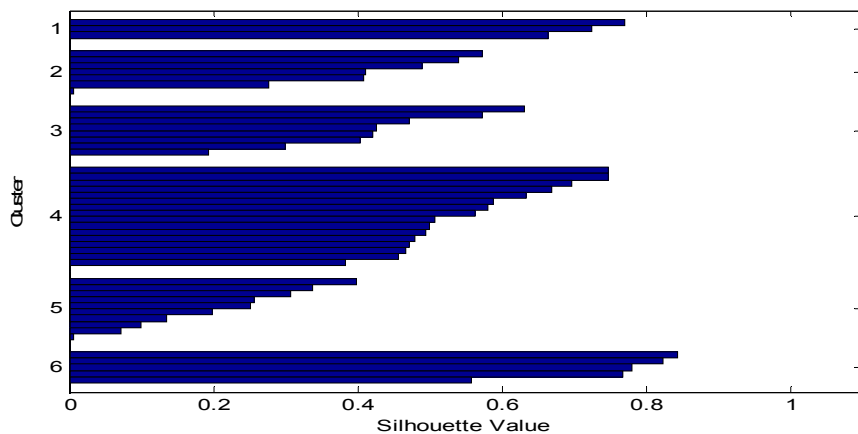
第一类	1969 1971 1978 1984 1987 1994 1995 1996 2003 2005
第二类	1964 1990 2004
第三类	1965 1968 1972 1981 1983 1986 1988 1989 1991 1992 1999 2001 2002 2012
第四类	1966 1967 1970 1973 1974 1975 1976 1977 1979 1980 1982 1985 1993 1997 1998 2000 2006 2007 2008 2009 2010 2011

③设置最终分类个数为5类时:

第一类	1969 1971 1978 1984 1987 1994 1996 2003 2005
第二类	1965 1967 1972 1976 1979 1981 1983 1986 1991 1993 1997 1999 2001 2006 2008 2012
第三类	1964 1990 2004
第四类	1966 1970 1973 1974 1975 1977 1980 1982 1985 1995 1998 2000 2007 2009 2010 2011
第五类	1968 1988 1989 1992 2002

④设置最终分类个数为6类时:

第一类	1964 1990 2004
第二类	1971 1978 1987 1994 1996 2003 2005
第三类	1966 1970 1974 1975 1982 1985 2009 2011
第四类	1965 1967 1972 1976 1979 1981 1983 1986 1991 1993 1997 1999 2001 2006 2008 2012
第五类	1969 1973 1977 1980 1984 1995 1998 2000 2007 2010
第六类	1968 1988 1989 1992 2002



六类分法K均值聚类的轮廓图

或6类较合适。6类分法可以描述为,全年降雨量与历年比较为:大、较大、中、一般、较小和小。比如,对2012年降雨采用6类分法,测报为“黄前水库流域全年降雨为一般水平”。

(3)采用五站降雨综合评判,能够较全面反映黄前水库流域历年降雨特征,

水库流域归类比较系统,特别是为今后水库调度提供基准比较案例决策支持。

(4)建立地区型和流域内K聚类分析比较对于区域水资源管理提供更加具体的决策数据。

3.2模糊C聚类黄前水库流域水文要素测报分析

模糊C-均值聚类算法fuzzy c-means algorithm(FCMA)或称(FCM),作为无监督机器学习的主要技术之一,是用模糊理论对重要数据分析和建模的方法,建立了样本类属的不确定性描述,能比较客观地反映现实世界,它已经有效地应用在大规模数据分析、数据挖掘、矢量量化、图像分割、模式识别等领域,具有重要的理论与实际应用价值,随着应用的深入发展,模糊聚类算法的研究不断丰富。在众多模糊聚类算法中,模糊C-均值(FCM)算法应用最广泛且较成功,它通过优化目标函数得到每个样本点对所有类中心的隶属度,从而决定样本点的类属以达到自动对数据样本进行分类的目的。FCM算法为,FCM把n个向量 $x_i (i=1, 2, \dots, n)$ 分为c个模糊组,并求每组的聚类中心,使得非相似性指标的价值函数达到最小。FCM与HCM的主要区别在于FCM用模糊划分,使得每个给定数据点用值在0, 1间的隶属度来确定其属于各个组的程度。与引入模糊划分相适应,隶属矩阵U允许有取值在0, 1间的元素。不过,加上归一化规定,一个数据集的隶属度的和总等于1。FCM的价值函数(或目标函数)就是所有个点隶属度乘以该点与中心的欧氏距离之和。在批处理方式运行时,FCM用下列步骤确定聚类中心 c_i 和隶属矩阵U。

步骤1: 用值在0, 1间的随机数初始化隶属矩阵U,使其满足设定的约束条件。

步骤2: 计算c个聚类中心 $c_i, i=1, \dots, c$ 。

步骤3: 计算价值函数。如果它小于某个确定的阈值,或它相对上次价值函数值的改变量小于某个阈值,则算法停止。

步骤4: 计算新的U矩阵。返回步骤2。

上述算法也可以先初始化聚类中心,然后再执行迭代过程。由于不能确保FCM收敛于一个最优解。算法的性能依赖于初始聚类中心。因此,我们要么

①模糊C聚类自动分类为5类时:

第一类	1964	1990	2004						
第二类	1968	1988	1989	1992	2002				
第三类	1969	1971	1978	1984	1987	1994	1995	1996	1998
	2003	2005							
第四类	1966	1970	1973	1974	1975	1977	1980	1982	1985
	2000	2007	2009	2010	2011				
第五类	1965	1967	1972	1976	1979	1981	1983	1986	1991
	1993	1997	1999	2001	2006	2008	2012		

②模糊C聚类自动分类为6类时:

第一类	1964	1990	2004						
第二类	1966	1974	1975	1977	1980	1985	1995	1998	2000
	2010	2011							
第三类	1965	1970	1972	1976	1981	1983	1986	1991	1999
	2001	2012							
第四类	1969	1971	1978	1984	1987	1994	1996	2003	2005
第五类	1968	1988	1989	1992	2002				
第六类	1967	1973	1979	1982	1993	1997	2006	2007	2008
								2009	

2003年-2012年隶属度值:

0.1580	0.8199	0.0865	0.0339	0.0535	0.0191	0.0467	0.0522	0.0697	0.0371
0.1912	0.0399	0.1532	0.1199	0.2323	0.0764	0.2572	0.4069	0.2489	0.1148
0.1165	0.0295	0.0771	0.1982	0.1699	0.1240	0.2266	0.1369	0.2176	0.3600
0.3155	0.0562	0.5421	0.0751	0.1490	0.0445	0.1195	0.1541	0.1508	0.0741
0.0800	0.0223	0.0488	0.0769	0.0780	0.0413	0.0808	0.0685	0.1106	0.1303
0.1389	0.0322	0.0923	0.4959	0.3173	0.6947	0.2693	0.1814	0.2024	0.2837

用另外的快速算法确定初始聚类中心,要么每次用不同的初始聚类中心启动该算法,多次运行FCM对历年降雨等水文要素进行测报,K聚类可以进行动态分析,当然历年数据系列对象间没有明确的界限,往往具体亦此亦彼的表现,如果严格划分到某类存在一定的不合理性。因此,使用模糊聚类分析处理这类问题,在模糊划分中,每一个样品不是严格地划分为某一类,而是以一定的隶属度属于某一类。

2012年隶属度最大值为0.3600,化为第三类。

通过以上模糊C聚类分析得出结论如下:

(1)使用模糊C聚类分析降雨等水文要素,适用于从归类隶属度大小分析历年降雨量序列,自动决定类别,得到最好的隶属度分类效果。

(2)同样,分类不宜过粗或过细,采用5类或6类较合适。6类分法可以描述为,

全年降雨量与历年比较为:大、较大、中、一般、较小和小。比如,对2012年降雨采用6类分法,测报为“黄前水库流域全年降雨为一般水平”。

(3)采用五站降雨综合评判,能够较全面反映黄前水库流域历年降雨特征,水库流域归类比较系统,特别是为今后水库调度提供基准比较案例决策支持。

(4)和K聚类互相比,建立地区型和流域内模糊C聚类分析比较对于区域水资源管理提供更加具体的决策数据。

[参考文献]

[1]周凯慧,刘霞,李化民,等.黄前水库水质监测及变化规律研究[J].水土保持学报,2004(04):150-153+157.

[2]张春荣,高宗军.黄前水库来水量特征与影响因素分析[J].地下水,2008(03):74-77.

[3]杨雪,赵得玉,申正涛,等.清水河降水分布特征及影响因素分析[J].湖北农机化,2019(05):64.