

DOI: 10.20040/j.cnki.1000-7709.2023.20221364

大渡河流域融雪期划分方法研究

李彬权, 荣伟光, 吴亚琪, 梁忠民

(河海大学水文水资源学院, 江苏 南京 210098)

摘要: 为准确判断融雪期起止时间, 以大渡河融雪径流主要来源区的丹巴以上流域为例, 基于 2009~2020 年的水文气象资料, 分别采用集对分析法、系统聚类法和 K-means 聚类法对流域枯季(11月~次年5月)融雪期进行分期计算, 并对分期结果进行合理性评估, 确定流域最终融雪期划分方案。结果表明, 大渡河丹巴以上流域的主融雪期为 1月11日~5月10日, 其前的 11月1日~1月10日为退水期, 其后的 5月11日~5月31日为降雨径流开始期(仍有部分融雪径流)。研究结果可为高纬度地区的融雪期划分提供参考。

关键词: 集对分析; 系统聚类; K-means 聚类; 融雪期; 大渡河流域

中图分类号: TV126⁺.6; P333

文献标志码: A

文章编号: 1000-7709(2023)05-0038-05

1 概况

大渡河流域地处川西南山地, 属高纬度地区, 积雪期最长可达 5 个月, 丹巴以上流域面积约为 52 763 km², 高程介于 3 000~7 000 m 之间, 在春季及初夏气温回升时期形成大量融雪径流, 这也是流域内众多大型调节性水库水位消落控制的关键期。一直以来, 关于大渡河流域融雪期的起止时间没有明确界定。因此, 明确融雪期起止时间、准确判断融雪期来水过程, 对于流域枯季防洪发电、水资源配置及风险管理等方面具有十分重要的理论与实践意义^[1]。聚类为一种无监督的分类方法, 即将由多个独立样本构成的数据集经相似性指标判定进而划分为若干个类簇的过程^[2], 基于聚类思想的水文分期问题通过对连续时间内的特征指标进行相似性计算, 从而得到携带有分类标签的各时段分期结果。目前, 关于此方面的研究已较多, 如陈守煜等^[3]提出了模糊集合理论并将其用于划分汛期起讫时间; 刘攀等^[4]将变点分析法引入到三峡水库汛期分期的研究中; 谢飞等^[5]提出将集对分析法应用于潘家口水库的汛期分期中; 蒋海艳等^[6]以潘家口水库坝址以上流域为例, 展开了基于灰色定权聚类法的汛期分期研究; 刘浅奎等^[7]采用熵权模糊集对分析法对溪洛

渡水库进行汛期分期研究。为此, 本文基于已有研究成果, 以大渡河上游丹巴以上流域为例, 根据旬时间尺度的多年平均流量、雨量、气温、最高气温、最低气温、地表太阳辐射及日照时数共 7 个指标, 分别采用集对分析法、系统聚类法、K-means 聚类法 3 种聚类分析法对流域融雪期进行划分, 并对分期结果展开合理性评估, 最终确定研究区的枯季融雪期划分方案, 为大渡河流域融雪径流预报及水资源调度等提供参考依据。

2 研究方法

2.1 集对分析法

对集合展开特性分析, 定义两个集合间联系度的表达式为:

$$\mu_{(X,Y)} = (S + Fi + Pj) / n \quad (1)$$

其中 $n = S + F + P$

式中, S 为集合 X, Y 在所有 n 种特性中所共有的特性; F 为既不共有也不相互对立的特性; P 为对立的特性; i 为差异度系数; j 为对立度系数。

基于集对分析法的流域融雪期划分流程见图 1。

2.2 系统聚类法

(1) 数据标准化。假设共有 n 个指标, 每个指标具有 m 个观测值, x_{ij} 为第 i 指标的第 j 个观测值, 以所有指标的观测值作为元素, 构成观测值

收稿日期: 2022-07-04, 修回日期: 2022-08-09

基金项目: 国家自然科学基金重点项目(41730750)

作者简介: 李彬权(1984-), 男, 博士、教授、博导, 研究方向为水文水资源, E-mail: libinquan@hhu.edu.cn

通讯作者: 梁忠民(1962-), 男, 博士、教授、博导, 研究方向为水文水资源, E-mail: zmliang@hhu.edu.cn

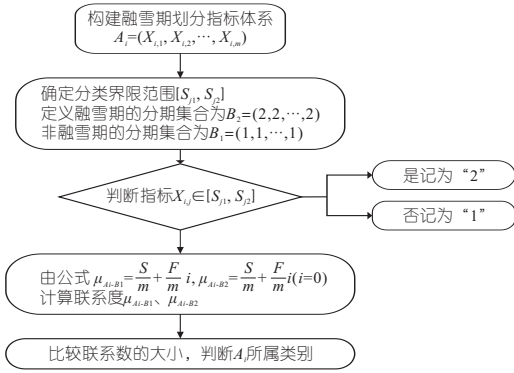


图 1 集对分析法划分融雪期步骤流程图

Fig. 1 Flow chart of dividing snowmelt period by set pair analysis

矩阵:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \quad (2)$$

标准化计算公式为:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (3)$$

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad (4)$$

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / s_j \quad (5)$$

式中, $i=1, 2, \dots, n$; $j=1, 2, \dots, m$ 。

经式(3)~(5)计算后,每一类指标的均值为 0, 标准差为 1。

(2)定义点间距离。假设指标 A 的观测值为 $(x_{A1}, x_{A2}, \dots, x_{Am})$, 指标 B 的观测值为 $(x_{B1}, x_{B2}, \dots, x_{Bm})$, 将这两个指标的观测值视为是 m 维空间内的两个点的坐标, 可对两个指标间的距离做出定义, 采用欧氏距离进行计算, 公式为:

$$d_{(A,B)} = \sqrt{\sum_{j=1}^m (x_{i,j} - x_{i_2,j})^2} \quad (6)$$

(3)定义类间距离。常用类间距离计算方法有最短距离法、最长距离法、中间距离法、重心法、类平均距离法、离差平方和法等。其中最短距离 $D(G_p, G_q)$ 法计算公式为:

$$D(G_p, G_q) = \min \{d_{ij} \mid x_i \in G_p, x_j \in G_q\} \quad (7)$$

最长距离 $D(G_p, G_q)$ 法计算公式为:

$$D(G_p, G_q) = \max \{d_{ij} \mid x_i \in G_p, x_j \in G_q\} \quad (8)$$

中间距离 $D^2(G_r, G_k)$ 法计算公式为:

$$D^2(G_r, G_k) = (D_{pk}^2 + D_{qk}^2) / 2 - D_{pq}^2 / 4 \quad (9)$$

重心 $D(G_p, G_q)$ 法计算公式为:

$$D(G_p, G_q) = \|\bar{X}_p - \bar{X}_q\|^2 \quad (10)$$

类平均距离 $D(G_p, G_q)$ 法计算公式为:

$$D(G_p, G_q) = \frac{1}{n_p n_q} \sum_{i \in G_k} \sum_{j \in G_r} d_{ij} \quad (11)$$

离差平方和 $D(G_p, G_q)$ 法计算公式为:

$$D(G_p, G_q) = \|\bar{X}_p - \bar{X}_q\|^2 / (1/n_p + 1/n_q) \quad (12)$$

2.3 K-means 聚类法

K-means 聚类法的准则函数 F 计算公式为:

$$F = \sum_{x \in X} \min \|x - C_i\|^2 \quad (13)$$

式中, X 为样本集合; x 为类中样本位置矢量; C_i 为类的中心点位置矢量。

3 大渡河流域融雪期划分方法

3.1 融雪期划分结果

3.1.1 集对分析法

以旬为单位, 选用丹巴以上流域 2009~2020 年 11 月上旬~次年 5 月下旬反映积融雪变化规律的主要特征指标多年旬平均值作为集对分析法的聚类输入项, 即流量(X_1)、雨量(X_2)、平均气温(X_3)、最高气温(X_4)、最低气温(X_5)、地表太阳辐射(X_6)、日照时数(X_7)共计 7 个指标($m=7$), 计算得到各时段指标特征值, 部分值见表 1。

表 1 融雪期划分指标特征值

Tab. 1 Characteristic values of indicators for snow melting period division

集 时段	流量	降雨 /mm	平均气温/°C	最高气温/°C	最低气温/°C	地表太阳辐射	日照时数/h
A ₈ 1月中	252	0.2	-2.6	7.3	-10.7	172	6.0
A ₉ 1月下	240	0.1	-1.4	8.6	-9.8	184	6.1
A ₁₀ 2月上	235	0.1	0.6	10.7	-7.7	197	5.9
A ₁₁ 2月中	207	0.2	1.2	10.6	-6.7	213	5.6
A ₁₂ 2月下	204	0.3	2.6	12.3	-5.6	231	6.2

注: 流量、地表太阳辐射单位分别为 m^3/s 、 W/m^2 。

构建融雪期分期指标体系 $X = (X_{i,j})$, $X_{i,j}$ 为第 i 旬第 j 个指标, $i=1, 2, \dots, 21$, $j=1, 2, \dots, 7$ 。

基于表 2 定义的融雪期划分标准, 对表 1 中各指标值是否位于主融雪径流期界限范围内进行判别, 若是则标记为 2, 反之标记为 1, 如此经符号量化后得到的各指标标记结果, 部分结果见表 3。

表 2 融雪期指标划分标准

Tab. 2 Classification standard of indicators in snowmelt period

类别	X_1	X_2	X_3	X_4	X_5	X_6	X_7
融雪径流期	[200,300]	[0.1,1.0]	[0,15]	[10,20]	[-10,5]	[170,300]	[5.5,6.5]

计算 $A_i (i=1, 2, \dots, 21)$ 与分期标准之间的联系度, 根据经验取值法可取 $i=0$, 将 i 值代入式

(5)、(6)计算联系系数,部分结果见表 4。

表 3 集合 A_i 的符号量化

Tab. 3 Symbol quantization of set A_i

集合	X_1	X_2	X_3	X_4	X_5	X_6	X_7
A_8	2	2	1	1	1	2	2
A_9	2	2	1	1	2	2	2
A_{10}	2	2	2	2	2	2	2
A_{11}	2	2	2	2	2	2	2
A_{12}	2	2	2	2	2	2	2

表 4 联系度、联系数及分期结果

Tab. 4 Contact degree, contact number and staging results

集合 时段	$\mu_{A_i-B_1}$			$\mu_{A_i-B_2}$			结果	
	a	b	联系数	a	b	联系数	所属标准	分期
A_8 1月中	0.43	0.57	0.43	0.57	0.43	0.57	B_2	融雪径流期
A_9 1月下	0.29	0.71	0.29	0.71	0.29	0.71	B_2	融雪径流期
A_{10} 2月上	0	1	0	1	0	1	B_2	融雪径流期
A_{11} 2月中	0	1	0	1	0	1	B_2	融雪径流期
A_{12} 2月下	0	1	0	1	0	1	B_2	融雪径流期

3.1.2 系统聚类法

同样选取 $X_1 \sim X_7$ 7 个因子作为聚类指标,采用欧氏距离计算点间距离,并分别以六种类间距离计算式(7)~(12)展开基于系统聚类分析法的流域枯季融雪期划分,谱系图见图 2,图 2 中 Y 坐标轴序号为 1~21,分别代表 11 月上旬~5 月下旬,共计 21 个旬。

本文选定系统聚类法分为三大类时为计算结果,其中位于中间时期的类簇即为融雪径流期时段。为将所有不同分期情况考虑在内,选定 6 种方法的并集作为最终融雪期分期结果。由图 2 可知,经系统聚类法分析得到的丹巴以上流域融雪期为 2 月 21 日~5 月 20 日。

3.1.3 K-means 聚类法

选取 $X_1 \sim X_7$ 7 个指标的多年旬平均值作为 K-means 聚类法的输入项。定义 K 的取值范围为 $[2, 10]$,最佳 K 值则通过计算轮廓系数 S_{C_i} ^[8] 选取,该指标取值在 $-1 \sim 1$ 之间,当轮廓系数达到最大值时,表示同类别最为接近,此时聚类效果最优,具体计算公式为:

$$S_{C_i} = (b_i - a_i) / \max(a_i, b_i) \quad (14)$$

式中, i 为样本点, $i = 1, 2, \dots, N$; a_i 为样本点 i 在同类别内到其他点的平均距离; b_i 为样本点到最近不同类别中样本的平均距离。

根据 K-means 聚类法方法原理,统计得 K-means 聚类法计算的融雪期划分结果见表 5,相应轮廓系数计算值见图 3。由图 3 可知,当聚类数 K 取 4 时,对应轮廓系数最大,故 K-means 聚类法划分的融雪期为(旬序号 8~19)1 月 11 日~5 月 10 日。

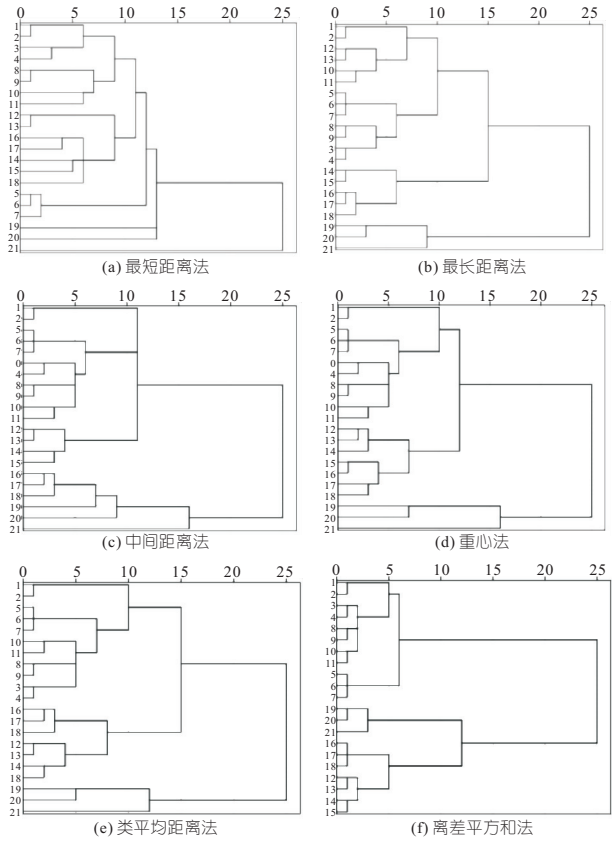


图 2 系统聚类法融雪期划分谱系图

Fig. 2 Classification pedigree of snowmelt period by systematic clustering method

表 5 K-means 聚类法融雪期划分结果统计

Tab. 5 Statistics of division results of snow melting period by K-means clustering method

聚类数	旬序号	日期	聚类数	旬序号	日期
2	2~19	11月11日~5月10日	7	9~16	1月21日~4月10日
3	4~18	12月1日~4月31日	8	10~16	2月1日~4月10日
4	8~19	1月11日~5月10日	9	11~15	2月11日~3月31日
5	9~17	1月21日~4月20日	10	14~16	3月11日~4月10日
6	8~16	1月11日~4月10日			

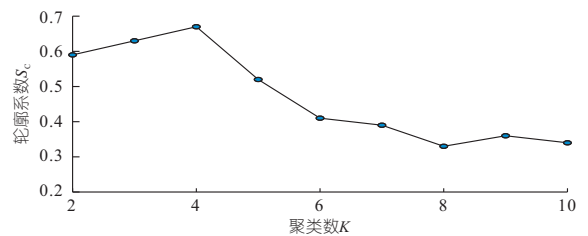


图 3 K-means 聚类法不同聚类数对应轮廓系数

Fig. 3 K-means clustering method contour coefficients corresponding to different cluster numbers

综合集对分析法、系统聚类法和 K-means 聚类法计算得到的丹巴以上流域融雪期分期结果见表 6。

3.2 分期结果合理性分析

3.2.1 实测降雨径流过程线分析

首先基于实测降雨径流过程线,并结合退水

表 6 融雪期划分结果

Tab. 6 Results of snow melting period

方法	融雪期划分日期
集对分析法	1月11日~5月10日
系统聚类法	2月21日~5月20日
K-means 聚类法	1月11日~5月10日

曲线对分期结果中的起始时间作进一步分析。根据流域 2009~2020 年 11 月~次年 5 月的实测降雨、流量数据绘制历年降雨径流过程线,取退水段根据最小二乘法计算退水系数,绘制退水曲线,再取流量过程线与退水曲线的分离点作为融雪期开始时间,以 2009 年 11 月 1 日~2010 年 5 月 31 日作为第一个场次,年份标注为 2010 年,统计结果见表 7。

表 7 退水曲线与流量过程线分离时间点统计表

Tab. 7 Statistical table of separation time point between recession curve and hydrograph

年份	日期	年份	日期	年份	日期
2010	1月7日	2014	1月5日	2018	1月15日
2011	1月10日	2015	1月16日	2019	1月9日
2012	1月19日	2016	1月3日	2020	1月8日
2013	1月1日	2017	1月17日		1月9日

取其中 4 场为例,其过程线见图 4。由图 4 可看出,丹巴以上流域枯季径流过程可大致分为三个阶段:前期为 11 月上旬~次年 1 月上旬,流域降水较少,实测径流过程与退水曲线拟合程度较高;中期为 1 月中旬~5 月上旬,流量值不再减小,径流过程线变缓,此时流域气温逐渐升高,表明开始有融雪径流产生;后期为 5 月中旬以后,流域内降雨明显增多,产生相应的降雨径流,流量值开始大幅增加。

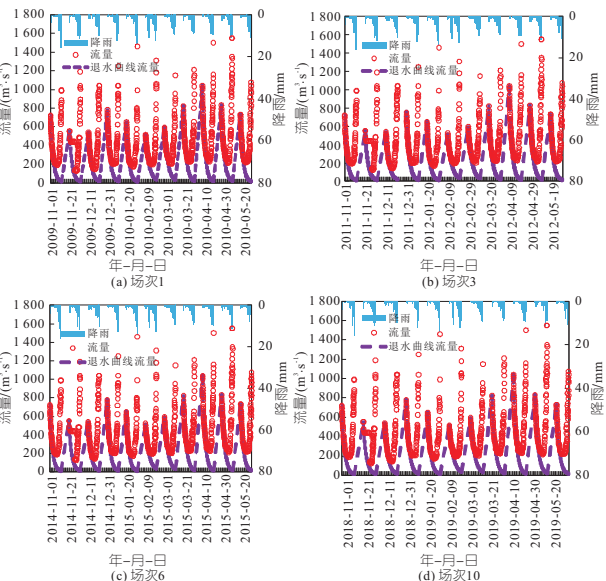


图 4 流量过程线与退水曲线分离点示例图

Fig. 4 Example diagram of separation points of flow hydrograph and recession curve

3.2.2 温度、积雪覆盖率与积雪深度变化分析

选用流域内 7 个气象站的日平均气温数据及欧洲中心天气预报中心的积雪深度再分析数据,在计算中,选用算数平均法统计流域内积雪深度日值,应用基于遥感监测技术的 MODIS 积雪产品。取积雪深度小于 10 mm 结束时间,积雪覆盖率小于 10% 结束时间及对应温度,部分结果见表 8。同时为更好了解这三种因子的变化趋势,对 2009~2020 年的数据计算多年平均值,结果见图 5。

表 8 基于积雪深度与积雪覆盖率的丹巴断面融雪期结束时间统计表

Tab. 8 Statistical table of end time of snowmelt in Danba section based on snow depth and snow cover

年份	积雪深度	积雪覆盖率	最早时间
	小于 10 mm	小于 10%	对应温度/℃
2009	04-20	06-15	9.7
2010	05-13	05-28	14.8
2011	05-17	06-02	10.5
2019	04-28	06-08	12.0
2020	05-04	06-07	13.5
平均值	05-09	05-31	12.2

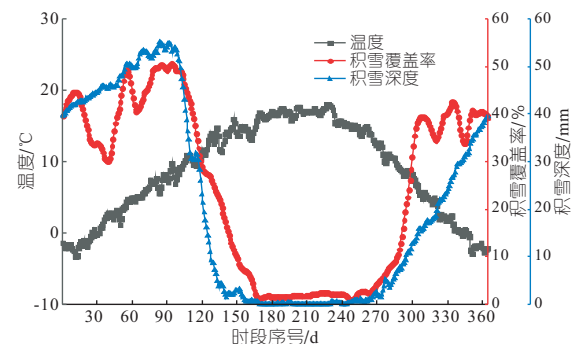


图 5 丹巴断面平均温度、积雪覆盖率、积雪深度多年平均值
Fig. 5 Multi year average value of Danba section average temperature snow cover and snow depth

由表 8、图 5 可知,丹巴以上流域主融雪径流期在 5 月 10 日左右结束,此后仍有部分融雪径流产生,但流量过程以降雨径流为主导。集对分析法与 K-means 聚类法得到的主融雪期结束时间为 5 月 10 日,而系统聚类法结束时间为 5 月 20 日,由于此前流域积雪深度已小于 10 mm,积雪覆盖率已小于 10%,无法产生较大的融雪径流。因此,集对分析法与 K-means 聚类法计算得到的融雪期结束时间更为合理。本文综合考虑枯季各项指标值与聚类分析的分期结果合理性,得到丹巴以上流域枯季融雪期的最终分期结果,即退水期为 11 月 1 日~1 月 10 日,融雪径流期为 1 月 11 日~5 月 10 日,降雨径流开始期为 5 月 11 日~5 月 31 日。

4 结论

a. 集对分析法得到的融雪径流分期为 1 月 11 日~5 月 10 日,系统聚类法的结果为 2 月 21

日~5月20日,K-means 聚类法的结果为1月11日~5月10日。集对分析法与 K-means 聚类法两种方法计算结果一致,系统聚类法的融雪期时段相对滞后。

b. 结合流域积雪深度与积雪覆盖率数据对分期结果的合理性分析表明,丹巴以上流域枯季径流过程可大致分为三个阶段,前期地表积雪增加,未产生融雪径流;中期积雪开始融化,此时以融雪径流为主;后期临近汛期,流域降水逐渐增多,流量增幅较大,以降雨径流为主,但仍有部分融雪径流。因此,最终确定的大渡河流域主融雪期为1月11日~5月10日。

c. 本文仅选取了11年枯季资料进行融雪期分析,结论具有一定的不确定性。未来随着资料的累积应做进一步的分析论证。

参考文献:

[1] 田琳. 中温带地区春季融雪径流研究[D]. 长春: 吉林

林大学,2019.

[2] AGGARWAL C C, REDDY C. Data clustering: algorithms and applications[M]. Chapman & Hall/CRC, 2013.

[3] 陈守煜, 赵瑛琪. 汛期的模糊集模式及其应用[J]. 水电能源科学, 1989,7(3): 211-219.

[4] 刘攀, 郭生练, 王才君, 等. 三峡水库汛期分期的变点分析方法研究[J]. 水文, 2005(1): 18-23.

[5] 谢飞, 王文圣. 集对分析在汛期分期中的应用研究[J]. 南水北调与水利科技, 2011,9(1): 60-63.

[6] 蒋海艳, 莫崇勋, 魏炜, 等. 灰色定权聚类法在水库汛期分期中的应用[J]. 水力发电, 2012,38(12): 8-10.

[7] 刘浅奎, 王文圣, 余思怡, 等. 熵权模糊集对分析法在溪洛渡水库汛期分期中的应用[J]. 水利水运工程学报, 2020(6): 16-22.

[8] KAUFMAN L, ROUSSEEUW P J. Finding groups in data: an introduction to cluster analysis [M]. John Wiley & Sons, 2009.

Classification of Snowmelt Periods in Daduhe River Basin

LI Bin-quan, RONG Wei-guang, WU Ya-qi, LIANG Zhong-min

(College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract: In order to accurately judge the start and end time of snowmelt period, the upstream of the Danba basin, the main source of snowmelt runoff in Daduhe River, was taken as an example. Based on hydrometeorological data from 2009 to 2020, set pair analysis method, systematic clustering method and K-means clustering method were used to calculate the snowmelt period in the dry season (November to the next May), and the rationality of the staging results was evaluated. The division scheme of final snowmelt period in the basin was determined. The results show that the main snowmelt period is from January 11 to May 10, the preceding period is from November 1 to January 10, and the following period is from May 11 to May 31, which is the beginning period of rainfall runoff (there is still some snowmelt runoff). The research results can provide reference for the classification of snowmelt periods in high latitude areas.

Key words: set pair analysis; hierarchical clustering; K-means clustering; snowmelt period; Dadu River Basin

(上接第 25 页)

Influence Analysis of Water Conservancy Engineering on Ecohydrological Regime in Qingjiang River Basin Based on Improved RVA Methods

DING Xin^{1a,1b}, CHEN Yan-fei^{1a,1b}, GUO Wei², YU Ting^{3,4}

(1a. Hubei Key Laboratory of Petroleum Geochemistry and Environment; 1b. College of Resources and Environment, Yangtze University, Wuhan 430100, China; 2. Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430000, China; 3. Hubei Water Resources Research Institute, Wuhan 430070, China; 4. Hubei Water Resources and Hydropower Science and Technology Promotion Center, Wuhan 430070, China)

Abstract: Scientific assessment of the hydrological regime is an important part of the assessment of the health of the river ecological environment. The IHA indicators were used to quantitatively evaluate the change of hydrological indicators of Gaobazhou Station during the construction period and operation period of water conservancy projects in the Qingjiang River Basin. On the basis of analyzing the change law of five different IHA indicators, the overall hydrological situation change of Qingjiang River was analyzed by selecting three different RVA methods, which include the Nemero index method, the RVA method based on the European distance method and the RVA method considering the comprehensive weight. It shows that the results of the analysis of the hydrological situation in the Qingjiang River Basin by the three methods during the construction period and the comprehensive utilization period are moderate and highly changed, respectively, and the hydrological situation change in the Qingjiang River Basin has an increasing trend. The evaluation results of the RVA method considering comprehensive weights are similar to the previous two methods and are reasonable, which can better reflect the changes in the river hydrological situation after the construction of the water conservancy project in the Qingjiang River Basin.

Key words: ecohydrological regime; RVA method based on Nemero exponential method; RVA method based on European distance method; improved RVA method; Qingjiang River Basin