

DOI: 10.20040/j.cnki.1000-7709.2023.20220723

# 基于系统聚类的分水江站水位分类预报和区间预报

姬战生<sup>1</sup>, 刘小勇<sup>2</sup>, 王 军<sup>3</sup>, 张振林<sup>1</sup>

(1. 杭州市水文水资源监测中心, 浙江 杭州 310016; 2. 江苏省水利工程科技咨询股份有限公司, 江苏 南京 210029; 3. 河海大学水文水资源学院, 江苏 南京 210098)

**摘要:** 研究分水江流域控制站——分水江站的水位预报方法, 提高预报精度, 对支撑区域防洪决策具有重要意义。利用分水江站 2012~2020 年典型历史洪水, 基于系统聚类分析方法, 以最大 1 d 洪量、洪峰流量、洪峰水位、洪水起涨历时占比作为聚类指标, 将洪水分为 I、II 类两类洪水, 采用流量演算—水位流量关系转换法对分水江站进行水位分类预报, 并构建不同类型洪水水位流量关系的上下边界, 提出一种分类洪水水位区间预报方法。结果表明, 两类洪水的洪峰水位预报精度有所提高, 预报区间宽度明显减小, 提高了区间预报结果的可利用性。

**关键词:** 分水江站; 系统聚类; 分类预报; 区间预报

**中图分类号:** [TV124]

**文献标志码:** A

**文章编号:** 1000-7709(2023)02-0011-04

## 1 引言

我国水旱灾害频发, 人民群众的生命财产安全受到严重威胁, 而精准的水文预报对防旱减灾具有极其重要的作用。随着洪水分类调度的深入研究, 采用分类和聚类预测洪水, 总结洪水规律, 提高洪水预报精度, 成为一种有效的实时洪水预报新方法<sup>[1]</sup>。传统水文预报多为流量预报, 如陈玲舫等<sup>[2]</sup>将分布式水文模型——流溪河模型用于松涛水库入库洪水预报中; 张宇等<sup>[3]</sup>采用贝叶斯预报系统(BFS)水文不确定性处理器(HUP)分析了水文预报不确定性, 实现概率洪水预报。然而, 对河道防洪而言, 水位抬升对堤防安全影响更直接<sup>[4]</sup>。常见的水位预报方法有相应水位法、统计相关法、经验分析法, 如胡炜等<sup>[5,6]</sup>分别采用统计相关方法、改进的最小二乘法预报太湖水位。然而, 水位预报相对流量预报研究成果偏少, 且在水位区间预报方面极少。鉴此, 本文以杭州市分水江流域控制站——分水江水文站为例, 提出一种基于聚类思想的水位分类预报和区间预报方法, 提高了水位预报精度, 亦为区域防洪决策提供了依据。

## 2 系统聚类法

### 2.1 基本思想

首先将  $n$  个样本看作一类, 规定各样本之间和类与类之间的距离; 然后将距离最近的 2 个样本划分为一个新的类, 重新计算新类与其他类的距离, 继续合并距离最近的两类, 直至全部样本完成归类, 最后形成具有亲疏关系的图谱, 关系图谱可清晰地反映分类结果, 即将样本分为几类、每个样本隶属于哪一类。

### 2.2 基本步骤

系统聚类法要同时定义样品之间的亲疏度及类与类之间的亲疏度, 采用 Matlab 对洪水样本分类, 基本步骤如下。

**步骤 1** 聚类指标确定。在聚类前, 需确定待聚类的  $n$  个样本及衡量聚类样本的特征、性质的指标, 即聚类指标。将  $n$  场洪水看作  $n$  个样本, 共提取  $m$  个聚类指标, 构成一个  $n \times m$  的指标矩阵  $\mathbf{X}$ :

$$\mathbf{X} = (x_{ik})_{n \times m} \quad (1)$$

式中,  $x_{ik}$  为第  $i$  场洪水的第  $k$  个指标。

**步骤 2** 样本距离定义。分类开始时, 将  $n$

**收稿日期:** 2022-04-12, **修回日期:** 2022-05-08

**基金项目:** 国家自然科学基金重点项目(41730750); 浙江省水利科技计划项目(RC1901, RB2102, RC2153); 杭州市科技计划引导项目(20211231Y086)

**作者简介:** 姬战生(1980-), 男, 硕士、高级工程师, 研究方向为水文监测预报预警, E-mail: 13514164@qq.com

**通讯作者:** 王军(1982-), 男, 博士、副教授, 研究方向为水文水资源, E-mail: wangjun\_hhu@163.com

个样本作为  $n$  个类, 计算  $n$  个样本间的距离。采用欧氏距离算法计算第  $i, j$  两个样本间的距离  $d_{ij}$  :

$$d_{ij} = \left[ \sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]^{1/2} \quad i, j = 1, 2, \dots, n \quad (2)$$

**步骤 3** 类间距离计算。定义样本距离后, 需定义类与类之间的距离, 即类间距离。常见的定义类间距离的方法有最长距离法、最短距离法、中间距离法、重心法、类平均距离法、离差平方和法等<sup>[7]</sup>。通过类间距离计算, 可实现洪水分类。

### 3 实例应用

#### 3.1 流域概况

分水江为钱塘江水系 1 级支流, 发源于安徽省绩溪县荆州乡方家湾村, 主流昌化溪在临安区潜川镇麻车埠村左纳天目溪后称分水江。流域地势西北部高、东南部低, 属于山丘区河流。流域总面积 3 443 km<sup>2</sup>, 主流长 167 km, 平均坡度 3.9‰。研究区域位于分水江流域中下游, 为分水江水库坝址至分水江水文站区间, 区间面积为 476 km<sup>2</sup>, 主流全长 12.07 km。流域内径流主要来源为分水江水库放水 and 区间降雨产流。

#### 3.2 洪水聚类分析

##### 3.2.1 聚类指标选取

洪水聚类分析需选取洪水样本和对应的聚类指标。为保证样本数据的一致性, 选取分水江站 2012~2020 年 12 场历史洪水作为样本, 记为  $Z_1, Z_2, \dots, Z_{12}$ 。为全面考虑洪水各特征要素对洪水分类和水位预报精度的影响, 选取最大 1 d 洪量、洪峰流量、洪峰水位、洪水起涨历时占比作为聚类指标。12 场洪水的聚类指标见表 1。

表 1 洪水样本聚类指标

Tab. 1 Cluster index of flood samples

样本编号	洪水编号	最大 1 d 洪量/10 <sup>8</sup> m <sup>3</sup>	洪峰流量/(m <sup>3</sup> · s <sup>-1</sup> )	洪峰水位/m	洪水起涨历时占比
#1	20130607	42.63	4 740	25.01	0.51
#2	20130627	39.25	4 720	24.95	0.55
#3	20150406	22.74	1 320	20.07	0.45
#4	20150608	27.30	2 580	21.72	0.43
#5	20150706	23.22	1 360	20.16	0.64
#6	20160407	15.12	1 290	19.91	0.45
#7	20160426	16.51	1 490	20.25	0.30
#8	20160529	23.16	2 840	22.11	0.44
#9	20170624	43.82	4 260	24.16	0.47
#10	20190713	57.82	4 790	24.75	0.35
#11	20190811	29.34	3 720	23.75	0.65
#12	20200707	60.51	4 730	24.96	0.47

##### 3.2.2 洪水聚类分析结果

对各场洪水的聚类指标预处理, 并进行聚类

分析, 聚类结果见图 1。由图 1 可知, 洪水可分为两类(定义为 I、II 类), 其中, # 1、# 2、# 9、# 10、# 12 洪水属于 I 类, # 3、# 4、# 5、# 6、# 7、# 8、# 11 洪水属于 II 类, 所有方法的相关系数均在 0.8 以上, 满足关联度要求。

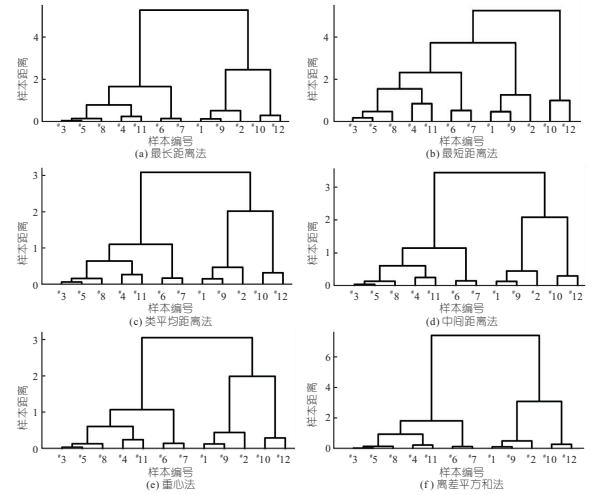


图 1 分水江站 12 场洪水聚类分析结果

Fig. 1 Clustering results of 12 floods at Fenshuijiang station

##### 3.3 分类洪水水位预报

因控制河段受下游漂流清理河道、水草生长等因素影响, 分水江水文站断面水位流量关系不稳定, 不同类型洪水采用同一水位流量关系的水位预报精度不高。因此, 采用分类预报方法, 以提高分水江站水位预报的精度水平。

##### 3.3.1 分析方案

采用流量演算—水位流量关系转换法分别预报两类洪水的水位。由于分水江站水位流量关系不稳定, 为综合考虑上一年度、当年和多年综合河道断面及河床变化对水位流量关系的影响, 分别采用上一年水位流量关系、当年水位流量关系、综合水位流量关系三种方案预报水位, 并对比分析三种方案在不同类型洪水中的预报结果, 提出两类洪水适用的较优方案。

##### 3.3.2 预报精度对比

采用三种方案分别预报两类洪水水位, 预报精度统计分析结果见表 2。由表 2 可知, 采用系统聚类法划分的两类洪水中, 不同类别洪水预报适用的水位流量关系不同。其中, I 类洪水采用综合水位流量关系预报水位的精度较高, 对应洪峰水位的绝对误差均值和绝对误差平均值的均值均最小, 分别为 0.13、0.50 m; II 类洪水采用当年水位流量关系预报水位的精度较高, 对应洪峰水位的绝对误差均值和绝对误差平均值的均值均最小, 分别为 0.13、0.55 m。究其原因因为 I 类洪水流量大、水位高, 过流断面面积大, 河床水草变化、

表 2 各方案预报精度对比

Tab. 2 Comparison of prediction accuracy of each scheme

洪水类别	洪水场次	洪峰流量 ( $m^3 \cdot s^{-1}$ )	洪峰水位/m				洪峰水位绝对误差/m			绝对误差平均值/m		
			实测	上年	当年	综合线	上年	当年	综合线	上年	当年	综合线
I 类洪水	20170624	4 260	24.16	22.97	23.99	24.41	-1.19	-0.17	0.25	0.50	0.45	0.47
	20190713	4 790	24.75	24.40	24.33	24.46	-0.35	-0.42	-0.29	0.62	0.59	0.59
	20130627	4 720	24.95	23.45	25.54	25.27	-1.50	0.59	0.32	0.63	0.59	0.50
	20200707	4 730	24.96	24.77	26.09	25.07	-0.19	1.13	0.11	0.35	0.45	0.34
	20130607	4 740	25.01	23.46	25.55	25.28	-1.55	0.54	0.27	0.79	0.71	0.61
	均值	4 650	24.77	23.81	25.10	24.90	-0.96	0.33	0.13	0.58	0.56	0.50
II 类洪水	20160407	1 290	19.91	20.27	20.05	20.33	0.36	0.14	0.42	0.66	0.55	0.63
	20150406	1 320	20.07	20.54	20.52	20.61	0.47	0.45	0.54	0.57	0.55	0.54
	20150706	1 360	20.16	20.46	20.45	20.53	0.30	0.29	0.37	0.45	0.43	0.42
	20160426	1 490	20.25	20.52	20.29	20.61	0.27	0.04	0.36	0.72	0.62	0.68
	20150608	2 580	21.72	21.79	21.78	22.00	0.07	0.06	0.28	0.48	0.47	0.45
	20160529	2 840	22.11	22.53	21.92	22.82	0.42	-0.19	0.71	0.49	0.38	0.46
	20190811	3 720	23.75	23.68	23.84	23.86	-0.07	0.09	0.11	0.85	0.83	0.85
	均值	2 090	21.14	21.40	21.26	21.54	0.26	0.13	0.40	0.60	0.55	0.58

自然冲淤或程度较轻的人工清理河道对过流断面影响占比较小,因此采用综合水位流量关系线更能反映断面的实际情况;相应地,II类洪水流量相对较小、水位低,过流面积小,河床水草变化、自然冲淤或程度较轻的人工清理河道对过流断面影响占比较大,采用当年的水位流量关系更符合实际情况。因此,建议 I 类洪水采用综合水位流量关系预报水位,II 洪水采用当年水位流量关系预报水位。

### 3.4 分类洪水水位区间预报

分类洪水水位预报在一定程度上提高了水位预报精度,但受多种不确定性因素的影响,部分场次洪水洪峰水位预报误差仍在 0.50 m 以上,精度难以满足生产实际的需要。本文提出分类洪水水位区间预报,既可进一步提高水位预报的精度,又提供预报水位的上下限。

#### 3.4.1 区间预报边界确定

洪水分类前,根据分水江站历史水位流量资料,采用最小二乘法拟合出上下边界的水位流量关系式,确定同一流量对应的最高与最低水位,作为水位区间预报的上下边界,见图 2。

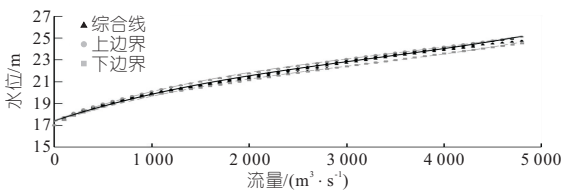


图 2 分水江站水位流量关系综合线及上下边界

Fig. 2 Comprehensive line and upper and lower boundary of water level and flow relation of Fenshuijiang station

洪水分类后,针对两类洪水,拟定不同的水位流量关系上下边界(图 3)。其中, I 类洪水,综合

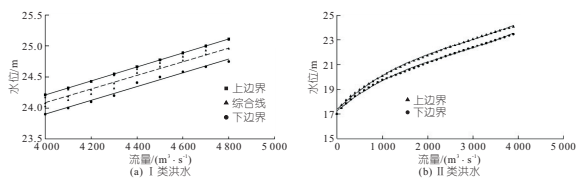


图 3 两类洪水新的上下边界水位流量关系

Fig. 3 New water level and flow relationship of upper and lower boundary for two types of floods

水位流量关系及上下边界水位流量关系采用分水江站历年洪峰流量在 4 000  $m^3/s$  以上的洪水过程重新拟定;II 类洪水,水位预报的上下边界采用分水江站历年洪峰流量在 4 000  $m^3/s$  以下的洪水过程重新拟定。

#### 3.4.2 水位区间预报

根据图 2、3 确定的分水江站上下边界的水位流量关系,对 12 场洪水进行水位区间预报。洪水分类前后各场次洪水水位区间预报结果见表 3。由表 3 可知,分类预报前,各场洪水洪峰水位预报绝对误差均值为 0.35 m,范围为 -0.46~0.71 m,预报区间宽度为 0.61~2.03 m。分类预报后, I 类洪水的洪峰水位预报绝对误差均值由 0.28 m 减至 0.10 m,范围由 -0.46~0.32 m 减至 -0.32~0.11 m,预报区间宽度由 1.78~2.03 m 减至 0.32~0.33 m,均值由 1.92 m 减至 0.32 m,平均减小 83.1%;II 类洪水预报绝对误差均值由 0.40 m 减至 0.18 m,范围由 0.11~0.71 m 减至 -0.19~0.45 m,预报区间宽度由 0.61~1.61 m 减至 0.49~0.89 m,均值由 0.94 m 减至 0.60 m,平均减小 29.2%。综上所述,洪水分类后,水位预报精度显著提升,洪峰水位预报区间宽度明显减小,水位预报区间更接近实测值,大大提高了预报结果的可利用性。

表 3 两类洪水洪峰水位分类区间预报结果

Tab. 3 Classified interval forecast results for two types of flood peak water level

洪水类别	洪水场次	实测值/m	绝对误差/m		预报区间/m		预报区间宽度			
			分类前	分类后	分类前	分类后	分类前/m	分类后/m	减小值/m	减小百分比/%
I类洪水	20170624	24.16	0.25	0.11	[22.89,24.67]	[24.07,24.39]	1.78	0.32	1.46	82.0
	20190713	24.75	-0.46	-0.32	[22.92,24.72]	[24.11,24.43]	1.80	0.32	1.48	82.2
	20130627	24.95	0.32	-0.02	[23.51,25.54]	[24.74,25.07]	2.03	0.33	1.70	83.7
	20200707	24.96	0.11	0.00	[23.36,25.33]	[24.60,24.92]	1.97	0.32	1.65	83.8
	20130607	25.01	0.27	-0.07	[23.52,25.54]	[24.75,25.08]	2.02	0.33	1.69	83.7
	均值		0.28	0.10			1.92	0.32	1.60	83.1
II类洪水	20160407	19.91	0.42	0.14	[19.98,20.59]	[20.04,20.53]	0.61	0.49	0.12	19.7
	20150406	20.07	0.54	0.45	[20.21,20.88]	[20.27,20.79]	0.67	0.52	0.15	22.4
	20150706	20.16	0.37	0.29	[20.14,20.80]	[20.20,20.72]	0.66	0.52	0.14	21.2
	20160426	20.25	0.36	0.04	[20.21,20.88]	[20.27,20.79]	0.67	0.52	0.15	22.4
	20150608	21.72	0.28	0.06	[21.28,22.31]	[21.32,22.02]	1.03	0.70	0.33	32.0
	20160529	22.11	0.71	-0.19	[21.81,23.12]	[21.93,22.69]	1.31	0.76	0.55	42.0
	20190811	23.75	0.11	0.09	[22.52,24.13]	[22.86,23.75]	1.61	0.89	0.72	44.7
	均值		0.40	0.18			0.94	0.60	0.31	29.2

### 4 结论

a. 以分水江站为例,采用系统聚类分析方法将洪水划分为两类,提出了一种水位分类预报和分类区间预报方法,在一定程度上克服了水位流量关系不稳定对水位预报精度的影响。

b. 本文方法适用于水位流量关系受冲淤或水草生长影响的山溪性河流水文测站水位预报。

c. 水位预报不确定性来源众多,未来可通过概率预报模型,进一步提高水位预报精度。

### 参考文献:

[1] 张改红,赵凡,牛云峰. 模糊数学在洪水分类预报中

的应用[J]. 人民黄河, 2010, 32(10): 65-67.

[2] 陈玲舫,陈洋波,李龙兵,等. 松涛水库入库洪水预报流溪河模型研究[J]. 水电能源科学, 2021, 39(7): 81-85.

[3] 张宇,梁忠民. BFS在洪水预报中的应用研究[J]. 水电能源科学, 2009, 27(5): 44-47.

[4] 戴清. 天然河道水位预报新方法初步研究[J]. 中国水利, 2010(13): 60, 62.

[5] 胡炜,董增川. 统计相关分析在太湖流域水位预报中的应用[J]. 水利经济, 2008(3): 55-58, 78.

[6] 赵兰兰,赵兵,阮燕云,等. 改进的偏最小二乘回归法在太湖水位预报中的应用[J]. 水电能源科学, 2021, 39(12): 44-46.

[7] 黄贤超. 基于聚类分析的我国 GDP 研究[D]. 济南: 山东大学, 2019.

## Classification Forecast and Interval Forecast for Water Level of Fenshuijiang Station Based on Systematic Clustering

JI Zhan-sheng<sup>1</sup>, LIU Xiao-yong<sup>2</sup>, WANG Jun<sup>3</sup>, ZHANG Zhen-lin<sup>1</sup>

(1. Hangzhou Hydrology and Water Resources Monitoring Center, Hangzhou 310016, China; 2. Hydraulic Engineering Science and Technology Consultation Limited Company of Jiangsu Province, Nanjing 210029, China;

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

**Abstract:** The research on water level prediction method of Fenshui River control station in Fenshui River basin that improves the prediction accuracy is of great significance to support regional flood control decision-making. The typical historical floods of Fenshuijiang hydrology station from 2012 to 2020 were selected and classified into category I and II flood based on systematic cluster analysis method, with the maximum 1-day flood volume, peak discharge, peak water level and the proportion of flood initiation duration as cluster indexes. The water level classification and prediction of water level stations were carried out by flow routing and water level and flow relation conversion method. The upper and lower boundaries of the water level and flow relationship of different flood types were constructed, and a classified interval prediction method of flood level was proposed. The results show that the accuracy of flood peak water level prediction for the two types of floods is improved effectively, the width of forecast interval is obviously reduced, and the availability of forecast results is improved.

**Key words:** Fenshuijiang station; systematic clustering; classified forecast; interval forecast