Journal of Water Resources Research 水资源研究, 2026, 15(4), 388-399
Published Online August 2026 in Hans. https://www.hanspub.org/journal/jwrr
https://doi.org/10.12677/jwrr.2026.154042

Hans汉斯

灰色关联度法与加权距离系数法在洪水相似性识别中的适用性分析

魏丹丹[1]，左雪飞[1]，李　念[1]，谈　新[1]，唐双英[1]，马丽梅[2]

[1]湖北省孝感市水文水资源勘测局，湖北 孝感
[2]湖北省水利水电科学研究院，湖北 武汉

收稿日期：2026年6月24日；录用日期：2026年8月16日；发布日期：2026年8月27日

摘　要

以澴水花园水文站46场历史洪水为研究对象，筛选10个易量化、易获取的暴雨洪水特征指标，开展基于1013种指标组合(不包含单指标组合)的洪水相似性研究。建立以专家经验识别为基准、以匹配率和洪峰流量合格率为量化指标的评定标准，采用灰色关联度法和加权距离系数法进行相似洪水识别，并引入差分进化算法优化加权距离系数法权重。结果表明：降雨总量P、前期影响雨量Pa、洪峰流量Q、洪量W是识别相似洪水的关键指标；差分进化算法优化权重后，加权距离系数法识别效果显著提升；灰色关联度法识别结果更为稳健，对指标组合变化不敏感；仅采用事前指标的识别效果明显不如全过程指标。

关键词

暴雨洪水特征指标，洪水相似性，灰色关联度，加权距离系数，差分进化算法，匹配率

Analysis on the Applicability of Grey Relational Analysis and Weighted Distance Coefficient Methods in Flood Similarity Identification

Dandan Wei[1], Xuefei Zuo[1], Nian Li[1], Xin Tan[1], Shuangying Tang[1], Limei Ma[2]

[1]Xiaogan Hydrology and Water Resources Survey Bureau of Hubei Province, Xiaogan Hubei
[2]Hubei Water Resources Research Institute, Wuhan Hubei

Received: June 24, 2026; accepted: August 16, 2026; published: August 27, 2026

Abstract

Using 46 historical flood events recorded at the Huayuan Hydrological Station on the Huanshui River as

作者简介：魏丹丹(1984-)，女，高级工程师，从事水文预报研究，Email: 396987750@qq.com

文章引用: 魏丹丹, 左雪飞, 李念, 谈新, 唐双英, 马丽梅. 灰色关联度法与加权距离系数法在洪水相似性识别中的适用性分析[J]. 水资源研究, 2026, 15(4): 388-399. DOI: 10.12677/jwrr.2026.154042

the study object, this study selected ten quantifiable and readily available rainfall-flood indicators and conducted a flood similarity analysis based on 1013 indicator combinations (excluding single-indicator combinations). An evaluation criterion was established using expert experience-based identification as the benchmark, with the matching rate and the qualified rate of peak discharge as quantitative metrics. Similar floods were identified using the grey relational analysis (GRA) method and the weighted distance coefficient (WDC) method; the differential evolution (DE) algorithm was introduced to optimize the weights of the latter. The results show that total rainfall (*P*), antecedent precipitation index (*Pa*), peak discharge (*Q*), and flood volume (*W*) are key indicators for identifying similar floods. After the weights were optimized by the DE algorithm, the performance of the WDC method improved markedly. The GRA method produced more robust results than the WDC method and was less sensitive to variations in indicator combinations. Using only pre-event indicators led to markedly poorer identification performance than using full-process indicators.

Keywords

Characteristic Indices of Rainstorm Flood, Flood Similarity, Grey Relational Analysis Method, Weighted Distance Coefficient, Differential Evolution Algorithm, Matching Rate

Copyright © 2026 by author(s) and Wuhan University & Bureau of Hydrology, Changjiang Water Resources Commission. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

暴雨洪水的形成与演变并非完全随机，而是具有一定规律性；历史洪水资料中，不同场次洪水过程常在若干特征上表现出相似性。研究这种相似性对实时洪水预报及调度决策具有重要的借鉴价值。然而，相似洪水识别精度高度依赖于特征指标的选择与度量方法的适配性，如何科学选取关键指标、合理确定相似性度量方法，是亟待深入研究的问题。

在特征指标筛选方面，国内学者已开展了大量研究。刘卫林等[1]从天气成因、暴雨过程及洪水过程三个方面构建指标体系；李春红等[2]将指标分为可量化和不可量化两类，指出不可量化指标筛选后会导致洪水样本迅速减少；王海潮等[3]按暴雨洪水发展阶段划分为 4 大类，选取时段累积降雨量、最高水位及其对应流量、累积洪量作为相似性分析指标；陈建等[4]在此基础上进一步细化为 7 大类共 67 个指标，经聚类分析和物理概念分析，优选出降水量、降水历时、起涨流量和洪水涨率 4 个指标用于相似洪水判别。

相似洪水识别的核心在于科学度量两场洪水间的相似程度。根据输入数据类型的不同，现有方法大致归为两类：一类是基于特征指标的相似性度量，将洪水抽象为特征向量，通过距离或关联度度量相似性，应用最广泛，如加权距离系数法[3] [5] [6]、灰色关联度法[7]-[9]、主成分分析法[10] [11]、欧氏距离法[12]等；另一类是基于过程序列，直接利用流量过程线，通过序列匹配算法度量形状相似性，以动态时间扭曲算法(*DTW*) [13]-[16]为代表。此外，还有一类方法通过聚类分析将特征相近的洪水归为一类，从群体角度揭示洪水分类结构，如 *K* 均值聚类法[17] [18]、模糊聚类法[1]、层次聚类法等。

上述方法各有侧重，在实际应用中可根据研究目的、数据特征和精度要求选择合适的方法或组合使用。本研究选取灰色关联度法与加权距离系数法两种方法相互验证，以提高相似洪水识别结果的可靠性。

2. 研究区域概况

澴水系长江中下游北岸支流府澴河一级支流，全长 150.8 km，流域面积 3618 km^2。花园水文站位于湖北省

孝感市孝昌县花园镇，为澴水中下游主要控制站，集雨面积 2601 km^2，主河道长 95.7 km，河道平均坡降 0.97‰。流域内有 7 座大中型水库，总库容 2.42 亿 m^3，调洪库容 0.73 亿 m^3，基本建成于六七十年代，总集雨面积 255 km^2，约占花园站流域面积的 10%，拦蓄作用有限。流域位于亚热带季风气候区，多年平均降雨量 1110 mm 多集中在 5~9 月，年内分配不均，年际变化较大。流域洪水多由暴雨形成，具有峰高量大、涨落迅速的特点。花园站洪水基本为天然洪水，受水库调蓄影响较小。

花园站是府澴河流域防洪体系的关键控制站，其预报流量直接作为下游孝感站和卧龙潭站相关图预报方案的输入条件。花园站相似洪水研究能提高该站洪水预报精准度，进而为下游站预报提供更可靠的输入，对孝感城区防洪调度决策和防洪安全具有重要意义。

3. 洪水特征指标筛选与数据标准化处理

3.1. 数据资料

本研究选取花园站 1995~2024 年共 46 场历史洪水作为样本。样本包含大、中、小洪水，代表性较好。场次洪水的降雨、水位、流量资料均经过湖北省水文水资源中心整编，成果规范可靠，为相似洪水研究提供了良好的数据基础。

3.2. 暴雨洪水特征指标筛选

暴雨洪水特征指标多达几十种，但选取过多指标会增加计算量和分析问题的复杂性。不同学者出于不同的研究目的，选择的特征指标也不尽相同。有的研究侧重于事后分析，选取洪水全过程发生后的特征指标[10] [11] [15] (简称“全过程指标”)，有的研究面向实时洪水预报，选取洪峰出现前可获取的特征指标[2]-[4] [6]-[9] (简称“事前指标”)。

本研究在前人研究的基础上，从天气成因、暴雨过程、洪水过程三个维度构建降雨洪水特征指标体系。天气成因指标反映洪水的天气背景与环流形势；暴雨指标按发展阶段分为暴雨发生发展指标和暴雨过程结束指标；洪水指标按发展阶段分为洪水发生发展指标和洪水过程结束指标。各项指标进一步区分为易量化指标与不易量化指标，共同构成多层次的相似性分析指标体系(见表 1)。

表 1. 暴雨洪水相似性分析指标体系

指标分类		指标类型	指标名称
天气成因指标		不易量化	环流背景、天气系统、对流云图、物理量场、暴雨成因
暴雨指标	暴雨发生发展指标	易量化	时段降雨量、累积时段降雨量
		不易量化	暴雨中心、降水移动方向
	暴雨过程结束指标	易量化	降雨历时、降雨总量、最大雨强、平均雨强、暴雨等级、雨型系数
		不易量化	暴雨发展过程、暴雨笼罩范围、降雨形状、降雨分布特点
洪水指标	洪水发生发展指标	易量化	起涨流量、起涨水位、T 时段内水位涨幅、T 时段内流量涨差、T 时段内洪水涨率、逐时段流量、时段洪量、累积时段洪量
	洪水过程结束指标	易量化	洪量、洪峰流量、洪峰水位、涨水历时、退水历时、洪水总历时、水位最大涨幅、流量涨差、洪水涨率、雨期平均流量、峰前洪量占比、洪峰时间偏度、退水速率、径流深、径流系数、径流模数、洪水量级、洪水重现期、超特征水位时长
		不易量化	洪水组成、洪水峰型
其他指标		易量化	前期影响雨量、雨洪峰值时间差、上游来水流量
		不易量化	回水顶托影响、水利工程调度影响

本研究遵循两个原则选取特征指标：① 对洪水过程有显著影响；② 易量化、易获取。据此筛选出 10 个特征指标，分别为：降雨总量 P、前期影响雨量 Pa、降雨历时 Tp、最大雨强 i、起涨流量 $Q0$、洪峰流量 Q、洪量 W、涨水历时 Tq、水位最大涨幅 ΔZ、洪水涨率 $\Delta Q/Tq$。其中，前 5 个为事前指标。需要指出的是，上述指标间存在不同程度的相关性。基于花园站 46 场洪水样本计算，洪峰流量与降雨总量、洪量、水位最大涨幅、洪水涨率存在较高的正相关性，相关系数分别为 0.824、0.902、0.923、0.819。有的学者认为高相关指标会造成信息重叠，导致计算结果失真[10]。然而，剔除高相关指标对相似洪水识别效果的影响尚缺乏系统验证。本文将在后续章节中结合具体方法对此进行对比分析。

3.3. 数据预处理

为消除量纲影响，采用均值化、Z-score 标准化和极差标准化三种标准化法对原始数据进行预处理。

1) 均值化

均值化处理是将各指标的原始数据除以该指标的均值。处理后，各指标的均值为 1。本次用于灰色关联度分析中的无量纲化处理。

$$x'_{ij} = \frac{x_{ij}}{\overline{x}_j} \tag{1}$$

式中，$i = 1,2,\cdots,n$ 为洪水场次(n = 46)；$j = 1,2,\cdots,k$ 为特征指标(k = 10)；x_{ij} 为第 i 场洪水第 j 个特征指标的原始值；$\overline{x}_j$ 为第 j 个洪水特征指标的均值；x'_{ij} 为 x_{ij} 标准化后的值。

2) Z-score 标准化

Z-score 标准化将数据转换为均值为 0、标准差为 1 的形式，保留了原始数据的离散程度，受异常值影响较小，本次用于加权欧氏距离计算。

$$x'_{ij} = \frac{x_{ij} - \overline{x}_j}{S_j} \tag{2}$$

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{3}$$

$$S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2} \tag{4}$$

式中，S_j 为第 j 个洪水特征指标的标准差；其他符号意义同上。

3) 极差标准化

极差标准化将数据线性变换到[0, 1]区间，压缩量纲差异，对异常值敏感，本次用于加权距离系数法的权重计算。

$$x'_{ij} = \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} \tag{5}$$

式中，$x_{j,\min}$ 和 $x_{j,\max}$ 分别是第 j 个特征指标的最小值和最大值；其他符号意义同上。

经均值化、Z-score 标准化、极差标准化处理后的数据矩阵分别为 X_1、X_2、X_3，均为 $n \times k$ 阶矩阵。

4. 相似洪水识别方法

4.1. 灰色关联度法

灰色关联度法通过计算母序列(目标洪水)与子序列(待比较洪水)的关联度来度量两者的相似程度。采用均值化后的数据矩阵 X_1，选取任意一场洪水作为母序列 $X_0 = \left(x_{01}, x_{02}, \cdots, x_{0k}\right)$，其他场次洪水作为子序列

$X_i=(x_{i1},x_{i2},\cdots,x_{ik})$，计算母序列与各子序列每个指标的绝对差：

$$\Delta_{0i}(j)=\left|x_{0j}-x_{ij}\right| \tag{6}$$

式中，x_{0j}、x_{ij}分别为母序列和第 i 场洪水的第 j 个指标；$\Delta_{0i}(j)$为两者之差的绝对值。

关联系数 $\xi_{0i}(j)$ 计算公式如下：

$$\xi_{0i}(j)=\frac{\Delta_{\min}+\rho\Delta_{\max}}{\Delta_{0i}(j)+\rho\Delta_{\max}} \tag{7}$$

式中，$\Delta_{\max}$ 和 $\Delta_{\min}$ 分别为绝对差 $\Delta_{0i}(j)$ 的最大值和最小值，ρ 为分辨系数，一般取 $\rho=0.5$。

各特征指标关联系数的均值即为关联度：

$$\gamma_{0i}=\frac{1}{k}\sum_{j=1}^{k}\xi_{0i}(j) \tag{8}$$

式中，γ_{0i} 为第 i 场洪水与母序列的关联度，关联度最大值对应的场次洪水，即为与目标洪水最相似洪水。

4.2. 加权距离系数法

加权距离系数法通过计算两场洪水在特征空间中的加权欧氏距离度量其相似程度。距离越小二者越相似。

采用 Z-score 标准化后的数据矩阵 X_2，计算母序列与各子序列之间的加权欧氏距离 d_{0i}：

$$d_{0i}=\sqrt{\sum_{j=1}^{k}\beta_j\left(x_{0j}-x_{ij}\right)^2} \tag{9}$$

式中，β_j 为第 j 个指标的权重，满足 $\sum_{j=1}^{k}\beta_j=1$；其他符号意义同上。

将加权欧氏距离转换为相似度 S_{0i}：

$$S_{0i}=\frac{1}{1+d_{0i}} \tag{10}$$

S_{0i} 越大，两场洪水越相似。相似度最大值对应的场次洪水即为最相似洪水。

指标权重确定方法主要分为主观赋权法和客观赋权法[19],常用的有专家打分法、层次分析法、变异系数法、熵权法等。为避免主观赋权的人为偏差，本研究采用变异系数法确定各指标权重，并引入差分进化算法进一步优化，以提升相似洪水识别精度。

采用极差标准化后的数据矩阵 X_3，计算各指标的变异系数 Cv_j：

$$Cv_j=\frac{S_j}{\bar{x}_j} \tag{11}$$

将变异系数归一化即得到权重 β_j：

$$\beta_j=\frac{Cv_j}{\sum_{j=1}^{k}Cv_j} \tag{12}$$

符号意义同上。

差分进化算法(*DE*)是一种基于群体差异的随机优化算法，通过变异、交叉、选择三个操作迭代寻优。其核心是利用种群中两个个体的差分向量扰动第三个个体生成变异个体，再与目标个体交叉生成试验个体，并通过贪婪策略保留更优个体，迭代至收敛。

本研究以相似洪水匹配率最大为目标函数，在[0, 1]区间内对各指标权重进行全局寻优。算法参数设置为：种群规模 $NP=20$，缩放因子 $F=0.5$，交叉概率 $CR=0.8$，最大迭代次数 100。

5. 相似洪水评定标准

现有研究对相似洪水识别结果的检验主要有三种方式：一是基于相似度阈值直接判定[2]-[4] [10]；二是以洪水预报精度(确定性系数、洪峰及洪量误差等)作为验证依据[7] [8] [12]；三是通过典型洪水个例直观对比[15] [16]。上述方法各有侧重，但均存在局限性。本研究在分析花园站历史洪水时发现，关联度或相似度高的洪水不一定真正相似，仅靠阈值判定不够可靠，需结合专家经验综合评定。

为评价两种方法的识别精度，本研究依据专家经验，综合对比洪水特征指标接近程度，人工判定每场目标洪水的相似洪水，作为基准参照。人工识别主要依据 6 个指标：Q、P、Pa、W、ΔZ、$\Delta Q/Tq$。人工识别结果见表 2，每场洪水匹配 1~3 场相似洪水。

在此基础上，进一步设置两项定量评价指标：相似洪水匹配率 R_M (算法识别的最相似洪水出现在人工相似洪水列表中的场次数占比)和洪峰流量合格率 R_Q (洪峰流量误差合格的场次数占比)。根据《水文情报预报规范》(GB/T 22482-2008)，洪峰流量预报许可误差一般为实测洪峰流量的 20%。该标准面向预报精度考核，而本研究旨在筛选量级相近的相似洪水，严格采用 20%会导致可匹配样本过少，故放宽至 30%，即相对误差绝对值小于 30%为合格。

6. 计算结果分析

6.1. 灰色关联度法计算结果

灰色关联度法识别步骤如下：① 计算 10 个指标 1013 种组合下每场洪水的最大关联度 γ_{0i}，并确定其对应的最相似洪水场次；② 计算最相似洪水与目标洪水的洪峰流量误差，小于 30%即为合格；③ 检查最相似洪水与人工识别结果是否匹配；④ 统计 1013 种指标组合下的相似洪水匹配率 R_M 和洪峰流量合格率 R_Q。

以 P、Pa、W、Q 这组指标为例，灰色关联度法在 46 场洪水中的相似洪水匹配率为 84.8%，洪峰流量合格率为 73.9%。

表 2. 灰色关联度法识别结果(以 P、Pa、W、Q 指标组合为例)

洪水场次	洪水代码	对应洪峰流量/$m^3 \cdot s^{-1}$	最大关联度/γ_{0i}	最相似洪水场次/次	对应洪峰流量/$m^3 \cdot s^{-1}$	洪峰流量误差绝对值/%	人工相似洪水	与人工结果是否匹配
1	*FA*19950708	3180	0.909	33	3110	2.2	6, 33, 19	1
2	*FA*19960703	3170	0.799	46	2190	30.9	34	0
3	*FA*19960708	1030	0.903	38	1200	16.5	38, 22	1
4	*FA*19970511	1170	0.886	5	1090	6.8	5, 11	1
5	*FA*19970607	1090	0.877	20	758	30.5	4, 11	0
6	*FA*19970701	3210	0.799	1	3180	0.9	1, 33, 19	1
7	*FA*19980412	1160	0.900	10	1130	2.6	10	1
8	*FA*19980502	681	0.934	12	366	46.3	16, 21	0
9	*FA*19980510	886	0.857	31	1600	80.6	31	1
10	*FA*19980601	1130	0.908	7	1160	2.7	7	1
11	*FA*19980703	1350	0.835	5	1090	19.3	4, 5	1
12	*FA*19980804	366	0.936	8	681	86.1	26, 44	0
13	*FA*19980808	684	0.903	21	690	0.9	21, 45, 17	1
14	*FA*19990628	526	0.896	24	495	5.9	24, 36	1

续表

15	*FA*20000602	519	0.905	20	758	46.1	20, 27, 30	1
16	*FA*20020506	800	0.885	14	526	34.3	35, 43	0
17	*FA*20020623	513	0.835	13	684	33.3	13, 21, 45	1
18	*FA*20020627	2920	0.803	42	3290	12.7	23, 42	1
19	*FA*20020724	3030	0.834	1	3180	5.0	1, 6, 33	1
20	*FA*20030623	758	0.927	27	924	21.9	15, 27, 30	1
21	*FA*20030627	690	0.906	13	684	0.9	13, 45, 8	1
22	*FA*20030701	1260	0.923	38	1200	4.8	38, 3	1
23	*FA*20030709	3000	0.755	41	1760	41.3	18, 42, 32	0
24	*FA*20030722	495	0.888	14	526	6.3	14, 36	1
25	*FA*20040718	1550	0.842	31	1600	3.2	31	1
26	*FA*20050830	357	0.883	44	349	2.2	44	1
27	*FA*20070531	924	0.922	20	758	18.0	20	1
28	*FA*20070709	1510	0.900	22	1260	16.6	29, 22, 38	1
29	*FA*20070714	1520	0.886	28	1510	0.7	28	1
30	*FA*20080420	611	0.893	20	758	24.1	15, 20	1
31	*FA*20080723	1600	0.852	25	1550	3.1	25	1
32	*FA*20080817	2430	0.824	41	1760	27.6	41	1
33	*FA*20080830	3110	0.913	1	3180	2.3	1, 19, 6	1
34	*FA*20100717	4040	0.796	39	4880	20.8	40, 39	1
35	*FA*20150617	903	0.875	43	835	7.5	43, 16	1
36	*FA*20150716	642	0.866	27	924	43.9	14, 24	0
37	*FA*20160620	986	0.804	27	924	6.3	27	1
38	*FA*20160624	1200	0.917	22	1260	5.0	3, 22	1
39	*FA*20160701	4880	0.865	40	4390	10.0	40, 34	1
40	*FA*20160720	4390	0.862	39	4880	11.2	39, 34	1
41	*FA*20200628	1760	0.831	32	2430	38.1	32	1
42	*FA*20200717	3290	0.788	18	2920	11.2	18, 23	1
43	*FA*20210515	835	0.884	35	903	8.1	16, 35	1
44	*FA*20210813	349	0.909	12	366	4.9	12, 26	1
45	*FA*20240705	729	0.880	21	690	5.3	13, 21	1
46	*FA*20240714	2190	0.785	2	3170	44.7	2, 33	1

对灰色关联度法 1013 种组合的相似洪水匹配率 R_M 从大到小排序，前 20 位见表 3。匹配率大于 80%的有 8 组，大于 70%的有 74 组。最佳指标组合为 *P*、*Pa*、*W*、*Q*、Δ*Q*/*Tq*，匹配率达 89.1%，洪峰流量合格率 82.6%。

表 3. 灰色关联度法——相似洪水匹配率前 20 位

排位	指标组合	特征指标个数/个	相似洪水匹配场次数/次	相似洪水匹配率/%	洪峰流量合格场次数/次	洪峰流量合格率/%
1	*P*, *Pa*, *W*, *Q*, Δ*Q*/*Tq*	5	41	89.1	38	82.6
2	*P*, *Pa*, *W*, *Q*	4	39	84.8	34	73.9
3	*P*, *Pa*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	6	38	82.6	37	80.4
4	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	8	38	82.6	36	78.3
5	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Q*/*Tq*	6	37	80.4	39	84.8
6	*Pa*, *Tp*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	6	37	80.4	37	80.4
7	*P*, *Pa*, *W*, *Q*, Δ*Z*, *Tq*, Δ*Q*/*Tq*	7	37	80.4	37	80.4
8	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Z*, *Tq*	8	37	80.4	36	78.3
9	*P*, *Pa*, *W*, *Q*, Δ*Z*, *Tq*	6	36	78.3	38	82.6
10	*P*, *Pa*, *W*, *Q*, Δ*Z*	5	36	78.3	37	80.4
11	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	7	36	78.3	37	80.4
12	*Pa*, *Tp*, *W*, Δ*Z*, Δ*Q*/*Tq*	5	36	78.3	36	78.3
13	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Z*	7	36	78.3	36	78.3
14	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Z*, *Tq*	7	36	78.3	36	78.3
15	*P*, *Pa*, *Q*	3	36	78.3	35	76.1
16	*P*, *Pa*, *W*, Δ*Z*, Δ*Q*/*Tq*	5	36	78.3	35	76.1
17	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Q*/*Tq*	7	36	78.3	35	76.1
18	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Z*, *Tq*, Δ*Q*/*Tq*	9	36	78.3	35	76.1
19	*P*, *Pa*, *Tp*, *W*, Δ*Z*, Δ*Q*/*Tq*	6	36	78.3	34	73.9
20	*P*, *Pa*, *Tp*, *W*, *Q*	5	36	78.3	33	71.7

6.2. 加权距离系数法计算结果

加权距离系数法识别步骤与灰色关联度法基本一致，区别在于其计算的是相似度 S_{0i}。该方法分别采用变异系数法权重和差分进化算法优化权重计算，并对相似洪水匹配率 R_M 从大到小排序。选取三组指标：① *P*、*Pa*、*W*、*Q*；② *P*、*Pa*、*W*、*Q*、Δ*Q*/*Tq*；③ *P*、*Pa*、*W*、*Q*、Δ*Q*/*Tq*、Δ*Z*，对比优化前后的权重及匹配率，见表 4。

表 4. 加权距离系数法权重优化前后的匹配率对比表

指标	第一组		第二组		第三组	
	优化前权重	优化后权重	优化前权重	优化后权重	优化前权重	优化后权重
P	0.2225	0.3054	0.1691	0.3151	0.1449	0.1741
Pa	0.1568	0.1238	0.1191	0.1292	0.1021	0.0718
W	0.2888	0.1093	0.2194	0.1000	0.1881	0.2689
Q	0.3319	0.4615	0.2523	0.4553	0.2162	0.2716
Δ*Q*/*Tq*	-	-	0.2400	0.0004	0.2057	0.1549
Δ*Z*	-	-	-	-	0.1430	0.0587
匹配率	82.6%	93.5%	80.4%	93.5%	80.4%	89.1%

权重优化前后，匹配率前 20 位分别见表 5 和表 6。

权重优化前：1013 种组合中匹配率大于 80%的有 6 组，大于 70%的有 19 组；最佳指标组合为 *P*、*Pa*、*W*、*Q*，匹配率 82.6%，洪峰流量合格率 80.4%。

权重优化后：匹配率大于 90%的有 31 组，大于 80%的有 286 组，大于 70%的有 502 组；最佳指标组合仍为 *P*、*Pa*、*W*、*Q*，匹配率 93.5%，洪峰流量合格率 89.1%。

对比发现，引入差分进化算法优化权重后，匹配率平均提升 27.5%，提升效果显著。

表 5. 加权距离系数法(权重优化前)——相似洪水匹配率前 20 位

排位	特征指标组合	特征指标个数/个	相似洪水匹配场次数/次	相似洪水匹配率/%	洪峰流量合格场次数/次	洪峰流量合格率/%
1	*P*, *Pa*, *W*, *Q*	4	38	82.6	37	80.4
2	*P*, *Pa*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	6	37	80.4	40	87.0
3	*P*, *Pa*, *W*, *Q*, Δ*Q*/*Tq*	5	37	80.4	38	82.6
4	*P*, *Pa*, *Q*	3	37	80.4	38	82.6
5	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	7	37	80.4	37	80.4
6	*P*, *Pa*, *Tp*, *W*, Δ*Z*, Δ*Q*/*Tq*	6	37	80.4	34	73.9
7	*P*, *Pa*, *Q*, Δ*Q*/*Tq*	4	36	78.3	37	80.4
8	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Q*/*Tq*	6	36	78.3	36	78.3
9	*P*, *Pa*, *W*, Δ*Q*/*Tq*	4	35	76.1	36	78.3
10	*P*, *Pa*, *W*, Δ*Z*, Δ*Q*/*Tq*	5	35	76.1	36	78.3
11	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Z*	6	35	76.1	36	78.3
12	*P*, *Pa*, *Q*0, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	7	34	73.9	37	80.4
13	*P*, *Pa*, *Tp*, *W*, *Q*	5	34	73.9	32	69.6
14	*Pa*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	5	33	71.7	39	84.8
15	*P*, *Q*0, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	6	33	71.7	39	84.8
16	*Pa*, *W*, *Q*	3	33	71.7	38	82.6
17	*Pa*, *Tp*, *W*, *Q*, Δ*Z*	5	33	71.7	37	80.4
18	*Pa*, *Tp*, *W*, *Q*, Δ*Z*, Δ*Q*/*Tq*	6	33	71.7	37	80.4
19	*P*, *Pa*, *Tp*, *Q*0, *W*, *Q*, Δ*Q*/*Tq*	7	33	71.7	37	80.4
20	*P*, *Pa*, *Q*, Δ*Z*, Δ*Q*/*Tq*	5	32	69.6	38	82.6

表 6. 加权距离系数法(权重优化后)——相似洪水匹配率前 20 位

排位	特征指标组合	特征指标个数/个	相似洪水匹配场次数/次	相似洪水匹配率/%	洪峰流量合格场次数/次	洪峰流量合格率/%
1	*P*, *Pa*, *W*, *Q*	4	43	93.5	41	89.1
2	*P*, *Pa*, *Tp*, *W*, *Q*	5	43	93.5	40	87.0
3	*P*, *Pa*, *W*, *Q*, Δ*Q*/*Tq*	5	43	93.5	41	89.1
4	*P*, *Pa*, *Tp*, *i*, *W*, *Q*	6	43	93.5	40	87.0
5	*P*, *Pa*, *Tp*, *W*, *Q*, Δ*Q*/*Tq*	6	43	93.5	41	89.1

续表

6	*P, Pa, Q0, W, Q, Tq*	6	43	93.5	41	89.1
7	*P, Pa, Tp, i, W, Q, Tq*	7	43	93.5	41	89.1
8	*P, Pa, Tp, Q0, W, Q, Tq*	7	43	93.5	39	84.8
9	*P, Pa, Tp, Q0, W, Q, ΔQ/Tq*	7	43	93.5	39	84.8
10	*P, Pa, i, Q0, W, Q, Tq*	7	43	93.5	41	89.1
11	*P, Pa, Tp, i, Q0, W, Q, Tq, ΔQ/Tq*	9	43	93.5	39	84.8
12	*P, Pa, Tp, Q, ΔQ/Tq*	5	42	91.3	40	87.0
13	*P, Pa, i, W, Q*	5	42	91.3	40	87.0
14	*P, Pa, Tp, Q0, Q, ΔQ/Tq*	6	42	91.3	40	87.0
15	*P, Pa, Tp, Q, Tq, ΔQ/Tq*	6	42	91.3	40	87.0
16	*P, Pa, W, Q, ΔZ, Tq*	6	42	91.3	40	87.0
17	*P, Pa, Tp, i, Q, ΔZ, ΔQ/Tq*	7	42	91.3	40	87.0
18	*P, Pa, i, W, Q, ΔZ, ΔQ/Tq*	7	42	91.3	40	87.0
19	*P, Pa, Tp, i, Q0, W, Q, Tq*	8	42	91.3	40	87.0
20	*P, Pa, Tp, Q0, W, Q, Tq, ΔQ/Tq*	8	42	91.3	40	87.0

7. 研究结论

7.1. *P*、*Pa*、*W*、*Q* 是识别相似洪水的关键指标

灰色关联度法匹配率最高的指标组合为 *P*、*Pa*、*W*、*Q*、Δ*Q*/*Tq*；加权距离系数法优化前后匹配率最高的指标组合均为 *P*、*Pa*、*W*、*Q*，均包含 *P*、*Pa*、*W*、*Q* 四个指标。从匹配率前 20 位各指标出现频次(见表 7)可以看出，*Pa*、*W*、*P*、*Q* 出现次数最多，前 20 位组合中均包含其中的 3 个或 4 个。这表明，*P*、*Pa*、*W*、*Q* 这四个指标是识别相似洪水的关键指标。而指标组合中加入最大雨强 *i*、涨水历时 *Tq*、起涨流量 *Q*0 等指标后，匹配率平均下降 10%~22%。

P、*Pa*、*W*、*Q* 这四个指标中，*P*、*W*、*Q* 虽高度相关，但本文结合灰色关联度法与加权距离系数法的对比分析表明，保留这些高相关指标并未造成明显的信息重叠和结果失真，反而使匹配率保持最高。这说明在相似洪水识别中，指标间的高相关并不必然等同于信息冗余。洪峰流量 *Q* 是预报精度评价的首要指标，降雨总量 *P* 是洪水的主要致洪因子，前期影响雨量 *Pa* 反映产流初始状态，洪量 *W* 反映洪水规模，这四个指标从不同侧面共同刻画洪水过程，构成判别相似洪水的关键信息。相反，若刻意剔除高相关指标，代之以 *i*、*Tq*、*Q*0 等相关性不高的指标，反而会因这些指标在相似量级洪水中波动较大、噪声较强而降低匹配率，干扰关键指标的相似性判断。

表 7. 匹配率前 20 位各指标出现频次统计表

方法	*Pa*	*P*	*W*	*Q*	Δ*Q*/*Tq*	Δ*Z*	*Tp*	*Q*0	*Tq*	*i*
灰色关联度法	20	18	19	17	12	14	12	5	4	0
加权距离系数法(优化前)	19	16	17	17	14	10	8	3	0	0
加权距离系数法(优化后)	20	20	16	20	10	3	13	8	9	8

7.2. 加权距离系数法权重优化后识别效果显著提升

采用差分进化算法优化权重后，加权距离系数法识别精度得到全面提升：匹配率大于 90%的由 0 组增至 31

组，大于 80%的由 6 组增至 286 组，大于 70%的由 19 组增至 502 组，平均提升 27.5%，验证了该优化策略的有效性。但该方法计算耗时较长，实际应用中需权衡算力成本与精度收益。

7.3. 灰色关联度法识别结果更稳健

灰色关联度法在 1013 种指标组合中表现稳定，最优组合匹配率 89.1%，前 20 位集中在 78.3%~89.1%之间，匹配率大于 70%的组合达 74 组，远超加权距离系数法优化前的 19 组。灰色关联度法的稳健性源于其计算原理：关联系数对各指标独立计算后取平均，单一指标的波动对关联度的影响有限；而加权距离系数法将各指标差异综合为单一距离值，任一指标的差异都会直接拉大整体距离，对指标变化更敏感。因此，灰色关联度法更适合指标选取存在不确定性的情形，加权距离系数法则更适合特征指标明确、可通过权重优化提升精度的情形，两者在稳健性与可优化性上形成互补。

7.4. 事前指标相似洪水识别效果明显不如全过程指标

仅采用事前指标 *P*、*Pa*、*Tp*、*i*、*Q*0 识别相似洪水，26 种组合中前 3 位最高匹配率仅 41.3%~56.5% (见表 8)，洪峰流量合格率仅 34.8%~58.7%，明显低于全过程指标。这一差距主要源于事前指标仅包含洪峰出现前的降雨和产流信息，相同面雨量条件下，可能因暴雨时空分布、下垫面异质性等产生差异显著的洪水响应，而全过程指标直接反映洪水规模与过程特征，两者在信息量上存在本质差异。但事前指标在实时洪水预报中具有不可替代的价值——洪峰到来之前，唯有依托这些可获取的信息进行相似性判断，才能为防洪调度提供决策参考。

表 8. 事前指标匹配率前 3 位统计表

方法	排位	组合	指标个数/个	匹配场次/次	相似洪水匹配率/%	合格场次/次	洪峰流量合格率/%
灰色关联度法	1	*P*, *Pa*, *Tp*, *Q*0	4	25	54.3	22	47.8
	2	*P*, *Pa*, *Tp*	3	21	45.7	21	45.7
	3	*P*, *Pa*	2	19	41.3	19	41.3
加权距离系数法 (权重优化前)	1	*P*, *Pa*, *Tp*	3	20	43.5	23	50.0
	2	*Pa*, *Tp*	2	19	41.3	16	34.8
	3	*P*, *Pa*, *Tp*, *Q*0	4	19	41.3	21	45.7
加权距离系数法 (权重优化后)	1	*P*, *Pa*, *Tp*, *i*	4	26	56.5	21	45.7
	2	*P*, *Pa*, *Tp*, *i*, *Q*0	5	26	56.5	27	58.7
	3	*P*, *Pa*, *Tp*, *Q*0	4	25	54.3	27	58.7

本次采用的五个事前指标主要反映流域降雨总量、降雨强度、降雨历时、前期产流条件及起涨流量等事前信息，对降雨空间分布及洪水起涨后的动态过程代表性不足。后续研究可引入暴雨中心位置、降雨空间分布特征及洪水起涨后 3 h、6 h、12 h 流量等事前指标，构建更完整、代表性更强的事前指标体系，进一步检验其对相似洪水识别精度和洪峰预报合格率的提升效果。

参考文献

[1] 刘卫林, 董增川, 梁忠民, 等. 暴雨洪水相似性分析及其应用研究[J]. 中国农村水利水电, 2007(2): 132-135, 148.

[2] 李春红, 陈建, 王建平. 基于量化指标的实时洪水相似性研究[C]//中国水力发电工程学会水电站运行管理专委会, 中国水力发电工程学会梯级调度控制专业委员会. 水电站运行管理及梯级调度控制研究(2019). 武汉: 长江出版社, 2019: 152-158.

[3] 王海潮, 董增川, 梁忠民, 等. 暴雨洪水相似性分析指标体系研究[J]. 水文, 2006, 26(2): 13-17.

[4] 陈建, 李春红, 王峰, 等. 长江上游流域暴雨洪水相似性判别指标研究[J]. 人民长江, 2019, 50(1): 58-63.
[5] 章四兵, 周美立. 系统相似性度量中的非平权距离系数法[J]. 合肥工业大学学报(自然科学版), 2004, 27(8): 903-906.
[6] 张艳平, 周惠成. 基于暴雨洪水相似性分析的洪水分类研究[J]. 水电能源科学, 2016, 30(9): 50-54.
[7] 吴业楠, 钟平安, 赵云发, 等. 基于灰色关联分析的相似洪水动态展延方法[J]. 南水北调与水利科技, 2014, 12(1): 126-130.
[8] 林子珩, 石朋, 钟平安, 等. 基于改进灰关联分析的相似洪水动态识别与预报方法[J]. 水电能源科学, 2020, 38(1): 48-51.
[9] 冯永修, 杨亚钤, 刘金涛. 新丰江水库流域相似洪水识别及洪量预报[J]. 中国农村水利水电, 2021(3): 58-62.
[10] 万新宇, 包为民, 荆艳东, 等. 基于主成分分析的洪水相似性研究[J]. 水电能源科学, 2007, 25(5): 36-39.
[11] 田逸飞, 杨海从, 纪国良, 等. 相似洪水分析与智能推荐模型研究[J]. 水资源研究, 2023, 12(5): 472-485.
[12] 王海军, 衣学军, 黄一昕, 等. 暴雨洪水相似性判别及外推预估方法[J]. 南水北调与水利科技(中英文), 2023, 21(5): 886-894.
[13] 杨艳林, 叶枫, 吕鑫, 等. 一种基于 DTW 聚类的水文时间系列相似性挖掘方法[J]. 计算机科学, 2016, 43(2): 245-249.
[14] 任继周, 彭德才, 乔伟, 等. 基于 DTW 算法的云南昭通地区场次洪水相似性研究[J]. 水利水电快报, 2017, 38(8): 35-38.
[15] 赵圆元, 张行南, 黄钰凯, 等. 基于动态权重的水文时间序列相似性分析方法[J]. 水电能源科学, 2021, 39(2): 9-12, 17.
[16] 李映辉, 钟平安, 钱睿智, 等. 基于动态时间弯曲算法的相似洪水识别方法[J]. 水电能源科学, 2018, 36(11): 51-55.
[17] 钱堃, 包为民, 李偲松, 等. K 均值聚类分析方法在洪水预报中的应用[J]. 水电能源科学, 2012, 30(5): 41-44.
[18] 李映辉, 钟平安, 杨敏芝, 等. 基于投影寻踪-K 均值的相似洪水动态识别方法[J]. 水电能源科学, 2020, 38(7): 78-82.
[19] 黄燕, 张静怡, 刘军涛, 等. 水文相似度的权值确定方法及其应用[J]. 中国水运, 2009, 9(18): 186-188.