

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. WANG H J, YI X J, HUANG Y X, et al. Similarity discrimination and extrapolation prediction methods of heavy rain and flood[J]. South-to-North Water Transfers and Water Science & Technology, 2023, 21(5): 886-894. (in Chinese)

# 暴雨洪水相似性判别及外推预估方法

王海军<sup>1</sup>, 衣学军<sup>1</sup>, 黄一昕<sup>2,3</sup>, 王军<sup>3</sup>, 梁忠民<sup>3</sup>

(1. 山东省水文中心, 济南 250000; 2. 广州市白云区水务工程建设管理中心, 广州 510405;  
3. 河海大学水文水资源学院, 南京 210098)

**摘要:** 超前洪水预估精度不高一直困扰防洪决策, 为解决这一技术难题, 提出一种基于暴雨洪水知识的相似性分析方法, 进行洪水预报预测。该方法从历史典型暴雨洪水知识中提取多要素特征指标, 基于欧氏距离进行当前与历史暴雨洪水的特征指标相似性判别, 根据判别出的最相似洪水, 经“峰-量”联合修正消除非一致性后, 实时外推预估未来洪水过程, 构成一套完整的“多要素特征指标提取-历史暴雨洪水相似性判别-实时洪水修正外推预估”技术。在沂河蒙阴站的应用结果表明, 基于“降雨-径流”关系, 对判别出的最相似洪水进行修正, 显著提高了外推预估洪水精度, 洪峰流量相对误差 $\delta_{Q_m}$ 的范围降至 10% 附近, 峰现时间绝对误差 $\Delta T$ 的范围降至 $\pm 2$  h 以内, 径流深相对误差 $\delta_r$ 的范围降至 $\pm 20$ % 以内, 且随着时间推移, 精度水平不断提高。该套技术方法能够挖掘隐含在历史暴雨洪水数据中的相似性, 超前预估当前洪水变化过程, 为洪水预报提供一种新的技术参考。

**关键词:** 暴雨洪水知识; 多要素特征指标; 相似性判别; 欧氏距离; 外推预估; 洪水预报预测

**中图分类号:** TV213   **文献标志码:** A   **DOI:** 10.13476/j.cnki.nsbdqk.2023.0086

洪水预报预测是防洪非工程措施的重要组成部分<sup>[1-3]</sup>。洪水预报预测的方法一般可分为过程驱动(水文模型)途径和数据驱动途径两类。传统上主要以水文模型途径<sup>[4-5]</sup>为主, 随着水文数据的不断累积及大数据分析技术的不断进步, 数据驱动途径<sup>[6-7]</sup>越来越受到重视。特别地, 随着人工智能技术<sup>[8-11]</sup>在水利行业应用的不断深化, 基于水文数据挖掘的预报预测方法<sup>[12]</sup>逐渐成为研究热点。

通过相似性原理进行水文知识挖掘及预报是其中的一个重要研究方向。牛俊等<sup>[13]</sup>通过寻找当前与历史场次暴雨洪水之间天气成因的相似性因子, 建立了历史气象信息与实时洪水预报耦合的理论框架, 从大尺度环流背景、天气尺度天气系统配置、中尺度云图特征量分析、物理量场 4 个层次寻找相似; 万新宇等<sup>[14]</sup>将原洪水指标综合为少数几个不相关的主成分, 提出了基于主成分分析的洪水相似性研究方法; Ouyang 等<sup>[15]</sup>通过对塔里木河流域的水文流量时间序列进行聚类及相似性搜索, 挖掘出水文过程隐含规律; 张艳平等<sup>[16]</sup>通过分析暴雨洪水在

形成过程中的相似性, 对暴雨洪水的天气成因进行定性分析, 分别采用相关系数法和加权距离系数法计算了暴雨洪水演化过程的相似系数, 从而对暴雨洪水进行分类; Sharma 等<sup>[17]</sup>将相似度模型与 k-NN 技术结合, 提出一种库存时间序列数据逼近与预测的降水预测方法; 吴业楠等<sup>[18]</sup>根据筛选出的洪水特征指标体系, 基于灰色关联分析原理, 构建了相似洪水动态识别办法和相似洪水展延实时洪水预报过程的展延效果评价方法; Hu 等<sup>[19]</sup>提出了一种新的局部对齐算法, 在历史数据库中查找相似的预警洪水; Stenta 等<sup>[20]</sup>研究了平原地区地表径流的水文相似性; Xiao 等<sup>[21-22]</sup>提出了基于“量”“型”相似指标、堆土机距离指标、暴雨分布相似指标的暴雨相似性综合度量方法, 在历史场次暴雨中寻找相似暴雨及其对应的洪水过程, 再根据暴雨量比值对典型洪水过程进行缩放, 以此实现洪水的超前预报预警; 李映辉等<sup>[23]</sup>将相似洪水动态展延相关理论与水库防洪优化调度模型相结合, 提出了基于相似洪水动态展延的防洪调度决策方法。

收稿日期: 2023-06-05   修回日期: 2023-09-28   网络出版时间: 2023-10-18

网络出版地址: <https://link.cnki.net/urlid/13.1430.TV.20231016.1418.001>

基金项目: 国家自然科学基金重点项目(41730750)

作者简介: 王海军(1983—), 男, 山东嘉祥人, 高级工程师, 主要从事水文学及水资源研究。E-mail: navy83@163.com

通信作者: 王军(1982—), 男, 安徽蒙城人, 副教授, 博士, 主要从事水文学及水资源研究。E-mail: wangjun\_hhu@163.com

由此可见,水文知识相似性分析可为深度挖掘隐含在降雨、洪水、流域下垫面信息中的内在规律提供一种新的技术手段,同时能够推进水利数据自动化和智能化处理,辅助提高洪水预报预测精度,对促进水利行业水利现代化、精准化具有重要意义。

理论上,水文数据资料系列越长,可挖掘的暴雨洪水知识越多,暴雨洪水知识相似性分析越准确。然而,实际情况下,由于气候变化和人类活动干扰,水文数据长系列常呈现非一致性,影响了基于暴雨洪水知识的相似性判别及洪水预报精度。目前,关于考虑非一致性的历史相似暴雨洪水知识指导实时洪水预报的研究较少。因此,迫切需要在此领域有所突破。鉴此,本文提出一种基于暴雨洪水知识的相似性修正外推预估方法,以沂河蒙阴站为例进行示例应用,论证了方法的有效性和适用性。

## 1 数据和方法

### 1.1 研究区域及数据来源

蒙阴站以上流域,位于山东省中南部,流域控制面积为  $442 \text{ km}^2$ ,年平均气温为  $13.4 \text{ }^\circ\text{C}$ ,多年平均降水量为  $810 \text{ mm}$ ,年际变化较大,容易出现极端暴雨洪水的现象。蒙阴站以上流域的水系和雨量站网分布见图 1。

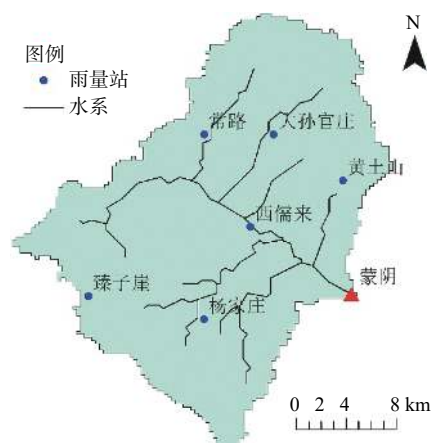


图 1 蒙阴站以上流域水系和雨量站网分布

Fig. 1 Distribution of water system and rainfall station network in the basin above Mengyin Station

从蒙阴站的历史暴雨洪水资料中,筛选出 1960—2020 年代表性较好且相关资料齐全的 71 场历史典型暴雨洪水过程,为暴雨洪水相似性判别及实时洪水推理预估提供数据支撑。

### 1.2 暴雨洪水相似性修正外推预估方法

提出的暴雨洪水相似性修正外推预估方法,主要包括多要素特征指标提取、暴雨洪水相似性判别和实时洪水修正外推预估等关键技术。具体流程见图 2。

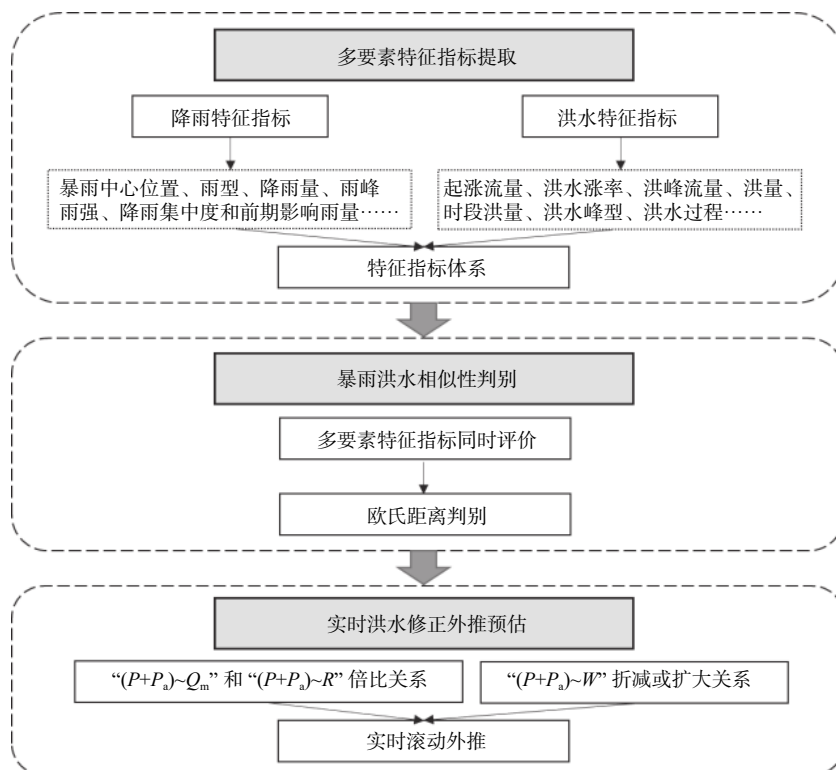


图 2 暴雨洪水相似性修正外推预估方法流程

Fig. 2 Flow chart of storm flood similarity modified extrapolation prediction method

### 1.2.1 多要素特征指标提取

特征指标体系。暴雨洪水多要素特征指标<sup>[24]</sup>, 分为两类: 一般性特征指标(基本指标)和特殊性特征指标(特定指标)。基本指标, 即不管对什么样的流域, 这些指标都必须采用, 才能概化暴雨洪水的基本特征, 如暴雨中心位置、降雨量、前期影响雨量、洪水起涨流量、洪量等。特定指标, 反映的是不同流域暴雨洪水特有的一些特征, 如沿海区的台风路径(影响暴雨中心移动路径)、水流入流方向、平原区的暴雨笼罩面积、暴雨的时-面-深关系, 山丘区的暴雨时空组成、洪水涨落率等。这两类指标共同构成完整的特征指标体系, 可定量表征暴雨洪水的时空变化规律。对某一研究流域, 一般均应包含基本指标, 但对特定指标, 可根据流域的气候、下垫面、资料条件等具体选用。

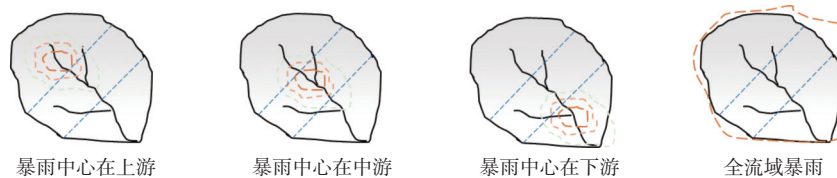


图 3 暴雨中心位置

Fig. 3 Location of rainstorm center

前期影响雨量是指在一场暴雨发生前, 流域土壤的干湿状况与本次降雨的下渗损失量直接有关, 对此次降雨产生径流的多少影响较大。因此, 在流域产流计算中一般都要考虑这一因素, 采用前期影响雨量来定量表示流域的干湿程度。前期影响雨量通常采用逐日连续计算的方法, 其初值往往根据流域特征和前期降雨情况人为给定, 初值的精度直接影响到预报的精度。

降雨量是表征降雨时间分配特征的关键指标之一, 可以直观地表示降雨的多少。在水文部门, 通常分为小雨、中雨、大雨、暴雨等<sup>[25]</sup>, 其中, 小雨指日降雨量在 10 mm 以下, 中雨日降雨量为 10~24.9 mm, 大雨降雨量为 25~49.9 mm, 暴雨降雨量为 50~99.9 mm, 大暴雨降雨量为 100~250 mm; 特大暴雨降雨量在 250 mm 以上。

降雨过程是表征降雨特征的综合型指标, 可综合体现雨型、雨强、降雨集中度、降雨时序组成等特征, 能够量化降雨在时间和空间上的非均匀性。流域面雨量过程, 一般根据流域内各个雨量站、水文站的实测降雨数据, 利用泰森多边形等方法计算

特征指标提取。由于环境变化下, 流域下垫面状况发生了较大改变, 水文序列表现出较强的非一致性, 不同年代相同降雨产生的洪水往往不同。因此, 在选取相似性指标时, 采取“暴雨相似”原则, 即只提取与降雨特征有关的指标, 以此判别出历史中最相似的一场暴雨洪水过程。

下面具体介绍暴雨中心位置、前期影响雨量、降雨量及降雨过程 4 个降雨特征指标的提取过程。

暴雨中心位置是表征降雨空间分布特征的关键指标之一, 能够定量判断降雨中心的空间位置及其数量。依据不同暴雨中心距流域出口断面的距离远近, 将历史暴雨洪水事件分为 4 类: 暴雨中心在上游、暴雨中心在中游、暴雨中心在下游、全流域暴雨(对于比较小的流域, 也可分为 3 类: 暴雨中心在上游、暴雨中心在下游、全流域暴雨), 见图 3。

得到。

### 1.2.2 历史暴雨洪水相似性判别

相似性判别是基于提取的暴雨洪水知识判别出历史相似性形成的历史事件典型时空属性及特征指标组合。因此, 暴雨洪水相似性判别是一个从局部到整体的映射过程, 属于多要素时间序列相似性分析问题。需要采用简洁的形式表达与多要素相关的特征指标, 表征历史典型暴雨洪水; 然后, 以特征表示的结果作为输入, 定量刻画两条时间序列的相似程度, 且能够较好地支持时间序列的多种形变; 最后, 以某种相似模式度量方法为基础, 在时间序列数据集中, 寻找与给定模式相似的时间序列。

本文提出的多指标同时相似性判别方法, 是通过挖掘历史相似性形成的历史事件典型时空属性及特征指标组合, 推演分析不同场景下的暴雨洪水演变过程。该方法的具体计算步骤如下:

(1) 选择  $n$  个代表性的多要素特征指标  $X_1, X_2, \dots, X_n$ 。例如, 本次研究区选取暴雨中心位置、前期影响雨量、面平均总雨量和降雨过程这 4 个特征指标。

(2)对当前暴雨洪水,采用欧氏距离<sup>[26]</sup>(Euclidean Distance, ED)同时判别评价4个特征指标。

(3)计算当前暴雨洪水与 $N$ 场历史暴雨洪水的相似度,找到欧氏距离最短 $D = \min(d_{12}, d_{13}, \dots, d_{1m})$ 的一场暴雨洪水,即为最相似的洪水。

### 1.2.3 实时洪水修正外推预估

实时洪水修正外推预估本质上是暴雨洪水模式的正向推理<sup>[27]</sup>。正向推理就是从初始事实数据出发,正向使用规则进行推理,即用规则前提与动态数据库中的事实匹配,或用动态数据库中的数据测试规则的前提条件,然后产生结论或执行动作,同时滚动修正,朝目标方向前进。

基于历史暴雨洪水相似性判别的实时洪水修正外推预估的主要步骤如下:

(1)基于多要素相似性判别方法,判别出与当前暴雨洪水过程第2相似的洪水过程,即除去本身外历史洪水中最相似的一场洪水过程。

(2)对于满足一致性的暴雨洪水过程,可以直接移用判别出的相似洪水外推预估。

(3)对于变化环境,水文系列往往难以满足一致性要求,例如:蒙阴站雨洪资料在1980年前后一致性发生明显突变,应该结合变异点前后的“降雨-洪峰流量( $P \sim Q_m$ )”或“降雨-径流( $P \sim R$ )”关系,将非一致性的暴雨洪水过程,经“峰-量”联合修正后,再进行相似洪水外推预估:

如果判别出的历史最相似洪水是发生突变前(1980年前)的洪水,则基于“ $(P+P_a) \sim Q_m$ ”关系或“ $(P+P_a) \sim R$ ”关系( $P_a$ 为前期雨量指数),计算1980年前后的洪水过程折减或扩大系数 $k$ ,进而建立当前洪水与历史最相似洪水之间的倍比关系。

$$k = \frac{Q_{m,H}}{Q_{m,C}}, \text{ 或 } k = \frac{R_H}{R_C} \quad (1)$$

式中: $k$ 为折减或扩大系数; $Q_{m,C}$ 为当前发生洪水的洪峰流量,  $m^3/s$ ;  $Q_{m,H}$ 为历史最相似洪水的洪峰流量,  $m^3/s$ ;  $R_C$ 为当前发生洪水的径流深,  $mm$ ;  $R_H$ 为历史最相似洪水的径流深,  $mm$ 。

如果判别出的历史最相似洪水是发生突变后(1980年后)的洪水,则基于“ $(P+P_a) \sim W$ ”关系( $W$ 为流域出口断面径流量),即发生在预报 $t_0$ 时刻前后径流系数 $\alpha = \frac{W}{P+P_a}$ 不变、折减或扩大修正。

$$\frac{\frac{W_{0,H}}{P_H + P_{a,H}}}{\frac{W_{0,C}}{P_C + P_{a,C}}} = \frac{\alpha_H}{\alpha_C} = \frac{\frac{W_{1,H}}{P}}{\frac{W_{2,C}}{P}} = \frac{W_{1,H}}{W_{2,C}} \quad (2)$$

$$W_{2,C} = \frac{W_{0,C} W_{1,H} (P_H + P_{a,H})}{W_0^H (P_C + P_{a,C})} \quad (3)$$

式中: $P_C$ 和 $P_{a,C}$ 分别为当前发生洪水的降雨量和前期雨量指数(反映前期土壤湿度),  $mm$ ;  $W_{0,C}$ 和 $W_{2,C}$ 分别为当前发生的洪水预报在 $t_0$ 时刻前和后的流域出口断面径流量,  $m^3$ ;  $\alpha_C$ 为当前发生洪水的径流系数;  $P_H$ 和 $P_{a,H}$ 分别为历史最相似洪水的降雨量和前期雨量指数(反映前期土壤湿度),  $mm$ ;  $W_{0,H}$ 和 $W_{1,H}$ 分别为历史最相似洪水的在预报 $t_0$ 时刻前和后的流域出口断面径流量,  $m^3$ ;  $\alpha_H$ 为历史最相似洪水的径流系数。

根据式(3)则可计算出预报 $t_0$ 时刻后的流域出口断面径流量。

(4)重复步骤(1)~(3),则可对当前实时暴雨洪水过程不断滚动修正并外推,快速预报预测流量,达到防汛目标。

## 2 结果和分析

对蒙阴站的所有场次洪水进行一致性分析,得到 $(P+P_a) \sim Q_m$ 关系和 $(P+P_a) \sim R$ 关系,见图4。从图4可以看出,在1980年前、1980年后,相同的 $(P+P_a)$ 对应不同的 $Q_m$ 和 $R$ ,即相同的降雨产生不同的洪水。由此表明,蒙阴站在1980年前、后出现了暴雨洪水时间序列的非一致性。因此,可按照1980年前和1980年后,对蒙阴站的所有历史场次洪水进行划分。导致1980年前、后暴雨洪水非一致性的原因很多,包括区域社会经济的发展、城市建设规模的扩大、耗水量的增多、水利工程的建设和运行等,使得即使降雨完全相同,但在不同年际也不可能产生相同洪水。

根据蒙阴站以上流域的特点,提取4个降雨特征指标:暴雨中心位置、前期影响雨量、面平均总雨量和降雨过程。假设所选的4个特征指标的重要性相同,即赋予它们相同的权重。基于欧氏距离相似性判别方法,同时对暴雨洪水4个特征指标进行相似性判别,找到最相似的历史典型暴雨洪水。为了验证该相似性判别方法的有效性,任意选择蒙阴站的1场暴雨洪水,看是否能在所有71场历史典型暴雨洪水中判别出最相似的本身。如图5所示是任选的20200813号洪水的多指标同时相似性判别结果,图中红色实线是当前20200813号洪水流量过程,黑

色实线是找到的历史最相似洪水流量过程,红色实线与黑色实线完全重合。结果表明:对任意洪水,

经过赋权制分级相似性判别,能够找到历史典型暴雨洪水中最相似的本身。

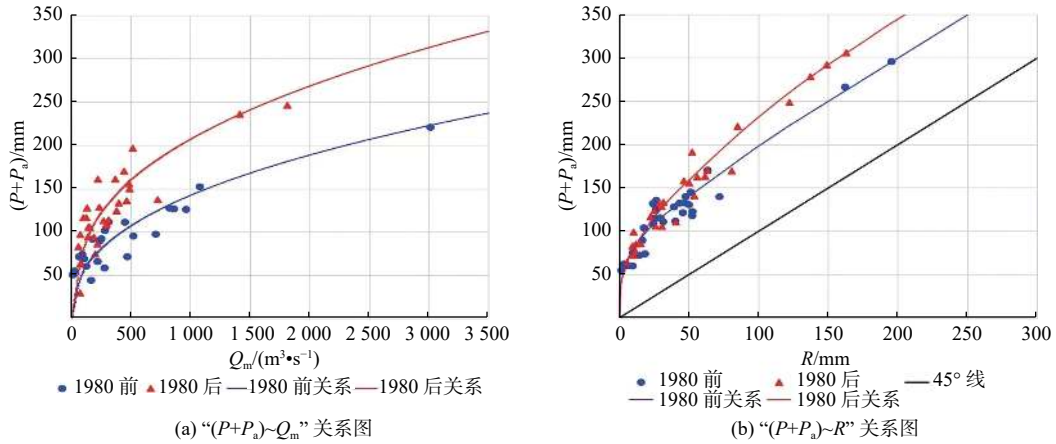


图 4 1980 年前、后的降雨-径流关系

Fig. 4 Rainfall-runoff relationship diagrams around 1980

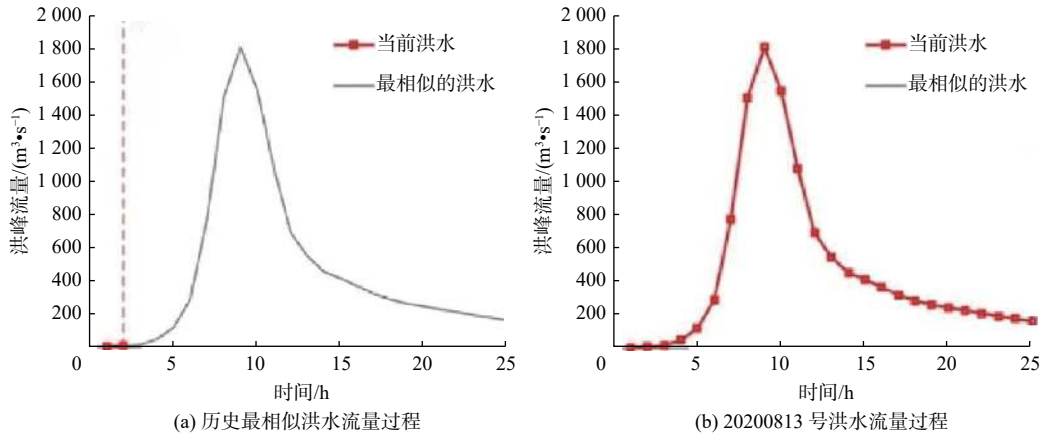


图 5 20200813 号暴雨洪水相似性判别结果

Fig. 5 Similarity discrimination results of rainstorm flood (20200813 #)

为模拟实际作业洪水预报过程,从蒙阴站的 71 场典型暴雨洪水中随机选择 1 场,如 20190810 号洪水,在该场暴雨洪水实时预报过程中,采用提出的多指标同时相似性判别方法,检索剩余的 70 场历史典型暴雨洪水中最为相似的 1 场暴雨洪水过程;基于“降雨-径流”关系,实时滚动修正并推理预估当前洪水的未来过程;实时滚动预报,从而获得不断更新的预报结果。20190810 号场次洪水在实时预报过程中,从 8 h 至 15 h 时刻的相似性洪水判别预估和修正外推预估结果见图 6,图中红色的实线是当前实时洪水过程,粉色的实线是当前洪水的未来洪水过程,点划线是从历史典型暴雨洪水中判别出的最相似洪水过程,虚线是修正外推预估得到的未来洪水过程。从图 6 可以看出:点划线与粉色实线的“型”相似,但“量”不够相似;虚线与粉色实线的

“型”相似,“量”也相似。这说明经一致性修正后的相似性外推预估洪水过程保证了“量-型”同时相似,更接近真实发生的洪水过程。除随机选取的 20190810 号洪水外,其余场次暴雨洪水结果也表明,修正外推预估对于洪峰流量和洪水过程都是一种理想的洪水预报预测方法。

表 1 是对图 6 中推理预估结果的精度统计。从表 1 可知:根据暴雨洪水相似性判别预估的未来洪水过程,精度不高;基于“降雨-径流”关系,对判别出的最相似洪水进行修正,极大提高了外推预估洪水精度,例如,洪峰流量相对误差  $\delta_{Q_m}$  的范围降至  $\pm 13.1\%$  以内,峰现时间绝对误差  $\Delta T$  的范围降至  $\pm 2$  h 以内,径流深相对误差  $\delta_r$  的范围降至  $\pm 17.7\%$  以内,确定性系数  $D_c$  的范围提高至 0.52 以上;随着预报时间推移,实测信息越来越多,修正外推预估

结果的精度也越来越高,在第 15 h 时刻,修正外推预估的  $D_c$  为 0.98。从图 6(h)也可以看出:在第 15 h 时刻的修正外推预估洪水过程与实际未来洪水过

程的趋势几乎完全一致;洪峰出现以前(第 12 h 以前),通过修正外推预估的未来洪水过程,洪峰和洪量精度都较高,可以满足实际防洪预报需求。

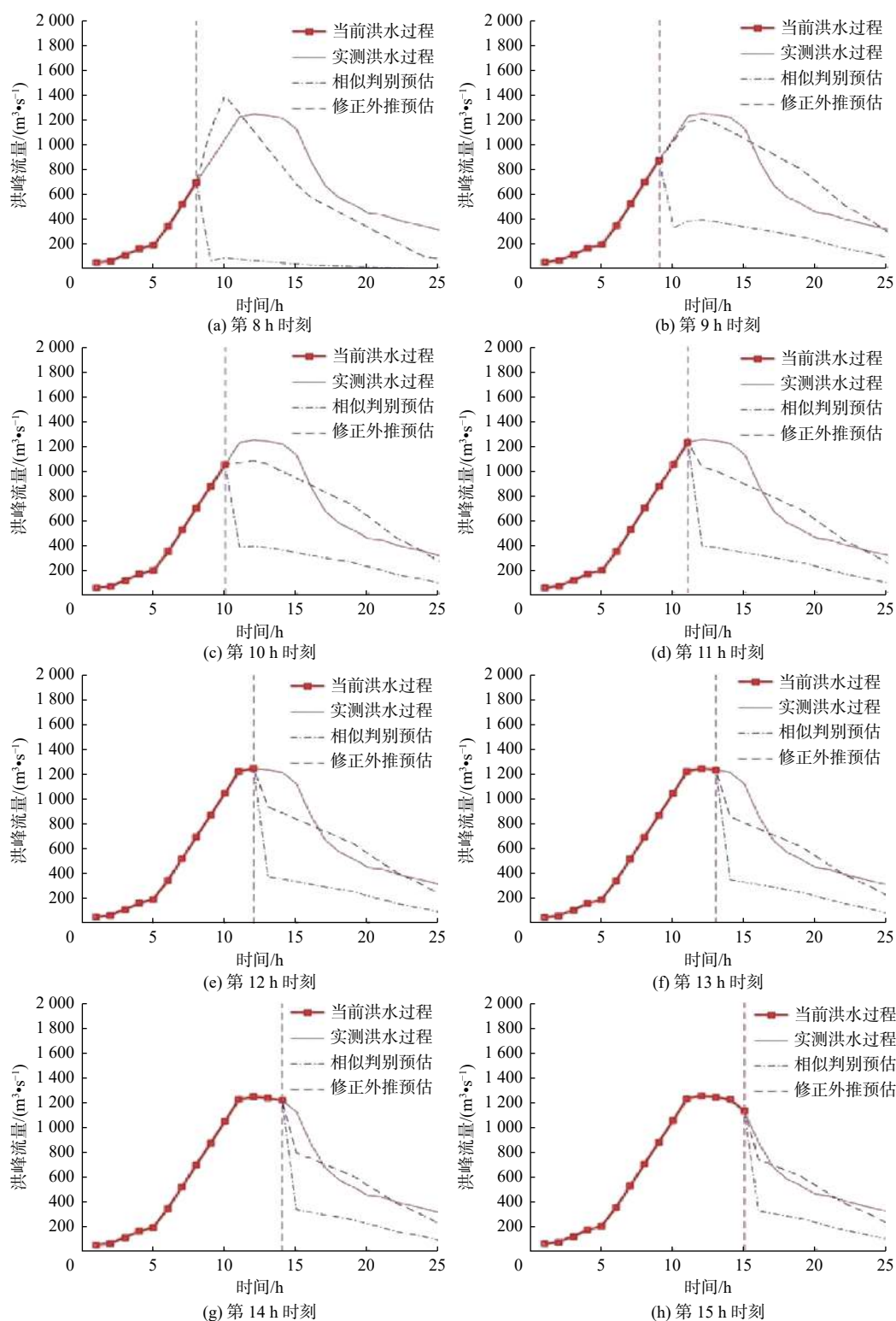


图 6 20190810 号暴雨洪水相似性修正外推预估结果

Fig. 6 Prediction results of rainstorm flood(20190810#) similarity correction extrapolation

综上,根据暴雨洪水相似性判别方法,找到了历史典型暴雨洪水过程中最相似的洪水。可直接移

用该相似洪水过程,对当前洪水进行外推预估;也可建立当前洪水与库中洪水的“降雨-径流”关系,

经“峰-量”联合修正判别出的相似洪水,再外推预估未来洪水过程。验证结果表明,经修正后的外推预估洪水过程更为准确。因此,建议将判别出的最

相似洪水,“峰-量”修正,消除数据非一致性的影响,再进行实时洪水滚动推理预估。

表 1 20190810 号暴雨洪水修正外推预估结果的精度统计

Tab. 1 Precision statistics of rainstorm flood (20190810#)correction extrapolation prediction results

时刻/h	方法	精度评价			
		$\delta_{Q_m}/\%$	$\Delta T$	$\delta_R/\%$	$D_c$
8	相似判别预估	-43.60	-4	-89.50	-4.14
	修正外推预估	11.90	-2	-17.70	0.52
9	相似判别预估	-29.80	-3	-56.70	-1.39
	修正外推预估	-3.40	0	8.50	0.82
10	相似判别预估	-16.00	-2	-51.50	-1.14
	修正外推预估	-13.10	0	-0.40	0.84
11	相似判别预估	-1.60	-1	-45.40	-0.8
	修正外推预估	-1.60	-1	-2.90	0.84
12	相似判别预估	0	0	-39.10	-0.45
	修正外推预估	0	0	-4.70	0.84
13	相似判别预估	0	0	-32.90	-0.1
	修正外推预估	0	0	-4.20	0.87
14	相似判别预估	0	0	-26.70	0.26
	修正外推预估	0	0	-2.50	0.92
15	相似判别预估	0	0	-20.90	0.56
	修正外推预估	0	0	-0.80	0.98

### 3 结论

建立了一套“多要素特征指标提取-历史暴雨洪水相似性判别-实时洪水修正外推预估”技术。以历史典型暴雨洪水为研究对象,从中提取降雨特征指标,如暴雨中心位置、前期影响雨量、面平均总雨量、降雨过程等;基于欧氏距离等判别准则,进行多要素特征指标同时相似性判别;根据相似性知识推理得到的历史典型洪水,结合变化前后“降雨-洪峰流量”或“降雨-径流”关系,经“峰-量”联合修正,实时滚动外推预估未来洪水过程,最终实现基于暴雨洪水知识的洪水预报预测。

在沂河蒙阴站的应用结果表明,对任意一场暴雨洪水,均能通过多要素特征,准确找到历史场次暴雨洪水中最相似的本身,保证了该技术流程在理论上的正确性。根据判别出的最相似洪水,经修正后(满足一致性)再外推预估未来洪水过程,相比于直接移用相似洪水过程进行预报,保证了“量-型”同时相似,能有效地提高洪水预报预测精度,满足实际防洪预报需求。该技术流程在蒙阴流域形成的示范应用,为暴雨洪水预报预测提供了新的技术

参考,可望在其他流域得到推广和应用。

作为示例并考虑到暴雨洪水资料的非一致性,本文只选用了降雨特征指标进行历史洪水相似性分析,未来可加入洪水特征做深入研究。

#### 参考文献:

- [1] 刘志雨. 洪水预测预报关键技术研究与实践[J]. 中国水利, 2020(17): 7-10. DOI: 10.3969/j.issn.1000-1123.2020.17.013.
- [2] 李红霞, 王瑞敏, 黄琦, 等. 中小河流洪水预报研究进展[J]. 水文, 2020, 40(3): 16-23,50. DOI: 10.19797/j.cnki.1000-0852.20190344.
- [3] 姜仁贵, 王思敏, 解建仓, 等. 变化环境下城市暴雨洪涝灾害应对机制[J]. 南水北调与水利科技(中英文), 2022, 20(1): 102-109. DOI: 10.13476/j.cnki.nsb-dqk.2022.0012.
- [4] 栾承梅, 梁忠民, 仇少鹏, 等. 山丘区小流域SCS-Nash产汇流模型应用[J]. 南水北调与水利科技(中英文), 2021, 19(2): 246-254. DOI: 10.13476/j.cnki.nsb-dqk.2021.0026.
- [5] 罗玮, 朱阳, 陈在妮, 等. 基于WRF模式的大渡河流域径流预报模型[J]. 南水北调与水利科技(中英文), 2021, 19(3): 469-476. DOI: 10.13476/j.cnki.nsb-dqk.

- 2021.0050.
- [6] 李鹏,徐宗学,赵刚,等.基于SWMM与LISFLOOD-FP模型的城市暴雨内涝模拟:以济南市为例[J].*南水北调与水利科技(中英文)*,2021,19(6):1083-1092. DOI: 10.13476/j.cnki.nsbdqk.2021.0111.
- [7] 欧阳文字,叶磊,王梦云,等.深度学习水文预报研究进展综述 I:常用模型与建模方法[J].*南水北调与水利科技(中英文)*,2022,20(4):650-659. DOI: 10.13476/j.cnki.nsbdqk.2022.0067.
- [8] ROSPOCHER M, VAN ERP M, VOSSSEN P, et al. Building event-centric knowledge graphs from news[J]. *Journal of Web Semantics*, 2016, 37: 132-151. DOI: 10.1016/j.websem.2015.12.004.
- [9] SHI L, LI S, YANG X, et al. Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services[J]. *BioMed Research International*, 2017(4): 1-12. DOI: 10.1155/2017/2858423.
- [10] 冯钧,杭婷婷,陈菊,等.领域知识图谱研究进展及其在水利领域的应用[J].*河海大学学报(自然科学版)*,2021,49(1):26-34. DOI: 10.3876/j.issn.1000-1980.2021.01.005.
- [11] 冯钧,徐新,陆佳民.水利信息知识图谱的构建与应用[J].*计算机与现代化*,2019,289(9):35-40. DOI: 10.3969/j.issn.1006-2475.2019.09.007.
- [12] 金菊良,陈鹏飞,陈梦璐,等.基于知识图谱的自然灾害风险防控研究进展的文献计量分析[J].*灾害学*,2019,34(2):145-152.
- [13] 牛俊,董增川,梁忠民.流域暴雨洪水天气成因相似性分析[J].*东北水利水电*,2006(5):20-22,72. DOI: 10.3969/j.issn.1002-0624.2006.05.008.
- [14] 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(5):36-39. DOI: 10.3969/j.issn.1000-7709.2007.05.010.
- [15] OUYANG R, REN L L, CHENG W M, et al. Similarity search and pattern discovery in hydrological time series data mining[J]. *Hydrological Processes*, 2010, 24(9): 1198-1210. DOI: 10.1002/hyp.7583.
- [16] 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*,2012,30(9):50-54.
- [17] SHARMA, A, BOSE M. Rainfall prediction using k-NN based similarity measure[J]. *Recent advances in information technology*, 2014, 266: 125-132. DOI: 10.1007/978-81-322-1856-2\_14.
- [18] 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. DOI: 10.3724/SP.J.1201.2014.01126.
- [19] HU W K, WANG J D, CHEN T W. A local alignment approach to similarity analysis of industrial alarm flood sequences[J]. *Control Engineering Practice*, 2016, 55: 13-25. DOI: 10.1016/j.conengprac.2016.05.021.
- [20] STENTA H, RICCARDI G, BASILE P. Grid size effects analysis and hydrological similarity of surface runoff in flatland basins[J]. *Hydrological Sciences Journal*, 2017, 62(11): 1736-1754. DOI: 10.1080/02626667.2017.1349315.
- [21] XIAO Z L., LIANG Z M., LI B Q, et al. New flood early warning and forecasting method based on similarity theory[J]. *Journal of Hydrologic Engineering*, 2019, 24(8): 04019023. DOI: 10.1061/(ASCE)HE.1943-5584.0001811.
- [22] 肖章玲,梁忠民,李彬权,等.一种基于混沌理论的“量”-“型”相似预测模型[J].*南水北调与水利科技(中英文)*,2021,19(5):894-899,929. DOI: 10.13476/j.cnki.nsbdqk.2021.0093.
- [23] 李映辉,钟平安,朱非林,等.基于相似洪水动态展延的水库防洪调度决策方法及其应用[J].*水电能源科学*,2020,38(10):39-43.
- [24] 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. DOI: 10.16232/j.cnki.1001-4179.2019.01.011.
- [25] GB/T 28592-2012,降水量等级[S].
- [26] 李正欣,张凤鸣,张晓丰,等.多元时间序列相似性搜索研究综述[J].*控制与决策*,2017,32(4):577-583. DOI: 10.13195/j.kzyjc.2016.1173.
- [27] 尹洁,施琴芬,李锋.面向应急决策的极端洪水关键情景推理研究[J].*管理评论*,2019,31(10):255-262. DOI: 10.14120/j.cnki.cn11-5057/f.2019.10.023.

## Similarity discrimination and extrapolation prediction methods of heavy rain and flood

WANG Haijun<sup>1</sup>, YI Xuejun<sup>1</sup>, HUANG Yixin<sup>2,3</sup>, WANG Jun<sup>3</sup>, LIANG Zhongmin<sup>3</sup>

(1. Shandong Provincial Hydrological Center, Jinan 250000, China; 2. Guangzhou Baiyun Water Engineering Construction and Management Center, Guangzhou 510405, China; 3. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

**Abstract:** Flood forecasting and prediction are integral components of non-structural flood management measures. Methods for flood forecasting and prediction can generally be classified into two categories: process-driven



approaches (hydrological models) and data-driven approaches. Traditionally, the focus has been on process-driven approaches, but with the accumulation of hydrological data and advancements in big data analytics, data-driven approaches have gained increasing attention. In particular, the application of artificial intelligence technology in the water industry has led to the emergence of hydrological data mining-based forecasting and prediction methods as a research hotspot. Conducting hydrological knowledge mining and prediction based on the principle of similarity has become an important research direction, offering a new technical means to uncover hidden patterns within rainfall, floods, and watershed surface information. This approach also promotes the automation and intelligence of water resources data processing, assisting in improving the accuracy of flood forecasting and prediction, thereby facilitating the modernization and precision of the water industry.

In theory, the longer the series of hydrological data, the more torrential rain-induced flood knowledge can be extracted. However, hydrological data series in a changing environment often exhibit inconsistencies, which affect the accuracy of flood forecasting and prediction based on torrential rain-induced flood knowledge. Currently, research on historical similar torrential rain-induced flood knowledge considering inconsistencies in guiding real-time flood forecasting is relatively limited. In this context, a methodology based on the knowledge of torrential rain-induced floods for real-time flood forecasting and prediction is proposed. The proposed method focuses on historical records of typical torrential rain-induced floods and extracts rainfall feature indicators, such as the position of the rainstorm center, antecedent precipitation, total average rainfall, and rainfall processes. Multiple feature indicators are simultaneously assessed for their similarity using criteria such as Euclidean distance. By inferring historical typical floods based on similarity knowledge and incorporating the "rainfall-peak flow" or "rainfall-runoff" relationship before and after the change, a combined "peak-flow" correction approach is applied to ensure consistency. Real-time rolling extrapolation is then performed to estimate future flood processes, forming a comprehensive "multi-feature indicator extraction-historical torrential rain-induced flood similarity determination-real-time flood correction and extrapolation" technique.

The application results at the Mengyin Station on the Yi River demonstrate the effectiveness of the proposed methodology. For any given torrential rain-induced flood event, the most similar historical flood event can be accurately identified through multiple feature indicators, ensuring the theoretical correctness of the technique. By considering the most identified similar flood event and applying suitable corrections to ensure consistency, the extrapolation and prediction of future flood processes significantly improve the accuracy of real-time flood forecasting compared to the direct application of similar flood processes.

In summary, the suggested methodology, grounded in torrential rain-induced flood knowledge, introduces an effective avenue for real-time flood forecasting and prediction. By extracting multiple feature indicators, evaluating their similarity, and incorporating correction and extrapolation steps, it enables accurate identification of similar historical flood events and enhances the precision of real-time flood forecasting. This study contributes to the progression of flood management and establishes the groundwork for further research aimed at enhancing flood forecasting accuracy and propelling the modernization of the water industry.

**Key words:** heavy rain and flood knowledge; multiple feature indicator; similarity discrimination; Euclidean distance; extrapolation prediction; flood forecasting