

DOI: 10.20040/j.cnki.1000-7709.2023.20222417

# 龙滩水电站相似性径流预报研究

谢金记<sup>1</sup>, 陈粉娥<sup>1</sup>, 杨 杉<sup>2</sup>, 李凌燕<sup>1</sup>, 钟 成<sup>1</sup>, 闻 昕<sup>2</sup>

(1. 广西桂冠电力股份有限公司, 广西 南宁 530000; 2. 河海大学水利水电学院, 江苏 南京 210098)

**摘要:** 针对岩溶地区产汇流机制复杂、地下水监测资料不足、径流预报精度低等问题, 提出了融合物理机制和数据驱动的相似性预测方法, 建立了不同流量量级、不同预报因子、不同前期影响滞时下最优组合预报方案, 可实现不同水雨情条件下多模式的自适应匹配和智能切换, 可生成具有可解释性的径流预报结果, 并将该方法应用于龙滩水电站径流预报中。结果表明, 该方法提高了岩溶地区的径流精度和有效预见期, 对保障大型电站科学调度和安全运行具有科学意义和实际价值。

**关键词:** 相似性识别; 最近邻抽样; 滚动预报; 岩溶; 多模式切换

**中图分类号:** TV743; P338

**文献标志码:** A

**文章编号:** 1000-7709(2023)11-0010-04

## 1 概况

龙滩水电站位于广西自治区天峨县境内, 是红水河梯级开发的龙头电站。龙滩水电站控制集水面积 98 500 km<sup>2</sup>, 多年平均流量 1 610 m<sup>3</sup>/s。该地区具有典型的岩溶地区雨源型河流特征。龙滩以上流域地形复杂, 同时位于中低纬度交界处, 受中小尺度天气系统的影响较大, 常发生局部性暴雨。区域径流分布与降雨分布相一致, 一般在 4~5 月出现中小洪水, 此后进入主汛期, 6~8 月发生大洪水, 8 月以后洪水渐小, 至 10 月底汛期结束。洪水大多为连续多峰, 持续时间长, 过程一般 6~8 d, 涨洪历时 0.8~3.0 d。岩溶地区的产汇流机制非常复杂<sup>[1]</sup>, 传统的水文预报方法难以达到较好的预报效果。集总式水文模型以系统为例, 结构简单, 对数据要求相对较低, 但往往具有多解性, 在该地区使用时难以充分考虑其复杂的地下水循环过程<sup>[2]</sup>。分布式水文模型考虑了气象、下垫面等影响因子的差异性, 能够较为准确地描述岩溶流域内的水文过程, 但要求研究区提供大量的数据, 制约了其在该地区的应用<sup>[3-5]</sup>。数据驱动模型不依赖真实的物理过程, 旨在寻找输入与输出之间最优的数学关系, 在该地区运用较为方便, 但模型属于“黑箱模型”, 水文规律和产汇流机制考虑不足, 限制了数据驱动模型的推广应用。对某一特定流域, 当制约降雨的主导天气系统、地

表水和地下水分布等水雨情条件相似时, 产生的径流过程也往往具有一定的相似性, 即岩溶地区复杂的产汇流规律实际隐藏在历史大量的水雨情数据中<sup>[6]</sup>。为此, 本文综合物理机制模型与数据驱动方法的优势, 提出基于多因素相似性的径流预报方法, 建立了不同流量量级、不同预报因子、不同前期影响滞时下最优组合预报方案, 可实现不同水雨情条件下多模式的自适应匹配和智能切换, 可生成具有可解释性的径流预报结果, 最后以龙滩水电站为例验证了该方法的可行性和有效性。

## 2 研究资料及方法

### 2.1 研究资料

龙滩水电站有两个上级水电站, 分别为南盘江支流的天生桥二级电站和北盘江支流的董箐电站。天二—董箐—龙滩区间共设有紫云等 70 个遥测雨量站点, 逐小时记录站点实测雨量数据, 详见图 1。采用流域区间降雨和电站的出入库流量数据, 降雨和径流数据的尺度均为小时。

### 2.2 研究方法

#### 2.2.1 历史样本库建立

分析龙滩水电站水雨情资料的完整程度和质量情况, 确定平班—董箐—龙滩区间的降雨径流样本库起止时间为 2016-01~2021-12。对平班—董箐—龙滩区间的径流进行预报, 区间径流通过龙滩入库流量减去天二和董箐的出库流量计算得到。

收稿日期: 2022-11-16, 修回日期: 2023-02-21

作者简介: 谢金记(1969-), 男, 高级工程师, 研究方向为水电水库调度、水工技术管理, E-mail: 718883133@qq.com

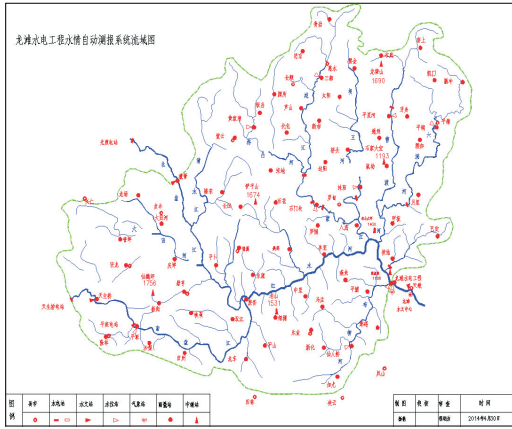


图 1 龙滩水电站控制流域范围及水雨情监测站点

Fig. 1 Watershed and hydrological monitoring system of Longtan Hydropower Station

根据广西气象局的气象分区划分,使用 70 个雨量站点资料计算各分区面雨量,建立降雨样本库。同时,由于降雨资料均为小时数据,为降低预报因子的输入维度,将降雨资料从小时数据换算为日数据。

### 2.2.2 相似性预报模型建立

建立多因子最近邻抽样回归模型,首先识别出与控制断面区间径流相关性较高的前滞降雨、径流特征指标矢量:

$$\mathbf{X}_j = (X_{1,j}, X_{2,j}, \dots, X_{p,j}, X_{p+1,j}, X_{p+2,j}, \dots, X_{p+q,j}) \quad (1)$$

式中,  $\mathbf{X}_j$  为预报因子;  $p$  为降雨特征指标个数;  $q$  为径流特征指标个数。

现在已知待预报时段的特征矢量  $\mathbf{X}_0$ , 预报其对应的  $y_0$ 。以欧氏距离作为衡量样本相似程度的指标,筛选与当前样本欧氏距离  $r_j$  最小的目标样本,欧氏距离越小说明  $\mathbf{X}_j$  与  $\mathbf{X}_0$  越相似,则  $y_j = y_0$  的可能性越大。降雨预报因子、径流预报因子欧氏距离计算公式分别为:

$$d_{r,t}^p = \left[ \sum_{p_s=1}^p (X_{p_s,j} - X_{p_s,0})^2 \right]^{1/2} \quad (2)$$

$$d_{r,t}^q = \left[ \sum_{q_s=1}^q (X_{q_s,j} - X_{q_s,0})^2 \right]^{1/2} \quad (3)$$

式中,  $d_{r,t}^p$ 、 $d_{r,t}^q$  分别为降雨特征矢量、径流特征矢量之间的欧氏距离;  $p_s$ 、 $q_s$  分别为降雨、径流特征指标索引。

由于降雨数据和径流数据的量级相差较大,本文引入降雨影响因子权重  $\alpha$ , 则径流影响因子的权重为  $(1-\alpha)$ , 提出综合欧氏距离:

$$d'_{r,t} = d_{r,t}^p \alpha + d_{r,t}^q (1-\alpha) \quad (4)$$

为避免筛选结果的随机性和偶然性,相似性预报模型会筛选出  $k$  个与待预报时段特征矢量相似的样本,预报值  $y_0$  为  $k$  个相似样本的后续

值  $y_j$  和抽样权重  $w_j$  的累加,各样本的抽样权重  $w_j$  通过其欧氏距离  $r_j$  计算得到,计算公式为:

$$w_j = \left( r_j \sum_{j=1}^k \frac{1}{r_j} \right)^{-1} \quad (5)$$

相似性径流预报模型涉及的参数有抽样样本数  $k$ 、预报因子的影响滞时、降雨影响因子权重系数  $\alpha$ 。参数之间相对比较独立,无明显的关联性,每个参数对模型输出结果均存在较大影响,各参数不需通过模型寻优确定。

### 2.2.3 预报方案

岩溶地区地表径流深受雨强和壤中流向地表径流转化过程的影响,地下水源的补给作用可通过前期的降水和径流时空分布特征来反映,因此选择流域降雨和径流作为预报因子。

预报方案是预报因子及其影响时间的组合,根据相关系数确定各类预报因子的前期影响滞时,在此基础上设置不同的预报因子和前期影响滞时组合作为输入因子。

(1)前期影响滞时。首先判断径流的前期影响时间,预报断面径流的自相关系数随滞时的延长而减小,在 3 d 之后,斜率较稳定,径流前期影响天数为 2、3 d。进一步分析区间总雨量与径流相关特性,该区间的降雨和径流的相关系数首先随滞时延长而增大,在滞时为 1 d 时到达最大值,随后随滞时的延长而减小。考虑到土壤含水率等影响因素,将降雨滞时数延长至 1~3 d。

(2)组合模式设置。岩溶地区由降水、地下水等不同水源主导,且洪水、枯水、涨水、退水下主导因素也不一致。根据流量量级提出多套预报模式,分析不同模式的前期降雨和径流等分布规律,建立多模式的预报方案和判别条件,可根据实际水雨情条件变化智能匹配和切换不同预报模式,提高预报精度和计算速度。对平班—董箐—龙滩区间的历史径流进行统计分析,分为大流量 ( $>1\,500\text{ m}^3/\text{s}$ )、中流量 ( $800\sim1\,500\text{ m}^3/\text{s}$ ) 和小流量 ( $<800\text{ m}^3/\text{s}$ ) 共三个等级。经方案测试和模型寻优计算,确定使用大流量模式预报的条件为  $P_t > 4$  且  $Q_{t-1} > 1\,800\text{ m}^3/\text{s}$  及  $P_t > 15$  且  $Q_{t-1} > 1\,000\text{ m}^3/\text{s}$ 。确定使用中流量模式预报的条件为  $Q_{t-1} > 800\text{ m}^3/\text{s}$ 。表示若前期降雨径流情况满足大流量模式预报的任一条件,认为预报的流量会超过  $1\,500\text{ m}^3/\text{s}$ ,使用大流量模式进行预报,中流量模式预报同理。

(3)组合方案确定。以相对误差最低为评价标准,经多方案比选,确定组合模式的预报方案,见表 1。

表 1 组合模式预报方案

Tab. 1 Combination model forecast scheme of runoff

预报模式	预报因子输入	$k$	最优降雨权重
大流量	7个面雨量( $t, t-1, t-2$ ) + 本站径流( $t-1, t-2, t-3$ )	5	0.989
中流量	7个面雨量( $t, t-1, t-2$ ) + 本站径流( $t-1, t-2$ )	3	0.979
小流量	7个面雨量( $t, t-1, t-2$ ) + 本站径流( $t-1, t-2$ )	3	0.003

2.2.4 滚动预报

为有效延长预见期,考虑在模型中不断加入预报降雨和预报径流信息作为下一预见期的输入因子,不断更新前期降雨和径流信息,实现7d预见期内的逐日滚动预报。

3 结果与分析

3.1 区间径流预报结果

2021年平班—董箐—龙滩区间的预报精度评价结果见表2。由表2可知,在预见期为未来1d时,平均相对误差  $M_{MARE}$  为 8.24%,平均绝对误差  $M_{MAE}$  为 47.09  $m^3/s$ 。

表 2 组合模式未来 1 d 预报精度评价

Tab. 2 Accuracy evaluation of 1-day runoff forecast results

项目	$M_{MAE}/(m^3 \cdot s^{-1})$	$M_{MARE}/\%$
整体平均	47.09	8.24
中、大流量	175.04	12.36
小流量	25.89	5.92

在1d预报方案的基础上,采用滚动预报的方式逐日预报未来7d的区间径流,结果见表3。由表3可知,在3、7d预见期下,预报平均相对误差分别为20.62%、39.71%,平均绝对误差分别为107.50、189.19  $m^3/s$ 。

表 3 区间径流 7 d 预报结果精度评价

Tab. 3 Precision evaluation of 7-day runoff forecast results

区间	纳什系数 $N_{NS}$	平均绝对误差 $M_{MAE}/(m^3 \cdot s^{-1})$	平均相对误差 $M_{MARE}/\%$
1 d	0.92	47.09	8.24
2 d	0.91	67.08	14.13
3 d	0.71	107.50	20.62
4 d	0.47	141.18	26.55
5 d	0.38	164.14	31.81
6 d	0.31	179.28	35.93
7 d	0.25	189.19	39.71

图2为区间降雨量、径流量2021-06-07相似样本分析。由图2可知,2021-06-07的实测径流量为635.88  $m^3/s$ ,相似性模型预报结果为652.21  $m^3/s$ ,预报准确率为97.43%,效果较好。该预报时段使用小流量模式进行预报,小流量模式降雨影响权重较小,搜索5个相似样本时以前

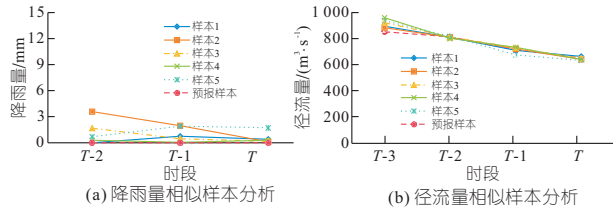


图 2 区间降雨量、径流量 2021-06-07 相似样本分析

Fig. 2 Similar sample analysis of rainfall and runoff on June 7, 2021

期径流相似为主导,模型搜索出的相似样本的确与实际情况较为相似,预报效果较好。

图3为平班—董箐—龙滩区间降雨量、径流量2021-07-02相似样本分析。由图3可知,2021-07-02的实测径流量为3321.99  $m^3/s$ ,相似性模型预报结果为2674.94  $m^3/s$ ,预报准确率为80.52%,预报结果比实测值偏低。该预报时段使用大流量模式进行预报,大流量模式降雨影响权重较大,搜索5个相似样本时以前期降雨相似为主导。除了样本5,其余时刻均无7月2日当天实测径流涨水过程,且从样本3开始,相似样本的前期降雨过程与实际预报过程相似性差,说明历史上缺乏真正与预报时段相似的降雨径流过程,从而导致预报效果较差。

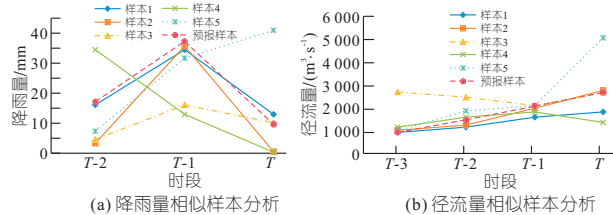


图 3 平班—董箐—龙滩区间径流 2021 年 7 月 2 日相似样本分析

Fig. 3 Similar sample analysis of runoff in Pingban-Dongqing-Longtan section on July 2, 2021

综上所述,当历史上发生降雨径流类似过程较多时,相似性预报模型能找到最相似的样本,得到准确率较高的预报结果,但当历史上发生降雨径流类似的过程较少时,预报精度则会降低。

3.2 与其他模型预报结果对比

为了比较相似性径流预报模型与机器学习方法的效果,还采用BP神经网络和支持向量机模型进行水文预报,精度评价见表4。由表4可知,3d预见期下,相似性预报模型、BP神经网络模型、支持向量机模型预报结果的平均相对误差分别为20.62%、23.76%、22.99%;5d预见期下,三种模型预报结果的平均相对误差分别为31.81%、36.18%、34.49%。在三种模型中,相似性径流预报模型精度最高,预报效果最好,支持向量机模型优于BP神经网络模型。

表 4 三种模型区间径流 7 d 预报结果精度评价

Tab. 4 Precision evaluation of 7-day runoff forecast results of three models

区 间	相似性预报模型			BP 神经网络模型			支持向量机模型		
	$N_{NS}$	$M_{MAE}$	$M_{MARE}$	$N_{NS}$	$M_{MAE}$	$M_{MARE}$	$N_{NS}$	$M_{MAE}$	$M_{MARE}$
1 d	0.92	47.09	8.24	0.91	47.96	8.39	0.91	47.72	8.35
2 d	0.91	67.08	14.13	0.89	71.62	15.09	0.9	69.79	14.70
3 d	0.71	107.5	20.62	0.67	123.89	23.76	0.69	119.84	22.99
4 d	0.47	141.18	26.55	0.4	161.81	30.43	0.43	149.08	28.04
5 d	0.38	164.14	31.81	0.3	186.68	36.18	0.33	177.98	34.49
6 d	0.31	179.28	35.93	0.22	205.5	41.18	0.24	194.12	38.90
7 d	0.25	189.19	39.71	0.16	221.69	46.53	0.17	209.87	44.05

注: $M_{MAE}$ 、 $M_{MARE}$  单位分别为  $m^3/s$ 、%。

### 3.3 入库径流预报结果

2021 年龙滩电站在滚动预报下的入库流量预报精度评价结果见表 5。龙滩入库流量通过区间径流预报结果和天二、董箐的出库计划相加得到。由表 5 可看出,在 1、3、7 d 预见期下,龙滩电站入库流量预报平均相对误差分别为 6.05%、11.33%、17.89%。

表 5 入库流量 7 d 预报结果精度评价

Tab. 5 Accuracy evaluation of 7-day forecast results of inflow

区 间	$N_{NS}$	$M_{MAE}$	$M_{MARE}$	区 间	$N_{NS}$	$M_{MAE}$	$M_{MARE}$
		$/(m^3 \cdot s^{-1})$	%			$/(m^3 \cdot s^{-1})$	%
1 d	0.93	77.15	6.05	5 d	0.38	196.26	14.62
2 d	0.82	108.31	8.12	6 d	0.32	206.44	15.26
3 d	0.66	152.57	11.33	7 d	0.24	240.33	17.89
4 d	0.52	183.64	13.61				

## 4 结论

a. 针对岩溶地区产汇流机制复杂、地下水监测资料不足、径流预测精度低等问题,提出了融合

物理机制和数据驱动的相似性预测方法,建立了不同流量量级、预报因子和前期影响滞时下最优组合预报方案,可实现不同水雨情条件下多模式的自适应匹配和智能切换,可生成具有可解释性的径流预报结果,并将该方法应用于龙滩水电站中。

b. 未来 1 d 时,龙滩电站区间径流平均相对误差为 8.24%,洪水时段平均相对误差为 12.36%,小流量时段平均相对误差为 5.92%;当预见期延长至 3、7 d 时,平均相对误差分别为 20.62%、39.71%。进一步考虑上游电站出库流量后,龙滩电站入库流量 1、3、7 d 预报平均相对误差分别为 6.05%、11.33%、17.89%。该结果提升了岩溶地区径流预报精度和有效预见期,可为大型电站科学调度运行提供参考和支撑。

### 参考文献:

- [1] 许波刘,董增川,洪娴. 集总式喀斯特水文模型构建及其应用[J]. 水资源保护, 2017, 33(2): 37-42, 58.
- [2] 郭生练,熊立华,杨井,等. 基于 DEM 的分布式流域水文物理模型[J]. 武汉水利电力大学学报, 2000, 33(6): 1-5.
- [3] 李晓,李致家,董佳瑞. SWAT 模型在伊河上游径流模拟中的应用[J]. 河海大学学报(自然科学版), 2009, 37(1): 23-26.
- [4] 林剑艺,程春田. 支持向量机在中长期径流预报中的应用[J]. 水利学报, 2006, 37(6): 681-686.
- [5] 谭乔凤,陈然,朱阳,等. 基于多因子最近邻抽样回归模型的径流相似性预报[J]. 河海大学学报(自然科学版), 2020, 48(6): 521-527.
- [6] 梁虹. 喀斯特流域地貌产流机制与产流特征[J]. 贵州师范大学学报(自然科学版), 1995(2): 23-28.

## Study on Similarity Runoff Forecasting of Longtan Hydropower Station

XIE Jin-ji<sup>1</sup>, CHEN Fen-e<sup>1</sup>, YANG Shan<sup>2</sup>, LI Ling-yan<sup>1</sup>, ZHONG Cheng<sup>1</sup>, WEN Xin<sup>2</sup>

(1. Guangxi Guiguan Electric Power Co., Ltd., Nanning 530000, China;

2. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

**Abstract:** In response to the complex mechanism of runoff yield and concentration, insufficient groundwater monitoring data, and low accuracy of runoff prediction in Karst areas, a similarity prediction method that integrates physical mechanisms and data-driven approaches has been proposed. The optimal combination prediction scheme has been established for different flow magnitudes, prediction factors, and preceding affected lag-time, which can achieve adaptive matching and intelligent switching of multiple modes under different water and rain conditions. It can generate interpretable runoff prediction results. This method is applied to the runoff prediction of Longtan Hydropower Station. The results indicate that this method improves the accuracy and effective prediction period of runoff in Karst areas, which has scientific significance and practical value for ensuring the scientific scheduling and safe operation of large power plants.

**Key words:** similarity identification; nearest neighbor sampling; rolling forecast; karst; multi-mode switching