

张达文,高玉丹.基于多尺度水文特征与 CatBoost 的日均流量预测[J].人民珠江,DOI:10.3969/j.issn.1001-9235.XXXX.XX.001.

基于多尺度水文特征与 CatBoost 的日均流量预测

张达文,高玉丹

(广东省水文局梅州水文分局,广东 梅州 514000)

摘要:针对短序列日径流预测前期状态刻画不足及单年验证代表性有限的问题,基于梅江横山(二)站 2021—2025 年水文气象及水库资料构建 33 维多尺度特征,采用 CatBoost 预测次日日均流量。以 2021—2024 年数据开展扩展窗口 5 折时序交叉验证和贝叶斯寻优,以 2025 年数据独立检验。结果表明:固定参数 CatBoost 检验集纳什效率系数为 0.918、Kling-Gupta 效率为 0.956,较 LightGBM 和 ExtraTrees 分别相对提升 8.0% 和 3.0%,均方根误差为 69.5 m³/s。消融结果表明,降雨湿润特征贡献 0.447 的纳什效率系数增益,流量记忆特征提供基线能力,水库状态特征改善枯水期平均绝对误差和百分比偏差。贝叶斯优化未提高独立检验精度,汛期预测区间宽度约为枯水期的 2.7 倍。该方法可为日尺度滚动调度和分级防汛会商提供依据。

关键词:日均流量预测;CatBoost;多尺度特征;贝叶斯优化;时序交叉验证;梅江

中图分类号: **文献标识码:** **文章编号:**1001-9235(XXXX)XX-0001-15

日均流量预测是防洪调度、水资源配置与生态管控的重要支撑。南方季风型中小流域受多水库联合调控,径流同时受前期湿润状态、水热条件和人工蓄泄影响,呈现非线性、时滞与季节异方差;业务资料年限偏短又限制了干湿周期刻画。概念水文模型虽具备明确机理,但复杂水库群调度难以参数化,在短资料流域的不确定性较大^[1]。当前日径流机器学习预测主要分为两类技术路线。第一类依托时序分解算法开展序列预处理:张力等^[2]采用季节分解集成模型预测非平稳径流;王秀杰等^[3]融合时变滤波分解、关联筛选与 LightGBM 取得较高模拟精度;谭永杰等^[4]对比多类分解-LSTM(Long Short-Term Memory)组合方案,证实多尺度信息可改善预报效果。但该类方法依赖分解重构操作,易引入额外误差,分解分量与流域退水、降雨累积等实际水文过程缺乏直观对应关系。第二类直接构建多因子输入并优化模型架构:黄一凡等^[5]采用贝

叶斯优化 LSTM 解析径流时滞规律;崔杰连等^[6]搭建深度学习入库径流预报模型;张玮等^[7]引入物理约束优化水库调度网络;黄泽禧等^[8]耦合概念模型与长短时记忆网络提升多预见期精度;王毅等^[9]验证 CatBoost 在单水库径流模拟中的优势,其有序提升结构可有效降低预测偏移。不过,国内同类研究多偏向 LSTM、EMD(Empirical Mode Decomposition)分解等技术路线,针对多水库调控、5 年短样本、多尺度表格化水文特征场景的算法适配性论证仍显不足。

相较 LSTM 等深度学习模型,CatBoost 更适配 5 年短样本、多尺度表格化特征:梯度提升树适合中小规模结构化数据,对稀疏降雨、非线性交互和缺失值较鲁棒;有序提升与对称树正则化可抑制预测偏移和过拟合^[10],然而,仅有算法适配并不足以系统性解决当前研究面临的三大共性瓶颈。现有研究仍有三点不足:一是随机 K 折或单年验证忽略水

文过程的路径依赖,随机打乱会使土壤含水量、基流等状态携带未来信息,造成时序泄露并违背“ $t-1$ 日”因果约束;二是特征工程缺少水文语义与递进消融,难以量化降雨湿润和水库调蓄的独立增益;三是仅用NSE等整体指标,易掩盖洪峰低估和枯水偏差,难以支撑分级会商。针对上述不足,区别于侧重单一算法改进或分解重构的研究,本研究构建了“物理机制驱动-业务约束验证-不确定性量化”的完整闭环范式,其创新点主要体现在以下3个方面:①物理机制驱动,搭建包含流量记忆、降雨湿润、蒸发季节、水库调蓄4类共33维具备明确水文语义的多尺度特征体系,通过分层消融实验定量甄别各类特征的独立贡献,厘清水库群聚合调蓄对枯水期模拟的偏差校正机制;②业务约束验证,构建严格遵循“ $t-1$ 日”业务预报因果约束的高质量日尺度样本集,建立“扩展窗口5折时序交叉验证+贝叶斯超参数寻优+2025年完全独立年份检验”的双重验证范式,从源头规避未来信息泄露,确立适用于业务化部署的稳健参数优选策略;③不确定性量化,从汛枯分期、流量分级、极端洪峰、概率预测区间多维度量化模型不确定性边界,揭示季节异方差特征与极端洪峰低估机理,为日尺度滚动调度与分级防汛会商提供可落地的量化依据。

1 数据与方法

1.1 研究区域数据来源

横山(二)水文站位于梅江干流下游,站以上集水面积12 954 km²,是梅江下游重要控制站。站以上流域地势总体由北部和东西两侧向中部河谷降低,梅江干流及主要支流构成流域汇流骨架。流域内布设雨量站、水库站、蒸发站等监测设施,流域及主要测站分布见图1。

研究采用2021年1月1日至2025年12月31日共1 826 d逐日资料。流域面平均降雨量由约184个雨量站资料经泰森多边形法计算;日均流量为横山(二)站实测整编成果;蒸发量来自流域内蒸发皿观测;14座水库逐日蓄水量相加得到流域总蓄水量。降雨、蒸发、蓄水和流量单位分别为mm、mm、

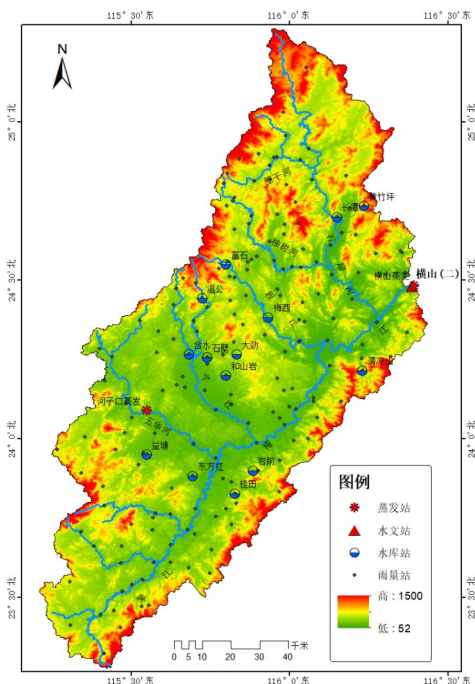


图1 横山(二)站以上流域概况

Fig. 1 Overview of the watershed upstream of the Hengshan (II) Station

10⁶ m³和m³/s。各变量均可在预报发布前获得,任务定义为利用 $t-1$ 日及以前信息预测 t 日日均流量。5年样本虽不足以支撑长周期气候态模拟,但完全满足短序列业务建模与滚动预报检验需求,见图2。

数据按日期统一到日尺度。流量均值为265.9 m³/s,最大值为6 432 m³/s;2021—2025年平均流量分别为73.5、287.2、276.9、468.1和223.3 m³/s,表明研究期内水文条件差异较大。水库日蓄水变化量为:

$$\Delta S_t = S_t - S_{t-1} \quad (1)$$

式中: S_t 为第 t 日流域水库总蓄水量; $\Delta S_t > 0$ 表示总体蓄水, $\Delta S_t < 0$ 表示总体消落。

1.2 数据预处理与样本构建

本研究区分短序列业务建模与气候态模拟:2021—2025年资料虽不足以代表长期气候变化,却覆盖丰、平、枯年型并满足“ $t-1$ 日”数据可得性要求,可用于检验滚动一步预报。

数据处理严格遵循“质量控制—因果移位—滚动统计”流程:①以日期为主键合并降雨、蒸发、流量和蓄水资料,检查日序连续性、重复记录、空值和

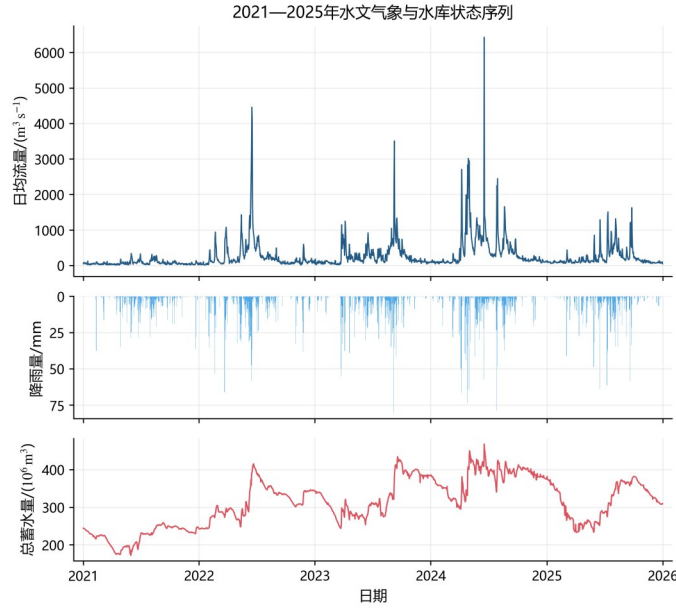


图2 2021—2025年水文气象与水库状态序列

Fig. 2 Hydrometeorological and reservoir status time series, 2021 - 2025

非负性, 1826 条原始记录未发现重复日期和缺测值, 不对洪水过程进行随机插值或平滑; ②按相邻日总蓄水差重建 ΔS_t , 修正原始表中变化量时间标记与预测日不一致的问题; ③所有输入先向后移动 1 d, 再计算滞后项或滚动统计, 确保特征仅使用发布时点已知信息, 验证期内 Q_{t-1} 采用前一日实测值, 符合滚动一步业务条件; ④最长窗口为 30 d, 删除初始化产生的空值后得到 1 795 组样本。Ridge 的标准化参数仅由相应训练集估计, 树模型保留原量纲; 负预测值按流量物理下限截断为 0。

样本集构建与计算过程: 原始数据集涵盖 2021 年 1 月 1 日至 2025 年 12 月 31 日, 共 1 826 条逐日观测记录。本研究构建了最长 30 d 的滚动统计特征 (如 Q_{30} 、 P_{sum30}), 需使用前 30 d 的历史数据进行计算, 因此序列前 30 d (2021 年 1 月 1 日—1 月 30 日) 无法生成完整的特征向量。因此, 删除初始化产生的 31 条空值记录 (含 2021 年 1 月 31 日, 因其无法构成完整的 30 d 回溯窗口), 最终得到有效样本 1 795 组。

2021 年 2 月 1 日至 2024 年 12 月 31 日的 1 430 组样本用于模型率定, 2025 年的 365 组样本作为独立检验集。模型率定集内部不随机打乱, 而采用扩展窗口时序切分, 使每个验证块均严格晚于对应训练

块。该设计同时保留季节转换和年际差异, 避免常规随机交叉验证将未来水文状态泄漏到过去。

1.3 研究框架

研究流程见图 3, 包括 4 个层次: 多源资料质控和因果对齐; 具有明确水文含义的 33 维多尺度特征构建; 持续性、Ridge、ExtraTrees、LightGBM 与 CatBoost 分层对比, 以及 CatBoost 贝叶斯寻优和 5 折时序验证; 最后通过全年与汛枯期消融、流量等级、洪峰误差、置换重要性和经验预测区间评价精度、稳定性与适用边界。

2 预测模型构建

2.1 多尺度预测特征

33 维特征并非无约束扩维, 而是围绕日尺度产汇流响应的 4 类可观测信息组织。特征窗口覆盖 1 ~ 30 d, 使模型同时获得快速降雨响应、河道退水、前期湿润和季节背景; 逐组消融用于判断每类信息是否产生独立增益。

a) 流量记忆特征, 共 11 项。前 1、2、3、7、14、30 d 流量, 前 3、7、14、30 d 平均流量和前 7 d 流量标准差。以 w 日平均流量为例:

$$\bar{Q}_{t,w} = \frac{1}{w} \sum_{i=1}^w Q_{t-i}, w \in \{3, 7, 14, 30\} \quad (2)$$

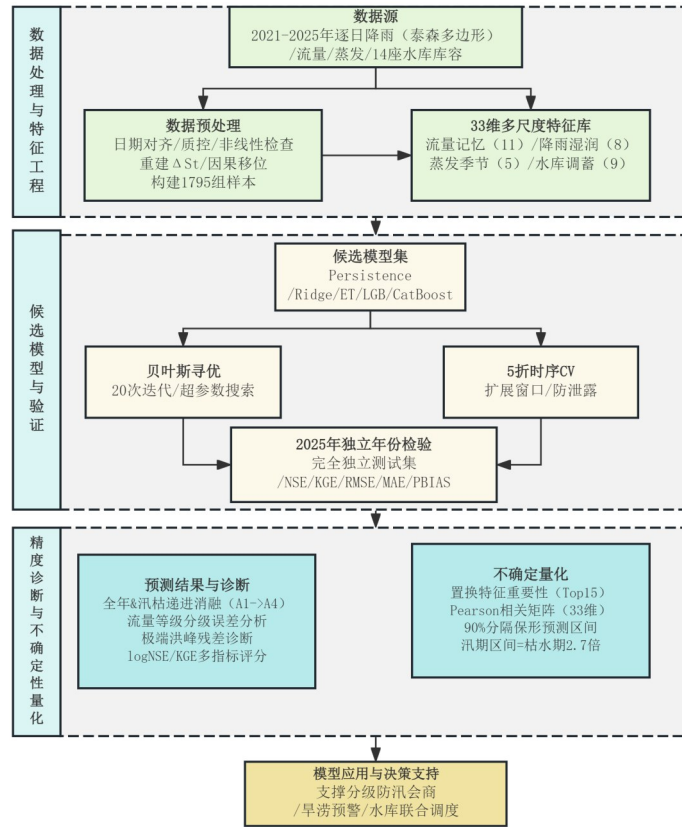


图3 基于多尺度特征与 CatBoost 的机器学习日均流量预测框架

Fig. 3 A machine learning framework for daily traffic forecasting based on multiscale features and CatBoost

短滞后项表示河道连续性,长滞后和滚动统计近似描述退水与基流记忆。

b)降雨湿润特征,共8项。前1、2、3、7 d降雨量,前3、7、14 d累计降雨量和前期降水指数(Antecedent Precipitation Index, API):

$$API_t = k \cdot API_{t-1} + P_{t-1} \quad (3)$$

式中: P_{t-1} 为前1 d降雨量, k 为衰减系数。上述多尺度特征的时间窗口选取具有明确的水文物理意义: 1~3 d窗口表征快速产流与河道响应, 7 d反映退水初期和表层土壤水变化, 14 d近似地下水补给缓变过程, 30 d描述基流稳定性与月尺度节律。 k 参考 GB/T 22482—2008《水文情报预报规范》的湿润地区取值, 并在 [0.70, 0.90] 内开展敏感性试验(表1); NSE变化小于0.005、RMSE变化小于1.5 m³/s, 表明 $k=0.85$ 适用于本研究且结果对其不敏感。

c)蒸发季节特征,共5项:前1、3 d蒸发量、前7 d平均蒸发量和年内日序的正余弦编码:

表1 API衰减系数 k 敏感性分析结果Tab. 1 Results of the sensitivity analysis for the API attenuation coefficient k

衰减系数	NSE	RMSE/(m ³ ·s ⁻¹)	MAE/(m ³ ·s ⁻¹)
0.7	0.917	70.1	39.0
0.85(推荐值)	0.918	69.5	38.7
0.9	0.918	69.8	38.9

$$D_{\sin} = \sin\left(\frac{2\pi d}{365.25}\right), D_{\cos} = \cos\left(\frac{2\pi d}{365.25}\right) \quad (4)$$

式中 d 为年内日序。循环编码避免年末与年初数值不连续,并提供季节性耗水和基流背景。

d)水库状态特征,共9项:前1、7、30 d总蓄水量,前1、3、7、14、30 d蓄水变化量和前7 d平均蓄水变化量。总蓄水表示慢变工程状态,蓄水变化表示近期蓄泄方向。由于缺少逐库入库、出库、闸门过程和传播时间,该组变量只能作为辅助状态,不能解释为完整的调度行为或因果贡献。

表 2 日均流量预测特征体系

Tab. 2 Feature system for forecasting average daily traffic

特征组	数量	主要变量	水文含义
流量记忆	11	$Q_{lag1}, Q_{lag2}, Q_{lag3}, Q_{lag7}, Q_{lag14}, Q_{lag30}, Q_{mean3}, Q_{mean7}, Q_{mean14}, Q_{mean30}, Q_{std7}$	河道连续性与退水记忆
降雨湿润	8	$P_{lag1}, P_{lag2}, P_{lag3}, P_{lag7}, P_{sum3}, P_{sum7}, P_{sum14}, API_{0.85}$	降雨输入与前期湿润程度
蒸发季节	5	$E_{lag1}, E_{lag3}, E_{mean7}, D_{sin}, D_{cos}$	季节背景和水分消耗
水库状态	9	$S_{lag1}, S_{lag7}, S_{lag30}, \Delta S_{lag1}, \Delta S_{lag3}, \Delta S_{lag7}, \Delta S_{lag14}, \Delta S_{lag30}, \Delta S_{mean7}$	水库群总体状态与蓄泄趋势
合计	33	预测时点可获得的多尺度特征	预测第 t 日日均流量

注: Q_{lagn} 表示预测日前 n 日流量, 即 Q_{t-n} ; P_{sumn} 表示前 n 日累计降雨量; 所有特征均在 $t-1$ 日及以前可得。

2.2 对比模型与 CatBoost

为区分时间连续性、特征工程和非线性算法的贡献, 设置 5 类模型: 持续性模型以 $\hat{Q}_t = Q_{t-1}$ 为严格业务基准; Ridge 回归检验 33 维特征的线性可预测性; ExtraTrees 利用随机分裂和多树平均检验方差降低效果; LightGBM 利用直方图、梯度单边采样和叶子优先生长拟合非线性阈值关系^[11]; CatBoost 通过有序提升减少梯度估计偏移, 并以对称树限制模型复杂度^[10]。所有模型使用相同数据边界和输入信息。

设输入为 x_t , CatBoost 加法模型为:

$$\hat{Q}_t = F_M(x_t) = F_0(x_t) + \sum_{m=1}^M \eta h_m(x_t) \quad (5)$$

式中: M 为迭代树数量, η 为学习率, h_m 为第 m 棵对称回归树。损失函数采用均方误差:

$$L = \frac{1}{n} \sum_{t=1}^n (Q_t - \hat{Q}_t)^2 \quad (6)$$

各模型的参数配置、搜索空间及调参预算、早停规则见表 3。Ridge、ExtraTrees、LightGBM 及主比较中的 CatBoost 均采用预设固定参数; 另对 CatBoost 开展贝叶斯优化以检验参数稳健性。优化过程仅使用 2021—2024 年资料, 2025 年独立检验数据不参与参数选择。

2.3 贝叶斯寻优与 5 折时序验证

贝叶斯优化以已评估参数及目标值拟合代理模型, 再由采集函数选择候选点。已有研究表明, 该方法可提高 LSTM 年径流预测的参数搜索效率^[12]; 其探索-利用平衡原理可参见相关综述^[13]。本研究采用 Matérn 核高斯过程和期望改进(EI)采

集函数。搜索空间为: 迭代树数 200 ~ 650、树深 4 ~ 8、学习率 0.015 ~ 0.120、L2 叶节点正则系数 1 ~ 20、随机强度 0 ~ 2、袋装温度 0 ~ 2; 学习率和 L2 系数按对数尺度映射。基准预算为 20 组(前 6 组随机初始化、后 14 组由 EI 选择), 随机种子为 42。

搜索空间兼顾 CatBoost 算法特性与短序列水文场景: 树深 4 ~ 8 平衡非线性表达与过拟合, 学习率—迭代次数采用低学习率、多迭代组合, L2 正则用于约束叶节点权重。沿同一搜索序列将预算由 20 组追加至 30 组和 50 组(表 4), 3 种预算均在第 3 组获得相同最优参数, 5 折平均 NSE 均为 0.797, 2025 年 NSE、KGE 和 RMSE 亦未变化, 说明 20 组在既定搜索空间和随机种子下已收敛, 该结论仅限于本研究数据条件。固定随机种子 42 确保了高斯过程代理模型初始化及采样的一致性, 满足机器学习实验可复现性的基本要求。

2.4 评价指标

采用 NSE、KGE、RMSE、MAE、PBIAS、logNSE 和相关系数 r 综合评价:

$$NSE = 1 - \frac{\sum_{t=1}^n (Q_t - \hat{Q}_t)^2}{\sum_{t=1}^n (Q_t - \bar{Q})^2} \quad (7)$$

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Q_t - \hat{Q}_t)^2}, MAE = \frac{1}{n} \sum_{t=1}^n |Q_t - \hat{Q}_t| \quad (9)$$

表3 各模型参数及调参设置

Tab. 3 Model parameters and hyperparameter settings

模型	最终参数设置	搜索空间与调参预算	早停规则
Ridge	StandardScaler,alpha=10.0,solver='auto' tol=1×10 ⁻⁴ ,max_iter=None,fit_intercept=True	未搜索;1组预设参数	不适用。
ExtraTrees	n_estimators=700,criterion=squared_error,max_depth=None, max_features=0.85,min_samples_split=2,min_samples_leaf=2,bootstrap=False,random_state=42,n_jobs=-1	未搜索;1组预设参数	未启用;完整训练700棵树
LightGBM	boosting_type=gbdt,n_estimators=600, learning_rate=0.03, num_leaves=31, max_depth=-1,min_child_samples=20, subsample=0.9,subsample_freq=0,colsample_bytree=0.9, reg_alpha=0,reg_lambda=1,random_state=42,n_jobs=-1	未搜索;1组预设参数	未启用;完整训练600轮
CatBoost	主比较: iterations=700,depth=6, learning_rate=0.03,l2_leaf_reg=3,loss_function=RMSE, random_seed=42 贝叶斯优化最优: iterations=490,depth=7, learning_rate=0.0377,l2_leaf_reg=1.9754,random_strength=1.1092,bagging_temperature=0.1276	高斯过程贝叶斯优化(Matérn核、EI采集函数); iterations为200-650,depth为4-8,learning_rate为0.015-0.120,l2_leaf_reg为1-20, random_strength和bagging_temperature均为0~2;共20次试验(前6次随机初始化,后14次EI选择);目标为5折扩展窗口时序交叉验证平均NSE	未启用;各试验按设定iterations完整训练

注:各模型未采用相同调参预算。固定参数用于主模型比较;CatBoost的20次贝叶斯优化仅作为参数稳健性分析。所有模型均未采用早停,未列参数采用相应默认值。

表4 贝叶斯搜索预算敏感性结果

Tab. 4 Bayesian search budget sensitivity results

搜索预算/组	最优试验	最优参数(树数/深度/学习率/1.2/随机强度/袋装温度)	5折NSE	2025年NSE	2025年KGE	2025年RMSE/(m ³ ·s ⁻¹)
20	3	490/7/0.0377/1.975/1.109/0.128	0.797	0.904	0.935	75.3
30	3	490/7/0.0377/1.975/1.109/0.128	0.797	0.904	0.935	75.3
50	3	490/7/0.0377/1.975/1.109/0.128	0.797	0.904	0.935	75.3

$$\text{PBIAS} = \frac{\sum_{t=1}^n (\hat{Q}_t - Q_t)}{\sum_{t=1}^n Q_t} \times 100\% \quad (10)$$

$$\ln \text{NSE} = 1 - \frac{\sum_{t=1}^n (\ln Q_t - \ln \hat{Q}_t)^2}{\sum_{t=1}^n (\ln Q_t - \overline{\ln Q})^2} \quad (11)$$

式中: Q_t 和 $\hat{Q}_t$ 分别为实测和预测流量, $\bar{Q}$ 为实测均值, $\alpha = \sigma_{\hat{Q}}/\sigma_Q$, $\beta = \mu_{\hat{Q}}/\mu_Q$ 。NSE和RMSE对洪峰平方误差敏感,MAE反映一般日期误差,KGE兼顾相关性、变幅和均值偏差,PBIAS表示总体高估或低估,lnNSE补充评价中低流量。

3 实验过程

3.1 模型训练与参数稳健性试验

实验先在2021—2024年开发集内运行5折时序验证,分别评价固定参数和贝叶斯优化参数;再使用全部开发集重新训练模型,对2025年作一次性独立预测。该顺序保证2025年结果不反馈到寻优过程。Ridge的标准化统计量按折重新估计,所有随机模型统一随机种子42。为避免将独立检验中的偶然高分解释为参数优势,同时报告5折均值、标准差、最小值和最大值,并保留两种CatBoost配置的2025年结果。

3.2 全年与汛枯期特征消融

设置4组递进特征：A1为3项近期流量；A2为11项多尺度流量；A3在A2基础上加入降雨、API、蒸发和季节项，共24项；A4再加入9项水库状态，共33项。消融模型采用固定 CatBoost 参数和相同训练-检验边界。除全年指标外，将4—9月定义为汛期、10月至翌年3月定义为枯水期，分别计算NSE、MAE和PBIAS，以识别不同特征在高变幅和低变幅时期的作用。

3.3 极端误差与不确定性诊断

按2025年实测流量20%、80%和95%分位数划分低、中、高和极高流量等级；对最大的5个实测流量日期，在前后1 d窗口内比较峰值相对误差和峰现时间误差。置换重要性通过逐项打乱特征并计算均方误差增量，表示模型对变量的预测依赖，不作因果解释。

点预测之外，采用两类残差分位区间评价不确定性。其一，按2021—2023年训练、2024年校准、2025年检验的顺序构建90%分割保形区间；其二，在贝叶斯参数稳健性试验中汇总5个时序验证块的折外绝对残差，分别计算全年以及汛、枯期90%分位数，形成时序折外经验区间。前者用于检验统一区间的覆盖率，后者用于诊断季节异方差。保形方

法不要求预设残差分布，但有限样本覆盖仍受时间非平稳性和校准样本代表性影响^[14-15]。

4 结果分析及模型评价

4.1 参数稳定性与独立验证

固定参数 CatBoost 的5折平均NSE为0.772±0.114(折间标准差)；贝叶斯优化后为0.797±0.083(折间标准差)，NSE折间标准差降低27.2%，5折验证集NSE实测范围为0.692~0.897。KGE均值由0.755升至0.769，RMSE和MAE结果见表5。联合寻优改善了开发期平均性能，但各折在水文状态转换或高流量时段仍有明显差异，尤其包含水文状态转换或高流量事件的验证块误差较大。

2025年独立检验中，固定参数与贝叶斯参数的NSE分别为0.918和0.904，KGE分别为0.956和0.935。贝叶斯参数针对2021—2024年单目标NSE精细适配，降低了开发期折间波动，却可能放大对该期降雨—径流关系和调度节律的依赖；2025年水文分布偏移时，泛化性能反而下降。由于2025年未参与寻优且不用于二次调参，后续比较采用预设固定参数，该结论适用于短序列业务滚动预报场景，不推广至长周期气候模拟，贝叶斯结果仅作为稳健性证据。

表5 CatBoost 参数方案的时序验证与独立检验结果

Tab. 5 Time-series validation and independent test results for the CatBoost parameter scheme

参数方案	5折平均NSE	5折平均KGE	5折RMSE/(m ³ ·s ⁻¹)	5折MAE/(m ³ ·s ⁻¹)
固定参数	0.772±0.114	0.755±0.103	166.3±133.3	70.1±30.5
贝叶斯优化	0.797±0.083	0.769±0.128	161.4±134.2	70.1±35.3
参数方案	2025年NSE	2025年KGE	2025年RMSE/(m ³ ·s ⁻¹)	2025年MAE/(m ³ ·s ⁻¹)
固定参数	0.918	0.956	69.5	38.7
贝叶斯优化	0.904	0.935	75.3	41.7

5类模型的2025年结果见表6。固定参数 CatBoost 的NSE、KGE分别为0.918、0.956，RMSE和MAE分别为69.5、38.7 m³/s，较持续性模型RMSE降低52.6%，较Ridge和ExtraTrees分别降低23.4%和13.4%。这说明仅依赖昨日流量不足以捕捉快

速涨水，且多尺度输入与有序梯度提升共同贡献于非线性预测能力。

4.2 多尺度特征及汛枯期贡献

全年消融中，A1仅用3项近期流量时NSE为0.326，A2加入长滞后与滚动统计后提高至0.472；

贝叶斯寻优、时序交叉验证与汛枯期特征贡献

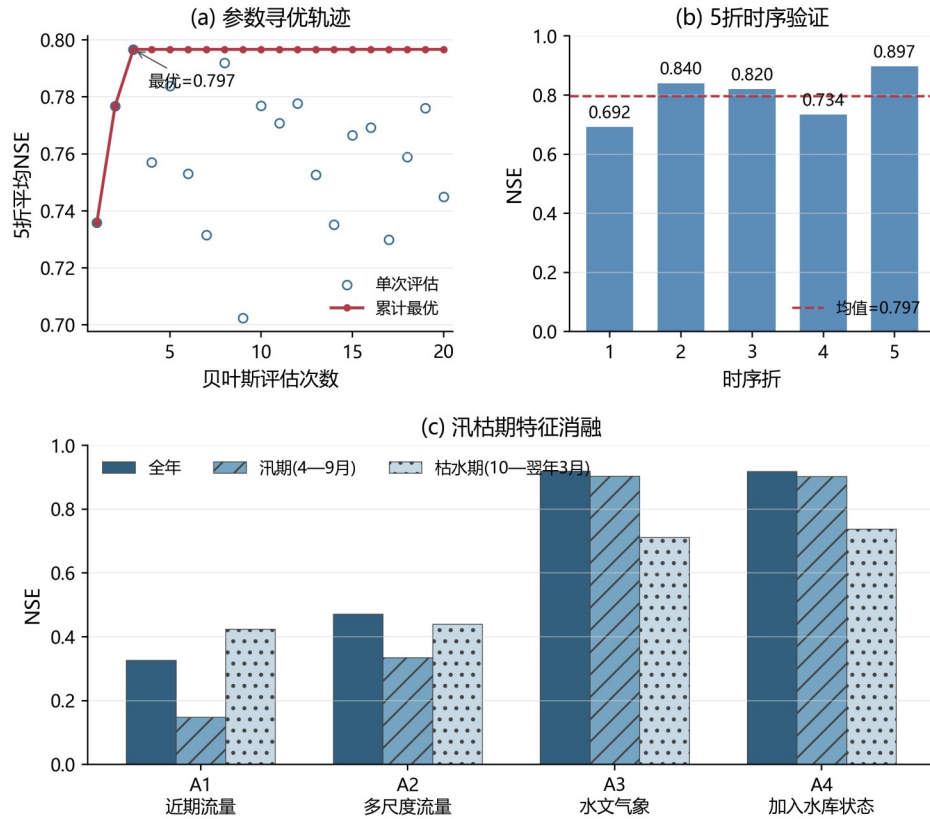


图4 贝叶斯寻优、时序交叉验证与汛枯期特征贡献

Fig. 4 Bayesian optimization, time-series cross-validation, and the contribution of flood and dry season features

表6 2025年独立检验期模型性能

Tab. 6 Model performance during the independent validation period in 2025

模型	NSE	KGE	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	MAE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	PBIAS/%
Persistence	0.636	0.818	146.8	69.5	0.06
Ridge	0.861	0.886	90.8	56.0	2.24
ExtraTrees	0.891	0.910	80.3	45.4	5.55
LightGBM	0.850	0.861	94.3	48.4	6.12
CatBoost	0.918	0.956	69.5	38.7	1.47

A3 加入降雨、API、蒸发和季节信息后跃升至 0.919, MAE 由 $78.6 \text{ m}^3/\text{s}$ 降至 $41.0 \text{ m}^3/\text{s}$ 。主要增益来自预报时点已知的降雨、前期湿润和季节背景, 33 维体系的价值在于水文语义分组及可验证的信息增益, 而非维数本身。

A4 加入水库状态后全年 NSE 由 0.919 变为 0.918, 差异可忽略, 但 MAE 由 $41.1 \text{ m}^3/\text{s}$ 降至 $38.7 \text{ m}^3/\text{s}$, KGE 由 0.949 提高至 0.956, PBIAS 由 3.09% 降至 1.47%。季节结果进一步揭示作用差异: 汛期 A3 与 A4 的 NSE 分别为 0.903 和 0.902, 水库状态未增加总体 NSE, 但 MAE 由 $61.5 \text{ m}^3/\text{s}$ 降至 $57.9 \text{ m}^3/\text{s}$, 同时 KGE 从 0.938 提升至 0.945, logNSE 从 0.877 提升至 0.889 (表 7), 表明过程结构拟合更优。枯水期 NSE 由 0.711 提高至 0.738, KGE 由 0.854 提升至 0.868, logNSE 由 0.584 跃升至 0.643, MAE 由 $20.5 \text{ m}^3/\text{s}$ 降至 $19.3 \text{ m}^3/\text{s}$, PBIAS 由 1.14% 降至 -0.04% 。上述 KGE 与 lnNSE 的同步改善, 有效解决了单一 NSE 指标对枯水期低流量刻画不足的问题, 证明聚合蓄水信息不仅优化了总量偏差, 更显著改善了低流量的比例关系与拟合精度。因此, 聚合蓄水信息更接近慢变状态和总量偏差校正因子, 对快速洪峰的边际信息

有限。

4.3 过程拟合与极端洪峰误差机理

2025 年过程线和散点-历时曲线见图 5、6。CatBoost 能够跟踪主要涨退水过程,多数高流量事件峰现误差为 0~1 d;中低流量历时曲线与实测较接近。但 9 月 25 日实测 $1\,630\text{ m}^3/\text{s}$ 、预测 $968.8\text{ m}^3/\text{s}$,相对低估 40.6%;8 月 5 日洪峰高估 33.3%。这种“峰现较准、峰量偏差较大”的现象表明,滞后流量和面平均降雨能够定位过程发生时间,却不足以稳定刻画极端降雨空间集中度、产流阈值和水库实时

泄放共同造成的峰量变化。

分级结果见图 7。CatBoost 在中、高、极高流量等级的 MAE 分别为 27.3、77.0 和 $149.4\text{ m}^3/\text{s}$;极高流量 NSE 仅 0.287,明显低于全年值。误差扩大主要有 3 方面原因:极高流量样本仅 19 d,平方损失仍由大量一般样本共同决定;树模型以训练样本叶节点均值组合预测,面对超出历史组合的事件外推能力有限;现有输入缺少小时雨强、降雨空间分布、土壤含水量、上游断面流量和逐库泄量。低流量组内方差很小,NSE 会出现较大负值,故枯水评价应优先

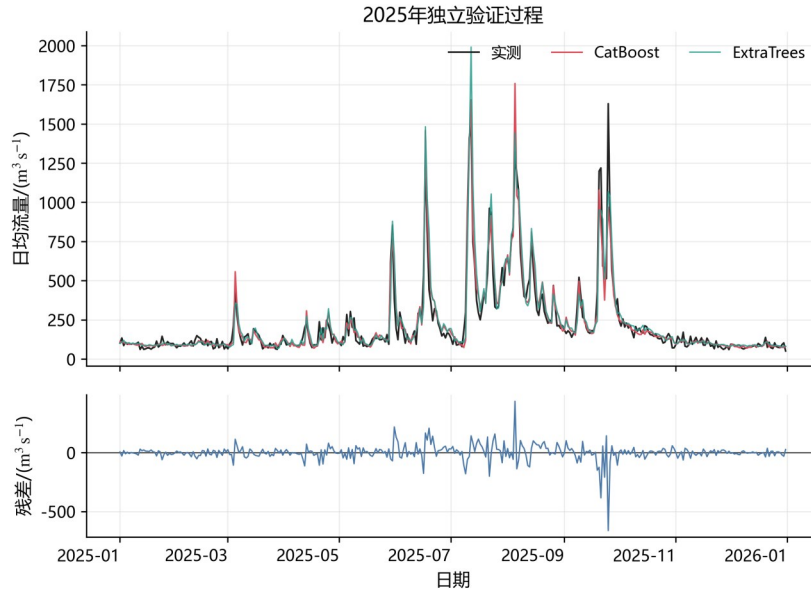


图 5 2025 年独立验证过程

Fig. 5 Independent verification process for 2025

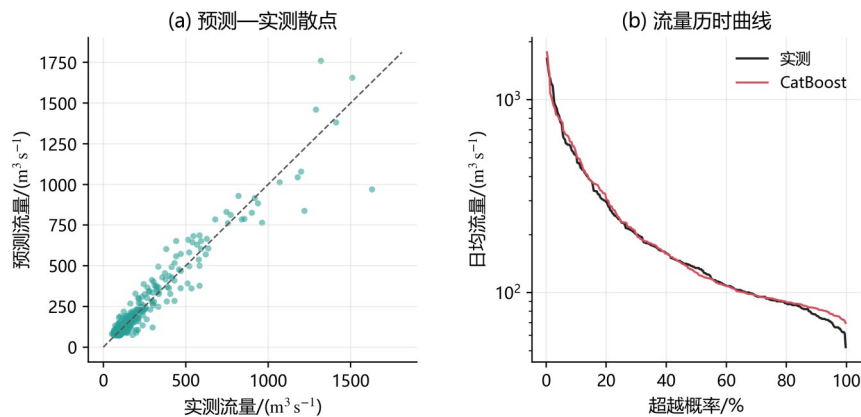


图 6 预测散点与流量历时曲线

Fig. 6 Predicted scatter plots and traffic duration curves

表7 全年及汛枯期特征消融多指标结果

Tab. 7 Results of multiple melting indices for the entire year and during the flood and dry seasons

特征组	时段	NSE	KGE	MAE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	PBIAS/%	logNSE
A1	全年	0.326	0.659	78.6	5.98	0.766
A1	汛期	0.148	0.592	131.9	6.24	0.697
A1	枯水期	0.424	0.713	24.9	5.23	0.452
A2	全年	0.472	0.742	75.0	2.77	0.762
A2	汛期	0.335	0.684	124.7	2.21	0.685
A2	枯水期	0.440	0.732	25.1	4.41	0.468
A3	全年	0.919	0.949	41.1	3.09	0.882
A3	汛期	0.903	0.938	61.5	3.75	0.877
A3	枯水期	0.711	0.854	20.5	1.14	0.584
A4	全年	0.918	0.956	38.7	1.47	0.895
A4	汛期	0.902	0.945	57.9	1.98	0.889
A4	枯水期	0.738	0.868	19.3	-0.04	0.643

结合 MAE、PBIAS 和对数 NSE。

4.4 不确定性与特征解释

模型预测不确定性呈现出显著的季节异方差特征(图8)。统一 90% 分割保形区间在 2025 年的覆盖率为 97.3%, 平均宽度为 $299.7 \text{ m}^3/\text{s}$, 整体偏保守。在贝叶斯参数方案的时序折外诊断中, 统一残

差分位区间覆盖率为 95.3%、平均宽度 $250.6 \text{ m}^3/\text{s}$; 改用汛枯期条件分位数后, 全年覆盖率为 96.7%, 汛期和枯水期平均宽度分别为 374.0 和 $139.0 \text{ m}^3/\text{s}$ 。汛期所需区间约为枯水期的 2.7 倍, 证明误差具有明显季节异方差。条件区间没有缩小全年平均宽度, 但避免用同一尺度描述洪水期与枯水期风险, 更适合业务分级展示。

为判别 33 维特征的有效性与冗余性, 本研究从统计关联、置换贡献和分组消融三方面诊断: ① Pearson 相关矩阵(图9)显示, 流量记忆组和降雨湿润组内部相关性较高, 存在组内重复, 水文气象与水库状态组之间相关性较弱, 提供了不同物理过程的互补信息; ② 置换重要性(图10)表明, 前 1 d 降雨、前 1 d 流量、API 和 3 d 累计降雨为主要预测因子; 部分长滞后蓄水变化特征的条件增量接近 0; ③ 共线特征并非全部删除, CatBoost 可在相关变量间自适应选择分裂, 递进消融显示, 保留多尺度变量有助于维持 KGE 和枯水期 PBIAS 的稳定性。

A2→A3 使 NSE 由 0.472 升至 0.919, 证明降雨湿润与季节信息是主要增益来源; A3→A4 虽未提高全年 NSE, 却使 MAE 由 41.1 降至 $38.7 \text{ m}^3/\text{s}$ 、枯水期 NSE 由 0.711 升至 0.738, 说明水库状态主要承担平稳期偏差校正。该格局也解释了汛期降雨波动主导高不确定性、枯水期水库状态辅助校正的季节差异。

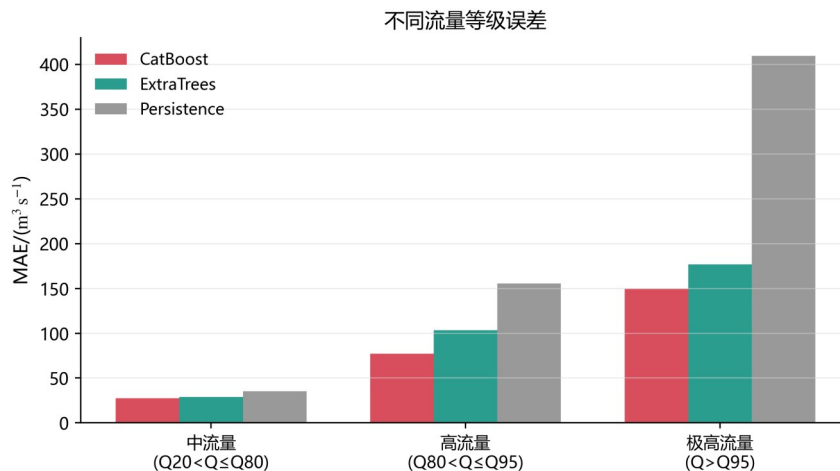


图7 不同流量等级误差

Fig. 7 Errors at different flow rates

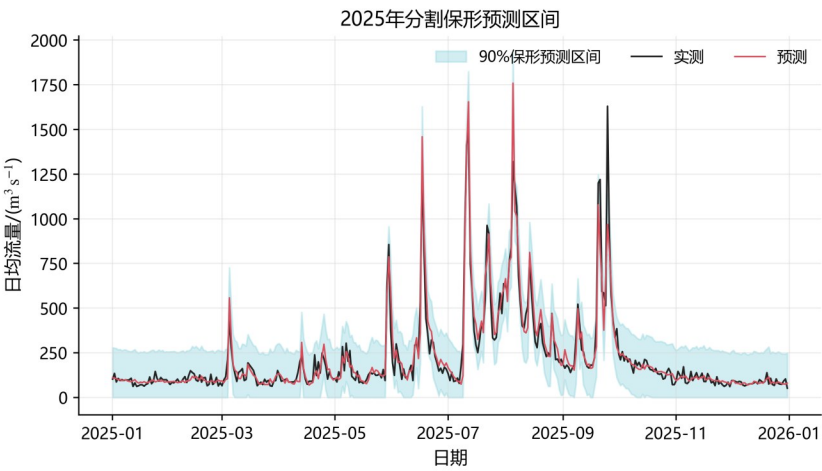


图8 2025年90%分割保形预测区间

Fig. 8 90% contour prediction interval for 2025

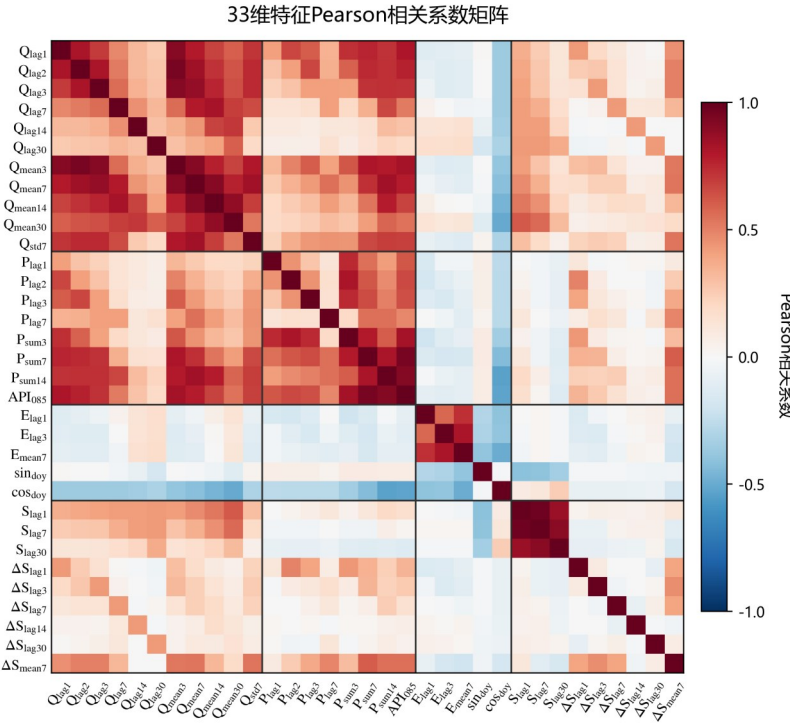


图9 33维特征 Pearson 相关系数矩阵

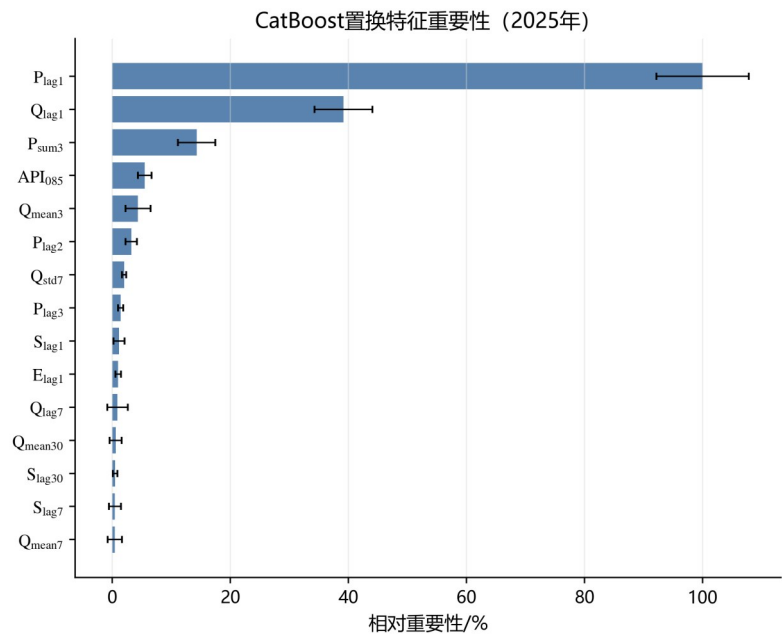
Fig. 9 33-dimensional feature pearson correlation coefficient matrix plot

4.5 同类研究对比与工程应用讨论

王秀杰等^[3]的TVF-EMD-GRA-LightGBM模型在利津站和高要站日径流预测中NSE为0.949和0.966;黄一凡等^[5]的BO-LSTM模型在汉江仙桃站检验期平均NSE为0.84;黄泽禧等^[8]的HEC-HMS-LSTM模型在高道站12 h预见期训练集NSE为

0.91。三者分别依赖模态分解、降水输入或物理模型输出,任务边界不同(表8)。

本研究仅使用业务可获得的日尺度降水、流量、蒸发和聚合蓄水数据,通过多尺度特征与CatBoost在2025年独立检验中取得NSE=0.918,相对持续性模型的误差降幅为52.6%,适用于多水库



注:图中仅列置换重要性前15项;其余长滞后蓄水和蒸发特征的单项贡献较低,保留依据见4.2节消融结果。

图 10 CatBoost 置换特征重要性
Fig. 10 CatBoost substitution feature importance

表 8 同类径流预测研究指标对比
Tab. 8 Comparison with related runoff forecasting studies

研究	对象与尺度	方法	报告NSE	核心差异与定位
王秀杰等 ^[3]	黄河利津站、珠江高要站日径流	TVF-EMD-GRA-LightGBM	0.949/0.966	依赖经验模态分解预处理,基于单站日径流序列建模
黄一凡等 ^[5]	汉江仙桃站日径流	BO-LSTM	0.84(检验期均值)	输入仅降水量,未构建多尺度特征体系
黄泽禧等 ^[8]	连江高道站3~12 h流量预测	HEC-HMS-LSTM耦合	0.91(12 h,训练集)	3h步长+物理模型耦合,面向洪水短时预报
本研究	横山(二)站次日日均流量	多尺度特征-CatBoost	0.918	仅用业务可获得的日尺度表格数据,构建33维多尺度特征

注:各研究流域特性、时间尺度、预见期及输入资料存在本质差异,绝对指标不可直接用于算法优劣排名。本研究的NSE基于2025年独立检验集实测所得;王秀杰等的NSE源于文献原文报告值。

调控流域“ $t-1$ 日”滚动预报。

当然,受限于日尺度面平均降雨输入,本研究对极端洪峰和长预见期(>1 日)的表达能力弱于拥有高分辨率过程输入或物理模型耦合的深度学习方法,这也是未来值得探索的方向。工程应用可采用“每日资料接入—自动预测与区间输出—阈值触发—人工复核发布”流程。点预测服务日尺度调配和趋势研判;汛期同步显示较宽区间,超过警戒阈值时转入高频监测或专业洪水预报。若前一日实测流量延迟或异常,应切换备用模型。日均值会平滑小时洪峰,故本模型不能替代现行洪水预报和调度会商。

研究资料仅覆盖5 a,5折验证不能替代长系列多周期检验;面平均降雨无法表达暴雨中心移动,聚合蓄水量无法区分单库库容、入出库与调度规则,预测区间也尚缺严格的非平稳覆盖保证。后续应滚动纳入新年份,并引入雷达降雨、上游流量、土壤含水量和逐库泄量开展跨年检验。

5 结论与展望

本研究基于梅江横山(二)站 2021—2025 年数据,构建了融合多尺度水文特征与 CatBoost 的日均流量滚动预报模型,并通过严格的时序验证与不确定性分析,得出如下主要结论。

a)方法优势与适用性明确。本研究提出的方案高度适配南方湿润区多水库调控的短序列场景。CatBoost 凭借有序梯度提升机制,在仅依赖日尺度常规数据的业务约束下,取得了 $NSE=0.918$ 的精度,较基准模型提升显著。特征消融表明,降雨湿润特征是精度跃升的核心(NSE 提升 0.447),而水库状态特征是枯水期偏差校正的关键($PBIAS$ 优化至 -0.04%),这为业务化滚动调度提供了可靠的技术手段。

b)不确定性边界明确。汛期强暴雨随机性、产汇流非线性及水库动态调度共同放大误差,枯水期则以平稳基流为主,故汛期预测区间宽度约为枯水期的 2.7 倍。极端洪峰误差主要归因于小时雨强、降雨空间分布及逐库泄量等信息缺失;极高流量 NSE 仅 0.287,最大低估 40.6%。因此,该方法适合日常调度与一般性洪水预警,极端洪水仍需结合高分辨率模型研判。

c)验证范式与参数策略可靠。研究证实,针对短序列数据,扩展窗口时序验证结合独立年份检验是评估模型泛化能力的必要手段。虽然贝叶斯优化提升了训练稳定性,但固定参数方案在未知水文年型下表现出更强的鲁棒性,推荐在业务化应用中优先采用固定参数方案以降低运维复杂度。

业务落地部署上可每日定时接入日尺度降水、流量、蒸发和蓄水数据,由固定参数 CatBoost 生成次日流量及预测区间;结合汛枯期、生态流量和警戒流量阈值触发预警,经人工复核后发布。后续可进一步接入欧洲中期天气预报中心(ECMWF)数值降水预报产品,沿用固定参数模型开展 7 天预见期滚动迭代计算,估算未来一周流量变化趋势,为周尺度调度研判提供趋势参考。

在此基础上,未来将引入雷达降雨、土壤含水

量和逐库调度信息,并进一步拓展概率预报与风险导向的学习机制;设置 $\tau = 0.1, 0.5, 0.9$ 的分位数 CatBoost,分别刻画枯水、常规和洪峰风险;设计面向生态流量与洪峰预警的非对称加权损失。两类方案均沿用扩展窗口时序验证和独立年份检验,其业务增益将通过 AB 测试评估。

参考文献:

- [1] 陆旻蛟. 新安江模型研究的回顾和展望[J]. 水利学报, 2021, 52(4): 432-441.
Lu Minjiao. Recent and future studies of the Xin'anjiang model [J]. Journal of Hydraulic Engineering, 2021, 52(4): 432-441. (in Chinese)
- [2] 张力, 王红瑞, 郭菲楠, 等. 基于时序分解与机器学习的非平稳径流序列集成模型与应用[J]. 水科学进展, 2023, 34(1): 42-52.
Zhang Li, Wang Hongrui, Guo Beinan, et al. Integrated model and application of non-stationary runoff based on time series decomposition and machine learning [J]. Advances in Water Science, 2023, 34(1): 42-52. (in Chinese)
- [3] 王秀杰, 乔鸿飞, 曾勇红, 等. 基于 TVF-EMD、GRA 和 LightGBM 的日径流预测组合模型[J]. 水资源保护, 2023, 39(5): 135-142, 151.
Wang Xiujie, Qiao Hongfei, Zeng Yonghong, et al. TVF-EMD, GRA, and LightGBM combined daily runoff prediction model [J]. Water Resources Protection, 2023, 39(5): 135-142, 151. (in Chinese)
- [4] 谭永杰, 王现勋, 段茗续, 等. 不同模态分解方法结合 LSTM 模型对日径流预报的影响[J]. 人民珠江, 2023, 44(12): 64-72, 79.
Tan Yongjie, Wang Xianxun, Duan Mingxu, et al. Impact of different modal decomposition methods combined with LSTM model on daily runoff forecasting [J]. Pearl River, 2023, 44(12): 64-72, 79. (in Chinese)
- [5] 黄一凡, 张翔, 邓梁堃, 等. 基于机器学习的汉江流域径流模拟与时滞变化分析[J]. 水资源保护, 2024, 40(6): 173-180.
Huang Yifan, Zhang Xiang, Deng Liangkun, et al. Runoff simulation and time-lag change analysis in the Han River Basin based on machine learning [J]. Water Resources Protection, 2024, 40(6): 173-180. (in Chinese)
- [6] 崔杰连, 常亮, 赵敏, 等. 耦合深度学习的水丰水库入库径流中长期预测方法研究[J]. 中国农村水利水电, 2024, (8): 73-80.

- Cui Jieliang, Chang Liang, Zhao Min, et al. Study on medium and long term forecasting methods for inflow runoff of Shuifeng reservoir coupled with deep learning[J]. China Rural Water and Hydropower, 2024(8): 73–80. (in Chinese)
- [7] 张玮, 郑雅莲, 刘志武, 等. 物理机制引导的水库调度深度学习模型研究[J]. 水力发电学报, 2023, 42(3): 13–25.
- Zhang Wei, Zheng Yalian, Liu Zhiwu, et al. Deep learning model guided by physical mechanism for reservoir operation[J]. Journal of Hydroelectric Engineering, 2023, 42(3): 13–25. (in Chinese)
- [8] 黄泽禧, 孙伟, 陈新林, 等. 耦合物理模型与深度学习的流量预测方法[J]. 人民珠江, 2025, 46(5): 51–62.
- Huang Zexi, Sun Wei, Chen Xinlin, et al. Flow prediction method combining physical model and deep learning: a case study of Gaodao Station along Lianjiang River[J]. Pearl River, 2025, 46(5): 51–62. (in Chinese)
- [9] 王毅, 胡凌云, 陈浩, 等. 基于 CatBoost 的水库入库径流预测及可解释分析[J]. 中国农村水利水电, 2026, (3): 50–56.
- Wang Yi, Hu Lingyun, Chen Hao, et al. Reservoir inflow runoff prediction and interpretability analysis based on CatBoost[J]. China Rural Water and Hydropower, 2026 (3): 50–56. (in Chinese)
- [10] Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features [C]// Advances in Neural Information Processing Systems 31. Red Hook, NY: Curran Associates, Inc., 2018: 6638–6648.
- [11] Ke Guolin, Meng Qi, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [C]//GUYON I, VON LUXBURG U, BENGIO S, et al. Advances in Neural Information Processing Systems 30. Red Hook, NY: Curran Associates, Inc., 2017: 3146–3154.
- [12] 徐冬梅, 王逸阳, 王文川. 基于贝叶斯优化算法的长短期记忆神经网络模型年径流预测[J]. 水电能源科学, 2022, 40(12): 42–46.
- Xu Dongmei, Wang Yiyang, Wang Wenchuan. Prediction of annual runoff using long short-term memory neural network model based on Bayesian optimization algorithm[J]. Water Resources and Power, 2022, 40(12): 42–46. (in Chinese)
- [13] Shahriari B, Swersky K, Wang Ziyu, et al. Taking the human out of the loop: a review of Bayesian optimization[J]. Proceedings of the IEEE, 2016, 104(1): 148–175.
- [14] 郑言峰. 水文模型不确定性研究方法耦合分析[J]. 人民珠江, 2022, 43(6): 108–115.
- Zheng Yanfeng. Coupling analysis of research methods for hydrological model uncertainty [J]. Pearl River, 2022, 43(6): 108–115. (in Chinese)
- [15] Angelopoulos A N, Bates S. Conformal prediction: a gentle introduction[J]. Foundations and Trends in Machine Learning, 2023, 16(4): 494–591.

Daily Discharge Forecasting Based on Multiscale Hydrological Features and CatBoost

ZHANG Dawen, GAO Yudan

(Meizhou Hydrology Branch, Guangdong Hydrology Bureau, Meizhou 514000, China)

Abstract: Accurate next-day mean daily discharge forecasts are fundamental to reservoir operation, flood preparedness, water resources allocation, and ecological management, particularly in regulated basins within monsoon climate regions. However, operational forecasting in such environments faces persistent challenges, including limited historical records, incomplete characterization of antecedent hydrological states (e. g., soil moisture and baseflow), and the risk of data leakage inherent in conventional random cross-validation schemes. To address these limitations, this study proposes a robust forecasting framework based on the CatBoost gradient boosting algorithm for the Hengshan (II) Hydrological Station in the Meijiang River Basin. Using daily hydrometeorological and reservoir data from 2021 to 2025, we constructed a 33-dimensional feature set with clear physical interpretability, organized into four groups: discharge memory, antecedent rainfall and catchment wetness (including a calibrated Antecedent Precipitation Index), evaporative demand and seasonality, and aggregated reservoir storage dynamics. A rigorous causality constraint was enforced by shifting all predictors so that only information available by day $t-1$ was used to forecast discharge on day t . Following a 30-day initialization period, 1 795 valid daily samples were retained. Model development employed an expanding-window five-fold time-series cross-validation approach on data from 2021 – 2024, coupled with Bayesian hyperparameter optimization, while all 365 observations from 2025 were strictly reserved for independent testing. Comparative analysis against Persistence, Ridge

Regression, ExtraTrees, and LightGBM benchmarks demonstrated the superiority of the fixed-parameter CatBoost model. On the independent test set, CatBoost achieved a Nash – Sutcliffe Efficiency (NSE) of 0.918, a Kling – Gupta Efficiency (KGE) of 0.956, a Root Mean Square Error (RMSE) of 69.5 m³/s, and a Mean Absolute Error (MAE) of 38.7 m³/s. Notably, its NSE exceeded those of LightGBM and ExtraTrees by 8.0% and 3.0%, respectively (corresponding to absolute NSE gains of 0.068 and 0.027), and its RMSE was 52.6% lower than that of the persistence benchmark. Feature ablation experiments elucidated the distinct roles of the feature groups: multiscale discharge-memory features established the predictive baseline, whereas rainfall and wetness variables contributed a significant NSE gain of 0.447. Crucially, the inclusion of reservoir-state features did not markedly alter annual NSE but substantially reduced MAE and Percentage Bias (PBIAS) during the dry season (PBIAS improved to -0.04%), highlighting their role in correcting low-flow biases. Bayesian optimization improved the mean cross-validation NSE from 0.772 ± 0.114 to 0.797 ± 0.083 and reduced fold-to-fold variability by 27.2%, yet it lowered the independent test NSE to 0.904, underscoring that optimization targeting development-period performance does not guarantee enhanced interannual generalization. Further uncertainty quantification via split-conformal prediction and out-of-fold residual analysis revealed pronounced seasonal heteroscedasticity; the mean 90% prediction interval width during the flood season was approximately 2.7 times that of the dry season. Although the model effectively captured the timing of most hydrograph limbs, its performance deteriorated for the 19 extremely high-flow days (peak NSE = 0.287), primarily because daily basin-average rainfall and aggregated storage metrics cannot represent sub-daily rainfall intensity, storm spatial heterogeneity, or individual reservoir release patterns. In conclusion, the proposed causality-constrained framework offers a reproducible and operationally feasible solution for next-day rolling forecasts in data-scarce, multi-reservoir basins. Nevertheless, for extreme flood events, the integration of higher-resolution radar rainfall data, soil moisture observations, and individual reservoir operation schedules remains essential for reliable risk assessment.

Keywords: daily average discharge forecasting; CatBoost; multiscale features; Bayesian optimization; time-series cross-validation; Meijiang River

(责任编辑: 向 飞)