

DOI:10.3876/j.issn.1000-1980.2026.05.010

# 融合机器学习与贝叶斯模型平均的水电站来水预测方法

陈 凯<sup>1</sup>, 罗煜宁<sup>2,3,4,5</sup>, 张 珂<sup>2,3,6,7,8</sup>, 吴 南<sup>2,3</sup>, 吴星宇<sup>2,3</sup>, 王宇昊<sup>2,3</sup>,  
王有香<sup>1</sup>, 彭 晶<sup>1</sup>

(1. 云南电网有限责任公司电力科学研究院; 2. 河海大学水灾害防御全国重点实验室; 3. 河海大学水文水资源学院;  
4. 水利部水文仪器及岩土工程仪器质量监督检验测试中心; 5. 水利部南京水利水文自动化研究所;  
6. 河海大学长江保护与绿色发展研究院; 7. 中国气象局水文气象重点开放实验室; 8. 水利部水利大数据重点实验室)

**摘要:** 针对水文模型在复杂环境条件下易出现建模不充分或未建模行为导致的模拟偏差问题, 提出了一种融合机器学习与贝叶斯模型平均(BMA)的水电站来水预测方法。该融合方法以 Hydromad 模型为基础计算来水及其残差序列, 利用随机森林(RF)模型、XGBoost 模型和支持向量回归(SVR)模型捕捉残差的非线性特征, 并通过 BMA 融合多模型预测结果, 实现模拟误差后处理修正与不确定性评估。金沙江梨园水电站实例分析结果表明, 提出的方法可以显著提升 Hydromad 模型在洪峰响应和涨、退水阶段的拟合精度, 修正后来水预测的 NSE、KGE 等指标显著提升, 水量偏差得到有效控制, 整体预测精度明显改善, 可为水电站来水预测及调度决策提供更科学的技术支撑。

**关键词:** 来水预测; 贝叶斯模型平均; 机器学习; Hydromad 模型; 不确定性评估; 金沙江梨园水电站  
**中图分类号:** P338<sup>+</sup>.2 **文献标志码:** A **文章编号:** 1000-1980(2026)05-0091-08

## Inflow forecasting method for hydropower stations by integrating machine learning and Bayesian model averaging

Chen Kai<sup>1</sup>, Luo Yuning<sup>2,3,4,5</sup>, Zhang Ke<sup>2,3,6,7,8</sup>, Wu Nan<sup>2,3</sup>, Wu Xingyu<sup>2,3</sup>, Wang Yuhao<sup>2,3</sup>, Wang Youxiang<sup>1</sup>, Peng Jing<sup>1</sup>

(1. Electric Power Research Institute, Yunnan Power Grid Co., Ltd.;

2. State Key Laboratory of Water Disaster Prevention, Hohai University;

3. College of Hydrology and Water Resources, Hohai University;

4. MWR Inspecting and Testing Centre for Hydrological & Geotechnical Instruments' Quality;

5. Nanjing Research Institute of Hydrology and Water Conservation Automation, Ministry of Water Resources;

6. Yangtze Institute for Conservation and Development, Hohai University;

7. Key Laboratory of Hydro-Meteorology China Meteorological Administration;

8. Key Laboratory of Water Big Data Technology of Ministry of Water Resources)

**Abstract:** To address the issue of simulation bias in hydrological models caused by insufficient modeling or unmodeled behaviors under complex environmental conditions, an inflow forecasting method for hydropower stations based on machine learning and Bayesian model averaging (BMA) was proposed. Based on the Hydromad model, the inflow and residual sequences were calculated; three machine learning models, namely random forest (RF) model, XGBoost model, and support vector regression (SVR) model, were used to capture the nonlinear characteristics of the residuals. Subsequently, the multi-model prediction results were fused via BMA to achieve post-processing error correction and uncertainty assessment of the simulation errors. The case study results of Liyuan Hydropower Station on Jinsha River indicate that the proposed method can significantly improve the fitting accuracy of the Hydromad model in peak flood responses and the rising and falling water stages. After correction, indicators of inflow prediction such as NSE and KGE are significantly improved; the water volume bias is effectively controlled; the overall prediction accuracy is obviously improved. This method can provide more scientific technical support for inflow forecasting and operational decision-making of hydropower stations.

**Key words:** inflow forecasting; Bayesian model averaging; machine learning; Hydromad model; uncertainty assessment; Liyuan Hydropower Station on Jinsha River

**基金项目:** 中国南方电网云南电网有限责任公司科技项目(YNKJXM20222329); 国家重点研发计划项目(2023YFC3006500)

**作者简介:** 陈凯(1990—), 男, 工程师, 博士, 主要从事电力电量平衡技术和水电优化调度研究。E-mail: chen kai09@ yn. csg. cn

**通信作者:** 罗煜宁(1999—), 女, 工程师, 博士, 主要从事水文预报研究。E-mail: yuning\_lynn@ 163. com

随着经济社会的发展,水资源利用效率不断提升,对水电站入库流量模拟和预报精度的要求也日益提高<sup>[1]</sup>。水文模型在流域径流预测中扮演着关键角色,而准确的径流预报后处理对优化水电调度至关重要<sup>[2-3]</sup>,尤其在青藏高原等水能资源丰富的大江大河源区,水文模拟的偏差直接影响调度方案制定与水能利用效率<sup>[4-5]</sup>。然而,由于降水过程具有高度复杂性和不确定性<sup>[6-8]</sup>及存在建模不充分或未建模的问题,在汛期或下垫面条件复杂的流域中<sup>[9]</sup>,水文模型常常低估极端高流量事件的发生频率与强度<sup>[10-12]</sup>,模拟结果存在与实测值之间的系统性残差,尤其在复杂的降雨-融雪径流过程中更为明显,水电站来水预测工作仍面临挑战。

近年来学者们开始关注基于残差的建模(即对水文模型模拟误差进行建模)方法<sup>[13-14]</sup>。残差是水文模型模拟结果与实测数据之间的差异,蕴含了模型未能捕捉的过程信息。通过对残差进行统计建模和预测,可以在保留原有物理模型的基础上校正其模拟结果,从而提升预测精度。然而,传统回归方法往往难以刻画残差的非线性和时序性特征,因此亟须引入数据驱动方法对残差进行建模与修正,以提升预测的稳定性与可信度。例如:Li 等<sup>[15]</sup>将长短期记忆神经网络(LSTM)用于分布式水文模型的残差建模,成功削弱了残差的异方差性并消除了几乎所有自相关结构;蒋晓蕾等<sup>[16]</sup>采用考虑误差异分布的 HRD 方法估计误差的概率分布,对赣江流域开展了径流概率预报。尽管上述机器学习方法和统计方法在提升水文模型精度方面取得了一定的成效<sup>[17]</sup>,但仍存在一定的局限性:一方面,残差误差的统计特征可能因时间、空间或来水条件而变化<sup>[18]</sup>,传统方法难以动态适应误差结构的变化;另一方面,不同方法的残差建模结果通常缺乏系统性的融合机制<sup>[19]</sup>,容易造成模型偏信或协同不足,从而影响径流预报精度。对此贝叶斯方法为残差建模提供了一种更为灵活且理论完备的处理框架,通过先验信息与观测数据,不断更新模型认知,从而获得更可信的预测结果。在多模型融合策略中,贝叶斯模型平均(Bayesian model averaging, BMA)<sup>[20-23]</sup>作为一种经典的贝叶斯集成方法,能够在兼顾多个模型模拟能力的同时,为每个模型依据后验概率分配权重,实现信息的最优融合。近年来,该方法已被广泛应用于气象、水文等领域<sup>[24]</sup>的预报修正和不确定性量化。例如:Zhao 等<sup>[25]</sup>构建了基于完全集合经验模态分解和 BMA 方法的降水集合预测模型(CEEMDAN-BMA),并在北京和广州站的月降水预估中取得了良好效果;周婷等<sup>[26]</sup>利用 BMA 对 3 种机器学习模型的中长期日径流预测结果进行融合,并开展预测不确定性评估,结果显示,经 BMA 融合后的模型较单一模型具备更高的预测精度。然而,当前 BMA 应用多集中于对流量或降水结果的直接后处理,面向残差修正的融合策略尚未得到深入研究。因此,开展基于 BMA 方法的水文模型残差建模与融合修正研究,尤其是通过集成多种机器学习模型捕捉残差的非线性特征<sup>[27]</sup>,并在此基础上进行来水预测的不确定性评估,可提升 BMA 在修正模拟误差和提高来水预测精度方面的应用能力。

本文基于 Hydromad 模型的模拟结果构建残差序列,利用多种机器学习模型对残差序列进行预测,并通过 BMA 方法对预测结果进行融合修正,从而提升来水预测精度并实现不确定性评估。

1 基于机器学习的水文模拟残差订正与 BMA 融合方法

本文基于自动优化率定的 Hydromad 模型生成的模拟径流结果,提取其与实测径流之间的残差序列,构建融合滞后特征的机器学习残差订正模型,以当前降水量、气温、模型模拟流量与历史残差信息作为输入特征,分别采用随机森林(RF)、XGBoost 和支持向量回归(SVR)3 种回归模型对残差进行预测。所有模型均在训练集上进行训练,并在测试集上评估其预测性能。随后,采用 BMA 方法融合 3 种模型预测的结果,以各模型在测试集上的性能作为先验概率,并结合模型残差的似然函数计算后验权重。在此基础上,对验证集数据的预测采用加权融合策略,生成残差预测值,与原模型模拟值叠加从而获得修正后的来水预测值,并通过模型融合过程中样本方差的估算获得不确定性区间,给出残差预测的上下界,从而实现了对水电站来水过程不确定性的定量评估。

1.1 基于 Hydromad 模型的径流预测

IHACRES 模型是一个集总式径流模型,通过将实测降雨转化为有效降雨,再由线性水库模块生成径流,适用于资料有限流域的径流模拟与预测。该模型对数据需求低,输入参数少且模拟精度高,可反映不同气候条件下的流域响应,在不同干湿气候区得到广泛应用<sup>[28-30]</sup>;缺点是物理机制简单,对强人类活动干扰的流域预测精度有限。本文采用由 Andrews 等<sup>[31]</sup>开发的 Hydromad 模型进行径流预测,该集成模型包含多个模型,

如 IHACRES 和度日因子模型<sup>[32]</sup>,并支持多模型的集成和参数自动优化率定。其中,度日因子模型基于气温与融雪量的线性关系,假设气温高于阈值时,积雪按一定“度日因子”融化,融水进入产流过程,适用于以积雪融水为主要补给的高寒山区,尤其是在缺乏详细能量平衡数据时。Hydromad 模型的主要模块与参数物理意义见表 1。

1.2 基于滞后特征的机器学习模型

为对 Hydromad 模型产生的残差进行修正,本文引入 3 种机器学习模型对模拟残差进行预测。

a. RF 模型<sup>[33]</sup>。RF 是一种基于决策树集成的非参数回归模型,通过构建多个决策树并对其结果进行平均以提升预测性能和稳定性。它在处理非线性关系、多变量输入及高噪声数据方面具有良好的鲁棒性,同时能有效缓解过拟合问题。

b. XGBoost 模型。XGBoost 是一种高效的梯度提升树模型,采用加权迭代的方式逐步拟合残差项,从而提升整体预测精度。它在建模过程中引入正则化项,具有良好的泛化能力,适用于对水文模型中的系统误差进行建模。

c. SVR 模型<sup>[34]</sup>。SVR 通过引入核函数将输入特征映射到高维空间,并在该空间中构造最优回归超平面。SVR 能够有效捕捉非线性关系,且对异常值不敏感,适用于对具有一定波动性的流量残差序列进行建模。

为评估各模型残差预测精度,本文采用平均绝对误差(MAE)、均方误差(MSE)和决定系数( $R^2$ )作为误差评价指标。

1.3 基于 BMA 的水电站来水预测方法

为综合上述 3 种机器学习模型对残差的预测能力,本文引入 BMA 方法对残差预测结果进行融合。该方法在保留各模型信息贡献的基础上,依据其预测能力进行加权,从而提升总体预测的稳健性与准确性。设第  $t$  天第  $m$  个残差模型预测的残差为  $\hat{\varepsilon}_{mt}$ ,对应的后验概率权重为  $w_m$ ,融合残差预测值  $\hat{\varepsilon}_t$  表示为:

$$\hat{\varepsilon}_t = \sum_{m=1}^3 w_m \hat{\varepsilon}_{mt}$$

(1)

模型的后验概率权重  $w_m$  由其先验概率  $P_m$  与似然函数  $L_m$  共同确定:

$$w_m = \frac{P_m L_m}{\sum_{j=1}^3 P_j L_j}$$

(2)

式中: $P_j$ 、 $L_j$  分别为求和项中第  $j$  个残差模型的先验概率和似然函数。先验概率  $P_m$  基于各模型在测试集上的决定系数  $R^2$  归一化得到:

$$P_m = \frac{R_m^2}{\sum_{j=1}^3 R_j^2}$$

(3)

式中: $R_m^2$  为第  $m$  个残差模型在测试集上的决定系数; $R_j^2$  为求和项中第  $j$  个残差模型在测试集上的决定系数。模型似然函数  $L_m$  假设残差服从正态分布  $N(\mu_m, \sigma_m^2)$ ,可由下式表示:

$$L_m = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left\{-\frac{(\varepsilon_t - \hat{\varepsilon}_{mt})^2}{2\sigma_m^2}\right\}$$

(4)

式中: $\varepsilon_t$  为实际残差; $\hat{\mu}_m$ 、 $\sigma_m^2$  分别为机器学习模型残差的均值与方差; $n$  为测试集中残差观测的总天数。

将融合残差与 Hydromad 模型模拟值叠加,可得修正后的预测流量  $\hat{Q}_{\text{corr},t}$ :

$$\hat{Q}_{\text{corr},t} = Q_{\text{sim},t} + \hat{\varepsilon}_t$$

(5)

式中: $\hat{Q}_{\text{corr},t}$  为修正后的预测流量,  $\text{m}^3/\text{s}$ ;  $Q_{\text{sim},t}$  为修正前的模拟流量,  $\text{m}^3/\text{s}$ 。同时,为评估预测结果的不确定

表 1 Hydromad 模型主要模块及参数

Table 1 Main modules and parameters of Hydromad model

| 模块    | 参数                | 单位       | 物理意义        |
|-------|-------------------|----------|-------------|
| 非线性产流 | $\tau_w$          | d        | 参考温度下流域变干速率 |
|       | $f$               | ℃        | 变干速率的温度调节因子 |
|       | $C$               | mm       | 流域水量平衡系数    |
| 线性汇流  | $\alpha_q$        |          | 快速径流消退系数    |
|       | $\alpha_s$        |          | 慢速径流消退系数    |
|       | $\beta_q$         |          | 快速径流洪峰响应系数  |
| 融雪    | $T_{\text{melt}}$ | ℃        | 融化阈值        |
|       | $c_s$             |          | 融雪修正系数      |
|       | $k_d$             | mm · ℃/d | 融化度日因子      |
|       | $k_f$             | mm · ℃/d | 再冻结度日因子     |
|       | $r$               |          | 液态水消退系数     |

性,引入残差方差融合估计,综合 3 个模型预测残差的方差可得水电站径流预测不确定性范围:

$$\sigma^2 = \sum_{m=1}^3 w_m^2 \sigma_m^2$$

(6)

$$\begin{cases} Q_{cl,t} = \hat{Q}_{corr,t} - 1.96\sigma \\ Q_{cu,t} = \hat{Q}_{corr,t} + 1.96\sigma \end{cases}$$

(7)

式中: $\sigma^2$  为 BMA 融合预测残差的总体方差; $\sigma$  为 BMA 的标准差; $Q_{cl,t}$ 、 $Q_{cu,t}$  分别为第  $t$  天校正径流预测值 95% 置信区间的下界和上界。

## 2 实例分析

### 2.1 研究区概况

金沙江梨园水电站以上流域(26°38'N~35°53'N, 90°34'E~100°24'E)地处青藏高原中东部,流域面积 21.42 万 km<sup>2</sup>,占金沙江流域面积的 42.7%,是该流域主要的产流区,水能资源丰富。流域地势呈北高南低,平均海拔 4500 m 以上,气候类型为高原大陆性气候,境内涵盖半湿润和半干旱区。水系发育良好,多年平均降水量为 452.6 mm,多年平均气温为 1.5℃。由于地理位置和气候条件的独特性,流域内降水形式具有显著的季节性规律:非汛期以降雪为主,夏季降水集中且丰沛。然而,由于横断山脉的阻隔作用,研究区属于少雨区,尤其是玉树巴塘河口至石鼓区间的金沙江纵向河谷地带,降水量的空间分布较为不均匀,这增加了水电站径流模拟与预测的难度。流域概况如图 1 所示。

### 2.2 数据来源及预处理

本文采用的输入数据包括日尺度降水、气温和实测入库流量,用于驱动 Hydromad 模型并构建残差建模数据集。降水 and 气温数据来源于国家气候中心研发的气候系统模式 CMA\_CPSv3 (<https://data.cma.cn/>),经国家气候中心降尺度处理后,时间分辨率为 1 d,空间分辨率为 0.05°×0.05°;日径流数据来源于梨园水电站实测入库流量。采用双线性插值法计算流域面平均降水量和气温。

选取 2010—2022 年的日尺度观测数据进行建模与验证,采用纳什效率系数(NSE)和克林-古普塔效率系数(KGE)评价模型模拟性能,其中 NSE 反映模拟值与实测值的整体拟合程度,KGE 综合衡量相关性、偏差和变异性。模型 NSE 为 0.91,KGE 为 0.89,模型参数率定结果如下: $\tau_w=25\text{ d}$ , $f=1.34^\circ\text{C}$ , $C=0.55\text{ mm}$ , $\alpha_q=0.96$ , $\alpha_s=0.99$ , $\beta_q=0.545$ , $T_{\text{melt}}=0^\circ\text{C}$ , $c_s=0.8$ , $k_d=5.5\text{ mm}\cdot^\circ\text{C/d}$ , $k_f=0.361\text{ mm}\cdot^\circ\text{C/d}$ , $r=0$ 。将 2011—2020 年作为 3 种机器学习模型的训练与测试期(按照 7:3 比例划分),用于训练和测试 RF、XGBoost 和 SVR 3 种模型;2021—2022 年作为验证期,采用 MBA 方法对预测残差进行加权融合,进一步完成对模拟误差的校正,并评估修正后来水预测结果的不确定性。

### 2.3 机器学习模型残差预测结果

为评估不同机器学习模型对 Hydromad 模型残差的预测能力,本文在训练完成后选取测试期数据进行精度验证与对比分析,3 种机器学习模型的残差预测性能如表 2 所示。由表 2 可知,XGBoost 模型表现最优, $R^2$  为 0.79,RF 模型次之,且性能表现与 XGBoost 相近, $R^2$  为 0.78,而支持向量机模型的性能最差, $R^2$  仅为 0.53。

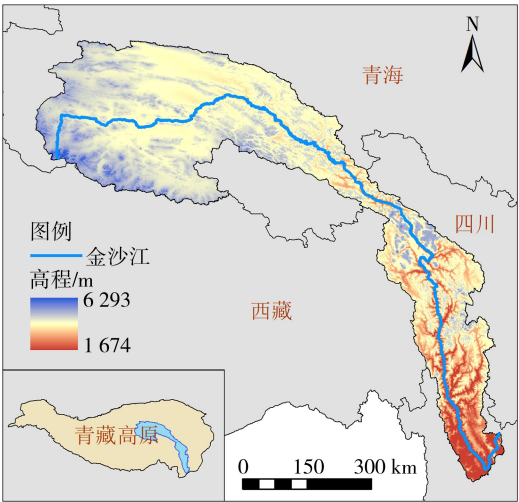


图 1 研究区地理位置及高程分布  
Fig. 1 Geographical location and elevation distribution of study area

表 2 3 种机器学习模型测试期残差预测性能

| Table 2 Residual prediction performance of three machine learning models during testing period |                         |                                      |                |
|------------------------------------------------------------------------------------------------|-------------------------|--------------------------------------|----------------|
| 模型                                                                                             | MAE/(m <sup>3</sup> /s) | MSE/(m <sup>3</sup> /s) <sup>2</sup> | R <sup>2</sup> |
| RF                                                                                             | 289.69                  | 186 784.47                           | 0.78           |
| XGBoost                                                                                        | 283.34                  | 177 742.00                           | 0.79           |
| SVR                                                                                            | 357.94                  | 232 110.40                           | 0.53           |

图 2 为 3 种机器学习模型在测试期对 Hydromad 模型模拟残差的预测结果与实际残差对比,可以看出,RF 模型和 XGBoost 模型整体上能够较好地拟合残差的时序变化趋势,有效捕捉残差的波动,预测曲线与真实值在大多数低值区以及大部分高值区较为贴合。而 SVR 模型的预测存在明显高频震荡现象,且在多个峰值和突变点附近拟合误差较大。此外,相较于 RF 模型,XGBoost 模型在部分极值时段的拟合略好,表现出更强的泛化能力,例如在第 319 步时,XGBoost 模型的模拟值远高于 RF 模型和 SVR 模型,与实际残差更为贴近;而 SVR 模型对非线性残差响应不足,难以准确捕捉复杂结构误差。综合残差时序趋势与表 2 中指标值可知,XGBoost 模型在残差预测中综合性能最佳,其次为 RF 模型,SVR 模型拟合能力相对有限。

2.4 BMA 方法残差融合结果及水电站来水不确定性分析

为提升残差预测的稳健性并对预测结果的不确定性进行量化,以 3 种机器学习模型在测试期中的性能表现( $R^2$ )作为先验概率,根据式(3)计算得到 RF 模型、XGBoost 模型和 SVR 模型的先验概率分别为 0.373、0.376、0.251。为估计模型的似然函数  $L_m$ ,本文在高斯误差假设下,使用 2021—2022 年验证期中各机器学习模型的残差序列,分别拟合其误差分布。假设各机器学习模型预测误差服从独立同分布的正态分布,则似然函数可表示为预测误差在该分布下的联合概率密度函数积。图 3 为 3 种机器学习模型残差误差分布直方图及正态分布拟合结果。

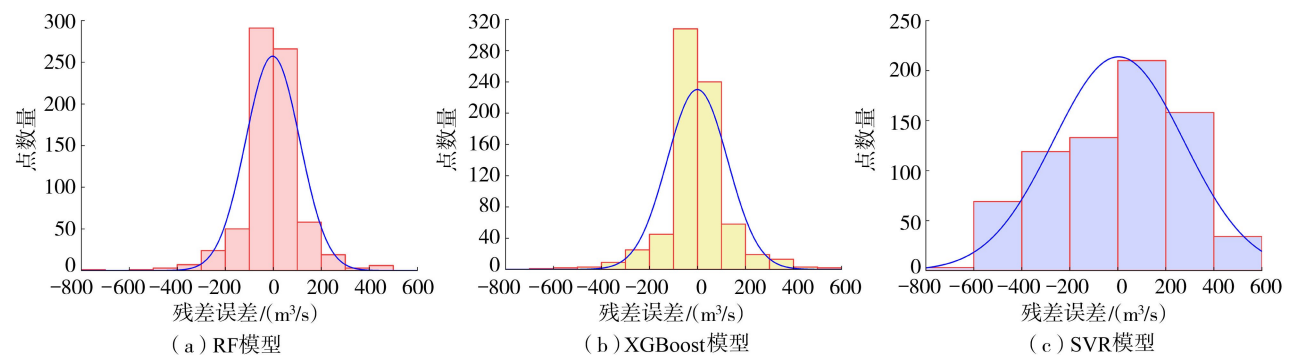


图 3 不同机器学习模型残差误差分布直方图

Fig.3 Histogram of residual error distribution of different machine learning models

从图 3 可以看出,RF 模型与 XGBoost 模型的残差误差分布基本呈对称性,拟合曲线形态接近正态分布,均值分别为 $-2.25\text{ m}^3/\text{s}$ 和 $-0.21\text{ m}^3/\text{s}$ ,标准差分别为 $113.05\text{ m}^3/\text{s}$ 和 $126.30\text{ m}^3/\text{s}$ ,二者在残差预测中波动较小,误差集中度较高。而 SVR 模型的残差误差分布较为离散,拟合正态分布曲线宽而扁平,标准差达 $272.08\text{ m}^3/\text{s}$ ,远高于其他两个模型,且均值偏正,表明其预测结果在多个时间步存在系统性高估。

在验证期(2021—2022 年)中,为评估这 3 种模型在水电站来水预测中的表现,进一步采用 BMA 方法对其残差预测结果进行加权融合。由于 BMA 方法综合考虑模型的先验能力与残差误差的置信程度,根据式(2)计算得到 RF 模型、XGBoost 模型和 SVR 模型的贝叶斯后验概率权重分别为 0.393、0.401、0.206。图 4 为残差预测结果及其 95%置信区间,可以看出,BMA 融合残差在大多数时间步上能够较好地模拟真实残差的变化趋势,尤其在残差波动剧烈的时段(如第 250~300 步和第 600~650 步),模型仍能保持稳定拟合能力,表明 BMA 融合残差具有较高可信度。

图 5 为 2021—2022 年验证期内梨园水电站来水模拟值与实测值的对比,可以看出,Hydromad 模型模拟结果在大部分时间段能够捕捉来水过程的总体趋势,但在多个高流量阶段,特别是丰水期,存在明显滞后或偏低现象,反映出一定的系统性偏差。经 BMA 融合残差修正后,流量过程线在绝大多数时间段内更接近实

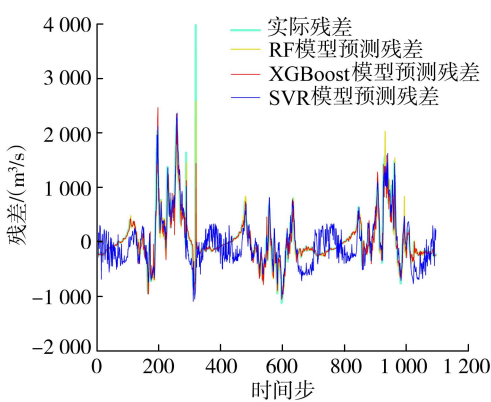


图 2 不同机器学习模型预测残差与实际残差对比

Fig.2 Comparison between predicted residuals of different machine learning models and actual residuals

测值,尤其是在4次主要洪峰过程(2021年7月中旬和8月下旬、2022年6月下旬和9月上旬)中,BMA融合残差在峰值响应、涨落趋势和退水过程上均较Hydromad模型拟合更为精细。相比初始模拟值的“矮胖”滞后特性,BMA融合方法能够更灵敏地反映实测流量的快速变化过程,增强了模型对极端来水事件的捕捉能力。

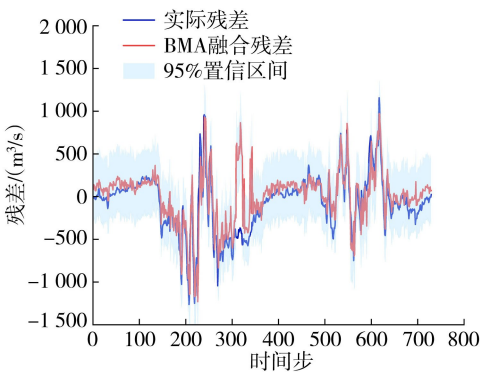


图4 残差预测结果及其95%置信区间

Fig. 4 Residual prediction results and their 95% confidence intervals

选取水文领域常见的NSE、KGE以及百分比偏差(PBIAS)作为评价指标,结果(表3)显示,经BMA融合残差对原模拟值修正后,NSE由0.832提高至0.983,KGE由0.866提升至0.987,说明该方法在拟合精度、相关性、一致性等方面均表现出更强的模拟能力;同时,PBIAS由原始模型的4.05%显著降低至-0.03%,表明系统性偏差被有效消除,预测结果在整体水量控制上更为准确。

综上所述,3种机器学习模型对Hydromad模型残差序列的预测性能呈现较大差异,XGBoost模型在MAE、MSE和 $R^2$ 等评价指标中均表现最优,RF模型次之,SVR模型的模拟残差波动相对较大,预测性能相对较弱。综合利用XGBoost模型的强非线性拟合能力和RF模型的稳健性,并减少SVR模型在高频波动区段的误差,可以显著提升来水预测精度。

3 结 语

本文提出了一种融合机器学习模型与贝叶斯模型平均的水电站来水预测方法,并以金沙江梨园水电站2021—2022年实测数据为例进行了残差建模修正与水电站来水预测不确定性评估。评估结果表明:①3种机器学习模型在残差预测中表现出不同的性能特征,其中XGBoost模型对复杂非线性残差拟合能力较强,RF模型具有较好的稳健性和泛化能力,SVR模型在高频波动区段易产生误差,BMA融合方法能够综合不同模型的预测优势,降低预测误差;②验证期内BMA融合残差修正结果在多个洪峰时段具有较好的响应能力,可识别复杂误差结构,提高水电站来水模拟精度,为水库调度提供更可靠的预测依据。

本文方法具有较强的通用性,可应用于任意水文模型的残差序列处理,并可进一步拓展至其他机器学习方法。但BMA方法权重分配受验证期数据质量和规模影响,未来可通过引入多源观测数据并开展跨区域、多情景的验证,进一步提升模型的适用性与推广潜力。

参考文献:

[1] 丁紫玉,方国华,毛莺池,等.考虑发电与弃水风险的梯级水电站水库风险调度模型[J].水资源保护,2024,40(2):100-106. (Ding Ziyu,Fang Guohua,Mao Yingchi,et al. Risk dispatching model of cascade hydropower station reservoirs considering power generation and water abandonment risks[J]. Water Resources Protection,2024,40(2):100-106. (in Chinese))

[2] 许月萍,陆豪楠,于欣廷,等.基于深度学习集合预报的水库闸门防洪优化调度[J].水科学进展,2024,35(6):900-913.

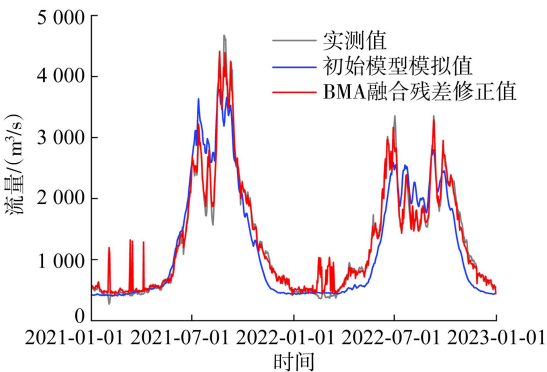


图5 梨园水电站来水模拟值与实测值对比

Fig. 5 Comparison of inflow simulated and measured value of Liyuan Hydropower Station

表3 BMA方法修正前后Hydromad模型预测性能对比

Table 3 Comparison of prediction performance of

Hydromad model before and after correction by BMA method

| BMA方法 | NSE   | KGE   | PBIAS/% |
|-------|-------|-------|---------|
| 修正前   | 0.832 | 0.866 | 4.05    |
| 修正后   | 0.983 | 0.987 | -0.03   |

- (Xu Yueping, Lu Haonan, Yu Xinting, et al. Optimization of gate operation for reservoir flood control based on deep learning ensemble forecasting[J]. *Advances in Water Science*, 2024, 35(6): 900-913. (in Chinese))
- [ 3 ] 钟加星,董增川,孟金玉,等. 基于机器学习的黄河上游梯级水库群多目标优化调度规则提取[J]. 河海大学学报(自然科学版), 2024, 52(6): 30-37. (Zhong Jiaxing, Dong Zengchuan, Meng Jinyu, et al. Extracting multi-objective optimal operation rules of cascade reservoirs in upper reaches of Yellow River based on machine learning[J]. *Journal of Hohai University (Natural Sciences)*, 2024, 52(6): 30-37. (in Chinese))
- [ 4 ] 罗煜宁,张珂,王宇昊,等. 金沙江上游降雨-融雪径流模拟与洪水风险评估[J]. 水资源保护, 2025, 41(3): 93-100. (Luo Yuning, Zhang Ke, Wang Yuhao, et al. Simulation of rainfall-snowmelt runoff and flood risk assessment in upper reaches of the Jinsha River[J]. *Water Resources Protection*, 2025, 41(3): 93-100. (in Chinese))
- [ 5 ] 郭珈源,孙旭杨,刘金涛,等. 基于 Budyko 理论的高寒山区流域径流演变归因分析[J]. 河海大学学报(自然科学版), 2025, 53(3): 22-29. (Guo Jiayuan, Sun Xuyang, Liu Jintao, et al. Attribution analysis of runoff evolution in cold alpine catchment based on Budyko theory[J]. *Journal of Hohai University (Natural Sciences)*, 2025, 53(3): 22-29. (in Chinese))
- [ 6 ] 鞠琴,吴金雨,王兴平,等. 机器学习算法在降水和气温多模式集成中的应用[J]. 水资源保护, 2024, 40(3): 106-115. (Ju Qin, Wu Jinyu, Wang Xingping, et al. Application of machine learning algorithms in multimodal integration of precipitation and temperature[J]. *Water Resources Protection*, 2024, 40(3): 106-115. (in Chinese))
- [ 7 ] 肖渝,孙若辰,王倩,等. 不同降水驱动的 SWAT 模型参数不确定性研究[J]. 河海大学学报(自然科学版), 2024, 52(2): 43-52. (Xiao Yu, Sun Ruochen, Wang Qian, et al. Study on parameter uncertainty in SWAT model driven by different precipitation[J]. *Journal of Hohai University (Natural Sciences)*, 2024, 52(2): 43-52. (in Chinese))
- [ 8 ] 李新,陈馨蕾,周义斌,等. 黄淮海和江淮平原小时极端降水时空演变及归因分析[J]. 水资源保护, 2025, 41(5): 285-295. (Li Xin, Chen Xinlei, Zhou Yibin, et al. Spatiotemporal evolution and attribution analysis of hourly extreme precipitation in the Huang-Huai-Hai Plain and Jiang-Huai Plain[J]. *Water Resources Protection*, 2025, 41(5): 285-295. (in Chinese))
- [ 9 ] 陈靖凯,张志才,陈喜,等. 喀斯特流域降雨-径流过程及其对干湿条件变化响应的模拟分析[J]. 河海大学学报(自然科学版), 2025, 53(1): 41-49. (Chen Jingkai, Zhang Zhicai, Chen Xi, et al. Simulation and analysis of rainfall-runoff processes in karst catchment and their response to changes in dry-wet conditions[J]. *Journal of Hohai University (Natural Sciences)*, 2025, 53(1): 41-49. (in Chinese))
- [ 10 ] 邓超,孙培源,尹鑫,等. 基于概念性水文模型与长短时记忆模型耦合的汉江上游流域径流模拟[J]. 湖泊科学, 2025, 37(1): 279-292. (Deng Chao, Sun Peiyuan, Yin Xin, et al. Runoff simulation in the upper Han River Basin using physics-informed machine learning model[J]. *Journal of Lake Sciences*, 2025, 37(1): 279-292. (in Chinese))
- [ 11 ] 杨肖丽,张佳乐,黄星怡,等. 气候变化和人类活动对长江流域极端径流的影响[J]. 水资源保护, 2025, 41(4): 1-9. (Yang Xiaoli, Zhang Jiale, Huang Xingyi, et al. Impacts of climate change and human activities on extreme runoff in the Yangtze River Basin[J]. *Water Resources Protection*, 2025, 41(4): 1-9. (in Chinese))
- [ 12 ] 张媛媛,姚成,王容,等. 基于地表地下双拦蓄网格模型的清水河流域“23·7”洪水模拟复盘[J]. 河海大学学报(自然科学版), 2026, 54(4): 61-68. (Zhang Yuanyuan, Yao Cheng, Wang Rong, et al. Retrospective simulation of “23·7” flood event in Qingshui River Basin based on surface-subsurface dual retention grid model[J]. *Journal of Hohai University (Natural Sciences)*, 2026, 54(4): 61-68. (in Chinese))
- [ 13 ] 张俊,冯宝飞,牛文静,等. 基于误差分布估计的三峡水库入库洪水概率预报方法[J]. 湖泊科学, 2023, 35(2): 722-729. (Zhang Jun, Feng Baofei, Niu Wenjing, et al. Probability forecasting method of Three Gorges Reservoir inflow flood based on error distribution estimation[J]. *Journal of Lake Sciences*, 2023, 35(2): 722-729. (in Chinese))
- [ 14 ] Jiang Xiaolei, Gupta H V, Liang Zhongmin, et al. Toward improved probabilistic predictions for flood forecasts generated using deterministic models[J]. *Water Resources Research*, 2019, 55(11): 9519-9543.
- [ 15 ] Li Dayang, Marshall L, Liang Zhongmin, et al. Characterizing distributed hydrological model residual errors using a probabilistic long short-term memory network[J]. *Journal of Hydrology*, 2021, 603: 126888.
- [ 16 ] 蒋晓蕾,沈静,郭洛夫杰,等. 考虑误差分布的赣江流域径流概率预报[J]. 水文, 2025, 45(3): 32-39. (Jiang Xiaolei, Shen Jing, Guo Luofujie, et al. Probability streamflow forecasting of Ganjiang River Basin considering the heterogeneous distribution characteristic of forecasting errors[J]. *Journal of China Hydrology*, 2025, 45(3): 32-39. (in Chinese))
- [ 17 ] 管亚硕,连炎清,金君良,等. 基于多模型神经网络的径流缺失数据重建[J]. 河海大学学报(自然科学版), 2026, 54(2): 38-44. (Guan Yashuo, Lian Yanqing, Jin Junliang, et al. Reconstruction of missing runoff data based on multi-model neural networks[J]. *Journal of Hohai University (Natural Sciences)*, 2026, 54(2): 38-44. (in Chinese))
- [ 18 ] 贺新月,曾献奎,王栋. 融合 AR 模型和 MCMC 方法的水文模拟不确定性分析[J]. 河海大学学报(自然科学版), 2020, 48

(2):116-122. (He Xinyue,Zeng Xiankui,Wang Dong. Uncertainty analysis of hydrological simulation with auto regressive model and MCMC method[J]. Journal of Hohai University (Natural Sciences),2020,48(2):116-122. (in Chinese))

[19] 徐嘉远,邹磊,夏军,等. TVGM-LSTM 耦合模型及其径流模拟效果分析[J]. 水资源保护,2023,39(6):104-110. (Xu Jiayuan,Zou Lei,Xia Jun,et al. TVGM-LSTM coupling model and its runoff simulation effect analysis[J]. Water Resources Protection,2023,39(6):104-110. (in Chinese))

[20] Zhong Yixuan,Guo Shenglian,Ba Huanhuan,et al. Evaluation of the BMA probabilistic inflow forecasts using TIGGE numeric precipitation predictions based on artificial neural network[J]. Hydrology Research,2018,49(5):1417-1433.

[21] 刘章君,郭生练,许新发,等. 贝叶斯概率水文预报研究进展与展望[J]. 水利学报,2019,50(12):1467-1478. (Liu Zhangjun,Guo Shenglian,Xu Xinfa,et al. Bayesian probabilistic hydrological forecasting: progress and prospects[J]. Journal of Hydraulic Engineering,2019,50(12):1467-1478. (in Chinese))

[22] Madadgar S,Moradkhani H. Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging[J]. Water Resources Research,2014,50(12):9586-9603.

[23] 祁海霞,彭涛,林春泽,等. 清江流域降水的多模式 BMA 概率预报试验[J]. 气象,2020,46(1):108-118. (Qi Haixia,Peng Tao,Lin Chunze,et al. Probabilistic forecasting of precipitation over the Qingjiang River Basin using BMA multimodel ensemble technique[J]. Meteorological Monthly,2020,46(1):108-118. (in Chinese))

[24] 尹家波,郭生练,王俊,等. 基于贝叶斯模式平均方法融合多源数据的水文模拟研究[J]. 水利学报,2020,51(11):1335-1346. (Yin Jiabo,Guo Shenglian,Wang Jun,et al. Blending multi-source data in hydrological simulations based on BMA method [J]. Journal of Hydraulic Engineering,2020,51(11):1335-1346. (in Chinese))

[25] Zhao Youyi,Luo Shangxue,Cai Jiafang,et al. Monthly precipitation prediction based on the CEEMDAN-BMA model[J]. Water Resources Management,2024,38(14):5661-5681.

[26] 周婷,温小虎,冯起,等. 基于 BMA 多模型组合的疏勒河径流预测研究[J]. 冰川冻土,2022,44(5):1606-1619. (Zhou Ting,Wen Xiaohu,Feng Qi,et al. Study on runoff prediction of Shule River based on BMA multi-model combination[J]. Journal of Glaciology and Geocryology,2022,44(5):1606-1619. (in Chinese))

[27] 欧阳文宇,叶磊,顾学志,等. 深度学习水文预报研究进展综述 II :研究进展及展望[J]. 南水北调与水利科技(中英文),2022,20(5):862-875. (Ouyang Wenyu,Ye Lei,Gu Xuezhi,et al. Review of deep learning for hydrological forecasting II : research progress and prospect[J]. South-to-North Water Transfers and Water Science & Technology,2022,20(5):862-875. (in Chinese))

[28] 梁瀚续,宋克超,张丹,等. IHACRES 模型在不同干湿气候区小流域的应用研究[J]. 水文,2023,43(5):76-83. (Liang Hanxu,Song Kechao,Zhang Dan,et al. Application of IHACRES model in small basins of different dry and wet climatic regions [J]. Journal of China Hydrology,2023,43(5):76-83. (in Chinese))

[29] Croke B F W,Jakeman A J. Use of the IHACRES rainfall-runoff model in arid and semi-arid regions [M]//Wheater H, Sorooshian S,Sharma K D. Hydrological Modelling in Arid and Semi-Arid Areas. Cambridge:Cambridge University Press,2007: 41-48.

[30] Jaberzadeh M,Saremi A,Ghorbanizadeh Kharazi H,et al. SWAT and IHACRES models for the simulation of rainfall-runoff of Dez watershed[J]. Climate Dynamics,2024,62(4):2823-2835.

[31] Andrews F T,Croke B F W,Jakeman A J. An open software environment for hydrological model assessment and development[J]. Environmental Modelling & Software,2011,26(10):1171-1185.

[32] Kokkonen T,Koivusalo H,Jakeman T,et al. Construction of a degree-day snow model in the light of the “ten iterative steps in model development”[C]//Proceedings of the IEMSs Third Biennial Meeting:Summit on Environmental Modelling and Software. Burlington:IEMSs,2006.

[33] Breiman L. Random forests[J]. Machine Learning,2001,45(1):5-32.

[34] Smola A J,Schölkopf B. A tutorial on support vector regression[J]. Statistics and Computing,2004,14(3):199-222.