

DOI:10.3880/j.issn.1004-6933.2023.02.016

基于集成学习模型的非平稳月径流预报

康 艳^{1,2}, 程 潇^{1,2}, 陈沛如^{1,2}, 向 悦^{1,2}, 张芳琴^{1,2}, 宋松柏^{1,2}

(1. 西北农林科技大学水利与建筑工程学院, 陕西 杨凌 712100;

2. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西 杨凌 712100)

摘要: 针对变化环境下月径流序列的非平稳性日益加剧, 传统径流预报模型采用普通学习算法的局限性, 基于 Bagging 和 Boosting 集成学习算法, 构建了随机森林 (RF)、梯度提升决策树 (GBDT) 和轻梯度提升机 (LightGBM) 3 种集成学习模型, 融合弹性网 (EN) 和变分模态分解 (VMD), 建立 VMD-EN-RF、VMD-EN-GBDT 和 VMD-EN-LightGBM 非平稳月径流组合预报模型, 并以黄河流域实测月径流为研究对象, 评估预报结果的不确定性。结果表明: 单一集成学习模型能够提供可靠的预报结果, 适用于非平稳月径流预报; 融合 VMD 和 EN 的集成学习模型预报性能较单一集成学习模型有了显著提高, 纳什效率系数提升了 15% ~ 20%, 均方根误差降低了 30% ~ 40%; 基于 Boosting 集成方法构建的集成学习模型优于 Bagging 集成方法, 其中 VMD-EN-LightGBM 预见期 3 月内的预报效果优于 VMD-EN-RF 和 VMD-EN-GBDT, 在 90% 置信度的区间预报覆盖率高于 90%, 表现出良好的性能。

关键词: 月径流预报; 集成学习算法; 弹性网; 变分模态分解; 黄河流域

中图分类号: TV213.4 **文献标志码:** A **文章编号:** 1004-6933(2023)02-0125-11

Non-stationary monthly runoff forecast based on ensemble learning model // KANG Yan^{1,2}, CHENG Xiao^{1,2}, CHEN Peiru^{1,2}, XIANG Yue^{1,2}, ZHANG Fangqin^{1,2}, SONG Songbai^{1,2} (1. School of Water Resource and Architectural Engineering, Northwest A&F University, Yangling 712100, China; 2. Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling 712100, China)

Abstract: In view of the increasing non-stationary of monthly runoff series under changing environment, the limitation of ordinary learning of traditional runoff forecasting models, and the accurate runoff forecasting is facing new challenges, the research on forecasting consistent with the changing characteristics of the runoff series has been carried out. Based on Bagging and Boosting integrated learning algorithms, three integrated learning models, namely random forest (RF), gradient boosting decision tree (GBDT) and light gradient boosting machine (LightGBM), are constructed. Integrated with elastic net (EN) and variational mode decomposition (VMD), combined forecasting models of VMD-EN-RF, VMD-EN-GBDT and VMD-EN-LightGBM non-stationary monthly runoff are established. The measured monthly runoff of the Yellow River Basin is taken as the research object to evaluate the uncertainty of the forecasting results. The results show that the single ensemble learning model can provide reliable forecasting results and is suitable for complex monthly runoff forecasting. The prediction performance of the integrated learning model integrating VMD and EN is significantly improved compared with the single integrated learning model. The Nash efficiency coefficient is increased by 15% ~ 20%, and the root mean square error is reduced by 30% ~ 40%. The ensemble learning model built based on Boosting is better than that built based on Bagging. Among them, the prediction effect of VMD-EN-LightGBM within 3-month forecast period is better than that of VMD-EN-RF and VMD-EN-GBDT, and the coverage rate is higher than 90% in the 90% confidence interval, showing good performance.

Key words: monthly runoff forecast; ensemble learning algorithm; elastic net; variational mode decomposition; Yellow River Basin

黄河流域水资源匮乏问题突出,精确预报河川径流可为流域水资源规划和管理提供依据,有效促进黄河流域生态保护和高质量发展。流域内不同地区气候差异显著,且人类活动对径流扰动强烈,径流的随机性和非平稳性日益加剧,径流预报面临新的挑战,科学精确的径流预报模型成为研究热点^[1]。国内外学者构建了许多水文模型用于径流预报,可大致分为两类:一类是过程驱动模型,另一类是数据驱动模型^[2]。过程驱动模型关注水文循环的物理机制^[3],但建模过程复杂、模型参数不易确定、输入数据量大。基于水文统计学的数据驱动模型受平稳性和线性假设的约束,只能提供有限的模拟精度^[4-5];基于机器学习的数据驱动模型能够捕捉变量之间的非线性信息,获得准确的径流预报^[6-7],如 Zuo 等^[8]利用 LSTM 神经网络构建了可预测高度非平稳径流的预报模型。这些径流预报模型依赖个体学习器的普通学习,适用于不同的数据,且受限个体学习器的模拟性能,泛化能力往往不如多学习器的集成。基于多学习器的集成学习模型克服了传统径流预报模型面临的统计问题和计算问题,在径流、地下水、洪水和干旱模拟研究中取得了较好的效果^[9]。随机森林(random forest, RF)是集成学习的代表模型之一,RF 基于 Bagging 集成学习算法,在预测地表水文过程时获得了较传统机器学习模型更低的误差和更高的稳定性^[10-11]。与 RF 不同,梯度提升决策树(gradient boosting decision tree, GBDT)模型采用 Boosting 集成学习算法,学习时更关注被前一个学习器错误预测的样本,以弥补前者不足,有利于减少偏差和方差,提高预测精度^[9]。轻梯度提升机(light gradient boosting machine, LightGBM)作为 GBDT 模型的改进模型,具有运算快速、准确度高、内存占用低等优点,受到了广泛关注,被应用于电力、交通等多个领域^[12-13],应借鉴到水文领域,以为时空异质性较强的大尺度流域的径流预报提供新方法。由于流域水循环的复杂性,水文预报存在不确定性,借助数据驱动模型开展径流预报属于确定性预报方法,无法刻画不确定性演化特征^[14]。考虑径流预报误差是不确定性的一种表现,可采用非参数估计对其分布近似拟合,进而量化径流预报的不确定性。目前,众多研究集中于传统径流预报模型开展径流预报研究,缺乏对 RF、GBDT、LightGBM 等多种集成学习模型的综合探讨及其预报不确定性分析,集成学习算法在水文预报领域的潜力还有待深入探索。

合理的预报因子有助于集成学习模型建立输入和输出之间的最优关系,进而提高模型精度和计算

速度。常用的岭回归、最小绝对收缩和选择算法(least absolutely shrinkage and selection operator, Lasso)等嵌入式选择方法具有较好的筛选效果^[15-16]。正则化方式有 L1 和 L2 两种,运用不同的方式会对特征系数产生不同的影响,L1 正则化会产生稀疏解,而 L2 使特征系数接近 0。岭回归采用 L2 正则化对特征系数进行压缩,但无法产生稀疏解;为了弥补岭回归的不足,Lasso 以 L1 为惩罚项使弱相关特征的系数压缩为 0,选择出强相关特征,如 Chu 等^[17]利用 Lasso 识别关键预报因子,有效地提高了模型的预测精度。但预报因子之间关系复杂,且具有不可忽视的相依性^[18],Lasso 存在不能处理多重共线性问题的缺陷^[19],可能无法有效分析相依月径流序列。为了克服这一缺点,Zou 等^[20]提出的弹性网(elastic net, EN)结合了 L1 和 L2 正则化的优点,在稳定产生稀疏解的同时,又能保证复共线性^[21]。EN 法选择的预报因子具有明确的物理意义,可解释性强^[22],受到了各领域学者的关注,但其对水文预报因子的识别能力还有待研究。

集成学习算法关注高误差预测样本,但同时也会导致对噪声敏感、模型泛化能力低等问题^[23]。分解技术对消除径流序列的噪声具有积极作用^[24]。小波变换在径流分解中应用较为广泛,可以对频域和时域多尺度分析,但分解效果取决于母小波的选择和分解层数,受人为因素影响较大^[25];经验模态分解(empirical mode decomposition, EMD)自适应性强,分解前无须设定任何基函数,但容易产生模态混叠问题;集合经验模态分解(ensemble empirical mode decomposition, EEMD)是基于 EMD 改进的信号分解方法,通过加入高斯白噪声,有效避免了模态混叠问题,但也存在重构后误差加大的风险^[26-27]。针对这些问题,Dragomiretskiy 等^[28]提出的变分模态分解(variational mode decomposition, VMD)是一种自适应、完全非递归的分解方法,具有更好的噪声鲁棒性、抗混叠能力和分解效果^[29]。研究表明序列分解技术可以有效提升模型预报精度^[2],现有研究多考虑单一集成学习模型预报,未充分探讨融合分解技术和预报因子识别方法对集成学习模型性能的影响。

基于此,本文以黄河流域兰州站、头道拐站、龙门站和高村站的实测月径流资料作为研究对象,将 EN 法、VMD 分解分别与 RF、GBDT、LightGBM 3 种集成学习模型融合,构建非平稳月径流组合预报模型,并借助核密度估计(kernel density estimation, KDE)刻画预报的不确定性。通过综合对比分析,揭示集成学习模型对非平稳径流序列的预报能力,

以期变化环境下径流预报精度的提升提供新思路。

1 研究方法

1.1 VMD

VMD 是 Dragomiretskiy 等^[28]于 2014 年提出的一种信号分解方法,其根据预设的模式数 K ,将输入信号分解为具有不同中心频率的有限带宽的模式分量^[30]。VMD 的主要框架是变分问题的构建和求解。若原始径流序列为非平稳输入信号 y , y 被分解为 K 个有限带宽的模式分量,同时每个模式分量的估计带宽之和最小。约束条件为所有模式分量之和等于原始信号 y ,变分约束表达式为

$$L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left\{ \left[\delta(t) + \frac{h}{\pi t} \right] \otimes u_{kt} \right\} e^{-hw_{kt}} \right\|_2^2 + \left\| y_t - \sum_{k=1}^K u_{kt} \right\|_2^2 + \langle \lambda_t, y_t - \sum_{k=1}^K u_{kt} \rangle \quad (1)$$

采用交替方向乘子法对 u_k 、 w_k 和 λ 进行迭代更新,直至满足式(2)的约束条件:

$$\sum_{k=1}^K \frac{\|u_{k,p+1} - u_{k,p}\|_2^2}{\|u_{k,p}\|_2^2} < \varepsilon \quad (2)$$

式中: u_k 为第 k 个模式; w_k 为第 k 个模式的中心频率; λ 为拉格朗日乘子; α 为二次惩罚项; ∂_t 为函数对时间的偏导; $\delta(\cdot)$ 为狄拉克函数; h 为虚数单位; t 为时间变量; ε 为精度收敛值; p 为迭代次数。

1.2 EN 法

EN 法是 Zou 等^[20]2005 年提出的嵌入式特征选择方法,其既含有岭回归中的 L_2 惩罚项,又含有 Lasso 中的 L_1 惩罚项,因此,EN 在对特征系数同比例压缩时,可以实现将部分特征系数压缩为 0 的目的,并兼具岭回归和 Lasso 的优点^[31]。若响应变量为 Y ,解释变量为 X ,EN 的表达式为

$$\hat{\beta} = \operatorname{argmin}(\|Y - X\beta\|^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2) \quad (3)$$

式中: $\hat{\beta}$ 为压缩系数估计值; λ_1 、 λ_2 为正则化参数; β 为压缩系数。令 $\alpha_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, $\lambda' = \lambda_1 + \lambda_2$, 式

(3) 等价于

$$\hat{\beta} = \operatorname{argmin} \{ \|Y - X\beta\|^2 + \lambda' [\alpha_1 \|\beta\|_1 + (1 - \alpha_1) \|\beta\|_2^2] \} \quad (4)$$

式中: λ' 为惩罚系数, $\alpha_1 \|\beta\|_1 + (1 - \alpha_1) \|\beta\|_2^2$ 为 EN 的惩罚项。

EN 的惩罚项是岭回归和 Lasso 惩罚项的凸组合,当 $\alpha_1 = 0$ 时,EN 退化为岭回归,当 $\alpha_1 = 1$ 时,EN 退化为 Lasso。EN 的惩罚力度与 λ' 的取值有

关, λ' 值越小,被选中的变量越多; λ' 越大,惩罚力度越大,被压缩为 0 的系数越多,被选中的变量越少。因此,需选择合适的 λ' 使模型获得可靠的解释变量和精确的预报结果。 α_1 和 λ' 通过交叉验证确定。

1.3 LightGBM

LightGBM 是基于 Boosting 算法的集成学习模型^[32],相较于 GBDT, LightGBM 采用梯度单边采样技术和互斥特征捆绑方法提高模型训练效率和速度^[33]。梯度单边采样技术认为梯度较大的数据对信息增益的贡献大,它会选择出梯度较大的数据,对梯度较小的数据随机采样,并赋予小梯度数据权重补偿,使 LightGBM 在不改变原始数据分布的基础上,更加关注训练不足的数据。在高维特征稀疏空间里,许多特征是互斥的,即两个特征不同时为 0。互斥特征捆绑方法可以将互斥的特征捆绑在一起,减少特征的维度,加快模型的训练速度,同时, LightGBM 通过直方图算法将连续的数据离散化,遍历离散的数据寻找最优分割点,代替了传统的预排序。

若 $D = \{(X_i, y_i) \mid i = 1, 2, \dots, n\}$, 每一个输入变量 X_i 都是 s 维变量,决策树为 f , 可得 LightGBM 表达式为

$$\tilde{V}_j(d) = \frac{1}{n} \left[\frac{(\sum_{X_i \in A_l} g_i + \frac{1-a}{b} \sum_{X_i \in B_l} g_i)^2}{n_{l,j}(d)} + \frac{(\sum_{X_i \in A_r} g_i + \frac{1-a}{b} \sum_{X_i \in B_r} g_i)^2}{n_{r,j}(d)} \right] \quad (5)$$

$$f_p = \sum_{p=1}^P f_p(X) \quad (6)$$

其中

$$\begin{aligned} A_l &= \{X_i \in A; X_{ij} \leq d\} & A_r &= \{X_i \in A; X_{ij} > d\} \\ B_l &= \{X_i \in B; X_{ij} \leq d\} & B_r &= \{X_i \in B; X_{ij} > d\} \end{aligned}$$

式中: $\tilde{V}(d)$ 为方差增益; j 为特征; d 为分裂点; g 为损失函数的负梯度方向; A 、 B 为数据子集, a 、 b 为对应比例; P 为树的数目; f_p 为第 p 棵树。

1.4 KDE

KDE 属于非参数估计方法^[34],不需要对误差分布进行假设,避免了参数估计对分布假设不合理的问题。设 $X_1, X_2, \dots, X_N$ 是未知密度函数 $f(x)$ 的独立同分布随机变量, $f(x)$ 的 KDE 为

$$F_K = \frac{1}{Nl} \sum_{i=1}^N K\left(\frac{x - X_i}{l}\right) \quad (7)$$

式中: F_K 为 KDE 值; N 为数据容量; l 为窗宽; $K(\cdot)$ 为核函数。

在给定样本数据的情况下,KDE 的估计精度取决于窗宽和核函数。常用的核函数有三角核函数、Epanechnikov 核函数、高斯核函数等。高斯核函数具有良好的可微性和光滑度^[35],本文选用高斯核函数,取渐进积分均方误差最小时的窗宽为最优窗宽。

1.5 预报精度评价

选取纳什效率系数 (Nash-Sutcliffe efficiency coefficient, NSE) 和均方根误差 (root mean squared error, RMSE) 作为确定性预报的精度评价指标。选取覆盖率 C_R 和平均相对带宽 R_B 作为概率区间预报的精度评价指标^[36],计算公式为

$$C_R = \frac{1}{n} \sum_{i=1}^n c_i \quad (8)$$

$$R_B = \frac{1}{n} \sum_{i=1}^n \left(\frac{U_i - L_i}{y_i} \right) \times 100\% \quad (9)$$

式中: y_i 为实测值; c_i 为布尔量,若 y_i 处于置信区间内,则 $c_i = 1$, 否则为 0; U_i 为区间上限; L_i 为区间下限。

2 实例应用

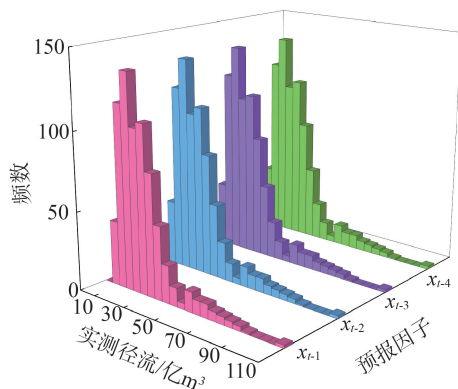
2.1 研究区概况

黄河发源于巴颜喀拉山脉,依次流经青海、四川、甘肃、宁夏、内蒙古、陕西、山西、河南和山东 9 个省区,干流全长约 5464 km,流域面积达 79.5 万 km^2 ^[37-38]。流域自西向东横跨青藏高原、内蒙古高原、黄土高原和江淮平原,地势西高东低,高差较大。受大气环流和季风环流的影响,流域内气候差异显著,大致可分为干旱、半干旱和半湿润气候。流域冬干春旱,夏秋多雨,多年平均年降水量约为 450 mm,且时空分布不均,全年降水量的 60% ~ 75% 集中在 6—9 月^[39]。近几十年来,径流年际年内分配差异增加,空间分布不均匀性增大^[40-41]。

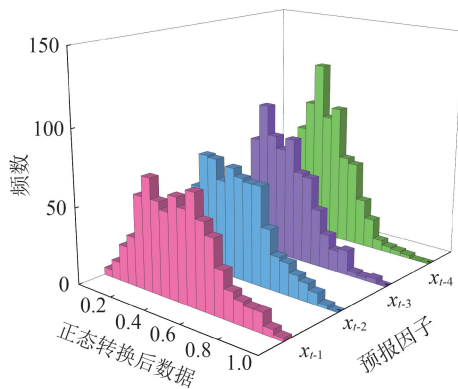
采用黄河流域兰州、头道拐、龙门和高村 4 个水文站 1956—2016 年实测月径流资料进行分析。气候变化和河道取水增加、实施水土保持措施、修建水利水电工程等高强度的人类活动对流域水文情势扰动显著,流域下垫面发生变化,极端水文事件的发生频率和严重程度加大,月径流的季节周期和年周期弱化^[42],径流序列的复杂性和非平稳性更加突出。在变化环境下,4 个水文站的径流序列均有显著减少趋势,兰州站径流序列在 1986 年发生突变,其余 3 站径流序列在 1990 年发生突变^[43]。为有效捕捉变异点后的径流变化特征,划分 1956—2006 年为训练期、2007—2016 年为检验期。

2.2 单一集成学习模型

构建模型之前,需对径流数据进行规范化处理。数据挖掘和分析过程要求数据呈正态化,Box-Cox 变换可以在不丢失原始序列信息的基础上,消除实测径流数据呈偏态分布的影响,具有较好的正态变换效果,因此,采用 Box-Cox 变换将初选预报因子(预报月前 1~12 月的月径流 x_{t-1} 、 x_{t-2} 、 $\dots$ 、 x_{t-12})转换为正态分布。图 1 为兰州站预报因子 x_{t-1} 、 x_{t-2} 、 x_{t-3} 、 x_{t-4} 正态变换前后的直方图。



(a) 转换前



(b) 转换后

图 1 兰州站预报因子的 Box-Cox 正态变换结果

Fig.1 Box-Cox transformation of predictors at Lanzhou hydrological station

正态变换后的预报因子对预报径流的影响程度各不相同,利用相关系数 R 分析预报径流与预报因子之间的相关性。 $|R| < 0.4$ 为低线性相关, $0.4 \leq |R| < 0.7$ 为显著相关, $0.7 \leq |R| < 1$ 为高线性相关,取 $|R| \geq 0.4$ 的预报因子作为关键预报因子。根据相关系数法筛选的关键预报因子如图 2 所示。

将筛选的关键预报因子分别作为 RF、GBDT 和 LightGBM 的输入,利用训练期月径流数据构建径流预报模型。借助网格搜索算法确定径流预报模型的最优参数组合。网格搜索算法最早被用于优化支持

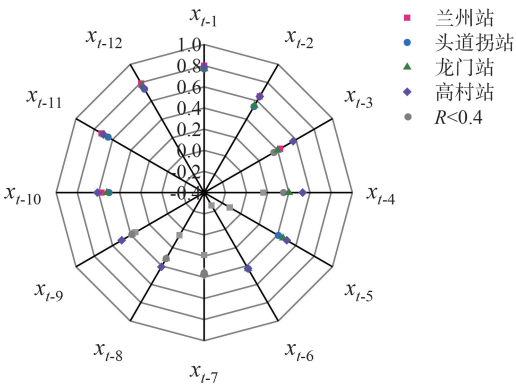


图2 相关系数法筛选关键预报因子

Fig. 2 Screening of key predictors based on correlation coefficient

向量机,本质上是一种穷举搜索算法,其可将各个参数的可能取值进行排列组合,排列组合的结果生成“网格”,遍历所有的参数组合进行寻优,同时结合 k 折交叉验证,每次选择 $k-1$ 组数据作为训练集,剩余的一组数据作为验证集,评估模型效果,如此反复更换验证集,直至每一组数据均作为验证集被预报过,模型训练结束^[44-45]。各个模型的参数寻优结果如表1和表2所示。将检验期的关键预报因子输入训练好的模型中,得到检验期径流预报结果。单一集成学习模型在检验期的径流预报效果如表3所示。

表1 RF和GBDT模型的最优参数取值

Table 1 Optimal parameter values of RF and GBDT models

水文站	模型	n_estimators	max_depth	random_state
兰州站	RF	110	7	2
	GBDT	100	3	6
头道拐站	RF	80	8	4
	GBDT	70	2	3
龙门站	RF	150	13	7
	GBDT	50	3	8
高村站	RF	110	8	10
	GBDT	50	3	1

表2 LightGBM模型的最优参数取值

Table 2 Optimal parameter values of LightGBM model

水文站	n_estimators	max_depth	colsample_bytree	reg_lambda	learning_rate
兰州站	140.00	2.00	0.80	0.60	0.05
头道拐站	90.00	1.00	0.93	1.00	0.10
龙门站	80.00	8.00	0.95	0.90	0.06
高村站	90.00	4.00	0.96	0.40	0.04

由表3可见,4个水文站的NSE不低于0.620, RMSE不高于6.300亿 m^3 , RF和GBDT模型的预报效果略优于LightGBM模型。总体来说,3个集成学习模型的预报精度尚可。集成学习模型的学习器是决策树,决策树属于有监督学习范畴,其将对象与属

表3 检验期单一集成学习模型月径流预报效果评价

Table 3 Evaluation on effect of monthly runoff prediction of single ensemble learning models during verification period

模型	NSE				RMSE/亿 m^3			
	兰州站	头道拐站	龙门站	高村站	兰州站	头道拐站	龙门站	高村站
RF	0.627	0.824	0.805	0.793	6.290	3.536	3.825	5.107
GBDT	0.710	0.803	0.794	0.804	5.540	3.750	3.934	4.967
LightGBM	0.700	0.758	0.797	0.794	5.636	4.150	3.908	5.098

性之间的映射关系表现为树状结构,每棵树由节点和分支组成,被认为具有处理高度非线性问题的能力^[46],因此在预报径流序列时更有优势。通过Bagging或Boosting集成学习算法集合多棵决策树,可以得到更优越的模型性能,结果验证了集成学习模型预报黄河流域径流的有效性。

2.3 融合EN的集成学习模型

EN法筛选预报因子的关键在于 α_1 和 λ' 的选择,本文利用10折交叉验证确定参数的取值。为避免岭回归算法带来的过拟合问题, α_1 的取值范围为 $[0.5, 1)$,对于确定的 α_1 ,基于MSE准则计算 λ' ,取MSE最小时的 $\lambda_{\min}$ 作为 α_1 对应的 λ' 。EN各个参数的取值如表4所示。

表4 EN的参数取值

Table 4 Parameter values of EN

水文站	α_1	$\lambda_{\min}$
兰州站	0.5000	0.0004
头道拐站	0.5000	0.0007
龙门站	0.5000	0.0002
高村站	0.5000	0.0004

预报因子的压缩系数随 λ' 的变化轨迹如图3所示。当 λ' 充分大时,4个水文站只有 x_{t-1} 被选择,其他预报因子均被剔除,说明 x_{t-1} 对预报径流的影响深刻,二者相关性强;随着 λ' 的减小,惩罚力度变弱,被压缩为0的系数越来越少,与预报径流相关的预报因子被依次选择。依据MSE准则, $\lambda' = \lambda_{\min}$ 时系数不为0的预报因子为关键预报因子,将关键预报因子输入集成学习模型,获得预报径流。融合EN的集成学习模型在检验期的径流预报效果评价如表5所示。

由表5可见,融合EN的集成学习模型的径流预报效果较单一集成学习模型有所提升。总体而言,EN-RF较RF,NSE提高了约1.52%,RMSE降低了约0.29%;EN-GBDT较GBDT,NSE提高了约1.25%,RMSE降低了约1.52%;EN-LightGBM较LightGBM,NSE提高了约3.19%,RMSE降低了约5.19%,表明有效输入可以提高模型的预报精度。

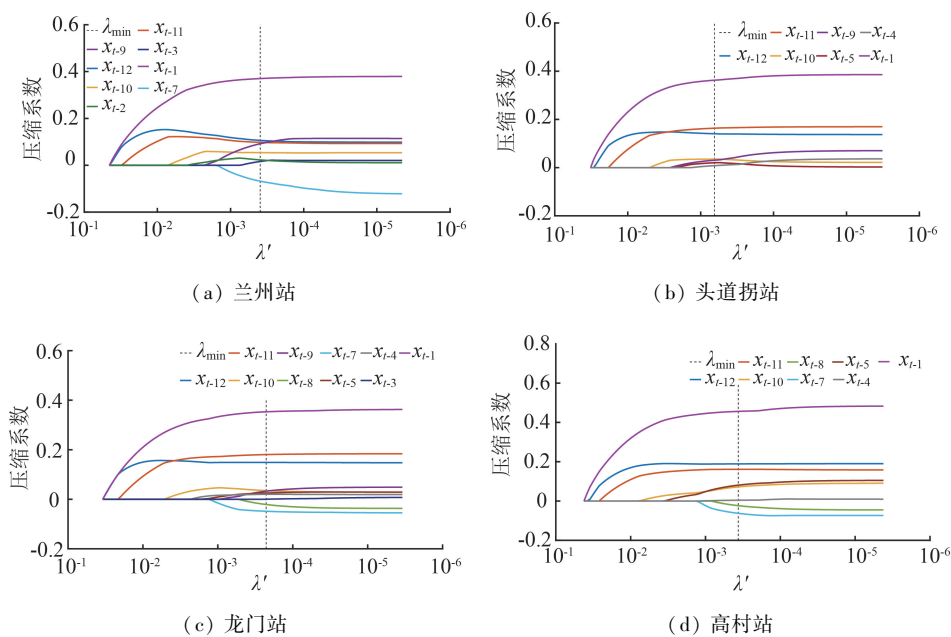


图3 预报因子的压缩系数随 λ' 的变化轨迹

Fig.3 Variation track of compressibility coefficient of predictors with λ'

表5 检验期融合 EN 的集成学习模型月径流预报效果评价

Table 5 Evaluation on effect of monthly runoff prediction of ensemble learning models integrated with EN during verification period

模型	NSE				RMSE/亿 m ³			
	兰州站	头道拐站	龙门站	高村站	兰州站	头道拐站	龙门站	高村站
EN-RF	0.680	0.794	0.807	0.801	5.825	3.536	3.811	5.015
EN-GBDT	0.746	0.800	0.818	0.782	5.190	3.750	3.701	5.241
EN-LightGBM	0.718	0.819	0.802	0.806	5.467	4.150	3.860	4.949

融合 EN 之后,基于 Boosting 算法的集成学习模型优于基于 Bagging 算法的集成学习模型,兰州站、头道拐站、龙门站和高村站的最优模型分别为 EN-GBDT、EN-LightGBM、EN-GBDT 和 EN-LightGBM。

由此可见,EN 法识别的预报因子有效地提高了预报精度和模型稳定性,在预报因子较少的情况下模型仍能提供准确的径流预报。EN 采用 L1 和 L2 联合惩罚,在确保组效应的同时,可以达到准确筛选关键预报因子的目的,减少冗余或无关因子对模型的干扰,有助于提高集成学习模型的预报精度。有效的预报因子筛选方法可以解决维数灾难和过拟合问题,还可以降低噪声影响^[47]。因此,融合 EN 的集成学习模型获得了较单一集成学习模型更好的预报效果。

2.4 融合 VMD 和 EN 的集成学习模型

利用 VMD 分解将各水文站的实测径流分解为不同频率的子序列。VMD 的参数有惩罚因子 α 、模态个数 K 、收敛条件 ε 和噪音容限 τ 。其中最重要的是 K , K 过小会导致原始序列的信息丢失, K 过大会使不同的子序列包含相同的频率,为获得合理

的 K 值,通过传统的 EMD 分解方法确定 K 取为 8,其他参数取默认值^[48]。VMD 分解得到的各子序列的关键预报因子采用 EN 法确定,将关键预报因子输入构建好的模型,得出各个模型在检验期的月径流预报结果。融合 VMD 和 EN 的集成学习模型在检验期的月径流预报效果评价如表 6。

由表 6 可见,与未使用分解技术相比,融合 VMD 的集成学习模型均获得了令人满意的预报效果,模型性能有了不同程度的提升。与 EN-RF 相比,VMD-EN-RF 的 NSE 提高了约 16.78%,RMSE 降低了约 30.77%;与 EN-GBDT 相比,VMD-EN-GBDT 的 NSE 提高了约 14.16%,RMSE 降低了约 29.90%;与 EN-LightGBM 相比,VMD-EN-LightGBM 的 NSE 提高了约 15.97%,RMSE 降低了约 33.80%。表明径流受到人类活动等的多重影响,预报模型在噪声的干扰下往往不能充分识别径流变化,制约了预报模型在中长期预报中的应用。数据分解技术对这一问题的解决起着至关重要的作用,其能够消除径流序列中的噪声,分离不同频率的子序列,使模型获取原始径流序列中隐藏的

表 6 检验期融合 VMD 和 EN 的集成学习模型月径流预报效果评价

Table 6 Evaluation on effect of monthly runoff prediction of ensemble learning models integrated with VMD and EN during verification period

预见期/ 月	模型	NSE				RMSE/亿 m ³			
		兰州站	头道拐站	龙门站	高村站	兰州站	头道拐站	龙门站	高村站
1	VMD-EN-RF	0.918	0.905	0.884	0.870	2.954	2.602	2.953	4.055
	VMD-EN-GBDT	0.918	0.902	0.889	0.877	2.952	2.648	2.892	3.943
	VMD-EN-LightGBM	0.928	0.913	0.905	0.889	2.757	2.495	2.669	3.743
2	VMD-EN-RF	0.898	0.885	0.880	0.839	3.282	2.865	3.007	4.500
	VMD-EN-GBDT	0.900	0.881	0.880	0.852	3.255	2.915	3.007	4.310
	VMD-EN-LightGBM	0.908	0.899	0.901	0.871	3.117	2.694	2.734	4.030
3	VMD-EN-RF	0.890	0.876	0.872	0.845	3.396	2.992	3.118	4.399
	VMD-EN-GBDT	0.896	0.864	0.871	0.856	3.308	3.131	3.127	4.237
	VMD-EN-LightGBM	0.898	0.883	0.889	0.861	3.275	2.910	2.905	4.176

信息,并利用这些信息进行准确预报,不仅降低了建模的难度,还有效地提高了模型的泛化能力和预报精度。

VMD 和 EN 使 3 个集成学习模型的性能有了显著提升,尤其是 LightGBM 模型。融合 VMD 和 EN 的 LightGBM 模型的 NSE 提升了约 18.98%,RMSE 降低了约 39.57%。从预报结果可以看出,VMD-EN-LightGBM 模型的径流预报效果明显优于 VMD-EN-RF 和 VMD-EN-GBDT,其 NSE 不低于 0.880, RMSE 不高于 3.750 亿 m³,获得了可靠的径流预报。VMD 和 EN 弥补了 Boosting 集成学习算法对噪声敏感的缺点,显著提高了 LightGBM 模型的径流预报能力;特别地,与大多数 GBDT 的 level-wise 分裂策略不同,LightGBM 决策树的生成遵循 leaf-wise 分裂策略,选择增益最大的叶子分裂,避免了分裂增益较小的叶子带来的消耗,能在有限的分裂次数下保证更低的预报误差。

为进一步验证基于集成学习算法的组合模型性能,对比分析不同预见期组合模型的预报效果。当预见期分别为 2 月和 3 月时,各个模型的预报误差较预见期 1 月的有所升高,但仍能提供较高的预报

精度。在检验期,各个模型的 NSE 不低于 0.830, RMSE 不高于 4.500 亿 m³。相较于 VMD-EN-RF 和 VMD-EN-GBDT, VMD-EN-LightGBM 模型具有更高的预报精度和模型稳定性,其 NSE 不低于 0.860, RMSE 不高于 4.180 亿 m³,表现出了良好的性能,是 4 个水文站的最优模型。

图 4 为预见期 1 月时检验期实测径流值和预报径流值的散点图,可以看出,对于高度复杂的实测径流序列,3 个组合模型都能提供可靠的径流预报。除少数极值之外,3 个组合模型对的预报值与实测值之间只存在轻微差异,预报效果较好。对于不同的站点,各预报模型在兰州站、头道拐站和龙门站的预报值最接近实测值,而高村站的预报值与实测值之间存在的偏差大于其他水文站。研究表明,在人类活动的影响下,从黄河上游至黄河下游,实测径流与天然径流的差值逐渐增大^[43],因此高村站的预报效果劣于其他 3 个站点。与 VMD-EN-RF 和 VMD-EN-GBDT 相比, VMD-EN-LightGBM 模型的预报结果最接近 45°线,表明其能充分掌握水文序列中的信息,进而稳定地、准确地预报径流。

综上所述,融合 VMD 和 EN 的 3 个集成学习模

-- 45°线 ● VMD-EN-RF ● VMD-EN-GBDT ● VMD-EN-LightGBM

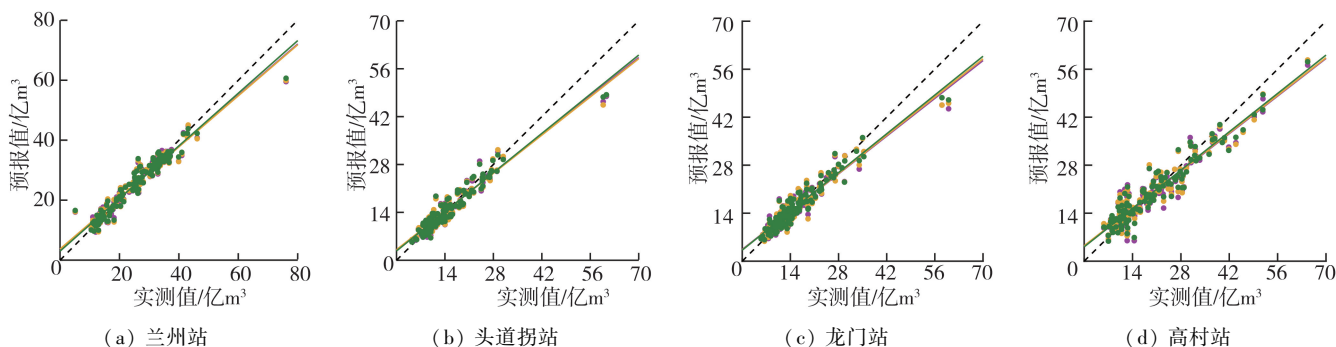


图 4 预见期 1 月时检验期实测径流值和预报径流值散点图

Fig. 4 Scatter diagram of measured runoff values and predicted runoff values during verification period with 1-month forecast period

型能稳定预报研究区预见期 3 月内的径流,其中,VMD-EN-LightGBM 模型更能充分捕捉预报因子与预报径流之间的依赖关系,准确预测动态非平稳的径流序列,提供可靠的径流预报。

2.5 概率区间预报结果

为深入探讨分析融合 VMD 和 EN 的集成学习模型对非平稳径流的预报能力,采用 KDE 对 VMD-EN-LightGBM 模型在检验期的预报误差进行拟合,获得特定置信度下的误差波动区间和预报区间。以 VMD-EN-LightGBM 模型在兰州站的预报误差为例,图 5 为 KDE 对预报误差的拟合结果。由图 5(a)(b)(c)可见,预报误差主要集中在 0 附近,随着预见期的增加,径流预报的不确定性也随之增加,预报误差越来越大。同时可以看出,采用 KDE 拟合预报误差效果较好,可以真实反映预报误差的分布特征。根据不同预见期的概率密度分布,可以得到相应的累积概率分布,如图 5(d)所示。基于各个水文站的累积概率分布,表 7 给出了 90% 置信度下的预报误

差波动区间。

根据表 7 的误差波动区间和 VMD-EN-LightGBM 模型的确定性预报结果,获得 90% 置信度下的预报区间,采用覆盖率 C_R 和平均相对带宽 R_B 对概率区间预报效果进行评价。表 8 为各水文站检验期的概率区间预报效果,图 6 为 1 月预见期 90% 置信度下的概率区间预报结果。由表 8 可见,预见期为 3 月内的 C_R 值不低于 0.915,90% 置信度下的概率区间覆盖了绝大多数的实测点,表明提取的置信区间是可靠的;随预见期的增加, R_B 值逐渐增大,表明预测精度随预见期的增加逐渐降低,不确定性也逐渐增大,符合确定性预报模型的预报规律。由图 6 可见,与其他 3 个水文站相比,受人类活动干扰强烈的流域下游的高村站预报区间最宽,表明流域下游径流非平稳特征更显著,预报不确定性比上游更高,但大部分实测值都落在了预报区间内。由此可见,VMD-EN-LightGBM 模型能够提供可靠的区间预报。

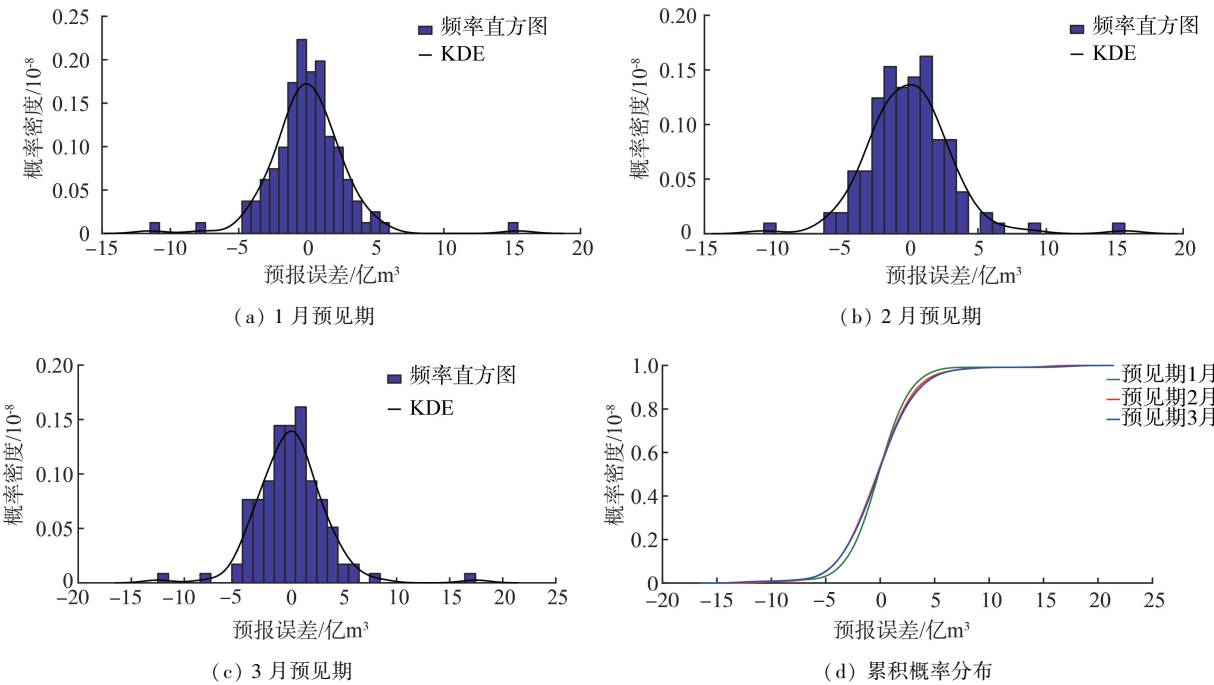


图 5 兰州站的预报误差分布

Fig. 5 Prediction error distribution at Lanzhou hydrological station

表 7 90% 置信度下的误差波动区间

Table 7 Error fluctuation intervals within 90% confidence level

预见期/月	误差/亿 m ³			
	兰州站	头道拐站	龙门站	高村站
1	[-4.328,4.047]	[-3.827,3.457]	[-4.374,4.147]	[-6.604,7.012]
2	[-5.152,4.789]	[-3.880,4.056]	[-4.227,4.538]	[-7.611,7.044]
3	[-5.100,5.053]	[-4.255,4.423]	[-4.545,5.118]	[-7.581,7.359]

表 8 各水文站检验期的概率区间预报效果

Table 8 Prediction effect of probability interval during verification period of different hydrological stations

预见期/ 月	C_R				$R_B/\%$			
	兰州站	头道拐站	龙门站	高村站	兰州站	头道拐站	龙门站	高村站
1	0.942	0.958	0.942	0.917	23.9	30.7	33.9	42.5
2	0.916	0.950	0.933	0.933	28.6	33.7	35.0	46.1
3	0.932	0.958	0.941	0.932	29.4	37.0	38.8	47.3

• 实测值 ■ 概率预测区间

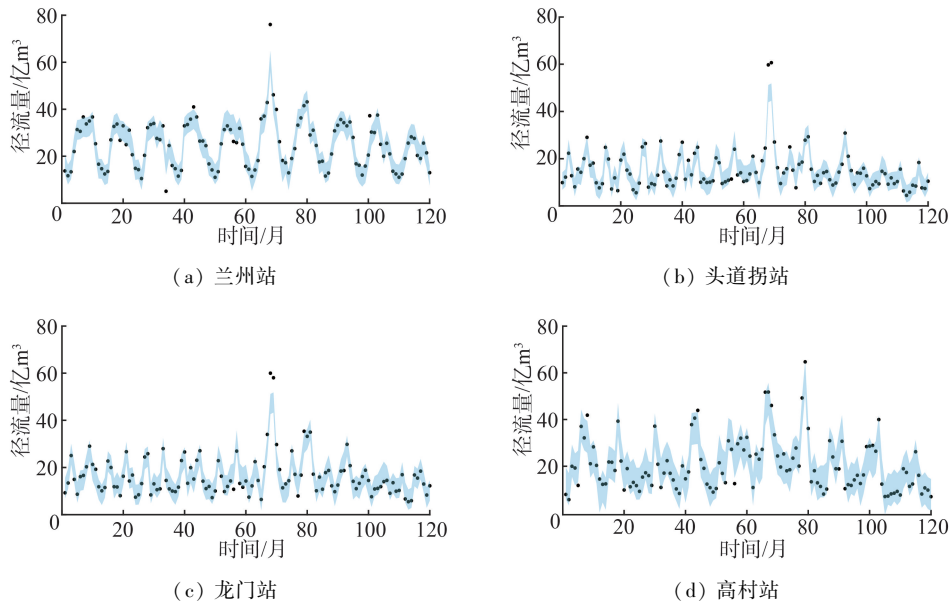


图 6 1 月预见期 90%置信度下的概率区间预报结果

Fig. 6 Prediction results of probability interval within 90% confidence level of 1-month forecast period

3 结 论

a. 单一集成学习模型在 4 个水文站的预报结果可靠且模型性能稳定,表明集成学习模型适用于黄河流域径流预报。

b. EN 法具有稳定产生稀疏解和保证复共线性的优点,其识别的预报因子有效地提高了单一集成学习模型的预报精度和模型稳定性,适合研究区的预报因子优选。

c. 融合 VMD 和 EN 的 VMD-EN-RF、VMD-EN-GBDT 和 VMD-EN-LightGBM 组合预报模型,其预报效果显著优于单一集成学习模型,模型性能大幅提升。随预见期的增加,模型的预报精度有所降低,但仍能提供准确的径流预报,4 个水文站的 NSE 均高于 0.830, RMSE 均小于 4.500 亿 m^3 。

d. Boosting 集成方法优于 Bagging 集成方法, VMD-EN-LightGBM 优于 VMD-EN-GBDT 和 VMD-EN-RF。VMD-EN-LightGBM 模型预见期在 3 月内的区间预报覆盖率高于 90%,表明 VMD-EN-LightGBM 能够精确预报复杂非平稳月径流序列。

参考文献:

- [1] 刘昌明,刘小莽,田巍,等.黄河流域生态保护和高质量发展亟待解决缺水问题[J].人民黄河,2020,42(9):6-9. (LIU Changming, LIU Xiaomang, TIAN Wei, et al. Ecological protection and high-quality development of the Yellow River Basin urgently need to solve the water shortage problem[J]. Yellow River, 2020, 42(9): 6-9. (in Chinese))
- [2] 梁浩,黄生志,孟二浩,等.基于多种混合模型的径流预测研究[J].水利学报,2020,51(1):112-125. (LIANG Hao, HUANG Shengzhi, MENG Erhao, et al. Runoff prediction based on multiple hybrid models[J]. Journal of Hydraulic Engineering, 2020, 51(1): 112-125. (in Chinese))
- [3] 张珂,牛杰帆,李曦,等.洪水预报智能模型在中国半干旱半湿润区的应用对比[J].水资源保护,2021,37(1):28-35. (ZHANG Ke, NIU Jiefan, LI Xi, et al. Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions[J]. Water Resources Protection, 2021, 37(1): 28-35. (in Chinese))

- [4] ADAMOWSKI J, CHAN H F, PRASHER S O, et al. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada[J]. Water Resources Research, 2012, 48(1): 273-279.
- [5] 荣艳淑, 胡玉恒, 冯瑞瑞, 等. 广义相加模型在乌江夏季径流预报中的应用[J]. 河海大学学报(自然科学版), 2021, 49(2): 121-126. (RONG Yanshu, HU Yuheng, FENG Ruirui, et al. Application of generalized additive model in summer runoff forecasting of Wujiang Basin[J]. Journal of Hohai University(Natural Sciences), 2021, 49(2): 121-126. (in Chinese))
- [6] ZHANG Xiaoli, PENG Yong, ZHANG Chi, et al. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? some experiment evidences [J]. Journal of Hydrology, 2015, 530: 137-152.
- [7] 王红瑞, 魏豪杉, 胡立堂, 等. 基于遗传算法的 SVM-AR 改进模型与应用[J]. 河海大学学报(自然科学版), 2020, 48(6): 488-497. (WANG Hongrui, WEI Haoshan, HU Litang, et al. Improved model and application of SVM-AR based on genetic algorithm [J]. Journal of Hohai University(Natural Sciences), 2020, 48(6): 488-497. (in Chinese))
- [8] ZUO Ganggang, LUO Jungang, WANG Ni, et al. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting[J]. Journal of Hydrology, 2020, 585: 124776.
- [9] ZOUNEMAT-KERMANI M, BATELAAN O, FADAEI M, et al. Ensemble machine learning paradigms in hydrology: a review[J]. Journal of Hydrology, 2021, 598: 126266.
- [10] CHEN Shijun, WEI Qin, ZHU Yanmei, et al. Medium- and long-term runoff forecasting based on a random forest regression model [J]. Water Supply, 2020, 20 (8) : 3658-3664.
- [11] GUO Wendar, CHEN Weibo, YEH Senhai, et al. Prediction of river stage using multistep ahead machine learning techniques for a tidal river of Taiwan[J]. Water, 2021, 13(7): 920.
- [12] 陈纬楠, 胡志坚, 岳菁鹏, 等. 基于长短期记忆网络和 LightGBM 组合模型的短期负荷预测[J]. 电力系统自动化, 2021, 45(4): 91-97. (CHEN Weinan, HU Zhijian, YUE Jingpeng, et al. Short-term load prediction based on combined model of long short-term memory network and light gradient boosting machine[J]. Automation of Electric Power Systems, 2021, 45(4): 91-97. (in Chinese))
- [13] 王芳杰, 王福建, 王雨晨, 等. 基于 LightGBM 算法的公交行程时间预测[J]. 交通运输系统工程与信息, 2019, 19(2): 116-121. (WANG Fangjie, WANG Fujian, WANG Yuchen, et al. Bus travel time prediction based on light gradient boosting machine algorithm [J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(2): 116-121. (in Chinese))
- [14] 谷洪钦, 刘开磊, 刘玉环, 等. 综合多源不确定性的洪水概率预报试验[J]. 河海大学学报(自然科学版), 2021, 49(2): 99-104. (GU Hongqin, LIU Kailei, LIU Yuhuan, et al. Experiments on flood probability forecasting accounting for multi-source uncertainty [J]. Journal of Hohai University (Natural Sciences), 2021, 49(2): 99-104. (in Chinese))
- [15] 李亚飞, 王磊斌, 毛慧琴, 等. 使用岭回归对哈萨克斯坦月平均气温的统计降尺度研究[J]. 气候与环境研究, 2016, 21(5): 567-576. (LI Yafei, WANG Leibin, MAO Huiqin, et al. Statistical downscaling of monthly mean temperature for Kazakhstan using ridge regression [J]. Climatic and Environmental Research, 2016, 21(5): 567-576. (in Chinese))
- [16] 谢师, 黄跃飞, 李铁键, 等. LASSO 回归和支持向量回归耦合的中长期径流预报[J]. 应用基础与工程科学学报, 2018, 26(4): 709-722. (XIE Shuai, HUANG Yuefei, LI Tiejian, et al. Mid-long term runoff prediction based on a Lasso and SVR hybrid method [J]. Journal of Basic Science and Engineering, 2018, 26(4): 709-722. (in Chinese))
- [17] CHU Haibo, WEI Jiahua, QIU Jun. Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection [J]. Water, 2018, 10(10): 1486.
- [18] 严宝文, 程妍. 渭河干流月径流的时历相依性研究[J]. 人民黄河, 2013, 35(11): 25-27. (YAN Baowen, CHENG Yan. Time dependency of monthly hydrologic data series of the main channel of Weihe River [J]. Yellow River, 2013, 35(11): 25-27. (in Chinese))
- [19] 刘建伟, 崔立鹏, 刘泽宇, 等. 正则化稀疏模型[J]. 计算机学报, 2015, 38(7): 1307-1325. (LIU Jianwei, CUI Lipeng, LIU Zeyu, et al. Survey on the regularized sparse models [J]. Chinese Journal of Computers, 2015, 38(7): 1307-1325. (in Chinese))
- [20] ZOU Hui, HASTIE T. Regularization and variable selection via the elastic net [J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(2): 301-320.
- [21] 梁丽军, 刘子先, 王化强. 基于弹性网-SVM 的疾病诊断关键特征识别[J]. 计算机应用研究, 2015, 32(5): 1301-1304. (LIANG Lijun, LIU Zixian, WANG Huaqiang, et al. Identification of critical features for disease diagnosis based on elastic net and SVM [J]. Application Research of Computers, 2015, 32(5): 1301-1304. (in Chinese))
- [22] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述[J]. 计算机工程与应用, 2019, 55(24): 10-19. (LI Zhiqin, DU Jianqiang, NIE Bin, et al. Summary of feature selection

- methods [J]. Computer Engineering and Applications, 2019, 55(24):10-19. (in Chinese))
- [23] 于玲, 吴铁军. 集成学习: Boosting 算法综述 [J]. 模式识别与人工智能, 2004, 17(1):52-59. (YU Ling, WU Tiejun. Assemble learning: a survey of boosting algorithms [J]. Pattern Recognition and Artificial Intelligence, 2004, 17(1):52-59. (in Chinese))
- [24] 王秀杰, 练继建, 费守明. 基于小波消噪的混沌神经网络径流预报模型 [J]. 水力发电学报, 2008, 27(5):37-40. (WANG Xiujie, LIAN Jijian, FEI Shouming. The chaotic neural network model of runoff forecast based on wavelet de-noising [J]. Journal of Hydroelectric Engineering, 2008, 27(5):37-40. (in Chinese))
- [25] 赵雪花, 陈旭, 袁旭琦. 基于 EMD 的数据驱动模型在径流预测中的应用 [J]. 系统工程, 2014, 32(9):150-154. (ZHAO Xuehua, CHEN Xu, YUAN Xuqi. Application of data-driven model based on empirical mode decomposition for runoff forecasting [J]. Systems Engineering, 2014, 32(9):150-154. (in Chinese))
- [26] 刘晓琼, 吴泽洲, 刘彦随, 等. 1960—2015 年青海三江源地区降水时空特征 [J]. 地理学报, 2019, 74(9):1803-1820. (LIU Xiaqiong, WU Zezhou, LIU Yansui, et al. Spatial-temporal characteristics of precipitation from 1960 to 2015 in the Three Rivers' Headstream Region, Qinghai, China [J]. Acta Geographica Sinica, 2019, 74(9):1803-1820. (in Chinese))
- [27] 张先起, 胡登奎, 刘斐. 基于 CEEMD-Elman 耦合模型的年降水量预测 [J]. 华北水利水电大学学报 (自然科学版), 2019, 40(4):32-39. (ZHANG Xianqi, HU Dengkui, LIU Fei. Prediction of annual precipitation based on CEEMD-Elman coupling model [J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2019, 40(4):32-39. (in Chinese))
- [28] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3):531-544.
- [29] 李文武, 石强, 王凯, 等. 基于变分模态分解和深度门控网络的径流预测 [J]. 水力发电学报, 2020, 39(3):34-44. (LI Wenwu, SHI Qiang, WANG Kai, et al. Runoff prediction based on variational mode decomposition and deep gated network [J]. Journal of Hydroelectric Engineering, 2020, 39(3):34-44. (in Chinese))
- [30] 包苑村, 解建仓, 罗军刚. 基于 VMD-CNN-LSTM 模型的渭河流域月径流预测 [J]. 西安理工大学学报, 2021, 37(1):1-8. (BAO Yuancun, XIE Jiancang, LUO Jungang. Monthly runoff prediction of Weihe River Basin based on VMD-CNN-LSTM model [J]. Journal of Xi'an University of Technology, 2021, 37(1):1-8. (in Chinese))
- [31] 张淑清, 杨振宁, 张立国, 等. 基于弹性网降维及花授粉算法优化 BP 神经网络的短期电力负荷预测 [J]. 仪器仪表学报, 2019, 40(7):47-54. (ZHANG Shuqing, YANG Zhenning, ZHANG Liguang, et al. Short-term power load forecast based on dimension reduction by elastic network and flower pollination algorithm optimized BP neural network [J]. Chinese Journal of Scientific Instrument, 2019, 40(7):47-54. (in Chinese))
- [32] 陈纬楠, 胡志坚, 岳菁鹏, 等. 基于长短期记忆网络和 LightGBM 组合模型的短期负荷预测 [J]. 电力系统自动化, 2021, 45(4):91-97. (CHEN Weinan, HU Zhijian, YUE Jingpeng, et al. Short-term load prediction based on combined model of long short-term memory network and light gradient boosting machine [J]. Automation of Electric Power Systems, 2021, 45(4):91-97. (in Chinese))
- [33] 谢勇, 项薇, 季孟忠, 等. 基于 Xgboost 和 LightGBM 算法预测住房月租金的应用分析 [J]. 计算机应用与软件, 2019, 36(9):151-155. (XIE Yong, XIANG Wei, JI Mengzhong, et al. An application and analysis of forecast housing rental based on Xgboost and LightGBM algorithms [J]. Computer Applications and Software, 2019, 36(9):151-155. (in Chinese))
- [34] 王文圣, 丁晶, 邓育仁. 非参数统计方法在水文水资源中的应用与展望 [J]. 水科学进展, 1999, 10(4):458-463. (WANG Wensheng, DING Jing, DENG Yuren. Non-parametric statistic approach: application and prospect in hydrology and water resource [J]. Advances in Water Science, 1999, 10(4):458-463. (in Chinese))
- [35] 徐玉琴, 张扬, 戴志辉. 基于非参数核密度估计和 Copula 函数的配电网供电可靠性预测 [J]. 华北电力大学学报 (自然科学版), 2017, 44(6):14-19. (XU Yuqin, ZHANG Yang, DAI Zhihui. Power supply reliability prediction of distribution network based on nonparametric kernel density estimation and Copula function [J]. Journal of North China Electric Power University (Natural Science Edition), 2017, 44(6):14-19. (in Chinese))
- [36] 刘章君. 基于 Copula 函数的贝叶斯概率水文预报方法研究 [D]. 武汉: 武汉大学, 2017.
- [37] 杨肖丽, 郑巍斐, 林长清, 等. 基于统计降尺度和 SPI 的黄河流域干旱预测 [J]. 河海大学学报 (自然科学版), 2017, 45(5):377-383. (YANG Xiaoli, ZHENG Weifei, LIN Changqing, et al. Prediction of drought in the Yellow River based on statistical downscale study and SPI [J]. Journal of Hohai University (Natural Sciences), 2017, 45(5):377-383. (in Chinese))
- [38] 杨玉霞, 闫莉, 韩艳利, 等. 基于流域尺度的黄河水生态补偿机制 [J]. 水资源保护, 2020, 36(6):18-23. (YANG Yuxia, YAN Li, HAN Yanli, et al. Compensation mechanism of the Yellow River water ecology based on watershed scale [J]. Water Resources Protection, 2020, 36(6):18-23. (in Chinese))

- coupling model[J]. Water Resources Protection, 2021, 37 (2): 81-88. (in Chinese))
- [27] 罗赞,董增川,刘玉环,等. 基于 Copula 函数的连河通江湖泊防洪安全设计:以洪泽湖为例[J]. 湖泊科学, 2021, 33 (3): 879-892. (LUO Yun, DONG Zengchuan, LIU Yuhuan, et al. Safety design for rivers-connected lake flood control based on Copula function: a case study of Lake Hongze[J]. Journal of Lake Sciences, 2021, 33 (3): 879-892. (in Chinese))
- [28] LUO Yun, DONG Zengchuan, LIU Yuhuan, et al. Research on stage-divided water level prediction technology of rivers-connected lake based on machine learning: a case study of Hongze Lake, China[J]. Stochastic Environmental Research and Risk Assessment, 2021, 35 (10): 2049-2065.
- [29] LUO Yun, DONG Zengchuan, LIU Yuhuan, et al. Safety design for water-carrying lake flood control based on copula function: a case study of the Hongze Lake, China [J]. Journal of Hydrology, 2021, 597: 126188.
- [30] ABBOTT M B, IONESCU F. On the numerical computation of nearly horizontal flows [J]. Journal of Hydraulic Research, 1967, 5 (2): 97-117.
- [31] GAO Wei, LIU Ruxun, LI Hong. A hybrid vertex-centered finite volume/element method for viscous incompressible flows on non-staggered unstructured meshes [J]. Acta Mechanica Sinica, 2012, 28 (2): 324-334.
- [32] JACKIEWICZ Z, TRACOGNA S. A general class of two-step Runge-Kutta methods for ordinary differential equations [J]. SIAM Journal on Numerical Analysis, 1995, 32 (5): 1390-1427.
- [33] 周洁,董增川,朱振业,等. 基于 MIKE FLOOD 的洪泽湖周边滞洪区洪水演进模拟[J]. 南水北调与水利科技, 2017, 15 (5): 56-62. (ZHOU Jie, DONG Zengchuan, ZHU Zhenye, et al. Flood routing simulation of Hongze Lake detention basin based on MIKE FLOOD[J]. South-to-North Water Transfers and Water Science & Technology, 2017, 15 (5): 56-62. (in Chinese))
- [34] ZHU Zhenye, DONG Zengchuan, YANG Wanhong, et al. Optimal operation research of flood retarding in plain river network region[J]. Water, 2017, 9 (4): 280.
- [35] 王蔚,周春飞,侍翰生,等. 基于水资源利用的洪泽湖汛期水位控制运用研究[J]. 江苏科技信息, 2021, 38 (14): 65-70. (WANG Wei, ZHOU Chunfei, SHI Hansheng, et al. Study on water level control of Hongze Lake in flood season based on water resources utilization [J]. Jiangsu Science & Technology Information, 2021, 38 (14): 65-70. (in Chinese))

(收稿日期:2022-01-25 编辑:施业)

(上接第 135 页)

- [39] 鲍振鑫,严小林,王国庆,等. 1956—2016 年黄河流域河川径流演变规律[J]. 水资源与水工程学报, 2019, 30 (5): 52-57. (BAO Zhenxin, YAN Xiaolin, WANG Guoqing, et al. The trend in streamflow of the Yellow River Basin during 1956-2016[J]. Journal of Water Resources & Water Engineering, 2019, 30 (5): 52-57. (in Chinese))
- [40] 景亚平. 黄河中游区四条一级支流径流量时空变化规律及预测研究[D]. 杨凌:西北农林科技大学, 2012.
- [41] 王光辉. 近 60 年黄河干流径流泥沙变异性分析[D]. 北京:清华大学, 2019.
- [42] ZHAO Guangju, MU Xingmin, STREHMEL A, et al. Temporal variation of streamflow, sediment load and their relationship in the Yellow River Basin, China [J]. PLoS ONE, 2017, 9 (3): e91048.
- [43] 于海超,张扬,马金珠,等. 1969-2018 年黄河实测径流与天然径流的变化[J]. 水土保持通报, 2020, 40 (5): 1-7. (YU Haichao, ZHANG Yang, MA Jinzhu, et al. Variation of observed and natural runoff of Yellow River from 1969 to 2018 [J]. Bulletin of Soil and Water Conservation, 2020, 40 (5): 1-7. (in Chinese))
- [44] 黄巧玲,栗晓玲. 基于小波支持向量机耦合的月径流预测方法[J]. 水力发电学报, 2015, 34 (3): 1-7. (HUANG Qiaoling, SU Xiaoling. Wavelet support vector machine-coupling method for monthly runoff forecasting[J]. Journal of Hydroelectric Engineering, 2015, 34 (3): 1-7. (in Chinese))
- [45] 周聂,侯精明,陈光照,等. 基于机器学习的山洪灾害快速预报方法[J]. 水资源保护, 2022, 38 (2): 32-40. (ZHOU Nie, HOU Jingming, CHEN Guangzhao, et al. A rapid forecasting method for mountain flood disaster based on machine learning algorithms [J]. Water Resources Protection, 2022, 38 (2): 32-40. (in Chinese))
- [46] 谭江红,陈伟亮,王珊珊. 一种机器学习方法在湖北定时气温预报中的应用试验[J]. 气象科技进展, 2018, 8 (5): 46-50. (TAN Jianghong, CHEN Weiliang, WANG Shanshan. Using a machine learning method for temperature forecast in Hubei Province [J]. Advances in Meteorological Science and Technology, 2018, 8 (5): 46-50. (in Chinese))
- [47] 崔鸿雁,徐帅,张利锋,等. 机器学习中的特征选择方法研究及展望[J]. 北京邮电大学学报, 2018, 41 (1): 1-12. (CUI Hongyan, XU Shuai, ZHANG Lifeng, et al. The key techniques and future vision of feature selection in machine learning [J]. Journal of Beijing University of Posts and Telecommunications, 2018, 41 (1): 1-12. (in Chinese))
- [48] FENG Zhongkai, NIU Wenjing, TANG Zhengyang, et al. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization[J]. Journal of Hydrology, 2020, 583: 124627.

(收稿日期:2022-03-13 编辑:王芳)