

基于 BP 和 LSSVM 的径流 分频预测模型构建

张炳林¹, 李俊², 宋松柏¹

(1.西北农林科技大学水利与建筑工程学院,陕西 杨凌 712100; 2.海南大学生态与环境学院,海南 海口 570100)

摘要:针对径流序列具有较强的随机性和波动性特征,提出一种短期月径流预测混合模型 CEEMDAN-VMD-(BP, LSSVM)-LSSVM。首先利用自适应白噪声完整集成经验模态分解(CEEMDAN, complete ensemble empirical mode decomposition with adaptive noise)将径流序列分解为高频、中频和低频分量,再利用变分模态分解(VMD, variational mode decomposition)方法进一步分解高频分量,并根据样本熵对两次分解得到的子序列进行整合,采用麻雀搜索算法优化的反向传播神经网络(BP, back-propagation neural network)和最小二乘支持向量机(LSSVM, least square support vector machine)分别预测高频分量和中低频分量,最后将不同频率分量训练期的拟合值作为 LSSVM 的输入,进行二次预测得到最终的径流预测结果。将提出的模型应用于黑河流域莺落峡站和祁连站的月径流预测,验证期相关系数和纳什效率系数均达到 0.99 以上,对比其他 8 组对照模型,该模型具有更高的预测精度,可以应用于实际的短期月径流预测。

关键词:径流预测;经验模态分解;变分模态分解;样本熵;神经网络;支持向量机

中图分类号:S273.29; P333.2 **文献标志码:**A

Construction of runoff frequency division prediction model based on BP and LSSVM

ZHANG Binglin¹, LI Jun², SONG Songbai¹

(1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China;

2. School of Ecology and Environment, Hainan University, Haikou, Hainan 570100, China)

Abstract: This paper proposes a short-term monthly runoff prediction hybrid model, CEEMDAN-VMD-(BP, LSSVM)-LSSVM, to address the strong randomness and volatility characteristics of runoff sequences. Firstly, a complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was used to decompose the runoff sequence into high-frequency, mid-frequency, and low-frequency components. Then, the variational mode decomposition (VMD) method was further applied to decompose the high-frequency component, and the obtained sub-sequences from the two decompositions were integrated based on sample entropy. The back-propagation neural network (BP) optimized by the sparrow search algorithm and the least square support vector machine (LSSVM) were employed to predict the high-frequency and mid-low-frequency components, respectively. Finally, the fitting values of different frequency components during the training period were used as inputs for LSSVM to obtain the final runoff prediction. The proposed model was applied to the monthly runoff prediction at Yingluoxia and Qilian stations in the Heihe River Basin. The correlation coefficients and Nash efficiency coefficients during the verification period are both above 0.99. Compared with other eight models, this model demonstrates better prediction accuracy and can be applied to practical short-term monthly runoff prediction.

Keywords: runoff prediction; empirical mode decomposition; variational mode decomposition; sample entropy; neural network; support vector machine

径流是水资源系统的重要组成部分,准确的径流预测对于防洪减灾、农业灌溉、水库调度等具有重要意义^[1]。径流的形成受水文气象条件、下垫面状况以及人类活动等多方面因素的影响^[2],是一个复杂的非线性过程,使得对径流的预测难以达到较高的精度^[3]。

目前径流预测方法通常可分为两类,即物理驱动模型和数据驱动模型^[4]。物理驱动模型需要大量的气象资料、土壤和植被参数等作为模型输入^[5],通过模拟径流形成的物理过程来进行径流预测,计算繁复庞杂且限制因素较多。数据驱动模型则不需考虑径流的形成机制,直接对时间序列进行分析,从数学角度寻找输入变量与输出变量对应的线性或者非线性关系^[6],相对于物理驱动模型更具优势。因此数据驱动模型多用于预测,以满足高准确性和可靠性的要求^[7]。数据驱动模型包括时间序列模型和人工智能算法。时间序列模型在径流预测领域应用较为广泛,常见的时间序列模型主要有自回归模型、滑动平均模型、自回归滑动平均模型等。时间序列具有平稳性,是现有时间序列模型建立的前提条件,然而径流序列往往具有随机性和非平稳性,通过差分得到的平稳序列在一定程度上会丧失原序列的变化特征,降低预测精度^[8]。随着计算机技术的发展,人工智能算法不断涌现,主要包括灰色模型、反向传播神经网络(BP, back-propagation neural network)、支持向量机(SVM, support vector machine)等方法,因其具有较强的非线性映射能力和泛化能力^[9],能够更好地对径流序列进行拟合和预测。灰色模型结构简单,计算方便。郭华等^[10]综合灰色模型和神经网络对入库流量进行预测,提高了预测精度。灰色模型在数据较少的情况下具有较高的预测精度,但当数据离散度增大时,预测精度会降低。BP作为人工神经网络的代表,越来越多地应用于径流预测。崔东文^[11]比较了BP与径向基神经网络和广义回归神经网络在径流预测上的性能,认为BP更优。Wang等^[12]将BP应用于新安江模型,验证了该方法在实时误差修正上的有效性。尽管BP具有较强的自学习能力和自适应能力,但存在收敛速度较慢且易陷入局部最优的缺点^[13]。依据VC维理论和结构风险最小原理提出的SVM算法具有较高的预测精度。廖杰等^[14]通过对比SVM和门限自回归模型的预测精度,论证了SVM用于径流预测的可行性。SVM结构清晰,准确

度高,但控制参数难以直接确定。针对SVM参数选择困难的问题,崔东文等^[15]利用遗传算法和粒子群算法对SVM的惩罚因子 c 和核函数参数 σ^2 进行优化,并应用于年径流预测,认为SVM具有更好的泛化能力和稳定性。最小二乘支持向量机(LSSVM, least square support vector machine)采用结构风险最小化原则^[16],将SVM的不等式约束转化为等式约束,是SVM的一种扩展。Shabri等^[17]将LSSVM用于径流预测发现精度优于SVM模型。采用数据分解方法对径流进行预处理能够降低径流序列的波动性,有利于提高径流预测精度。张洪波等^[18]利用经验模态分解(EMD, empirical mode decomposition)与RBF神经网络耦合,对降水和径流序列进行预测,认为耦合模型能够提高预测精度。Parisouj等^[19]借助EMD与机器学习方法进行耦合,在年径流和月径流的模拟过程中表现出卓越的性能。在EMD的基础上发展起来的CEEMDAN通过在每个阶段添加有限次的自适应白噪声,可以有效解决EMD分解易出现模态混叠和端点效应的不足^[20],但分解得到的子序列中仍包含噪声且高频分量(特别是 IMF_1)仍具有较大的波动性。Huang等^[21]认为如果分解过程中一直存在非平稳非线性现象,就会导致出现不规则的模态分量,降低预测精度。因此有必要对CEEMDAN分解得到的高频分量进行二次分解^[22]。变分模态分解(VMD, variational mode decomposition)不同于EMD和CEEMDAN,是一种新的自适应时频分析方法,在频率搜索和分离上性能优异^[23-24],具有较好的鲁棒性,但VMD的分解个数难以直接确定,可能出现欠分解和过分解。

针对上述预报模型方法的不足,首先采用CEEMDAN对径流序列进行分解,并将分解后的各模态分量($IMF_1, IMF_2, \dots, IMF_n$)划分为高频分量、中频分量和低频分量,再利用VMD对高频分量进行分解($VMF_1, VMF_2, \dots, VMF_n$),提取其主要的波动特征。计算各分量的样本熵(SE, sample entropy),并利用K-means方法进行聚类整合,减小计算量。由于BP能够以任意精度逼近任意非线性函数,因此采用BP来预测径流序列的高频分量。LSSVM将SVM中的不等式约束转化为等式约束,将非线性问题转化为线性问题,有利于预测中低频分量。最后,高频、中频和低频的预测值再由LSSVM进行二次预测,得到径流序列最终的预测结果。

1 研究方法

1.1 自适应白噪声完整集成经验模态分解

CEEMDAN 是在 EMD 的基础上发展起来的,旨在解决 EMD 中模态混叠现象。CEEMDAN 通过在分解过程中添加有限次自适应白噪声,使得在集成次数较少时能够保证重构误差近似为 0,相较于集合经验模态分解方法在添加噪声后完备性较差具有明显优势^[25]。CEEMDAN 的具体步骤如下:

步骤一:首先对目标序列添加白噪声。

$$x^i(t) = x(t) + \varepsilon_0 \gamma^i(t) \quad (1)$$

式中, $x(t)$ 为时间序列; $x^i(t)$ 为添加噪声后的时间序列; $\varepsilon_0 \gamma^i(t)$ 为噪声分量; $i=1, 2, \dots, m$; m 为添加白噪声的个数。

步骤二:根据 EMD 方法计算每一个 $x^i(t)$ 的 $IMF_1^i(t)$, 并求平均值, 得到第一个模态分量 $IMF_1(t)$ 和残差分量 $r_1(t)$ 。

$$IMF_1(t) = \frac{1}{m} \sum_{i=1}^m IMF_1^i(t) \quad (2)$$

$$r_1(t) = x(t) - IMF_1(t) \quad (3)$$

步骤三:对残差分量 $r_1(t)$ 添加噪声并进行 EMD 分解, 得到第二个模态分量 $IMF_2(t)$ 和残差分量 $r_2(t)$ 。

$$IMF_2(t) = \frac{1}{m} \sum_{i=1}^m E[r_1(t) + \varepsilon_1 E(\gamma_1^i(t))] \quad (4)$$

重复上述步骤, 得到第 K 个模态分量 $IMF_k(t)$ 和残差分量 $r_k(t)$ 。

$$r_2(t) = r_1(t) - IMF_1(t) \quad (5)$$

$$IMF_{k-1}(t) = \frac{1}{m} \sum_{i=1}^m E[r_{k-1}(t) + \varepsilon_{k-1} E(\gamma_{k-1}^i(t))] \quad (6)$$

$$r_k(t) = r_{k-1}(t) - IMF_{k-1}(t) \quad (7)$$

式中, $E[\cdot]$ 为 EMD 方法得到的模态分量, ε_k 为信噪比, $\gamma_k^i(t)$ 为噪声分量。

CEEMDAN 的主要控制参数为噪声标准差、实现次数和允许最大筛选迭代次数。由于 IMF_1 属于不规则的高频分量, 直接预测精度较低, 因此采用 VMD 对 IMF_1 进一步分解。

1.2 变分模态分解

VMD 分解是一种基于变分问题的非递归分解方法, 具有严密的数学理论体系, 能够有效地分解非平稳非线性序列, 避免出现模态混叠现象^[26]。VMD 分解的主要过程如下:

步骤一:通过将固有模态函数 IMF_k 转化为有限带宽的调幅调频信号 u_k , 并以固有模态函数估计带

宽之和最小作为分解的约束条件。

$$f = \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[(\delta(t) + \frac{j}{nt}) \mu_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \quad (8)$$

且满足

$$\sum_k u_k(t) = x(t) \quad (9)$$

式中, $x(t)$ 为原始序列; k 为分解得到的模态分量个数; u_k 为原始序列分解后得到的模态分量; ω_k 为 u_k 的中心频率。

步骤二:引入拉格朗日乘子 $\lambda(t)$ 和二次惩罚项 α , 将约束优化转化为非约束优化。

步骤三:采用交替乘子方向法对模态函数 u_k 、中心频率 ω_k 以及拉格朗日乘子 $\lambda(t)$ 迭代更新, 得到最优值。

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i=k} \hat{u}_i + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (10)$$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (11)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{x}(\omega)) - \sum_k \hat{u}_k^{n+1}(\omega) \quad (12)$$

式中, $\hat{u}_k^{n+1}(\omega)$ 为剩余分量 $[\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega)]$ 的维纳滤波结果; $\hat{\omega}_k^{n+1}$ 为模态函数功率谱的中心频率; τ 为更新因子。

步骤四:重复步骤一、二、三, 当满足 $\sum_k (\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2) < f$ 时, 即认为收敛, 停止迭代, 其中 f 为阈值。

VMD 的核心参数为数据保真度约束参数 α 和分解个数 K , 本文中取 α 为径流序列的长度, K 为 CEEMDAN 分解 IMF_1 产生的模态函数的个数。

1.3 样本熵与 K-means 聚类

样本熵不依赖数据长度, 计算时具有较好的一致性, 可以有效度量时间序列的复杂程度, 时间序列越复杂, 样本熵值越大。样本熵的计算步骤见参考文献^[27]。

K-means 聚类是一种无监督聚类算法, 旨在依据多个聚类中心将数据划分为不同的聚类簇, 使得同一簇内方差最小, 不同簇间方差最大^[28]。K-means 聚类的计算步骤见参考文献^[29]。

1.4 不同频率分量的预测模型与优化算法

BP 和 LSSVM 是机器学习领域具有代表性的算

法。其中 BP 具有优秀的学习能力和非线性拟合能力,可以有效地对复杂序列进行分析处理,在此基础上选择 BP 对高频序列进行预测。LSSVM 具有结构简单、计算速度快、泛化性能好的特点,因此选用 LSSVM 预测中低频分量。鉴于 BP 和 LSSVM 在径流预测中应用广泛,具体的计算过程可参考文献 [30-31]。

麻雀搜索算法 (SSA, sparrow search algorithm) 是 2020 年提出的一种智能优化算法,基于模拟麻雀觅食和反捕食行为的过程进行参数寻优。与传统的优化算法相比,SSA 在全局优化问题中具有更强的搜索能力和更快的收敛速度^[32]。SSA 算法的原理和参数可参考文献 [33]。

1.5 基于 BP 和 LSSVM 月径流分频预测模型

本文采用“分解-预测-重构”的方法对月径流序列进行预测,预见期为一个月。首先采用 CEEMDAN 对径流序列分解,得到多个模态函数 $IMF_1, IMF_2, \dots, IMF_n$, 并将其划分为高频、中频和低频分量。由于 IMF_1 具有强烈的不规律性和波动性,直接预测会产生较大误差,因此选用 VMD 对 IMF_1 进一步分解,充分提取 IMF_1 的波动特征。进而将 BP 和 LSSVM 优势互补,采用 BP 预测高频分量, LSSVM 预测中低频分量。通过偏自相关函数 (PACF, partial autocorrelation function) 确定各频率分量的模型输入长度,采用不同频率分量在训练期的数据对 BP 和 LSSVM 模型进行率定,得到各频率分量在训

练期的模拟值和验证期的预测值,但不同频率分量的预测精度不同,直接相加可能会降低模型精度。为进一步提高模型精度,将各分量在训练期的拟合值作为 LSSVM 模型的输入,训练期的实测值作为 LSSVM 的输出,重新率定 LSSVM 模型,此过程即为根据各频率分量预测值准确性的不同赋予其不同的叠加权重。将各频率分量在验证期的预测值输入率定好的 LSSVM 模型进行预测,最终得到径流序列在验证期的预测值。BP 和 LSSVM 分频径流预测模型流程见图 1。

1.6 评价指标

本文选用均方根误差 (RMSE, root mean squared error)、平均绝对百分比误差 (MAPE, mean absolute percentage error)、相关系数 (R, correlation coefficient)、纳什效率系数 (NSE, nash-sutcliffe efficiency coefficient) 作为径流预测精度的评价指标。

$$RMSE = \sqrt{\frac{\sum_{t=1}^n [x(t) - \hat{x}(t)]^2}{n - 1}} \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x(t) - \hat{x}(t)}{x(t)} \right| \times 100\% \quad (14)$$

$$R = \frac{\sum_{t=1}^n [x(t) - \bar{x}(t)] [\hat{x}(t) - \bar{\hat{x}}(t)]}{\sqrt{\sum_{t=1}^n [x(t) - \bar{x}(t)]^2 \sum_{t=1}^n [\hat{x}(t) - \bar{\hat{x}}(t)]^2}} \quad (15)$$

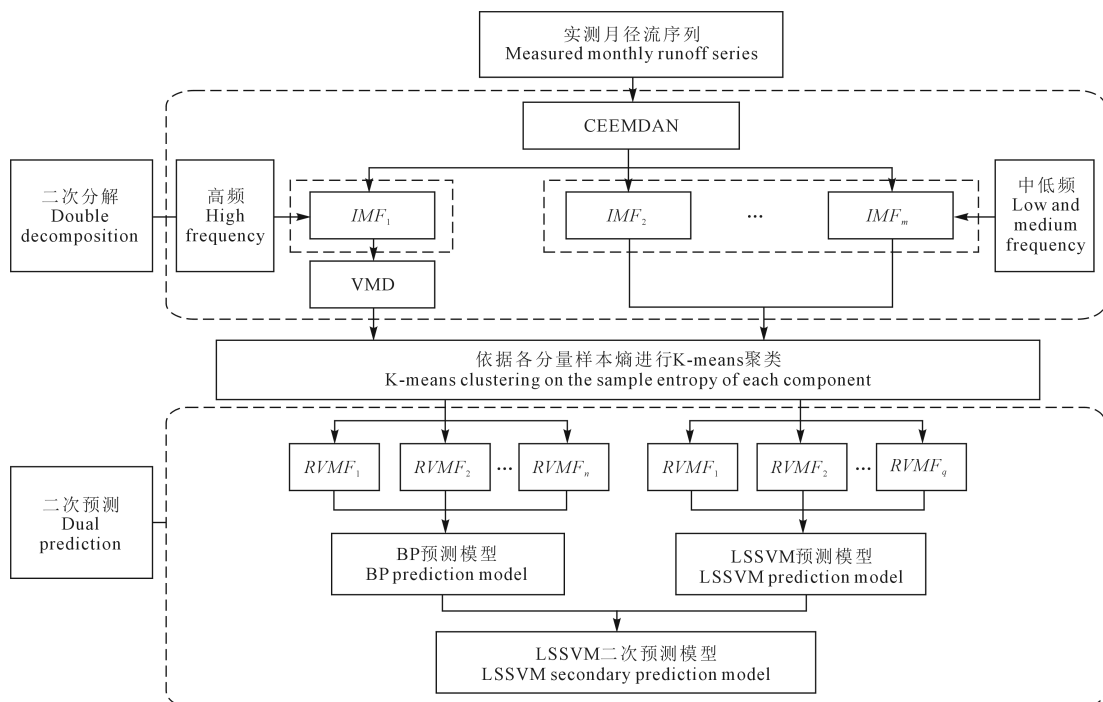


图 1 CEEMDAN-VMD-(BP, LSSVM)-LSSVM 预测模型流程图

Fig.1 CEEMDAN-VMD-(BP, LSSVM) - LSSVM prediction model flowchart

$$NSE = 1 - \frac{\sum_{t=1}^n [x(t) - \hat{x}(t)]^2}{\sum_{t=1}^n [x(t) - \bar{x}(t)]^2} \quad (16)$$

式中, $x(t)$ 为径流序列实测值, $\hat{x}(t)$ 为径流序列实测值的均值, n 为预测长度。 $RMSE$ 和 $MAPE$ 的取值范围是 $[0, +\infty)$, R 的取值范是 $[-1, 1]$, NSE 的取值范围是 $(-\infty, 1]$ 。在评价模型精度时 $RMSE$ 和 $MAPE$ 越小, R 和 NSE 越大则精度越高。

2 研究实例

2.1 研究区概况

黑河是我国西北第二大内陆河流, 准确的径流预测对黑河流域的水资源调度和农业灌溉具有重要意义, 流域的区域位置如图 2(见 262 页) 所示。本文选择莺落峡站月径流(1945 年 1 月—2009 年 12 月, 共计 780 个月) 和祁连站月径流(1968 年 1 月—2010 年 12 月, 共计 516 个月) 来构建模型, 以 5:1 比例确定训练期和验证期长度。莺落峡站和祁连站月径流量如图 3 所示, 月径流序列的统计特征见表 1。从各个统计参数可以看出, 训练集包含了

大量原始径流序列的统计特征, 训练集和验证集在统计特性上无明显差异, 因此可以认为训练集和验证集的划分较为合理。

2.2 分解径流序列

首先利用 CEEMDAN 分解莺落峡站和祁连站的原始径流序列, 得到的 IMF_1 和 IMF_{10} 分量如图 4 所示, $IMF_2 \sim IMF_8$ 省略。从图 4 可以看出 IMF_1 具有较强的随机性和波动性, 可以看作是高频分量, $IMF_2 \sim IMF_{10}$ 可以看作中低频分量。再利用 VMD 对高频分量 IMF_1 进行分解, 提取主要的频率特征。为简化计算结构, 减少累计误差, 计算 VMF 分量和 IMF 分量的 SE 值, 并利用 K-means 方法进行重构, 各分量的 SE 值见表 2, 重构后的各分量见图 5(见 262 页)。

2.3 确定模型的输入

准确的模型输入有利于提高模型的精度。通过计算各分量的 PACF, 并根据 PACF 的截尾性来确定各分量的最优输入长度。以莺落峡站 $RIMF_2$ 分量为例, 该分量的 PACF 图如图 6 所示, 当滞时大于 9 时, PACF 的值相对平稳且处于不显著范围内, 因此最优输入长度可选择为 9。按照同样的方法可确定两站不同分量的最优输入长度, 结果见表 3。

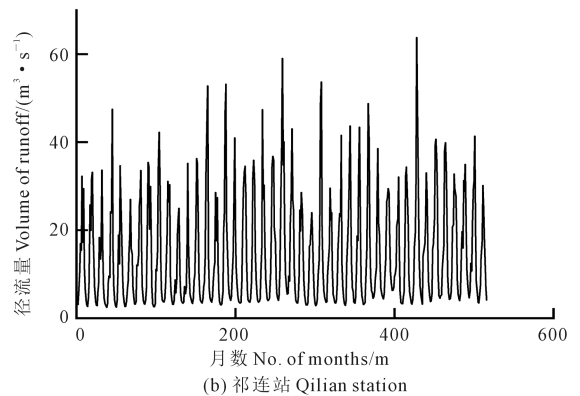
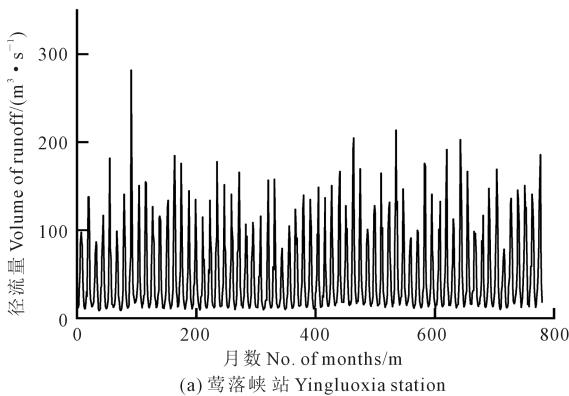


图 3 两站月径流过程

Fig.3 Monthly runoff process of two stations

表 1 莺落峡站和祁连站月径流统计参数

Table 1 Monthly runoff statistical parameters of Yingluoxia and Qilian stations

站点 Station	数据集 Data set	数据 长度 Length/m	最大值 Maximum /($m^3 \cdot s^{-1}$)	最小值 Minimum /($m^3 \cdot s^{-1}$)	平均值 Average /($m^3 \cdot s^{-1}$)	变差系数 Coefficient of variation	偏态系数 Coefficient of skew
莺落峡站 Yingluoxia station	原序列 Raw data	780	282.00	8.48	49.362	0.893	1.444
	训练集 Training set	648	282.00	8.48	48.855	0.904	1.486
	验证集 Validation set	132	203.00	9.62	51.741	0.844	1.262
祁连站 Qilian station	原序列 Raw data	516	63.84	2.42	14.453	0.827	1.213
	训练集 Training set	432	63.84	2.42	14.183	0.853	1.304
	验证集 Validation set	84	41.40	3.17	15.842	0.705	0.724

2.4 不同频率分量的预测

确定各分量的模型输入后,采用单隐含层的 BP 模型对归一化后高频分量 $RVMF_1 \sim RVMF_4$ 进行预测,隐含层的节点个数范围由式 (16) 与试算确定。

$$l = \sqrt{m + n} + a \quad (16)$$

式中, l 为隐含层节点个数, m 为输入层节点个数, n 为输出层节点个数, a 为 1~10 的整数。将各分量的预测值作为 LSSVM 模型的输入进行二次预测,得到径流序列的预测值。计算结果见表 4、5。

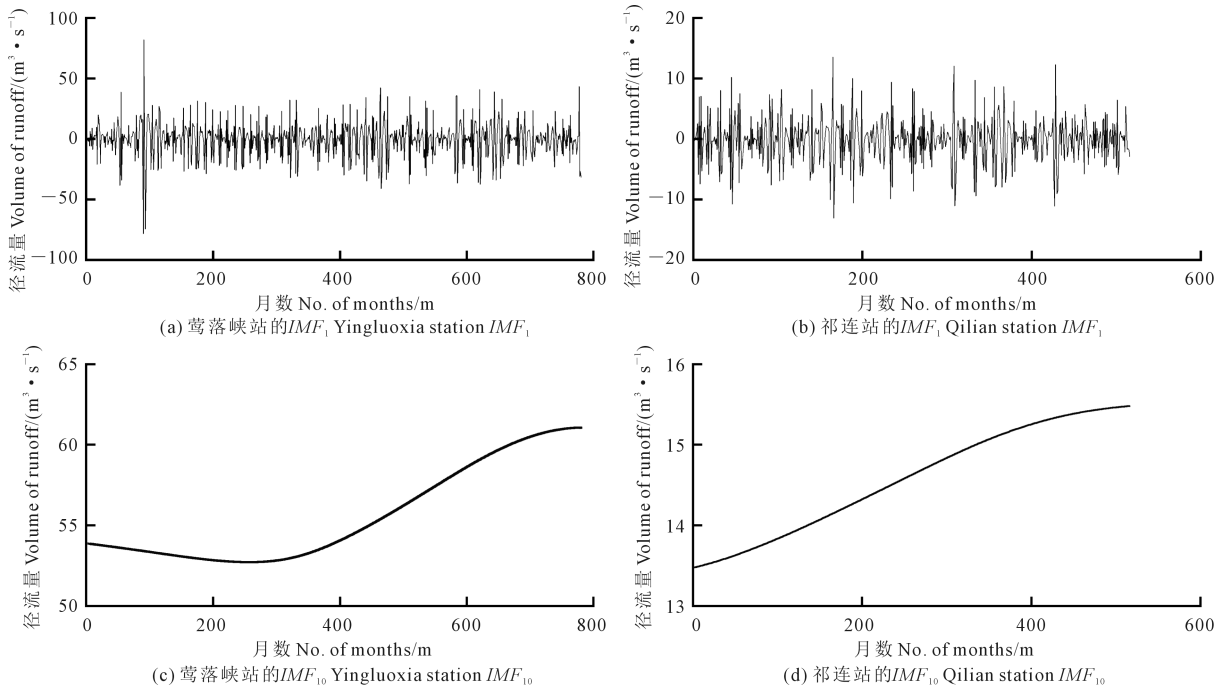


图 4 莺落峡站和祁连站 CEEMDAN 分解结果

Fig.4 CEEMDAN decomposition results at Yingluoxia and Qilian stations

表 2 子序列样本熵与模态分量重构

Table 2 Subsequence sample entropy and modal component reconstruction

站点 Station	模态分量 Modal component	样本熵值 Sample entropy	K-means	重构后的模态分量 Reconstructed modal components	站点 Station	模态分量 Modal component	样本熵值 Sample entropy	K-means	重构后的模态分量 Reconstructed modal components
莺落峡站 Yingluoxia station	VMF_1	0.481	VMF_1	$RVMF_1$	祁连站 Qilian station	VMF_1	0.511	VMF_1	
	VMF_2	0.533	VMF_5			VMF_2	0.629	VMF_6	$RVMF_1$
	VMF_3	0.544	VMF_2			VMF_3	0.355	VMF_8	
	VMF_4	0.279	VMF_3	$RVMF_2$		VMF_4	0.431	VMF_3	$RVMF_2$
	VMF_5	0.489	VMF_7			VMF_5	0.300	VMF_5	
	VMF_6	0.292	VMF_{10}			VMF_6	0.513	VMF_2	$RVMF_3$
	VMF_7	0.555	VMF_4	$RVMF_3$		VMF_7	0.476	VMF_9	
	VMF_8	0.646	VMF_6			VMF_8	0.566	VMF_7	
	VMF_9	0.621	VMF_8	$RVMF_4$		VMF_9	0.633	VMF_4	$RVMF_4$
	VMF_{10}	0.525	VMF_9			VMF_{10}	0.418	VMF_{10}	
	IMF_2	1.046	IMF_2	$RIMF_2$		IMF_2	1.110	IMF_2	$RIMF_2$
	IMF_3	0.659	IMF_3	$RIMF_3$		IMF_3	0.682	IMF_3	
	IMF_4	0.574	IMF_4	$RIMF_3$		IMF_4	0.646	IMF_4	$RIMF_3$
	IMF_5	0.548	IMF_5	$RIMF_4$		IMF_5	0.561	IMF_5	
	IMF_6	0.474	IMF_6			IMF_6	0.403	IMF_6	
	IMF_7	0.271	IMF_7	$RIMF_5$		IMF_7	0.297	IMF_7	$RIMF_4$
	IMF_8	0.089	IMF_8	$RIMF_5$		IMF_8	0.094	IMF_8	$RIMF_5$
	IMF_9	0.051	IMF_9	$RIMF_6$		IMF_9	0.016	IMF_9	$RIMF_6$
	IMF_{10}	0.006	IMF_{10}	$RIMF_7$		IMF_{10}	0.001	IMF_{10}	$RIMF_7$

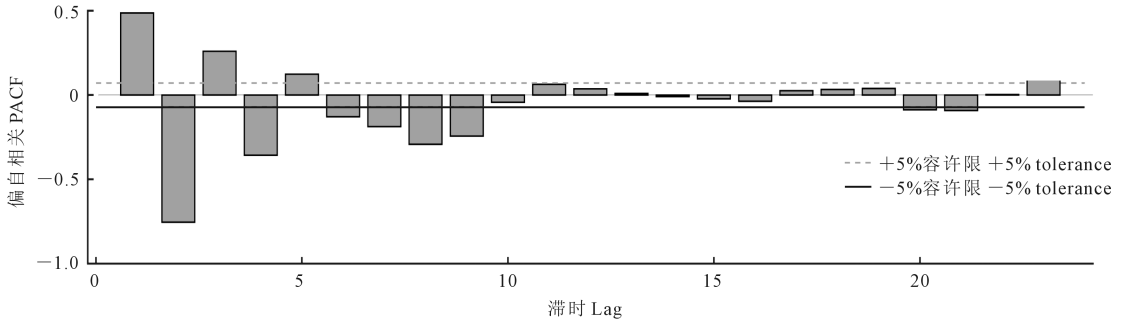


图 6 莺落峡站 $RIMF_2$ 分量的 PACF 图

Fig.6 PACF diagram of $RIMF_2$ component at Yingluoxia station

表 3 重构后子序列的最优输入长度

Table 3 Optimal input length of reconstructed subsequences

站点 Station	重构后的模态分量 Reconstructed modal component	最优输入维数 Optimal input dimension	站点 Station	重构后的模态分量 Reconstructed modal component	最优输入维数 Optimal input dimension
莺落峡站 Yingluoxia station	$RVMF_1$	5	祁连站 Qilian station	$RVMF_1$	10
	$RVMF_2$	5		$RVMF_2$	10
	$RVMF_3$	8		$RVMF_3$	12
	$RVMF_4$	12		$RVMF_4$	7
	$RIMF_2$	9		$RIMF_2$	9
	$RIMF_3$	8		$RIMF_3$	8
	$RIMF_4$	5		$RIMF_4$	5
	$RIMF_5$	10		$RIMF_5$	9
	$RIMF_6$	9	$RIMF_6$	2	
	$RIMF_7$	2	$RIMF_7$	2	

表 4 预测高频分量的 SSA-BP 模型结构

Table 4 SSA-BP model structure for predicting high frequency components

站点 Station	重构后的模态分量 Reconstructed modal component	最优模型结构 Optimal model structure	站点 Station	重构后的模态分量 Reconstructed modal component	最优模型结构 Optimal model structure
莺落峡站 Yingluoxia station	$RVMF_1$	5-10-1	祁连站 Qilian station	$RVMF_1$	10-5-1
	$RVMF_2$	5-12-1		$RVMF_2$	10-13-1
	$RVMF_3$	8-7-1		$RVMF_3$	12-12-1
	$RVMF_4$	12-4-1		$RVMF_4$	7-16-1

注:最优模型结构列数字代表三层 BP 模型的输入层、隐含层和输出层的神经元个数。

Note: The numbers in the optimal model structure column represent the amount of neurons in the input layer, hidden layer and output layer of the three-layer BP model.

表 5 预测中低频分量的 SSA-LSSVM 模型参数

Table 5 SSA-LSSVM model parameters for predicting low-frequency components in prediction

站点 Station	重构后的模态分量 Reconstructed modal component	最优参数 Optimal parameter		站点 Station	重构后的模态分量 Reconstructed modal component	最优参数 Optimal parameter	
		c	σ^2			c	σ^2
莺落峡站 Yingluoxia station	$RIMF_2$	41.173	8.993	祁连站 Qilian station	$RIMF_2$	38.847	10.750
	$RIMF_3$	51.258	10.229		$RIMF_3$	54.439	33.250
	$RIMF_4$	66.470	45.218		$RIMF_4$	61.405	98.374
	$RIMF_5$	76.879	1.599		$RIMF_5$	65.875	16.750
	$RIMF_6$	70.565	24.938		$RIMF_6$	94.712	4.280
	$RIMF_7$	199.036	27.897		$RIMF_7$	226.588	14.033

注: c 代表惩罚因子, σ^2 代表核函数参数

Note: c represents the penalty factor, σ^2 represents the kernel function parameter.

2.5 不同模型的对比分析

为验证本文所提出 CEEMDAN-VMD-(BP, LSSVM)-LSSVM 模型的优越性,设置 BP、LSSVM、EMD-(BP, LSSVM)-LSSVM、CEEMDAN-(BP, LSSVM)-LSSVM、VMD-(BP, LSSVM)-LSSVM、CEEMDAN-VMD-(BP)-LSSVM、CEEMDAN-VMD-(LSSVM)-LSSVM、CEEMDAN-VMD-(BP, LSSVM) 模型进行对比。各模型的预测结果见图 7、8。评价指标可以直观有效地评判模型的预测能力,莺落峡站和祁连站关于上述模型的评价指标值分别见表 6、7。

根据表 6 和表 7 可知,本文提出的模型 CEEMDAN-VMD-(BP, LSSVM)-LSSVM 的 *RMSE* 和 *MAPE* 均最小, *R* 和 *NSE* 均最大,相较于其他对比模型性能最优。对比 M1、M2 和其他模型可以看出,莺落峡站和祁连站采用数据分解的混合模型性能均明显优于直接预测的单一模型,这是因为分解技术将具有随机性和波动性的径流序列分解为若干相对平稳的子序列,有利于提高预测精度。通过对比 M3~M5 发现采用不同数据分解方法的混合模型在精度上也存在较大差异,两站利用 CEEMDAN 分

解和 VMD 分解的模型预测精度均明显高于 EMD 分解的模型精度,说明 CEEMDAN 和 VMD 均能有效避免模态混叠现象,且采用 VMD 分解的混合模型 M5 精度要高于采用 CEEMDAN 分解的 M4。而 M6 利用 VMD 对 CEEMDAN 分解后的高频分量进行二次分解,莺落峡站及祁连站的 *MAPE* 相对于 M3 的降低幅度分别为 16.7% 和 35.2%,证明了二次分解可以提取高频分量的主要特征,进一步提高模型的预测精度。对比 M6~M8 可以看出分频预测模型 M6 性能优于单频预测模型 M7、M8,主要因为分频预测模型能够对不同频率的分量匹配更适宜的预测模型,结合了不同模型的优点,因此分频预测模型相对单频预测模型更具优势。对比 M6 和 M9 可以看出,采用 LSSVM 根据各频率分量预测值进行二次预测的方法优于将各分量预测结果直接叠加。莺落峡站 M6 的 *RMSE* 和 *MAPE* 相对 M9 分别降低 22.7% 和 21.4%,祁连站 M6 的 *RMSE* 和 *MAPE* 相对 M9 分别降低 34.4% 和 36.4%,主要原因是 LSSVM 在整合的过程中根据各分量预测值准确性的不同赋予其不同的叠加权重,相较于平均赋权直接累加的精度更高。

表 6 莺落峡站不同径流预测模型结果比较

Table 6 Comparison of results of different runoff prediction models at Yingluoxia station

模型编号 Model number	模型 Model	<i>RMSE</i> /($\text{m}^3 \cdot \text{s}^{-1}$)	<i>MAPE</i> /%	<i>R</i>	<i>NSE</i>
M1	BP	18.319	22.584	0.911	0.826
M2	LSSVM	18.386	21.254	0.911	0.825
M3	EMD-(BP, LSSVM)-LSSVM	8.474	15.748	0.982	0.963
M4	CEEMDAN-(BP, LSSVM)-LSSVM	3.638	8.088	0.997	0.993
M5	VMD-(BP, LSSVM)-LSSVM	3.287	7.341	0.997	0.994
M6	CEEMDAN-VMD-(BP, LSSVM)-LSSVM	1.693	4.811	0.999	0.998
M7	CEEMDAN-VMD-(BP)-LSSVM	4.330	9.837	0.996	0.990
M8	CEEMDAN-VMD-(LSSVM)-LSSVM	5.862	11.656	0.992	0.982
M9	CEEMDAN-VMD-(BP, LSSVM)	2.191	6.118	0.999	0.997

表 7 祁连站不同径流预测模型结果比较

Table 7 Comparison of results of different runoff prediction models at Qilian station

模型编号 Model number	模型 Model	<i>RMSE</i> /($\text{m}^3 \cdot \text{s}^{-1}$)	<i>MAPE</i> /%	<i>R</i>	<i>NSE</i>
M1	BP	5.319	21.152	0.917	0.773
M2	LSSVM	5.497	23.377	0.888	0.758
M3	EMD-(BP, LSSVM)-LSSVM	4.545	15.011	0.968	0.864
M4	CEEMDAN-(BP, LSSVM)-LSSVM	2.475	12.775	0.982	0.951
M5	VMD-(BP, LSSVM)-LSSVM	1.102	8.914	0.995	0.990
M6	CEEMDAN-VMD-(BP, LSSVM)-LSSVM	0.773	5.776	0.998	0.995
M7	CEEMDAN-VMD-(BP)-LSSVM	1.267	10.295	0.994	0.987
M8	CEEMDAN-VMD-(LSSVM)-LSSVM	1.191	7.044	0.995	0.987
M9	CEEMDAN-VMD-(BP, LSSVM)	1.179	9.078	0.995	0.989

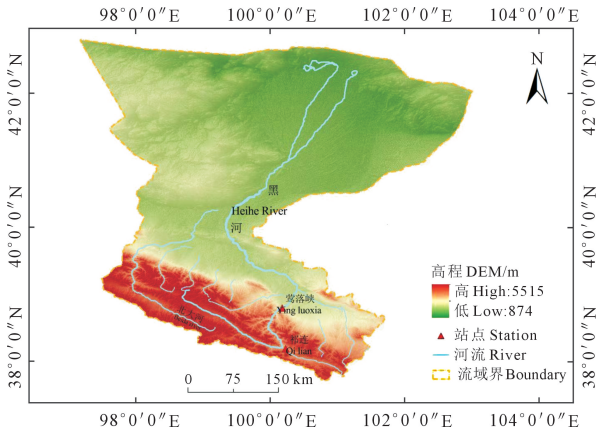


图 2 研究区域及水文测站位置图

Fig.2 Location of the study area and hydrological stations

3 结论与展望

本文将 CEEMDAN-VMD 分解与 BP 和 LSSVM 耦合,建立了月径流序列预测模型。利用多种模型对莺落峡站和祁连站月径流进行预测,优选出 CEEMDAN-VMD-(BP,LSSVM)-LSSVM 模型。研究表明:

(1)“分解-重构”预测方法可有效提升单一模型的预测精度。通过分解方法将具有较强波动特征的径流序列分解得到若干更为平稳的子序列,降低了预测难度。

(2)相较于单一分解方法(EMD, CEEMDAN, VMD),CEEMDAN-VMD 进一步分解高频分量,可以

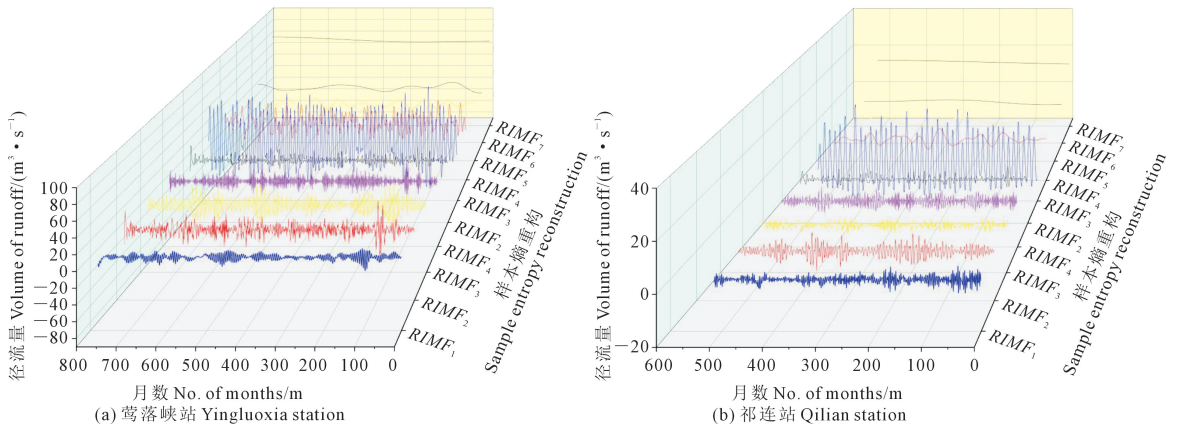


图 5 CEEMDAN-VMD 分解后的聚类整合结果

Fig.5 Cluster integration results after CEEMDAN-VMD decomposition

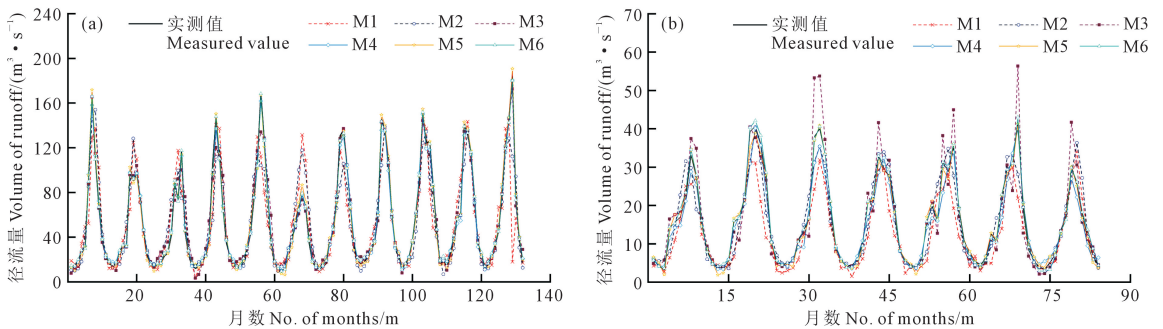


图 7 基于单一模型和混合模型的径流预测结果

Fig.7 Runoff prediction results based on single model and mixed model

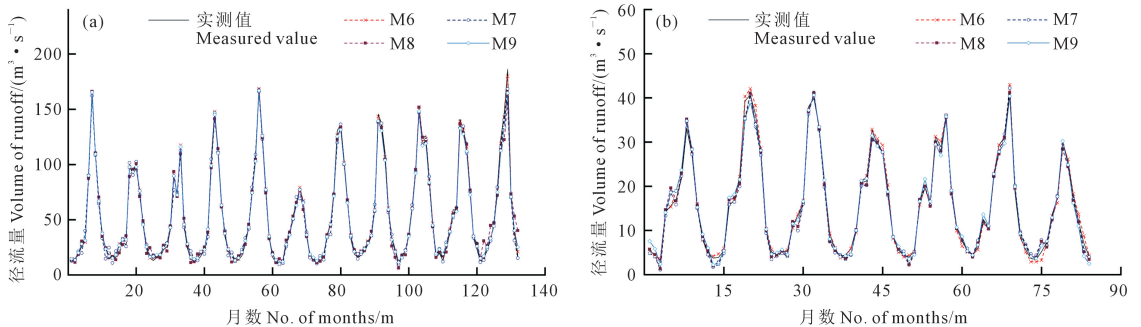


图 8 基于不同混合模型的径流预测结果

Fig.8 Runoff prediction results based on different mixed models

精细地揭示径流序列的变化特征,提高模型预测精度。

(3)将径流序列分解后的子序列根据频率进行划分,针对不同的频率分量采用BP和LSSVM进行预测,并将各分量在训练期得到的拟合值作为LSSVM输入进行二次预测可有效改善预测精度。

(4)月径流序列的波动性弱于日径流序列,后续应考虑该方法在日径流序列预报中的可行性,扩大模型的适用范围。

传统的统计方法如自回归和自回归滑动平均等模型只能抓住水文的稳态特征,但是在变化环境下,气候和下垫面变化导致的水文响应具有高度非线性特征,统计方法难以刻画这种非线性变化。而当前的基于机器学习和深度学习的非线性方法能够有效地描述水文序列非线性特征,能够显著提高预测精度,但同时也应注意到数理统计模型难以解释水文过程中的物理机制。因此在未来的研究中,应基于机器学习和深度学习最新的时间序列预测方法,并与物理模型相耦合,在保证预测精度的同时增强模型的可解释性。

参考文献:

[1] 王丽丽,李新,冉有华,等.基于奇异谱分析—灰狼优化—支持向量回归混合模型的黑河正义峡月径流预测[J].遥感技术与应用,2020,35(2):355-364.
WANG L L, LI X, RAN Y H, et al. Monthly runoff prediction of zhengyixia in the heihe river based on singular spectrum analysis-grey wolf optimizer-support vector regression hybrid model [J]. Remote Sensing Technology and Application, 2020, 35(2): 355-364.

[2] WANG Y H, SHI F Z, YAO P, et al. Assessing the evolution and attribution of watershed resilience in arid inland river basins, Northwest China[J]. Science of the Total Environment, 2024, 906: 167534.

[3] WANG X, WANG Y, YUAN P, et al. An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach [J]. Hydrological Sciences Journal, 2021, 66(9): 1488-1502.

[4] 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 H, HUANG S Z, MENG E H, et al. Runoff prediction based on multiple hybrid models[J]. Journal of Hydraulic Engineering, 2020, 51(1): 112-125.

[5] SIBTAIN M, LI X S, SALEEM S. A multivariate and multistage medium-and long-term streamflow prediction based on an ensemble of signal decomposition techniques with a deep learning network[J]. Advances in Meteorology, 2020, 2020: 8828664.

[6] HE X X, LUO J G, LI P, et al. A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting [J]. Water Resources Management, 2020, 34(2): 865-884.

[7] KUMAR S, TIWARI M K, CHATTERJEE C, et al. Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method[J]. Water Resources Management, 2015, 29(13): 4863-4883.

[8] ZHOU Z Q, DING Y B, ZHAO Y Y, et al. A new perspective for assessing hydro-meteorological drought relationships at large scale based on causality analysis [J]. Environmental Research Letters, 2023, 18(10): 104046.

[9] 岳兆新,艾萍,熊传圣,等.基于改进深度信念网络模型的中长期径流预测[J].水力发电学报,2020,39(10):33-46.
YUE Z X, AI P, XIONG C S, et al. Mid-and long-term runoff forecasting based on improved deep belief networks model[J]. Journal of Hydroelectric Engineering, 2020, 39(10): 33-46.

[10] 郭华,陈勇,马耀光.组合灰色预测模型在入库流量预测中的应用[J].干旱地区农业研究,2012,30(3):96-100.
GUO H, CHEN Y, MA Y G. Application of combined grey forecasting model in the prediction of incoming flow[J]. Agricultural Research in the Arid Areas, 2012, 30(3): 96-100.

[11] 崔东文.多隐层BP神经网络模型在径流预测中的应用[J].水文,2013,33(1):68-73.
CUI D W. Application of hidden multilayer BP neural network model in runoff prediction[J]. Journal of China Hydrology, 2013, 33(1): 68-73.

[12] WANG J, SHI P, JIANG P, et al. Application of BP neural network algorithm in traditional hydrological model for flood forecasting[J]. Water, 2017, 9(1): 48.

[13] HE S, SANG X F, YIN J X, et al. Short-term runoff prediction optimization method based on BGRU-BP and BLSTM-BP neural networks[J]. Water Resources Management, 2023, 37(2): 747-768.

[14] 廖杰,王文圣,李跃清,等.支持向量机及其在径流预测中的应用[J].四川大学学报(工程科学版),2006,38(6):24-28.
LIAO J, WANG W S, LI Y Q, et al. Support vector machine method and its application to prediction of runoff[J]. Journal of Sichuan University (Engineering Science Edition), 2006, 38(6): 24-28.

[15] 崔东文,金波.基于改进的回归支持向量机模型及其在年径流预测中的应用[J].水力发电学报,2015,34(2):7-14.
CUI D W, JIN B. Improved support vector machine regression model and its application to annual runoff forecasting[J]. Journal of Hydroelectric Engineering, 2015, 34(2): 7-14.

[16] LI P B, TIAN J, ZHOU Z Y, et al. Detection of internal wire broken in mining wire ropes based on WOA-VMD and PSO-LSSVM algorithms [J]. Axioms, 2023, 12(10): 995.

[17] SHABRI A, SUHARTONO. Streamflow forecasting using least-squares support vector machines [J]. Hydrological Sciences Journal, 2012, 57(7): 1275-1293.

[18] 张洪波,王斌,兰甜,等.基于经验模态分解的非平稳水文序列预测研究[J].水力发电学报,2015,34(12):42-53.
ZHANG H B, WANG B, LAN T, et al. A modified method for non-stationary hydrological time series forecasting based on empirical mode decomposition[J]. Journal of Hydroelectric Engineering, 2015, 34(12): 42-53.

[19] PARISOUJ P, JUN C, BATENI S M, et al. Machine learning models coupled with empirical mode decomposition for simulating monthly and yearly streamflows: a case study of three watersheds in Ontario, Canada [J]. Engineering Applications of Computational Fluid Mechanics, 2023, 17(1): 2242445.

[20] GUO S L, WEN Y H, ZHANG X Q, et al. Runoff prediction of lower Yellow River based on CEEMDAN-LSSVM-GM (1,1) model[J]. Scientific Reports, 2023, 13(1): 1511.

[21] HUANG S Z, CHANG J X, HUANG Q, et al. Monthly streamflow prediction using modified EMD-based support vector machine[J]. Journal of Hydrology, 2014, 511: 764-775.