

朱波,杨琼波,崔东文.基于小波包分解的CFO-LSTELM日供水量时间序列预测研究[J].人民珠江,2026,47(3):88-97.

基于小波包分解的CFO-LSTELM日供水量 时间序列预测研究

朱波¹,杨琼波²,崔东文^{3*}

(1.红河州泸西县惠民供水有限公司,云南红河 651400;2.云南省水文水资源局红河分局,云南红河 651400;3.云南省文山州水务局,云南文山 663000)

摘要:为提高日供水量时间序列预测精度,避免提前使用“未来信息”可能导致预测结果“失真”问题,提出基于小波包分解(Wavelet Packet Decomposition, WPD)的虫草优化(Caterpillar Fungus Optimizer, CFO)算法-最小二乘孪生极限学习机(Least Squares Twin Extreme Learning Machine, LSTELM)预测模型。首先将日供水量时间序列划分为训练集、验证集和预测集,利用1层WPD分别对“训练集+验证集”和预测集进行分解处理,分别得到1个低频分量和1个高频分量,以避免提前使用“未来信息”;其次基于训练集各分量构建LSTELM超参数优化目标函数,利用CFO对各目标函数进行LSTELM超参数寻优,并建立WPD-CFO-LSTELM日供水量时间序列预测模型,同时构建WPD-灰狼优化(Grey Wolf Optimizer, GWO)算法/鲸鱼优化算法(Whale Optimization Algorithm, WOA)-LSTELM模型、WPD-CFO-最小二乘极限学习机(Least Squares Extreme Learning Machine, LSELM)/孪生极限学习机(Twin Extreme Learning Machine, TELM)/极限学习机(Extreme Learning Machine, ELM)/最小二乘支持向量回归机(Least Squares Support Vector Regression, LSSVR)模型和WPD-LSTELM/LSELM/TELM/ELM/LSSVR模型共11种对比模型,以对比验证WPD-CFO-LSTELM模型性能;最后通过云南省红河州旧城水厂、金马水厂日供水量时间序列预测实例对12种模型进行验证。结果表明:①WPD-CFO-LSTELM模型对2个实例训练集、验证集、预测集拟合、预测的平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)小于等于0.34%、平均绝对误差(Mean Absolute Error, MAE)小于等于11.3 m³/d、均方根误差(Root Mean Square Error, RMSE)小于等于14.9 m³/d、决定系数(R²)大于等于0.9996,具有最好的模型性能和最小的拟合、预测误差;②CFO寻优精度显著优于GWO、WOA,元启发式算法寻优能力越好,日供水量拟合、预测精度越高;③LSTELM/LSELM/TELM/ELM/LSSVR超参数优劣对模型性能具有重要影响,通过CFO寻优超参数,可以显著提升日供水量预测精度;④提出的WPD-CFO-LSTELM模型及预测方法可为日供水量的精准预测提供了有效的解决途径。

关键词:日供水量预测;小波包分解;虫草优化算法;最小二乘孪生极限学习机;未来信息;超参数优化

中图分类号:TV697.4⁺;TP18 **文献标识码:**A **文章编号:**1001-9235(2026)03-0088-10

提高日供水量时间序列预测精度对于优化配置水资源、降低水厂运营成本、保障城乡供水安全等具有重要意义。近年来,高精度且高效的日供水量预测模型及方法越来越受到学者的关注,并在日

供水量预测研究中取得较好的预测效果,如长短时记忆(Long Short-Term Memory, LSTM)神经网络^[1-2]、回声状态网络(Echo State Network, ESN)^[3]、BP(Back Propagation)神经网络^[4]、XGBoost模型^[5]、

基金项目:国家重点研发计划项目(2021YFC300205-06);滇池湖泊生态系统云南省野外科学观测研究站(202305AM340008)

收稿日期:2025-08-13 **修回日期:**2025-09-10 **网络首发日期:**2025-09-30

作者简介:朱波(1984—),男,主要从事农村饮水安全工作。

通信作者:崔东文(1978—),男,正高级工程师,主要从事水资源管理保护及智能算法在水文水资源系统中的应用研究等工作。

联系作者:致信rmzj@pearlwater.gov.cn,由期刊转接。

多元线性回归模型^[6]等。然而,由于日供水量时间序列受天气突变(如极端降水和持续高温)、节假日效应(居民用水模式骤变)、突发公共事件(如疫情封控)、经济活动等多重复杂因素影响,通常呈现出显著的非线性、非平稳性特征,未经数据分解处理的传统时间序列预测模型难以充分捕捉日供水量序列内在的复杂动态规律与外部扰动综合效应,从而导致预测精度低、鲁棒性差。当前,基于“分解技术+预测器”或“分解技术+元启发式算法+预测器”架构的时间序列混合模型已在各行业领域及水文预测预报方面得到广泛应用,也尝试应用于日供水量时间序列预测研究,如牟天蔚等^[7]利用Symlets小波对新开河日供水量数据进行分解处理,并构建小波深度信念网络(SW-DBN)对新开河各分解分量进行预测和加和重构,结果表明该模型具有较好的预测效果,是一种有效的预测方法;李唯特等^[8]利用小波分解(Wavelet Decomposition, WD)处理日供水量时间序列,构建WD-人工蜂群优化(Artificial Bee Colony, ABC)算法-多变量门控循环单元(Multivariate Gated Recurrent Unit, MGRU)预测模型,通过某市日供水量预测实例对模型进行验证,结果表明该模型能有效捕捉多变量多尺度日供水量时序特征,从而提升模型日供水量预测精度;白云等^[9]引入小波变换(Wavelet Transform, WT)分解技术和随机森林回归(Random Forest, RF)方法,构建WT-RF预测模型,通过重庆某水厂日供水量预测实例验证了该模型较高的预测精度。

虽然基于“分解技术+预测器”或“分解技术+元启发式算法+预测器”架构的时间序列预测模型应用广泛,但在实际应用中,文献[7-9]将SW(Sliding Window)、WD等分解技术应用于整个时间序列,这意味着在训练阶段模型已提前使用“未来信息”,可能导致预测结果“虚高”和“失真”,从而失去应用价值。为提高日供水量时间序列预测精度,解决提前使用“未来信息”可能导致预测结果“失真”问题,本文创新性地提出小波包分解(Wavelet Packet Decomposition, WPD)-虫草优化(Caterpillar Fungus Optimization, CFO)算法-最小二乘孪生极限学习机

(Least Squares Twin Extreme Learning Machine, LSTELM)日供水量时间序列预测模型,并构建WPD-灰狼优化(GWO)算法-LSTELM、WPD-鲸鱼优化算法(WOA)-LSTELM对比模型,旨在对比验证CFO对LSTELM超参数的寻优效果;构建WPD-CFO-最小二乘极限学习机(LSELM)、WPD-CFO-孪生极限学习机(TELM)、WPD-CFO-极限学习机(ELM)、WPD-CFO-最小二乘支持向量回归机(LSSVR)对比模型,旨在对比验证在相同分解和优化情形下,LSTELM预测器的预测性能;构建WPD-LSTELM、WPD-LSELM、WPD-TELM、WPD-ELM、WPD-LSSVR对比模型,旨在对比验证预测器超参数优化与否对日供水量预测精度的影响。最后通过云南省红河州旧城水厂、金马水厂日供水量时间序列预测实例对12种模型进行验证,旨在检验在不提前使用“未来信息”情形下,WPD-CFO-LSTELM模型的优势与用于日供水量时间序列预测的可行性和普适性。

1 研究方法

1.1 小波包分解(WPD)

WPD是WD的扩展形式,相比于WD仅对低频部分进行递归分解,WPD能同时对高频和低频部分进行多层次分解,从而获得更精细的信号分解^[10-12]。WPD分解步骤包括:①选择小波基函数,根据信号特性选择合适的小波基,并确定分解层数;②逐层分解,将原始信号通过低通滤波器(得到低频子带)和高通滤波器(得到高频子带)进行第1层信号分解,对上一层所有子带分别进行同样的滤波操作,……,重复此过程,直到达到预设层数;③重构信号,根据需要选择特定子带进行重构。WPD分解公式见式(1):

$$\begin{cases} d_l^{j,2n} = \sum_k h_{k-2l} d_k^{j-1,n} \\ d_l^{j,2n+1} = \sum_k g_{k-2l} d_k^{j-1,n} \end{cases} \quad (1)$$

重构公式为:

$$\bar{d}_l^{j,n} = \sum_k (\bar{h}_{l-2k} d_k^{j+1,2n} + \bar{g}_{l-2k} d_k^{j+1,2n+1}) \quad (2)$$

式中各参数见文献[10-11]。

1.2 虫草优化(CFO)算法

CFO算法是Yang等^[13]于2025年受冬虫夏草独特的形成过程启发而提出的一种新型元启发式算法。该算法基于以下原则建模求解待优化问题:①在每一轮虫草生命周期繁殖过程中,每个子囊孢囊仅释放一个子囊孢子,每个子囊孢子仅能感染一只幼虫;②在寄生阶段,每个虫草在最多尝试两次寄生行为后可视为寄生成功;③寄生成功的幼虫总数保持恒定,同时设定寄生成功概率50%。

a)初始化。CFO采用式(3)初始化虫草空间位置。

$$X_{i,j} = \text{rand}(N, \text{Dim}) \times (\text{UB}_j - \text{LB}_j) + \text{LB}_j \quad (3)$$

式中: $X_{\text{CF},i}^{\text{so}}$ 为探索阶段第*i*个虫草更新位置; $X_{\text{CF},i}$ 、 $X_{\text{CF},i-1}$ 为第*i*个、第*i-1*个虫草位置(由于虫草个体适应度值按降序排列,因此第*i-1*个虫草性能优于第*i*个虫草性能); r_1 为 $[0, 1]$ 范围内随机数; X_{best} 为虫草

$$X_{\text{CF},i}^{\text{so}} = \begin{cases} X_{\text{CF},i} - r_1 * \{X_{\text{best}} - X_{\text{CF},i}\} + \alpha * \{X_{\text{best}} - X_{\text{CF},i}\}, & i = 1 \\ X_{\text{CF},i} - r_1 * \{X_{\text{CF},i-1} - X_{\text{CF},i}\} + \alpha * \{X_{\text{best}} - X_{\text{CF},i}\}, & i = 2, 3, \dots, N \end{cases} \quad (4)$$

$$X_{\text{CF},i}^{\text{so}} = \begin{cases} X_{\text{best},i} - r_3 * \{X_{\text{best}} - X_{\text{CF},i}\} + \beta * \{X_{\text{best}} - X_{\text{CF},i}\}, & i = 1 \\ X_{\text{best}} - r_3 * \{X_{\text{CF},i-1} - X_{\text{CF},i}\} + \beta * \{X_{\text{best}} - X_{\text{CF},i}\}, & i = 2, 3, \dots, N \end{cases} \quad (5)$$

式中: r_3 为 $[0, 1]$ 范围内随机数; $X_{\text{best},i}$ 为第*i*个虫草最佳位置; β 为虫草螺旋上升阶段搜索步长,描述为 $\beta = 2 \cos(\pi * r_4) * |(t/T)^{r_5}|$, r_4 为 $[0, 1]$ 范围内均匀分布的随机数, t 、 T 为当前迭代次数和最大迭代次数, r_5 为随机取值为1或2的随机整数。

c)寄生行为。虫草在最优寄生和再寄生之间表现出一种战略性选择。这种选择性寄生行为结合了全局搜索再寄生特性和局部搜索最优寄生特性,有利于CFO在搜索空间内进行更全面的探索。①再寄生行为。在再寄生行为阶段,虫草继续与同一幼虫保持寄生关系。但幼虫会被具有更优适应度的虫草重新寄生,以提高寄生的可靠性,其数学描述为:

$$X_{\text{CF},i}^{\text{pl}} = X_{\text{CF},i}^{\text{so}} + 3 * r_6 * \{\text{rand} * X_{\text{best}} - \text{rand} * X_{\text{CF},i}^{\text{so}}\}, \quad i = 1, 2, \dots, N \quad (6)$$

式中: $X_{\text{CF},i}^{\text{pl}}$ 为第*i*个虫草寄生阶段位置; r_6 为 $[0, 1]$ 范围内随机数。②最优寄生行为。与再寄生行为不同,最优寄生行为多发生在虫草与被寄生幼虫相遇

式中: $X_{i,j}$ 为第*i*个虫草第*j*维位置; UB_j 、 LB_j 为搜索空间上、下限值; rand 为 $[0, 1]$ 范围内随机数; N 为虫草数量; Dim 为问题维度。

b)探索行为。在该阶段,虫根据当前适应度值进行降序排序。随后,每个虫草个体将从波动推进算子、螺旋上升算子中随机选择一种算子进行位置更新,2种算子选择概率相等。①波动推进算子。该算子模拟虫草在土壤水平维度上的扩散行为,其搜索路径具有一定的非线性与周期性振荡特征,以增加对解空间的广域探索能力。该算子数学描述为:

最佳位置; α 为探索行为搜索系数,描述为 $\alpha = 2.5 * r_2 * |\cos(\pi * r_2)|$, r_2 为 $[0, 1]$ 范围内随机数。②螺旋上升算子。该算子模拟虫草在土壤垂直维度上的螺旋探索行为。该算子数学描述为:

的情况下。因此,在最优寄生行为阶段,虫草倾向于寻找最有利的生长条件,其数学描述为:

$$X_{\text{CF},i}^{\text{pl}} = X_{\text{best}} + 3 * \lambda * \{\text{rand} * X_{\text{best}} - \text{rand} * X_{\text{CF},i}^{\text{so}}\}, \quad i = 1, 2, \dots, N \quad (7)$$

式中: λ 为最优寄生阶段虫草搜索步长,描述为 $\lambda = r_7 * \{(t/T)^2 - 2 * (t/T) + 1\}$, r_7 为 $[0, 1]$ 范围内均匀分布的随机数。

1.3 最小二乘孪生极限学习机(LSTELM)

LSTELM是针对ELM普遍存在对噪声敏感和鲁棒性差等缺点而提出的一种创新性改进算法,与ELM不同,LSTELM采用2个非平行的超平面进行分类或预测,而不是单个超平面,同时采用松弛变量L2范数和等式约束设定,有效提高了分类或预测精度^[14-15]。LSTELM问题求解由式(8)给出:

$$\begin{cases} \min \frac{1}{2} \|Y - \varepsilon_1 e_1 - H\beta\|_2^2 + \frac{C_1}{2} \xi^T \xi \\ \text{s.t. } Y - H\beta = \varepsilon_1 e_1 \end{cases} \quad (8)$$

$$\begin{cases} \min \frac{1}{2} \|Y - \varepsilon_2 e_2 - H\beta_2\|_2^2 + \frac{C_2}{2} \eta^T \eta \\ \text{s.t. } Y - H\beta_2 = \varepsilon_2 e_2 \end{cases} \quad (9)$$

将等式约束替换到目标函数中,得到:

$$L(\beta_1, \xi) = \min \left[\frac{1}{2} \|Y - \varepsilon_1 e - H\beta_1\|_2^2 + \frac{C_1}{2} \|H\beta_1 + \varepsilon_1 e - Y\|_2^2 \right] \quad (10)$$

$$L(\beta_2, \eta) = \min \left[\frac{1}{2} \|Y - \varepsilon_2 e - H\beta_2\|_2^2 + \frac{C_2}{2} \|H\beta_2 + \varepsilon_2 e - Y\|_2^2 \right] \quad (11)$$

在 β_1, β_2 梯度为0时,可求解得到 β_1, β_2 的最优值为:

$$\beta_1 = C_1 (H^T H + C_1 H^T H + \omega I)^{-1} H^T (Y - \varepsilon_1 e) \quad (12)$$

$$\beta_2 = C_2 (H^T H + C_2 H^T H + \omega I)^{-1} H^T (Y - \varepsilon_2 e) \quad (13)$$

最终LSTELM模型分类或预测结果如下:

$$f(x) = \beta h(x) = \frac{\beta_1 + \beta_2}{2} h(x) \quad (14)$$

式中: Y 为输出; H 为隐含层; $\varepsilon_1, \varepsilon_2$ 为不敏感区域参数; β_1, β_2 为待求解的决策变量; e_1, e_2 为单位向量; C_1, C_2 为正权重因子; ξ, η 为松弛变量; $h(x)$ 为输入项的非线性映射; I 为单位矩阵; ω 为趋于无穷小的正数,一般取值 10^{-6} 。

2 实例应用

2.1 数据来源

本文日供水量时间序列来源于云南省红河州

旧城水厂和金马水厂。旧城、金马水厂实际日供水量序列为2023年1月1日至2025年6月30日,共912组。其中,旧城水厂设计规模4100 m³/d,供水人口约4.0万人;金马水厂设计规模6400 m³/d,供水人口约4.5万人。旧城、金马水厂日供水量变化曲线见图1。

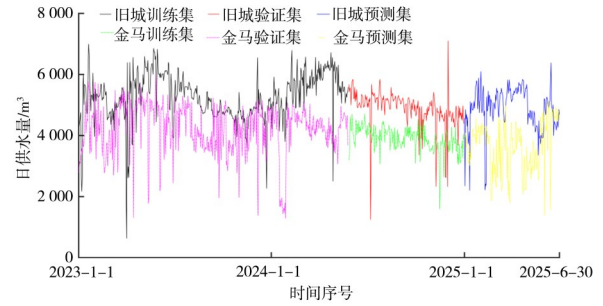


图1 日供水量变化曲线

Fig. 1 Daily water supply variation curve

2.2 模型构建

步骤一 WPD分解。为解决提前使用“未来信息”问题,利用1层WPD对旧城、金马水厂2023年1月1日至2014年12月31日共731组日供水量序列进行分解处理,分别得到1个低频分量和1个高频分量,选取各分量序列的70%作为训练集,剩余30%作为验证集;利用1层WPD对旧城、金马水厂2025年1月1日至2025年6月31日共181组日供水量序列进行分解处理,同样得到1个低频分量和1个高频分量,并将各分量作为预测集。旧城水厂WPD分解效果见图2,限于篇幅,金马水厂分解图略;预测流程见图3。

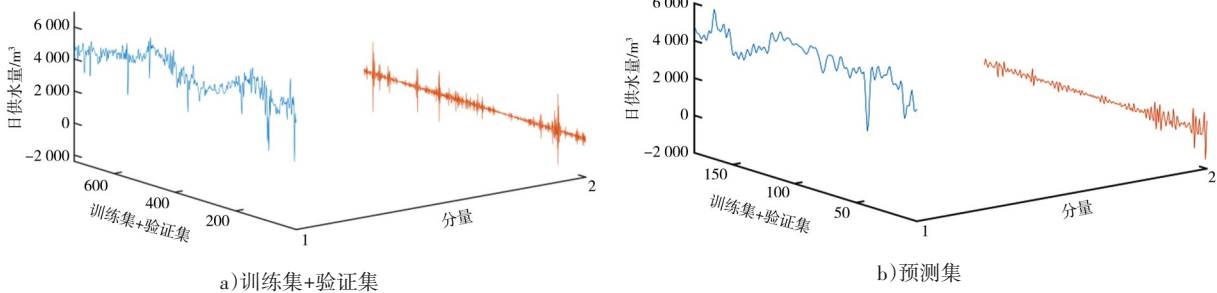


图2 旧城水厂日供水量序列分解3D效果

Fig. 2 3D rendering of daily water supply sequence decomposition for Old City Water Plant

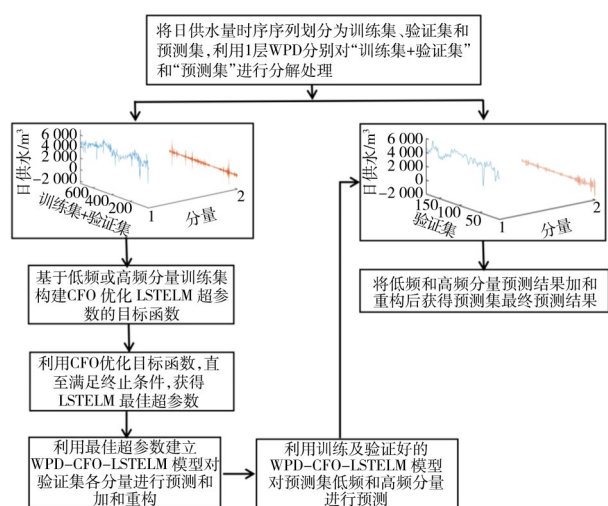


图3 预测流程

Fig. 3 Predictive flowchart

步骤二 嵌入维度 m 。采用改进假近邻法(Cao方法)确定旧城、金马水厂日供水量时间序列的 m 均为 28, 即分别利用前 28 日日供水量预测当日供水量。模型的输入 Q_{input} 、输出 Q_{output} 表示为:

$$Q_{input} = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_m \\ Q_2 & Q_3 & \cdots & Q_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{M-m} & Q_{M-m-1} & \cdots & Q_M \end{bmatrix}, Q_{output} = \begin{bmatrix} Q_{m+1} \\ Q_{m+2} \\ \vdots \\ Q_{M+1} \end{bmatrix} \quad (15)$$

式中: M 为样本数量; m 为嵌入维度。

步骤三 目标函数 F_{MSE} 。基于旧城、金马水厂各低频分量和高频分量训练集的拟合值与实际值构建均方误差 f_{MSE} 作为 CFO 优化 LSTELM 超参数的目标函数(CFO 优化 LSELM/TELM/ELM/LSSVR 超参数及 GWO、WOA 优化 LSTELM 超参数的目标函数及优化过程可参考实现), 见式(16):

$$\begin{cases} \min f(W, B) = \sqrt{\frac{1}{l} \sum_{i=1}^l (\hat{Q}_i - Q_i)^2} \\ \text{s.t. } W \in [W_{\min}, W_{\max}], B \in [B_{\min}, B_{\max}] \end{cases} \quad (16)$$

式中: \hat{Q}_i 为分解分量日供水量; Q_i 为分解分量日供水量拟合值; l 为训练集数量; W 为 LSTELM 输入层权重; B 为 LSTELM 隐藏层偏置。

步骤四 模型参数。12 种模型主要参数及寻优范围设置如下。

WPD-CFO/GWO/WOA-LSTELM 模型: 选取 WPD 小波基为“demy”函数, 设定分解层数为 1 层; 设置 CFO/GWO/WOA 最大迭代次数 200、虫草/灰狼/鲸鱼数量 30, 其他采用算法默认值; 设置 LSTELM 隐藏层神经元数 30、第一、第二非平行超平面正权重因子 C_1 、 C_2 均为 10^8 、不敏感区域参数 10^{-6} , 激活函数采用 sig; 输入层权重和隐藏层偏置搜索空间设置为 $[-1, 1]$ 。

WPD-CFO-LSELM 模型: WPD、CFO 参数同上; 设置 LSELM 隐藏层神经元数 30、正权重因子 10^8 、不敏感区域参数 10^{-6} , 激活函数采用 sig; 输入层权重和隐藏层偏置搜索空间设置为 $[-1, 1]$ 。

WPD-CFO-TELM 模型: WPD、CFO 参数同上; 设置 TELM 隐藏层神经元数 30、第一、第二非平行超平面正权重因子 C_1 、 C_2 均为 10^8 , 不敏感区域参数 10^{-6} 、激活函数 sigmoid; 输入层权重和隐藏层偏置搜索空间 $[-1, 1]$ 。

WPD-CFO-ELM 模型: WPD、CFO 参数同上; 设置 ELM 激活函数 sigmoid, 隐藏层神经元数 30, 输入层权值和隐含层偏置搜索空间 $[-1, 1]$ 。

WPD-CFO-LSSVR 模型: WPD、CFO 参数同上; 选取高斯函数作为 LSSVR 核函数; 设置 LSSVR 高斯函数宽度系数搜索空间 $[10^{-3}, 10^3]$ 、正则化参数搜索空间 $[10, 10^4]$ 。

WPD-LSTELM/LSELM/TELM/ELM/LSSVR 模型: 除 CFO 参数外, 其余参数设置均与上述对应模型相同。

步骤五 超参数寻优。①依据式(3)初始化虫草空间位置。依据式(16)评估虫草目标函数值, 根据最优目标函数值确定最佳虫草位置 X_{best} 。②执行虫草探索行为: 在 $[0, 1]$ 范围内生成随机数 r , 若 $r < 0.5$, 依据式(4)更新虫草位置; 否则依据式(5)更新虫草位置; 计算探索阶段所有虫草目标函数值, 确定探索阶段最佳虫草位置。③执行虫草寄生行为: 同样在 $[0, 1]$ 范围内生成随机数 r , 若 $r < 0.5$, 依据式(6)更新虫草位置; 否则依据式(7)更新虫草位置; 计算寄生阶段所有虫草目标函数值, 确定寄生阶段最佳虫草位置。④比较探索阶段和寄生阶段虫草

目标函数值,确定迄今为止最佳虫草位置 X_{best} 。⑤ 重复直至满足终止条件。输出 X_{best} ,即 LSTELM 最佳超参数。

步骤六 模型构建及评估指标。基于最佳超参数建立模型对各分解分量进行训练、预测和加和重构。并利用平均绝对百分比误差 MAPE(%)、平均绝对误差 MAE(m³/d)、均方根误差 RMSE(m³/d)和决定系数 R^2 对各模型进行评估。

2.3 实例验证

依据上述步骤建立 12 种模型对旧城、金马水厂日供水量各分解分量进行训练、预测与加和重构,预测结果见表 1;CFO/GWO/WOA 对旧城、金马水厂各低频、高频分量构建的目标函数寻优曲线见图 4;WPD-CFO-LSTELM 模型(模型 1)对旧城、金马水厂日供水量训练、验证和预测结果散点图见图 5;12 种模型对旧城、金马水厂预测集日供水量预测的绝对误差线箱图见图 6。

表 1 日供水量训练、验证和预测结果

Tab. 1 Daily water supply training, validation, and prediction results

站名	模型	序号	训练集				验证集				预测集			
			MAPE/ %	MAE/ (m ³ ·d ⁻¹)	RMSE/ (m ³ ·d ⁻¹)	R^2	MAPE/ %	MAE/ (m ³ ·d ⁻¹)	RMSE/ (m ³ ·d ⁻¹)	R^2	MAPE/ %	MAE/ (m ³ ·d ⁻¹)	RMSE/ (m ³ ·d ⁻¹)	R^2
旧城 水厂	WPD-CFO-LSTELM	1	0.16	7.5	10.3	0.999 8	0.19	8.9	12.3	0.999 6	0.17	8.0	10.9	0.999 7
	WPD-GWO-LSELM	2	1.03	49.3	69.4	0.991 2	1.35	62.6	96.8	0.973 0	1.27	59.9	79.4	0.985 3
	WPD-WOA-TELM	3	2.37	117.2	169.7	0.947 4	2.94	137.6	194.4	0.890 9	2.91	132.8	196.1	0.910 1
	WPD-CFO-LSELM	4	0.20	9.9	13.8	0.999 7	0.23	10.6	14.7	0.999 4	0.22	10.2	14.2	0.999 5
	WPD-CFO-TELM	5	0.27	13.1	18.3	0.999 4	0.32	14.6	20.0	0.998 8	0.29	13.5	18.8	0.999 2
	WPD-CFO-ELM	6	0.38	18.9	26.5	0.998 7	0.43	20.2	26.8	0.997 9	0.35	15.9	22.4	0.998 8
	WPD-CFO-LSSVR	7	1.40	68.4	109.0	0.978 3	2.45	111.7	158.0	0.927 9	2.37	107.3	151.0	0.946 7
	WPD-LSTELM	8	2.85	140.8	194.1	0.931 2	3.98	176.4	258.7	0.806 8	4.17	190.3	265.6	0.835 0
	WPD-LSELM	9	3.02	145.2	206.6	0.922 0	4.48	202.3	282.6	0.769 5	4.29	192.9	288.1	0.805 9
	WPD-TELM	10	3.29	163.3	230.1	0.903 3	4.69	210.9	293.6	0.751 2	4.42	201.5	276.3	0.821 5
	WPD-ELM	11	3.47	173.0	240.6	0.894 2	4.58	204.1	305.1	0.731 2	4.84	224.2	300.4	0.788 9
	WPD-LSSVR	12	3.04	145.3	216.3	0.914 5	4.50	186.1	284.1	0.767 0	4.47	196.3	273.7	0.824 7
金马 水厂	WPD-CFO-LSTELM	1	0.23	8.9	11.9	0.999 8	0.13	4.9	6.3	0.999 7	0.34	11.3	14.9	0.999 7
	WPD-GWO-LSELM	2	1.47	56.1	73.9	0.991 7	0.83	31.6	43.0	0.987 9	2.76	92.4	119.7	0.980 9
	WPD-WOA-TELM	3	3.74	148.2	197.6	0.940 8	2.06	78.0	107.2	0.925 0	5.90	182.3	253.7	0.914 1
	WPD-CFO-LSELM	4	0.32	12.8	16.6	0.999 6	0.20	7.5	9.6	0.999 4	0.49	16.3	21.2	0.999 4
	WPD-CFO-TELM	5	0.46	17.6	22.5	0.999 2	0.27	10.3	13.6	0.998 8	0.68	22.1	29.3	0.998 9
	WPD-CFO-ELM	6	0.61	23.8	31.8	0.998 5	0.32	12.2	16.4	0.998 3	0.71	23.2	30.6	0.998 7
	WPD-CFO-LSSVR	7	2.45	97.0	130.8	0.974 1	1.89	72.0	98.3	0.936 8	6.67	202.1	262.2	0.908 2
	WPD-LSTELM	8	5.06	198.6	259.1	0.898 3	2.79	106.2	141.4	0.869 3	10.23	319.4	420.8	0.763 5
	WPD-LSELM	9	5.56	220.2	294.8	0.868 3	3.12	118.9	154.5	0.844 1	10.13	316.4	406.1	0.779 7
	WPD-TELM	10	6.10	241.4	319.9	0.844 9	3.48	132.4	171.2	0.808 5	10.88	343.0	433.8	0.748 7
	WPD-ELM	11	6.14	239.1	314.6	0.849 9	3.38	128.3	166.2	0.819 6	11.05	346.5	448.4	0.731 5
	WPD-LSSVR	12	4.98	188.3	250.4	0.905 0	3.33	123.3	167.8	0.816 0	12.34	355.5	464.0	0.712 5

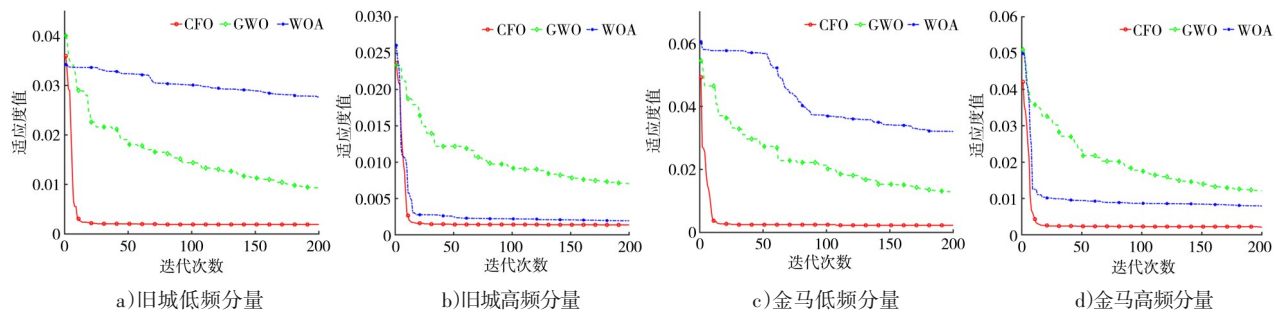


图4 CFO/GWO/WOA 寻优曲线

Fig. 4 Optimization curve of CFO/GWO/WOA

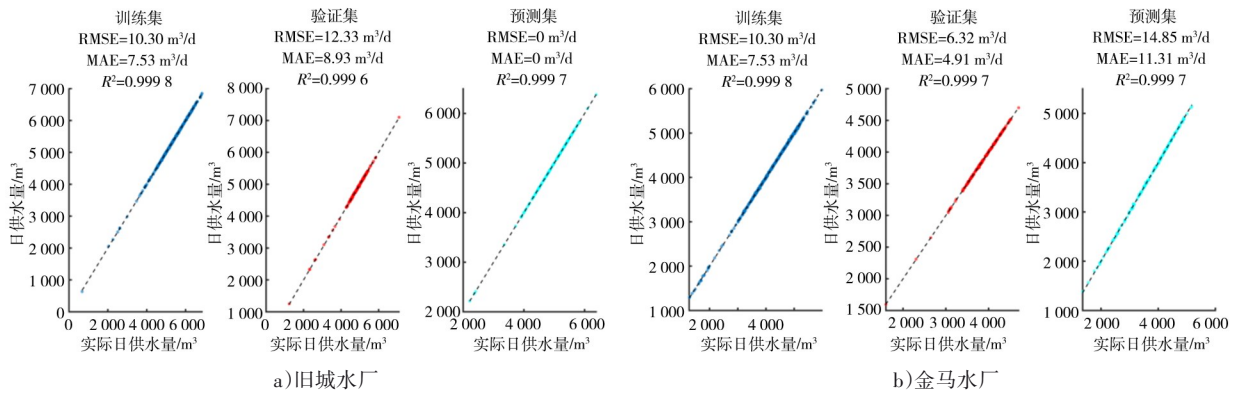


图5 模型1训练集、验证集、预测集散点图

Fig. 5 Model 1 training set, validation set, and prediction hub map

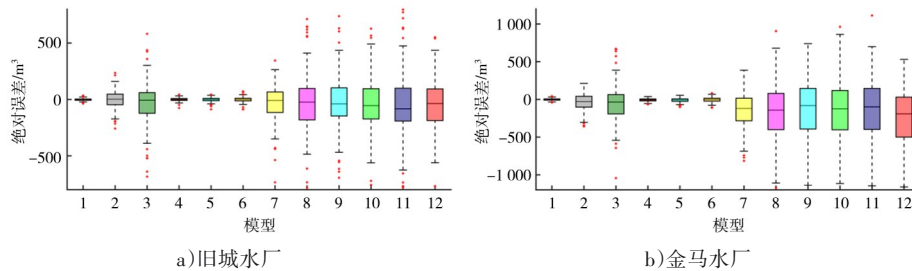


图6 日供水量预测集绝对误差线箱图

Fig. 6 Absolute error box chart of daily water supply prediction set

a)从表1可以看出,模型1对旧城水厂训练集拟合的MAPE、MAE、RMSE、 R^2 分别为0.16%、7.5 m³/d、10.3 m³/d、0.999 8,对验证集预测的MAPE、MAE、RMSE、 R^2 分别为0.19%、8.9 m³/d、12.3 m³/d、0.999 6,对预测集预测的MAPE、MAE、RMSE、 R^2 分别为0.17%、8.0 m³/d、10.9 m³/d、0.999 7;模型1对金马水厂训练集拟合的MAPE、MAE、RMSE、 R^2 分别为0.23%、8.9 m³/d、11.9 m³/d、0.999 8,对验证集预测的MAPE、MAE、RMSE、 R^2 分别为0.13%、4.9 m³/

d、6.3 m³/d、0.999 7,对预测集预测的MAPE、MAE、RMSE、 R^2 分别为0.34%、11.3 m³/d、14.9 m³/d、0.999 7,预测精度略优于模型4—6,优于模型2、3、7,远优于其他对比模型。可见,模型1不但具有更高的拟合、验证和预测精度,而且具有较好的普适性,将其用于水厂日供水量预测是可行和可靠的。

b)从图3可以得出,CFO寻优化精度显著最优,其寻优目标函数值远低于GWO、WOA,均能在50次迭代前获得全局最优,表明CFO在整体寻优精度上

具有绝对优势,具有优秀的全局搜索能力、收敛速度和求解精度,适合处理复杂优化问题。结合表1可以得出:元启发式算法寻优能力越好,超参数越佳,模型性能越好,日供水量拟合、验证和预测精度越高。

c)从图4可以看出,模型1对旧城、金马水厂训练集、验证集、预测集的拟合、验证和预测值与实际值具有强烈的线性正相关关系($R^2 \geq 0.9996$),数据点密集分布于 45° 对角线上,说明模型1的拟合度最高、预测误差最小。

d)从图6可以得出,模型1预测的绝对误差中位数均处于0值附近波动,其四分位距最窄、误差范围最小、误差分布最为集中,预测效果最好;模型4—6次之;模型8—12四分位距较宽、误差范围较大,整体预测效果较差。

e)从模型1与8、4与9、5与10、6与11、7与12对比可以看出,在相同数据处理和同一预测器情形下,采用CFO寻优LSTELM/LSELM/TELM/ELM/LSSVR超参数,可以显著提升模型性能和日供水量的拟合、验证和预测精度。

f)从模型1与模型4—6的拟合、验证和预测结果对比来看,在相同数据处理和优化情形下,LSTELM性能要优于LSELM/TELM/ELM/LSSVR。

3 结论

为解决时间序列预测提前使用“未来信息”问题,提出基于“训练集+验证集”、预测集两次分解的WPD-CFO-LSTELM时间序列预测模型,并构建11种对比分析模型,通过云南省红河州旧城、金马水厂日供水量时间序列预测实例进行验证。

a)在不提前使用“未来信息”情形下,WPD-CFO-LSTELM模型的拟合、验证和预测精度均优于其他对比模型,明显能增强日供水量特征表达能力,灵活地适应不同尺度的日供水量数据分布,具有较好的普适性和实际应用价值。

b)CFO对4个实例目标函数的寻优效果显著优于GWO、WOA,表现出优秀的全局搜索能力、收敛速度和求解精度,适合处理复杂优化问题。

c)通过CFO寻优LSTELM/LSELM/TELM/ELM/LSSVR超参数可以显著提升模型性能和日供水量预测精度;元启发式算法寻优能力越优,基于此建立的组合模型性能越好,日供水量拟合、验证和预测精度越高。

d)在相同数据分解和超参数优化情形下,基于双非平行超平面架构和最小二乘法改进的LSTELM性能要优于LSELM/TELM/ELM/LSSVR。

e)与传统“分解技术+元启发式算法+预测器”架构下的混合模型相比,本文WPD-CFO-LSTELM模型采用“训练集+验证集”“预测集”两段式分解,解决了传统模型提前使用“未来信息”可能导致预测结果“失真”的问题。

参考文献:

- [1] 周欣磊,顾海挺,刘晶,等. 基于集成学习与深度学习的日供水量预测方法[J]. 浙江大学学报(工学版),2023,57(6):1120-1127.
ZHOU X L, GU H T, LIU J, et al. Daily water supply prediction method based on integrated learning and deep learning [J]. Journal of Zhejiang University (Engineering Edition), 2023, 57(6): 1120-1127. (in Chinese)
- [2] 苟非洲,程玉婷. 基于长短期记忆网络的日供水量预测方法研究[J]. 中国给水排水,2019,35(17):79-83.
GOU F Z, CHENG Y T. Daily water supply forecasting method based on long short-term memory network [J]. China Water & Wastewater, 2019, 35(17): 79-83. (in Chinese)
- [3] 白云,严政杰,张晋,等. 基于多粒度泄漏积分回声状态网络的日供水量预测[J]. 中国给水排水,2023,39(9):50-56.
BAI Y, YAN Z J, ZHANG J, et al. Prediction of daily water supply based on multi-granularity leakage integral echo state network [J]. China Water & Wastewater, 2023, 39(9): 50-56. (in Chinese)
- [4] 张君君. 北票市城市自来水厂供水量预测分析[J]. 水利科学与寒区工程,2022,5(10):47-49.
ZHANG J J. Analysis on water supply forecast of urban waterworks in Beipiao City [J]. Water Conservancy Science and Cold Region Engineering, 2022, 5(10): 47-49. (in Chinese)
- [5] 贺波,马静,高赫余. 基于多粒度特征和XGBoost模型的城市日供水量预测[J]. 长江科学院院报,2020,37(5):43-49.
HE B, MA J, GAO H Y. Predicting urban daily water supply

- based on multi-granularity feature and XGboost integrated model [J]. *Journal of Changjiang River Scientific Research Institute*, 2020, 37(5): 43-49. (in Chinese)
- [6] 鞠佳伟, 宋良喜, 马晓明. 苏州市吴江区日供水量预测模型的建立与应用[J]. *中国给水排水*, 2017, 33(23): 141-144.
- JU J W, SONG L X, MA X M. Development and application of a daily water demand forecast model in Wujiang District [J]. *China Water & Wastewater*, 2017, 33(23): 141-144. (in Chinese)
- [7] 牟天蔚, 蒋白懿, 沈丹玉, 等. 深度学习框架对城市日供水量预测的研究[J]. *人民黄河*, 2018, 40(9): 58-62.
- MU T W, JIANG B Y, SHEN D Y, et al. Study on the prediction of deep learning framework for city daily water supply [J]. *Yellow River*, 2018, 40(9): 58-62. (in Chinese)
- [8] 李唯特, 陈宇, 于忠清. 基于小波分解的ABC-MGRU城市供水量预测模型[J]. *中国农村水利水电*, 2022(2): 184-189, 198.
- LI W T, CHEN Y, YU Z Q. ABC-MGRU model predicting urban water supply based on wavelet decomposition [J]. *China Rural Water and Hydropower*, 2022 (2): 184-189, 198. (in Chinese)
- [9] 白云, 陈国强. 基于尺度特征融合的随机森林日供水量预测模型[J]. *长江科学院院报*, 2022, 39(3): 33-37.
- BAI Y, CHEN G Q. Forecasting daily water supply based on random forest model with scale feature fusion [J]. *Journal of Changjiang River Scientific Research Institute*, 2022, 39(3): 33-37. (in Chinese)
- [10] 王永顺, 崔东文. 基于数据分解与NARX优化的滇池COD_{Mn}时间序列预测[J]. *人民珠江*, 2024, 45(7): 92-100.
- WANG Y S, CUI D W. Time series prediction of COD_{Mn} in Dianchi Lake based on data decomposition and NARX optimization [J]. *Pearl River*, 2024, 45 (7) : 92-100. (in Chinese)
- [11] 高雪梅, 崔东文. 基于WPT-ARO-DBN/WPT-EPO-DBN模型的月含沙量多步预测[J]. *人民珠江*, 2024, 45(3): 69-78.
- GAO X M, CUI D W. Multi-step prediction of monthly sediment concentration based on WPT-ARO-DBN/WPT-EPO-DBN model [J]. *Pearl River*, 2024, 45(3): 69-78. (in Chinese)
- [12] 李菊, 崔东文. 基于WPT-IDBO-RELM和WPT-IDMO-RELM模型的日径流预测[J]. *水利水电科技进展*, 2024, 44(6): 48-55, 85.
- LI J, CUI D W. Daily runoff prediction using WPT-IDBO-RELM and WPT-IDMO-RELM models [J]. *Advances in Science and Technology of Water Resources*, 2024, 44(6): 48-55, 85. (in Chinese)
- [13] YANG B, LIANG B X, ZHOU S, et al. A novel bio-inspired caterpillar fungus (*Ophiocordyceps sinensis*) optimizer for SOFC parameter identification via GRNN [J]. *Renewable Energy*, 2026, 256. DOI: 10.1016/j.renene.2025.123995.
- [14] 李旻, 孙大雁, 梁志峰, 等. 基于最小二乘孪生极限学习机的水电系统发电能力预测方法[J]. *水利水电技术(中英文)*, 2025, 56(8): 162-174.
- LI M, SUN D Y, LIANG Z F, et al. Prediction method for hydropower system generation capacity based on least squares twin extreme learning machine [J]. *Water Resources and Hydropower Engineering (in Chinese and English)*, 2025, 56(8): 162-174. (in Chinese)
- [15] ANAND P, BHARTI A, RASTOGI R. Time efficient variants of twin extreme learning machine [J]. *Intelligent Systems with Applications*, 2023, 17. DOI: 10.1016/j.iswa.2022.200169.

Research on CFO-LSTELM Daily Water Supply Time Series Prediction Based on Wavelet Packet Decomposition

ZHU Bo¹, YANG Qiongb², CUI Dongwen^{3*}

(1. Huimin Water Supply Co., Ltd., Luxi County, Honghe Prefecture, Honghe 651400, China; 2. Honghe Branch of Yunnan Provincial Hydrological and Water Resources Bureau, Honghe 651400, China; 3. Wenshan Water Bureau, Wenshan 663000, China)

Abstract: To improve the accuracy of daily water supply time series prediction and avoid the problem of distortion of prediction results caused by using future information in advance, a cordyceps optimization (CFO) algorithm based on wavelet packet decomposition (WPD)-least squares twin extreme learning machine (LSTELM) prediction model was proposed. Firstly, the daily water supply time series was divided into a training set, a validation set, and a prediction set. One layer of WPD was used to decompose the training set + validation set and the prediction set, respectively, so as to obtain one low-frequency component and one high-frequency component and avoid using future information in advance. Secondly, based on the components of the training set, an LSTELM hyperparameter optimization objective function was constructed. The CFO was used to optimize the LSTELM hyperparameters of each objective

function, and a WPD-CFO-LSTELM daily water supply time series prediction model was established. At the same time, 11 comparative models were constructed, including the WPD grey wolf optimization (GWO) algorithm/whale optimization algorithm (WOA)-LSTELM model, WPD-CFO-Least squares extreme learning machine (LSELM)/twin extreme learning machine (TELM)/extreme learning machine (ELM)/least squares support vector regression (LSSVR) model, and WPD-LSTELM/ELM/LSSVR model, so as to compare and verify the performance of the WPD-CFO-LSTELM model. Finally, 12 models were validated through time series prediction examples of daily water supply at the Old City Water Plant and Jinma Water Plant in Honghe Prefecture, Yunnan Province. The results show that: ① The WPD-CFO-LSTELM model has the best model performance and the smallest fitting and prediction errors for two instance training sets, validation sets, and prediction sets, with an average absolute percentage error (MAPE) of $\leq 0.34\%$, an average absolute error (MAE) of $\leq 11.3 \text{ m}^3/\text{d}$, a root mean square error (RMSE) of $\leq 14.9 \text{ m}^3/\text{d}$, and a coefficient of determination (R^2) of ≥ 0.9996 . ② The optimization accuracy of CFO is significantly better than GWO and WOA, and a better optimization ability of metaheuristic algorithm indicates a higher fitting and prediction accuracy of daily water supply. ③ The quality of LSTEL/LSELM/TELM/ELM/LSSVR hyperparameters has a significant impact on model performance. By optimizing hyperparameters through CFO, the accuracy of daily water supply prediction can be significantly improved. ④ The WPD-CFO-LSTELM model and prediction method proposed in this article provide an effective solution for the accurate prediction of daily water supply.

Keywords: daily water supply forecast; wavelet packet decomposition; caterpillar fungus optimization algorithm; least squares twin extreme learning machine; future information; hyperparameter optimization

(责任编辑:向飞)