

DOI: 10.20040/j.cnki.1000-7709.2023.20230042

联合 ICEEMDAN 与多种智能优化算法的 径流预测及对比研究

毛钦男^{a,b}, 刘招^{b,c}, 李杰^{a,b}, 王舒民^{a,b}, 张庭豪^{a,b}

(长安大学 a. 水利与环境学院; b. 旱区地下水文与生态效应教育部重点实验室; c. 水与发展研究院, 陕西 西安 710054)

摘要: 为提高径流预测的精度及可靠性, 引入 EMD 处理非稳态时间序列的优势, 建立改进自适应噪声完整集合经验模态分解(ICEEMDAN)和鲸鱼算法(WOA)优化的 BP 神经网络预测模型。以陕西省黑河金盆水库入库径流预测为例, 建立基于多种智能优化算法的模拟模型对水库入库径流进行预测, 同时分别选用降水、径流等不同时间序列历史资料作为输入因子, 对比在相同输入因子条件下 BP、WOA-BP、ICEEMDAN-BP、ICEEMDAN-WOA-BP 模型的预测能力。结果表明, 就输入序列而言, 以降水为输入因子的模型预测效果优于以径流为输入因子的模型; 就不同算法而言, ICEEMDAN-WOA-BP 模型具有较好的稳定性, 纳什系数可达 80%~90%, 预测精度更高。说明所提出的 ICEEMDAN-WOA-BP 模型可为河川径流预测、水库水文预报和流域水资源管理等提供技术支持。

关键词: 径流预测; ICEEMDAN; 鲸鱼算法; BP 神经网络

中图分类号: P338; TV121

文献标志码: A

文章编号: 1000-7709(2023)10-0023-04

1 引言

径流预测作为水文预报中的一项重要工作, 对合理配置水资源、洪涝和干旱预防等起到关键作用^[1]。近年来, 人工神经网络方法及其改进或优化的新方法^[2]较多, 其中基于多种算法和神经网络的组合模型备受关注^[3-4]。为进一步提高河川径流预测的精度, 针对 BP 神经网络在应用中存在易陷入局部极值及初始权值和阈值过于随机的问题, 引入 ICEEMDAN 分解, 以降低原始径流系列的复杂性和不稳定性, 同时结合 WOA 算法的优点, 对 BP 神经网络预测模型的输入因子、初始权值和阈值进行优化, 提出 ICEEMDAN-WOA-BP 神经网络模型并应用于径流预测实例中, 将所提模型的预测结果与 BP、WOA-BP、ICEEMDAN-WOA-BP 模型预测结果进行对比, 评估模型不同输入因子和不同组合所获得的月径流预报效果, 旨在为提高河川径流预测精度提供新的可行路径。

2 研究方法

2.1 ICEEMDAN 分解法

已有时间序列的处理方法自适应噪声的完整集合经验模态分解(CEEMDAN)虽然有抑制模态混叠现象、减少残留噪声和降低重构误差等优点, 但在其分解前期阶段仍包含大量噪声和信号的虚假分量。对此, COLOMINAS M A 等^[5]提出了改进自适应噪声的完整集合经验模态分解(ICEEMDAN)方法。ICEEMDAN 主要分解步骤如下。

步骤 1 将待分解序列信号 $x(t)$ 添加均值为 0、方差为 1 的白噪声, 构造分解序列 $x_i(t)$, 经过 EMD 得到 1 阶残差分量 $r_1(t)$, 设 $E_k(\cdot)$ 为 EMD 分解后的第 k 个分量, $M(\cdot)$ 为信号局部均值处理, α_k 为噪声标准差, δ_i 为添加的第 i 个白噪声。

$$x_i(t) = x(t) + \alpha_0 E_1(\delta_i) \quad (1)$$

$$E_1(x_i(t)) = x(t) - M(x_i(t)) \quad (2)$$

收稿日期: 2023-01-09, **修回日期:** 2023-02-08

基金项目: 陕西省自然科学基金联合基金项目(2021JLM-54); 国家自然科学基金项目(41273104); 陕西省自然科学基金基础研究计划项目(2012K12-03-06)

作者简介: 毛钦男(1998-), 男, 硕士研究生, 研究方向为水文学及水资源, E-mail: 18758860500@163.com

通讯作者: 刘招(1975-), 男, 博士、教授, 研究方向为水文学及水资源, E-mail: lz975@163.com

$$r_1(t) = \text{AVERAGE} [M(x_i(t))] \quad (3)$$

步骤 2 $k = 1$ 时, 计算第 1 个模态分量

$I_{\text{IMF}_1}(t)$:

$$I_{\text{IMF}_1}(t) = x(t) - r_1(t) \quad (4)$$

步骤 3 $k = 2$ 时, 计算 2 阶残差分量 $r_2(t)$

和第 2 个模态分量 $I_{\text{IMF}_2}(t)$:

$$I_{\text{IMF}_2}(t) = r_1(t) - r_2(t) \quad (5)$$

$$r_2(t) = \text{AVERAGE} [M(r_1(t) + \alpha_1 E_2(\delta_i))] \quad (6)$$

步骤 4 相应可得到 k 阶残差分量 $r_k(t)$ 和第 k 个模态分量 $I_{\text{IMF}_k}(t)$:

$$I_{\text{IMF}_k}(t) = r_{k-1}(t) - r_k(t) \quad (7)$$

$$r_k(t) = \text{AVERAGE} [M(r_{k-1}(t) + \alpha_{k-1} E_k(\delta_i))] \quad (8)$$

步骤 5 重复步骤 1~4, 得到所有残差分量和模态分量。

2.2 鲸鱼优化算法

鲸鱼优化算法(WOA)^[5]具有易操作、调整参数少和全局寻优能力强等特点, 近年来常被用于优化其他智能算法。WOA 算法主要过程如下。

步骤 1 包围目标阶段。定义与目标最接近的鲸鱼个体初始位置作为最佳位置, 其他鲸鱼朝最佳位置不断聚集更新自己的位置, 即:

$$\vec{S}_1 = |\vec{c}\vec{X}^*(t) - \vec{X}(t)| \quad (9)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}\vec{S}_1 \quad (10)$$

$$\vec{A} = 2a\vec{r} - \vec{a} \quad (11)$$

$$\vec{c} = 2\vec{r} \quad (12)$$

式中, \vec{X}^* 为当前最优位置; \vec{X} 为个体鲸鱼位置; t 为迭代次数; \vec{c} 、 \vec{A} 均为系数; \vec{S}_1 为鲸鱼个体与目标食物距离; \vec{a} 为控制参数(从 2 到 0 线性变化); \vec{r} 为 $[0, 1]$ 产生的随机向量。

步骤 2 气泡攻击阶段。鲸鱼群体进行收缩包围或螺旋游动行为, 此时 $|\vec{A}| < 1$ 。假定鲸鱼进行两种行为的概率都为 0.5, 该阶段的数学表达式为:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A}\vec{S}_1 & p < 0.5 \\ \vec{X}^*(t) + \vec{S}_2 e^{bl} \cos 2\pi l & p \geq 0.5 \end{cases} \quad (13)$$

其中 $\vec{S}_2 = |\vec{X}^*(t) - \vec{X}(t)|$

式中, \vec{S}_2 为鲸鱼个体到食物的距离; b 为螺旋形状参数; l 为 $[-1, 1]$ 产生的随机数。

步骤 3 搜索食物阶段。鲸鱼种群随机选择个体更新位置, 此时的 $|\vec{A}| > 1$, 是全局寻优过

程, 数学表达式为:

$$\vec{S}_{\text{rand}} = |\vec{c}\vec{X}_{\text{rand}}(t) - \vec{X}(t)| \quad (14)$$

$$\vec{X}(t+1) = \vec{X}_{\text{rand}}(t) - \vec{A}\vec{S}_{\text{rand}} \quad (15)$$

式中, $\vec{X}_{\text{rand}}(t)$ 为群体中随机鲸鱼个体位置; \vec{S}_{rand} 为随机鲸鱼个体到目标食物的距离。

2.3 BP 神经网络

BP 神经网络属于前馈神经网络^[6], 基础构架见图 1。其信号为正向传播, 误差经过逆向传播, 神经网络的权值和阈值逐渐调整更新, 使网络输出层结果接近期望值。

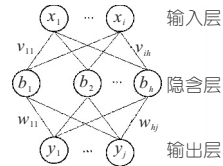


图 1 单隐含层 BP 神经网络结构

Fig. 1 Single hidden layer BP neural network structure

信号正向传播过程为:

$$\alpha_h = \sum x_i v_{ih} + \theta_h \quad (16)$$

$$\beta_j = \sum b_h w_{hj} + \theta_j \quad (17)$$

$$y_j = \sum b_i w_{ij} + \theta \quad (18)$$

式中, α_h 为输入层到隐含层过程; β_j 为隐含层到输出层过程; y_j 为第 j 个输出层结果; θ 为激活函数。

误差反向传播过程为:

$$w_{ij}(t+1) = lE_{y_k} + w_{ij}(t) \quad (19)$$

式中, $w_{ij}(t+1)$ 、 $w_{ij}(t)$ 分别为 $t+1$ 、 t 时刻隐含层到输出层的权值; l 为学习率; E_{y_k} 为第 j 个输出层的误差。

输入层到隐含层的权值更新类似, 达到误差设定结束时结束算法。

2.4 ICEEMDAN-WOA-BP 径流预测

根据算法及神经网络对径流系列的处理顺序, 将 ICEEMDAN-WOA-BP 径流预测模型分为四个过程(图 2): ① 先利用 ICEEMDAN 对径流系列进行分解, 分别得到 IMF 1、IMF 2、...、IMF N 若干个特征分量和一个趋势分量 RES。② 再通过给定的因子作为 BP 神经网络的输入层(研究所使用的输入因子为流域降水系列和前期径流系列)。③ 通过 WOA 算法优化 BP 神经网络的权值和阈值, 得到各分量的预测值。④ 合成各分量的预测值得到预测结果。

2.5 模型评价指标

选用平均绝对误差(M_{MAE})、平均绝对百分比误差(M_{MAPE})、均方根误差(R_{RMSE})和纳什效率系

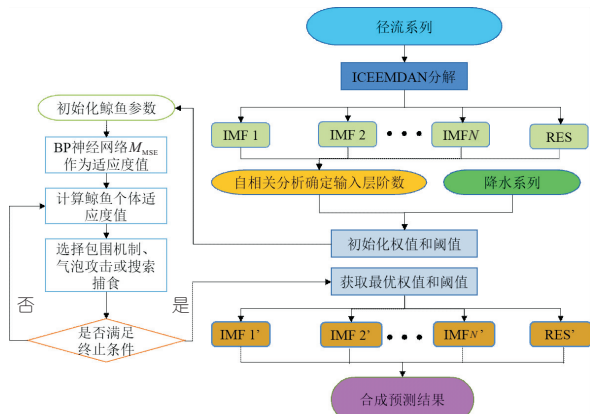


图 2 ICEEMDAN-WOA-BP 模型径流预测流程

Fig. 2 Flow of ICEEMDAN-WOA-BP model runoff prediction

数(N_{NSE})评估模型的预测结果。

3 算例研究

因不考虑产汇流物理学机制,基于数据驱动和智能算法的径流预测模型具有极强的普适性,与流域大小、地形和下垫面等因素无关。以陕西省黑河径流为例,对黑河金盆水库入库径流进行预测。金盆水库作为西安市重要水源地,肩负为西安地区年供水 $4.2 \times 10^8 \text{ m}^3$ 的重大任务,因此预测其来水情况极为重要。采用陈河水文站月平均入库流量系列及水库集水面积内东老君岭、康家垭子、板房子、钓鱼台、南天门等 15 个雨量站的月降雨量资料,二者系列长度均为 2005~2021 年。

3.1 数据处理与模型设置

径流和降水系列长度共 204 个月,选择前 168 个月作为预测模型的训练期,后 36 个月作为验证期。ICEEMDAN 分解参数中附加噪声标准差与原噪声标准差之比为 0.2,信号平均次数为 100,最大迭代数为 1 000。径流分解后各分量结果及原始径流系列见图 3。

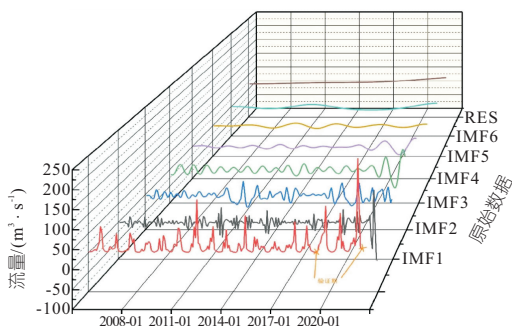


图 3 径流系列 ICEEMDAN 分解结果

Fig. 3 Decomposition results of runoff series ICEEMDAN

选择降水和前期径流分别作为神经网络的输入层进行预测。其中以上游集水区降水为输入层

的模型在选择输入层节点时,以站点个数作为输入层节点个数;以前期径流为输入层的模型考虑前期径流滞时,对径流量选取各自相关系数较高的阶数作为模型输入层节点数。

神经网络隐含层和输出层的激励函数分别选择 tansig 正切函数和 purelin 函数,训练算法为 Trainlm 函数。最大训练次数设定为 1 000,学习速率为 0.1,目标最小误差为 0.001。

对于模型隐含层节点数的选择以经验公式确定。分别以计算得到的隐含层节点数进行初步预测,根据不同隐含层所得结果的 M_{MSE} 大小进行隐含层节点数的筛选,选择误差最小的情况作为最终的隐含层节点数,径流各分量的节点数设置见表 1。

表 1 神经网络节点设置情况

Tab. 1 Setting of neural network nodes

设置 系列	降水为输入		前期径流为输入	
	输入层节点数	隐含层节点数	输入层节点数	隐含层节点数
原始数据	15	9	3	10
IMF1	15	7	4	7
IMF2	15	8	4	9
IMF3	15	6	8	8
IMF4	15	11	15	12
IMF5	15	13	5	7
IMF6	15	7	6	12
RES	15	10	7	9

3.2 结果与分析

为验证模型预测效果,在分别以降水和前期径流作为模型输入因子的前提下,建立 BP、WOA-BP、CEEMDAN-BP 和 CEEMDAN-WOA-BP 模型,对验证期径流进行预测。验证期不同输入因子经过不同组合模型得到的预测结果见图 4。由图 4 可知,就整体预测系列的拟合效果而言,以降水作为输入因子的各模型得到的曲线与实际值更接近,曲线总体趋势基本一致。而以前期径流为输入因子得到的曲线,在某些时段与实际值相差较大,尤其是系列结尾径流减少部分,曲线增减

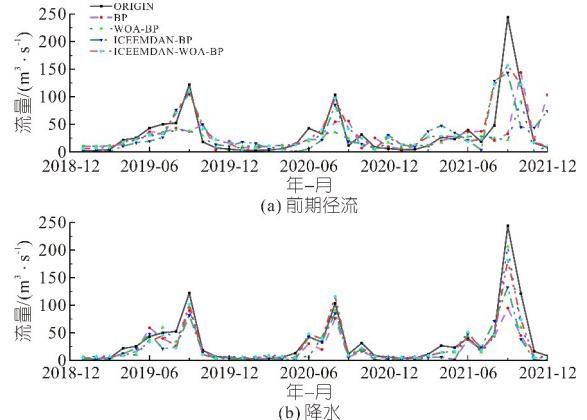


图 4 验证期不同输入因子各模型的预测结果

Fig. 4 Prediction results of various models with different input factors during the validation period

趋势明显不同。非汛期径流较少时的预测结果显示,以降水为输入因子时所得到的结果更加稳定,而以前期径流为输入因子得到的曲线具有明显的波动性。预测结果中三处峰值的平均相对误差分别为 10.01%(降水为输入)、17.12%(径流为输入)。

表 2 为不同输入因子和不同模型预测效果。由表 2 可看出,以降水为模型输入因子的条件下, BP、ICEEMDAN-BP 模型的 R_{RMSE} 、 M_{MAE} 明显大于 WOA-BP、ICEEMDAN-WOA-BP 模型; M_{MAPE} 结果显示 BP、ICEEMDAN-WOA-BP 模型的预测准确性相对更高;验证期 WOA-BP、ICEEMDAN-WOA-BP 模型的纳什系数明显优于其他两个模型,分别达到 88.76%、89.04%。

表 2 不同输入因子和不同模型预测效果

Tab.2 Prediction effects of different input factors and models

输入因子	模型	评价指标							
		R_{RMSE}		M_{MAE}		$M_{MAPE}/\%$		$N_{NSE}/\%$	
		训练期	验证期	训练期	验证期	训练期	验证期	训练期	验证期
降水	BP	8.52	30.23	5.70	13.88	71.06	39.33	81.84	58.91
	WOA-BP	9.62	15.81	6.37	9.94	57.48	54.41	76.82	88.76
	ICEEMDAN-BP	11.32	26.55	8.58	14.00	58.34	48.55	67.93	68.30
	ICEEMDAN-WOA-BP	8.72	15.61	5.63	8.92	51.69	38.80	80.96	89.04
径流	BP	18.92	44.21	12.69	22.14	97.71	136.26	15.38	12.11
	WOA-BP	17.66	42.77	11.86	19.06	85.47	94.25	21.91	17.75
	ICEEMDAN-BP	8.01	31.75	5.88	21.61	77.21	133.95	83.94	54.67
	ICEEMDAN-WOA-BP	7.84	22.14	5.75	12.88	76.85	90.37	84.63	77.96

在径流量为模型输入因子的条件下, ICEEMDAN-BP、ICEEMDAN-WOA-BP 模型的 R_{RMSE} 明显比 BP、WOA-BP 模型小; ICEEMDAN-WOA-BP 模型在训练期和验证期的 M_{MAE} 都表现良好,分别为 5.75、12.88 m^3/s ;从 M_{MAPE} 的结果来看,验证期预测指标值均低于训练期, BP、ICEEMDAN-BP 模型最差; ICEEMDAN-WOA-BP 模型的纳什系数较好,接近 80%,但 BP、WOA-BP 模型验证期的纳什系数仅为 12.11%、17.75%,其原因为当径流为输入时,模型对流量突增的敏感性远没有对降雨变化的强,导致峰值点的误差增大。

在同一模型组合下,降水为输入因子的各模型所得预测效果优于以径流为输入因子的模型。其中 ICEEMDAN-WOA-BP 模型在两种不同输入因子的条件下都表现较好,预测精度均高于其他三个模型。黑河流域径流变化受降水变化的影响更大,降雨径流的相关性也较高,使得以降水为输入因子的各模型的预测误差和预测精确度都优于以前期径流为输入因子的模型。前期径流经过 ICEEMDAN 分解,减小了序列不稳定带来的影响,使得以径流为输入因子的各模型中 ICEEM-

DAN-BP、ICEEMDAN-WOA-BP 模型的预测误差更小。2021 年 9 月径流高峰处以降雨为输入因子时的相对误差约为 20%,但以径流为输入因子的预测结果并不理想,相对误差达到 35%。根据历史资料可知该月份陕西地区暴雨频发,降雨量属 1961 年以来最大,洪水量极大。当降雨径流相关性较高且仅使用前期径流进行预测时,模型在此类极端条件下仍存在不足。但考虑到目前的预报多以降雨为输入因子,ICEEMDAN-WOA-BP 模型仍可以获得较好的预报效果。

4 结论

a. 本文基于 ICEEMDAN 分解法、WOA 优化算法和 BP 神经网络模型,考虑不同输入因子(降水和前期径流)分别构建不同算法组合模型(BP、WOA-BP、ICEEMDAN-BP 和 ICEEMDAN-WOA-BP),对黑河金盆水库入库径流系列进行预测。结果表明,以降水为输入因子的各模型预测误差整体小于以前期径流为输入因子的模型,四种模型的平均绝对误差分别减少了 8.26、9.12、7.61、3.96 m^3/s 。以前期径流为输入因子时,对径流系列进行 ICEEMDAN 分解,所得纳什系数可提高至 80%。

b. ICEEMDAN-WOA-BP 模型相比其他三种模型适应性更强,在不同的输入条件下具有更好的稳定性和泛化能力,满足在基础资料不完备时,基于不同输入因子的较高精度径流预测。

参考文献:

- [1] 聂敏,刘志辉,刘洋,等. 基于 PCA 和 BP 神经网络的径流预测[J]. 中国沙漠,2016,36(4):1144-1152.
- [2] BARGE J T, SHARIF H O. An ensemble empirical mode decomposition, self-organizing map, and linear genetic programming approach for forecasting river streamflow[J]. Water, 2016, 8: 247.
- [3] 周有荣,王凯. 改进鲸鱼算法优化混合核支持向量机在径流预测中的应用[J]. 中国农村水利水电, 2020(7): 50-53.
- [4] 王肖鑫,岑威钧,李昭辉,等. 基于人工电场算法优化的大型灌区径流预测模型研究[J]. 水资源与水工程学报, 2022, 33(4): 79-84.
- [5] COLOMINAS M A, SCHLOTTHAUER G, TORRES M E. Improved complete ensemble EMD: a suitable tool for biomedical signal processing[J]. Biomedical signal processing and control, 2014, 14: 19-29.
- [6] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in engineering software, 2016, 95: 51-67.

