

DOI:10.3880/j.issn.1004-6933.2023.05.016

基于 TVF-EMD、GRA 和 LightGBM 的日径流预测组合模型

王秀杰^{1,2}, 乔鸿飞^{1,2}, 曾勇红^{1,2}, 田福昌^{1,2}, 张 帅^{1,2}

(1. 天津大学水利工程仿真与安全国家重点实验室, 天津 300350; 2. 天津大学建筑工程学院, 天津 300350)

摘要:针对径流过程的非线性和非平稳性特点及预报精度低的问题,提出了结合时变滤波器的经验模态分解(TVF-EMD)、灰色关联度分析(GRA)和轻量级梯度提升机(LightGBM)的日径流预测组合模型。以黄河利津站和珠江高要站实测日径流序列为例,建立 TVF-EMD-GRA-LightGBM(TGL)组合模型,并将其预测结果与多种单一或组合预测模型的预测结果进行了对比分析。结果表明:TGL 组合模型高效且预测性能最佳,利津站和高要站日径流预测结果的纳什效率系数分别为 0.949 和 0.966,相关系数分别为 0.974 和 0.984,峰值流量预测误差分别小于 0.078 和 0.073。TGL 组合模型具有预测精度高、运行效率高、适用性强等优势,可用于日径流预测。

关键词:日径流预测;轻量级梯度提升机;TVF-EMD;灰色关联度分析

中图分类号:P338 **文献标志码:**A **文章编号:**1004-6933(2023)05-0135-08

TVF-EMD, GRA, and LightGBM combined daily runoff prediction model//WANG Xiujie^{1,2}, QIAO Hongfei^{1,2}, ZENG Yonghong^{1,2}, TIAN Fuchang^{1,2}, ZHANG Shuai^{1,2}(1. *State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China*; 2. *School of Civil Engineering, Tianjin University, Tianjin 300350, China*)

Abstract: In view of the nonlinear and non-stationary characteristics of runoff processes and low prediction accuracy, a daily runoff prediction model combining time-varying-filter-based empirical mode decomposition (TVF-EMD), grey relation analysis (GRA), and light gradient boosting machine (LightGBM) was proposed. Taking the measured daily runoff series of Lijin station on the Yellow River and Gaoyao station on the Pearl River as examples, the TVF-EMD-GRA-LightGBM (TGL) combined model was established, and its prediction results were compared with those of other single or combined models. The results show that the TGL combined model is efficient and has the best prediction performance, and the prediction results of the TGL combined model for Lijin and Gaoyao stations have the Nash-Sutcliffe efficiency coefficient of 0.95 and 0.97, respectively, the correlation coefficient of 0.97 and 0.98, respectively, and the peak flow prediction error less than 0.078 and 0.073, respectively. The TGL combined model has the advantages of high prediction accuracy, high operation efficiency, and strong applicability, providing a new way for daily runoff prediction and scientific regulation of water resources system.

Key words: daily runoff prediction; light gradient boosting machine; TVF-EMD; grey relation analysis

准确可靠的径流预测是防洪减灾、水资源配置、电力生产等流域管理和调度决策的关键依据,也是全球水文水资源领域研究的热点^[1]。受气候变化和人类活动等因素影响,径流过程更加复杂,预测精度面临更大挑战。随着经济社会的发展,对于径流预测的精确性和效率等方面的要求也越来越高。20 世纪中叶以来,人们已经提出了多种水文过程模拟和预测的方法^[2]。

径流预报模型主要分为过程驱动模型和数据驱动模型两大类。过程驱动模型基于水文、气象专业理论分析,与时空降水分布、气象条件和下垫面条件紧密结合^[3]。但由于过程驱动模型对数据的要求较高,以及产流和汇流过程的复杂性,很难建立精确的水文模型,从而限制了其应用^[4]。数据驱动模型主要分为水文统计模型和人工智能模型。水文统计模型基于概率论和数理统计法,根据大量历史水文

基金项目:国家重点研发计划项目(2022YFC3202501);天津大学自主创新基金项目(2022XHX-0013, 2022XSU-0019)

作者简介:王秀杰(1973—),女,副教授,博士,主要从事水文预报研究。E-mail:wangxiujie@tju.edu.cn

通信作者:田福昌(1989—),男,助理研究员,博士,主要从事智慧水利研究。E-mail:tju_tsgroup@163.com

资料研究预报对象和预报因子之间的潜在关系,或探索水文要素自身的演变规律建立模型^[5]。常用的水文统计模型包括自回归滑动平均模型^[6]、差分整合移动平均自回归模型^[7]及 Logistic 多元回归模型^[8]等,但这些模型在捕捉水文时间序列的非线性特征时存在一些缺陷,不适用于具有非线性和非平稳特征的径流预测^[9]。随着人工智能的发展,越来越多的学者将其引入径流预测的研究中,如反向传播(back propagation, BP)神经网络^[10]、支持向量机(support vector machine, SVM)^[11]和自适应神经模糊推理系统(adaptive neuro fuzzy inference system, ANFIS)^[12]等,但在应用过程中,这些方法存在一些缺点。如 BP 神经网络和 ANFIS 参数设置较多,易陷入局部最优;SVM 参数敏感性强,在大容量样本预报中表现不佳等^[13]。近年来,循环神经网络因其在处理复杂水文要素间高度非线性相互作用中表现出的优势而被应用于径流预测中^[14],长短期记忆(long short-term memory, LSTM)模型和门控循环单元(gate recurrent unit, GRU)模型是目前研究最广泛的两种深度学习神经网络。陶思铭等^[15]将 LSTM 模型应用于黄河源区唐乃亥站中长期径流预测,并与 BP 模型进行对比,结果表明 LSTM 模型预测精度较高。GRU 模型比 LSTM 模型构造简单,将遗忘门和输入门合成了一个更新门,在训练大量数据时可以节省时间。Zhao 等^[16]利用 GRU 模型进行月径流预测,并采用灰狼优化算法对其进行改进,模型达到了令人满意的预报效果。然而,LSTM 和 GRU 模型存在参数较多、计算量大和运行时间较长等缺陷^[17],且不善于存储信息,无法处理长期依赖关系^[18]。针对上述模型存在的缺点,集成学习模型,如轻量级梯度提升机(light gradient boosting machine, LightGBM)具有更多的优势,目前已广泛应用于金融、医学和风力发电等领域的研究中。LightGBM 基于决策树的梯度提升框架,具有准确率高、训练效率高、收敛速度快以及可以处理大规模数据等优点^[19],为高效并行运算和数据特征的有效提取提供了可靠的理论依据,具有高效准确预测日径流的应用潜力。但 LightGBM 对于强非线性时间序列数据的处理效果不理想,限制了其在径流预测中的应用。

基于时变滤波器的经验模态分解(time-varying filter-based empirical mode decomposition, TVF-EMD)方法有效解决了模态混叠和噪声干扰问题,将数据序列分为局部高频和低频分量,已广泛应用于数据的预处理。王文川等^[20]验证了 TVF-EMD 方法可以提高 LSTM 模型的预测性能和精度;赵磊娜等^[21]发现 TVF-EMD 方法与最小二乘支持向量机模

型结合可提供更好的交通流预测。

通过信号分解方法将原始序列分解为多个不同频率的子序列,可以消除其非平稳性,降低预测难度,但是过多的子序列会增加序列重构误差,进而影响模型的预测精度。将子序列合并为一个新的子序列是目前减小重构误差的有效途径。常用的重组方法有排列熵、模糊熵和 K 均值聚类等,本文引入灰色关联度分析(grey relation analysis, GRA)将与原始序列灰色关联系数相同的子序列进行重新组合以提高整体预测精度。

目前研究主要采用变分模态分解、互补集合经验模态分解和极点对称模态分解(extreme-point symmetric mode decomposition, ESMD)等方法构建组合模型应用于径流预测,而 TVF-EMD 方法在径流预测领域内的应用研究还相对较少,且较少采用 GRA 将子序列合并以提升总体预测精度。本文提出一种结合 TVF-EMD、GRA 和 LightGBM 的日径流预测组合模型(TGL 组合模型),目的是克服 LightGBM 对高分辨率数据变化的敏感性,提高非线性时间序列数据的模拟精度,从而实现日径流的高效、精确预测。

1 模型构建和评价指标

1.1 LightGBM

LightGBM 是 2017 年微软亚洲研究院基于梯度提升决策树算法框架提出的升级算法。在大数据应用场景中,面对大样本、高维度的数据集,LightGBM 模型具有训练速度快、计算效率高和预测较为精确的特点,因此特别适合大数据分类、预测等问题^[22]。

本文采用 LightGBM 模型进行日径流预测研究,由于 LightGBM 模型的超参数较多,参考文献[22]并结合日径流预测对精度和效率的需求,选取叶子节点中最小的样本数量、树的深度、学习率和最大叶子节点数作为超参数优化的研究对象。其中,树的深度、学习率和最大叶子节点数影响日径流预测的精度,最小样本数量可以提高模型的泛化能力。通过树状结构 Parzen 估计方法在推荐的取值空间内对超参数进行寻优,以提升 LightGBM 模型的预测精度。

1.2 TVF-EMD 方法

为了解决经验模态分解(empirical mode decomposition, EMD)存在的模态混叠和低采样率下效率较低的问题,Heng 等^[23]提出一种 TVF-EMD 方法。TVF-EMD 方法通过 B 样条逼近滤波器来完成截止频率的选择,也改进了停止准则,可以有效解决

模态混叠问题,使得 TVF-EMD 方法在低采样率下有很好的鲁棒性,被逐步应用于各种工程机械振动信号的分离。本文利用 TVF-EMD 方法将实测日径流序列分解为若干个较为线性平稳的子序列,再对各子序列进行预测,最后叠加获得日径流的预测结果。通过 TVF-EMD 方法有效降低实测径流过程非线性和非平稳性的影响,进一步提高日径流的预测精度。

1.3 GRA

GRA^[24]是灰色系统理论的一个重要分支,其基本思想是根据序列的几何形状判断不同序列之间联系的紧密程度。GRA 对样本的需求量较小,在一定程度上可以排除系统分析时的主观性,且计算过程中占用内存空间不大,因此广泛应用于各个研究领域。本文利用 GRA 计算日径流实测序列与经过 TVF-EMD 方法分解得到的子序列之间的灰色关联系数,进而将灰色关联系数相同的子序列进行叠加重构。

1.4 模型预测流程

为了解决现有径流预测模型预测精度较低、普适性不强和效率不高的问题^[25],本文提出了 TGL 组合模型,模型预测流程如图 1 所示。

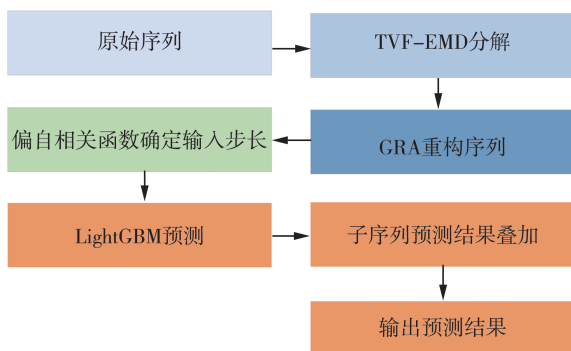


图 1 TGL 组合模型预测流程

Fig.1 Flowchart of TGL combined model prediction

TGL 组合模型预测流程为:① 利用 TVF-EMD 将日径流实测序列分解为若干个固有模态函数 (intrinsic mode function, IMF),即为日径流子序列 IMF;② 通过 GRA 计算子序列与原始序列的灰色关联系数,将灰色关联系数相同的 IMF 合并,重构成新的子序列 IMF*;③ 通过偏自相关分析确定每个重构序列 IMF* 的输入因子;④ 使用 LightGBM 对每个重构序列 IMF* 进行预测,并将预测结果叠加得出最终的预测值。

1.5 评价指标

为了评价 TGL 组合模型对于日径流预测的性能,采用纳什效率系数 (NSE)、相关系数 (R)、均方根误差 (RMSE)、平均绝对误差 (MAE) 和峰值流量误差 (top flow prediction error, TPE)^[26] 作为模型预

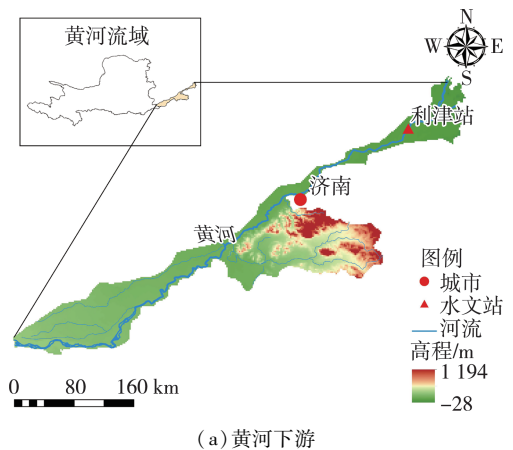
测结果的精度评价指标。

NSE 和 R 是评估预测值和实测值之间差异的有效指标,NSE 最大值为 1, R 取值范围为 $[-1, 1]$,NSE 和 R 值越大,预测值与实测值越接近;RMSE 和 MAE 反映预测结果的准确性;TPE 反映峰值流量预测结果的准确性,TPE 最小值为 0,其值越小,预测结果越准确。

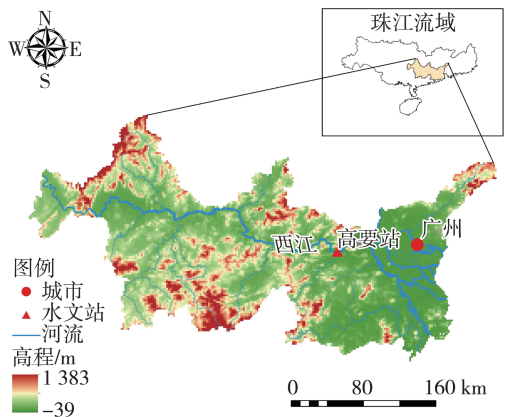
2 实例验证

2.1 研究区概况及数据来源

为了验证本文所提出的模型在日径流预测中的性能特点和适用性,选取我国黄河下游的利津站和珠江下游的高要站为研究对象,对其日径流进行预测,研究区域如图 2 所示。



(a) 黄河下游



(b) 珠江下游

图 2 研究区概况

Fig.2 Schematic diagram of study areas

利津水文站是黄河最下游的水文站,是全国大江大河重要水文站,位于山东省东营市利津县,集水面积 75.2 万 km^2 ,年均径流量为 200.8 亿 m^3 ,历史最大流量 10 400 m^3/s 。利津站监测黄河入海水量、沙量,为黄河下游河道治理和水资源规划管理等提供水文资料,也为黄河下游防洪、防凌、水资源统一调度提供水情信息,对黄河河口地区的治理开发及经济发展起到关键作用。

高要水文站位于广东省肇庆市,是西江中下游国家级重要控制站,集水面积为 35.15 万 km²,占西江集水面积的 99.6%。西江是珠江流域的主干流,为中国第四大河流,其水资源丰富,为沿岸地区的农业灌溉、河运、发电等做出了巨大贡献。高要站为珠江三角洲地区的洪水预报与防洪减灾提供重要的决策依据;在枯水期间,实时水量、水质数据对指导深受咸潮上溯之苦的珠三角地区的取水时机选择具有非常重要的作用。

本文分别采用利津站 2010—2014 年和高要站 2004—2008 年的日径流实测序列进行预测研究。两测站的日均流量数据分别来自黄河水利委员会和珠江水利委员会,所有数据均按照国家标准进行测量和检验,具有可靠性和真实性。两测站的日径流实测序列如图 3 所示,统计特征如表 1 所示。选取每个水文站前 80% 的数据作为训练集,后 20% 的数据作为测试集,进行预测研究。

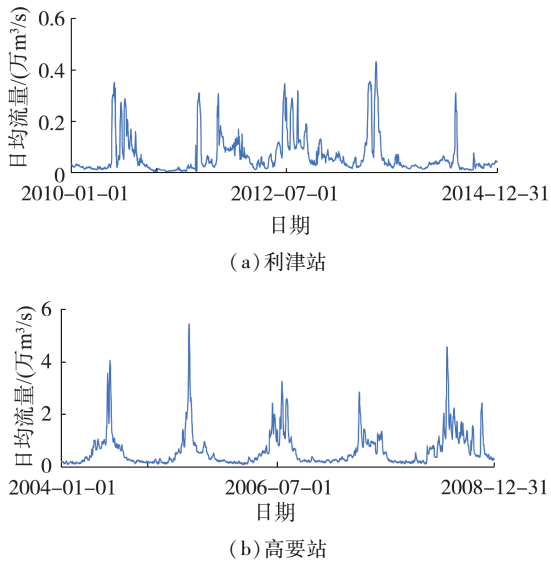


图 3 两测站日径流实测序列

Fig. 3 Daily runoff measurement sequence at two stations

表 1 两测站日径流序列统计特征

Table 1 Statistical characteristics of daily runoff series at two stations

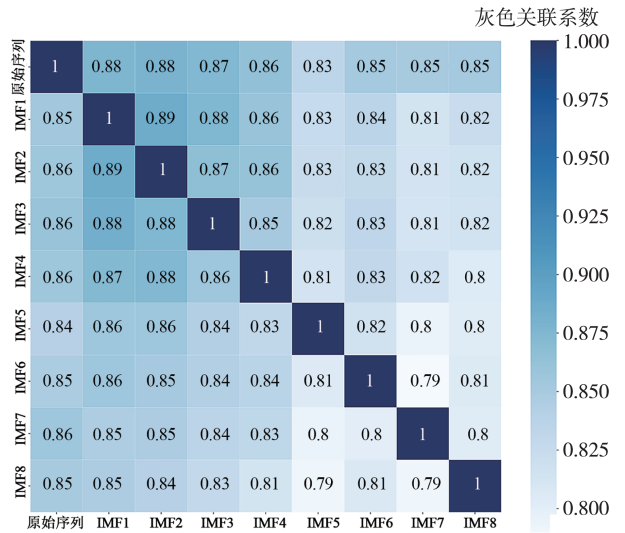
测站	序列长度/d	日径流序列统计值/(m ³ /s)			
		最大值	最小值	平均值	标准差
利津站	1826	4330	34.3	641.41	738.15
高要站	1827	54500	872.0	6330.26	6648.99

2.2 模型构建

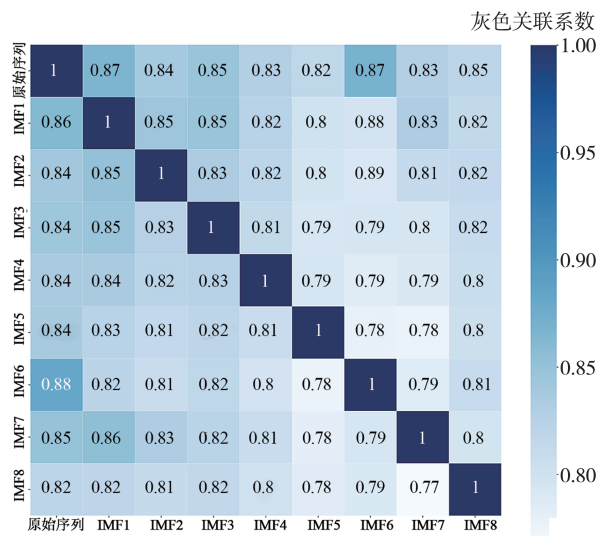
2.2.1 数据分解与序列重构

为了解决数据噪声对预测结果造成的不利影响,采用 TVF-EMD 方法对利津站和高要站的日径流实测序列进行分解,并通过 GRA 计算每个日径流子序列 IMF 与原始日径流序列的灰色关联系数(图

4),进而重构子序列 IMF 以减小预测误差。由于选取原始序列作为参考序列,各子序列 IMF 作为比较序列,采用图 4 中第一行数据所示的灰色关联系数作为序列重构的依据,将与原始序列灰色关联系数相同的子序列 IMF 进行重构(表 2),得到重构序列 IMF* (图 5)。



(a) 利津站



(b) 高要站

图 4 子序列 IMF 与原始序列的灰色关联系数

Fig. 4 Grey correlation coefficient between IMF and original sequence

表 2 两测站子序列 IMF 重构的组合方式

Table 2 Combination of IMF sequence reconstruction for two stations

重构序列	利津站 IMF 分量组合	高要站 IMF 分量组合
IMF1*	IMF1 + IMF2	IMF1 + IMF6
IMF2*	IMF3	IMF2
IMF3*	IMF4	IMF3
IMF4*	IMF6 + IMF7 + IMF8	IMF4 + IMF7
IMF5*	IMF5	IMF5 + IMF8

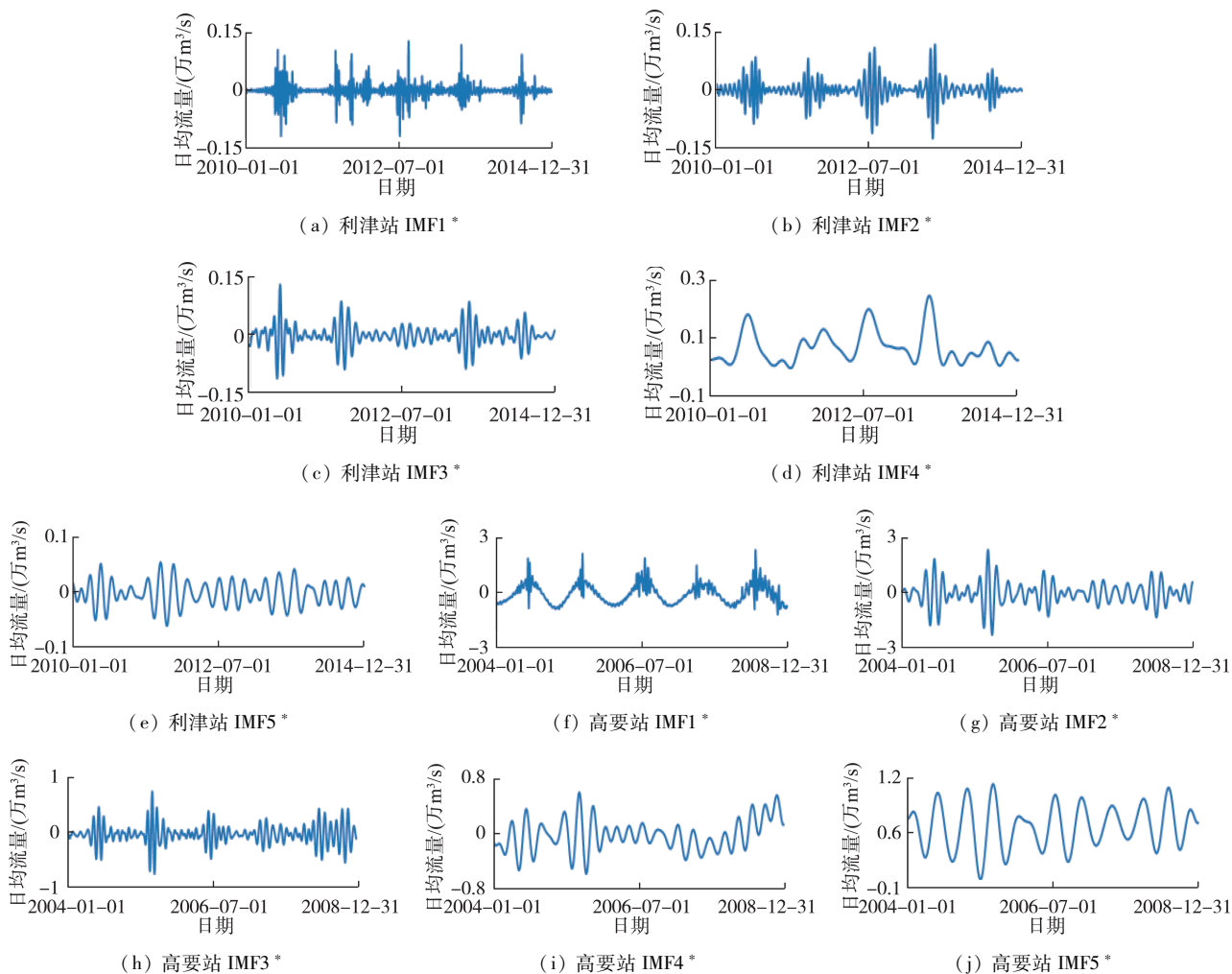


图 5 两测站重构序列 IMF*

Fig. 5 Reconstructed IMF* sequences of two stations

2.2.2 输入因子及模型参数

不同的预处理方法、输入变量和数据大小会在一定程度上影响模型的性能^[27]。本文通过采用偏自相关分析,在 95% 置信区间下确定每个重构序列 IMF* 的输入步长 s 。当预测第 i 个重构序列 IMF* 第 t 天的日径流 x_t 时,输入因子为第 $(t-1)$ 天到第 $(t-s)$ 天的日径流序列,即 $x_{i(t-1)}$ 、 $x_{i(t-2)}$ 、 \dots 、 $x_{i(t-s)}$,各重构序列 IMF* 的输入步长及输入因子如表 3 所示。

由于模型参数对于预测效果起着关键作用,本文通过寻优获得每个重构序列 IMF* 预测模型参数的最优值,如表 4 所示。首先将测试集输入到具有最优超参数的 LightGBM 模型中,利用经过训练的 N 棵决策树对日径流进行预测,输出各叶子节点未经 sigmoid 函数规格化的预测值;再将每棵决策树中得到的预测值叠加得到日径流的初始预测值;最后通过模型中的函数变换输出日径流预测值。

表 3 各重构序列 IMF* 的输入步长及输入因子

Table 3 Input steps and input factors of each reconstructed IMF* sequence

测站	IMF* 分量	输入步长	输入因子
利津站	IMF1*	3	$x_{1(t-1)}$ 、 $x_{1(t-2)}$ 、 $x_{1(t-3)}$
	IMF2*	6	$x_{2(t-1)}$ 、 $x_{2(t-2)}$ 、 $x_{2(t-3)}$ 、 $x_{2(t-4)}$ 、 $x_{2(t-5)}$ 、 $x_{2(t-6)}$
	IMF3*	5	$x_{3(t-1)}$ 、 $x_{3(t-2)}$ 、 $x_{3(t-3)}$ 、 $x_{3(t-4)}$ 、 $x_{3(t-5)}$
	IMF4*	3	$x_{4(t-1)}$ 、 $x_{4(t-2)}$ 、 $x_{4(t-3)}$
	IMF5*	9	$x_{5(t-1)}$ 、 $x_{5(t-2)}$ 、 $x_{5(t-3)}$ 、 $x_{5(t-4)}$ 、 $x_{5(t-5)}$ 、 $x_{5(t-6)}$ 、 $x_{5(t-7)}$ 、 $x_{5(t-8)}$ 、 $x_{5(t-9)}$
高要站	IMF1*	4	$x_{1(t-1)}$ 、 $x_{1(t-2)}$ 、 $x_{1(t-3)}$ 、 $x_{1(t-4)}$
	IMF2*	6	$x_{2(t-1)}$ 、 $x_{2(t-2)}$ 、 $x_{2(t-3)}$ 、 $x_{2(t-4)}$ 、 $x_{2(t-5)}$ 、 $x_{2(t-6)}$
	IMF3*	5	$x_{3(t-1)}$ 、 $x_{3(t-2)}$ 、 $x_{3(t-3)}$ 、 $x_{3(t-4)}$ 、 $x_{3(t-5)}$
	IMF4*	6	$x_{4(t-1)}$ 、 $x_{4(t-2)}$ 、 $x_{4(t-3)}$ 、 $x_{4(t-4)}$ 、 $x_{4(t-5)}$ 、 $x_{4(t-6)}$
	IMF5*	4	$x_{5(t-1)}$ 、 $x_{5(t-2)}$ 、 $x_{5(t-3)}$ 、 $x_{5(t-4)}$

表 4 各重构序列 IMF* 预测模型对应 LightGBM 的最优参数

Table 4 Optimal parameters of LightGBM corresponding to each reconstructed IMF* sequence prediction model

测站	IMF* 分量	最小样本数量	树的深度	学习率	最大叶子节点数
利津站	IMF1*	4	4	0.265	2
	IMF2*	7	6	0.203	8
	IMF3*	6	3	0.236	4
	IMF4*	4	3	0.121	5
	IMF5*	10	3	0.161	5
高要站	IMF1*	5	5	0.450	2
	IMF2*	7	6	0.180	5
	IMF3*	6	9	0.205	5
	IMF4*	7	9	0.186	6
	IMF5*	5	5	0.135	6

2.3 结果与分析

基于本文建立的 TGL 组合模型分别对利津站和高要站进行日径流预测。为了验证模型的性能,分别建立了 LightGBM、GRU、SVM、BP 等单一模型和 ESMD-GRA-LightGBM(以下简称为 EGL)、ESMD-GRA-GRU(以下简称为 EGG)组合模型进行对比分析。EGL 和 EGG 组合模型构建过程利用 ESMD 分解方法对原始序列进行分解,通过 GRA 对分解序列重构,分别利用 LightGBM 和 GRU 对各个重构序列进行预测,最后叠加获得预测结果。这些模型对不同测站日径流的预测结果如图 6 所示,对应的评价指标及运行时间如表 5 所示。

由图 6 和表 5 可知,TGL 组合模型对于利津站和高要站的预测结果整体上优于 EGL 和 EGG 组合模型以及 LightGBM、GRU、SVM 和 BP 模型。对于不同流域且量级差异明显的利津站和高要站的日径流实测序列,TGL 组合模型的预测精度均优于其他模型,具有更强的预测性能。

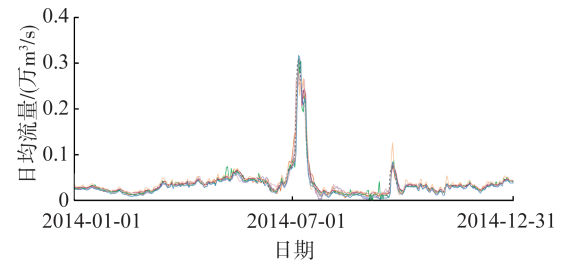
对于利津站,可以清楚地看出相较于其他模型,TGL 组合模型的预测效果最佳。TGL 组合模型预测

表 5 各模型对两测站日径流预测结果的评价指标

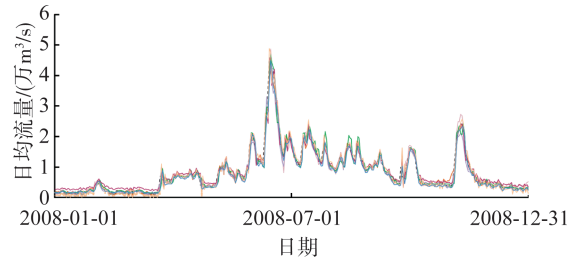
Table 5 Evaluation indexes of daily runoff prediction results of two stations by different models

测站	模型	NSE	RMSE/(m ³ /s)	R	MAE/(m ³ /s)	TPE	t/s
利津站	TGL	0.949	85.17	0.974	40.96	0.0780	16.65
	EGL	0.945	88.37	0.972	48.79	0.1127	18.56
	EGG	0.938	93.64	0.970	48.75	0.0811	638.32
	LightGBM	0.933	97.41	0.966	46.44	0.1206	4.11
	GRU	0.937	94.97	0.969	42.86	0.0821	152.33
	SVM	0.915	106.71	0.963	58.34	0.0808	3.83
	BP	0.928	98.96	0.970	50.49	0.0967	3.71
高要站	TGL	0.966	1358.84	0.984	864.15	0.0726	18.32
	EGL	0.964	1407.43	0.982	871.64	0.0756	20.16
	EGG	0.961	1461.16	0.984	867.26	0.0944	599.61
	LightGBM	0.953	1600.19	0.977	983.65	0.0779	4.38
	GRU	0.958	1511.29	0.983	1094.82	0.1012	166.49
	SVM	0.954	1681.03	0.977	1157.06	0.0999	3.69
	BP	0.950	1553.62	0.982	929.16	0.0995	4.02

--- 实测值 — TGL组合模型 — EGL组合模型 — EGG组合模型
— LightGBM模型 — GRU模型 — SVM模型 — BP模型



(a) 利津站



(b) 高要站

图 6 各模型对两测站日径流的预测结果

Fig. 6 Prediction results of daily runoff of two stations by different models

结果的 NSE 和 R 值分别为 0.949 和 0.974,分别比 LightGBM 模型提高了 1.71% 和 0.82%;RMSE 和 MAE 分别为 85.17 m³/s 和 40.96 m³/s,相较于 LightGBM 模型分别降低了 14.37% 和 13.39%。与 GRU 模型预测结果相比,TGL 组合模型预测结果的 NSE 和 R 值分别提升了 1.26% 和 0.51%,RMSE 和 MAE 分别降低了 11.51% 和 4.64%;与 BP 模型预测结果相比,TGL 组合模型预测结果的 NSE 和 R 值分别提升了 2.21% 和 0.41%,RMSE 和 MAE 分别降低了 16.19% 和 23.26%;与 SVM 模型预测结果相比,TGL 组合模型预测结果的 NSE 和 R 值分别提升了 3.58% 和 1.13%,RMSE 和 MAE 分别降低了

42.43% 和 25.29% ,SVM 模型整体的预测效果最差。TGL 组合模型与 EGL 和 EGG 组合模型相比,也表现出较好的预测效果。相较于 EGL 组合模型,TGL 组合模型预测结果的 NSE 和 R 值分别提升了 0.53% 和 0.21% ,RMSE 和 MAE 分别降低了 3.76% 和 19.12% ;相较于 EGG 组合模型,TGL 组合模型预测结果的 NSE 和 R 值分别提升了 1.16% 和 0.41% ,RMSE 和 MAE 分别降低了 9.94% 和 19.02% 。

对于高要站,TGL 组合模型对于日径流序列整体上的预测效果在 7 个模型中表现最佳,较好地拟合了实测序列的变化趋势,其预测结果的 NSE 和 R 值分别为 0.966 和 0.984 ,RMSE 和 MAE 分别为 $1358.84 \text{ m}^3/\text{s}$ 和 $864.15 \text{ m}^3/\text{s}$ 。与 LightGBM、GRU、SVM 和 BP 模型结果相比,TGL 组合模型预测结果的 NSE 分别提升了 1.35%、0.83%、1.24% 和 1.66% , R 值分别提升了 0.71%、0.1%、0.71% 和 0.2% ,RMSE 分别降低了 17.76%、11.22%、23.71% 和 14.33% ,MAE 分别降低了 13.83%、26.69%、33.9% 和 7.53% 。与 EGL 和 EGG 组合模型相比,TGL 组合模型的预测精度也有一定程度的提升。

从图 7 所示的利津站预测流量和实测流量散点图可以看出,TGL 组合模型的预测结果更加接近实测日径流。高要站的散点图也呈现出类似的分布情况,进一步证明 TGL 组合模型优异的预测性能。

TGL 组合模型在 LightGBM 模型的基础上预测性能得到进一步提升,预测结果的各项指标相较于其他模型均达到最佳。对于短期预测,减少模型在训练和测试中所需的计算时间具有重要意义^[28]。从表 5 可以看出 TGL 组合模型的运行时间远低于 GRU 模型和 EGG 组合模型,预测效率有显著的提

升。虽然与 LightGBM、SVM 和 BP 模型相比,计算时间有所增加,但在可以接受的范围内。

峰值流量预测是径流预测最为关注的方面,准确的峰值流量预测对防洪减灾和水资源规划管理起着十分重要的作用。为了验证 TGL 组合模型对于日径流峰值流量预测的准确性,利用峰值流量预测误差指标 TPE 来评价其预测性能。由表 5 可知,对于利津站的日径流峰值流量预测,TGL 组合模型预测结果的 TPE 相较于 EGL 和 EGG 组合模型以及 LightGBM、GRU、SVM 和 BP 模型分别减小约 44.49%、3.97%、54.62%、5.26%、3.59% 和 23.97% ;对于高要站的日径流峰值预测,TGL 组合模型预测结果的 TPE 相较于 EGL 和 EGG 组合模型以及 LightGBM、GRU、SVM 和 BP 模型分别减小了约 4.13%、30.03%、7.3%、39.39%、37.6% 和 37.05% 。与其他模型相比,TGL 组合模型具有良好的峰值流量预测性能。

综上,TGL 组合模型解决了 LightGBM 模型在处理强非线性时间序列数据时的缺点,一定程度上避免了冗余信息对 LightGBM 模型预测产生的干扰,并且采用 GRA 对分解序列进行重构,减小了误差的累积。TGL 组合模型与 LightGBM、GRU、SVM 和 BP 等单一模型相比,提升了抗噪声干扰的稳定性,预测的准确性大幅度提高,且相较于 EGL、EGG 组合模型仍具有较好的预测性能。TGL 组合模型对于不同流域且量级差异明显的日径流预测均具有很强的鲁棒性,在保证精度最佳的情况下,相较于深度学习模型 GRU 和组合模型 EGG 仍然具有较高的效率。同时,TGL 组合模型对于不同量级的峰值流量均获得了良好的模拟效果,可以为防范峰值流量产生的洪水风险和危害发挥重要作用。

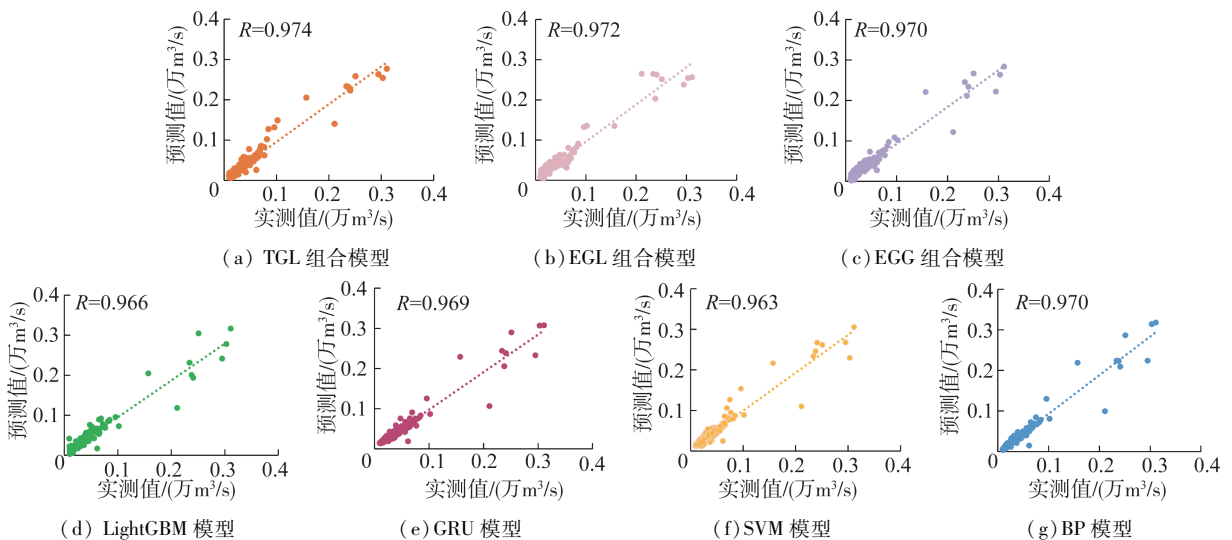


图 7 利津站预测流量与实测流量散点图

Fig. 7 Scatter plots of predicted flow rate and measured flow rate at Lijin station

3 结 论

a. LightGBM 模型具有精度高、收敛速度快、内存消耗小和缓存优化友好等集成模型结构特点,可满足日径流预测在预测精度和运行效率方面的需求,在短期径流预测中具有极高的应用价值,未来可进一步应用于实时降雨预测和山洪预报中。

b. TVF-EMD 方法可有效分离趋势项、波动项和噪声分量,提高了 LightGBM 模型在处理非线性序列时的性能,并通过 GRA 对子序列重组,减小重构误差。基于 TVF-EMD、GRA 及 LightGBM 方法建立的日径流预测组合模型 TGL 相对于其他模型,对不同流域且量级差异明显的日径流预测均具有更高的精度、更强的鲁棒性和广泛的适用性,尤其相对于深度学习模型 GRU 仍然表现出显著的效率优势。

c. TGL 组合模型对不同量级峰值流量的预测性能优于其他模型,进一步验证了该模型具有良好的稳定性和适应性, TGL 组合模型可为缓解洪峰流量、智能精细化水资源调控和数字孪生流域建设提供有力支撑。

参考文献:

- [1] 谭乔凤,陈然,朱阳,等. 基于多因子最近邻抽样回归模型的径流相似性预报[J]. 河海大学学报(自然科学版), 2020, 48(6): 521-527. (TAN Qiaofeng, CHEN Ran, ZHU Yang, et al. Runoff similarity forecast based on multi-factor nearest neighbor bootstrapping regressive model [J]. Journal of Hohai University (Natural Sciences), 2020, 48(6): 521-527. (in Chinese))
- [2] FENG Zhongkai, NIU Wenjing, TANG Zhengyang, et al. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction[J]. Journal of Hydrology, 2021, 595: 126062.
- [3] LEE D, LEE G, KIM S, et al. Future runoff analysis in the Mekong River Basin under a climate change scenario using deep learning[J]. Water, 2020, 12(6): 1556.
- [4] YUAN Ruifang, CAI Siyu, LIAO Weihong, et al. Daily runoff forecasting using ensemble empirical mode decomposition and long short-term memory[J]. Frontiers in Earth Science, 2021, 9: 621780.
- [5] 王富强,霍风霖. 中长期水文预报方法研究综述[J]. 人民黄河, 2010, 32(3): 25-28. (WANG Fuqiang, HUO Fenglin. A review of medium and long term hydrological forecasting methods [J]. Yellow River, 2010, 32(3): 25-28. (in Chinese))
- [6] ZHANG Jinping, XIAO Honglin, FANG Hongyuan.

Component-based reconstruction prediction of runoff at multi-time scales in the source area of the Yellow River based on the ARMA model [J]. Water Resources Management, 2022, 36(1): 433-448.

- [7] BARTHOLOMEW D J, BOX G E P, JENKINS G M. Time series analysis forecasting and control [J]. Journal of the Operational Research Society, 1971, 22(2): 199-201.
- [8] 王琪,张亭亭,游海林,等. 基于多元回归分析的大伙房水库径流中长期预报[J]. 水力发电, 2014, 40(5): 17-20. (WANG Qi, ZHANG Tingting, YOU Hailin, et al. Mid-long term runoff forecasting model for Dahuofang Reservoir based on multiple regression analysis [J]. Water Power, 2014, 40(5): 17-20. (in Chinese))
- [9] 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))
- [10] 何映颖,陈晓宏,张云,等. BP 神经网络在小流域径流模拟中的应用[J]. 水文, 2015, 35(5): 35-40. (HE Yiyang, CHEN Xiaohong, ZHANG Yun, et al. Application of BP neural network model in runoff simulating for small watershed [J]. Journal of China Hydrology, 2015, 35(5): 35-40. (in Chinese))
- [11] 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))
- [12] 马盼盼,白涛,武连洲,等. 黄河源区河川径流短期预测的 ANFIS 模型[J]. 西北农林科技大学学报(自然科学版), 2018, 46(6): 145-154. (MA Panpan, BAI Tao, WU Lianzhou, et al. Short-term runoff prediction by ANFIS model in source region of the Yellow River [J]. Journal of Northwest A&F University (Natural Science Edition), 2018, 46(6): 145-154. (in Chinese))
- [13] 张亚杰,崔东文. 基于奇异谱分析的 SPBO-ANFIS 月径流组合预测模型[J]. 人民珠江, 2022, 43(5): 137-144. (ZHANG Yajie, CUI Dongwen. SPBO-ANFIS model of combined monthly runoff forecasting based on singular spectrum analysis [J]. Pearl River, 2022, 43(5): 137-144. (in Chinese))
- [14] ZHANG Jiangwei, CHEN Xiaohui, KHAN Amirul, et al. Daily runoff forecasting by deep recursive neural network [J]. Journal of Hydrology, 2021, 596: 126067.

(下转第 151 页)

- 版), 2023, 51(1):1-9. (WU Zhiyong, BAI Boyu, HE Hai, et al. Temporal and spatial characteristics of hydrological drought in the Pearl River Basin from 1981 to 2020 [J]. Journal of Hohai University (Natural Sciences), 2023, 51(1):1-9. (in Chinese))
- [37] HAO Zengchao, HAO Fanghua, SINGH V P, et al. Probabilistic prediction of hydrologic drought using a conditional probability approach based on the meta-Gaussian model [J]. Journal of Hydrology, 2016, 542: 772-780.
- [38] SATTAR M N, LEE J Y, SHIN J Y, et al. Probabilistic characteristics of drought propagation from meteorological to hydrological drought in South Korea [J]. Water Resources Management, 2019, 33(7):2439-2452.
- [39] WANG Fei, WANG Zongmin, YANG Haibo, et al. Comprehensive evaluation of hydrological drought and its relationships with meteorological drought in the Yellow River Basin, China [J]. Journal of Hydrology, 2020, 584:124751.
- [40] SHAO Jun. An asymptotic theory for linear model selection [J]. Statistica Sinica, 1997, 7(2):221-262.
(收稿日期:2022-10-18 编辑:施业)
-
- (上接第 142 页)
- [15] 陶思铭, 梁忠民, 陈在妮, 等. 长短期记忆网络在中长期径流预报中的应用 [J]. 武汉大学学报(工学版), 2021, 54(1):21-27. (TAO Siming, LIANG Zhongmin, CHEN Zaini, et al. Application of long-short term memory network in medium-long term runoff forecast [J]. Engineering Journal of Wuhan University, 2021, 54(1):21-27. (in Chinese))
- [16] ZHAO Xuehua, LYU Hanfang, LYU Shujin, et al. Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer [J]. Journal of Hydrology, 2021, 601:126607.
- [17] 宋欣瑞, 张宪琦, 张展, 等. 多传感器数据融合的复杂人体活动识别 [J]. 清华大学学报(自然科学版), 2020, 60(10):814821. (SONG Xinrui, ZHANG Xianqi, ZHANG Zhan, et al. Multi-sensor data fusion for complex human activity recognition [J]. Journal of Tsinghua University (Science and Technology), 2020, 60(10):814-821. (in Chinese))
- [18] YU Yong, SI Xiaosheng, HU Changhua, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural Computation, 2019, 31(7):1235-1270.
- [19] KE Guolin, MENG Qi, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [C]//Proceedings of the 31th International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017:3149-3157.
- [20] 王文川, 高畅, 徐雷. 基于 TVF-EMD 与 LSTM 神经网络耦合的月径流预测研究 [J]. 中国农村水利水电, 2022(2):76-81. (WANG Wenchuan, GAO Chang, XU Lei. Monthly runoff forecast based on TVF-EMD and LSTM neural network coupling [J]. China Rural Water and Hydropower, 2022(2):76-81. (in Chinese))
- [21] 赵磊娜, 王延鹏, 邵毅明, 等. 利用时变经验模态分解的主干道短时交通量预测 [J]. 重庆交通大学学报(自然科学版), 2022, 41(3):37-44. (ZHAO Leina, WANG Yanpeng, SHAO Yiming, et al. Short-term traffic volume prediction of arterial road using TVF-EMD method [J]. Journal of Chongqing Jiaotong University (Natural Science), 2022, 41(3):37-44. (in Chinese))
- [22] CUI Zhongjie, QING Xiaoxia, CHAI Hongxiang, et al. Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis [J]. Journal of Hydrology, 2021, 603:127124.
- [23] HENG Li, ZHI Li, WEI Mo. A time varying filter approach for empirical mode decomposition [J]. Signal Processing, 2017, 138:146-158.
- [24] 刘思峰, 蔡华, 杨英杰, 等. 灰色关联分析模型研究进展 [J]. 系统工程理论与实践, 2013, 33(8):2041-2046. (LIU Sifeng, CAI Hua, YANG Yingjie, et al. Advance in grey incidence analysis modelling [J]. Systems Engineering-Theory & Practice, 2013, 33(8):2041-2046. (in Chinese))
- [25] HAO Fanghua, SUN Mingze, GENG Xiaojun, et al. Coupling the Xinanjiang model with geomorphologic instantaneous unit hydrograph for flood forecasting in northeast China [J]. International Soil and Water Conservation Research, 2015, 3(1):66-76.
- [26] YIN Hanlin, ZHANG Xiuwei, WANG Fandu, et al. Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model [J]. Journal of Hydrology, 2021, 598:126378.
- [27] MOOSAVI V, FARD Z G, VAFAKHAH M. Which one is more important in daily runoff forecasting using data driven models: input data, model type, preprocessing or data length? [J]. Journal of Hydrology, 2022, 606:127429.
- [28] BERKHAHN S, FUCHS L, NEUWEILER I. An ensemble neural network model for real-time prediction of urban floods [J]. Journal of Hydrology, 2019, 575:743-754.
(收稿日期:2022-10-04 编辑:施业)