

武少振,任智慧,赵雪花,等.基于多模型比选耦合的降水预测[J].南水北调与水利科技(中英文),2024,22(1):99-109. WU S Z, REN Z H, ZHAO X H, et al. Precipitation prediction based on decomposition algorithm-based models[J]. South-to-North Water Transfers and Water Science & Technology, 2024, 22(1): 99-109. (in Chinese)

基于多模型比选耦合的降水预测

武少振^{1,2,3},任智慧^{1,5},赵雪花²,杨默远³,桑燕芳^{1,4,5}

(1. 中国科学院地理科学与资源研究所陆地水循环及地表过程重点实验室,北京 100101; 2. 太原理工大学水利科学与工程学院,太原 030024; 3. 北京市水科学技术研究院,北京 100038; 4. 复合链生自然灾害动力学应急管理部重点实验室,北京 100085; 5. 中国科学院大学,北京,100049)

摘要:变化环境下水文时间序列的模拟预测难度不断加大,以往研究大多聚焦在模型的不同组合尝试与应用探索,但缺乏针对不同组合模型适用性与稳定性的系统研究。选择经验模态分解(empirical mode decomposition, EMD)、总体平均经验模态分解(ensemble empirical mode decomposition, EEMD)、改进的总体平均经验模态分解(modified ensemble empirical mode decomposition, MEEMD)和变分模态分解(variational mode decomposition, VMD)4种常用的分解算法,与多元线性回归(multivariable linear regression, MLR)、随机森林(random forest, RF)、BP神经网络(back propagation, BP)、卷积神经网络(convolutional neural networks, CNN)和长短期记忆神经网络(long short-term memory, LSTM)5种具有代表性的模型结合,构建20种基于“分解-预测-重构”模式的组合模型,并以华北地区密云、官厅两流域年和汛期降水为例,进行模型适用性与稳定性综合对比分析。结果表明:单一模型对密云流域年降水和汛期降水的预测结果优于官厅流域,但整体预测结果均不理想;结合分解算法后的组合模型预测结果明显优于单一模型,且该预测结果存在正负误差抵消现象,因此有助于进一步提高组合模型的整体预测精度;与基于EMD系列的分解算法相比,VMD算法对模型预测精度提升效果最显著,组合模型适用性和稳定性整体上表现为VMD-MLR>VMD-LSTM>VMD-BP>VMD-CNN。

关键词:中长期预测;数据驱动模型;组合模型;时间序列分解;非平稳性

中图分类号: P333; TV125 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdqk.2024.0012

揭示流域水文过程的复杂变化规律,掌握未来水文演变情势,是研究解决实际水文水资源问题的基本依据和必要前提^[1]。由于受到诸多因素的共同作用和影响,实际水文过程复杂多变,实测水文时间序列也呈现出十分复杂的变化特征,对其精准预测十分困难^[2-3]。为掌握水文过程的未来演变情势,水文学者相继提出了许多模拟预测模型,其大致可分为过程驱动模型和数据驱动模型。过程驱动模型需考虑水循环物理过程及其影响因素,且对数据的数量和质量均有较高要求,因此在实际应用中时常受限。与过程驱动模型相比,数据驱动模型以气象水文要素之间的相互关系为基础,仅简单考虑流

域水文循环的物理过程,通过对数据进行训练和拟合形成最优化的决策模型,具有较好的稳定性和泛化能力,因此在中长期水文模拟预测中得到较好的应用^[4-6]。

受到气候变化和人类活动的持续影响,降水序列逐渐呈现出明显的非平稳性、非线性等复杂变化特征,单一模型在很多情况下已经无法取得理想的模拟预测效果^[7-8]。众多研究^[9-10]证明,基于“分解-预测-重构”模式的组合模型较单一模型模拟效果有明显优势。此类组合模型将分解算法与预测模型结合,处理水文时间序列的非平稳性之后再行预测,以进一步提高模拟预测精度^[11]。自Bates等

收稿日期:2023-08-16 修回日期:2023-12-02 网络出版时间:2023-12-29

网络出版地址:https://link.cnki.net/urlid/13.1430.TV.20231228.0016.002

基金项目:国家重点研发计划项目(2019YFA0606903);国家自然科学基金项目(41971040;52279020);山西省科技创新人才团队专项资助项目(202204051002027)

作者简介:武少振(1998—),男,河南濮阳人,主要从事水文水资源研究。E-mail:913489880@qq.com

通信作者:桑燕芳(1983—),男,山西黎城人,研究员,博士生导师,主要从事水文气象与自然灾害研究。E-mail:sangyf@igsrr.ac.cn

1969 年提出组合预测理论后,许多学者^[12-14]在分解算法和组合模式等方面进行了深入探讨与应用尝试,基于“分解-预测-重构”模式的组合模型研究取得了诸多成果。目前组合模型的相关研究已成为水文模拟预测的研究热点,但大多数研究集中在模型的不同组合尝试与应用探索等方面;针对不同组合模型的适用性评估与模拟优选等问题,当前缺乏系统性的研究与科学认识。

为此,在考虑分解算法和预测模型代表性的基础上,选用经验模态分解(empirical mode decomposition, EMD)、总体平均经验模态分解(ensemble empirical mode decomposition, EEMD)、改进的总体平均经验模态分解(modified ensemble empirical mode decomposition, MEEMD)和变分模态分解(variational mode decomposition, VMD)4种分解算法,并与基于回归分析的多元线性回归(multivariable linear regression, MLR)、基于 Bagging 算法的随机森林(random forest, RF)、BP 神经网络(back propagation, BP)、卷积神经网络(convolutional neural networks, CNN)和长短期记忆神经网络(long short-term memory, LSTM)5种数据驱动模型进行组合,构建多种“分解-预测-重构”组合模型,并以降水非平稳特征显著的华北地区密云、官厅两流域为例,对流域年降水和汛期降水进行模拟预测。通过综合对比不同单一模型和组合模型,分析各类模型适用性和稳定性的差异,以期为合理构建基于“分解-预测-重构”模式的组合模型,以及提高变化环境下的降水预测精度提供有效的方法参考。

1 研究方法

1.1 分解算法

作为一种自适应的时频局部化分析算法,EMD 依据序列自身时间尺度特征、本征模态函数(intrinsic mode function, IMF)概念,将复杂的时间序列自适应地分解为多个包含时间尺度局部信息的本征模态分量和一个趋势项^[15]。相较于目前较常用的其他分解算法(如小波分解等),EMD 无须定义基函数,操作简单,且处理非平稳、非线性时间序列时具有明显优势。EMD 分解原序列 $y(t)$ 可表示为

$$y(t) = \sum_{i=1}^n r_i(t) + q(t) \quad (1)$$

式中: $r_i(t)$ 为本征模态分量, $i = 1, 2, \dots, n$; n 表示分量个数; $q(t)$ 为趋势项。

为克服 EMD 分解过程中出现的模态混叠现象, Wu 等^[16]在待分析时间序列中加入白噪声,使不同时间尺度的时间序列可以自动分离到与其相适应的参考尺度上,进而提出了 EEMD。该算法在原始序列 $y(t)$ 中加入了 m 组白噪声序列 $n_j(t)$ ($j = 1, 2, \dots, m$), 要求 m 组白噪声的均值为零。待分析序列变为 $z_j(t)$:

$$z_j(t) = y(t) + n_j(t) \quad (2)$$

然后对 $z_j(t)$ 分别进行 EMD 分解,对得到的 IMF 求平均后作为最终序列分解结果。

EEMD 虽然能够在一定程度上抑制模态混淆,但计算量较大,添加的白噪声有时不能被完全中和。MEEMD 通过结合 EEMD 和基于排列熵的序列随机性检测,在检测出 EEMD 的异常分量之后,直接进行 EMD 分解,不仅抑制了 EMD 分解过程中的模态混淆问题,而且能够减小计算量,缩小重构误差^[17]。

VMD 是一种新型自适应、完全非递归的序列分解算法,能够有效降低复杂序列的非线性和非平稳性^[18]。相对于经验模态分解和小波分解, VMD 能够更好地还原原始序列,具有更好的噪声鲁棒性。使用该算法可以根据需要将时间序列分解为多个不同频段且相对平稳的子序列,每个子序列都具有局部化的时间和频率特性,可以更好地反映时间序列的局部特征^[19]。

1.2 预测模型

MLR 善于拟合多个自变量和因变量之间的线性关系,通过数据样本训练得到模型参数,从而确定具体的 MLR 方程进行预测^[20]。MLR 可解释性强,运行速度快,对小数据集尤为适用,但处理特征冗余的数据集时结果并不稳定。MLR 模型的一般形式为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (3)$$

式中: y 为因变量; β_i 为待定参数; $i = 0, 1, 2, \dots, k$; x_i 为自变量, $i = 1, 2, \dots, k$; ε 为随机变量, $\varepsilon \sim N(0, \sigma^2)$ 。根据最小二乘法,求解回归系数 β_i , 使误差项的平方和最小。

RF 是一种基于集成学习的监督式机器学习算法,在不同或相同类型算法的集成下,模型预测能力更加强大。RF 模型基于 Bootstrap 抽样生成决策树,可以克服多重共线性、数据噪声的影响,能良好

规避输入因子间的相关性或不符合正态分布等问题,从而对影响因子重要性进行无偏估计^[21]。

BP神经网络模型具有较高的非线性拟合精度、较强的自适应能力和一定的容错性^[22]。多层结构的BP神经网络可以对任意非线性连续函数进行精确拟合,很好地解决模型内部规律未知的多输入多输出拟合与模拟问题^[23]。

CNN是一种常见的前馈神经网络。在深度学习的背景下,CNN的平移不变性使其具有较强的模型泛化能力,同时避免了梯度爆炸或消失、缺乏内存保留等缺陷。此外,CNN还可以进行并行计算输出,很大程度上提高了模型计算效率。

LSTM作为一种为解决长时间序列训练过程中的梯度消失和梯度爆炸问题而优化设计的深度学

习网络,LSTM模型对序列数据拥有优异的处理能力,已在不同领域的模拟预测中得到广泛应用^[24]。

LSTM作为循环神经网络的发展,拥有同样的自循环反馈的全连接神经网络结构。此外,LSTM在神经元中被加入了3个独特的“门”结构,使得模型结构状态可以随序列流动,从而达到长期记忆的目的。

1.3 模型组合方案

为综合对比分析不同组合模型的模拟预测性能,收集密云、官厅两流域的年降水和汛期降水数据,首先采用M-K趋势检验、水文变异诊断系统等算法对降水序列进行非平稳特征分析;在此基础上,将4种常用分解算法与5种典型预测模型结合,共构建20种“分解-预测-重构”组合模型(图1),并对不同组合模型的适用性和稳定性进行综合对比分析。

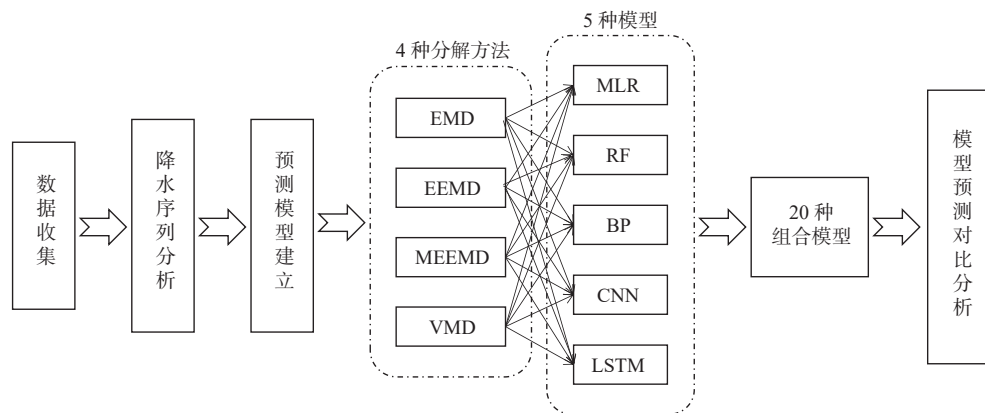


图1 利用不同组合模型的降水预测方案框架

Fig. 1 Prediction scheme of precipitation by different combined models

2 实例分析

2.1 研究区基本概况

密云、官厅两水库地处华北半湿润半干旱地区,控制流域面积分别为15 788 km²、43 402 km²。两流域均属于温带季风气候,降水丰枯变化明显且主要集中于夏季^[25-26]。受气候变异和人类活动的共同影响,两流域水文过程复杂多变,降水序列非平稳特性显著,模拟预测难度较高。

收集密云流域内小坝子、上黄旗、五道营、大阁、石人沟、安纯沟门、虎什哈、云州水库、雕鄂、黑龙江、白草、东万口、三道营、千家店、喇叭沟门共15个气象站1960—2021年的逐月降水序列,以及官厅流域及周边地区张北、右玉、集宁、大同、五台山、蔚县、五寨、原平、张家口、怀来、北京共11个气象站1960—2019年的逐月降水序列,通过泰

森多边形处理后得到流域汛期(6—9月)降水量及年降水量序列,见图2。

利用M-K趋势检验法对密云、官厅两流域降水进行趋势分析,结果显示:密云流域年降水和汛期降水趋势统计量均为-0.7,官厅流域年降水和汛期降水趋势统计量分别为-0.4和-0.7,均未通过5%置信水平的显著性检验,表明两流域降水下降趋势不显著。经水文变异诊断系统与标准化离散小波谱方法综合诊断,密云、官厅两流域降水序列均未诊断出显著变异点和显著周期特征,结果表明降水序列表现为较强的随机变化特征。进一步分析降水序列的自相关特性,结果表明密云流域、官厅流域年和汛期降水序列均存在 $t=5$ 年的时间滞后效应,因此在对研究区进行降水预测时,需要重点考虑降水的自相关特性。

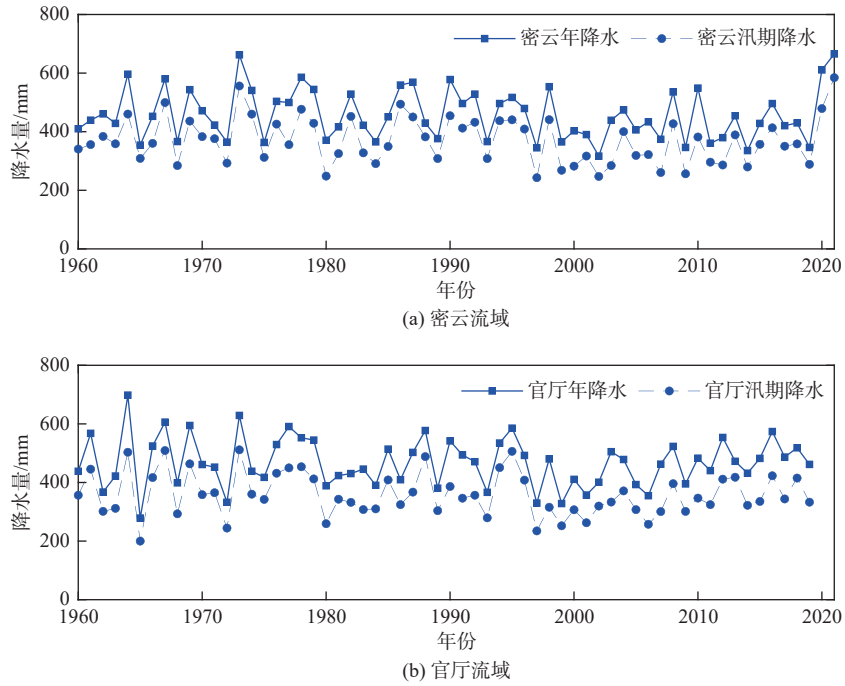


图 2 年和汛期降水时间序列

Fig. 2 Time series of the annual and flood season precipitation

2.2 单一模型降水预测

考虑降水时间序列的自相关特性,利用 5 种单一模型分别对密云、官厅两流域的年和汛期降水进行预测。数据处理方面,当以某年降水作为目标因子时,以其对应的前 5 年降水数据作为模型输入因子,以此类推将降水序列划分为模型输入集和目标集,最终以 1960—2004 年降水序列作为率定期数据,以 2005—2021 年(官厅 2005—2019 年)降水序列作为验证期数据。模型处理方面,对各模型进行参数率定:BP 模型为 3 层、最大训练次数 800 次、学习率 0.01、误差阈值 10^{-6} ;CNN 模型最大训练次

数为 200 次、学习率下降周期 100、学习率下降因子 0.6;LSTM 模型默认隐含层数为 1、隐含层神经元数量 20、最大训练次数为 800 次、学习率下降周期 400、学习率下降因子 0.2;RF 模型决策树数目和最小叶子数分别为 200、5;MLR 模型的参数估计在要求误差平方和为最小的前提下,用最小二乘法求解参数。此外,选用均方根误差(E_{RMS})、纳什效率系数(E_{NS})和平均绝对百分比误差(E_{MAP})作为评价指标,对各单一模型的降水预测结果进行评价。基于不同单一模型得到的降水预测结果见表 1。

表 1 单一模型得到的降水预测结果

Tab. 1 Prediction results of precipitation by single models

项目	率定期			验证期			
	E_{RMS}/mm	E_{NS}	E_{MAP}	E_{RMS}/mm	E_{NS}	E_{MAP}	
MLR	密云年	77.1	0.162	0.146	89.6	0.081	0.160
	密云汛期	74.5	0.121	0.182	77.7	0.131	0.166
	官厅年	81.7	0.101	0.154	69.5	-0.426	0.126
	官厅汛期	75.9	0.081	0.183	62.3	-0.561	0.172
BP	密云年	66.4	0.378	0.123	87.4	0.125	0.154
	密云汛期	55.5	0.513	0.101	78.5	0.111	0.182
	官厅年	73.7	0.267	0.143	73.6	-0.601	0.121
	官厅汛期	60.8	0.410	0.144	59.8	-0.438	0.152
CNN	密云年	39.5	0.780	0.077	79.0	0.284	0.136
	密云汛期	37.4	0.779	0.091	74.4	0.202	0.169

表 1(续)
Tab. 1 (Continued)

项目		率定期			验证期		
		E_{RMS}/mm	E_{NS}	E_{MAP}	E_{RMS}/mm	E_{NS}	E_{MAP}
CNN	官厅年	48.4	0.685	0.086	68.7	-0.395	0.133
	官厅汛期	39.5	0.751	0.095	63.8	-0.639	0.156
LSTM	密云年	44.2	0.724	0.077	86.0	0.152	0.148
	密云汛期	32.6	0.831	0.069	58.5	0.506	0.139
	官厅年	49.8	0.666	0.084	60.1	-0.066	0.098
	官厅汛期	32.4	0.833	0.068	60.1	-0.452	0.147
RF	密云年	33.8	0.839	0.066	76.0	0.338	0.131
	密云汛期	31.8	0.840	0.078	77.8	0.127	0.180
	官厅年	35.9	0.826	0.067	69.7	-0.433	0.126
	官厅汛期	31.6	0.841	0.072	64.8	-0.691	0.172

表 1 结果显示:模型整体降水预测精度依 MLR、BP、CNN、LSTM、RF 次序递增,且密云流域降水预测结果普遍优于官厅流域,汛期降水预测普遍优于年降水。各单一模型率定期的预测结果较好,但验证期均不理想。基于上述预测结果,采用麻雀搜索算法和粒子群优化算法对模型参数作进一步优化,并再次进行预测。结果发现:参数优化并不能明显改善各单一模型的降水预测结果,即单一模型的降水预测结果均不理想。

2.3 组合模型降水预测

2.3.1 基于不同分解算法的 CNN 组合模型降水预测

考虑到降水时间序列蕴含的复杂信息,分别采用 EMD、EEMD、MEEMD、VMD 算法对降水数据进行分解处理,并结合单一模型构建 20 种基于“分解-预测-重构”模式的组合模型,将分解后得到的具有规律性的降水子序列分别作为模型输入进行预测,最后求和计算得到最终的降水预测结果。利用 20 种组合模型分别对 4 组降水序列模拟预测,结果显示基于不同分解算法的 CNN 组合模型的结果最为稳定。因此,以 CNN 组合模型得到的官厅流域汛期降水的预测结果为例进行分析,结果见图 3 和表 2。

对比官厅流域汛期降水各子序列组的预测结果,经 EMD 和 MEEMD 分解后, CNN 模型的预测误差均主要集中在高频项 IMF1 和低频项 IMF3,其余子序列的预测误差较小,趋势项预测结果最好。利用 EEMD 分解后得到更多的降水子序列,但 CNN 模型的预测结果并未因此得到较大改善,模型对加入

的白噪声序列 IMF1 和 IMF7 的预测误差较大,进而影响了整体预测结果的精度。

此外,图 3 中分析结果显示模型结合分解算法后预测结果存在正负误差抵消现象。例如: EEMD-CNN 模型对 IMF1 的预测值偏高,与对应年份段 IMF7 预测值偏低抵消; MEEMD-CNN 模型对 IMF2、IMF3 整体预测误差为负,但与其余项的正误差相互抵消。需要指出的是,利用各分解算法同其他单一模型组合对降水进行预测时,同样存在不同子序列的正负误差抵消现象。

相比单一的 CNN 模型,结合分解算法后的 CNN 组合模型的降水预测结果得到明显改善。EMD、EEMD、MEEMD、VMD 等分解算法结合的 CNN 组合模型的预测精度依次提升,但基于 EMD 的分解算法的提升效果并不明显;而降水序列经过 VMD 分解后,模型预测结果达到最优,验证期 E_{NS} 为 0.879、 E_{MAP} 为 0.043。

2.3.2 基于 VMD 的不同组合模型降水预测

基于上述不同分解算法得到的降水预测结果,选择在本研究区适用性最佳的 VMD 算法,通过与各单一模型结合建立组合模型进行降水预测,结果见表 3 和图 4。经过 VMD 分解再进行降水预测,各模型的预测效果均取得了极大改善,率定期 E_{NS} 均在 0.919 以上、 E_{MAP} 最大为 0.045,验证期平均 E_{NS} 为 0.849、 E_{MAP} 最小仅为 0.020。

对比不同组合模型的降水预测结果发现: VMD-MLR 对 4 组降水序列的整体预测结果最优,验证期 E_{NS} 平均高达 0.936、 E_{MAP} 平均为 0.037; VMD-BP、VMD-LSTM 和 VMD-CNN 预测结果差距不明显;

此前预测效果较理想的 RF 模型在降水序列分解后的结果改善幅度较小, 预测结果最差, E_{NS} 平均值仅

为 0.677、 E_{MAP} 平均值为 0.074。

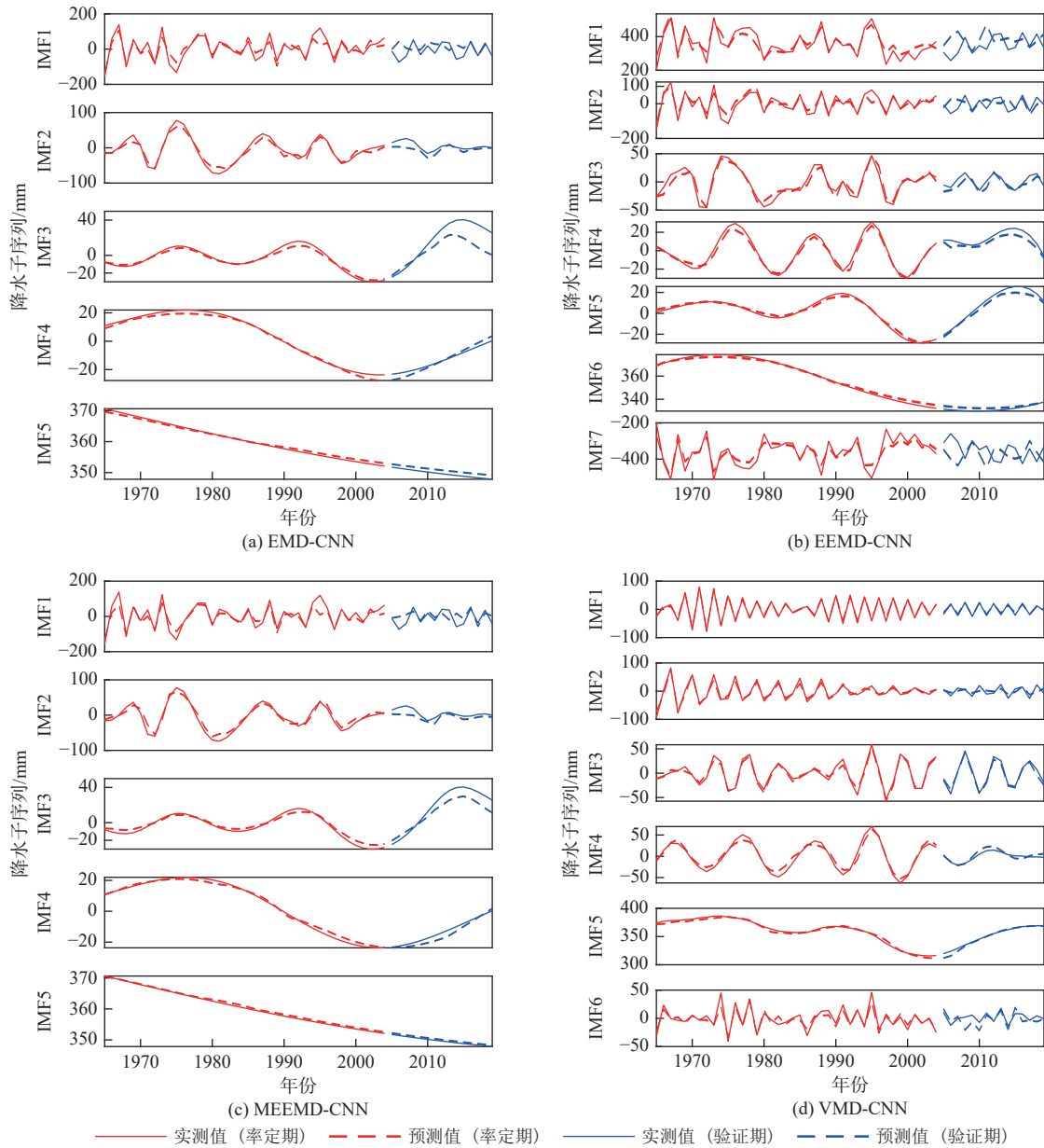


图 3 基于不同分解算法的 CNN 组合模型得到的降水子序列预测结果

Fig. 3 Prediction results of precipitation by CNN-combined models based on different decomposition algorithms

表 2 基于不同分解算法的 CNN 组合模型得到的降水预测结果

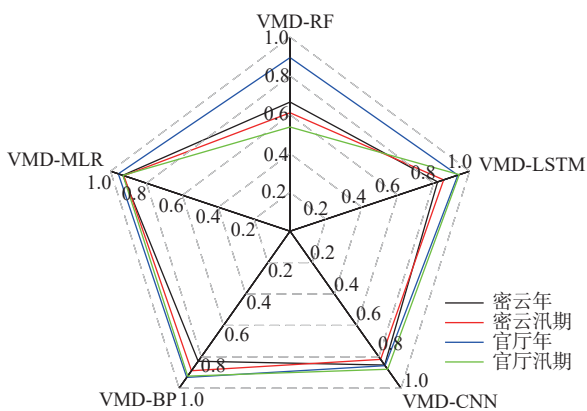
Tab. 2 Prediction results of precipitation by CNN-combined models based on different decomposition algorithms

CNN组合模型	率定期			验证		
	E_{RMS}/mm	E_{NS}	E_{MAP}	E_{RMS}/mm	E_{NS}	E_{MAP}
CNN	39.5	0.751	0.095	63.8	-0.639	0.156
EMD-CNN	33.7	0.819	0.070	44.6	0.199	0.106
EEMD-CNN	35.2	0.803	0.078	39.6	0.369	0.098
MEEMD-CNN	29.9	0.858	0.061	38.7	0.397	0.097
VMD-CNN	18.7	0.944	0.045	17.3	0.879	0.043

表 3 基于 VMD 的不同组合模型得到的降水预测结果

Tab. 3 Prediction results of precipitation by different VMD-combined models

VMD组合模型		率定期			验证期		
		E_{RMS}/mm	E_{NS}	E_{MAP}	E_{RMS}/mm	E_{NS}	E_{MAP}
VMD-MLR	密云年	13.8	0.973	0.023	24.5	0.931	0.047
	密云汛期	15.9	0.960	0.035	22.7	0.926	0.052
	官厅年	16.6	0.963	0.033	12.1	0.957	0.020
	官厅汛期	18.3	0.947	0.045	13.1	0.931	0.029
VMD-BP	密云年	18.0	0.954	0.032	38.7	0.828	0.063
	密云汛期	15.8	0.960	0.039	27.8	0.889	0.054
	官厅年	15.8	0.966	0.030	15.3	0.931	0.025
	官厅汛期	17.7	0.950	0.043	13.8	0.923	0.035
VMD-CNN	密云年	14.8	0.969	0.026	35.9	0.852	0.069
	密云汛期	18.4	0.946	0.043	35.7	0.816	0.088
	官厅年	24.5	0.919	0.045	22.0	0.856	0.039
	官厅汛期	18.7	0.944	0.045	17.3	0.879	0.043
VMD-LSTM	密云年	14.9	0.969	0.025	39.4	0.822	0.056
	密云汛期	14.5	0.967	0.032	31.6	0.856	0.062
	官厅年	14.9	0.970	0.028	14.9	0.935	0.027
	官厅汛期	18.9	0.943	0.045	11.9	0.943	0.025
VMD-RF	密云年	17.7	0.956	0.034	54.0	0.665	0.078
	密云汛期	14.7	0.966	0.033	52.0	0.610	0.116
	官厅年	18.9	0.952	0.036	18.9	0.894	0.032
	官厅汛期	16.3	0.958	0.039	33.9	0.537	0.069

图 4 基于 VMD 的不同组合模型验证期 E_{NS} 结果Fig. 4 E_{NS} results obtained from different VMD-combined models during the validation period

2.3.3 组合模型对比检验

选择性能较好的 VMD-MLR、VMD-BP、VMD-CNN 和 VMD-LSTM 进行模型适用性与稳定性检验。以预测中反应较灵敏的 N_{SE} 作为模型评价指标,利用各模型分别进行 100 次模拟预测,计算验证期 E_{NS} 的均值和离散系数,结果见表 4。

表 4 结果显示: VMD-MLR 对密云、官厅两流

域的年和汛期降水预测效果最优, E_{NS} 整体均值为 0.938, 离散系数整体均值为 0.005; VMD-CNN 模型的适用性最差, E_{NS} 整体均值为 0.807; VMD-BP 稳定性最差, 离散系数整体均值为 0.106。模型适用性以 VMD-CNN、VMD-BP、VMD-LSTM、VMD-MLR 次序递增, 稳定性按 VMD-BP、VMD-CNN、VMD-LSTM、VMD-MLR 顺序递增。模型预测效果整体表现为 VMD-MLR>VMD-LSTM>VMD-BP>VMD-CNN。VMD-MLR、VMD-LSTM 模型得到的验证期降水预测结果见图 5。

2.4 结果讨论

采用 MLR、BP、CNN、LSTM 和 RF 模型对华北地区密云、官厅两流域年降水和汛期降水序列进行模拟预测, 结果表明模型整体预测效果以 MLR、BP、CNN、LSTM、RF 次序依次递增, 对密云流域年和汛期降水的预测效果优于官厅流域, 但整体预测结果均不理想。经分析认为主要原因是: 本研究中降水序列长度较短且随机特征强, 导致单一模型难以对降水的复杂变化规律进行有效把握。

表 4 不同组合模型验证期的 E_{NS} 检验结果

Tab. 4 E_{NS} results obtained from different combined models during the validation period

验证期	均值				离散系数			
	VMD-MLR	VMD-BP	VMD-CNN	VMD-LSTM	VMD-MLR	VMD-BP	VMD-CNN	VMD-LSTM
密云年	0.938	0.841	0.837	0.826	0.012	0.070	0.059	0.078
密云汛期	0.926	0.739	0.753	0.839	0	0.167	0.080	0.090
官厅年	0.956	0.900	0.869	0.940	0.009	0.069	0.070	0.037
官厅汛期	0.931	0.823	0.769	0.884	0	0.118	0.093	0.070
整体均值	0.938	0.826	0.807	0.872	0.005	0.106	0.075	0.069

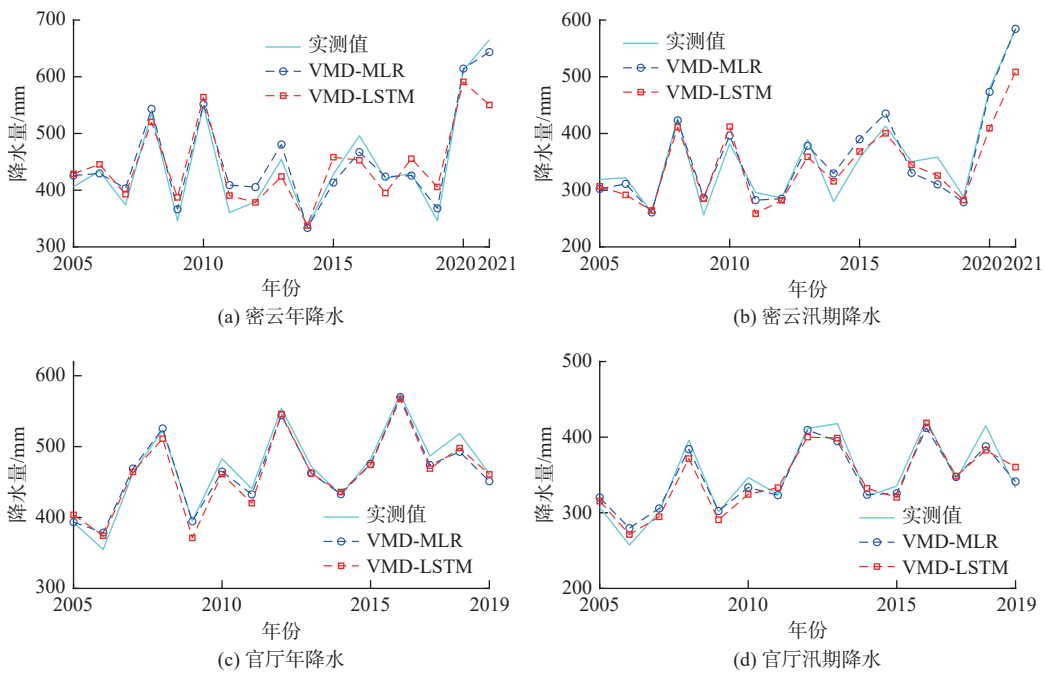


图 5 不同组合模型验证期的降水预测结果

Fig. 5 Prediction results of precipitation by using different combined models during the validation periods

相较单一模型,结合分解算法后的组合模型的预测效果明显改善,整体以 EMD、EEMD、MEEMD 和 VMD 次序逐渐提升。相较于 EMD 及其改进分解算法存在的模态混淆和端点效应等问题,具有更好数学理论依据的 VMD 方法具有更强的抗模态混淆能力和更好的噪声鲁棒性。基于 VMD 算法的组合模型适用性和稳定性整体上表现为 VMD-MLR>VMD-LSTM>VMD-BP>VMD-CNN。本研究中降水序列长度较短,导致需要大量数据训练支撑的深度机器学习模型的强大数据处理能力的优势并未充分体现,而基于最小二乘法的 MLR 模型对 VMD 分解后得到的相对平稳的周期性降水子序列具有更好的模拟性能。

组合模型的降水预测性能改善效果显著。对降水序列进行分解处理可有效降低数据区域化特征和非平稳特征,使得模型具有更高的预测精度以及

更强的稳定性和泛化能力。此外,模型结合分解算法后对各子序列的预测结果重构时存在正负误差抵消现象,该现象在以“分解-预测-重构”模式的组合模型预测中普遍存在,因此有助于进一步提高重构后模型整体的预测精度。

3 结论

以华北地区密云流域、官厅流域为例,对基于 4 种分解算法和 5 个预测模型的 20 个组合模型进行了综合对比分析。研究表明:在所选研究区内,单一模型难以对降水序列的复杂变化规律做出精准把握,而基于分解算法的组合模型可以有效改善模型预测结果;通过引入优化算法对模型参数优化可以在一定程度上提高预测精度,但在本文所选研究区内效果并不明显;在组合模型中,分解算法对原序列的分解效果直接影响模型预测结果,如

EMD、VMD 等不同分解算法下,模型预测结果存在明显差异;结合分解算法后,模型性能改善效果显著,适用性和稳定性增强,即使使用简单的模型(如MLR)对变化规律复杂的降水序列同样能够准确预测。

基于分解算法的组合模型预测效果明显优于单一模型,对于不同分解算法或预测模型的组合模型,其模拟预测效果也存在差异。为此,未来研究过程中还应考虑更多物理影响因子(如气候因子、大气环流因子、海温指标等)和更为复杂的“分解-预测-重构”模式的组合模型(如引入数据降维算法等),进一步提高模拟预测过程的物理依据,优化预测模型和预测过程,以进一步提高该区域降水的模拟预测精度与合理可靠性。

参考文献:

- [1] 唐兵,肖昌虎,黄金凤,等.太湖流域长江边界水文特性研究[J]. *水利水电技术(中英文)*, 2022, 53(2): 52-59. DOI: 10.13928/j.cnki.wrahe.2022.02.005.
- [2] 李新华,崔东文.基于WPD-RSA-ELM模型的水文时间序列多步预测[J]. *水利水电技术(中英文)*, 2022, 53(11): 69-77. DOI: 10.13928/j.cnki.wrahe.2022.11.007.
- [3] 宋晓猛,张建云,占车生,等.气候变化和人类活动对水文循环影响研究进展[J]. *水利学报*, 2013, 44(7): 779-790. DOI: 10.13243/j.cnki.slxb.2013.07.001.
- [4] MOHAMMADI B, MOAZENZADEH R, CHRISTIAN K, et al. Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models[J]. *Environmental Science and Pollution Research*, 2021, 28: 65752-65768. DOI: 10.1007/s11356-021-15563-1.
- [5] 王雪松.机器学习理论、方法及应用[M].北京:科学出版社,2009.
- [6] 徐盼盼,张奇莹,钱会,等.宝鸡地区降水特征分析及降水量预测[J]. *南水北调与水利科技*, 2017, 15(5): 63-69. DOI: 10.13476/j.cnki.nsbdqk.2017.05.010.
- [7] HU J, LIU Y, SANG Y F, et al. Precipitation variability and its response to urbanization in the Taihu Lake basin, China[J]. *Theoretical and Applied Climatology*, 2021, 144: 1205-1218. DOI: 10.1007/s00704-021-03597-x.
- [8] HUANG H P, LIANG Z M, LI B Q, et al. Combination of multiple data-driven models for long-term monthly runoff predictions based on bayesian model averaging[J]. *Water Resources Management*, 2019, 33(9): 3321-3338. DOI: 10.1007/s11269-019-02305-9.
- [9] 李昂,张坤,桑宇婷,等.基于ESMD-VMD-ESN二次分解组合模型的水位预测[J]. *人民珠江*, 2022, 43(11): 116-121,131. DOI: 10.3969/j.issn.1001-9235.2022.11.015.
- [10] ZHANG Y, LI C, JIANG Y, et al. Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model[J]. *Journal of Cleaner Production*, 2022, 354: 131724. DOI: 10.1016/j.jclepro.2022.131724.
- [11] 桑宇婷.汾河上游中长期径流预测方法研究[D].太原:太原理工大学,2020. DOI: 10.27352/d.cnki.gylgu.2020.001624.
- [12] SONG C, CHEN X H, WU P, et al. Combining time varying filtering based empirical mode decomposition and machine learning to predict precipitation from nonlinear series[J]. *Journal of Hydrology*, 2021, 603: 126914. DOI: 10.1016/j.jhydrol.2021.126914.
- [13] 姚欣明,陈元芳,顾圣华,等.EEMD-NNBR模型在降水预测中的应用[J]. *水电能源科学*, 2014, 32(12): 191-202. DOI: CNKI:SUN:SDNY.0.2014-12-003.
- [14] 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. DOI: 10.1080/02626667.2021.1937631.
- [15] 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. DOI: 10.13243/j.cnki.slxb.20190434.
- [16] WU Z, HUANG N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method[J]. *Advances in adaptive data analysis*, 2009, 1(1): 1-41. DOI: 10.1142/S1793536909000047.
- [17] 郑近德,程军圣,杨宇.改进的EEMD算法及其应用研究[J]. *振动与冲击*, 2013, 32(21): 21-26. DOI: 10.13465/j.cnki.jvs.2013.21.007.
- [18] 熊怡,周建中,孙娜,等.基于自适应变分模态分解和长短期记忆网络的月径流预报[J]. *水利学报*, 2023, 54(2): 172-198. DOI: 10.13243/j.cnki.slxb.20220459.
- [19] 吴松梅,蒋建东,燕跃豪,等.基于VMD-PSO-多核极限学习机的短期负荷预测[J]. *电力系统及其自动化*

- 化学报, 2021, 34(5): 18-25. DOI: [10.19635/j.cnki.csu-epsa.000864](https://doi.org/10.19635/j.cnki.csu-epsa.000864).
- [20] KOUADRI S, PANDE C B, PANNEERSELVAM B, et al. Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models[J]. *Environmental Science and Pollution Research*, 2022, 29(14): 21067-21091. DOI: [10.1007/s11356-021-17084-3](https://doi.org/10.1007/s11356-021-17084-3).
- [21] NAGHIBI S A, AHMADI K, DANESHI A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping[J]. *Water Resources Management*, 2017, 31: 2761-2775. DOI: [10.1007/s11269-017-1660-3](https://doi.org/10.1007/s11269-017-1660-3).
- [22] 李永坤, 马旭, 潘兴瑶, 等. 基于芒种日分析的BP模型在中长期汛期降雨量预报中的应用[J]. *南水北调与水利科技*, 2019, 17(3): 1-6,39. DOI: [10.13476/j.cnki.nsbdk.2019.0053](https://doi.org/10.13476/j.cnki.nsbdk.2019.0053).
- [23] LI J, CHENG J, SHI J, et al. Brief introduction of back propagation (BP) neural network algorithm and its improvement[C]//Advances in Computer Science and Information Engineering: Volume 2. Springer Berlin Heidelberg, 2012: 553-558.
- [24] VAN HOUTD G, MOSQUERA C, NÁPOLES G. A review on the long short-term memory model[J]. *Artificial Intelligence Review*, 2020, 53: 5929-5955. DOI: [10.1007/s10462-020-09838-1](https://doi.org/10.1007/s10462-020-09838-1).
- [25] 周连兄, 崔万晶, 赵云杰. 北京密云水库集水区降水及干旱时空分布特征[J]. *中国水土保持科学*, 2020, 18(5): 35-42. DOI: [10.16843/j.sswc.2020.05.005](https://doi.org/10.16843/j.sswc.2020.05.005).
- [26] 杜勇, 李建柱, 牛凯杰, 等. 1982—2015年永定河山区植被变化及对天然径流的影响[J]. *水利学报*, 2021, 52(11): 1309-1323. DOI: [10.13243/j.cnki.slxb.20210201](https://doi.org/10.13243/j.cnki.slxb.20210201).

Precipitation prediction based on decomposition algorithm-based models

WU Shaozhen^{1,2,3}, REN Zhihui^{1,5}, ZHAO Xuehua², YANG Moyuan³, SANG Yanfang^{1,4,5}

(1. Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. College of Water Resources Science and Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 3. Beijing Water Science and Technology Institute, Beijing 100038, China; 4. Key Laboratory of Compound and Chained Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China; 5. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Prediction of hydrological time series is a challenging issue due to complicated hydrologic processes, which would greatly impact the water resources management and hydraulic engineering design. Related studies indicated that combined models, which are based on the decomposition-prediction-reconstruction mode usually perform much better for the prediction of hydrological time series than single models. A great number of studies have been conducted on diverse combinations and applications of combined models, however, a comprehensive evaluation of the applicability and stability of different combined models is lacking, leaving a research gap for this important issue.

Four commonly used decomposition methods were applied, namely empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), modified ensemble empirical mode decomposition (MEEMD), and variational mode decomposition (VMD). The four decomposition methods were further combined with five representative prediction models, namely multivariable linear regression (MLR), random forest (RF), back propagation (BP), convolutional neural networks (CNN), and long short-term memory (LSTM), to establish a total of 20 combined models. These 20 combined models were used to predict the annual precipitation and flood season precipitation and conducted a comparative analysis of the models in the Miyun basin and Guanting basin in North China.

Results showed that: (1) The single models predicted both annual precipitation and flood season precipitation more accurately in the Miyun basin than in the Guanting basin, but the single model's performances were overall poor in the two basins. (2) The prediction results from the combined models after coupling with decomposition algorithms become much better than those from the single models, and the positive errors could be offset by the negative errors during the prediction processes when using the combined models, which could improve the overall

prediction accuracy of precipitation. (3) Compared with the EMD and other algorithms, the VMD algorithm has the most significant effect on improving the prediction accuracy of precipitation, and the applicability and stability of the combined models is VMD-MLR> VMD-LSTM> VMD-BP> VMD-CNN.

Moreover, the results indicated a single model can not accurately grasp the complex characteristics of the precipitation time series. The prediction accuracy of a single model could be approved through parameters optimization, however, the effect is not obvious. Compared with a single model, the combined models based on decomposition algorithms can effectively improve the prediction results. In the combined models, the effectiveness of decomposition algorithms (such as EMD and VMD) in decomposing the original time series directly affects the models' prediction results. After combining with the decomposition algorithm, the models' performance improves significantly, and their applicability and stability are greatly enhanced. After combining with the decomposition algorithm, even some simple models (such as MLR) can be used to accurately predict precipitation time series with complex variability patterns. Different model combinations and predictors lead to differences in prediction results among combined models. Therefore, more influencing factors (such as climate indicators) and more complex combined models based on the decomposition-prediction-reconstruction mode should be explored in future research to optimize the prediction model and prediction process, to further improve the prediction accuracy and reliability of the precipitation in this study area.

Key words: medium-to-long-term prediction; data-driven model; combined models; time series decomposition; non-stationarity characteristics