

DOI:10.13869/j.cnki.rswc.2023.04.021.

胡作龙, 高鹏. 基于 EEMD-SVM 模型的北洛河上游径流预测[J]. 水土保持研究, 2023, 30(4): 98-102, 109.

HU Zuolong, GAO Peng. Runoff Prediction in the Upper Reaches of Beiluo River Based on EEMD-SVM Model[J]. Research of Soil and Water Conservation, 2023, 30(4): 98-102, 109.

基于 EEMD-SVM 模型的北洛河上游径流预测

胡作龙¹, 高鹏²

(1.河南水环境勘测设计有限公司, 河南 三门峡 472000; 2.西北农林科技大学
水土保持研究所 黄土高原土壤侵蚀与旱地农业国家重点实验室, 陕西 杨凌 712100)

摘要: [目的] 提高北洛河上游径流预报精度, 为流域管理及水资源合理配置提供依据。[方法] 以 1971—2014 年北洛河上游吴旗水文站实测径流资料为基础, 采用 EEMD-SVM 耦合模型对吴旗站月径流序列进行了模拟预测, 并与 EEMD-ARIMA 和 EEMD-NAR 两种耦合模型的预测结果进行对比。[结果] EEMD-SVM 模型的平均绝对误差 (MAE) 和均方根误差 (RMSE) 最低, 决定系数 (R^2) 和纳什系数 (NSE) 最高。其中, 相比于 EEMD-ARIMA 和 EEMD-NAR 模型, EEMD-SVM 模型的决定系数 (R^2) 分别提高了 186.63%, 49.49%。[结论] EEMD-SVM 模型具有更高的预测精度和更强的非线性拟合能力, 可以成功地应用于北洛河上游的月径流预报。同时, 研究表明 EEMD-NAR 模型的预测性能高于 EEMD-ARIMA 性能。

关键词: 径流预测; 集合经验模态分解; 支持向量机; 北洛河

中图分类号: P338

文献标识码: A

文章编号: 1005-3409(2023)04-0098-05

Runoff Prediction in the Upper Reaches of Beiluo River Based on EEMD-SVM Model

HU Zuolong¹, GAO Peng²

(1. Henan Water Environment Survey and Design Co., Ltd., Sanmenxia, Henan 472000, China; 2. State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: [Objective] The aim of this study is to improve the accuracy of runoff forecast in the upper reaches of Beiluo River and provide basis for watershed management and rational allocation of water resources. [Methods] Based on the measured runoff data of Wuqi hydrological station in the upper reaches of Beiluo River from 1971 to 2014, the monthly runoff series was predicted by using EEMD-SVM coupling model, and compared with the prediction results of EEMD-ARIMA and EEMD-NAR models. [Results] EEMD-SVM model has the lowest mean absolute error and root mean square error, and the highest coefficient of determination (R^2) and Nash coefficient. Compared with EEMD-ARIMA and EEMD-NAR models, the R^2 value of EEMD-SVM model increases by 186.63% and 49.49%, respectively. [Conclusion] EEMD-SVM model has higher prediction accuracy and stronger nonlinear fitting ability, and can be successfully applied to the monthly runoff prediction in the upper reaches of Beiluo River. At the same time, the prediction performance of EEMD-NAR model is higher than that of EEMD-ARIMA model.

Keywords: runoff prediction; ensemble empirical mode decomposition; support vector machine; Beiluo River

收稿日期: 2022-05-23

修回日期: 2022-06-03

资助项目: 国家自然科学基金“黄土高原水土保持措施潜力及其对河流水沙的调控机制”(U2243211)

第一作者: 胡作龙 (1975—), 男, 甘肃泾川人, 高级工程师, 主要从事水利工程规划设计工作。E-mail: 531565936@qq.com

通信作者: 高鹏 (1976—), 男, 陕西扶风人, 博士, 研究员, 主要从事水土保持和流域生态水文研究。E-mail: gaopeng@ms.iswc.ac.cn

<http://stbcj.paperonce.org>

河川径流是水文水资源的重要组成部分,是水循环研究的主要对象,也是防洪、规划、水资源保护利用及水土保持工作中必须考虑的重要因素^[1]。受气候变化与人类活动的综合影响,近年来黄河中游区间干支流的径流量呈显著下降趋势^[2-3],水资源时空分布不均等问题愈发突出。在生态建设背景下,实现黄河干支流泥沙通量的准确预测,对区域生态保护、水资源合理利用及经济可持续发展具有重要意义^[4]。

径流变化受降水、蒸发等多种水文过程的影响,呈现出非线性、非平稳性和多尺度等特征,这些特征限制了径流的准确预测^[5]。目前,径流预测方法大致可分为两类:物理模型和数据驱动模型。使用基于物理机制的水文模型进行径流模拟和预测需要大量的实测资料,这在许多地区难以实现^[6]。相比之下,数据驱动模型不考虑降雨径流过程的物理机制,适用性较强。其中,“分解-预测-重构”模式的模型得到广泛应用,相比于传统单一模型,该模型可有效提高预测精度^[7]。练继建等^[8]将经验模态分解(Empirical Mode Decomposition, EMD)与自回归模型(Autoregressive, AR)组合,发现模型能较好地模拟丹江口水库年径流序列,且预报精度符合要求。郑芳芳等^[9]利用改进的集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)耦合最近邻抽样回归模型(Nearest Neighbor Bootstrapping Regressive, NNBR)对屏山站年径流进行预测,并验证了模型的可行性。席东洁等^[10]基于 EMD-Elman 模型预测了黄河上游唐乃亥水文站的月径流,取得了较好的模拟效果。孟二浩^[11]将 EMD 和支持向量机(Support Vector Machine, SVM)模型有机结合,并应用于月径流序列的预测,模拟效果较单一模型显著提高。研究结果显示混合模型可有效提高径流预测精度。本文将 EEMD 与 SVM 耦合,建立 EEMD-SVM 混合模型,对北洛河上游吴旗水文站 1971—2014 年的月径流进行预测,并与 EEMD 耦合的差分自回归移动平均模型(Autoregressive Integrated Moving Average, ARIMA)和非线性自回归动态神经网络(Nonlinear Autoregressive, NAR)的预测效果进行对比分析,拟确立 EEMD-SVM 混合模型对径流时间序列的预测精度和适应性。

1 研究区概况

北洛河是渭河的一级支流,黄河的二级支流,流

域总面积 26 905 km²,其中上游吴旗水文站控制面积约 3 408 km²,属于典型的黄土丘陵沟壑区,地势西北高、东南低。研究区地处半干旱、半湿润向干旱气候过渡区域,具有温带大陆性季风气候特征。多年平均降雨量 454.9 mm,年内分配不均,集中在汛期(7—9 月)且多暴雨。土壤类型以黄绵土为主,另有黑垆土和灰褐土。流域沟壑纵横,水土流失严重。退耕还林(草)工程实施后,区域植被覆盖度显著提高,与退耕前相比,2010 年的输沙模数减少了约 90%^[12]。

2 数据与方法

2.1 数据来源

本研究选择北洛河上游吴旗水文观测站,1971—2014 年逐月实测径流量数据摘抄自水利部水文局《黄河流域水文年鉴》及《黄河泥沙公报》。数据序列经过严格校验与检查,确保其准确可靠。

2.2 研究方法

2.2.1 集合经验模态分解(EEMD) 经验模态分解(EMD)是 Huang 等^[13]提出的 Hilbert-Huang 变换(HHT)的核心算法。该算法是一种基于经验的自适应数据处理方法,可用于分析非线性和非平稳时间序列,具有较高的信噪比。假设原始信号中存在多种振幅的不同模式,EMD 可以将复杂信号逐级分解为频率由高到低的本征模态函数(IMF)和一个残余项。然而,EMD 方法存在“模式混叠”问题,Wu 等^[14]在 EMD 的基础上提出集合经验模态分解(EEMD)的方法,该方法在 EMD 分解过程中多次加入白噪声序列,使随机集成得到的白噪声序列与添加的白噪声序列相互抵消,减少了“模式混叠”出现的概率。增加的白噪声效果通过预设白噪声标准差(ϵ)和加入白噪声序列的数量(n)来调节。

2.2.2 支持向量机(SVM)及其回归算法 SVM 被认为是预测领域最有效的机器学习方法之一,它在很大程度上解决了传统人工神经网络等方法的模型选择与过学习问题、非线性和高维问题、局部极小点问题等,被广泛应用于各个领域^[15]。本研究将 SVM 用于径流预测的非线性回归问题,首先通过一个非线性映射函数把原样本空间映射到高维特征空间,然后在高维特征空间进行线性回归分析,得到回归函数为:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \times K(x, x_i) + b \quad (1)$$

式中: α_i 和 α_i^* 为拉格朗日乘子,对应的样本点就是支

持向量; $K(x; x_i)$ 为核函数; b 为常数。

采用核函数进行预测时, 其中的核参数(g)和惩罚因子(c)对预测精度至关重要。核参数的值太小或惩罚因子的值太大会导致训练集过学习, 反之则会造成欠学习现象^[16]。本研究采用网格法对核参数和惩罚因子进行优化。

2.2.3 差分自回归移动平均模型(ARIMA) ARIMA模型是一种经典的时间序列预测方法, 又称为 Box-Jenkins 模型^[17]。ARIMA 模型假设时间序列具有潜在平稳性, 通过差分将非平稳时间序列转化为平稳时间序列, 然后将因变量仅对它的滞后值及随机误差项的现值和滞后值进行回归, 描述时间序列发展规律。ARIMA 模型的建模过程可概括为以下 3 个步骤: (1) 检查平稳性; (2) 确定模型阶数; (3) 预测趋势并返回结果。

2.2.4 非线性自回归(NAR)动态神经网络 NAR 动态神经网络是一种特殊的人工神经网络, 它基于 NAR 非线性自回归模型, 以自身作为回归变量, 通过一段时间内的随机变量的线性组合来表示之后某一时刻的随机变量。NAR 动态神经网络一般由输入层、时滞层、隐藏层与输出层构成, 具有反馈与记忆功能。NAR 神经网络模型可以描述为:

$$y(t) = f[(y(t-1), y(t-2), \dots, y(t-d))] \quad (2)$$

式中: $y(t)$ 为当前时刻的输入值; $y(t-1), y(t-2), \dots, y(t-d)$ 为历史时刻的输出值; d 为延迟阶数。

3 月径流预测模型构建

首先利用 EEMD 方法对吴旗站 1971—2014 年的月径流时间序列进行分解, 本研究添加了 100 组白噪声序列到原始序列中, 每组噪声的幅值设置为 0.4, 即标准差 ϵ 和数量 n 分别设置为 0.4, 100, 可得到 8 个由高频到低频、变幅从大到小的 IMF 分量和 1 个残差项 R (图 1), 其中残差项表示序列的变化趋势。然后, 利用 SVM 模型分别对 IMF 和残差项 R 进行预测, 最后对每个 SVM 模型预测的结果求和, 获得最终的径流预测结果。EEMD-ARIMA 和 EEMD-NAR 模型的构建流程与 EEMD-SVM 类似。本文将前 39 a (1971 年 1 月—2009 年 12 月, 共 468 个数据) 的月径流数据作为模型的训练集, 将最近 5 a (2010 年 1 月—2014 年 12 月, 共 60 个数据) 作为模型的验证集。

4 模型结果对比分析

4.1 模型性能指标

选用决定系数(R^2)、平均绝对误差(MAE)、均方根误差(RMSE)以及纳什系数(NSEC)作为模型性能的评价指标, 计算公式如下:

$$R^2 = \frac{\sum_{i=1}^N (y_i - \bar{y}) \times (y_i^* - \bar{y}^*)}{\left[\sum_{i=1}^N (y_i - \bar{y})^2 \right]^{0.5} \times \left[\sum_{i=1}^N (y_i^* - \bar{y}^*)^2 \right]^{0.5}} \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - y_i^*| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2} \quad (5)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - y_i^*)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (6)$$

式中: N 为测试样本数; y_i 和 y_i^* 分别为样本的实测值和预测值; \bar{y} 和 \bar{y}^* 分别为样本实测值的平均值和样本预测值的平均值。

4.2 结果与分析

以 2010 年 1 月—2014 年 12 月为验证期, 将 EEMD-SVM, EEMD-ARIMA 和 EEMD-NAR 模型在验证期的月径流预测值与实际值进行比较 (图 2)。可以看出, EEMD-SVM 模型的模拟效果最好, 预测值与实测值最为接近。对模型的预测值与实测值进行线性拟合 (图 3), 发现 EEMD-SVM 模型的决定系数(R^2)达到了 0.986 ($p < 0.01$), 远高于 EEMD-ARIMA ($R^2 = 0.344$) 和 EEMD-NAR ($R^2 = 0.498$)。表 1 显示了 4 个指标对 3 个模型的预测性能评价结果。其中 EEMD-ARIMA 模型的预测值与实测值的平均绝对误差(MAE)为 219.575 m^3 , 均方根误差(RMSE)为 359.623 m^3 , 纳什系数(NSE)为 -0.282; EEMD-NAR 模型的预测值与实测值的平均绝对误差(MAE)为 199.864 m^3 , 均方根误差(RMSE)为 264.389 m^3 , 纳什系数(NSE)为 0.307; EEMD-SVM 模型的预测值与实测值的平均绝对误差(MAE)为 50.775 m^3 , 均方根误差(RMSE)为 64.648 m^3 , 纳什系数(NSEC)为 0.959。这些结果显示 EEMD-SVR 具有最低的平均绝对误差(MAE)和均方根误差(RMSE), 以及最高的决定系数(R^2)和纳什系数(NSE), 表明 EEMD-SVR 模型的模拟精度最高。EEMD-NAR 模型的预测性能高于 EEMD-ARIMA 模型。

基于以上分析和评价, EEMD-SVR 耦合模型可以成功地应用于北洛河上游的月径流预报。

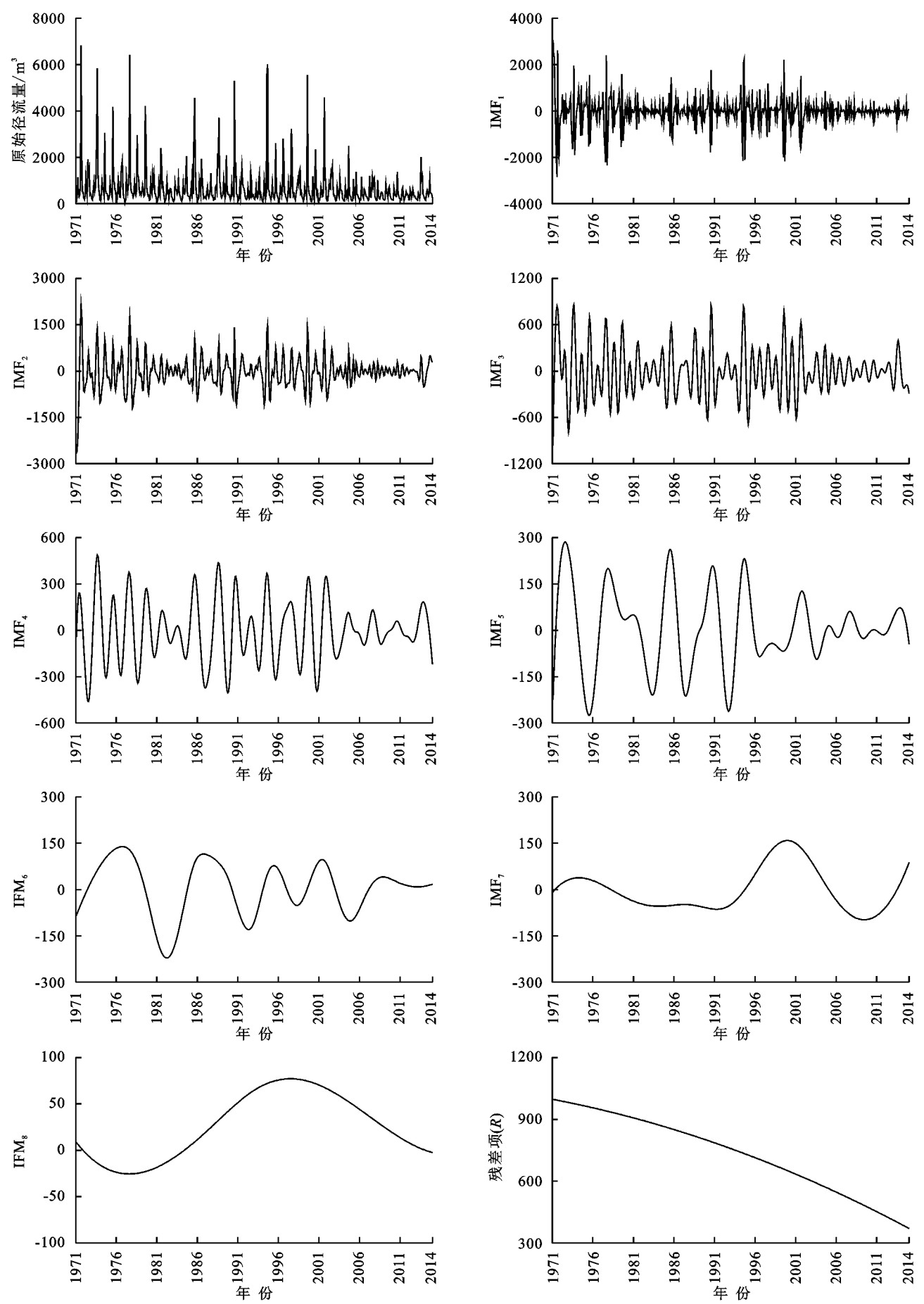


图 1 北洛河上游吴旗站 1971—2014 年的月径流 EEMD 分解结果

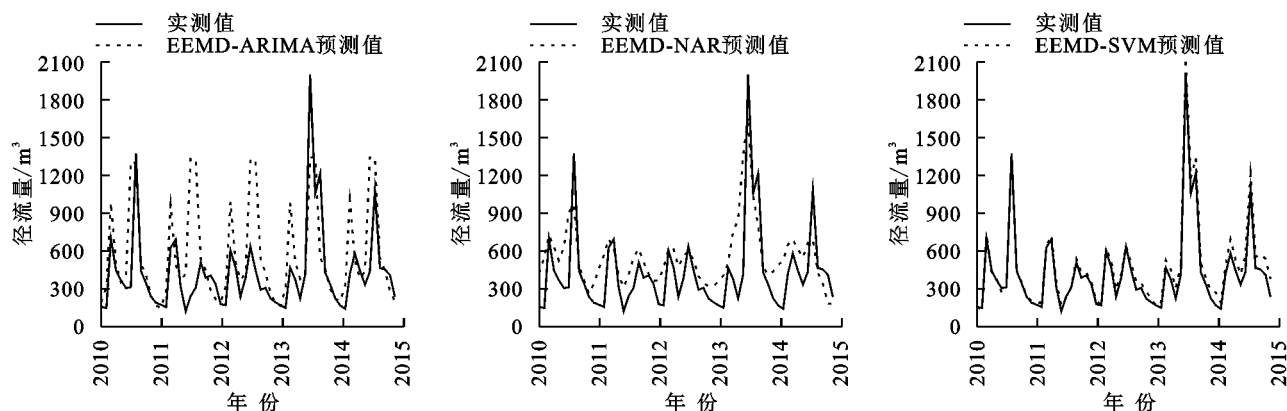


图 2 北洛河上游吴旗站 1971—2014 年的月径流实测值与模拟值对比

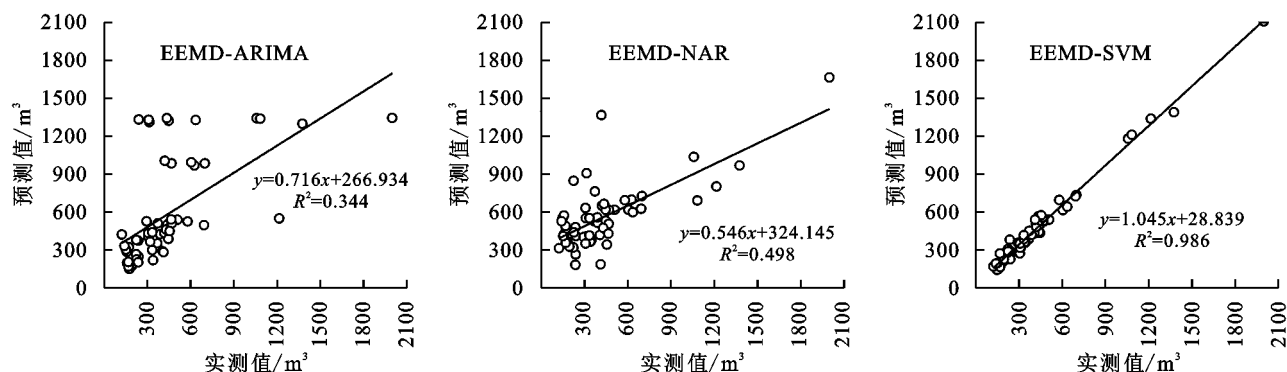


图 3 北洛河上游吴旗站 1971—2014 年的月径流实测值与模拟值拟合结果

表 1 模型的预测性能评价

| 模型 | R^2 | MAE/ m^3 | RMSE/ m^3 | NSE |
|------------|-------|-------------------|--------------------|--------|
| EEMD-ARIMA | 0.344 | 219.575 | 359.623 | -0.282 |
| EEMD-NAR | 0.498 | 199.864 | 264.389 | 0.307 |
| EEMD-SVM | 0.986 | 50.775 | 64.648 | 0.959 |

5 结论

非线性和非平稳性是径流时间序列的重要特征。为了提高径流预测的预测精度,利用 EEMD 方法去除非平稳信号中的噪声。基于 EEMD-SVM 耦合模型对径流序列进行“分解—预测—重构”,并将其与 EEMD-ARIMA 和 EEMD-NAR 两种耦合模型的预测结果进行对比。结果表明,EEMD-SVM 模型的平均绝对误差(MAE)和均方根误差(RMSE)最低,分别为 50.775,64.648 m^3 ,而决定系数(R^2)和纳什系数(NSE)最高,分别为 0.986,0.959。相比于 EEMD-ARIMA 和 EEMD-NAR 模型,EEMD-SVM 模型的决定系数(R^2)分别提高了 186.63%,49.49%。因此,在对径流序列进行 EEMD 分解的前提下,SVM 比传统时间序列模型(ARIMA)和人工神经网络模型(NAR)的预测精度更高,非线性拟合能力更强,可以成功地应用于北洛河上游的月径流预报。

参考文献:

[1] 丛晓红,拾兵,于西达,等.基于 PSO-BP 神经网络的黄河利

津站输沙量预测[J].人民黄河,2020,42(S1):1-3,8.

- [2] Gao P, Geissen V, Ritsema C J, et al. Impact of climate change and anthropogenic activities on stream flow and sediment discharge in the Wei River basin, China[J]. Hydrology and Earth System Sciences, 2013,17:961-972.
- [3] 李二辉,穆兴民,赵广举.1919—2010 年黄河上中游区径流量变化分析[J].水科学进展,2014,25(2):155-163.
- [4] 穆兴民,赵广举,高鹏,等.黄河未来输沙量态势及其适用性对策[J].水土保持通报,2020,40(5):328-332.
- [5] 赵益平,王文圣,刘浅奎.R/S-BP 网络耦合模型及其在年径流预测中的应用[J].华北水利水电大学学报:自然科学版,2019,40(4):40-45.
- [6] Noori N, Kalin L. Coupling SWAT and ANN models for enhanced daily streamflow prediction[J]. Journal of Hydrology, 2016,533:141-151.
- [7] 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.
- [8] 练继建,孙萧仲,马超,等.基于 EEMD-AR 模型的丹江口水库年径流随机模拟与预报[J].水利水电科技进展,2017,37(5):16-21.
- [9] 郑芳芳,王文圣,张岚婷.改进的 EEMD-NNBR 耦合模型在年径流预测中的应用[J].人民珠江,2021,42(2):1-6.
- [10] 席东洁,赵雪花,张永波,等.基于经验模态分解与 Elman 神经网络的月径流预测[J].中国农村水利水电,2017(7):112-115.

(下转第 109 页)

- 速的区域变化特征分析[J].水土保持通报,2011,31(3):40-43.
- [17] 师玮,周俊菊,胡子瑛,等.甘肃省陇东地区近55年风速时空变化特征[J].生态学杂志,2017,36(12):3594-3601.
- [18] 魏凤英.现代气候统计诊断与预测技术[M].2版.北京:气象出版社,2007.
- [19] IPCC. Climate Change 2013: The Physical Science Basis [M]. New York: Cambridge University Press, 2013.
- [20] 何毅,杨太保,陈杰,等.1960—2013年南北疆风速变化特征分析[J].干旱区地理,2015,38(2):249-259.
- [21] Xu Xingkui, Levy Jason K, Lin Zhaohui, et al. An investigation of sand-dust storm events and land surface characteristics in China using NOAA NDVI data[J]. Global and Planetary Change, 2006,52(1/4):182-196.
- [22] 毛东雷,蔡富艳,杨余辉,等.新疆策勒4个典型下垫面近地层风速脉动特征[J].干旱区研究,2017,34(1):126-135.
- [23] 郭永强,王乃江,褚晓升,等.基于 Google Earth Engine 分析黄土高原植被覆盖变化及原因[J].中国环境科学,2019,39(11):4804-4811.
- [24] 宋永永,马蓓蓓,李武斌,等.黄土高原县域城镇化的地域分异与动因机制[J].陕西师范大学学报:自然科学版,2019,47(4):9-21.
- [25] 黄晓龙,杨雪,李涛,等.建站至2016年四川气象台站元数据变动特征分析[J].高原山地气象研究,2017,37(4):84-90.
- [26] 刘学锋,梁秀慧,任国玉,等.台站观测环境改变对我国近地面风速观测资料序列的影响[J].高原气象,2012,31(6):1645-1652.
- [27] 王恰.中国风电产业40年发展成就与展望[J].中国能源,2020,42(9):28-32,9.

(上接第97页)

- [20] 丁文峰,张平仓,任洪玉.近50年来嘉陵江流域径流泥沙演变规律及驱动因素定量分析[J].长江科学院院报,2008,25(3):23-27.
- [21] Gao P, Deng J C, Chai X K, et al. Dynamic sediment discharge in the Hekou-Longmen region of Yellow River and soil and water conservation implications[J]. Science of the Total Environment, 2017,578:56-66.
- [22] Hu B Q, Wang H J, Yang Z S, et al. Temporal and spatial variations of sediment rating curves in the Changjiang(Yangtze River) basin and their implications [J]. Quaternary International, 2011,230(1):34-43.
- [23] Searcy J K, Hardison C H. Double-Mass Curves[M]. Washington D C: United States Government Printing Office, 1960.
- [24] 穆兴民,张秀勤,高鹏,等.双累积曲线方法理论及在水文气象领域应用中应注意的问题[J].水文,2010,30(4):47-51.
- [25] Zhao G J, Mu X M, Jiao J Y, et al. Assessing response of sediment load variation to climate change and human activities with six different approaches[J]. Science of the Total Environment, 2018,639:773-784.
- [26] Zhang L, Dawes W R, Walker G R. Response of mean annual evapotranspiration to vegetation change at catchment scale[J]. Water Resources Research, 2001,37(3):701-708.
- [27] 张春林,凡炳文,刘国华.近58年来洮河流域水沙演变特征与驱动力分析[J].人民黄河,2014,36(8):10-14.
- [28] 牛最荣,崔亮.气候变化和人类活动对洮河输沙量变化的影响分析[J].水利水电技术,2016,47(5):80-83,88.

(上接第102页)

- [11] 孟二浩.基于改进的 EMD 混合模型月径流预测研究[D].西安:西安理工大学,2018.
- [12] Zhang X P, Lin P F, Chen H, et al. Understanding land use and cover change impacts on run-off and sediment load at flood events on the Loess Plateau, China [J]. Hydrological Processes, 2018,32(4):576-589.
- [13] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences, 1998,454(1971):903-995.
- [14] Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method [J]. Advances in Adaptive Data Analysis, 2009,1:1-41.
- [15] 高隽.人工神经网络原理及仿真实例[M].北京:机械工业出版社,2003.
- [16] 茆美琴,龚文剑,张榴晨,等.基于 EEMD-SVM 方法的光伏电站短期出力预测[J].中国电机工程学报,2013,33(34):17-24,15.
- [17] Box G E P, Jenkins G M, Reinsel G C, et al. Time Series Analysis: Forecasting and Control[M]. John Wiley & Sons, 2015.