

# 考虑气象因素和BP神经网络的中长期径流预报

王伟<sup>1</sup>, 金新峰<sup>1\*</sup>, 尹华政<sup>2</sup>, 刘洋<sup>1</sup>, 余豪<sup>1</sup>, 胡朋瑞<sup>1</sup>

<sup>1</sup>中国电建集团中南勘测设计研究院有限公司, 湖南 长沙

<sup>2</sup>水电水利规划设计总院, 北京

Email: \*258988544@qq.com

收稿日期: 2021年4月9日; 录用日期: 2021年8月23日; 发布日期: 2021年8月31日

## 摘要

中长期径流预报结果对中长期发电计划制定至关重要, 为了提高中长期径流预报精度, 提出了考虑气象因素与BP神经网络相结合的预报方法。首先根据偏互信息法获得影响中长期入库流量变化的关键因子, 并根据模糊聚类方式对其影响径流的相关因子进行聚类分析, 最后采用BP神经网络以聚类的因子为输入对中长期日平均径流进行预测。以湖南省凤滩水库为例, 预测结果表明: 1) BP神经网络预测模型预报效果优于支持向量基、时间序列预测模型的预报结果; 2) 考虑气象因素的径流预报优于未采用气象相似相的径流预报; 3) 采用气象相似性与BP神经网络相结合的径流预报效果较好, 预报精度均在丙级以上。

## 关键词

径流预报, 中长期, BP神经网络

# Medium and Long Term Runoff Forecast Considering Meteorological Factors and BP Neural Network

Wei Wang<sup>1</sup>, Xinfeng Jin<sup>1\*</sup>, Huazheng Yin<sup>2</sup>, Yang Liu<sup>1</sup>, Hao Yu<sup>1</sup>, Pengrui Hu<sup>1</sup>

<sup>1</sup>PowerChina Zhongnan Engineering Corporation Limited, Changsha Hunan

<sup>2</sup>China Renewable Energy Engineering Institute, Beijing

Email: \*258988544@qq.com

Received: Apr. 9<sup>th</sup>, 2021; accepted: Aug. 23<sup>rd</sup>, 2021; published: Aug. 31<sup>st</sup>, 2021

作者简介: 王伟, 出生于 1984 年 7 月, 硕士研究生, 高级工程师, 研究方向为水文算法研究、水电站水库调度、水环境等。  
\*通讯作者。

文章引用: 王伟, 金新峰, 尹华政, 刘洋, 余豪, 胡朋瑞. 考虑气象因素和 BP 神经网络的中长期径流预报[J]. 水资源研究, 2021, 10(4): 391-398. DOI: 10.12677/jwrr.2021.104042

## Abstract

The results of medium and long-term runoff forecast are very important for medium and long-term power generation planning. In order to improve the accuracy of medium and long-term runoff forecast, a forecast method combining meteorological factors with BP neural network is proposed. Firstly, the key factors affecting the medium and long-term inflow are obtained by partial mutual information method, and the related factors affecting runoff are analyzed by fuzzy clustering method. Finally, the BP neural network is used to predict the medium and long-term daily average runoff with the clustering factors as input. Taking Fengtan reservoir in Hunan Province as an example, the prediction results show that: 1) The prediction effect of BP neural network prediction model is better than that of support vector basis and time series prediction model; 2) The runoff prediction considering meteorological factors is better than that without meteorological similarity; 3) The runoff prediction effect of combining meteorological similarity and BP neural network is better, and the prediction accuracy is in the third above grade.

## Keywords

Runoff Forecast, Medium and Long Term, BP Neural Network

Copyright © 2021 by author(s) and Wuhan University.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

中长期径流预报的预见期通常为 3 天以上, 1 年以内。与短期径流预报相比, 中长期径流预报对研判水库中长期水情变化, 制定更加科学的防洪调度、发电调度方案具有重要指导意义[1]。随着区域水资源的统一调度管理的加强, 对中长期径流预报精度要求也越来越高[2]。受气象因素、下垫面等多种因素的影响, 径流序列呈现出随机性与不确定性, 而中长期优化调度结果十分依赖径流预报精度, 因此寻求合适的方法对中长期入库径流进行准确预报意义重大。

根据预报机理不同, 中长期径流预报可分为基于模型驱动的机理分析模型和数据分析的数据驱动模型两大类。前者将气象预报信息输入到具有产汇流机制的水文预报模型中进行径流预报, 然后受前期影响雨量与气象预报精度的影响, 预报精度有待提高。后者为基于数据分析的数据驱动模型, 通过构建径流与相关影响因子的统计模型, 以影响因子的实测值作为输入, 从而开展中长期径流预报。随着计算机技术的发展与普及, 基于数据分析的数据驱动模型逐渐成为中长期径流预报研究的重点。基于时间序列分析的数据驱动模型(如滑动平均自动回归模型等)以及其改进策略最早应用与中长期径流预报[3] [4], 之后神经网络、支持向量机等模型广泛应用于中长期径流预报, 并取得了较为精确的预报结果[5] [6]。

基于数据驱动的中长期径流预报, 大都以实测历史径流作为模型的输入, 较少的考虑中长期气象预报对径流预报的影响, 随着数值天气预报技术与计算机技术的发展, 中长期气象预报精度越来越高。因此在中长期径流预报过程中很有必要考虑气象预报因素对中长期径流预报精度的影响[7]。因此本文提出一种基于气象相似性和 BP 神经网络相结合的中长期径流预报方法, 首先基于信息熵因子筛选法获得影响中长期入库流量变化的关键因子, 并对各个因子权重进行校核后模糊聚类分析, 再将聚类后的因子集输入到 BP 神经网络中进而对中长期径流进行预报, 最后开展一个月的中长期径流预报, 并与支持向量机、滑动平均自动回归模型预报结果相比验证了所提方法的有效性。

## 2. 求解方法与策略

### 2.1. 偏互信息法

偏互信息法(partial mutual information, PMI), 既能度量输入变量与输出结果的线性与非线性关系, 也可消除互信息的耦合对预报结果的影响, 从而保证了变量选择的可靠性与准确性[8] [9]。针对中长期径流预报影响因子筛选来说, 偏互信息能有效剔除已选预测因子对新加入预测因子的影响, 能有效计算出新加入的预测因子(特征量)与预测流量的相关性。偏互信息可按下述策略进行求解。

$$PMI = \frac{1}{N} \ln \left[ \frac{f_{x',y'}(x'_i, y'_i)}{f_{x'}(x'_i) f_{y'}(y'_i)} \right] \quad (1)$$

$$x' = x - E[x|B] \quad (2)$$

$$y' = y - E[y|B] \quad (3)$$

式中:  $f_x(x_i)$ 与  $f_y(y_i)$ 分别表示随机变量  $X$ 、 $Y$  的边缘概率密度函数;  $f_{x,y}(x_i, y_i)$ 表示  $X$ 、 $Y$  的联合概率密度函数;  $E$  为随机变量的数学期望;  $x$  为选择的输入变量;  $y$  为需要预报值;  $B$  表示选择的随机变量集合;  $x$ 、 $y$  分别表示排除  $B$  影响的  $x$  与  $y$  的残差值。

针对偏互信息中随机变量的边缘密度分布函数和联合分布密度函数估计, 采用高斯函数作为核函数估计样本的概率密度[10]。

$$\hat{f}_x(x_i) = \frac{1}{N} \sum_{i=1}^N \frac{\exp\left(-\left(x_i - x_j\right)^T S^{-1} \left(x_i - x_j\right)\right)}{2\lambda^2 (2\pi)^{d/2} \lambda^d (\det S)^{1/2}} \quad (4)$$

式中:  $f$  表示随机变量  $X$  在  $x_i$  处的密度函数估计值;  $d$  为  $x$  的维数;  $S$  为随机变量  $X$  的协方差矩阵;  $\det S$  为  $S$  的行列式;  $\lambda$  为核密度估计的窗口宽度, 根据经验[11],  $\lambda$  可表示为:

$$\lambda = \left( \frac{4}{d+2} \right)^{1/(4+d)} N^{-1/(d+4)} \quad (5)$$

偏互信息的显著性检验标准可按 Hampel 显著性检验, 从而确定选择边界[12]。计算公式为:

$$L_x = \frac{M_x}{1.483M_x^{50}} \quad (6)$$

$$M_x = |PMI - PMI^{50}| \quad (7)$$

式中:  $M_x^{50}$  与  $PMI^{50}$  分别表示为  $M_x$  与  $PMI$  的中位数;  $L_x$  为显著性边界对于备选因子  $x$ , 若  $L_x > 3$  则认为备选输入因子与预报对象显著相关。

### 2.2. 聚类分析

聚类分析的基本原理是认为所研究的样本的影响因子之间存在相似性, 通过一定的手段和方法, 将数据分类到不同的类或者簇这样的一个过程, 使得相同类或簇间差异尽可能的小, 不同类或簇间差异尽可能大的方法。聚类分析具有广泛的实用价值, 然而实际过程中, 分类往往具有模糊性, 分类并不是两者是否有关, 而是关系的深浅, 故采用模糊数学语言更加贴合实际。因此本文采用模糊数学的方法确定样本的亲疏程度, 从而实现样本分类。假设共有  $n$  个影响因子, 影响中长期径流预报结果, 并把  $M_i = [m_1, m_2, \dots, m_n]$  记为第  $i$  预测阶段的特征向量。为了描述任何两个阶段的相似程度, 引入了相似度的概念, 设共有  $i$ 、 $j$  两个阶段, 其因素量化指标  $M_i$ 、 $M_j$  则这两个阶段的相似性为:

$$R_{i,j} = \frac{\sum_{k=1}^n m_{ik} m_{jk}}{\sqrt{\sum_{k=1}^n m_{ik}^2 \sum_{k=1}^n m_{jk}^2}} \quad (8)$$

通常来讲,  $i, j$  两个阶段影响因素越接近, 相似度  $R_{i,j}$  就越大。

### 2.3. BP 神经网络

BP 神经网络是按照误差逆传播算法训练的多层前置网络, 能根据误差值不断调整网络的权重和阈值, 从而使预测值逼近实际值。BP 神经网络具有三层结构, 具有很强的映射能力, 被广泛应用于水文预报中。详情参考文献[13] [14]。

### 2.4. 评价指标

为评价中长期流量预测结果的合理性, 引入归一化的平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)、均方根误差(Root Mean Squared Error, RMSE)、平均百分比误差(Mean Absolute Percentage Errors, MAPE)与确定性系数(Deterministic Coefficient, DC)、合格率(Qualified Rate, QE)对预测结果进行评价。根据《水文情报预测规范》(GB/T 22482-2008)规定, 中长期径流预报误差小于 20%为合格。依据预报等级评价表, 对预报精度进行进一步评价。

## 3. 应用分析

以我国南方地区西水流域凤滩水库为例, 验证所提方法的有效性。该流域面积为 17,500 km<sup>2</sup>, 处于亚热带季风气候, 平均温度 17.6 摄氏度。降雨主要集中在 3~6 月, 约占全年的 60%, 雨量充沛, 降雨集中, 中长期径流预报难度较大。流域内的预报降雨与蒸发数据根据流域内的 27 个气象站, 采用泰森多边形计算得到。本次计算选取 2011 年 1 月~2020 年 12 月共计 10 年的数据实测径流与气象预报数据进行研究对未来 15 日的平均径流进行预报。

### 3.1. 预测因子选取

采用偏互信息法对备选输入变量的遴选分为两个阶段。首先确定可能影响径流的关键影响因子, 经过分析预报降雨、降雨时间、蒸发等信息可能影响中长期日平均径流预报结果, 根据经验备选的影响因子如表 1 所示。在此基础上将上一阶段选出的输入变量重新组合备选输入变量, 再次采用偏互信息法剔除相互影响的冗余变量, 计算结果如表 2 所示,  $\bar{p}$ 、 $\overline{p_6}$ 、 $t_6$ 、 $\overline{R_{15}}$  的  $L_x$  均大于 3, 可认为其对中长期日平均径流预报结果影响较大。最终计算结果见表 2。

**Table 1.** Candidate factors affecting long-term runoff process change

**表 1.** 中长期径流过程变化影响候选因子

候选因子	说明	候选因子	说明
$\bar{p}$	逐日平均降雨	$\bar{T}$	内逐日平均温度
$\overline{p_3}$	逐日 3 小时最大降雨量	$w$	逐日最大风力等级
$t_3$	逐日 3 小时最大降雨量的起始时间	$\overline{R_{15}}$	预测前 15 日逐日平均径流
$\overline{p_6}$	逐日 6 小时最大降雨量	$\overline{R_{30}}$	预测前 30 日逐日平均径流
$t_6$	逐日 6 小时最大降雨量的起始时间	$\bar{E}$	逐日平均蒸发量

**Table 2.** Final selection results of input variables by partial mutual information method in this calculation  
**表 2.** 偏互信息法对输入变量的最终遴选结果

候选因子	$L_x$ 计算结果	候选因子	$L_x$ 计算结果
$\bar{p}$	32.15	$t_6$	28.13
$\bar{p}_6$	47.16	$\bar{R}_{15}$	5.62

### 3.2. 聚类分析与权重校核

在聚类分析前为了使中长期径流预报结果具有代表性, 分别在汛期、过渡期、枯水期进行权重率定。首先确定初始解, 由于逐日平均降雨量、逐日 6 小时最大降雨量、逐日 6 小时最大降雨量的起始时间、预测前 15 日逐日平均径流对径流预报影响较大, 其初始解的范围均设置为 $[0.5, 1.0)$ 。按逐日平均降雨量、逐日 6 小时最大降雨量、逐日 6 小时最大降雨量的起始时间、预测前 15 日逐日平均径流特征量的顺序采用逐步优化的方式分别对其权重进行校核, 给定精度  $\varepsilon = 0.01$ 。表 3 给出了最优预报结果的权重值。

**Table 3.** Optimal weight of each eigenvalue  
**表 3.** 各特征值最优权重

项目	$\bar{p}$	$\bar{p}_6$	$t_6$	$\bar{R}_{15}$
丰水期	0.32	0.27	0.22	0.19
平水期	0.35	0.25	0.19	0.21
枯水期	0.36	0.24	0.17	0.23

### 3.3. 预测结果分析

训练集取 2011 年 1 月~2016 年 12 月的实测径流数据与气象预报数据, 验证集取 2017 年 1 月~2018 年 12 月的实测径流数据与气象预报数据, 测试集选取 2019 年 1 月~2020 年 12 月的实测径流数据与气象预报数据。为验证所提方法的有效性, 取 2020 年 12 月份测试集数据进行验证, 三种模型的逐日滚动预测效果如表 4 所示。

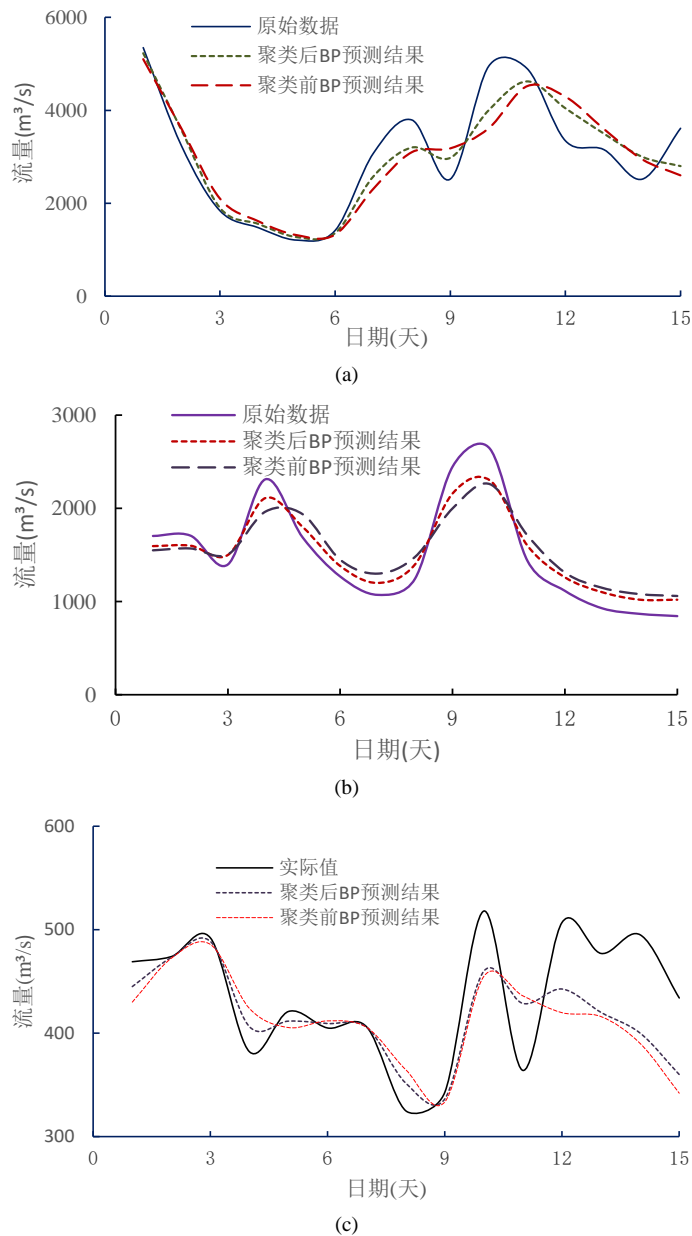
**Table 4.** Comparison between the prediction results of each model and the actual situation  
**表 4.** 各模型预报结果与实际对比

模型	RMSE	MAE	MAPE	最大值偏差	最小值偏差
BP	9.76%	1.78%	6.64%	-26	7
支持向量基	13.03%	2.61%	11.20%	-33	16
数据驱动	18.20%	3.82%	14.20%	-52	21

注: 最大值差、最小值差: 表示径流处于最大值、最小值时预测值与实际值之差。

从表 4 中可以直观看出, BP 神经网络的预测效果最好, 这是由于 BP 神经网络的容错能力与泛化能力较强。采用 BP 神经网络的 RMSE 为 9.76%, 小于支持向量基的 13.03%、数据驱动模型的 18.20%。采用 BP 网络预测最大值与最小值预测效果均优于支持向量机、数据驱动模型, 说明 BP 预测结果更能反应预测曲线拐点变化情况, 更接近真实预测效果。

为了进一步验证聚类前后, BP 神经网络的预测效果, 分别选取丰水期( $A_1$ )、平水期( $A_2$ )、枯水期( $A_3$ )未来 15 日逐日径流预报, 选取最优的 15 日径流预报结果, 如图 1 所示:



**Figure 1.** (a) Forecast results of wet season; (b) Forecast results of normal water period; (c) Forecast results in low water period  
**图 1.** (a) 丰水期预报结果; (b) 平水期预报结果; (c) 枯水期预报结果

**Table 5.** Prediction results of  $A_1, A_2, A_3$  BP before and after clustering  
**表 5.** 聚类前后  $A_1, A_2, A_3$  BP 预测结果

时段	聚类前 BP 预测结果			聚类后 BP 预测结果		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
$A_1$	20.52%	5.30%	16.42%	15.45%	3.99%	11.40%
$A_2$	8.24%	2.12%	16.79%	5.72%	1.47%	11.63%
$A_3$	1.78%	0.45%	9.71%	1.48%	0.38%	7.76%

根据图 1, 对时段  $A_1$ ,  $A_2$ ,  $A_3$  三个时段分析, 可以看出, 采用本文提出的方法, 预测得到的日平均明显比只采用 BP 神经网络预测效果, 预测曲线更加接近真实值。其原因是聚类后增加的 BP 神经模型的泛化能力, 使得预测效果更好, 具体结果见表 5。本章采用的方法充分的利用历史资料与气象预报资料, 对 15 日平均径流进行预测, 预测精度相对较高。

**Table 6.** Qualified rate and certainty coefficient of BP prediction results before and after clustering  
**表 6.** 聚类前后 BP 预测结果合格率与确定性系数

模型	丰水期		平水期		枯水期	
	合格率	确定性系数	合格率	确定性系数	合格率	确定性系数
聚类前 BP	80%	0.86	86.67%	0.89	86.67%	0.90
聚类后 BP	86.67%	0.91	93.33%	0.93	100%	0.94

由表 6 可以看出, 丰水期、平水期、枯水期聚类后采用 BP 网络预测合格率为 86.67%、93.33%、100%, 确定性系数为 0.91、0.93、0.94 预测等级达到甲级标准。聚类后采用神经网络对日平均径流预测结果的合格率、确定性系数均高于聚类前。

## 基金项目

特别感谢“水电工程水文气象重大关键技术应用研究课题一——水文气象及流域水电开发大数据平台研究设计”提供的资金支持。

## 参考文献

- [1] 李伶杰, 王银堂, 胡庆芳, 等. 基于随机森林与支持向量机的水库长期径流预报[J]. 水利水运工程学报, 2020(4): 33-40. LI Lingjie, WANG Yintang, HU Qingfang, et al. Long term runoff forecast of reservoir based on random forest and support vector machine. Journal of Water Conservancy and Water Transport Engineering, 2020(4): 33-40. (in Chinese)
- [2] 晋健, 刘育, 王琴慧, 等. 基于小波去噪和 FA-SVM 的中长期径流预报[J]. 人民长江, 2020, 51(9): 67-72. JIN Jian, LIU Yu, WANG Qinhui, et al. Medium and long term runoff forecasting based on wavelet denoising and FA-SVM. Renmin Changjiang River, 2020, 51(9): 67-72. (in Chinese)
- [3] CARLSON, R. F., MACCORMICK, A. J. A. and WATTS, D. G. Application of linear random models to four annual stream-flow series. Water Resources Research, 1970, 6(4): 1070-1078. <https://doi.org/10.1029/WR006i004p01070>
- [4] 谢帅, 黄跃飞, 李铁键, 刘朝云, 王建华. LASSO 回归和支持向量回归耦合的中长期径流预报[J]. 应用基础与工程科学学报, 2018, 26(4): 709-722. XIE Shuai, HUANG Yuefei, LI Tiejian, LIU Chaoyun and WANG Jianhua. Medium and long term runoff forecasting based on the coupling of LASSO regression and support vector regression. Journal of Applied Basic and Engineering Sciences, 2018, 26(4): 709-722. (in Chinese)
- [5] KALTEH, A. M. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Computers & Geosciences, 2013, 54(4): 1-8. <https://doi.org/10.1016/j.cageo.2012.11.015>
- [6] CHU, H., WEI, J., LI, T., et al. Application of support vector regression for mid- and long-term runoff forecasting in “Yellow River Headwater” region. Procedia Engineering, 2016(54): 1251-1257. <https://doi.org/10.1016/j.proeng.2016.07.452>
- [7] 杨强. 耦合 CFS 集合预报和深度学习的中长期径流预测方法研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2019. YANG Qiang. Research on medium and long term runoff forecasting method coupled with CFS ensemble forecasting and deep learning. Dalian: Dalian University of Technology, 2019. (in Chinese)
- [8] HE, J., VALEO, C., CHU, A., et al. Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection. Journal of Hydrology, 2011, 400(1-2): 10-23. <https://doi.org/10.1016/j.jhydrol.2011.01.024>
- [9] MAY, R. J., DANDY, G. C., MAIER, H. R., et al. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Environmental Modelling & Software, 2008, 23(10-11): 1289-1299. <https://doi.org/10.1016/j.envsoft.2008.03.008>
- [10] 刘本希, 廖胜利, 冯仲恺, 等. 耦合偏互信息的贫资料地区小水电发电能力预测[J]. 电力系统自动化, 2015, 39(19): 149-154.

- LIU Benxi, LIAO Shengli, FENG Zhongkai, et al. Prediction of small hydropower generation capacity in poor data area based on coupling partial mutual information. *Power System Automation*, 2015, 39(19): 149-154. (in Chinese)
- [11] SCOTT, D. W. *Multivariate density estimation: Theory, practice, and visualization*. Indian Apolis: Wiley, 2008.
- [12] MAY, R. J., MAIER, H. R., DANDY, G. C., et al. Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, 2008, 23(10-11): 1312-1326.  
<https://doi.org/10.1016/j.envsoft.2008.03.007>
- [13] 周雨婷, 张善亮, 邬雪松, 等. 基于改进 BP 神经网络的非平稳水文时间序列长期预报研究[C]//河海大学. 2020 年(第八届)中国水利信息化技术论坛论文集: 2020 年卷. 北京: 北京沃特咨询有限公司, 2020: 8.  
ZHOU Yuting, ZHANG Shanliang, WU Xuesong, et al. Study on long-term forecasting of non-stationary hydrological time series based on improved BP neural network//Hehai University. *Proceedings of the 8th China Water Conservancy Information Technology Forum: 2020*. Beijing: Beijing Water Resources Consulting Co., Ltd., 2020: 8. (in Chinese)
- [14] 朱静思, 侯天宇, 张治倩. 海河流域逐月降水预测研究[C]//中国气象学会. 第 35 届中国气象学会年会 S10 水文气象灾害形成机理、预报预测预警与风险评估新技术: 2018 年卷. 中国气象学会, 2018: 14.  
ZHU Jingsi, HOU Tianyu and ZHANG Zhiqian. Research on monthly precipitation prediction in Haihe River Basin//China Meteorological Society. *The 35th Annual Meeting of China Meteorological Society. S10 New Technologies of Hydrometeorological Disaster Formation Mechanism, Prediction, Early Warning and Risk Assessment: 2018*. China Meteorological Society, 2018: 14. (in Chinese)