

DOI:10.3880/j.issn.1004-6933.2022.04.019

长江上游水库入库流量的中长期预报

张 轩^{1,2}, 张行南^{2,3}, 王高旭^{1,2}, 吴 巍^{1,2}, 许 怡^{1,2}

(1. 南京水利科学研究院水文水资源研究所, 江苏 南京 210029; 2. 水文水资源与水利工程科学国家重点实验室, 江苏 南京 210029; 3. 河海大学水文水资源学院, 江苏 南京 210098)

摘要:为分析数理统计模型与机器学习模型在中长期径流预报中的特点与适用性, 挑选逐步回归与随机森林两种方法构建入库流量中长期预报模型, 以气象因子的物理机制为基础, 结合单相关系数及随机森林重要性分析识别关键气象因子并输入模型。利用长江上游乌东德、瀑布沟两个水库1959—1998年的入库流量训练了模型, 并且预测了两个水库1999—2014年的入库流量。结果表明: 两种模型的训练效果良好, 稳定性强, 随机森林的预测结果比逐步回归的精度高, 但精度的差距较小; 随机森林能减少预测因子值的异常变化带来的拟合误差, 但过拟合问题更为明显。

关键词:入库流量; 中长期预报; 逐步回归; 随机森林; 气象因子; 乌东德水库; 瀑布沟水库; 长江上游

中图分类号: P338⁺.2 **文献标志码:** A **文章编号:** 1004-6933(2022)04-0131-06

Medium and long term forecast of reservoir inflow in upper reaches of the Yangtze River // ZHANG Xuan^{1,2}, ZHANG Xingnan^{2,3}, WANG Gaoxu^{1,2}, WU Wei^{1,2}, XU Yi^{1,2} (1. *Hydrology and Water Resources Department, Nanjing Hydraulic Research Institute, Nanjing 210029, China*; 2. *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing 210029, China*; 3. *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*)

Abstract: In order to analyze the characteristics and applicability of the mathematical statistical model and the machine learning model in medium and long term runoff forecast, stepwise regression and random forest were selected to build a medium and long term forecast model. Based on the physical mechanism of meteorological factors, combined with single correlation coefficient and random forest importance analysis, key meteorological factors were identified and input into the model. The model was trained based on the inflow runoff of Wudongde and Pubugou reservoirs in the upper reaches of the Yangtze River from 1959 to 1998, and the inflow runoff of the two reservoirs from 1999 to 2014 was predicted. The results show that the two models have good training effect and strong stability. The prediction result of random forest is higher than that of stepwise regression, but the difference of accuracy is small. Random forest can reduce the fitting error caused by the abnormal change of predictor value, but the overfitting problem is more obvious.

Key words: reservoir inflow; medium and long term forecast; stepwise regression; random forest; meteorological factors; Wudongde Reservoir; Pubugou Reservoir; upper reaches of the Yangtze River

中长期预报技术可以服务于水库复杂的工作任务, 对防汛抗旱、科学调度、蓄水发电、航道运输有着重大意义。国内外关于中长期预报方法的研究很多, 从诸多气象因子中选取合适的因子集合, 建立预报模型训练流域径流, 是目前中长期预报的主流方法之一^[1]。该类方法常见的模型有数理统计模型以及近年出现的机器学习模型, 数理统计模型较早

应用于径流预报, 主要有多元线性回归^[2-3]、逐步回归^[4]等模型; 机器学习模型应用较晚, 但是发展迅速, 以人工神经网络^[5]、支持向量机^[6]、随机森林^[7]为代表的模型在中长期预报中都取得了较好的效果。赵鹏雁等^[8]基于最近邻抽样回归模型、周期叠加模型、逐步回归模型和神经网络模型等10种模型建立了澜沧江流域小湾水库月径流预报方案; 陶思

基金项目: 国家重点研发计划(2016YFA0601703, 2016YFC0401005); 国家自然科学基金(91847301, 92047203, 52009080, 42075191)

作者简介: 张轩(1996—), 男, 助理工程师, 硕士, 主要从事水资源管理与水利信息化研究。E-mail: xzhang@nhri.cn

通信作者: 张行南(1960—), 男, 教授, 博士, 主要从事水文物理规律模拟研究。E-mail: zxn@hhu.edu.cn

铭等^[9]采用长短期记忆神经网络,构建了未来 30 d 的逐日流量预报模型,预报效果良好;谢帅等^[10]引入 Lasso 回归方法筛选径流预报因子,以其选择结果作为支持向量回归模型中的预报因子,构建了 Lasso 回归和支持向量回归耦合(Lasso and support vector regression,LSVR)模型,并应用到龙羊峡水库进行月入库径流预报。综上,中长期径流预报中模型的实际应用已经非常成熟,然而对比于传统的模型,机器学习模型有什么优势与缺点,适用性与稳定性如何,仍然需要进一步探索。在数理统计方法中,逐步回归法可以按照自变量对因变量的作用大小,剔除不显著的变量,在预报因子优选方面有独特的优势^[11];而与其他机器学习模型相比,随机森林模型训练速度快^[12],对噪声及异常值的容忍性较强,对缺失数据和非平衡的数据的预测更为稳健。这两种模型在各自领域具有一定的代表性和优势性,故对这两种模型的特点和适用性进行研究具有一定意义。

1 研究区概况

选择长江上游流域作为研究区,为了更好地验证结果,选取长江干流重要水库乌东德水库与岷江的瀑布沟水库作为研究对象(图 1)。乌东德水库(101°14'E ~ 103°3'E, 25°20'N ~ 27°12'N)位于四川省会东县和云南省禄劝县交界处的金沙江峡谷内,属于低纬度高原季风气候,气候特征为高温、干旱、少雨;水库上距观音岩水电站 253 km,下距白鹤滩大坝 180 km,坝址下至宜宾河道长 562 km,控制集水面积 406 142 km²;水库正常蓄水位 950 m;死水位 920 m,相应库容为 33.06 亿 m³,校核洪水位 959.79 m,相应库容为 42.18 亿 m³。乌东德水电站是流域开发的重要梯级工程,有一定的防洪、航运和拦沙作用。瀑布沟水库(102°21'E ~ 102°49'E, 29°14'N ~ 29°24'N)位于岷江支流大渡河的下段,在四川省雅安市汉源县和凉山彝族自治州甘洛县境内,属于干热河谷气候;水库实行夏季放水、冬季蓄水的水

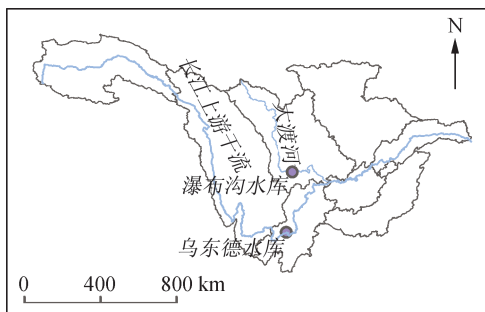


图 1 研究区域
Fig. 1 Study area

位调节模式,是一座以发电为主,兼有防洪、拦沙等综合利用效益的大型水电工程;瀑布沟水库坝址以上控制流域面积 68 512 km²,水库正常蓄水位 850.00 m,汛限水位 841.00 m,死水位 790.00 m,总库容 53.90 亿 m³,从历史洪水资料分析得知,瀑布沟最大日均流量多出现在 7 月、8 月下旬和 9 月上旬。

2 研究方法

选取逐步回归与随机森林两种方法,基于气象因子影响的物理机制,结合数理统计分析以及随机森林重要性分析,筛选关键气象因子作为模型输入,对长江上游的乌东德、瀑布沟两个水库 1959—1998 年的入库流量进行模型训练,验证两个水库 1999—2014 年的入库流量。

2.1 逐步回归法

逐步回归法在建立回归方程时,需定义一个指标用来衡量因子的重要程度,再依次逐步挑选预报因子。在所有可供挑选的因子之中,被选中的因子需要使残差平方和下降达到最大值。首先根据方差贡献大小挑选预报因子 x_1 ,再根据剩余因子的方差贡献大小挑选出第二个因子 x_2 ,第一、二步过渡方程分别为

$$y = b_0 + b_1 x_1 \quad (1)$$

$$y = b_0 + b_1 x_1 + b_2 x_2 \quad (2)$$

式中 b_0 、 b_1 、 b_2 为回归系数。

用同样的方法继续挑选因子进入方程,随着因子的逐渐增加,因子之间的共线性关系使得后面的因子的显著性越来越小。因此,在逐步回归的每一步都要对因子作剔除检验,直至方程 F 检验不能通过为止,最终得到的方程即为拟合最佳的回归方程。

2.2 随机森林算法

随机森林是一种集成学习算法^[13],该算法主要思路是通过某种策略综合大量的子预报模型的结果。随机森林就是以决策树模型为子模型,基于袋装法^[14-15]构造而成。随机森林在以决策树为子学习器的前提下,在训练过程中进一步改进了属性特征的随机选择^[16]。传统决策树在划分属性特征时,在当前节点的属性集合中选择最优的属性;而在随机森林中,需要先从事决策树的每个结点的属性集合中随机选择一个包含若干个属性的子集,再从该子集中选择最优属性用于决策树的划分。随机森林中子学习器除了来自袋装法的样本扰动外,还有自身的属性扰动,因此随着学习时间的增加,泛化误差会逐渐减小,集成学习的泛化性能也因此得到进一步的提升^[17]。随机森林可以计算出特征为随机森

林的每棵树的贡献,取均值后对不同特征的贡献进行排序。因此,该模型除了可以挖掘众多解释因子与目标变量之间的关系,亦可评价解释因子的重要程度,为挑选预报因子提供了新的方法。

2.3 预报方案

a. 关键预报因子筛选方案。乌东德水库和瀑布沟水库都位于长江上游流域,但位置距离较远,下垫面与气候条件不同,因此挑选的预报因子不同,从而避免因为预报因子相似而造成结果的偶然性。本文收集整理了130项气象因子数据,包括88项大气环流指数、26项海温指数以及16项其他指数。数据来源于中国气象局国家气候中心(<https://cmdp.ncc-cma.net/Monitoring/cn-index-130.php>),国家气象局已核查检验了数据系列,数据的真实性与可靠性较高。本文选用1958—2013年的130项气象因子,通过查阅大量文献^[18-21],结合已有研究,就130项气象因子逐一分析其具体的物理机制,排除无显著影响的因子,实现对各项预报因子的初步筛选。以初步筛选的因子为基础,结合单相关系数分析与随机森林的重要性分析进行关键预报因子的识别,据此对各水库逐月径流量的关键预报因子进行筛选,最终2个水库每月选取重要程度前5的关键气象因子作为该水库月径流量的预报因子。采用相关系数(r)进行相关性检验,利用 t 检验确定其显著性。

b. 预报模型构建方案。针对乌东德水库和瀑布沟水库进行独立建模,分别采用逐步线性回归与随机森林两种方法构建预报模型,建模时间为12个月,共构建了48个子模型。模型从每年年初起预报,利用上一年的关键气象因子预报该年的月平均径流,每个月的预见期不一致,预见期最长为1a。选取了数据系列的70%作为训练样本,即训练期为1959—1998年,验证期为1999—2014年。逐步回归计算由SPSS软件实现。随机森林核心内容基于Python语言机器学习算法库sklearn中的随机森林分类器(random forest classifier)实现^[22]。

c. 结果精度评定方案。中长期预报的预见期长,不确定因素多,缺乏确定的精度评定标准。本文参考前人在中长期预报中所采用的相关精度评定方法,选取了相关系数(r)、均方根误差(RMSE)、确定性系数(R^2)、平均相对误差(MRE)4个指标进行误差分析。其中, r 和 R^2 越大表明预测效果越佳,而RMSE值和MRE值越小表明预报效果越佳。RMSE值因水库集水面积不同而变化较大,仅作为参考;MRE无量纲,可消除水库集水面积的影响,准确反映不同流量量级的误差大小,进而较为直观地反映误差水平。

3 结果与分析

3.1 关键预报因子识别结果

乌东德水库与瀑布沟水库各月关键预报因子识别结果如表1所示,分析各分区的因子筛选结果可以发现,同一分区临近月份关键预报因子相似。例如乌东德水库1—3月,11月黑潮区海温指数与11月NINO A区海表温度距平指数都是关键预报因子,这符合气象因子对降雨径流影响的长时效性的规律,同时也反映了气象因子挑选结果的可靠性。此外,两个水库的同一月份的关键预报因子可能相似,例如乌东德与瀑布沟水库11月都含有5月北极涛动指数的关键预报因子,体现了大气环流对降雨径流的大尺度范围影响。

3.2 乌东德水库预报结果

表2为乌东德水库预报结果精度,可见,两种模型在训练期的训练效果良好,稳定性强,预报能力差别不大。在验证期时,两种模型的精度与模型稳定性有所下降,而随机森林的预测结果比逐步回归的精度高,但总体上来说两种模型的精度区别不大。计算两种模型训练期与验证期每个月的MRE值,结果见表3。由表3可见,两种预报模型在乌东德水库上游非汛期预测结果比汛期的结果更好,整体上误差随着径流量的增加而增加。训练期,随机森林预测结果的逐月平均相对误差更低一些;验证期,两种模型有高有低,难以具体比较。

图2为1959—2014年乌东德水库径流量逐月预测结果。在验证期2017年8月,逐步回归预测出的入库径流量与实测流量有大幅偏差,这也导致了8月逐步回归法预测结果的MRE值达到了34.59%。据分析,2017年8月,其关键预报因子印度副高强度指数陡增至1959—2014年内最高,其他时间该指标都比较低,所以导致了模型预测结果的巨大误差。随机森林这种情况下体现了优势,预测因子通过子学习器的分类后,按照不同数量级别对应不同的拟合流量值,形成的预测流量范围总是在历史流量的范围内,从而避免了某一预测因子值的突变带来的预报误差。

分析图2可知,两种模型在训练期时拟合实测还原流量效果非常优异,验证期时效果变差,而随机森林这一点特点更为明显,可见随机森林模型的过拟合现象较为严重。机器学习的过拟合主要原因是非线性模型对数据外延方面能力的不足,训练的数据一旦缺乏代表性,就会对预测精度造成较大影响。尽管在建模时已经多次调参,将模型调至最佳拟合效果,但由于水文数据时间序列较短,数据量不足,过拟合现象仍无法避免。

表1 关键预报因子识别结果

Table 1 Identification results of key forecasting factors

预报月份	乌东德水库关键预报因子	瀑布沟水库关键预报因子
1	11月黑潮区海温指数;11月NINO A区海表温度距平指数;8月亚洲区极涡面积指数;2月斯堪的纳维亚遥相关型指数;9月赤道中东太平洋200hPa纬向风指数	12月亚洲纬向环流指数;1月冷舌型 ENSO 指数;9月太平洋区极涡面积指数;9月南海副高脊线位置指数;4月NINO W区海表温度距平指数
2	11月黑潮区海温指数;11月NINO A区海表温度距平指数;2月斯堪的纳维亚遥相关型指数;5月北大西洋-欧洲环流型C型指数;10月热带印度洋海温偶极子指数	12月南海副高面积指数;12月太平洋区极涡面积指数;11月西太平洋暖池面积指数;3月西风漂流区海温指数;8月斯堪的纳维亚遥相关型指数
3	11月黑潮区海温指数;11月NINO A区海表温度距平指数;1月东亚槽强度指数;5月斯堪的纳维亚遥相关型指数;9月赤道中东太平洋200hPa纬向风指数	3月西太平洋副高脊线位置指数;4月亚洲纬向环流指数;11月亚洲区极涡强度指数;6月印度洋暖池面积指数;11月北极涛动指数
4	9月东亚槽位置指数;9月赤道中东太平洋200hPa纬向风指数;1月东亚槽强度指数;6月黑潮区海温指数	11月850hPa中太平洋信风指数;7月印度副高强度指数;4月北半球极涡中心强度指数;11月南海副高脊线位置指数;1月北极涛动指数
5	3月南海副高脊线位置指数;9月北半球极涡面积指数;5月南极涛动指数;9月赤道中东太平洋200hPa纬向风指数;8月亚洲区极涡面积指数	2月北半球极涡强度指数;8月印缅槽强度指数;1月NINO W区海表温度距平指数;11月北大西洋-欧洲环流型C型指数;12月黑潮区海温指数
6	2月北太平洋副高脊线位置指数;9月北半球极涡面积指数;7月北太平洋遥相关型指数;6月黑潮区海温指数;6月NINO A区海表温度距平指数	11月印缅槽强度指数;12月NINO C区海表温度距平指数;7月西太平洋副高脊线位置指数;4月赤道中东太平洋200hPa纬向风指数;5月850hPa中太平洋信风指数
7	4月欧亚经向环流指数;11月类 ENSO 指数;10月斯堪的纳维亚遥相关型指数;5月西太平洋暖池面积指数;7月北太平洋遥相关型指数	12月亲潮区海温指数;12月亚洲区极涡面积指数;12月850hPa中太平洋信风指数;8月NINO W区海表温度距平指数;12月南方涛动指数
8	5月东大西洋-西俄罗斯遥相关型指数;2月北半球极涡强度指数;2月北半球极涡中心强度指数;10月印度副高强度指数;11月热带印度洋海温偶极子指数	10月太平洋区极涡强度指数;10月东亚槽强度指数;4月斯堪的纳维亚遥相关型指数;3月西风漂流区海温指数;6月50hPa纬向风指数
9	5月北极涛动指数;2月北半球副高脊线位置指数;5月东大西洋-西俄罗斯遥相关型指数;4月太平洋区极涡面积指数;12月欧亚纬向环流指数	9月北半球副高脊线位置指数;12月太平洋区极涡面积指数;12月NINO 4区海表温度距平指数;7月青藏高原-1指数;10月亚洲纬向环流指数
10	2月北半球极涡中心强度指数;5月北极涛动指数;3月亚洲经向环流指数;6月西太平洋副高脊线位置指数;1月亲潮区海温指数	12月北太平洋副高脊线位置指数;2月北半球极涡中心经向位置指数;2月欧亚经向环流指数;12月NINO 3区海表温度距平指数;11月斯堪的纳维亚遥相关型指数
11	5月北极涛动指数;1月NINO A区海表温度距平指数;1月黑潮区海温指数;6月亚洲经向环流指数;2月斯堪的纳维亚遥相关型指数	5月北极涛动指数;5月亚洲区极涡强度指数;8月南海副高脊线位置指数;6月冷舌型 ENSO 指数;10月东亚槽位置指数
12	2月斯堪的纳维亚遥相关型指数;1月NINO A区海表温度距平指数;1月黑潮区海温指数;5月北极涛动指数;2月北半球副高脊线位置指数	10月北半球极涡中心强度指数;11月极地-欧亚遥相关型指数;2月850hPa西太平洋信风指数;10月北极涛动指数;10月冷舌型 ENSO 指数

表2 乌东德水库预报结果精度

Table 2 Accuracy of forecast results of Wudongde Reservoir

模型	训练期			验证期		
	r	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	R^2	r	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	R^2
逐步回归	0.95	967.02	0.91	0.87	1625.96	0.74
随机森林	0.95	968.24	0.91	0.87	1584.79	0.75

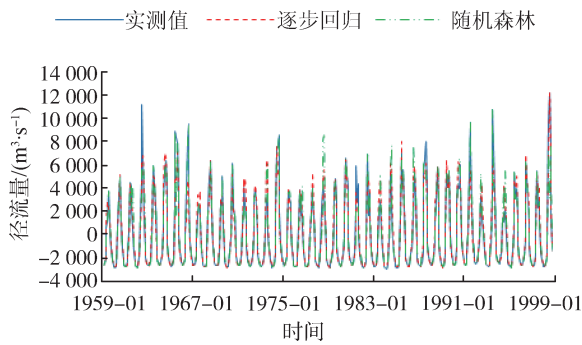
表3 乌东德水库预报结果逐月 MRE 值

Table 3 Monthly MRE values of forecast results of Wudongde Reservoir

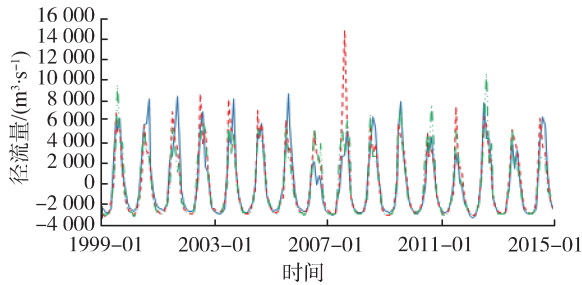
月份	MRE 值/%			
	训练期		验证期	
	逐步回归	随机森林	逐步回归	随机森林
1	7.03	1.40	17.85	16.65
2	5.99	2.87	18.16	16.87
3	4.95	2.70	15.87	16.50
4	7.26	3.76	12.67	13.77
5	12.12	4.04	21.07	14.00
6	14.86	3.97	19.37	29.28
7	14.20	5.08	29.66	21.91
8	20.59	12.62	34.59	27.76
9	16.30	7.25	22.85	28.57
10	16.40	13.52	14.46	26.19
11	9.15	5.79	14.85	16.68
12	9.05	5.33	9.96	14.44

3.3 瀑布沟水库预报结果

表4为瀑布沟水库预报结果精度,可见,在训练期,随机森林的预测效果比逐步回归更加优秀,3项精度评定指标都优于逐步回归的预测结果。尽管训练期有过拟合倾向,但在验证期,逐步回归的预测能力更加优秀,验证期结果与乌东德水库验证期相似。两种模型在验证期精度与模型稳定性有所下降,而随机森林的预测结果比逐步回归精度高,但总体上来说两种模型的精度差距不大。



(a) 训练期



(b) 验证期

图2 乌东德水库入库流量逐月预测结果

Fig.2 Monthly forecast results of Wudongde

Reservoir inflow

表4 瀑布沟水库预报结果精度

Table 4 Accuracy of forecast results of Pubugou Reservoir

模型	训练期			验证期		
	r	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	R^2	r	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	R^2
逐步回归	0.95	298.32	0.90	0.89	429.59	0.78
随机森林	0.99	122.04	0.98	0.89	420.87	0.79

计算两种模型训练期与验证期每个月的 MRE 值,如表 5 所示。由表 5 可见,逐步回归预报模型在瀑布沟水库上游非汛期时预测结果比汛期时的结果

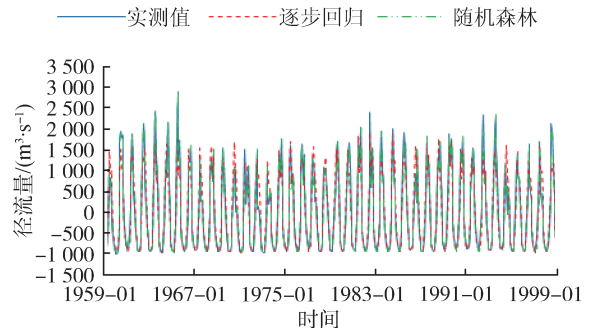
表5 瀑布沟水库预报结果逐月 MRE 值

Table 5 Monthly MRE values of forecast results of Pubugou Reservoir

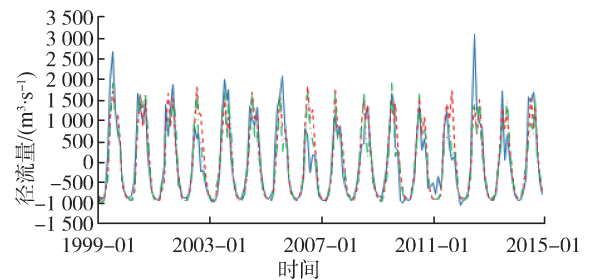
月份	MRE 值/%			
	训练期		验证期	
	逐步回归	随机森林	逐步回归	随机森林
1	5.59	2.20	14.30	14.29
2	4.16	3.06	7.86	9.11
3	7.44	2.14	13.98	14.69
4	15.58	3.50	14.55	24.39
5	15.44	4.74	17.54	17.27
6	13.68	2.66	16.57	19.43
7	15.86	3.02	26.40	27.07
8	17.42	0.00	28.03	27.40
9	20.87	0.77	34.68	20.62
10	14.47	5.33	18.62	20.28
11	9.31	5.31	27.52	29.84
12	7.86	4.57	13.87	15.03

更好,整体上误差随着径流量的增加而增加。而随机森林预报结果在训练期出现了 MRE 值减小的反规律状况,其在 6—9 月的平均相对误差降到了相当低的范围。据分析,出现这种状况的原因主要是在预测过程中随机森林模型的学习深度过高,导致了过拟合的问题。

图 3 为 1959—2014 年瀑布沟水库径流量逐月预测结果。可以看到,2012 年 7 月的月流量较大,达到了 $4\,415\text{ m}^3/\text{s}$,模型拟合效果较差。在该月的 20—23 日,受较强的冷空气和西南涡的共同影响,在岷沱江流域出现了 1 次强降雨过程,导致了该流量值是瀑布沟水库 1959—2014 年中最大的月入库流量。在瀑布沟水库 7 月挑选的因子为 12 月亲潮区海温指数、12 月亚洲区极涡面积指数、12 月 850 hPa 中太平洋信风指数、8 月 NINO W 区海表温度距平指数与 12 月南方涛动指数,该类因子在训练期的相关性较高,但 2012 年的 7 月流量偏高的主要影响因素为冷空气与西南低涡,所以预测结果精度较差。



(a) 训练期



(b) 验证期

图3 瀑布沟水库入库流量逐月预测结果

Fig.3 Monthly forecast results of Pubugou

Reservoir inflow

分析图 3 发现,在 2011 年 1 月、2 月、3 月出现了远远高于往年的流量,因为训练期未发生如此大的流量过程导致了模型拟合效果较差。同理,还有 2009 年的 11 月与 12 月的流量过程,异常低于以往的平均流量,结果拟合效果同样不佳。除了上述 2 段流量过程之外,整体上的流量拟合效果良好,两种模型均能准确训练出流量的变化过程,验证期相对误差在 40% 范围内,平均相对误差为 17%。

4 结 论

a. 针对长江流域上游的乌东德水库和瀑布沟水库的中长期径流预报,随机森林和逐步回归两种模型的模拟效果良好,稳定性强。随机森林的预测结果比逐步回归的精度高,但精度的差距不大。在验证期时,两种模型的精度与稳定性较训练期均有所下降,精度结果依然满足预报要求。预测误差随着预测流量的增加而增大,存在汛期误差大、非汛期误差小的现象。

b. 随机森林由于模型自身特点,不能够做出超越训练集数据范围的预测。在训练数据量充足的条件下,该特点可以避免某一预测因子值的异常变化带来的拟合误差,以逐步回归为代表的线性模型则相较不足。

c. 两种模型均存在过拟合现象,但随机森林更为明显。具体表现为:在训练期时拟合实测还原流量异常优异,到了验证期时变差。由于水文数据时间序列较短,数据量不足,导致了过拟合现象无法通过调参避免。这也体现了以随机森林为代表的非线性模型在数据外延方面能力的不足。

d. 相较于传统的数理统计模型,以随机森林为代表的机器学习模型预报精度更高,稳定性更强,在长江上游流域适用性良好,但模型本身存在着过拟合的问题,需要在接下来的研究中进一步探索。

参考文献:

[1] 刘勇,王银堂,陈元芳,等. 基于物理成因的中长期水文预报方法应用研究[M]. 南京:河海大学出版社,2017.

[2] 宋润虎. 多元线性回归与逐步回归在洪水预报中的应用:以南汀河中长期洪水预报为例[J]. 人民长江, 2012, 43 (增刊1): 122-125. (SONG Runhu. Application of multiple linear regression and stepwise regression in flood forecasting: taking the mid-long term flood forecasting of Nanting River as an example [J]. Yangtze River, 2012, 43 (Supl): 122-125. (in Chinese))

[3] 张幼成,李琼,徐汪洋,等. 基于导向标准断面法的受冲淤影响洪峰水位预报模型[J]. 水资源保护,2019,35 (3): 53-56. (ZHANG Youcheng, LI Qiong, XU Wangyang, et al. Forecasting model of flood peak water level affected by scour and silt based on guide standard section method [J]. Water Resources Protection, 2019, 35 (3): 53-56. (in Chinese))

[4] 葛朝霞,薛梅,宋颖玲. 多因子逐步回归周期分析在中长期水文预报中的应用[J]. 河海大学学报(自然科学版), 2009, 37 (3): 255-257. (GE Zhaoxia, XUE

Mei, SONG Yingling. Application of multi-factor stepwise regression period analysis in mid- and long-term hydrological forecasting [J]. Journal of Hohai University (Natural Sciences), 2009, 37 (3): 255-257. (in Chinese))

[5] 张轩,张行南,江唯佳,等. 秦淮河流域东山站水位预报研究[J]. 水资源保护, 2020, 36 (2): 41-46. (ZHANG Xuan, ZHANG Xingnan, JIANG Weijia, et al. Study on water level forecast of Dongshan Station in Qinhuai River Basin [J]. Water Resources Protection, 2020, 36 (2): 41-46. (in Chinese))

[6] 林剑艺,程春田. 支持向量机在中长期径流预报中的应用[J]. 水利学报, 2006, 37 (6): 681-686. (LIN Jianyi, CHENG Chuntian. Application of support vector machine in mid- and long-term runoff forecast [J]. Journal of Hydraulic Engineering, 2006, 37 (6): 681-686. (in Chinese))

[7] 许斌,杨凤根,酆于杰. 两类集成学习算法在中长期径流预报中的应用[J]. 水力发电, 2020, 46 (4): 21-24. (XU Bin, YANG Fenggen, LI Yujie. Application of two types of integrated learning algorithms in mid- and long-term runoff forecast [J]. Water Power, 2020, 46 (4): 21-24. (in Chinese))

[8] 赵鹏雁,张利平,王旭,等. 澜沧江流域中长期径流预报方法研究[J]. 武汉大学学报(工学版), 2018, 51 (7): 565-569. (ZHAO Pengyan, ZHANG Liping, WANG Xu, et al. Study on medium and long term runoff forecasting method of Lancang River Basin [J]. Engineering Journal of Wuhan University, 2018, 51 (7): 565-569. (in Chinese))

[9] 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 term and short term memory network in medium and long term runoff forecast [J]. Engineering Journal of Wuhan University, 2021, 54 (1): 21-27. (in Chinese))

[10] 谢帅,黄跃飞,李铁键,等. 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, et al. Mid-long term runoff prediction based on a lasso and SVR hybrid method [J]. Journal of Basic Science and Engineering, 2018, 26 (4): 709-722. (in Chinese))

[11] 顾巍巍,张卫国,张焱,等. 台风实时信息防汛服务云平台设计与应用[J]. 水资源保护, 2019, 35 (2): 85-89. (GU Weiwei, ZHANG Weiguo, ZHANG Yan, et al. Design and application of cloud platform for real time flood control information of typhoon [J]. Water Resources Protection, 2019, 35 (2): 85-89. (in Chinese))

(下转第 165 页)

- [70] 叶海, 李森, 薛峰, 等. 表面负荷对多介质土壤层系统处理污染河水的影响 [J]. 中国给水排水, 2012, 28(19): 74-77. (YE Hai, LI Sen, XUE Feng, et al. Effect of surface load on treatment of polluted river water by MSL system [J]. China Water & Wastewater, 2012, 28(19): 74-77. (in Chinese))
- [71] 陈佳利. 多级土壤渗滤系统处理污染河水的试验研究 [D]. 长沙: 湖南大学, 2012.
- [72] GUAN Y, ZHANG Y, ZHONG C N, et al. Effect of operating factors on the contaminants removal of a soil filter: multi-soil-layering system [J]. Environmental Earth Sciences, 2015, 74(3): 2679-2686.
- [73] PERALTA A L, MATTHEWS J W, FLANAGAN D N, et al. Environmental factors at dissimilar spatial scales influence plant and microbial communities in restored wetlands [J]. Wetlands, 2012, 32(6): 1125-1134.
- [74] 郭俊元, 周禹伶, 江世林, 等. 多级土壤渗漏系统处理农村生活污水 [J]. 中国环境科学, 2018, 38(9): 182-192. (GUO Junyuan, ZHOU Yuling, JIANG Shilin, et al. Multi soil layering filter and its performance in rural domestic sewage treatment [J]. China Environmental Science, 2018, 38(9): 182-192. (in Chinese))
- [75] 宋颖. 水产养殖污染源强及多介质土壤层技术废水处理效果与机理研究 [D]. 杭州: 浙江大学, 2016.
- [76] SBAHI S, OUAZZANI N, HEJJAJ A, et al. Neural network and cubist algorithms to predict fecal coliform content in treated wastewater by multi-soil-layering system for potential reuse [J]. Journal of Environmental Quality, 2021, 50(1): 144-157.
- [77] HO C C, WANG P H. Efficiency of a multi-soil-layering system on wastewater treatment using environment-friendly filter materials [J]. International Journal of Environmental Research and Public Health, 2015, 12(3): 3362-3380.
- (收稿日期:2021-04-22 编辑:王芳)

(上接第 136 页)

- [12] 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))
- [13] DIETTERICH T G. Ensemble learning [J]. The Handbook of Brain Theory and Neural networks, 2002, 2: 110-125.
- [14] BREIMAN L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123-140.
- [15] 赵钢铁, 杨大文, 蔡喜明, 等. 基于随机森林模型的长江上游枯水期径流预报研究 [J]. 水力发电学报, 2012, 31(3): 18-24. (ZHAO Tongtiegang, YANG Dawen, CAI Ximing, et al. Research on runoff forecast of the Upper Yangtze River in low water period based on random forest model [J]. Journal of Hydroelectric Power, 2012, 31(3): 18-24. (in Chinese))
- [16] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- [17] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [18] 龚振淞, 何敏. 长江流域夏季降水与全球海温关系的分析 [J]. 气象, 2006, 32(1): 56-61. (GONG Zhensong, HE Min. Relationship between Summer Rainfall in Changjiang River Valley and SSTA of various seasons [J]. Meteorological Monthly, 2006, 32(1): 56-61. (in Chinese))
- [19] 宋文玲, 杨义文. 长江三峡地区夏季旱涝特征及气候预测 [J]. 气象, 2003, 29(7): 13-18. (SONG Wenling, YANG Yiwu. Characteristics of drought/flooding in the Yangtze Gorges and climate prediction [J]. Meteorological Monthly, 2003, 29(7): 13-18. (in Chinese))
- [20] BOTHE Oliver, FRAEDRICH Klaus, ZHU Xiuhua, 等. 青藏高原夏季旱涝与大尺度环流的关系 [J]. 李耀辉, 翟颖佳, 译. 干旱气象, 2013, 31(4): 845-858. (BOTHE Oliver, FRAEDRICH Klaus, ZHU Xiuhua, et al. Relationship between summer drought and flood over the Tibetan Plateau and large scale circulation [J]. Journal of Arid Meteorology, 2013, 31(4): 845-858. (in Chinese))
- [21] 马振锋, 高文良. 青藏高原季风年际变化与长江上游气候变化的联系 [J]. 高原气象, 2003, 22(增刊1): 8-16. (MA Zhenfeng, GAO Wenliang. Relationship between interannual change over Qinghai-Xizang Plateau Monsoon and climate changing upper reach of Changjiang River [J]. Plateau Meteorology, 2003, 22(Sup1): 8-16. (in Chinese))
- [22] KOMER B, BERGSTRA J, ELIASMITH C. Hyperopt-sklearn: automatic hyperparameter configuration for scikit-learn [C]//ICML Workshop on AutoML. Austin T X: Citeseer, 2014: 50.
- (收稿日期:2021-01-19 编辑:王芳)