

基于机器学习方法的东江三角洲咸潮长期预测模型研究

谢雨航¹, 廖梓瑾¹, 王晨乃², 周喆¹, 王京晶¹

(1. 广东省水文局, 510150, 广州; 2. 广东省水文局广州水文分局, 510150, 广州)

摘要: 长期水文预测是水文学和水资源管理中的关键领域, 通过对未来一段时间内的水文变量进行预测, 对水资源规划、洪水预警、农业灌溉和城市供水等工作提供重要支撑。围绕咸潮影响下水资源配置中面临的关键问题和难点, 探讨了基于机器学习的咸潮预测模型在东江三角洲的应用, 重点研究了特征因子的选取和模型在不同条件下对年际尺度咸潮活动的中长期预测。主要通过建立咸潮影响等级评价体系, 筛选出关键特征因子, 并采用随机森林、梯度提升树和集成模型等学习技术进行咸潮预测。研究表明, 非汛期降雨量、龙滩水库汛末蓄水量、三水站分流比以及新丰江水库汛末蓄水量是影响咸潮活动的关键因素。基于这些关键特征因子构建的机器学习集成模型提高了咸潮预测的准确性和稳定性。此外, 结合实际设定多种场景开展咸潮活动的预测工作, 可为东江三角洲地区的水资源管理以及咸潮预警提供科学支撑。

关键词: 咸潮评价体系; 机器学习; 特征因子; 咸潮预测模型; 东江三角洲

Research on a long-term salinity intrusion forecasting model for the Dongjiang River Delta based on machine learning approaches//XIE Yuhang, LIAO Zijin, WANG Chennai, ZHOU Zhe, WANG Jingjing

Abstract: Long-term hydrological forecasting is a key area in hydrology and water resources management. It provides essential support for water resources planning, flood warning, agricultural irrigation, and urban water supply by predicting hydrological variables over extended time periods. Focusing on the key challenges in water resources allocation under salinity intrusion, this study investigates the application of machine learning-based salinity intrusion forecasting models in the Dongjiang River Delta. Emphasis is placed on the selection of characteristic factors and the medium- to long-term prediction of salinity intrusion on an interannual scale under varying conditions. A salinity intrusion impact evaluation system is developed to identify critical characteristic factors. Random forest, gradient boosting tree, and ensemble models are employed for forecasting. The results indicate that non-flood season rainfall, end-of-flood season storage in Longtan Reservoir, diversion ratio at Sanshui Station, and end-of-flood season storage in Xinfengjiang Reservoir are the key factors influencing salinity intrusion. The ensemble model constructed with these critical factors enhances the accuracy and stability of salinity intrusion forecasting. Additionally, predictions conducted under multiple practical scenarios provide scientific support for water resources management and salinity intrusion warning in the Dongjiang River Delta.

Keywords: salinity intrusion evaluation system; machine learning; characteristic factors; salinity intrusion forecasting model; Dongjiang River Delta

中图分类号: TV1+P731 文献标识码: B 文章编号: 1000-1123(2025)14-0056-10

DOI: 10.3969/j.issn.1000-1123.2025.14.010

收稿日期: 2025-07-07 修回日期: 2025-07-24

作者简介: 谢雨航, 助理工程师, 主要从事水文与水资源方向的研究。

通信作者: 廖梓瑾, 高级工程师, 主要从事水文与水资源方向的研究。E-mail: 13610294792@163.com

基金项目: 广东省水利科技创新项目“北江流域-城市复合洪涝过程模拟与风险评估”(2025-23)。

长期水文预测是水文学和水资源管理中的关键领域,通过对未来一段时间内的水文变量进行预测,对水资源规划、洪水预警、农业灌溉和城市供水等工作提供重要支撑。在咸潮预测领域,以往的研究多集中在月或日尺度的模拟,而对年际尺度的模拟研究相对较少。年际尺度的咸潮模拟能够反映咸潮在不同年份之间的变化规律和趋势,揭示气候变化、人类活动等因素对咸潮的长期影响,为应对气候变化和合理开发利用水资源提供科学依据。此外,年际尺度的模拟还可以帮助预测未来咸潮可能对沿海地区供水系统造成的威胁,从而提前制定应对措施,保障供水安全。水库调度作为调控河流径流的关键人为因素,直接影响河口咸潮入侵的强度和范围。年际尺度模拟能够揭示水库运行的长期累积效应与气候变化的协同影响,而短期尺度(日、月)难以捕捉此类复杂相互作用。

近年来,机器学习和人工智能技术在水文预测中的应用日益广泛,包括神经网络、随机森林(RF)、支持向量机等,这些技术能够处理较为复杂的非线性关系,长短期记忆网络(LSTM)等循环神经网络在序列预测方面展现出显著的优势,而集成学习技术则通过融合多个模型的优势,进一步提升了预测的准确性。机器学习技术的发展为提高长期水文预测的精确度和可靠性提供了新的可能。在咸潮模拟与预测研究领域,机器学习技术的应用日益广泛。陈记臣等构建了集成Ada Boost算法的BP神经网络模型,对广东省珠海市平岗泵站取水口的日均氯浓度进行了模拟计算。刘德地等通过结合偏最小二乘回归与支持向量机算法,开发了咸潮预测模型(PLS-SVM),并应用于模拟和预测珠海平岗站月尺度的咸潮超标时间。袁从贵等基于偏互信息和固定尺度最小二乘支持向量机,建立了小时单变量大样本咸潮预测模型,并在东江流域进行了预测研究。苏程佳等利用随机森林算法建立了咸潮的日预测模型。机器学习方法近年来在水文预测的生产实践中得到了广泛且有效的应用。

本研究以东江南支流的泗盛围站为研究对象,运用机器学习方法构建咸潮预测模型,并引入西北江来水因子,探究东江咸潮活动与西北江来水之间可能存在的潜在联系。本研究重点是构建机器学习咸潮预测模型,探索特征因子选取的难点问题,针对珠江三角洲尤其是东江三角洲开展年际尺度的咸潮预警预报关键技术研究;目的在于实现对咸潮上溯严重程度的科学评估,有效筛选适宜的特征因子,以及对咸潮现象进行中长期模拟与预测,为跨年度水资源管理提供科学工具。

一、咸潮评价体系建立

1. 研究区域概况

东江三角洲位于广东省中部、东江下游,与西、北江三角洲共同构成珠江三角洲,年平均气温约 20°C ,年平均降雨量为 1947 mm ,雨量集中于4—9月,暴雨较多,东江三角洲网河区北面以东江北干流为界,东南到南支流,西面到狮子洋,总面积为 319.5 km^2 。其中河涌水面积 58 km^2 ,河网密度达 18.2% 。东江三角洲上游控制断面博罗站多年平均流量为 $738\text{ m}^3/\text{s}$,下游泗盛围站潮动力较珠江河口其他口门而言相对较强,多年平均涨落潮差为 1.57 m 。东江三角洲地势宽广平坦,平均坡降仅为 0.011% ,整体高程多在 $0\sim 3\text{ m}$,受降雨径流特征以及地形影响,东江三角洲极易受洪、涝、潮威胁,枯季咸潮可上溯至石龙以上。东江三角洲内包含广州、东莞、深圳等经济核心城市,经济繁荣且人口、产业密集,对于咸潮入侵等灾害敏感性较高,实施压咸调度措施的需求尤为迫切。

2020年至2022年年初,东江流域经历了60年一遇的严重旱情和近10年来最严重的咸潮灾害。此次咸潮灾害表明东江三角洲的咸潮活动可能受到东江和西江、北江来水的共同作用,泗盛围站为东江西南支流的入海口站,站点受潮汐影响显著,且有较长序列的咸潮监测整编资料,可作为年际尺度上预判咸潮活动严重程度的指标站点。在西江、东江干流水库建成后,上游径流受水库调节影响显著,对珠江三角洲枯季来水影响较大,在进行年际尺度上的咸潮预测时,除研究天然降雨、径流外,也需考虑汛末水库蓄水量。龙滩水库位于西江干流,控制集雨面积 $98\ 500\text{ km}^2$,总库容 188.09 亿 m^3 ,兴利库容 111.5 亿 m^3 ,是西江流域水资源调度的关键性工程。新丰江水库位于东江支流新丰江上,控制集雨面积 5734 km^2 ,总库容 138.96 亿 m^3 ,兴利库容 64.91 亿 m^3 ,是东江水资源调度的重点工程,本文以龙滩水库和新丰江水库作为指标水库。

2. 基础数据

研究所采用的资料主要来源于广东省水文部门的监测资料,包括站点常规监测资料及专项测验监测资料,站点常规监测资料时间序列超过30年,测验规范、资料经过合理性分析及整编,可对咸潮上溯的严重程度进行评价以及对咸潮上溯进行年际尺度上的预测建模。

3. 咸潮影响等级评价体系

我国现行标准《地表水环境质量标准》(GB 3838—

2002)和《生活饮用水水源水质标准》(CJ 302—1993)均把氯化物浓度是否大于250 mg/L作为水质是否超标的判断标准。按照《水资源调查评价技术细则》,可根据年降水量分为丰水年(降水频率 $P < 12.5\%$)、偏丰年($12.5\% \leq P < 37.5\%$)、平水年($37.5\% \leq P < 62.5\%$)、偏枯年($62.5\% \leq P < 87.5\%$)、枯水年($87.5\% \leq P < 95\%$)、特枯年($P \geq 95.0\%$)6个等级。本文参考以上水文年评价标准,设定咸潮活动影响为不严重、偏不严重、一般、偏严重、严重、特严重6个等级(见表1)。

丰枯水年的评价以年降水量作为评价指标,但水体中咸潮是动态变化的过程,难以用单一指标衡量其影响程度,咸潮影响评价需同时考虑量级和时间两个因素。含氯度涨憩日最大值是纳入整编的数据,在一个非汛期求平均值,得到非汛期含氯度涨憩日最大值的平均值,同时考虑了影响时间的评价指标。

泗盛围站历史序列超过30年,计算每年10月至次年3月泗盛围站含氯度涨憩日最大值的算术平均值,其中1986年、2005年及2021年数据由于监测原因仅观测当年数据,取当年10月至12月含氯度涨憩日最大值的平均值。用适线法进行频率分析,完成频率分析后可对年际尺度上的咸潮影响程度进行等级评价。本文使用武汉大学水资源与水电工程科学国家重点实验室的水文频率分布曲线适线软件进行频率适线分析,对站点历年咸潮影响程度进行等级评价,各级频率特征值及历年评价结果见表2和图1。

二、模型特征因子确定

选择合适的特征影响因子对于机器学习模型的精度和泛化能力至关重要,合适的特征因子应当满足3个条件:相互独立、对模型预测结果有显著影响、易于获取和计算。

东江三角洲的咸潮活动从物理角度来看,受到东江来水和西

北江来水共同影响,对咸潮活动影响较显著的因子为降雨、径流、水库汛末蓄水量、潮汐和三角洲分流比。西北江在三水站以下分流,经过珠江前后航道、沙湾水道流入狮子洋,与其中的海水充分混合,这一过程决定了泗盛围站的含氯度,进而影响东江咸潮的上溯,基于此,引入西江来水因子、泗盛围站逐时含氯度因子作为控制条件构建模型,在考虑特征物理意义的基础上,采用过滤法进行特征因子选择,过滤法通过计算特征的差异性、相关性等指标来评估选择特征因子。由于非汛期径流量数据难以准确获取,可通过易直接获取的水库汛末蓄水量数据,叠加枯季天气的中长期预报,构成年际尺度的预报依据。

由于西江上游龙滩水库的数据序列较短,仅有14个

表1 咸潮影响等级评价标准

影响等级	频率范围
不严重	$P \geq 87.5\%$
偏不严重	$62.5\% \leq P < 87.5\%$
一般	$37.5\% \leq P < 62.5\%$
偏严重	$12.5\% \leq P < 37.5\%$
严重	$5\% \leq P < 12.5\%$
特严重	$P < 5\%$

表2 泗盛围站含氯度频率分析结果

序号	1	2	3	4	5
频率	87.5%	62.5%	37.5%	12.5%	5%
特征值	1.98	2.75	3.47	4.60	5.37

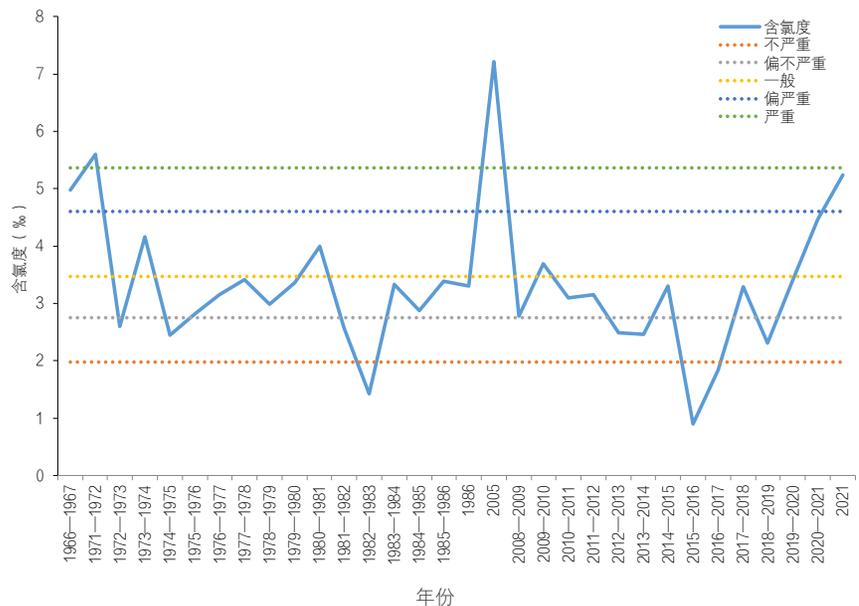


图1 泗盛围站咸潮等级评价结果

数据序列,而年际尺度上其他因子的数据序列有57个,需对龙滩水库数据序列进行预处理。由14个样本系列建立龙滩水库汛末蓄水量与梧州站非汛期流量的线性关系式,按照水量尺度 $W_0=205.3 \times 10^8 \text{ m}^3$ (龙滩水库有效蓄水量),流量尺度 $Q_0=2350 \text{ m}^3/\text{s}$ (梧州水文站非汛期多年平均流量)对序列进行无量纲化处理,得到无量纲线性关系:

$$W_{\text{龙}}^* = 1.51 \times 10^{-3} Q_{\text{梧}}^* + 4.15 \times 10^{-3}$$

式中, $W_{\text{龙}}^*$ 为无量纲化后龙滩水库汛末蓄水量, $Q_{\text{梧}}^*$ 为无量纲化后梧州站非汛期流量。

根据上式得到的14个样本系列与真实值的平均相对误差为17.91%,在20%的许可误差范围之内。基于梧州站非汛期的流量序列估算出龙滩水库汛末的蓄水量,将数据序列扩展至57个并进行方差分析,方差计算结果见表3。

序列长度 n 为14时,除三水站、马口站分流比的

方差很小,其他特征因子方差都远大于置信度95%的F检验值。 $n=57$ 时,方差增大,但仍未通过F检验,从物理意义上来看分流比对咸潮是有一定影响的。

三水站、马口站分流比互为完全负相关,故选择其中一项参与Pearson相关分析,相关因子结果如表4所示,结果显示龙滩水库汛末蓄水量与梧州站非汛期流量显著相关;三水站分流比与石角站非汛期流量、非汛期降雨量均极显著相关。由于数据序列短,对除了龙滩水库的其他6个因子进行相关分析。

对57个数据序列的Pearson相关分析结果,结果显示,新丰江水库汛末蓄水量与博罗站非汛期流量极显著相关;非汛期降雨量与梧州、石角、博罗3站的非汛期流量均极显著相关;三水站分流比与梧州站、石角站非汛期流量显著相关(见表5)。

综上所述,在构建泗盛围站模型时,筛选出了龙滩

表3 方差计算结果

特征因子	龙滩水库	新丰江水库	梧州站非汛期流量	石角站非汛期流量	博罗站非汛期流量	非汛期降雨	三水站分流比	马口站分流比
序列长度 $n=14$, 方差 $f=4.747$	8.86×10^2	2.27×10^2	1.54×10^6	1.05×10^5	2.34×10^4	3.75×10^4	2.22×10^{-4}	2.22×10^{-4}
序列长度 $n=57$, 方差 $f=4.016$	/	3.90×10^2	9.06×10^5	9.68×10^4	2.09×10^4	2.69×10^4	2.00×10^{-3}	2.00×10^{-3}

表4 年际尺度预测各影响因子相关性检验

$n=14$	龙滩水库	新丰江水库	梧州站非汛期流量	石角站非汛期流量	博罗站非汛期流量	非汛期降雨量	三水站分流比
龙滩水库	1.000						
新丰江水库	0.227	1.000					
梧州站非汛期流量	0.548*	0.043	1.000				
石角站非汛期流量	-0.049	0.159	0.602*	1.000			
博罗站非汛期流量	0.156	0.517	0.536*	0.756**	1.000		
非汛期降雨量	-0.158	-0.055	0.576*	0.847**	0.610*	1.000	
三水站分流比	-0.135	0.256	0.337	0.713**	0.497	0.730**	1.000

注: *为显著相关($a=0.05$, $|r_a|=0.532$); **为极显著相关($a=0.01$, $|r_a|=0.661$)。

表5 不考虑龙滩水库的影响因子相关性检验

$n=57$	新丰江水库	梧州站非汛期流量	石角站非汛期流量	博罗站非汛期流量	非汛期降雨量	三水站分流比
新丰江水库	1.000					
梧州站非汛期流量	0.101	1.000				
石角站非汛期流量	0.206	0.643**	1.000			
博罗站非汛期流量	0.499**	0.468**	0.741**	1.000		
非汛期降雨量	0.008	0.630**	0.834**	0.629**	1.000	
三水站分流比	0.405**	0.276*	0.262*	0.102	0.197	1.000

注: *为显著相关($a=0.05$, $|r_a|=0.262$); **为极显著相关($a=0.01$, $|r_a|=0.340$)。

水库汛末蓄水量、新丰江水库汛末蓄水量、非汛期降雨量以及三水站分流比这4个关键特征因子,边界概化见图2。

三、基于机器学习的咸潮预报集成模型

为了筛选出合适算法,首先用多种分类方法进行试算,本文比较了K最近邻回归、支持向量机、随机森林、梯度提升树(GBT)等多种分类算法,给定随机种子30,使用默认参数的计算结果见表6,结果显示当咸潮评价分类为6个等级时,算法的预测精度较低,最高仅为0.71,当目标变量精简为3个等级后,几乎所有算法的精度都提高了。为了提升模型的预测精度,并考虑到在实际应用中,年际尺度咸潮强度预测更关注于咸潮较为严重的情况,同时避免因等级划分导致的样本数量分布不均,不利于模型训练等问题,将年际尺度

上的咸潮系列简化为3个等级(1级、2级、3级),作为模型的目标变量。根据表7的计算结果可知,随机森林及梯度提升树两类算法在训练集和测试集表现均较优,具有不错的稳定性,因此本文考虑采用随机森林、梯度提升树构建集成模型。

如果汇总一组的预测模型的预测结果,理论上可以获得到比最优的单个预测器更好的预测结果,这一组预测器称为一个集成,这种技术称为集成学习。集成模型拓宽了假设空间,在函数表示上比任一原始模型更加灵活,因此一般可以取得更好的预测性能。采用软投票的方式将随机森林、梯度提升树组合成集成模型,集成模型的建模框架见图3。

四、咸潮模型特征因子重要性分析

本研究构建咸潮预测模型,旨在为河口科学调度、降低咸潮灾害风险提供技术支撑,故模型需要能够反映关键水文过程及物理规律,脱离物理基础的数学模型可能影响水文预测价值,且难以指导实际工作。如果一个机器学习模型的精度很高,但是其模型的构建基础与其要预测的问题毫无关系或关联甚小,即使得到了良好的模拟效果也是不可信的。

本文采用基尼指数来评价特征因子重要性,绘制了5种方案(5种不同的随机种子)下不同模型的4种特征因子的重要性排序及数值大致分布范围块图,各项特征因子的和为1。同时,对5种方案下各特征因子的重要性进行定量计算,以箱线图来表征(见图4)。

由特征因子重要性分析结果可以得出以下结论:

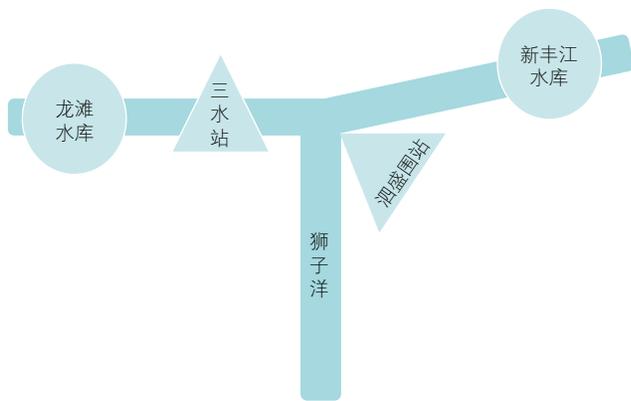


图2 泗盛围站年际尺度模型边界概化

表6 咸潮影响评价的数字处理

咸潮影响程度	不严重	偏不严重	一般	偏严重	严重	特严重
数字表达(简化前)	1	2	3	4	5	6
数字表达(简化后)	1		2	3		

表7 分类算法优选表

目标变量表达	算法	K最近邻回归	支持向量机	LinearSVC	Bagging_tree	随机森林	AdaBoost	梯度提升树	多层感知器	Logistic回归
6级	RMSE	1.07	1.25	1.0	0.53	0.53	0.65	1.0	0.76	1.25
	训练集R2	0.48	0.56	0.6	1.0	1.0	0.72	1.0	0.72	0.60
	测试集R2	0.29	0.29	0.43	0.71	0.71	0.57	0.43	0.43	0.29
精简为3级	RMSE	0.53	0.76	0.53	0.53	0.38	0.93	0.53	0.53	0.53
	训练集R2	0.60	0.60	0.60	0.96	1.0	0.68	1.0	0.80	0.60
	测试集R2	0.71	0.43	0.71	0.71	0.86	0.57	0.71	0.71	0.71

①图4(a)(b)(c)显示,无论怎样划分测试集和训练集,对年际尺度上咸潮活动预测起关键作用的因子排序基本一致,这说明本文所建模型的水文意义和物理意义是可信的,模型关键结构没有随着不同训练

集的划分使得各特征因子的增益发生剧烈变化,样本的水文特性被模型所学习,且具有良好的稳定性。

②图4(d)(e)(f)表明,5种方案下特性因子重要性从大到小排序依次为:非汛期降雨量>龙滩水库蓄水量>三水站分流比>新丰江水库蓄水量。这一结论与之前对特征重要性的定性分析结果吻合,证明了模型坚实的物理基础。

③非汛期降雨量的影响明显高于其他因子,非汛期降雨量越大,产生的河道径流量越大,越能抑制咸潮上溯运动。

④龙滩水库的重要性排在新丰江水库之前,这一排序凸显了西江来水对东江区域咸潮活动的实质性影响,其影响程度甚至超过了新丰江水库对东江下游咸潮活动的直接影响。

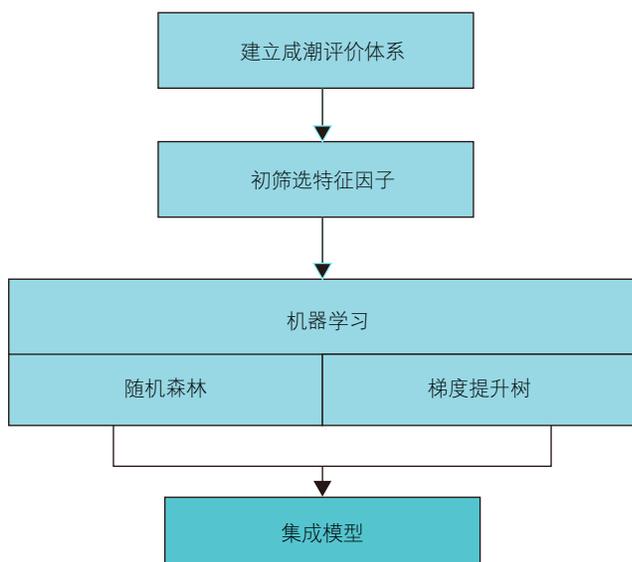


图3 集成模型建模框架示意图

五、泗盛围站年际尺度模型评价

1. 泗盛围站年际尺度模型稳定性评估

采用交叉验证 (cross-validation, CV) 的方法评估不同模型稳定性。在数据处理与模型训练过程中,数据的交叉使用能够增强模型捕捉训练集特征的能力,进而

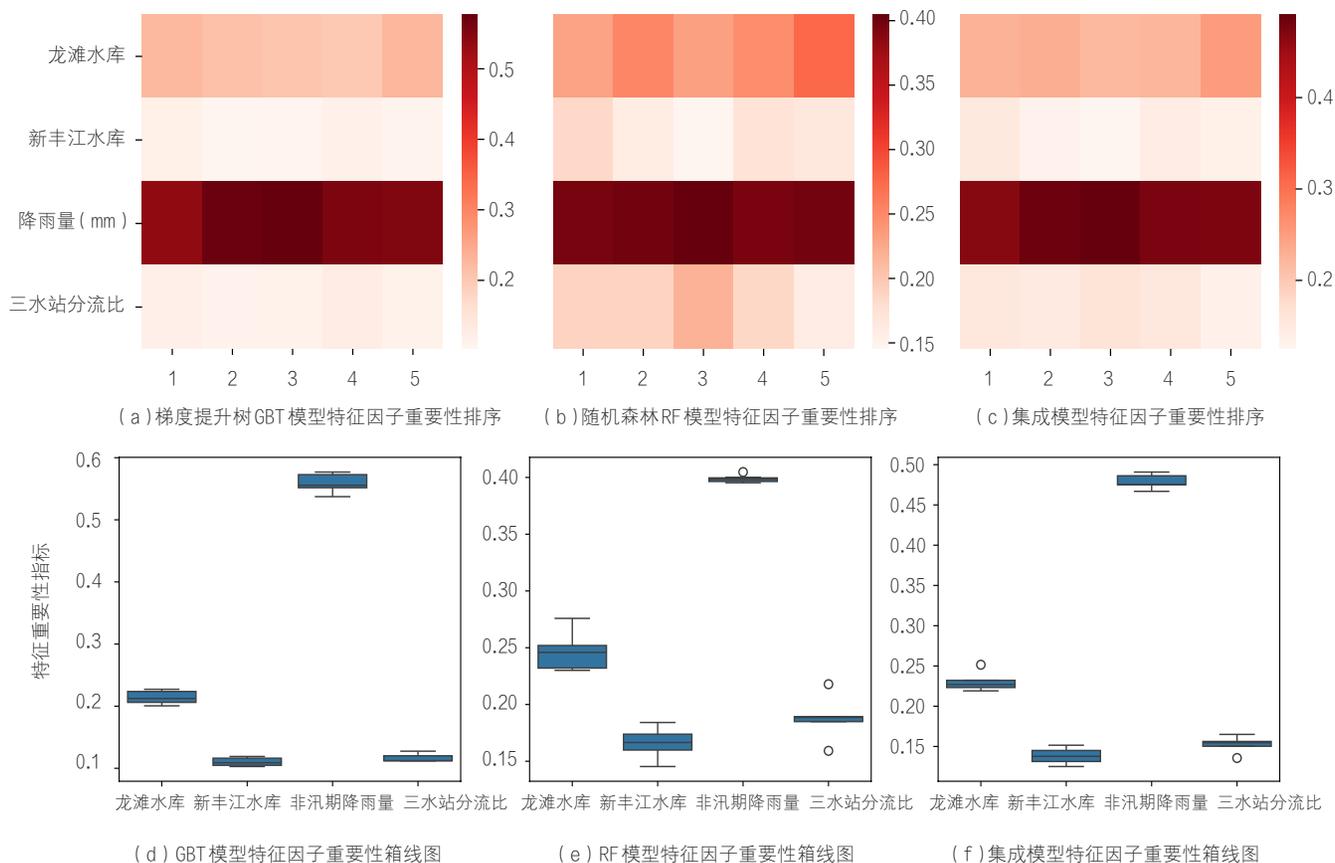


图4 不同方案下模型特征因子重要性分析

有效缓解选择偏差问题,同时降低过拟合风险。交叉验证方法所得的平均结果能够较为客观、准确地映射模型在实际应用中的表现,进而提升模型的泛化能力。

为考察模型拟合性能随数据集规模的变化,将训练集按15%、30%、50%、70%、90%的比例抽样,进行5折交叉验证,得到图5中3种模型的学习曲线。图5中,红色圆点表示训练集的平均分类准确率,绿色圆点表示验证集的平均分类准确率,线两侧区域显示方差大小。

根据图5可看出:①随机森林模型在数据集过小时存在明显的过拟合,泛化性能较差,且在验证集上表现稳定性不足,表明模型对数据集选取较为敏感。②梯度提升树模型在小数据量下的预测效果比随机森林模型稳定,随着训练数据量的增加,模型精度也逐渐提升。③集成模型的验证集表现总体优于子模型。这表明集成模型吸取了各个子模型的优势,即增强了小数据量下的稳定性。④总体来看,利用集成模型进行泗

盛围站年际尺度咸潮预报方案是稳定可靠的。

2. 泗盛围站年际尺度咸潮模拟精度评定

对于分类算法,常采用混淆矩阵进行结果评价。混淆矩阵可以很方便地评价分类问题的预测结果。混淆矩阵显示了在进行预测时,二分模型的哪些部分会被混淆,不仅能够揭示分类模型所犯的错误,还能对错误类型进行分类。

混淆矩阵利用准确率、精确率和召回率来描述模型的性能。根据率定好的泗盛围站年际尺度随机分类模型,对该站2008—2021年非汛期的咸潮活动严重程度进行模型性能评估,将数据按照80%训练集和20%测试集进行划分。本模型有3个目标变量,因此本混淆矩阵是一个三元矩阵(图6)。

由图6可知,三种模型的预测准确率均较高,其中随机森林模型表现较差,GBT模型与集成模型大多数预测值与真值一致,这在一定程度上反映了模型在处理数据时的稳定性。由表8可知,集成模型在训练集

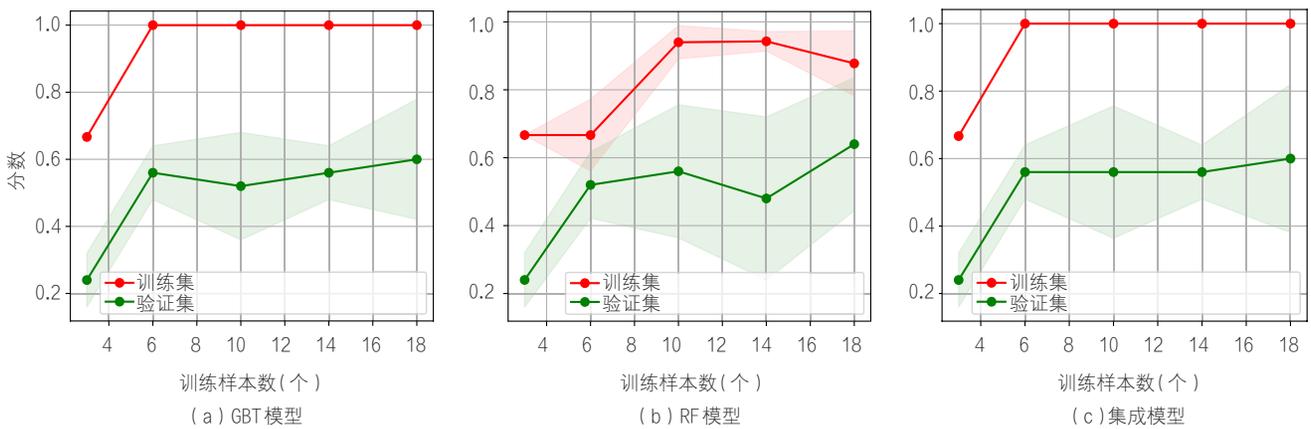


图5 不同方案下模型学习曲线

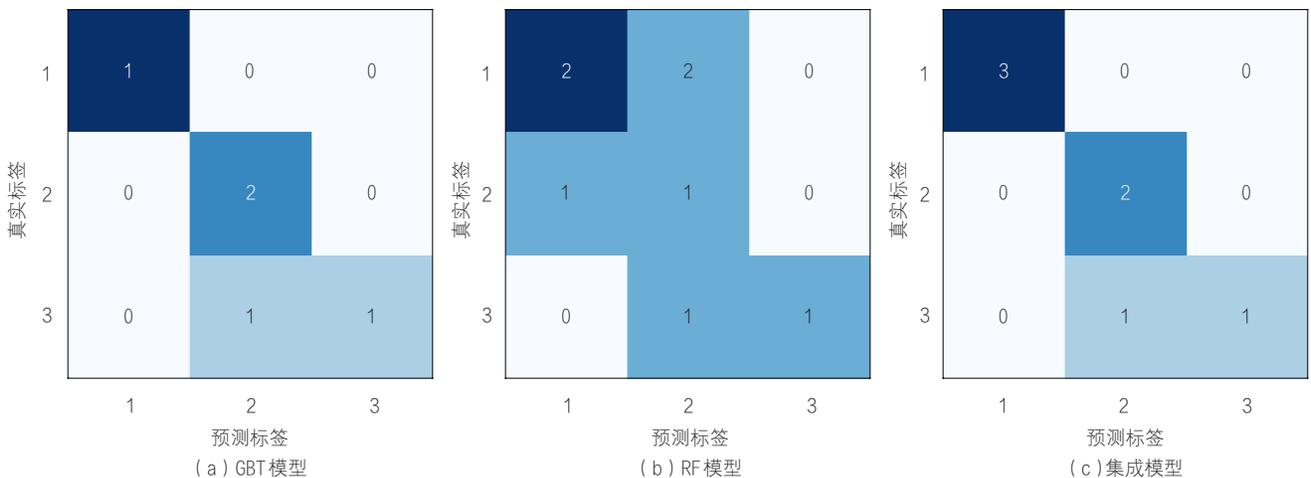


图6 泗盛围站年际尺度模拟混淆矩阵结果

和测试集上的表现均优于随机森林模型和梯度提升树模型,随机森林模型在训练集和测试集上的性能指标较为接近,表明其泛化能力较好,没有明显过拟合或欠拟合现象。梯度提升树模型在训练集上表现较好,但在测试集上的性能有所下降,表明该模型在面对新数据时适应性稍弱。集成模型则在训练集和测试集上均表现出色,表明其不仅在训练数据上具有强大的学习能力,而且能够将这种性能稳定地迁移到测试数据上,具有良好的平衡性和泛化能力。

六、模拟结果与讨论

基于珠江流域(广东省)非汛期的丰枯评价结果,选取临界雨量值作为非汛期降雨量的输入参数,并结合实际设定多种水库汛末蓄水量的场景进行分析,以期获得在不同咸潮影响程度下,上游水库汛末蓄水量的最优阈值组合。

1. 水库蓄水量组合设置

基于龙滩水库和新丰江水库的多年汛末蓄水量均值(分别为130.5亿 m^3 和76.3亿 m^3)以及最小蓄水量(分别为80.5亿 m^3 和46.3亿 m^3),设计了以下4种场景:①以多年蓄水量均值作为水库输入值;②龙滩水库水量逐次递减10亿 m^3 ,直到龙滩水库达到历年最小汛末蓄水量为止,新丰江水库保持多年均值,作为水库输入值;③龙滩水库保持多年蓄水量均值,新丰江水库逐次递减10亿 m^3 ,直到新丰江水库达到历年最小汛末蓄水量为止,作为水库输入值;④龙滩水库、新丰江水库均逐次递减10亿 m^3 ,先达到历年最小汛末蓄水量的水库维持最小蓄水量,另一水库仍逐次递减。

2. 临界雨量值

将非汛期划分为:非汛期丰水年($P < 12.5\%$)、非汛期偏丰年($12.5\% \leq P < 37.5\%$)、非汛期平水年($37.5\% \leq P < 62.5\%$)、非汛期偏枯年($62.5\% \leq P < 87.5\%$)、非汛期枯水年($87.5\% \leq P$

$< 95\%$)及非汛期特枯年($P \geq 95\%$)6类非汛期水文年。对珠江流域(广东省)1962—2020年的非汛期面雨量值序列进行频率分析计算,根据划分标准得到相应的临界雨量特征值(见表9)。

3. 模拟结果分析

分别用以上4种场景,以5个临界雨量值作为输入条件,对年际尺度咸潮活动严重程度进行预测(见图7)。

表8 不同模型泗盛围站咸潮年尺度模型性能评估

方案	训练集			测试集		
	F1分数	准确率	精确率	F1分数	准确率	精确率
随机森林	0.88	0.88	0.896	0.86	0.86	0.83
梯度提升树	1.0	1.0	1.0	0.82	0.86	0.89
集成模型	1.0	1.0	1.0	0.86	0.89	0.89

表9 珠江流域(广东省)临界面雨量特征值

非汛期丰枯水年	频率范围(%)	特征雨量值(mm)	非汛期丰枯水年评价(个)
丰水年	$P < 12.5$	$R \geq 540.03$	4
偏丰年	$12.5 \leq P < 37.5$	$378.71 \leq R < 540.03$	20
平水年	$37.5 \leq P < 62.5$	$297.02 \leq R < 378.71$	14
偏枯年	$62.5 \leq P < 87.5$	$234.91 \leq R < 297.02$	13
枯水年	$87.5 \leq P < 95$	$215.84 \leq R < 234.91$	3
特枯年	$P \geq 95$	$R < 215.84$	4
总计	/	/	58

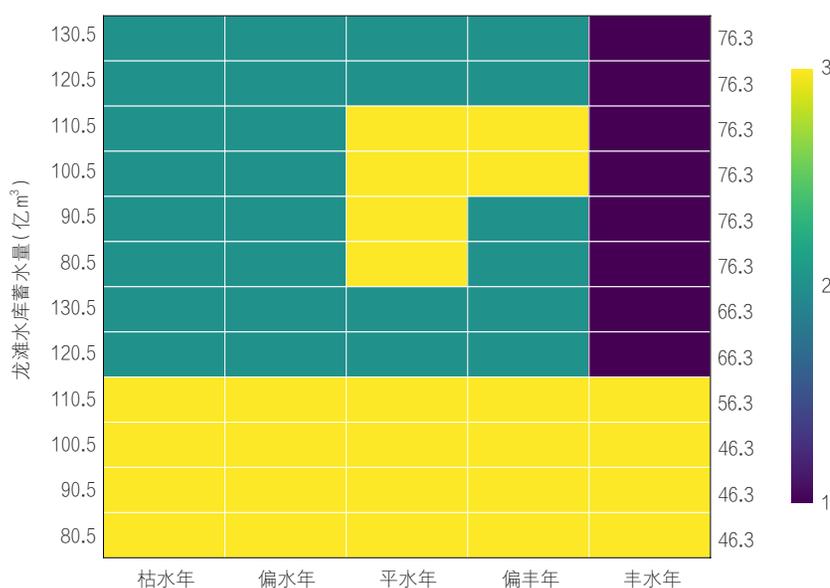


图7 不同设计场景下泗盛围站咸潮年际尺度预测结果

图7为不同设计场景下泗盛围站咸潮年际尺度预测结果(数字1、2、3依次代表年际尺度上泗盛围站的咸潮活动严重程度为不严重到偏不严重、一般严重、偏严重到特严重)。

如图7所示,当龙滩、新丰江水库为汛末平均蓄水量时,只要非汛期雨量大于特枯雨量临界值,东江南支流的咸潮活动就能保持一般到不严重;当新丰江水库汛末蓄水量达到历年平均值76.3亿 m^3 、龙滩水库汛末蓄水量小于等于110.5亿 m^3 时,东江南支流的咸潮活动可能达到偏严重程度;若龙滩水库汛末蓄水量达到历年平均值130.5亿 m^3 ,东江南支流的咸潮活动可能保持一般到不严重;当龙滩水库汛末蓄水量小于等于110.5亿 m^3 、新丰江水库汛末蓄水量小于等于56.3亿 m^3 时,东江南支流的咸潮活动就可能达到偏严重到特严重程度。

当龙滩水库和新丰江水库汛末蓄水量同时变化时,咸潮活动变化最为敏感,为了得到更精确的水库蓄水量组合,将龙滩水库(120.5,66.3)亿 m^3 到(110.5,56.3)亿 m^3 之间各以1为递减变化量进行试算,得到预测结果见图8。

从图8可以看出,当水库蓄水量组合小于(110.5,56.3)亿 m^3 时,东江咸潮活动可能达到偏严重以上。虽然在(117.5,63.3)亿 m^3 的水库组合得到了3(严重)的预测结果,但当蓄水量更小时,预测结果跳动至2(一般),故可以认为3(严重)的结果是计算误差,仍以稳定出现3(严重)结果的(110.5,56.3)亿 m^3 为阈

值组合。

由图中结果可知,当龙滩水库汛末蓄水量小于等于110.5亿 m^3 、新丰江水库汛末蓄水量小于等于56.3亿 m^3 时,东江南支流的咸潮活动就可能达到偏严重到特严重程度。2021年新丰江水库汛末蓄水量仅50.56亿 m^3 ,龙滩水库为110.03亿 m^3 ,2021—2022年东江发生了严重的咸潮上溯,可见本文得到的结论是合理的。该结论对根据水库蓄水量及中长期天气预测来预判当年的咸潮严重程度具有重要意义。

七、总结与展望

本文通过建立咸潮影响等级评价体系、筛选关键特征因子和构建集成模型,有效提升了咸潮预测模型的精度和稳定性,为东江乃至珠江三角洲的水资源管理和咸潮预警提供了科学依据。未来可进一步通过遥感技术获取更多高频率的水文和气象数据,引入深度学习模型、探索模型参数自动优化方法以及结合短期气象预报数据等方法,实现兼具时效性和准确性的动态预测,进一步提升咸潮预测模型的精度和实用性,为水资源管理和防灾减灾提供更加有力的科技支撑。

参考文献:

[1] 钟桂芳.灰色变基模型在密云水库长期水文预报中的应用[J].北京水利,1996(3):47-50.
 [2] 华莉娟,俞永强.海洋环流的长期变化和预估[J].气候变化研究进展,2022,18(1):19-30.
 [3] 位帅,黄宇铭,林凯荣,等.磨刀门水道咸潮演变规律及咸界变化分析[J].中国农村水利水电,2025(2):88-96.
 [4] 严鑫,孙昭华,谢翠松,等.基于经验模型的长江口南支上段压咸临界流量[J].地理学报,2019,74(5):935-947.
 [5] 孙娜.机器学习理论在径流智能预报中的应用研究[D].武汉:华中科技大学,2019.
 [6] 陈记臣,查悉妮,卓文珊,等.基于AdaBoost算法和BP神经网络的咸潮模拟研究[J].人民珠江,2017,38(1):5-10.
 [7] 刘德地,陈晓宏.基于偏最小二乘

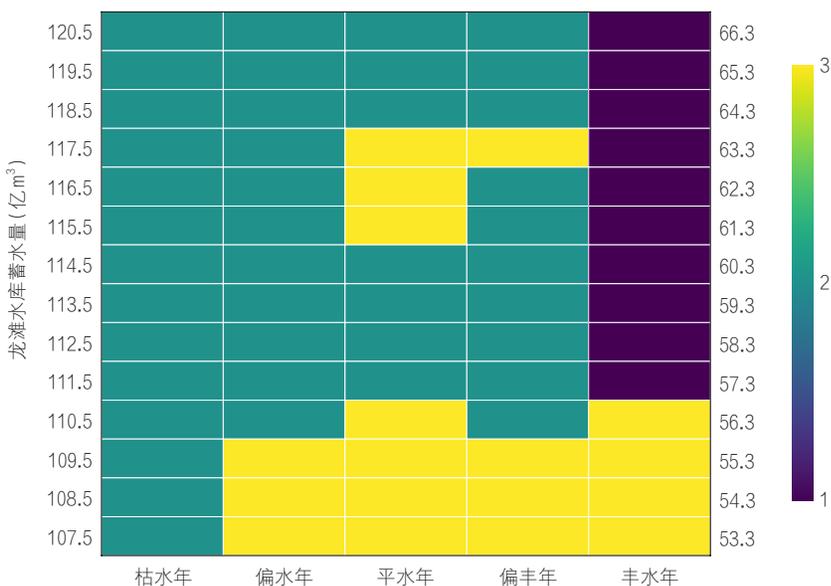


图8 精确设计场景下泗盛围站咸潮年际尺度预测结果

- 回归与支持向量机耦合的咸潮预报模型[J]. 中山大学学报(自然科学版), 2007(4):89-92.
- [8] 苏程佳, 陈莎, 陈晓宏. 基于随机森林模型的咸潮预报[J]. 热带地理, 2018, 38(3):432-439.
- [9] 李平日, 林晓东, 黄光庆. 东江三角洲地貌特征[J]. 地理研究, 1991(2):11-19.
- [10] 吴宏旭, 张庆通, 曾维汉. 东江三角洲咸潮活动现状初探[J]. 广东水利水电, 2001(5):34-36.
- [11] 王凤恩, 林中源, 邹华志, 等. 珠江流域东江三角洲压咸流量探讨[J]. 中国防汛抗旱, 2024, 34(11):44-50.
- [12] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016.
- [13] 王连华. 冰峪沟水文站洪峰流量频率计算分析[J]. 东北水利水电, 2017, 35(1):37+47.
- [14] 孔兰, 陈晓宏, 闻平, 等. 2009/2010年枯水期珠江口磨刀门水道强咸潮分析[J]. 自然资源学报, 2011, 26(11):1858-1865.
- [15] WANG Q, LUO Z H, HUANG J C, et al. A Novel Ensemble Method for Imbalanced Data Learning: Bagging of Extrapolation-SMOTE SVM.[J]. Computational intelligence and neuroscience, 2017: 1827016.
- [16] 赵铜铁钢, 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+38.
- [17] 余胜男, 陈元芳, 顾圣华, 等. 随机森林在降水量长期预报中的应用[J]. 南水北调与水利科技, 2016, 14(1):78-83.
- [18] BREIMAN L. Random Forests.[J]. Machine Learning, 2001, 45(1):5-32.
- [19] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3):32-38.
- [20] FABIAN PEDREGOSA, VAROQUAUX_G, ALEXANDRE GRAMFORT, et al. Scikit-learn: Machine Learning in Python.[J]. JOURNAL OF MACHINE LEARNING RESEARCH, 2011(12): 2825-2830.
- [21] SYLVAIN ARLOT, ALAIN CELISSE. A survey of cross-validation procedures for model selection[J]. Statistics surveys, 2010(4):40-79.
- [20] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine Learning in Python. [J]. Journal of Machine Learning Research, 2011, 12:2825.
- [22] 殷兆凯, 廖卫红, 王若佳, 等. 基于长短时记忆神经网络(LSTM)的降雨径流模拟及预报[J]. 南水北调与水利科技, 2019, 17(6):1-9+27.
- [23] 鲍振鑫, 张建云, 王国庆, 等. 基于水文模型与机器学习集合模拟的水沙变异归因定量识别——以黄河中游窟野河流域为例[J]. 水科学进展, 2021, 32(4):485-496.
- [24] 孙娜. 机器学习理论在径流智能预报中的应用研究[D]. 武汉:华中科技大学, 2019.
- [25] 苏辉东, 贾仰文, 倪广恒, 等. 机器学习在径流预测中的应用研究[J]. 中国农村水利水电, 2018(6):40-43+48.

责任编辑 刘磊宁