



# 基于机器学习的长江航道水位 集合概率预报模型\*

李雪<sup>1</sup>, 蔡孝燕<sup>1</sup>, 林妙丽<sup>2,3</sup>, 范雯霆<sup>4</sup>

- (1. 长江涪陵航道处, 重庆 408000; 2. 长江航道规划设计研究院, 湖北 武汉 430040;  
3. 长江航道勘察设计院(武汉)有限公司, 湖北 武汉 430040;  
4. 南京水利科学研究院, 江苏 南京 210029)

**摘要:** 针对目前水位预测模型预报不确定性以及水位预测精度不高等问题, 通过综合多个模型预报结果, 提出一种融合机器学习与贝叶斯模型平均法(BMA)的航道水位集合概率预报框架。以长江航道上荆江河段为研究区域, 采用随机森林(RF)、支持向量机(SVM)以及人工神经网络(ANN)进行了沙市站和新厂站的水位预测, 结果表明水位预测精度表现为 RF>SVM>ANN, 3种模型预测精度整体均处于较优状态。基于3种机器学习模型预测结果, 采用贝叶斯模型平均法进行了水位集合概率预报, BMA模型在沙市站和新厂站的水位预测结果相较于RF得到进一步提升, 并准确获取了未来水位在90%概率下可能的出现范围。研究方法有效提升了航道水位预测精度并实现了概率预报, 能够为船舶通航安全提供技术支持。

**关键词:** 长江航道; 水位预测; 机器学习; 贝叶斯模型平均法; 概率预报

中图分类号: U612; P338

文献标志码: A

文章编号: 1002-4972(2024)10-0158-06

## Machine learning-based ensemble probability forecast model of Yangtze River channel water level

LI Xue<sup>1</sup>, CAI Xiaoyan<sup>1</sup>, LIN Miaoli<sup>2,3</sup>, FAN Wenting<sup>4</sup>

- (1. Changjiang Fuling Waterway Division, Chongqing 408000, China;  
2. Changjiang Waterway Institute of Planning and Design, Wuhan 430040, China;  
3. Changjiang Waterway Survey Design and Research Institute (Wuhan), Wuhan 430040, China;  
4. Nanjing Hydraulic Research Institute, Nanjing 210029, China)

**Abstract:** Aiming at the uncertainties and low accuracy of current water level prediction models, this paper proposes a probabilistic forecast framework of channel water level by integrating machine learning and Bayesian model averaging method. Taking Jingjiang River section of the upper Yangtze River channel as the research area, random forest (RF), support vector machine (SVM) and artificial neural network (ANN) are used to predict the water level of Shashi station and Xinchang station. The results show that RF has the highest prediction accuracy, followed by SVM, and ANN has a relatively lower prediction accuracy, but all of the three machine learning models provide relatively excellent prediction accuracy. Based on the prediction results of three machine learning models, Bayesian model averaging method is used to forecast the water level ensemble probability. Compared with RF, BMA water level prediction results are further improved, and the possible range of future water level under 90% probability is accurately obtained. The method proposed in this study effectively improves the forecast accuracy of waterway water level and realizes probabilistic forecast, which can provide technical support for ship navigation safety.

**Keywords:** Yangtze River channel; water level forecast; machine learning; bayesian model averaging; probabilistic forecast

收稿日期: 2024-01-10

\*基金项目: 长江航道规划设计研究院自主立项项目(2023-026-6-Z-Y)

作者简介: 李雪(1994—), 女, 硕士, 工程师, 从事航道水位模拟与预测研究。

长江航道是全球货运量最大的内河航道, 在水路交通运输中有着极其重要的地位, 而航道的水位变化情况对于水路运输过程中载货船舶通航安全具有重要的影响<sup>[1]</sup>。因此, 开展长江航道的水位预测研究对于保障船舶通航安全具有重要意义。

数学模型是进行水位预测的重要手段, 可以分为基于水动力过程的机理模型和基于数据驱动的统计模型。机理模型由于输入数据类型要求多, 模型构建过程复杂, 水位预测精度很大程度上取决于边界条件的预测精度, 使得其应用受到了一定的限制<sup>[2]</sup>; 统计模型通过数理统计方法挖掘预测变量与其影响因子之间的关系, 进而实现水位的预测, 数据类型需求较少, 操作简便。早期应用主要以线性回归为主, 但回归方法对于变量间的非线性关系捕捉能力较差, 预测精度较低<sup>[3-4]</sup>。随着人工智能的发展, 机器学习方法如人工神经网络(ANN)<sup>[5]</sup>、支持向量机(SVM)<sup>[6]</sup>以及随机森林(RF)<sup>[7]</sup>等模型有效提升了水位的预测精度, 王蒙蒙等<sup>[8]</sup>采用支持向量机进行洞庭湖水位的快速预测, 刘晓阳等<sup>[9]</sup>利用 RF、SVM 和 ANN 进行三峡水库小时尺度的坝前水位预测, 均取得了较好的预测效果。然而, 由于输入数据、模型参数以及自身结构设计的不确定性, 采用单个模型预报

会导致预报结果存在不确定性<sup>[10]</sup>。贝叶斯模型平均法(BMA)是一种集合概率预报方法<sup>[11]</sup>, 通过将单个模型预报结果的后验概率分布进行加权平均获得综合多个模型优势的预报值和不确定区间, 进而实现高精度的集合概率预报, 将其与多个水位预测模型结合, 具备进一步提升预报精度的潜力。

本研究提出一种融合机器学习模型和 BMA 的航道水位集合概率预报框架, 以长江航道为研究对象, 首先采用多个机器学习模型(ANN、SVM、RF)分别进行长江中游碍航段(荆江河段)代表性站点的水位预测, 对比分析机器学习模型的水位预测效果, 再基于 BMA 进行代表性站点的水位集合概率预报, 进一步提升水位预报精度, 并给出未来水位可能的出现范围, 量化水位预报风险, 以期为长江航道船舶通航安全提供技术支撑。

## 1 模型原理

### 1.1 模型基本原理

在进行长江航道水位预测时, 以 3 个不同设计原理的代表性机器学习方法(ANN、SVM 和 RF)为基础, 构建水位预测模型。3 个模型的结构见图 1。

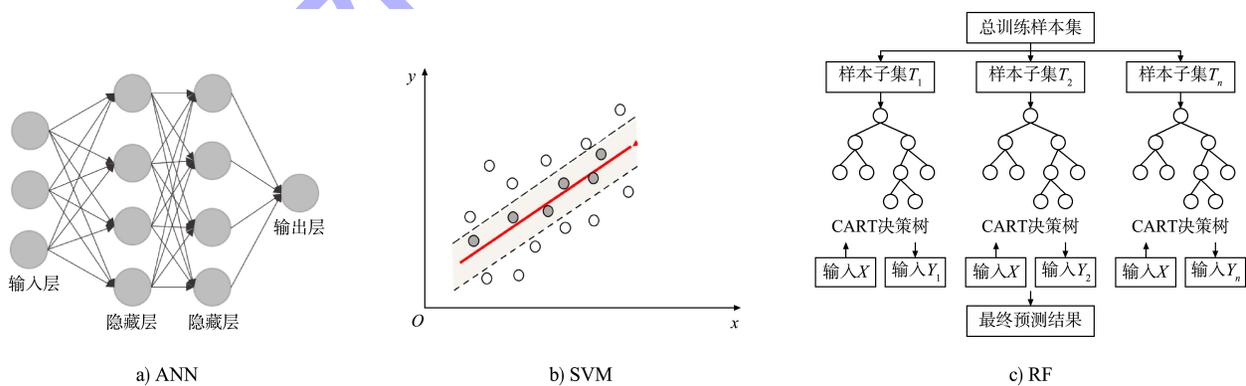


图 1 ANN、SVM 和 RF 结构

BMA 是一种基于贝叶斯理论的统计分析方法, 能够考虑模型结构不确定性对预测结果的影响<sup>[12]</sup>。它以实测样本隶属于某一模型的后验概率为权重, 将各个模型预测结果的后验分布进行加权平均, 进而获得预报量的概率分布函数, 并推

导均值集合预报结果和不确定性区间。

假设  $y$  为待预测变量,  $D_{\text{obs}} = \{y_1, y_2, y_3, \dots, y_T\}$  为给定的实测数据资料,  $M = \{M_1, M_2, M_3, \dots, M_k\}$  代表所有用于预测的模型所组成的模型空间。则待预测变量综合预报结果的后验分布  $p(y | D_{\text{obs}})$  为:

$$p(y | D_{\text{obs}}) = \sum_{i=1}^k P(M_i | D_{\text{obs}}) p(y | M_i, D_{\text{obs}}) \quad (1)$$

式中： $M_i$  表示第  $i$  个模型， $P(M_i | D_{\text{obs}})$  为单个模型预测结果的后验概率，即 BMA 模型的权重； $p(y | M_i, D_{\text{obs}})$  为单个模型预测结果的后验分布。

基于 BMA 模型获得待预测变量的后验分布后，其均值预报结果  $E[y | D_{\text{obs}}]$  和方差  $V[y | D_{\text{obs}}]$  如式(2)、(3)所示。基于式(3)推导的方差，采用期望最大化算法<sup>[13]</sup>和蒙特卡洛抽样方法则可以获得待预报变量处于某一置信水平的不确定性区间，实现概率预报。

$$E[y | D_{\text{obs}}] = \int_{-\infty}^{+\infty} y \cdot \left( \sum_{i=1}^k P(M_i | D_{\text{obs}}) p(y | M_i, D_{\text{obs}}) \right) dy \quad (2)$$

$$\text{VAR}[y | D_{\text{obs}}] = \int_{-\infty}^{+\infty} [y - E(y | D_{\text{obs}})]^2 \cdot \left( \sum_{i=1}^k P(M_i | D_{\text{obs}}) p(y | M_i, D_{\text{obs}}) \right) dy \quad (3)$$

### 1.2 精度评价指标

进行基于机器学习以及 BMA 的长江航道水位确定性预测时，预测精度评价指标为平均绝对误差  $M$ 、均方根误差  $R$  以及纳什系数  $N$ ：

$$M = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

$$R = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

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

式中： $n$  为样本数量， $y_i$  为实测水位数据， $\bar{y}$  为实测水位数据平均值， $\hat{y}_i$  为预测水位数据。

进行基于 BMA 的长江水位概率预测时，概率预测精度评价指标为区间覆盖度  $C$  和平均相对带宽  $B$ ：

$$C = \frac{N_c}{N} \times 100\% \quad (7)$$

$$B = \frac{1}{N} \sum_{i=1}^N \left( \frac{\bar{h}_i - h_i}{h_i} \right) \quad (8)$$

式中： $N_c$  为落在置信区间的水位实测值样本数量，

$N$  为所有的水位实测值样本数量， $\bar{h}_i$  和  $h_i$  为第  $i$  个时刻预报区间的上下限， $h_i$  为第  $i$  个时刻的实测水位。

## 2 模型应用

### 2.1 研究区域

长江航道上起宜宾，下至长江口，全长 2 380 km，其中，长江航道中游荆江河段以枝城为起点，以城陵矶为终点，全长 347 km，是典型的弯曲型河道，包括石首、调关等 10 个弯曲段，且大部分属于急弯段，通航条件相对较差。

### 2.2 数据收集与处理

研究进行上荆江河段 2 个急弯段(沙市站和新厂站)的水位预测，由于下游水位会受到上游流量影响，研究选用的数据资料包括上游干流宜昌站、枝城站和支流高坝洲站 2017—2021 年的逐日流量监测数据以及沙市站和新厂站 2017—2021 年的逐日水位监测数据，水文监测站点分布见图 2。在利用机器学习模型进行水位预测时，需要对流量水位数据进行归一化处理，避免量纲差异导致的预报精度降低，采用 min-max 方法将数据归一化至 [0, 1]。研究以 2017—2019 年共 3 a 的数据作为训练集，以宜昌站、枝城站和高坝洲站的流量作为模型输入，以同时刻沙市站和新厂站的水位作为输出，进行 3 个机器学习模型的训练。以 2020—2021 年的数据作为测试集，进行模型预测效果的验证。

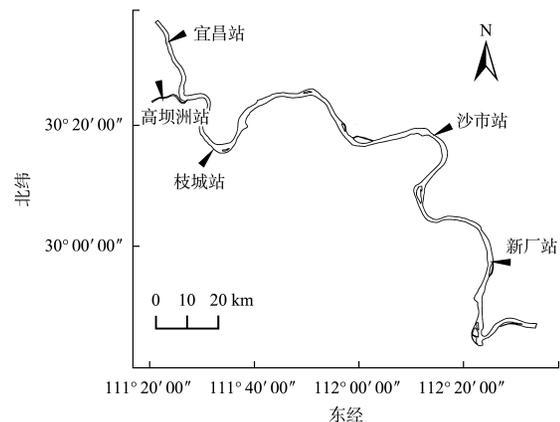


图 2 研究区域与水文监测站点分布

2.3 结果与分析

模型参数是影响水位预测精度的重要因素,

采用网格搜索法进行参数寻优后, 3 个模型在 2 个站点的键参数最优取值见表 1。

表 1 3 个机器学习模型在 2 个站点处的键参数取值

水文 站点	ANN					SVM			RF	
	隐藏层层数	隐藏层神经元	激活函数	学习率	梯度下降算法	惩罚因子	核参数	核函数	节点数	回归数数量
沙市站	3	64	tanh	0.005	Adam	1.15	1.23	rbf	2	106
新厂站	3	64	tanh	0.005	Adam	0.81	3.03	rbf	2	135

3 个机器学习模型在沙市站和新厂站训练期和测试期的预测水位和实测水位对比情况见图 3, ANN 模型的预测水位整体略低于 1:1 线, 表明 ANN 的预测结果偏低, 一方面可能是由于在进行 ANN 模型训练时采用的是同时刻的输入流量和输出水位, 未充分考虑上下游流量水位关系的滞时特征, 使得预测水位低于下游实测值; 另一方面可能是由于训练集仅为 3 a, 训练样本量数据相对较少, ANN 对于流量水位数据之间的关系学习不

够充分。而 RF 和 SVM 的预测水位则集中在 1:1 线附近, 预测水位与实测水位吻合较好。3 个模型在 2 个站点训练期和测试期的精度评价结果见表 2, 3 个模型在测试期和精度评价指标结果与训练期的差距较小, 表明模型参数取值合理, 不存在过拟合现象。总体而言, RF 模型的水位预测精度最高, 其次是 SVM, ANN 模型的预测精度相对较低, 但 3 个模型的水位预测效果均处于较优状态。

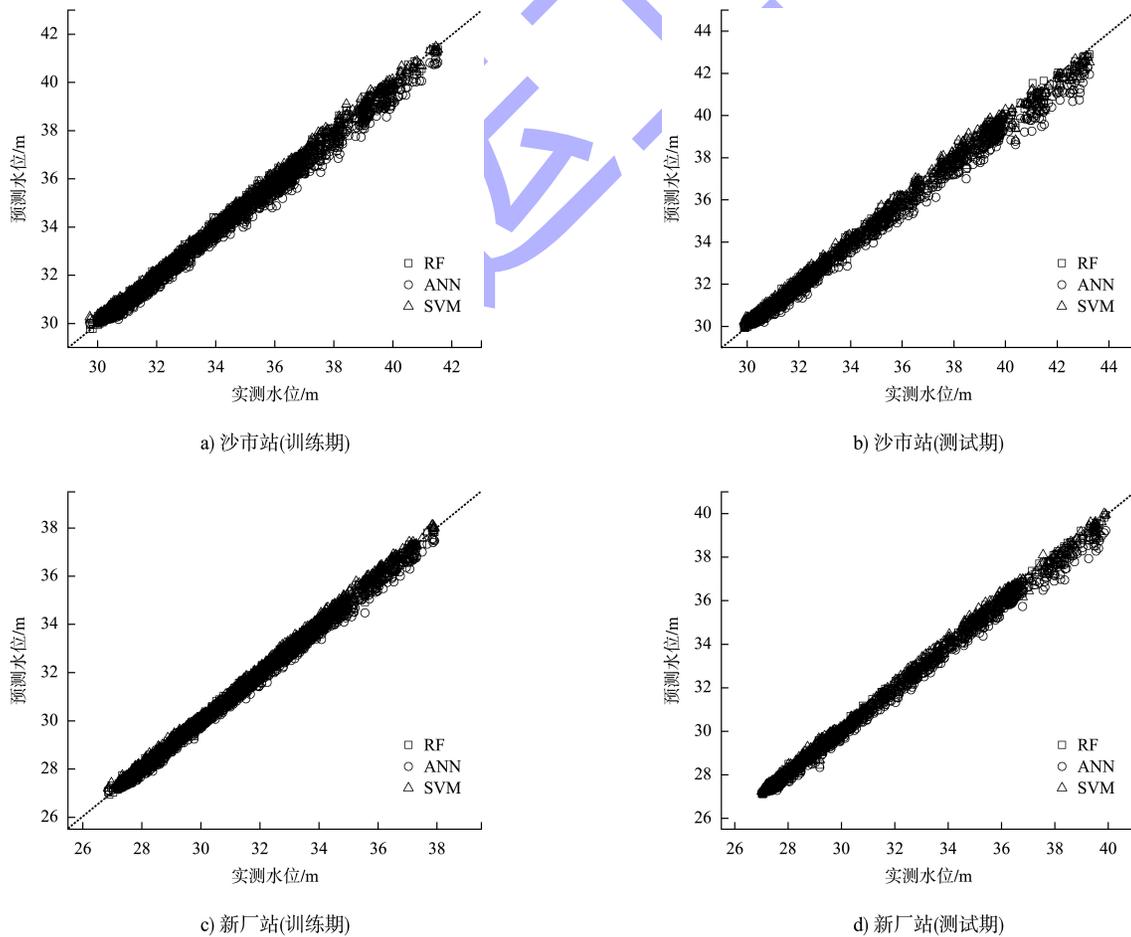


图 3 RF、ANN 和 SVM 在两个站点训练期和测试期的预测结果对比

表2 RF、ANN、SVM 在两个站立训练期和测试期的精度评价结果

水文站点	时期	模型	精度指标		
			M	R/m	N
沙市站	训练期	RF	0.026 4	0.161 0	0.997 7
		ANN	0.205 2	0.452 9	0.976 8
		SVM	0.027 2	0.164 9	0.996 9
	测试期	RF	0.048 8	0.220 9	0.996 5
		ANN	0.297 8	0.545 7	0.978 9
		SVM	0.082 6	0.287 3	0.994 1
新厂站	训练期	RF	0.014 2	0.119 1	0.998 3
		ANN	0.073 7	0.271 4	0.991 2
		SVM	0.030 1	0.173 6	0.996 4
	测试期	RF	0.024 0	0.154 8	0.998 2
		ANN	0.112 6	0.335 6	0.991 6
		SVM	0.047 4	0.217 7	0.996 5

采用 BMA 进行多个模型预报结果综合时，各个模型的权重分配情况见图 4，沙市站 RF、SVM 和 ANN 的权重分别为 0.459 5、0.378 0 和 0.162 5，新厂站 RF、SVM 和 ANN 的权重分别为 0.436 9、0.385 5 和 0.177 6，3 个模型在 2 个站点的权重整体表现为 RF>SVM>ANN，这与 3 个模型在 2 个站点的水位预测精度保持一致，表明 BMA 模型能够较好地识别出优势模型。

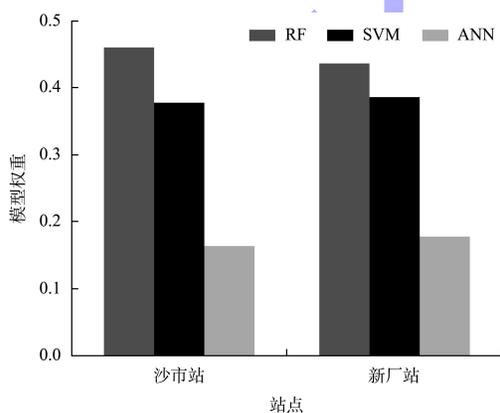


图 4 RF、SVM 和 ANN 模型权重

BMA 模型的水位集合预报结果、不确定性区间及精度评价结果见图 5，沙市站水位集合预报结果的  $M$ 、 $R$  和  $N$  分别为 0.042 7、0.211 3 m 和 0.997 1，新厂站水位集合预报结果的  $M$ 、 $R$  和  $N$  分别为 0.023 5、0.150 4 m 以及 0.998 4，相较于 RF 模型，BMA 模型的水位集合预报结果在沙市和新厂站均有不同程度的提升，这主要是因为 BMA

本质是一种基于概率的加权平均方法，能够综合多个模型的预测结果使得 BMA 水位集合预报结果更加接近实测值，然而，当多个模型的预测结果整体趋势均表现为高于或低于实测值时，可能会限制 BMA 模型对于水位预测精度的提升幅度。但总体上采用 BMA 模型进行水位集合预测时，无需进行模型间复杂的对比与选择，即可获得较好的水位预测效果。同时，BMA 模型 90% 置信区间下的概率预报结果表明：沙市站和新厂站的区间覆盖度均为 100%，区间平均相对带宽分别为 0.026 4 和 0.021 1，均处于较小状态，基于 BMA 的水位预测模型能够准确地给出未来水位在 90% 的概率下可能的出现范围，实现较好的概率预报，为风险管理提供支撑。

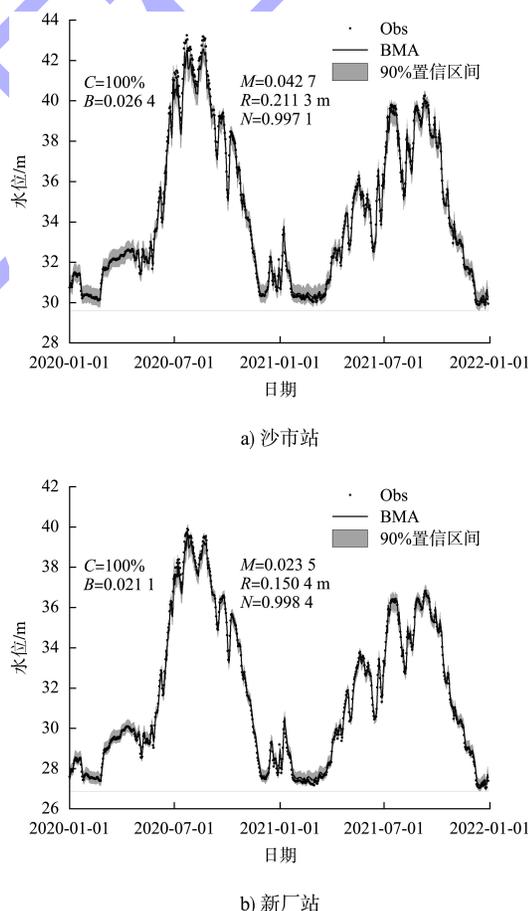


图 5 BMA 集合预报结果及 90% 置信区间

### 3 结论

1) 3 种机器学习模型在长江上荆江航道沙市站和新厂站的水位预测中，预测精度表现为 RF>SVM>ANN。

2) 基于贝叶斯模型平均法的长江航道水位预测精度相较于 3 个机器学习模型有进一步的提升, 同时提供了未来水位在 90% 概率下的可能出现范围, 实现了概率预报。

#### 参考文献:

- [1] 李瀛, 雷雪婷, 黄纲, 等. 长江下游武汉—安庆河段中洪水期航道尺度潜力研究[J]. 水运工程, 2023(7): 133-139, 145.
- [2] LI G, ZHU H Y, JIAN H, et al. A combined hydrodynamic model and deep learning method to predict water level in ungauged rivers [J]. Journal of hydrology, 2023, 625: 130025.
- [3] 范先友, 李朝阳, 杨传华, 等. 基于多元线性回归分析的水位短期预测方法[J]. 水运工程, 2014(1): 159-162.
- [4] 熊波. 三峡蓄水后长江中游沿程河道水位的预测方法[J]. 水运工程, 2018(9): 25-32.
- [5] AGATONOVIC KUSTRIN-S, BERESFORD R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research[J]. Journal of pharmaceutical and biomedical analysis, 2000, 22(5): 717-727.
- [6] BORG A, BOLDT M. Using VADER sentiment and SVM for predicting customer response sentiment [J]. Expert systems with applications, 2020, 162: 113746.
- [7] LU H, MA X. Hybrid decision tree-based machine learning models for short-term water quality prediction [J]. Chemosphere, 2020, 249(6): 126169.
- [8] 王蒙蒙, 戴凌全, 戴会超, 等. 基于支持向量回归的洞庭湖水位快速预测[J]. 排灌机械工程学报, 2017, 35(11): 954-961.
- [9] 刘晓阳, 姚华明, 张海荣, 等. 基于机器学习的三峡水库小时尺度坝前水位预测[J]. 人民长江, 2023, 54(2): 147-151.
- [10] 金君良, 舒章康, 陈敏, 等. 基于数值天气预报产品的气象水文耦合径流预报[J]. 水科学进展, 2019, 30(3): 316-325.
- [11] 刘章君, 郭生练, 许新发, 等. 贝叶斯概率水文预报研究进展与展望[J]. 水利学报, 2019, 50(12): 1467-1478.
- [12] 巴欢欢, 郭生练, 钟逸轩, 等. 考虑降水预报的三峡入库洪水集合概率预报方法比较[J]. 水科学进展, 2019, 30(2): 186-197.
- [13] 蒋晓蕾, 梁忠民, 胡义明, 等. 洪水概率预报评价指标研究[J]. 湖泊科学, 2020, 32(2): 539-552.

(本文编辑 王传瑜)

## · 消 息 ·

### 中交集团召开战略研究院研究中心成立会议

近日, 中交集团在京召开战略研究院研究中心成立会议并举行授牌仪式。

会议充分肯定了战略研究院成立以来所取得的工作成绩。会议指出, 进一步推进中交集团新型智库建设, 发挥战略研究服务作用, 必须坚持党的领导, 把握正确方向; 坚持围绕大局, 服务中心工作; 坚持价值创造, 引领发展方向; 坚持科学精神, 鼓励大胆探索。

会议要求, 各部门、各单位要统一思想、形成合力, 共同将中交集团战略研究工作做深做实, 做出成效。一是要强化协同体系, 为战略研究和智库建设等工作赋能; 二是要创新人才机制, 鼓励研究中心与战略研究院形成畅通的人才双向流动格局; 三是要结合新型智库建设特点与要求, 建立激励机制; 四是要加强成果转化, 搭建研究成果转化平台。

会上, 中国港湾、中交二航局、中交一公局集团、中交公规院、中交水规院、中交城投、中交资本 7 家单位被授予中交集团战略研究院相关研究中心标牌。

战略研究院成立的研究中心包括未来交通研究中心、未来城市研究中心、海外研究中心、绿色交通研究中心、高质量发展研究中心、新质生产力发展研究中心、产业金融研究中心等 7 个研究中心, 旨在进一步增强中交集团战略研究力量, 完善战略研究布局, 依托相关单位进一步推动企业重点课题研究, 打造具有中交集团特色的国内一流创新型智库。

[https://www.ccccltd.cn/news/gsyw/202410/t20241005\\_216452.html](https://www.ccccltd.cn/news/gsyw/202410/t20241005_216452.html) (2024-10-04)