

罗虎, 马荣. 基于机器学习的磁场强化型人工湿地水质的预测[J]. 净水技术, 2026, 45(1): 157–163.

LUO H, MA R. Prediction of water quality for magnetic field enhanced constructed wetlands based on machine learning[J]. Water Purification Technology, 2026, 45(1): 157–163.

## 基于机器学习的磁场强化型人工湿地水质的预测

罗 虎, 马 荣\*

(西南林业大学机械与交通学院, 云南昆明 650224)

**摘 要** 【目的】 磁场强化人工湿地(MF-CW)作为绿色环保的污水处理技术, 由于其在污水脱氮方面表现突出, 近年来被用于对生活污水进行处理, 并展现出了显著的优势和应用潜力, 机器学习(ML)技术通过模拟和分析水质中各种影响参数在提高预测准确性方面表现出了极佳效果, 为此通过 ML 建立一个高精度的模型用来预测 MF-CW 系统出水氨氮浓度, 对于推进其技术发展至关重要。【方法】 本文基于磁场强化人工湿地, 以实际测量数据作为模型输入, 然后构建不同的机器学习算法模型进行对比, 本文建立了 4 种机器学习算法模型[反向传播神经网络(BPNN)、粒子群算法优化 BPNN(PSO-BPNN)、长短期记忆(LSTM)、随机森林(RF)]用来预测 MF-CW 系统的氨氮浓度, 并使用 3 个性能指标[决定系数( $R^2$ )、均方误差(RMSE)、平均绝对误差(MAE)]对 4 种模型的预测性能进行评估。【结果】 RF 模型在预测 MF-CW 系统出水质量方面的准确性优于其他 3 种模型, 实现了 0.945 9 的  $R^2$ 、4.197 1 mg/L 的 RMSE 和 2.805 9 mg/L 的 MAE, 相较于 BPNN、PSO-BPNN 和 LSTM, RF 模型的  $R^2$  分别提高了 37.17%、16.82%、19.06%, RMSE 降低了 50.30%、43.10%、45.78%, MAE 降低 62.63%、50.53%、53.47%。【结论】 证实了 RF 模型采用的集成学习方法在处理复杂的水质非线性数据关系方面非常出色, 并且具有很强的抗噪性和鲁棒性, 本文为推进磁场强化人工湿地污水处理技术发展应用提供了重要技术支持。

**关键词** 磁场(MF) 人工湿地(CW) 水质预测 机器学习(ML) 随机森林(RF)

中图分类号: X703 文献标志码: A 文章编号: 1009-0177(2026)01-0157-07

DOI: 10.15890/j.cnki.jsjs.2026.01.019

## Prediction of Water Quality for Magnetic Field Enhanced Constructed Wetlands Based on Machine Learning

LUO Hu, MA Rong\*

(School of Mechanical Engineering and Transportation, Southwest Forestry University, Kunming 650224, China)

**Abstract** [Objective] The magnetic field-enhanced constructed wetland (MF-CW) is an environmentally friendly wastewater treatment technology. Due to its outstanding performance in nitrogen removal of wastewater, it has been increasingly used for domestic wastewater treatment in recent years, demonstrating significant advantages and application potential. Machine learning (ML) techniques have shown excellent effectiveness in improving prediction accuracy by simulating and analyzing various influencing parameters in water quality. Therefore, establishing a high-precision ML model to predict the effluent ammonia nitrogen concentration of the MF-CW system is crucial for advancing its technological development. [Methods] Based on the magnetic field-enhanced constructed wetland system and using actual measured data as model inputs, this paper constructed and compared different machine learning algorithm models. Four ML models were developed, including backpropagation neural network (BPNN), particle swarm optimization-based BPNN (PSO-BPNN), long short-term memory (LSTM), and random forest (RF), to predict the ammonia nitrogen

[收稿日期] 2025-08-10

[基金项目] 云南省科学技术厅科技计划(202501BD070001-097)

[作者简介] 罗虎(2000—), 男, 硕士研究生, 研究方向为污水处理, E-mail: 1507510588@qq.com。

[通信作者] 马荣(1987—), 男, 副教授, 研究方向为污水处理, E-mail: mr@swfu.edu.cn。

concentration of the MF-CW system. Three performance metrics including coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the predictive performance of the four models. [Results] The RF model outperformed the other three models in predicting the effluent quality of the MF-CW system, achieving an  $R^2$  of 0.945 9, an RMSE of 4.197 1 mg/L, and an MAE of 2.805 9 mg/L. Compared to the BPNN, PSO-BPNN, and LSTM models, the  $R^2$  of the RF model increased by 37.17%, 16.82%, and 19.06%, respectively; RMSE decreased by 50.30%, 43.10%, and 45.78%, respectively; and MAE decreased by 62.63%, 50.53%, and 53.47%, respectively. [Conclusion] The results confirm that the ensemble learning method adopted by the RF model excels in handling complex nonlinear relationships in water quality data and demonstrates strong noise resistance and robustness. This study provides important technical support for advancing the development and application of magnetic field-enhanced constructed wetland technology in wastewater treatment.

**Keywords** magnetic field (MF) constructed wetland (CW) water quality prediction machine learning (ML) random forest (RF)

人工湿地(CW)作为绿色环保的污水处理技术逐渐被各个地区广泛应用<sup>[1]</sup>。然而,由于CW天然存在着脱氮效率低、对污染物处理效果不佳等问题,极大地限制了CW的开发和应用<sup>[2]</sup>,为提高CW的污水处理能力,人们通过磁场(MF)强化技术改善了CW污水处理工艺的长期性能,增强了污水的物理化学特性,加快了物质交换和微生物吸收<sup>[3]</sup>。研究<sup>[4]</sup>表明:MF能显著影响水质中氨氮的去除性能并促进碳代谢和氮循环,研究该系统需要大量的水质数据作为参考,然而传统的试验收集需要大量的人力物力投入,为此通过建立一个高精度的模型来预测该系统出水氨氮浓度变化,对推进MF-CW系统的研究和治理至关重要。机器学习(ML)技术的发展在环境领域的应用展现了相当大的潜力<sup>[5]</sup>,并且在模拟和优化不同水质变量之间错综复杂的非线性关系方面获得了很大的成功<sup>[6]</sup>。ML算法模型被应用于各种类型水质预测,Xu等<sup>[7]</sup>使用反向传播神经网络(BPNN)构建了海洋牧场水质参数预测模型,表明BPNN优秀的非线性映射能力和存在容易陷入局部极值的特性;Yan等<sup>[8]</sup>采用粒子群优化(PSO)算法优化BPNN,对北京北海湖的水质进行预测,表明该算法具有较好的收敛性和鲁棒性;Huang等<sup>[9]</sup>采用长短期记忆(LSTM)及其改进模型对湖南省洞庭湖水质进行预测,展示了LSTM模型在对水质进行时间序列预测方面发挥着重要作用,Zanoni等<sup>[10]</sup>使用随机森林(RF)算法建立了水温、溶解氧(DO)、砷、硫酸盐和氯化物浓度以及电导率(EC)的区域模型,结果表明:RF在检测非线性关系方面更加灵活和有效。对MF-CW系统出水氨氮预测方面的研究还未发展完善。本文的目的是通过4种常见的ML算法[BPNN、粒子群优化BPNN(PSO-

BPNN)、LSTM、RF]来模拟和预测MF-CW的出水水质。通过对模型的预测效果进行比较,以便寻找预测效果最佳模型,相关人员可以根据预测数据对MF-CW系统实施更加科学的管理研究和优化措施。

## 1 材料和方法

### 1.1 数据来源

本文在传统人工湿地的基础上加入了磁场强化,构建磁场强化人工湿地(MF-CW)系统,系统装置如图1所示。系统尺寸为40 cm×30 cm×30 cm,系统分为2个单元:生物降解单元和磁场强化单元,生物降解单元中放置了28 cm生物膜球(直径为50 mm)。磁场强化单元基质配置由左到右填充:26 cm的砾石层(直径为4~6 mm)和260 cm的浮水陶粒层(直径为3~4 mm),磁板尺寸厚度为0.5 cm,宽度为28 cm,高度为28 cm,两磁板间距为120 cm,磁场强度设置为30 mT,并且种植有茭草。污水来源于西南林业大学污水处理厂,其水质为西南林业大学生活污水。MF-CW系统污水运行采用连续循环运行模式。

装置稳定运行后,设置的水力停留时间(HRT)为24、48、72、96、144 h。每6 d更换系统装置的进水,收集水样时间为2023年4月1日—2024年3月16日,共计350 d,因装置维修造成了部分数据缺失,共收集有效数据330组,监测数据主要有以下几个指标:pH、氧化还原电位(ORP)、DO、EC、温度(T)、总磷(TP)、氨氮、化学需氧量(COD),其水质数据如表1所示。由于该套装置污水来源于校园污水,学校人数变化和降雨等因素均会造成污水浓度波动较大等原因,原始数据按照70%训练集和30%测试集的比例划分。

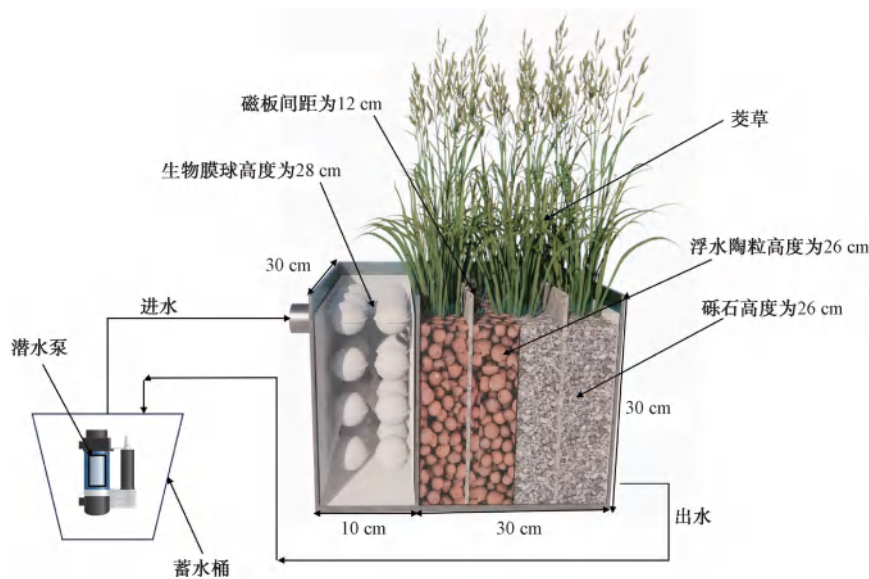


图 1 MF 强化型人工湿地装置

Fig. 1 Schematic of MF Enhanced Constructed Wetland Set-Up

表 1 水质数据  
Tab. 1 Water Quality Data

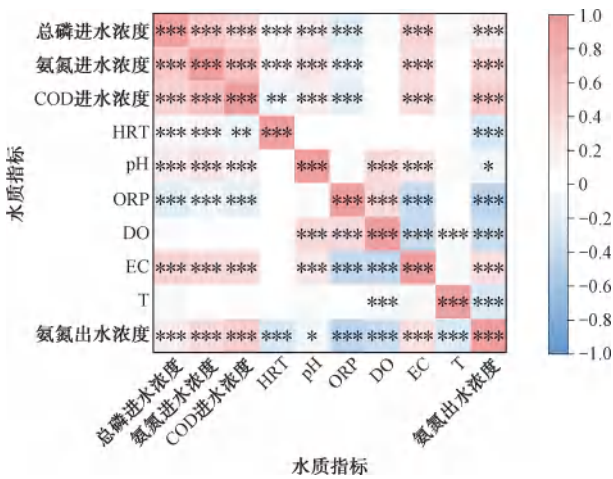
指标	最大值	最小值	平均值	标准偏差
pH 值	8. 730	6. 560	7. 814	0. 413
ORP/mV	418. 000	-187. 300	-14. 369	78. 217
DO/(mg·L <sup>-1</sup> )	5. 810	1. 430	3. 058	1. 076
EC/(mS·cm <sup>-1</sup> )	1 197. 000	264. 000	723. 551	225. 925
T/℃	25. 120	17. 200	21. 756	1. 697
进水 TP/(mg·L <sup>-1</sup> )	10. 898	0. 936	4. 397	2. 424
进水氨氮/(mg·L <sup>-1</sup> )	98. 061	9. 139	47. 403	18. 754
进水 COD/(mg·L <sup>-1</sup> )	458. 000	56. 278	212. 599	91. 744
出水氨氮/(mg·L <sup>-1</sup> )	58. 439	0. 048	16. 565	17. 282

1. 2 相关性分析

相关性分析有助于理解变量之间的关系,确定有利于改善数据的模式和趋势,以及评估各种输入属性,通过相关性分析筛选模型输入指标,可有效降低模型的过拟合风险和提升训练效率<sup>[11]</sup>,确保模型的预测准确性。为此本文针对氨氮出水浓度的数据进行皮尔逊相关性分析,筛选与其相关性较强的指标作为模型输入指标( $P<0.01$ ),其相关性热图如图 2 所示。根据相关性分析结果,本文选取以下 8 个水质指标作为模型的输入参数,分别为进水 TP 浓度、进水氨氮浓度、进水 COD 浓度、HRT、ORP、DO、EC、T。

1. 3 原始数据预处理

在对原始数据的预处理中,为了减少由于输入



注: \* 指代相关性系数  $P<0.05$ , \*\* 和 \*\*\* 分别代表相关性系数  $P<0.02$  和  $P<0.01$ 。

图 2 输入特征皮尔逊相关性热图

Fig. 2 Pearson Correlation Heatmap of Input Features

和输出数据之间的值存在巨大差异而引起的误差,数据集使用规定的方程进行了归一化,将数据线性缩放到 0~1。归一化如式(1)所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

其中: $x'$  ——缩放值;  
 $x$  ——原始值;  
 $x_{\max}$  和  $x_{\min}$  ——数据序列的最大值和最小值。



## 1.4 预测模型

### 1.4.1 LSTM

LSTM 神经网络作为针对序列预测而设计的改进模型,通过将门控函数与状态动态学相结合来解决传统 RNN 的消失梯度问题<sup>[12]</sup>。与简单的 RNN 相比,LSTM 引入了更复杂的单元结构。它有 3 个主门:输入门、忘记门和输出门,以及帮助保存长期信息的细胞状态和保存短期记忆的隐藏状态组成<sup>[13]</sup>。

### 1.4.2 RF

决策树(DT)算法通过对目标变量采用决策规则来自动细分数据集,RF 算法是一种基于 DT 的 ML 算法。RF 生成大量独立树,并采用投票计算最终预测。同样,RF 衡量每个特征对预测的相对重要性,这是通过最小化森林中所有树木的均方误差来估计的<sup>[14]</sup>。RF 对异常值不太敏感,并且考虑到特征和引导重采样的选择过程的随机性,通常能提供更好的概括性能。此外,RF 之所以有吸引力,是因为它能够利用数据之间的非线性关系<sup>[10]</sup>。

### 1.4.3 BPNN

BPNN 具有从输入因素中提取非线性关系的能力,具有很高的可解释性。BPNN 从输入到输出由 3 层组成,分别为输入层、隐藏层和输出层<sup>[15]</sup>。BPNN 模型特点是采用模拟预测前向推进、误差修正后向传播的过程分离运行方式,根据预测输出值与实际值之间误差减小的方向,逐层修正节点与神经元之间的连接权值和阈值,保证网络仿真和预测精度的不断提高<sup>[16]</sup>。

### 1.4.4 PSO-BPNN

PSO 算法是在研究鸟群觅食行为中发现的一种随机搜索算法,它通过模仿鸟群成员之间的相互协作和竞争来实现全局优化。在 PSO 算法中,每个个体都被视为一个粒子<sup>[17]</sup>。当粒子在预条件空间中运动时,它会根据获得的个体和全局极端不断改变自己的位置,然后通过不断修正自己的位置来更新解,从而在预条件空间中寻找最佳值<sup>[18]</sup>,因此本文采用 PSO 算法来优化传统 BPNN。

## 1.5 模型性能评估

在此次研究中,我们通过软件 MATLAB 2019a 中构建了 BPNN、PSO-BPNN、LSTM、RF 模型。此外,为了更清晰地观察上述模型的预测效果,本文通过相对均方根误差(RMSE)、决定系数( $R^2$ )和平均

绝对误差(MAE)3 个性能评估指标对模型性能进行评估。3 种指标计算分别如式(2)~式(4)。

$$R_{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t^{\text{act}} - y_t^{\text{pre}})^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t^{\text{act}} - y_t^{\text{pre}})^2}{\sum_{t=1}^n (y_t^{\text{act}} - y_t^{-\text{act}})^2} \quad (3)$$

$$M_{\text{AE}} = \frac{1}{n} \sum_{t=1}^n |y_t^{\text{act}} - y_t^{\text{pre}}| \quad (4)$$

其中: $R_{\text{MSE}}$ ——RMSE 值,mg/L;

$M_{\text{AE}}$ ——MAE 值,mg/L;

$n$ ——样本数量;

$y_t^{\text{act}}$ ——实际值;

$y_t^{\text{pre}}$ ——预测值;

$y_t^{-\text{act}}$ ——实际数据值的平均值。

## 2 结果与讨论

### 2.1 测试集的预测性能

通过比较  $R^2$ 、RMSE 和 MAE,可以更加直观地看出 4 种模型的可预测性强度。4 种算法模型的预测性能如表 2 所示。由表 2 可知,不同类型的算法模型具有不同的预测效果。BPNN 容易产生局部最小值和收敛速度慢等缺点,导致模型效果不佳,其测试集中对出水氨氮浓度的预测  $R^2$  值小于 0.7。为提高 BPNN 预测精度,本文采用粒子群算法来优化 BPNN,相比于传统的 BPNN,经过粒子群算法优化后的 BPNN 预测精度得到了很大的提高,对污染物预测  $R^2$  值提升到了 0.8 以上,提升效果很明显。相比于单一的 BPNN 模型,作为时间序列模型中的经典算法模型,LSTM 模型的预测效果要好一些,对污染物的  $R^2$  值达到 0.794 5, LSTM 因继承了 RNN 模型的记忆性、参数共享性和图灵完全性等特点,能够高效学习序列的非线性特征,这对于具有长期序列依赖性问题的数据非常有效<sup>[19]</sup>。虽然预测效果有所改进,但这还远不能达到所需的预测精度。RF 在所有指标的预测精度均优于其他模型,由表 2 可知,与其他 3 种模型相比,RF 模型的  $R^2$  值在 0.9 以上,预测精度远远超过其他模型。RF 模型性能最佳的原因可能是 RF 是一种用于回归应用的集成学习方法,其中预测是通过聚集多个决策树的输出来得出的,

这种方法对于处理非线性关系和具有复杂交互作用的水质数据集特别有效<sup>[20]</sup>。

表 2 模型性能评估指标参数  
Tab. 2 Parameters of Model Performance Evaluation

模型	$R^2$		RMSE/( $\text{mg} \cdot \text{L}^{-1}$ )		MAE/( $\text{mg} \cdot \text{L}^{-1}$ )	
	训练集	测试集	训练集	测试集	训练集	测试集
BPNN	0.754 0	0.689 6	9.898 9	8.444 9	6.070 5	7.508 8
PSO-BPNN	0.859 9	0.809 7	6.512 9	7.376 5	4.494 5	5.671 8
LSTM	0.802 3	0.794 5	7.648 2	7.740 2	5.609 7	6.029 7
RF	0.978 1	0.945 9	2.501 0	4.197 1	1.561 4	2.805 9

## 2.2 模型对比分析

4 种机器算法模型对 MF-CW 系统出水氨氮的预测效果如图 3 所示。图 3 中显示了出水氨氮浓度的预测数据和真实数据,4 种模型中 BPNN 预测效果最差,训练集  $R^2$  值为 0.754 0、测试集  $R^2$  值为 0.689 6;训练集 RMSE 值为 9.898 9  $\text{mg/L}$ 、测试集 RMSE 值为 8.444 9  $\text{mg/L}$ ;训练集 MAE 值为 6.070 5  $\text{mg/L}$ 、测试集 MAE 值为 7.508 8  $\text{mg/L}$ 。虽然该模型的拟合效果最差,但 BPNN 作为基于误差反向传播算法训练的多层前馈网络,已经被广泛应用于各种水质预测研究<sup>[21]</sup>。基于 BPNN 模型,PSO-BPNN 模型无论训练集还是测试集的拟合精度都大幅增长,其中 BPNN 训练集的  $R^2$  在经过粒子群算法优化后由 0.754 0 以下增长至 0.859 9 以上,增长了 14.05%;测试集同样有较大提升,相较于 BPNN 增

加了 17.42%,同时 RMSE 和 MAE 也有较大幅度的下降, RMSE 下降了 12.65%, MAE 下降了 24.46%。 PSO 作为一种基于随机值的智能进化算法,通过连续迭代找到最优解,并且 PSO 通过群体搜索机制,在全局空间内优化 BPNN 的初始权值和阈值,显著降低了 BPNN 陷入局部最优的风险,提升了预测精度<sup>[22]</sup>。 LSTM 模型相较于 BPNN 模型预测精度有了略微增加,测试集  $R^2$  值增长了 15.21%,但相较于 PSO-BPNN 模型, LSTM 模型预测精度就略差,无论在训练集还是测试集中的表现均不如 PSO-BPNN 模型。 LSTM 作为一种能解决梯度消失或爆炸问题的变体<sup>[9]</sup>,分别通过 3 种类型的门负责控制信息流,输入门调节新信息的流动,而忘记门确定应该从每个存储单元中丢弃哪些信息,最后,输出门确定应该将哪些信息发送到 LSTM 的输出<sup>[23]</sup>。正是因为

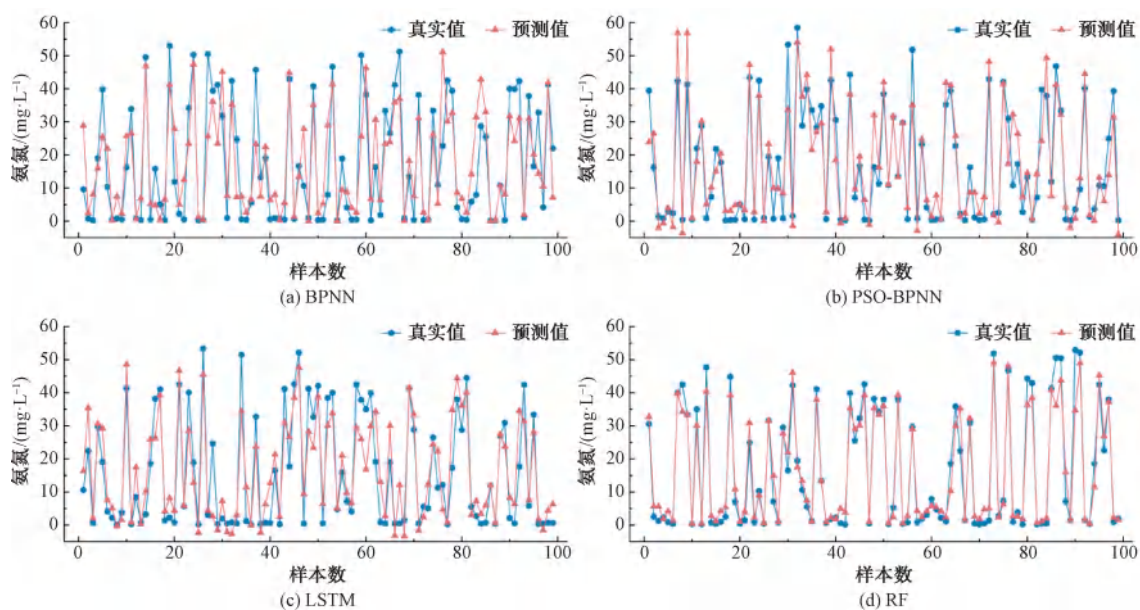


图 3 氨氮真实值与预测值拟合图

Fig. 3 Fitting Plots of Actual and Predicted Ammonia Nitrogen Values

这种特性使 LSTM 模型在捕获顺序数据中的长期依赖关系并准确预测长时间的数据方面优于其他方法,这是 LSTM 模型用于各种水质数据预测的关键原因<sup>[13,24]</sup>。导致 LSTM 模型预测效果不佳的原因可能是 LSTM 可以考虑过去结果对现在的影响,这在时间序列问题中发挥着重要作用,但在本文中将其应用于对水质数据的回归预测,造成收敛速度慢、最终收敛的损失值较高等问题,使得模型的预测精度降低。

结合图 3 的拟合结果与表 2 的指标参数来看,RF 模型无论是训练集还是测试集的  $R^2$ 、RMSE 和 MAE 都说明 RF 模型的拟合精确度优于传统的 BPNN 模型、PSO-BNN 模型和 LSTM 模型。RF 模型具有这么高的预测精度,可能因为与其他算法相比,RF 是一种数据驱动的非线性建模工具,可以直接处理多维数据,无需降维或特征选择,并且完全数据驱动<sup>[25]</sup>。4 种算法模型中 RF 模型中对氨氮浓度预测中训练集  $R^2$  为 0.978 1、测试集  $R^2$  为 0.945 9、相较于效果最差的 BPNN 模型,训练集  $R^2$  和测试集  $R^2$  分别提升了 29.72% 和 37.17%。相对于 RF 模型对  $R^2$  值的提升效果,其对 RMSE 和 MAE 值的降低作用更加突出。RF 模型的训练集 RMSE 为 2.501 0 mg/L、测试集 RMSE 为 4.197 1 mg/L,与 BPNN 模型对比分别降低了 74.73% 和 50.30%。RF 模型的训练集 MAE 为 1.561 4 mg/L、测试集 MAE 为 2.805 9 mg/L,与 BPNN 模型对比分别降低了 74.28% 和 62.63%。RF 模型测试集的  $R^2$  相较于 PSO-BPNN 和 LSTM,分别提高了 16.82% 和 19.06%,RMSE 降低了 43.10% 和 45.78%,MAE 降低了 50.53% 和 53.47%。RF 模型的  $R^2$  值最接近于 1,并且 RMSE 和 MAE 值均最小,表明该模型的预测值与水质真实数据很好地吻合。此外,由图 3(d)中测试集中真实值与预测值的拟合效果可知,预测值在波峰和波谷处的变化与真实值的变化趋势保持高度一致,更适合于用来建立 MF-CW 系统出水水质的预测模型。RF 模型作为基于集成学习的回归方法,通过构建多个决策树并集成它们的预测,RF 模型的最终预测作为所有树预测的平均值,虽然平均值本身对极端值比较敏感,但 RF 平均的是许多树的预测值,而不是原始数据值,通过多次平均可平滑预测结果。RF 并不试图去“修正”极端值,而是通过“集体决策”来削弱极端值的影响力,将数据中极

端值对模型的影响分散化和稀释化,因此其具有很强的抗噪性和鲁棒性<sup>[26]</sup>。研究结果表明:RF 模型在处理复杂的水质非线性数据关系方面非常有效,这使得它非常适合推进 MF-CW 系统实际应用研究进程,同时这项工作的结果进一步验证了集成学习方法在污水处理中的研究潜力。

### 3 结论

随着数据分析技术的持续成熟,推动了 ML 技术在环境科学领域的广泛应用。本文中通过 ML 算法模型对 MF-CW 系统出水氨氮预测,得出了 3 个主要研究结论:(1)通过对构建的 4 种预测模型性能评估指标的比较,RF 模型的预测精度及与模型拟合程度均优于其他 3 种模型,证明该模型的强鲁棒性和抗噪性能够为 MF-CW 系统预测出水提供较准确的预测结果;(2)PSO 算法可以改进传统的 BP 模型,PSO 通过群体搜索机制可以减少 BPNN 陷入局部最优等问题提高了模型的预测精度;(3)本文构建的 ML 模型,实现了良好的水质参数预测性能,通过 RF 算法模型可以大幅减少 MF-CW 系统在试验收集数据的过程中浪费大量资源投入,通过 ML 进行水质预测的方法,对推动 MF-CW 系统研究发展有着重要作用,同时也为其他水质参数预测提供了新思路。

### 参考文献

- [1] LIU T T, LI D, TIAN Y, et al. Enhancing nitrogen removal in constructed wetlands: The role of influent substrate concentrations in integrated vertical-flow systems[J]. *Environmental Science and Ecotechnology*, 2024, 21: 100411. DOI: 10.1016/j.ese.2024.100411.
- [2] LI C, FENG L K, LIAN J J, et al. Enhancement of organics and nutrient removal and microbial mechanism in vertical flow constructed wetland under a static magnetic field[J]. *Journal of Environmental Management*, 2023, 330: 117192. DOI: 10.1016/j.jenvman.2022.117192.
- [3] XIONG L C, MA R, YIN F J, et al. Simulation and optimisation of magnetic and experimental study of magnetic field coupling constructed wetland[J]. *Environmental Technology*, 2024, 45(24): 5083–5103.
- [4] WANG Q, LIU Y G, YANG S L, et al. Treatment performance and microbial community dynamics in vertical flow constructed wetlands integrated with gradient magnetic fields[J]. *Bioresource Technology*, 2025, 434: 132756. DOI: 10.1016/j.biortech.2025.132756.
- [5] LÜ J Q, YIN W X, XU J M, et al. Augmented machine learning



- for sewage quality assessment with limited data [ J ]. Environmental Science and Ecotechnology, 2025, 23: 100512. DOI:10.1016/j.ese.2024.100512.
- [ 6 ] DAWOOD T, ELWAKIL E, NOVOA H M, et al. Toward urban sustainability and clean potable water; Prediction of water quality via artificial neural networks[J]. Journal of Cleaner Production, 2021, 291: 125266. DOI:10.1016/j.jclepro.2020.125266.
- [ 7 ] XU H, LÜ B, CHEN J, et al. Research on a prediction model of water quality parameters in a marine ranch based on LSTM-BP [J]. Water, 2023, 15(15): 2760.
- [ 8 ] YAN J Z, XU Z B, YU Y C, et al. Application of a hybrid optimized BP network model to estimate water quality parameters of Beihai Lake in Beijing[J]. Applied Sciences, 2019, 9(9): 1863.
- [ 9 ] HUANG S, XIA J, WANG Y L, et al. Water quality prediction based on sparse dataset using enhanced machine learning[J]. Environmental Science and Ecotechnology, 2024, 20: 100402. DOI:10.1016/j.ese.2024.100402.
- [ 10 ] ZANONI M G, MAJONE B, BELLIN A. A catchment-scale model of river water quality by machine learning[J]. Science of the Total Environment, 2022, 838: 156377. DOI:10.1016/j.scitotenv.2022.156377.
- [ 11 ] UDDIN M G, RAHMAN A, TAGHIKHAH F S, et al. Data-driven evolution of water quality models: An in-depth investigation of innovative outlier detection approaches—A case study of Irish Water Quality Index (IEWQI) model[J]. Water Research, 2024, 255: 121499. DOI: 10.1016/j.watres.2024.121499.
- [ 12 ] BARZEGAR R, AALAMI M T, ADAMOWSKI J. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model[J]. Stochastic Environmental Research and Risk Assessment, 2020, 34(2): 415–433.
- [ 13 ] DODIG A, RICCI E, KVASCEV G, et al. A novel machine learning-based framework for the water quality parameters prediction using hybrid long short-term memory and locally weighted scatterplot smoothing methods [ J ]. Journal of Hydroinformatics, 2024, 26(5): 1059–1079.
- [ 14 ] GEORGESCU P L, MOLDOVANU S, ITICESCU C, et al. Assessing and forecasting water quality in the Danube River by using neural network approaches [ J ]. Science of the Total Environment, 2023, 879: 162998. DOI:10.1016/j.scitotenv.2023.162998.
- [ 15 ] CHEN L X, WU T H, WANG Z C, et al. A novel hybrid BPNN model based on adaptive evolutionary artificial bee colony algorithm for water quality index prediction [ J ]. Ecological Indicators, 2023, 146: 109882. DOI: 10.1016/j.ecolind.2023.109882.
- [ 16 ] LU Y B, LI T, DENG Y. Predication of water pollution peak concentrations by hybrid BP artificial neural network coupled with genetic algorithm[J]. Applied Artificial Intelligence, 2024, 38(1): 2341356.
- [ 17 ] YIN F J, MA R, LIU Y G, et al. Prediction study of pollutants in artificial wetlands enhanced by electromagnetic fields [ J ]. Sustainability, 2024, 16(23): 10327.
- [ 18 ] WANG W, LIANG R, QI Y, et al. Prediction model of spontaneous combustion risk of extraction borehole based on PSO-BPNN and its application[J]. Scientific Reports, 2024, 14: 5. DOI: 10.1038/s41598-023-45806-9.
- [ 19 ] YU Z Y, YANG K, LUO Y, et al. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network [ J ]. Journal of Hydrology, 2020, 582: 124488. DOI:10.1016/j.jhydrol.2019.124488.
- [ 20 ] NISHAT M H, KHAN M H R B, AHMED T, et al. Comparative analysis of machine learning models for predicting water quality index in Dhaka's rivers of Bangladesh [ J ]. Environmental Sciences Europe, 2025, 37(1): 31–31.
- [ 21 ] WANG P. Research on PSO-BPNN based model for forecasting typhoon rainfall [ J ]. Journal of Physics: Conference Series, 2020, 1650(3): 032026.
- [ 22 ] ZHANG X M, TANG Z S. Construction of computer model for enterprise green innovation by PSO-BPNN algorithm and its impact on economic performance [ J ]. PLOS One, 2022, 17(1): e0262963.
- [ 23 ] JAMSHIDZADEH Z, EHTERAM M, SHABANIAN H. Bidirectional Long Short-Term Memory ( BiLSTM )-Support Vector Machine: A new machine learning model for predicting water quality parameters [ J ]. Ain Shams Engineering Journal, 2024, 15(3): 102510.
- [ 24 ] YUAN X H, CHEN C, LEI X H, et al. Monthly runoff forecasting based on LSTM-ALO model [ J ]. Stochastic Environmental Research and Risk Assessment, 2018, 32(8): 2199–2212.
- [ 25 ] FANG X, LI X Y, ZHANG Y F, et al. Random forest-based understanding and predicting of the impacts of anthropogenic nutrient inputs on the water quality of a tropical lagoon [ J ]. Environmental Research Letters, 2021, 16(5): 055003.
- [ 26 ] SUN J J, GUAN X, SUN X J, et al. Water quality prediction and carbon reduction mechanisms in wastewater treatment in Northwest cities using random forest regression model [ J ]. Scientific Reports, 2024, 14(1): 31525.