J. Lake Sci.(湖泊科学),2026,38(6):000-000
DOI 10.18307/2026.0624
©2026 by Journal of Lake Sciences

# 基于机器学习的水生植物去除营养盐及 PFAS 效能研究*

关擎宇，李英豪，孟庆龙，张静怡，李颖**
（河海大学 环境学院，江苏南京，210000）

**摘 要**：为解决传统实验方法在多因素复杂体系下进行优化筛选时存在实验量大，浪费时间、人力和物力的局限性。本研究以营养盐及全氟和多氟烷基物质（Per- and Polyfluoroalkyl Substances，PFAS）污染水体为对象，集成已有文献记录，构建了包含环境因子、植物种植密度及污染物去除率的综合数据集，通过对比多目标随机森林（Multi-target Random Forest，MTRF）、多层感知机（Multi-layer Perceptron，MLP）、随机森林（Random Forest，RF）和极端梯度提升（Extreme Gradient Boosting，XGBoost）算法综合评估了各因素对水环境中 PFAS 去除率的影响，利用可解释机器学习 SHAP（Shapley Additive Explanations）框架对模型进行深度解释，筛选最优水生植物并预测其最佳处理工况。结果表明：MTRF 表现出良好的预测性能，其对营养盐和 PFAS 去除率的决定系数 $R^2$ 值普遍高于其他模型，能够准确预测不同工况下营养盐和 PFAS 的去除率。基于模型预测，苦草被筛选为净化营养盐及 PFAS 污染的最佳优势植物。通过工况优化预测，苦草去除营养盐的最佳种植密度为 19 株/m²，最佳修复时间为 69 天；去除 PFAS 的最佳种植密度为 59 株/m²，最佳修复时间为 34d。本研究建立的机器学习模型不仅能够为复杂水质条件下的植物修复效率提供精准预测，也为水体复合污染的低成本、系统性治理提供了理论依据与工程决策支持。
**关键词**：水生植物；全氟和多氟化合物（PFAS）；机器学习；SHAP；营养盐

## On the Efficiency of Removing Nutrients and Per- and Polyfluoroalkyl substances from Aquatic Plants Based on Machine Learning*

Guan Qingyu , Jiang Chenxue , Li Yinghao , Meng Qinglong , Zhang Jingyi & Li Ying**
（*College of Environment, Hohai University, Nanjing, 210098, P.R.China*）

**Abstract**： Global concerns over nutrient and per- and polyfluoroalkyl substances (PFAS) pollution are growing, leading to a demand for more efficient and low-cost remediation strategies. This study explored the remediation of nitrogen, phosphorus, and multiple PFAS components using aquatic plants as a sustainable alternative to conventional physical and chemical methods. To overcome the limitations of traditional experimental approaches, we compiled a comprehensive dataset by systematically retrieving literature from the CNKI, ScienceDirect, and Web of Science databases for the period 2015–2025. The final dataset included 128 independent experimental records covering pollutant remediation data under diverse experimental conditions, which supports the generalizability of our findings. We used 14-dimensional input features, including aquatic plant types (submerged, emergent, and floating plants), planting density, ambient temperature, and initial concentrations of nitrogen, phosphorus, and eight specific PFAS components (e.g., PFOS, PFOA, and PFBA). Strict quality control protocols were applied to reduce systematic biases inherent in multi-source literature data. Missing values of secondary environmental parameters were filled using RF-based imputation, which better retains the nonlinear structure of the original data than simple mean replacement. All numerical features were standardized with StandardScaler to remove dimensional differences. In addition, one-way analysis of variance (ANOVA) and effect size (η2) were used to quantitatively assess data heterogeneity. The results showed that planting density and temperature were the primary sources of statistical heterogeneity, explaining 41.67% and 41.47% of the total variation, respectively ($P$ < 0.001). Four algorithms, namely Multi-objective Random Forest (MTRF), Multilayer Perceptron (MLP), Random Forest (RF),

*2025-12-12 收稿;2026-02-10 收修改稿。
国家重点研发计划项目（2022YFC3204105）资助。
**通信作者： hj6688@hhu.edu.cn

and Extreme Gradient Boosting (XGBoost), were applied to comprehensively evaluate the effects of various factors on PFAS removal rates. $R^2$ and Root Mean Square Error (RMSE) were selected as evaluation indicators of model performance. The MTRF model yielded average $R^2$ values of 0.68 for nitrogen removal and 0.72 for PFNA prediction, demonstrating reliable predictive capacity. Its performance was significantly better than that of the MLP model, which produced negative $R^2$ values and failed to capture complex data patterns. Further mechanism analysis using the SHAP framework indicated that initial phosphorus concentration positively affected nitrogen absorption. Phosphorus can enhance plant metabolic capacity and thus promote nitrogen uptake. A threshold inhibition effect was detected for PFAS remediation. High initial pollutant concentrations and extreme temperatures reduced remediation efficiency by inducing reactive oxygen species (ROS) accumulation, which triggers oxidative stress and inhibits enzyme activity. Virtual screening identified *Vallisneria natans* as the dominant remediation species. Its superior performance is attributed to strong root-to-stem transport capacity and high PFAS bioaccumulation factor. Condition optimization determined the optimal parameters for *Vallisneria natans* remediation. The optimal planting density and remediation time for nutrient removal are 19 plants/m² and 69 days, while those for PFAS removal are 59 plants/m² and 34 days. The machine learning model established in this study can accurately predict aquatic plant remediation efficiency under complex water environmental conditions. It also provides a theoretical basis and engineering decision support for the low-cost and systematic treatment of combined water pollution.

Keywords：Aquatic plants; Per- and Polyfluorinated Substances (PFAS); Machine learning; SHAP; Nutrients

全氟和多氟烷基物质（Per- and Polyfluoroalkyl Substances，PFAS）因其独特的碳-氟键，具备优异的化学稳定性和疏水疏油特性，被广泛应用于工业和消费品中[1,2]。然而，PFAS 极强的持久性、生物累积性和毒性使其成为一种全球性污染物，对生态系统和人类健康构成严重威胁[3–5]。与此同时，由氮、磷等营养盐过量输入导致的水体富营养化问题依然严峻[6,7]。当富营养化与 PFAS 这两种污染同时存在于同一水体时，其治理难度远超单一污染。传统的物理化学处理技术，如活性炭吸附或反渗透[8,9]，虽能有效去除 PFAS，但通常成本高、能耗大，且对营养盐去除能力有限。因此，亟需一种环境友好、成本效益高的替代或辅助技术。

利用水生植物进行生态修复，是一种具有巨大潜力的“绿色”技术[10,11]。它利用植物及其根际微生物的生命活动来吸收、降解或固定污染物。水生植物不仅能有效去除水体中的氮、磷[12–14]，还能通过吸收和富集作用去除部分 PFAS[11]，为协同净化富营养化和 PFAS 复合污染水体提供了可能。近年来，学术界对水生植物修复 PFAS 污染水体进行了初步探索，证实了多种水生植物具有吸收和富集 PFAS 的能力[15]。然而，现有研究存在显著局限性，比如体系单一，多数研究集中于单一污染物或单一植物种类[16]，缺乏对多种 PFAS 混合物和不同植物物种去除能力的系统性比较。另外，很多工作侧重于机理探讨[17,18]，缺乏复杂环境中优化多因素交互作用以达到最佳修复效果的研究。传统实验方法难以全面评估植物种类、种植密度、环境条件等多因素及其非线性交互作用的影响，需要海量的实验组合，耗时耗力。

为克服上述瓶颈，以机器学习为代表的数据驱动方法为环境科学领域提供了新途径[19]。机器学习模型擅长处理多因素的非线性关系，通过训练现有数据构建预测模型，实现对不同修复方案效果的“虚拟筛选”和参数优化，从而大大减少实验成本和时间[20–22]。温诗林[23]用机器学习的方法筛选出强化厌氧氨氧化工艺的最优生物炭种类，并通过特征分析得出原料类型和热解温度是影响生物炭分类的重要变量。李珊珊等人[24]利用机器学习的方法筛选出一种提升乙酰丙酸转化率和 γ-戊内酯产率的催化剂，不仅减少实验试错成本还提高了实验效率。此外，结合 SHAP（Shapley Additive Explanations）等可解释性工具[25]，不仅能预测结果，还能量化每个特征对预测结果的具体贡献，为数据驱动解析植物修复机理提供解释路径。

本研究旨在构建基于机器学习的预测与优化框架，以解析水生植物对营养盐与 PFAS 污染的修复规律。通过整合文献数据，利用 $R^2$ 与 RMSE 等指标筛选最优模型，并引入 SHAP 可解释框架量化各环境变量及污染物物性对去除效率的贡献。研究克服了传统实验法无法处理多因素复杂协同关系的问题，量化了环境因子与植物生理特性对修复效能的非线性交互作用；通过模型训练、虚拟筛选、参数优化的流程，实现了从理论认知向工程实践的跨越，为复杂环境背景下水体复合污染的高效治理提供科学指导与工程参考。

## 1 材料与方法

### 1.1 数据来源与数据集构建

本研究通过系统检索知网、Web of Science 和 Science Direct 等学术数据库，检索关键词包括“水生植物(Aquatic plants)”、“植物修复(Phytoremediation)”、“全氟和多氟烷基物质(PFAS)”、“营养盐(Nutrients)”等。经过筛选及数据提取，最终集成了来自 2015 年至 2025 年间发表的 11 篇论文中的 128 条实验记录。数据集包含水生植物（如金鱼藻、鸢尾、苦草等）对 PFAS（如 PFOS、PFOA、PFBA 等）及营养盐（总氮、总磷）的去除效果数据。数据集包含输入特征和输出目标两部分，其中输入特征涵盖了影响植物修复效率的关键生物及非生物因素，共 14 个维度，包括植物种类、种植密度、种植温度、种植时间、初始污染物浓度（总氮、总磷、8 种 PFAS）等。输出目标包括 8 种 PFAS 去除率、总氮去除率和总磷去除率。

### 1.2 数据预处理

针对文献中单位不一致的问题，本研究建立了统一的量纲转换标准：所有营养盐单位统一换算成 mg/L，PFAS 浓度统一换算成 μg/L，种植密度统一换算成株/$m^2$，去除时间统一换算成天。只记录关键因子大于 3 个的数据，对于部分参数缺失问题，本研究采用了基于 Scikit-learn 的随机森林填补法。剔除异常值（比如去除率>100%或者<0）。为消除不同环境特征间量纲不一致导致的预测偏差，调用 Scikit-learn 库中的 StandardScaler 模块对所有数值型变量进行了标准化处理。所有操作均在 Python 环境下借助 pandas 和 scikit-learn 库完成。

### 1.3 机器学习模型

本研究选用多种机器学习模型进行比较，包括：(1)多目标随机森林（MTRF）：一种能够同时预测多个连续变量的模型，能捕捉不同目标变量之间的内在联系，提高预测精度[26]。(2)多层感知机（MLP）：一种经典的神经网络模型，通过多层神经元连接来处理复杂的非线性关系，常用于分类和回归任务[27]。(3)随机森林（RF）：这是一种集成学习算法，通过构建多棵决策树并取其平均结果来提高预测精度和稳定性，但一次只能处理一个目标变量[28]。(4)XGBoost：作为一种高效且强大的集成学习算法，它通过迭代地训练弱模型（通常是决策树）并进行加权组合[25]。

模型训练采用 K-折交叉验证（K-Fold），将数据集随机划分成 k 个互斥子集，依次取其中 k-1 个子集进行模型训练，余下 1 个子集作为验证集[29]。使用留一法进行敏感性分析，对比单一实验记录与子集模型训练之间的差异。使用决定系数（$R^2$）和均方根误差（RMSE）作为模型性能评估指标。

### 1.4 特征重要性与模型解释（SHAP）

为解决传统特征重要性方法的局限性，本研究采用基于合作博弈论的 SHAP 方法。SHAP 能够为每一个预测公平地分配每个特征的贡献值，实现局部和全局的可解释性[30]。通过 SHAP Summary Plot 可视化工具，可以直观地分析每个特征对模型预测的平均影响大小及其影响方向。

### 1.5 最佳工况筛选流程

优化修复方案包含筛选最佳植物和优化种植密度与时间两个部分。通过固定其他环境因子为平均水平，预测不同植物种类的综合净化效果，从而筛选出最佳植物。针对筛选出的最佳植物，创建种植密度和种植时间的二维网格，并利用训练好的模型对网格上的每个点进行预测。通过热力图直观展示综合净化效果得分，找到得分最高的点，即为理论最佳工况。

### 1.6 数据异质性分析

为识别并量化多源文献数据中的异质性来源，本研究采用单因素方差分析（ANOVA）检验碳链长度、植物类型、环境因子对污染物去除率的显著性。引入效应量指标 $\eta^2$（Eta-squared）[31]量化各特征对数据异质性的贡献比例，计算公式如下：

$$\eta^2 = \frac{SS_{effect}}{SS_{total}} \quad (1)$$

式中，$SS_{effect}$ 代表组间平方和，$SS_{total}$ 代表总平方和。该指标用于衡量观测到的数据波动中有多少比例可由分类特征所解释，为评估多源数据的稳定性提供定量依据。

## 2 结果与讨论

### 2.1 机器学习模型性能评估

模型的性能评估采用两个关键指标：$R^2$ 和 RMSE。$R^2$ 用于衡量模型对数据的拟合优度，其值越接近 1 表明模型对目标变量的解释能力越强。RMSE 则量化了预测值与真实值之间的平均偏差，其值越低表示模型的预测精度越高。本研究将 MTRF、MLT、RF、和 XGBoost 算法四种机器学习模型应用于处理好的相同数据集进行对比评估，对每个模型分别计算并报告目标污染物去除率的 $R^2$ 和 RMSE。

表 1 机器学习模型对氮磷去除率模拟效果比较

Tab.1 Comparison of simulation effects of four machine learning models on nitrogen and phosphorus removal rates

| 模型 | 氮去除率 | | 磷去除率 | |
|---|---|---|---|---|
| | $R^2$ | *RMSE* | $R^2$ | *RMSE* |
| MTRF | 0.68(0.12) | 7.79(2.38) | 0.55(0.16) | 9.10(1.33) |
| MLP | -0.17(0.32) | 15.01(3.707) | -0.27(0.43) | 15.49(3.00) |
| RF | 0.65(0.12) | 8.23(2.48) | 0.56(0.18) | 8.92(1.23) |
| XGBoost | 0.62(0.16) | 8.37(2.18) | 0.52(0.13) | 9.66(2.48) |

注：括号内是标准误差

表 2 机器学习模型对不同 PFAS 去除率模拟效果比较

Tab.2 Comparison of simulation effects of four machine learning models on PFAS removal rates

| 模型 | PFOS 去除率 | | PFOA 去除率 | | PFBA 去除率 | | PFPeA 去除率 | | PFNA 去除率 | | PFDA 去除率 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2$ | *RMSE* | $R^2$ | *RMSE* | $R^2$ | *RMSE* | $R^2$ | *RMSE* | $R^2$ | *RMSE* | $R^2$ | *RMSE* |
| MTRF | 0.65 (0.27) | 9.23 (1.89) | 0.22 (0.45) | 7.52 (3.38) | 0.62 (0.19) | 5.43 (2.17) | 0.54 (0.22) | 6.48 (2.13) | 0.72 (0.20) | 5.80 (4.15) | 0.56 (0.56) | 8.33 (3.83) |
| MLP | 0.57 (0.29) | 10.38 (2.10) | 0.19 (0.41) | 8.33 (4.70) | 0.56 (0.23) | 5.73 (2.42) | 0.22 (0.64) | 8.55 (4.44) | 0.61 (0.19) | 6.70 (3.60) | 0.53 (0.37) | 10.05 (2.97) |
| RF | 0.53 (0.46) | 9.98 (3.06) | 0.11 (0.56) | 7.64 (3.71) | 0.66 (0.17) | 5.06 (1.72) | 0.47 (0.18) | 7.20 (1.57) | 0.68 (0.22) | 6.08 (4.22) | 0.45 (0.70) | 9.10 (4.48) |
| XGBoost | 0.58 (0.19) | 11.00 (1.91) | 0.02 (0.77) | 7.52 (2.54) | 0.64 (0.24) | 5.27 (2.32) | 0.53 (0.22) | 6.39 (1.57) | 0.64 (0.19) | 6.40 (3.28) | 0.49 (0.63) | 9.16 (3.82) |

注：括号内为标准误差

为定量评估模型的可靠性，本研究参考了环境建模领域的权威评价准则。Moriasi 等[32]认为 $R^2$>0.5 的模型预测效果处于满意水平的参照范围，Henseler 等[33]将 $R^2$>0.67 视为模型有实质性的解释力。如表 1 和表 2 所示，MTRF 模型对氮去除率预测的平均 $R^2$ 为 0.68，具有实质解释力；对磷去除率预测的平均 $R^2$ 为 0.55，处于满意水平。相比之下，MLP 模型的 $R^2$ 仅为-0.17 和-0.27，远低于 0.5，且其 RMSE(15.01~15.49)较 MTRF(7.79~9.10)高出 70%~93%，表明其无法捕获复杂的数据规律。在 6 种 PFAS 的预测中，MTRF 在 5 种组分上的平均 $R^2$ 均高于 0.50，其中对 PFNA 的预测效果有实质解释力($R^2$=0.72)。定量对比显示，MTRF 对氮去除率的 $R^2$(0.68)较 XGBoost(0.62)高出近 10%，且其 RMSE(7.79)较表现最差的 MLP(15.01)降低了约 48.1%。在 PFOS 预测上，MTRF 的平均 $R^2$(0.65)比 RF(0.53)高出约 22.6%。

综上所述，MTRF 模型的 $R^2$ 值较高且 RMSE 波动较小，展现出较强的集成学习和非线性建模的能力，能够有效地捕捉水体净化过程中复杂的环境因子与污染物去除率之间的非线性交互关系。因此，本研究选用 MTRF 模型进行后续的虚拟实验和工况优化分析。

### 2.2 污染物去除的异质性分析

为了深入探讨多源文献数据集的统计异质性来源，采用 ANOVA 结合 $\eta^2$ 统计量，定量评估了各关键

实验变量对污染物去除率异质性的贡献比例。PFAS 的碳链长度对异质性的贡献率为 6.43%。相比于短链 PFAS，长链组分在植物体内的生物累积潜能更高，这种由结构差异引起的去除率差异是多组分模拟中不可忽视的异质性来源。挺水植物与沉水植物之间的差异解释了 1.96%的异质性。尽管该比例相对较低，但仍证实了不同植物类型在根系构造与转运机制上的差异。

表 3 不同实验变量对污染物去除率的 P 值及 $\eta^2$

Tab.2 Removing efficiency (*P* value and $\eta^2$) of pollutants under different experimental parameters

| 因素 | *P* 值 | $\eta^2$ |
|---|---|---|
| 链长 | <0.001 | 6.43% |
| 去除时间 | <0.001 | 32.6% |
| 种植密度 | <0.001 | 41.67% |
| 温度 | <0.001 | 41.47% |
| 植物类型 | <0.05 | 1.96% |

上述定量结果表明，数据集的异质性主要源于环境条件的波动，而非单一的实验误差。本研究通过在 MTRF 模型中显式引入这些高贡献度特征（$\eta^2$>40%），有效地将噪声转化为可学习的预测信号，从而降低系统偏差，确保不同研究背景下预测结论的客观性与稳健性。

### 2.3 基于 SHAP 值的关键影响因素分析

MTRF 虽然预测性能优越，但其内部决策过程如同“黑箱”[30]，难以直观理解。为了增强模型的可解释性并揭示污染物去除的内在机制，本研究引入了基于合作博弈论的 SHAP 算法。SHAP 值将每个特征对单个预测结果的贡献进行量化，确保了局部解释的准确性。通过对测试集所有样本的 SHAP 值进行汇总分析，进而揭示影响污染物去除效果的关键影响因素。

对营养盐的影响，如图 1（A）和（B）所示，影响水生植物去除氮磷的重要性排序依次为：去除时间、初始总磷浓度、初始总氮浓度、种植密度和植物苦草。图 1（E）和（F）进一步揭示了环境变量与氮磷去除率之间的非线性关系：去除营养盐初期，曲线斜率陡峭，净化效率高，与根系表面的活性位点吸附及根际微生物膜的定植代谢有关。随着种植时间的延长，SHAP 值增速放缓，表现出边际效应递减现象，可能因为植物生物量趋于平稳，植物的相对生长速率放缓导致对营养盐的吸收效率减弱。图 1（C）和（D）的 SHAP 散点图显示，较高的种植时间（红色点）与高的正 SHAP 值（向右）相关联。这表明在一定范围内，延长种植时间是提高总氮去除率的最直接有效手段，这是因为植物的生长发育需要吸收氮磷，随着植物的生长，氮磷等营养盐转化成化合物形态储存在植物体内。初始氮磷浓度与氮磷去除率也呈正比关系，即初始氮磷浓度越高去除率越高，低浓度下植物吸收速率受扩散限制，而高浓度营养盐则驱动植物主动运输的潜力。氮磷浓度对彼此去除具有协同效应[34]，初始磷浓度的增加不仅会提升磷的去除效率，还会通过改善植物代谢能力促进对氮的吸收。伍钦江[35]在基于人工浮岛对湖泊营养盐去除的实验中也发现去除时间和去除效率有明显的正相关关系。

对 PFAS 的影响，如图 2 所示，影响 PFAS 去除率最重要的因素依次为：种植密度、处理时间、初始 PFAS 浓度和种植温度。图 3 的 SHAP 散点图显示，处理时间和种植密度与 PFAS 去除率呈正比关系。初始 PFAS 浓度和种植温度与去除率则存在反比关系，PFAS 去除受疏水性影响显著，长链 PFAS 更易被根系截留，而短链 PFAS 则表现出更强的向地上部转运的趋势[36]。图 4 的 SHAP 依赖图进一步揭示了环境变量与 PFAS 去除效能之间的非线性关系。在处理初期，PFBA、PFHxA、PFNA 去除的 SHAP 值随时间快速上升，反映该阶段植物根系的物理吸附与主动摄取为去除 PFAS 的主要过程。随着时间延长，斜率逐渐平缓，表现出边际效益递减，可能因为植物根系表面有效吸附位点趋于饱和。SHAP 值的垂直离散度揭示了种植密度与去除时间之间的交互效应。高种植密度工况下，时间增加对去除率的贡献度显著高于低密度组。这表明高生物量不仅提供更多吸附位点强化净化能力，还可能通过形成的复杂根际微环境增强对 PFAS 的截留效率。高种植密度可以加快去除水中的 PFAS。过高的温度会诱发植物释放活性氧，抑制植物生长和酶活性。肖君[37]在研究 PFOA 在人工湿地系统中的去除实验中发现，高 PFOA 浓度（10mg/L）会明显抑

制人工湿地对总氮总磷的去除效率，植物体内与去除营养盐有关的酶活性明显收到抑制，印证了 PFAS 对植物的毒害作用。

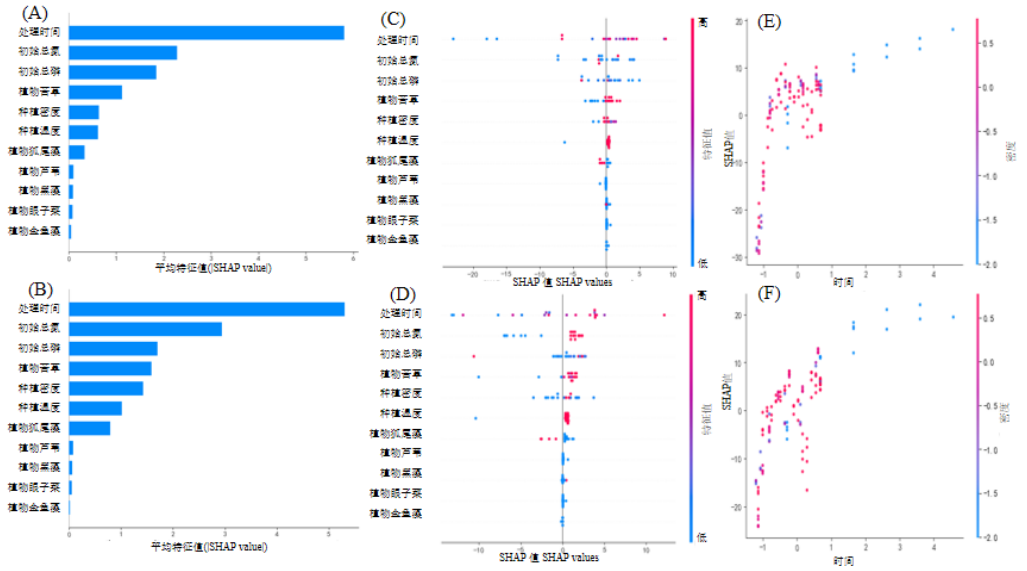


图 1 水生植物去除氮磷特征重要性排序、蜂群图及依赖图

(A)氮和(B)磷去除模型特征重要性排序图; (C)氮和(D)磷去除模型蜂群图; (E)氮和(F)磷去除依赖图

Fig.1 Feature Importance Ranking and SHAP Summary Plot of nitrogen and phosphorus removal by aquatic plants(A) Feature Importance Ranking of nitrogen removal model and(B) phosphorus removal model; (C) SHAP Summary Plot of nitrogen removal model and (D) phosphorus removal model; (E)SHAP Dependence Plot of nitrogen removal and(F)phosphorus removal

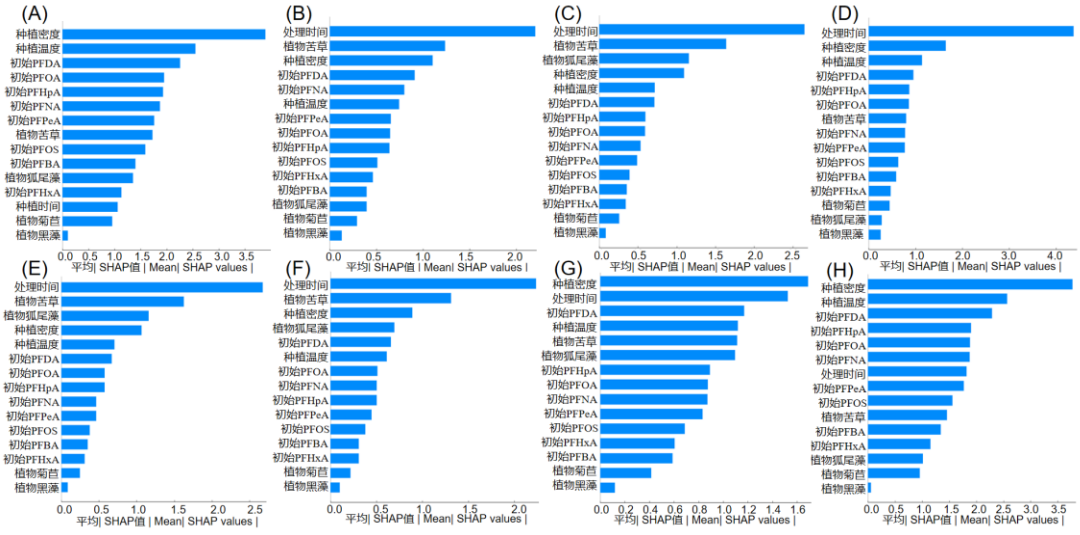


图 2 水生植物去除 PFAS 特征值重要性排序

(A) PFOS; (B)PFOA; (C)PFBA; (D)PFPeA; (E)PFHxA; (F)PFHpA; (G)PFNA; (H)PFDA

Figure 2 Feature Importance Ranking for PFAS removal by aquatic plants(A) PFOS ; (B) PFOA ; (C) PFBA ; (D) PFPeA ; (E) PFHxA; (F) PFHpA ; (G) PFNA ; (H) PFDA

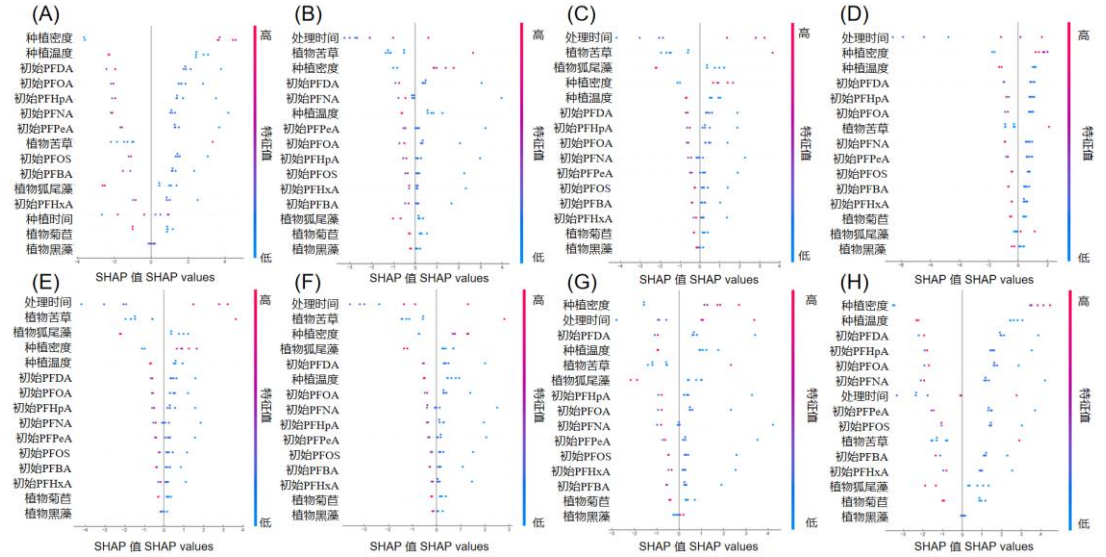


图 3 水生植物去除 PFAS 蜂群图

(A) PFOS; (B)PFOA; (C)PFBA; (D)PFPeA; (E)PFHxA; (F)PFHpA; (G)PFNA; (H)PFDA

Figure 3 SHAP Summary Plot for PFAS removal by aquatic plants(A) PFOS; (B) PFOA; (C) PFBA; (D) PFPeA; (E) PFHxA; (F) PFHpA; (G) PFNA; (H) PFDA

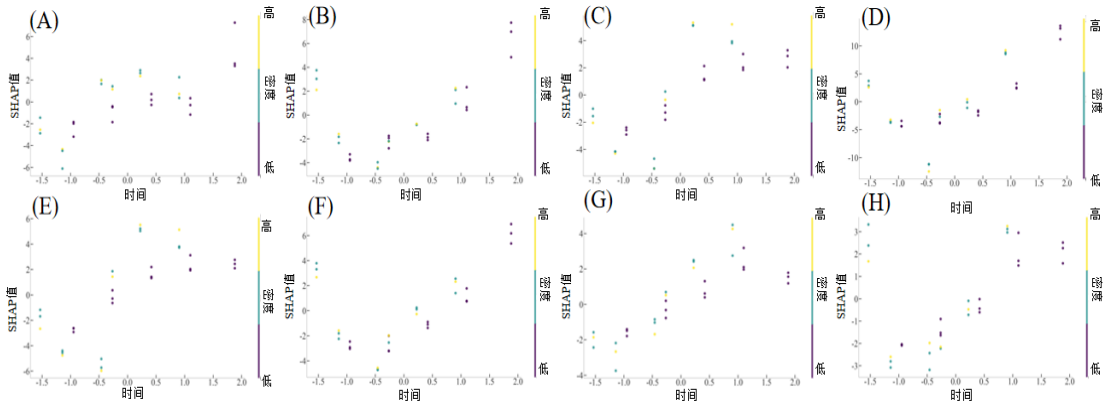


图 4 水生植物去除 PFAS 依赖图

(A)PFOS; (B)PFOA; (C)PFBA; (D)PFPeA; (E)PFHxA; (F)PFHpA; (G)PFNA; (H)PFDA

Figure 4 SHAP Dependence Plot for PFAS removal by aquatic plants(A) PFOS; (B) PFOA; (C) PFBA; (D) PFPeA; (E) PFHxA; (F) PFHpA; (G) PFNA; (H) PFDA

## 2.4 最佳水生植物筛选

为了筛选出具有最高净化潜能的沉水植物，本研究设计了一个基于机器学习模型的虚拟实验。将除植物种类外的所有数值型特征固定在数据集的平均水平，将每一种植物种类分别代入固定工况，输入至训练好的模型中，获取其预测的去除率结果。为了进行统一排序，采用综合得分进行量化，为真实反映 PFAS 毒性对生态风险健康的影响贡献，参考加拿大卫生部饮用水筛查值，根据 PFAS 最大可接受浓度（MACs）对不同种类 PFAS 进行加权，权重公式如下：

$$W_i = \frac{1}{MAC_{si}} \qquad (2)$$

其中 $W_i$ 为赋予第 i 种 PFAS 的毒性权重，$MAC_{si}$ 为第 i 种 PFAS 最大可接受浓度，并对每种植物的污染物去除率预测进行标准化处理，并将其求和作为最终的综合净化得分。得分越高，表明该植物的综合净化效果越优。

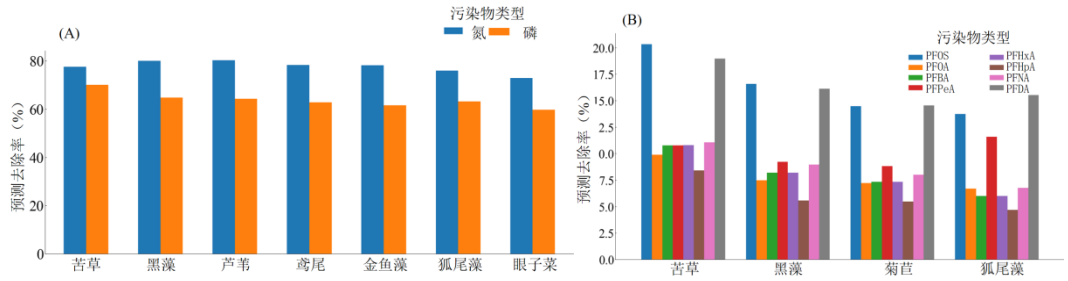


图 4 不同水生植物在固定环境条件下对氮磷及 PFAS 的预测净化效果

(A)氮磷去除率预测图; (B)PFAS 去除率预测图

Fig.4 Prediction of purification effect of different aquatic plants on nitrogen and phosphorus under fixed environmental conditions(A) Prediction of nitrogen and phosphorus removal rate ; (B) Prediction of PFAS removal rate

如图 4 所示，在多植物筛选中，苦草表现出最优的综合净化潜力。相较于金鱼藻等针状叶沉水植物，苦草的带状叶在动态水体中为生物膜提供良好的附着位点；区别于芦苇等挺水植物侧重于营养存储的木质化匍匐茎，苦草的匍匐茎具有更高的分裂能力[38]。通过克隆生长，苦草能产生具有高代谢能力的新生组织，这不仅加速营养盐的摄取，还为 PFAS 的吸收更新未饱和吸附位点。此外，苦草能利用其根际泌氧机制营造了微氧环境，从而强化了附着微生物对营养盐及有机污染物的降解效能[39]。综合得分排名也表明苦草在这些水生植物中最具潜力。洪瑜等[40]对比了 12 种水生植物对氮磷的去除效果，发现苦草对农田退水氮磷去除效果较好。因此，苦草被确定为综合净化效果最佳的优势植物。

## 2.5 最佳工况优化

确定最佳植物苦草后，对 PFAS 毒性加权且将除密度和时间之外的所有环境因子继续固定在典型工况的平均值上。在模型训练数据的范围内，构建一个二维网格，覆盖了种植密度和种植时间所有可能的组合。采用综合去除效果得分作为目标函数，寻找得分最高的网格点。

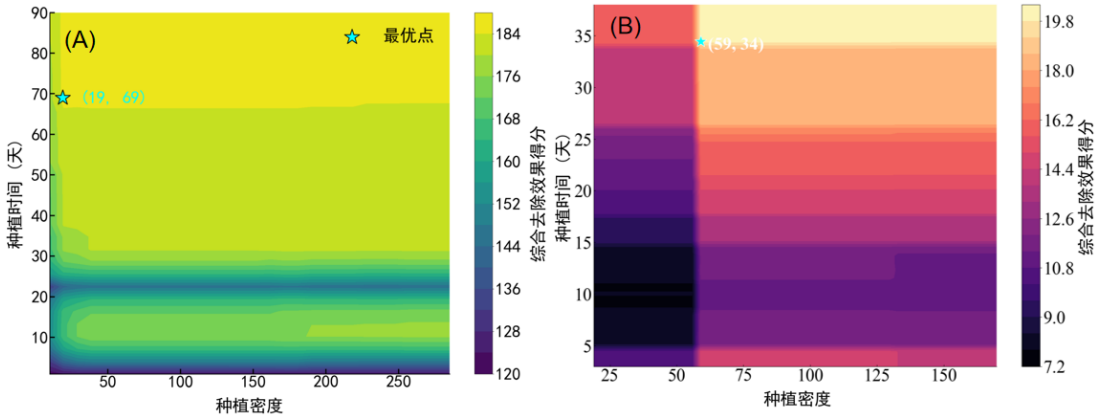


图 5 苦草去除氮磷及 PFAS 的综合净化效果热力图(A)氮 磷; (B)PFAS

Fig.5 Thermogram of comprehensive purification effect of *Vallisneria natans* on nitrogen and phosphorus and PFAS removal(A) nitrogen and phosphorus; (B) PFAS

图 5 展示了苦草在不同种植密度和种植时间组合下的综合去除效果热力图。热力图显示：对营养盐去除：最佳工况为种植密度 19 株/m²，处理时间 69 天。对 PFAS 去除：最佳工况为种植密度 59 株/m²，处理时间 34 天。

本研究预测的最佳种植密度与处理时间主要基于中宇宙的实验数据，旨在反映植物在生长阶段对污染物去除的最大潜力。在实际人工湿地工程中，系统通常呈现长期连续运行态势。因此，本模型输出的最佳去除时间可被解读为植物达到最佳净化效果的建议收割周期。每 34 天，针对 PFAS 重污染区域进行留茬收割；而对于高浓度营养盐污染区，每 69 天进行留茬收割；对于 PFAS 与营养盐负荷都高的水域可采取 34 天的收割周期，这个周期与 Verhofstad 等人[41]的建议采集频率接近，能够保证污染物去除效率。为确保系统的长期稳定的净化效能性，建议采取空间轮割的策略，留茬高度建议为 20cm。适当频率的收割能够触发植物补偿性生长机制，新生组织拥有更高的代谢活性同时为 PFAS 的吸收提供更多未饱和的吸附位点。通过及时移除已达到吸附饱和的组织，能够有效避免凋落物分解导致的污染物二次污染。最优初始种植密度的确定，有助于在湿地构建初期快速建立优势植物种群，缩短系统恢复周期。

## 3 结论

本研究利用机器学习方法对水生植物修复污染水体进行预测和优化。MTRF 在处理氮、磷及多种 PFAS 去除率的拟合中表现出最优的预测性能和稳定性，其 $R^2$ 显著高于其他模型。利用 SHAP 可解释性框架量化各因子对 MTRF 预测结果的贡献程度，揭示了去除时间、初始污染物浓度及种植密度是驱动净化效率的核心因素，其中，修复时间和种植密度与污染物的去除率均呈显著正相关，验证了生物量积累在净化过程中的关键作用。通过污染物综合去除得分筛选最优水生植物，苦草在多种水生植物中表现出最佳的净化潜力，被识别为修复营养盐与 PFAS 污染的优势物种。通过工况优化显示，净化营养盐的最佳种植密度为 19 株/$m^2$，对应修复时间为 69 天；而净化 PFAS 的最佳种植密度为 59 株/$m^2$，对应修复时间为 34 天。研究为水生植物修复工程的规模化应用提供精准的参数指导，对降低环境修复的试错成本具有重要的实际价值。

## 4 参考文献


[1] Jin XH, Yan N. Advances in researches on migration and transformation behavior of per- and polyfluoroalkyl substances precursors in the environment. Hydrogeology & Engineering Geology, 2024, 51(2): 35-49. DOI:10.16030/j.cnki.issn.1000-3665.202311005. [金宵卉, 阎妮. 全氟及多氟烷基化合物前体物质在环境中迁移与转化行为研究进展. 水文地质工程地质, 2024, 51(2): 35-49. ]

[2] Shi YL, Zhang BN, Zheng Z, et al. Research Progress on Existence and Behavior of Per- and Polyfluoroalkyl Substances in Atmospher. Reasearch of Environmental Science, 2022, 35(9): 2037-2046. DOI:10.13198/j.issn.1001-6929.2022.06.14. [史亚利, 张博钠, 郑哲, 等. 全氟和多氟烷基类物质在大气环境中的存在和行为研究进展. 环境科学研究, 2022, 35(9): 2037-2046.]

[3] Wang HD, Jiang H, Cai JQ, et al. Systematic transcriptomic analysis reveals differential toxicity of three representative per- and polyfluoroalkyl substances. China Environmental Science, 2025. DOI：10.19674/j.cnki.issn1000-6923.20250729.005. [王慧东, 姜浩, 蔡佳琪, 等. 系统转录组分析揭示三种典型全氟和多氟烷基物质毒性差异. 中国环境科学, 2025. ]

[4] Yin LT, Chen XY, Qin ZF. Research Progress of Hepatotoxicity and Lipid Metabolism Disrupting Effects of Typical Per-and Polyfluoroalkyl Substances. Asian Journal of Ecotoxicology, 2025, 20(3): 80-92. DOI: 10.7524/AJE.1673-5897.20241114001. [尹乐陶, 陈宣樾, 秦占芬. 典型 PFAS 的肝毒性和脂代谢干扰效应研究进展. 生态毒理学报, 2025, 20(3): 80-92.]

[5] 陈欣. 金属/类金属及 PFAS 职业暴露与类固醇激素的关联研究[学位论文]. 广东：广东工业大学, 2025.

[6] Shen CY, Zuo Q, Cheng GM, et al. Spatiotemporal Variation in Water Quality and Eutrophication Status of Dianchi Lake. Journal of Changjiang River SciencesResearch Institute, 2025, 42(9): 58-66. DOI: 10.11988/ckyyb.20240741. [沈春颖, 左黔, 程乖梅, 等. 滇池水质时空变化及富营养状况分析. 长江科学院院报, 2025, 42(9): 58-66.]

[7] Su H, Shi J, Sun J, et al. Study on Spatiotemporal Variation Characteristics of Occasional Blue-green Algae Blooms Accumulation Pollution in Eutrophic Water Bodies. Environmental Science and Management, 2025, 50(9): 158-162. DOI: 10.3969/j.issn.1673-1212.2025.09.033. [苏华, 石军, 孙璟, 等. 富营养化水体偶发性蓝藻水华堆积污染时空变化特征研究. 环境科学与管理, 2025, 50(9): 158-162.]

[8] Wang T, Bao YX, Zhong JK, et al. Research progress on removal of PFAS from water by activated carbon and resin adsorption. Modern Chemical Industry, 2024, 44(S1): 21-25, 29. DOI:10.16606/j.cnki.issn0253-4320.2024.S1.005. [王菟, 包一翔, 钟金魁, 等. 活性炭和树脂吸附去除水中 PFAS 的研究进展. 现代化工, 2024, 44(S1): 21-25, 29.]

[9] Zhuang XJ, Feng G. Progress of membrane separation technology dealing with the removal of per-and polyfluoroalkyl substances in water. Water and Wastewater Engineering, 2023, 59(S1): 920-925, 935. DOI:10.13789/j.cnki.wwe1964.2023.03.29.0006. [庄晓杰, 冯光. 膜分离技术在水体 PFAS 去除的研究进展. 给水排水, 2023, 59(S1): 920-925, 935.]
[10] Liu B. Application Research and Prospect of Aquatic Plants in Water Pollution Remediation. Horticulture and Seed, 2024, 44(8): 108-110. DOI:10.16530/j.cnki.cn21-1574/s.2024.08.041. [刘兵. 水生植物在水污染修复中的应用研究及展望. 园艺与种苗, 2024, 44(8): 108-110. ]
[11] Lei JM, Li D, Li J, et al. Bioaccumulation of short-chain per-and polyfluoroalkyl substances in submerged macrophytes. China Environmental Science, 2025, 45(7): 3905-3915. DOI:10.19674/j.cnki.issn1000-6923.20250121.004. [雷嘉敏, 李冬, 李静, 等. 沉水植物对短链全(多)氟烷基化合物的生物富集. 中国环境科学, 2025, 45(7): 3905-3915.]
[12] Yuan NX, Wang AN, Xu H, et al. Research on the purification effects of seventeen aquatic plants and their combinations on eutrophic water. Environmental Pollution and Control, 2025, 47(7): 72-78. DOI:10.15985/j.cnki.1001-3865.202405210. [原宁欣, 汪安宁, 徐红, 等. 17 种水生植物及其组合对富营养化水体的净化效果研究. 环境污染与防治, 2025, 47(7): 72-78.]
[13] 苗婧怡, 李倩倩, 苏雷, 等. 植物浮岛技术在富营养化水体治理中的应用及研究进展. 清洗世界, 2025, 41(10):151-153. DOI: 10.3969/j.issn.1671-8909.2025.10.051
[14] Li XZ, Li W, Huo BT, et al. Study on the Combined Use of Composite Algicide and Submerged Ppants to Conctrol Algal Blooms and Remediate Eutrophic Water Bodies. Environmental Engineering, 2025. [黎啸振, 李威, 霍斌堂, 等. 复合除藻剂联合沉水植物抑制水华和修复富营养化水体研究. 环境工程, 2025.]
[15] He Q, Yan Z, Zhi Y, et al. Research progress on bioaccumulation of per-and polyfluoroalkyl compounds by plants. China Environmental Science, 2022, 42(11): 5395-5407. DOI:10.19674/j.cnki.issn1000-6923.2022.0207. [何强, 颜正, 智悦, 等. 植物对 PFAS 的富集机制研究进展. 中国环境科学, 2022, 42(11): 5395-5407.]
[16] Jiang T, Pervez MN, Ilango AK, et al. Stabilizing per- and polyfluoroalkyl substances (PFAS) through amending an adsorbent to contaminated soil planted with alfalfa. Science of The Total Environment, 2025, 1000: 180430. DOI:10.1016/j.scitotenv.2025.180430.
[17] Li XQ, Hua ZL, Wu JY, et al. Removal of perfluoroalkyl acids (PFAAs) in constructed wetlands: Considerable contributions of submerged macrophytes and the microbial community. Water Research, 2021, 197: 117080. DOI:10.1016/j.watres.2021.117080.
[18] Hua ZL, Li XQ, Zhang JY, et al. Removal potential of multiple perfluoroalkyl acids (PFAAs) by submerged macrophytes in aquatic environments: Tolerance of Vallisneria natans and PFAA removal in submerged macrophyte-microbiota systems. Journal of Hazardous Materials, 2022, 424: 127695. DOI:10.1016/j.jhazmat.2021.127695.
[19] Qian YF, Pei CH, Lv C, et al. Predicting uptake risk of PFAS in plant roots using transfer learning from hydroponic to soil systems. Journal of Agro-Environment Science, 2024, 43(11): 2516-2524. DOI: 10.11654/jaes.2024-0854.[钱一凡, 裴晨浩, 吕陈, 等. 水培到土培体系植物根系 PFAS 吸收风险的迁移机器学习研究. 农业环境科学学报, 2024, 43(11): 2516-2524.]
[20] He XM, Guo JY, Shen DK, et al.Application of machine learning in anaerobic digestion of organic solid waste:Recent advances and prospects. Journal of Southeast University (Natural Science Edition), 2025, 55(3): 743-750. DOI: 10.3969/j.issn.1001-0505.2025.03.014. [何晓满, 郭静远, 沈德魁, 等. 机器学习在有机固废厌氧消化中的应用进展与展望. 东南大学学报（自然科学版）, 2025, 55(3): 743-750.]
[21] 林鑫波, 田文哲, 陈炳坚, 等. 机器学习在抗生素环境污染领域的应用研究进展. 环境科技, 2024, 37(6): 71-76.]
[22] Song KH, Liu LY, Li YY.Advances in the Application of Machine Learning to the Design of Electrocatalysts. Journal of Henan University(Natural Science), 2025, 55(4): 442-452. DOI:10.15991/j.cnki.411100.20250611.001. [宋柯含, 刘乐园, 李园园. 机器学习在电催化剂设计中的应用进展. 河南大学学报（自然科学版）, 2025, 55(4): 442-452.]
[23] 温诗林. 基于机器学习的生物炭筛选及其对厌氧氨氧化效能优化研究[学位论文]. 浙江：杭州电子科技大学, 2025.
[24] 李珊珊, 陶玮庆, 柳栋宇, 等. 机器学习筛选乙酰丙酸加氢制 γ-戊内酯催化剂. 青岛大学学报（工程技术版）, 2025.
[25] Yang J, Hu RY, Zhao HP, et al. Analysis of Trade-offs and Driving Forces of Ecosystem Services in EcologicalBarrier Zone of Three Gorges Reservoir Area Based on XGBoost-SHAP Model. Journal of Soil and Water Conservation, 2025, 39(6): 127-138. DOI:10.13870/j.cnki.stbcxb.2025.06.038. [杨佳, 斛如媛, 赵海平, 等. 基于 XGBoost-SHAP 的三峡库区生态屏障区生态系统服务权衡及驱动力分析. 水土保持学报, 2025, 39(6): 127-138.]

[26] Li Y, Liao MJ, Zhang J.Interpretability enhancement model of random forest using ensemble pruning and multi-objective evolutionary algorithm. Application Research of Computers, 2024, 41(10): 2947-2954. DOI:10.19734/j.issn.1001-3695.2024.02.0047. [李扬, 廖梦洁, 张健. 利用集成剪枝和多目标优化算法的随机森林可解释增强模型. 计算机应用研究, 2024, 41(10): 2947-2954. ]
[27] Cui BQ, Gao HY, Han F. Model Correction Method for Rotating Machinery Based on Multi-Layer Perceptron. Chinese Journal of Theoretical and Applied Mechanics, 2025, 57(10)：2446-2456. DOI: 10.6052/0459-1879-25-283. [崔炳乾, 高海洋, 韩放. 基于多层感知机的旋转机械模型修正方法. 力学学报, 2025, 57(10)：2446-2456.]
[28] 任文宗, 刘婕, 邓玉婷, 等. 基于 RF 回归和 LSTM 神经网络的空气污染耦合性研究及预测——以北京市为例. 科学技术创新, 2025(10): 22-26.
[29] Gonçalves Saraiva FM, Rodrigues Guimarães AC, Da Silva Júnior OS, et al. Performance analysis of Gradient Boosting Regressor Using K-fold and Group K-fold for Predicting Permanent Deformation in Tropical Soils. Journal of Materials Research and Technology, 2026, 40: 3555-3567. DOI:10.1016/j.jmrt.2025.12.152.
[30] Qiu DE , Zhang JY, Qi KL, et al. Spatiotemporal evolution charactenistics and critical driving factor threshold analysis of habitat quality in Sichuan province based on SHAP-interpretable machine learning. Aeta Eeologica Sinica 2025 , 45(24): 1-18. DOI:10.20103/j.stxb.202503270704. [邱大鹅, 张军以, 齐渴路, 等. 基于 SHAP 可解释性机器学习的四川生境质量时空演变特征及关键驱动因子阈值分析. 生态学报, 2025 , 45(24): 1-18.]
[31] Kiani-Harchegani M, Sadeghi SH, Singh VP, et al. Effect of rainfall intensity and slope on sediment particle size distribution during erosion using partial eta squared. Catena, 2019, 176: 65-72. DOI:10.1016/j.catena.2019.01.006.
[32] Moriasi DN, Arnold JG, Van Liew MW, et al. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 2007, 50(3): 885-900. DOI:10.13031/2013.23153.
[33] Henseler J, Ringle CM, Sinkovics R R. The use of partial least squares path modeling in international marketing. Advances in International Marketing, 2009, 20: 277-320. DOI:10.1108/S1474-7979(2009)0000020014.
[34] Wu J, He Y, Li Y, et al. Revealing removal potential for nitrogen and phosphorus: Comparative analysis of three wetland terrestrial plants in purification effect and microbial community structure. Environmental Technology and Innovation, 2025, 40: 104493. DOI:10.1016/j.eti.2025.104493.
[35] 伍钦江, 潘文斌. 基于人工浮岛的湖泊营养盐去除效果研究. 当代化工研究, 2025(8): 91-93.
[36] Marzi D, Valente F, Luche S, et al. Phytoremediation of perfluoroalkyl and polyfluoroalkyl substances (PFAS): Insights on plant uptake, omics analysis, contaminant detection and biomass disposal. Science of The Total Environment, 2025, 959: 178323. DOI:10.1016/j.scitotenv.2024.178323.
[37] 肖君. 全氟辛酸（PFOA）在人工湿地系统中的影响和去除研究[学位论文]. 东南大学, 2024.
[38] 王澳宇. 水生植物—活性炭耦合人工湿地去除 PFAS 研究[学位论文]. 广东：广东工业大学, 2025.
[39] Dai Y, Ding D, Cui M, et al. Biotransformation of per- and polyfluoroalkyl substances (PFAS) in organisms. Chemical Engineering Journal, 2025, 524: 168618. DOI:10.1016/j.cej.2025.168618.
[40] Hong Y, He ZQ, Fang X, et al. The removal effect of 12 aquatic plants on nitrogen and phosphorus in the return flow of farmland. Journal of Central South University of Forestry and Technology, 2024, 44(10): 105-115, 180. DOI:10.14067/j.cnki.1673-923x.2024.10.011. [洪瑜, 何紫琪, 方晰等. 12 种水生植物对农田退水氮磷的去除效果. 中南林业科技大学学报, 2024, 44(10): 105-115, 180.]
[41] Verhofstad MJJM, Poelen MDM, Kempen MML, et al. Finding the harvesting frequency to maximize nutrient removal in a constructed wetland dominated by submerged aquatic plants. Ecological Engineering, 2017, 106: 423-430. DOI:10.1016/j.ecoleng.2017.06.012.