DOI:10.11918/202509041

基于机器学习的加载絮凝效能优化与机理解析

马 珍[1,2],叶 程[1,2],魏 卿[1,2],罗嘉程[1,2],王 品[1,2],钟士发[1,2],楚文海[1,2],徐祖信[1,2]

(1. 同济大学 环境科学与工程学院,上海 200092;
2. 水污染控制与资源绿色循环全国重点实验室(同济大学),上海 200092)

摘 要:利用加载絮凝技术强化溢流污水颗粒物去除,是控制雨天排水污染的关键手段。然而,传统优化方法依赖经验与单因素试验,难以系统揭示多参数间的复杂作用,导致工艺设计缺乏理论指导,运行效能不稳定。为此,引入机器学习方法,构建加载絮凝工艺的数据驱动优化与机制解析框架,以实现关键参数识别与澄清性能提升。通过整合580组试验数据,涵盖药剂特性、操作条件和介质特性等14个特征参数,构建随机森林(Random Forest,RF)、极端梯度提升(eXtreme Gradient Boosting,XGBoost)及轻量级梯度提升机(Light Gradient Boosting Machine,LightGBM)等多种机器学习模型预测加载絮凝的澄清性能。结果表明,LightGBM模型的预测准确性和泛化性能最佳(均方根误差 E_{RMS} 为7.31%,R^2 =0.74)。可解释的SHAP分析揭示,运行参数对加载絮凝澄清效果的影响占主导作用(贡献度占38.6%),显著高于介质参数和药剂参数,其中,沉淀时间与混凝强度为关键驱动因子。同时,结合水处理化学与动力学原理,阐明初始浊度与混凝强度之间的动态耦合效应,并揭示沉淀时间与介质相对密度之间存在显著的协同效应。溢流污水实验验证了经优化的参数组合可实现95%以上的颗粒物去除率,预测值与实测值高度吻合。本研究构建的"数据驱动建模—参数解释与推荐—实验验证"的框架,为加载絮凝工艺的精细化调控与智能决策提供了理论依据与方法支撑。

关键词:加载絮凝;机器学习;溢流污水;轻量级梯度提升机;合流制系统
中图分类号:X703 **文献标志码**:A **文章编号**:0367-6234(2026)06-0001-14

Machine learning-based efficiency optimization and mechanistic analysis of ballasted flocculation

MA Zhen[1,2], YE Cheng[1,2], WEI Qing[1,2], LUO Jiacheng[1,2], WANG Pin[1,2],
ZHONG Shifa[1,2], CHU Wenhai[1,2], XU Zuxin[1,2]

(1. College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China;
2. State Key Laboratory of Water Pollution Control and Green Resource Recycling (Tongji University), Shanghai 200092, China)

Abstract: Enhancing particle removal from combined sewer overflows using ballasted flocculation is a critical strategy for drainage pollution on rainy days. However, traditional optimization methods, which rely on empirical knowledge and one-factor-at-a-time experiments, struggle to systematically unravel the complex interactions among multiple parameters, resulting in a lack of theoretical guidance for process design and inconsistent operational performance. This study established a machine learning-based method for data-driven optimization and mechanistic interpretation of the ballasted flocculation process, aiming to identify key parameters and improve clarification efficiency. It integrated a comprehensive dataset of 580 experimental records, encompassing 14 feature parameters across reagent properties, operational conditions, and medium characteristics. Multiple machine learning models, including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), were developed to predict clarification efficiency. Results demonstrate that the LightGBM model achieves the best predictive accuracy and generalization capability, with a root mean square error (E_{RMS}) of 7.31% and a coefficient of determination (R^2) of 0.74. Interpretable shapley additive explanations analysis reveals that operational parameters contribute most significantly to clarification performance (contribution rate of 38.6%), substantially outperforming medium and reagent parameters. Settling time and coagulation velocity gradient are identified as the key drivers. Furthermore, by integrating water treatment chemistry and kinetic principles, the analysis elucidates the dynamic coupling effect between initial turbidity and coagulation velocity gradient and reveals

收稿日期:2025-09-10;**录用日期**:2025-10-31;**网络首发日期**:2025-11-17
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20251117.1421.005
基金项目:国家重点研发计划(2021YFC3200700,2021YFC3200702)
作者简介:马 珍(1999—),男,博士研究生;徐祖信(1956—),女,教授,博士生导师
通信作者:徐祖信,xzx@tongji.edu.cn

a significant synergistic effect between settling time and specific gravity of media. Experimental verification using actual combined sewer overflows confirms that the optimized parameter combination achieves over 95% particulate removal, with strong agreement between predicted and actual values. This study establishes a research framework comprising "data-driven modeling, parameter interpretation and recommendation, and experimental validation", providing a theoretical basis and methodological support for the precise regulation and intelligent decision-making in ballasted flocculation.

Keywords: ballasted flocculation; Machine Learning; combined sewer overflow; Light Gradient Boosting Machine; combined sewer system

合流制系统溢流污染是高截污率城市水环境污染的重要原因,对中国城市黑臭水体治理成效构成严峻挑战[1]。溢流污水中高浓度的颗粒物不仅显著降低水体清澈度,更对现有水处理系统产生巨大压力。颗粒物的高效去除,可以实现溢流污水中胶体物质以及附着在颗粒物上的多种污染物(如重金属、磷、有机微污染物)的同步去除,从而实现对溢流污水中大部分以颗粒态形式存在的核心污染物的有效去除[2]。混凝技术因其颗粒污染物去除效果优异、工艺简单、运行可靠等优点,已广泛用于溢流污染控制[3]。然而,常规混凝沉淀工艺存在启动时间长、絮体沉降速度慢等固有缺陷[4],导致占地面积大,难以满足城市溢流污水快速高效处理的需求。相比之下,加载絮凝工艺通过形成高密度、大尺寸絮体,可有效提升出水水质和系统单位面积处理能力(处理负荷可达80 $m^3/(m^2 \cdot h)$)[5],已在饮用水、市政污水、雨水及溢流污水处理等领域展现出良好的应用前景[6-8]。

通常,加载絮凝工艺的澄清效果受加载介质、悬浮颗粒、絮凝药剂之间多重作用机制的影响,包括静电中和、吸附架桥、网捕卷扫以及表面官能团络合等[8-10]。在加载絮凝工艺中,加载介质通过直接掺入絮体,显著影响系统的处理负荷。常用加载介质主要包括石英砂和磁粉。采用石英砂的加载絮凝工艺技术成熟、成本较低[11],而磁粉因其相对密度更大,能形成沉降速度更快的加载絮体[12]。然而,其他材料(如活性炭、火山石、沸石)同样具备作为加载介质的潜力,相关研究尚不充分。同时,絮凝药剂(种类、投加量)通过中和胶体颗粒表面电荷以及促进絮体聚集生长,对加载絮凝工艺的澄清效率具有显著影响[13]。此外,运行参数的优化是确保加载絮凝工艺完整发挥效能的关键步骤。如对混凝/絮凝接触时间、搅拌强度等参数进行合理设计与优化,可以有效减少中试或全规模试验的轮次,从而在工艺设计和优化过程中大幅节约成本[8]。然而,现有文献中关于实现最佳澄清效果的工艺条件(介质选择、药剂组合与投加量、运行参数等)报道不一。通过传统的实验方法系统评估各因素的定量贡献,仍是一个耗时且繁琐的过程。

相较于传统的控制变量法,机器学习(Machine Learning,ML)能够同时处理更多相关特征与更广泛的数据,并有效揭示工艺运行中的复杂非线性关系[14-16]。目前,机器学习技术已在常规絮凝领域取得应用进展。Lin 等[17]利用时间序列模型预测自来水厂的混凝剂剂量和出水浊度,该模型能有效捕捉不同特征间的隐含关系,并充分考虑特征自身的历史演变规律。Yokoyama 等[18]结合深度推理模型 ConcNeXt 与图像传感器实现了污泥脱水过程中絮凝剂的自动投加,实现了混凝/絮凝过程自动化与精细化控制。Lu[19]利用随机森林(Random Forest,RF)模型预测壳聚糖的重金属去除效能,并基于特征重要性分析识别关键影响因子,从而为新型絮凝剂的设计与配方优化提供理论依据。此外,Bankole 等[20]利用多层感知器(Multilayer Perceptron,MLP)模型模拟了絮体的动力学演变过程,为实时监测与智能控制提供参考。综上,机器学习已在常规絮凝领域形成较为系统的研究体系,其应用涵盖工艺参数预测、图像识别到材料建模等多层次应用。这些研究展现了机器学习在解析复杂非线性系统的强大能力,为深入理解并精准优化水处理过程的整体动态行为提供了新视角与决策支持。然而,时序模型通常在满足平稳假设的数据上表现良好,而对非线性、非平稳及复杂交互特征的数据,其性能则显著下降。许多模型在特定厂站或实验条件下表现良好,但若仅给出预测值(如投加量)而缺乏特征重要性分析,则难以获得实际操作人员的信任。尽管机器学习已经在常规絮凝领域得到成功应用,但其在加载絮凝工艺中的研究尚属空白。此外,相比常规絮凝,加载絮凝过程的瞬态性和非线性更强,其关键动态特征(如絮体-加载介质结合效率、沉降速率分布)难以通过传统模型捕捉。特别是高质量数据集的缺乏以及对"加载介质-药剂-运行参数"与出水水质间复杂耦合机制认识的不足,严重制约了机器学习方法在加载絮凝工艺优化中的应用。

基于前期研究,通过系统的数据采集与筛选,构建了一个包含580组样本的加载絮凝数据集。为准

确预测加载絮凝的澄清效率，通过交叉验证的方法并考虑不同水质背景，建立了多层感知机（MLP）、随机森林（RF）、极端梯度提升（eXtreme Gradient Boosting，XGBoost）和轻量级梯度提升机（Light Gradient Boosting Machine，LightGBM）4 种机器学习模型。进一步地，采用 Shapley 可加性解释（SHapley Additive exPlanations，SHAP）方法进行定量分析，以深入解析各输入变量（包括介质参数、药剂参数、运行参数和水质参数）对模型输出的贡献度，从而优化工艺参数组合并指导加载介质与药剂的高效筛选。最后，通过独立的溢流污水实验结果，验证了优选模型的预测准确性与泛化能力。通过上述步骤，构建了“数据驱动建模—参数解释与推荐—实验验证”的闭环研究框架，为提升加载絮凝工艺的处理效能、实现参数精准调控提供了重要的数据支撑和理论依据。

1　基本理论和方法

1.1　数据集构建

通过 Web of Science 核心数据库对加载絮凝相关文献进行了系统性检索与分析。鉴于颗粒物浓度是表征溢流污染负荷及多种附着污染物的关键替代参数[2,10,21-22]。检索使用“Ballasted flocculation”和“Turbidity”作为关键词，以聚焦颗粒物去除效果的研究。同时，将沉降方式明确限定为重力沉降，以排除旋流分离等其他技术。选取颗粒物去除率（particle removal efficiency，E_R）作为加载絮凝工艺核心评价指标，其计算公式为

$$E_R = \frac{c_0 - c_i}{c_0} \tag{1}$$

式中：c_0为初始颗粒物浓度，c_i为沉降后的颗粒物浓度。

对于以图表形式呈现颗粒物去除率的研究，采用 Web Plot Digitizer（https://apps.automeris.io/wpd/）从图表中提取响应数据。对于因实验设计而未涉及的步骤，其对应特征值进行特殊编码。具体而言，未进行絮凝步骤的样本，其“絮凝剂种类”被标记为“blank”，“絮凝剂剂量”“搅拌强度”“絮凝时间”被记录为“0”。通过上述流程，从最终遴选的文献[7-8,12,21,23-28]中提取并构建了一个包含 580 组有效样本的数据集，用于后续建模分析。该数据集已作为附件提供（data_model.csv）。

1.2　特征工程与水质背景分组

系统梳理了影响加载絮凝工艺性能的关键因素，并将其归纳为 4 大类，即运行参数、加载介质参数、药剂参数以及水质参数。各类别所涵盖的详细特征变量见表 1。

表 1　14 项加载絮凝特征类别与数据类型

Tab.1　Characteristic categories and data types of ballasted flocculation

类别	特征	单位	数据类型
操作参数	混凝强度	s^{-1}	数值
	混凝时间	min	数值
	絮凝强度	s^{-1}	数值
	絮凝时间	min	数值
	沉淀时间	min	数值
加载介质参数	介质尺寸	μm	数值
	介质相对密度	—	数值
	介质剂量	$m^2 \cdot L^{-1}$	数值
药剂参数	混凝剂类型	—	类别
	混凝剂剂量	$mg \cdot L^{-1}$（以 Fe 或 Al 计）	数值
	絮凝剂类型	—	类别
	絮凝剂剂量	$mg \cdot L^{-1}$	数值
水质参数	酸碱度	—	数值
	初始浊度	NTU	数值
目标变量	颗粒物去除率 E_R	%	数值

为评估水质背景的数据异质性对模型泛化能力的影响，按水源类型将数据集划分为市政污水、地表水、模拟水样与溢流污水 4 个子集。通过在各个子集上测试已训练模型的预测性能，以判定系数（coefficient of determination，R^2）为评价指标，系统验证模型在不同水质条件下的适应性与稳健性。

1.3　模型选择与训练

为提高模型训练效率并确保评估的客观性，采

用分层随机划分法将数据集按8∶2分为训练集(80%,464个样本)和测试集(20%,116个样本)[29]。通过Scikit-Learn库(版本1.1.3)中的StandardScaler对输入特征进行标准化处理,使各特征满足均值为0、标准差为1的分布,以消除量纲差异并改善模型收敛性。

本研究开发并评估了4个基于不同原理的机器学习模型,即多层感知器(Multilayer Perceptron, MLP)、随机森林(Random Forest, RF)、极端梯度提升(eXtreme Gradient Boosting, XGBoost)、轻量级梯度提升机(Light Gradient Boosting Machine, LightGBM)。所有模型均采用完整数据集($N=580$)进行训练和测试,输入特征为1.2节所述的14项工艺参数,输出变量为颗粒物去除率(E_R)。在模型训练阶段,采用了五折交叉验证策略。该方法通过多次随机划分训练集与验证集来评估模型性能,有效降低数据抽样波动和异质性带来的不确定性,从而确保所模型结果的稳健性。

模型性能通过判定系数(R^2)和均方根误差(root mean square error, RMSE, E_{RMS})进行量化,计算公式见式(2)和(3)。一般而言,较低的E_{RMS}和较高的R^2值代表模型具备更优的预测精度与拟合能力[30-31]。

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

$$E_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3}$$

式中:$\hat{y}_i$为输出的预测值,y_i为从文献中收集的输出的真实值,$\bar{y}$为所有输出值的平均值,N为训练或测试数据集中的数据样本数。

1.4 参数优化与推荐机制

将训练完成后的模型作为目标函数代理,并基于模型预测结果,应用SHAP方法进行可解释分析,以识别关键参数及其最优区间。SHAP是一种基于博弈论的模型解释方法,能够统一解释各类机器学习模型的输出[32-33],计算方法见式(4)。其核心思想是将每个预测值视为所有输入特征的联合贡献,并通过计算SHAP值来公平分配每个特征的边际贡献。为评估不同类别参数的整体重要性,计算了运行参数、加载介质参数、药剂参数和水质参数这4类特征其SHAP绝对值的均值,以进行重要性排序与比较[34-35]。

$$f(x) = \phi_0 + \sum_{i=1}^{M}\phi_i \tag{4}$$

式中:$f(x)$为模型对样本的预测输出,ϕ_0为训练数据上模型输出的期望值,ϕ_i为第i个特征对预测的边际贡献(即SHAP值)。

首先,通过SHAP条形图量化并比较各特征的全局重要性。随后,利用SHAP蜂群图进一步揭示重要特征对输出影响的方向(正向或负向)以及其特征值的分布规律。其次,利用部分依赖图(partial dependence plot, PDP)分析关键输入变量与颗粒物去除率之间的偏依赖关系[36]。同时,分析多个特征间的交互影响。最后,综上为实际水质条件下实现高颗粒物去除率提供优化的参数组合建议。

1.5 实验验证与反馈

为验证模型的预测性能,开展了20组(LightGBM和XGBoost各10组)针对溢流污水的加载絮凝处理实验。实验选取了5种不同相对密度的加载介质,每种介质的相对密度均通过相对密度瓶法重复测量2~3次确定。随后,将每种加载介质过筛,并使用激光粒度仪测量获取3种粒径的加载介质。所用介质的相对密度与尺寸见表2。溢流污水收集自上海市的一处泵站,水质如表3所示。加载絮凝实验程序参考了Lapointe和Barbeau的流程[21],首先,将水样快速混合($G=300\ s^{-1}$)10 s;随后,注入混凝剂并在设定搅拌强度下快速搅拌相应时间;接着注入絮凝剂,转为慢速搅拌相应时间;最后,静置沉淀预定时间,使用广口移液器于液面下2~3 cm取样。所有实验均在室温(23 ℃±1 ℃)下进行,且每组条件均重复2~3次,以控制实验误差,保证结果的可靠性。

表2 加载介质特性

Tab.2 Properties of ballast media

加载介质	相对密度	粒径1 (D50)/μm	粒径2 (D50)/μm	粒径3 (D50)/μm
磁粉	4.16	100	130	160
石英砂	2.65	100	130	160
沸石	2.20	100	130	160
火山石	1.79	100	130	160
活性炭	1.50	100	130	160

表3 溢流污水水质

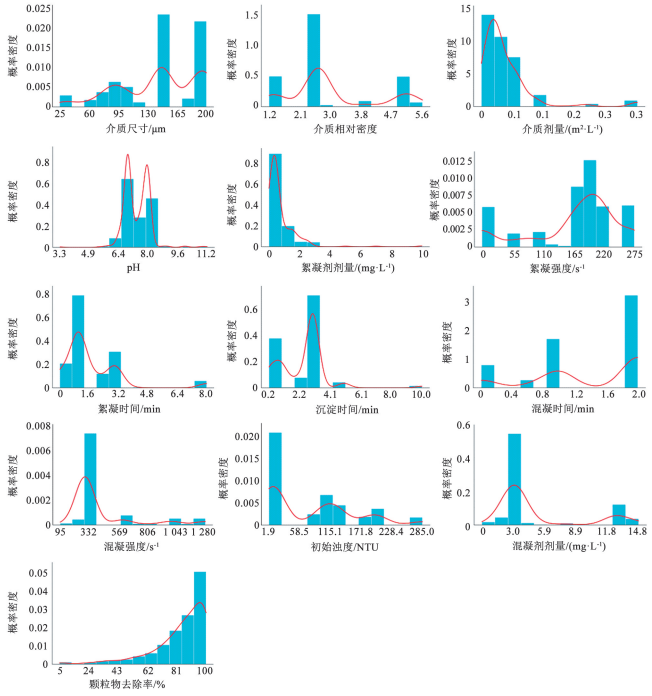
Tab.3 Water quality of combined sewer overflows

水体	浊度/NTU	pH	化学需氧量/(mg·L^{-1})	总磷/(mg·L^{-1})	总有机碳/(mg·L^{-1})	氨氮/(mg·L^{-1})
溢流污水	352.1±8.2	7.9±0.4	126.5.0±1.0	5.2±0.2	58.6±3.2	7.5±0.4

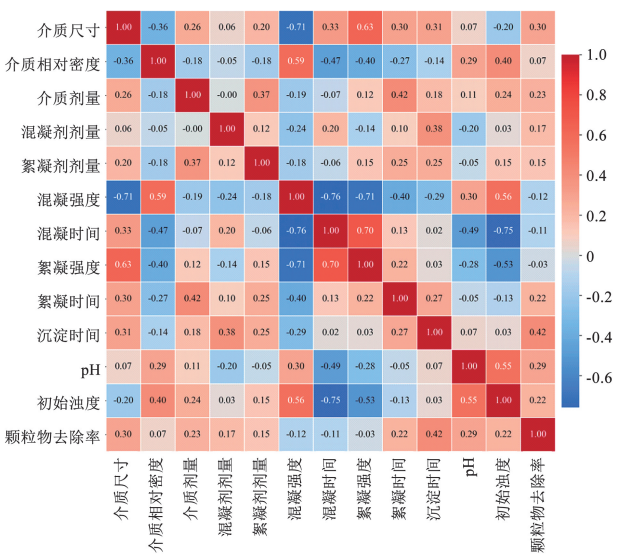
2　结果与讨论

2.1　加载絮凝数据统计性分布

如图 1(a)所示，首先对加载絮凝过程中涉及的运行、药剂、介质及水质参数进行了数据密度和分布分析。加载介质的投加剂量主要集中在 0 ~ 0.03 m^2/L。介质尺寸和相对密度是影响加载絮凝澄清效果的关键因素。本数据集中的加载介质尺寸主要分布在约 143 μm 和 185 μm 附近，已有研究指出，当介质尺寸大于 125 μm 时，体系更能承受较高的水力负荷[8]。加载介质相对密度集中在 2.65，这主要源于数据集中所用介质以石英砂为主。药剂的种类及用量对絮体的表面电荷及沉降性能有重要影响。数据集中的絮凝剂包括阴离子聚丙烯酰胺、阳离子聚丙烯酰胺、非离子聚丙烯酰胺和淀粉。研究表明，絮凝剂的类型会显著影响絮体的尺寸、强度及恢复能力[23]。数据集中大部分絮凝剂的投加量低于 1 mg/L，这与其作用机制相符，即高分子链状结构通过架桥作用促进加载介质与微絮体结合，从而加速絮体沉降[5]。数据集涵盖的混凝剂为铝盐和铁盐。研究表明，铁盐所形成的絮体沉降速度更快，铝盐则对处理设备的腐蚀性相对较低[37]。混凝剂的投加量主要集中在 3 mg/L 和 12 mg/L 附近。得益于加载介质和絮凝剂对絮体的强化效果，数据集中大部分工况的沉淀时间可缩短在 5 min 内。加载絮凝数据集的颗粒物去除率大多集中在 90% 左右，这充分体现了加载介质与高分子絮凝剂的共同作用，即通过形成高密度、大尺寸的加载絮体，实现了工艺的高效澄清。



(a) 加载絮凝参数分布



(b) 皮尔逊相关性

图 1　加载絮凝数据集特征

Fig. 1　Characteristics of ballasted flocculation dataset

加载絮凝性能受加载介质、药剂、运行参数等多种因素影响。为初步评估这些特征与颗粒物去除率之间的线性关联，绘制了所有 12 项数值特征与颗粒物去除率的皮尔逊相关性图(图 1(b))。线性相关分析结果表明，各特征与去除率之间以及部分特征两两之间的线性关系较弱(没有特征的相关性系数超过 0.8)，且部分特征的关联模式(如混凝强度–混凝时间、混凝时间–絮凝强度)与水动力学机制不符。这表明过程加载絮凝内在的复杂非线性特性使得常规线性回归模型可靠性不足。鉴于线性方法的局限性，在构建预测模型时转向能够捕捉复杂非线性关系的机器学习算法。为避免因过早剔除特征而丢失潜在的重要交互信息，初始阶段保留了全部 14 个特征用于后续建模。

2.2　4 种机器学习模型性能对比

图 2 比较了 4 种机器学习模型(MLP、RF、XGBoost、LightGBM)在测试集上的预测性能。具体而言，MLP、RF、XGBoost、LightGBM 在测试集上的 E_{RMS} 分别为 10.33%、7.53%、7.47%、7.31%，对应的测试集 R^2 为 0.49、0.73、0.73、0.74。MLP 在所

有 4 个模型中表现最差,其训练集 R^2(0.61)显著高于测试集 R^2(0.49),表明 MLP 模型在数据训练过程中产生了典型的过拟合,泛化能力不足。这可能是 MLP 模型对加载絮凝数据集中复杂的非线性关系刻画能力有限。尽管 MLP 模型在本数据集上表现不佳,但将其作为基准模型进行对比,其相对较弱的性能反而从侧面印证了树集成模型(RF、XGBoost、LightGBM)在此特定数据集上的优势。相比之下,RF 模型表现出优秀的泛化能力,其训练集和测试集的 E_{RMS} 值非常接近。这表明 RF 模型学习到普适性规律,测试集性能是真实可靠的。然而,RF 对训练集的拟合相对保守(训练集 R^2 = 0.83),这可能意味着其对数据中某些更复杂的非线性模式捕捉不足。XGBoost 和 LightGBM 同属梯度提升模型,其算法通过顺序地构建决策树来不断修正前序模型的误差,通常能在训练集上达到极高的精度。这解释了二者为何在训练集上性能优异($R^2 \approx$ 0.9)。值得注意的是,XGBoost 在训练集上取得最低的 E_{RMS}(5.33%)和最高的 R^2(0.92),但测试集性能与 RF 相近,这表明其在训练阶段存在一定的过拟合。相比之下,LightGBM 展现出最佳的综合性能,其在测试集上实现最低 E_{RMS}(7.31%)和最高 R^2(0.74),同时保持了训练集与测试集的良好一致性。这表明 LightGBM 在模型复杂度与泛化能力之间取得了更优的平衡。综上,在本研究构建的 4 种机器学习模型中,LightGBM 在预测精度与泛化稳健性方面表现最为出色,被确定为预测加载絮凝澄清性能的最优模型,其优化后的超参数配置详见表 4。

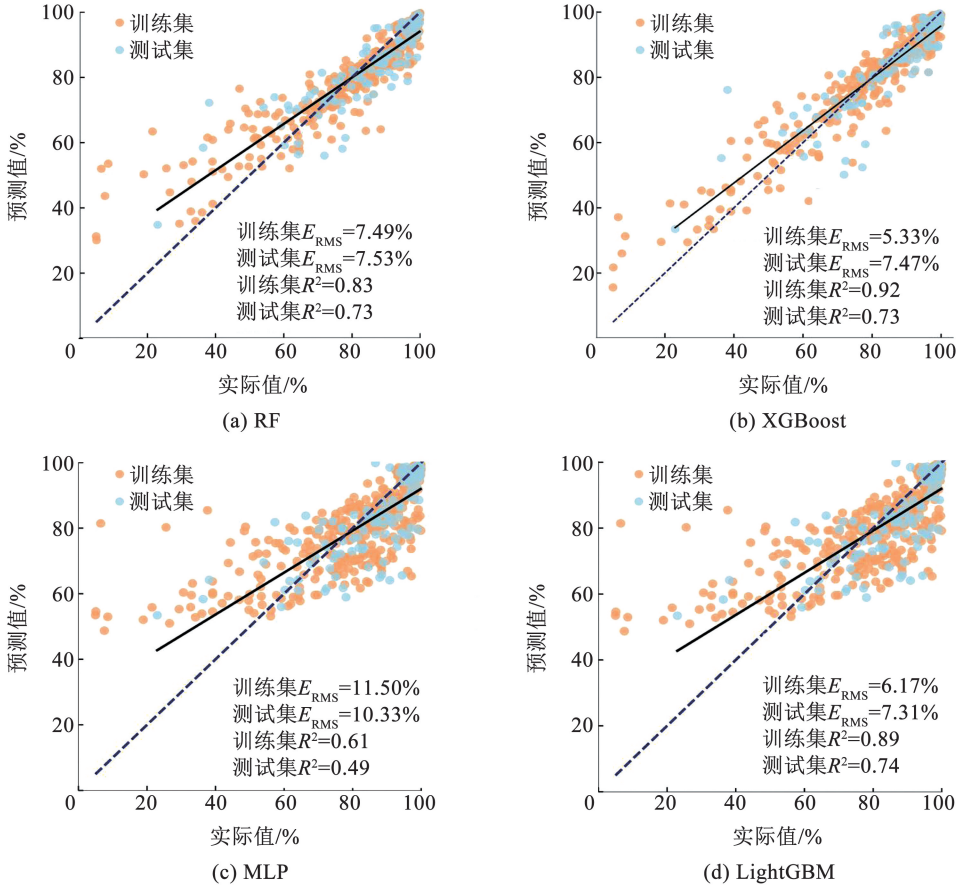


图 2　模型性能对比

Fig. 2　Model performance comparison

2.3　LightGBM 模型在不同水质子集上的性能对比

为评估数据异质性对模型性能的影响,根据水质类型(市政污水、地表水、模拟水体、溢流污水)将加载絮凝数据集划分为 4 个子集,并分别测试了 LightGBM 模型的预测性能。如图 3 所示,LightGBM 在所有 4 个水质组别的测试集上均表现出了优异且稳定的预测能力,R^2 均高于 0.69,且集中分布于 0.69 ~0.79 这一狭窄的高性能区间内。然而,本研究观察到模型在地表水和模拟水样上性能略低于市政污水和溢流污水,这可能是不同水质背景下的“特征–响应关系”发生了漂移。与成分相对单一的模拟水样/地表水样相比,成分复杂的市政污水或溢流污水絮凝机制、中间过程及颗粒物去除率间的关系可能存在本质差异。要求模型试图用同一套参数化框架去拟合这些存在差异的物理化学过程,可能导致其在特定水质子集上表现出预测性能的波动。

尽管存在上述性能差异，LightGBM 模型在所有子集上均保持了良好的预测性能，这表明模型成功克服了数据异质性带来的挑战，并未产生系统性的预测偏差。尤为重要的是，在本研究聚焦的核心应用场景“溢流污水”处理上，模型取得了最佳的测试集性能（R^2 =0.79）。这强有力地证明，尽管不同水质背景数据集具有异质性，LightGBM 模型已成功捕捉到了针对溢流污水加载絮凝处理最关键且最稳健的规律。这为其在实际溢流污水控制中的应用提供了有力支持。后续也将在 2.5 节通过独立的实际溢流污水进一步验证模型可靠性。

表 4　LightGBM 模型超参数设置

Tab. 4　Hyperparameter settings of LightGBM model

模型	超参数	取值
LightGBM	n_estimators	1 000
	learning_rate	0.05
	max_depth	-1
	num_leaves	31
	min_child_samples	30
	subsample	0.8
	colsample_bytree	0.8
	reg_alpha	0.1
	reg_lambda	0.1
	random_state	50
	n_jobs	-1

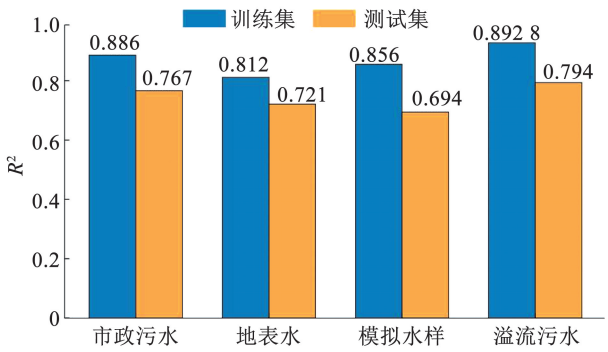


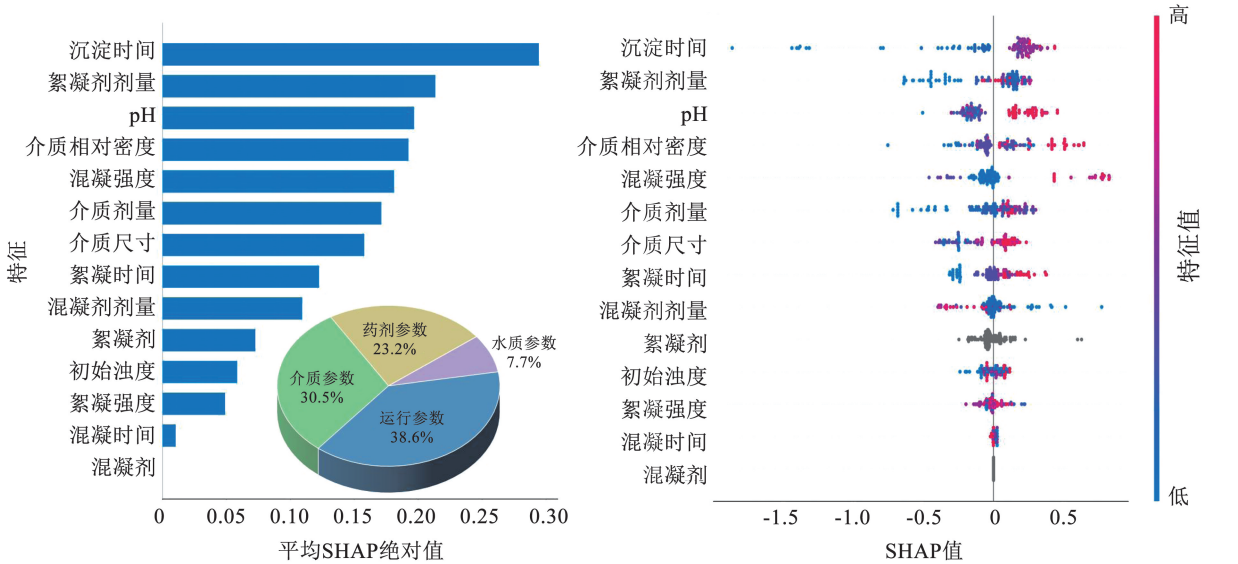
图 3　不同水质背景下 LightGBM 模型的训练集与测试集性能对比

Fig. 3　Performance comparison of training and test sets of LightGBM model under different water quality backgrounds

2.4　SHAP 解释和特征分析

为解析 LightGBM 模型的决策机制，采用 SHAP 方法进行可解释分析。图 4 展示了根据 SHAP 值大小排序的 14 项输入特征的全局重要性，其中，每个特征的 SHAP 值量化了其对模型预测颗粒物去除率的平均边际贡献。

为深入解析不同类别参数的影响，本研究将输入特征归为 4 类，即加载介质参数、药剂参数、运行参数和水质参数。基于分类特征的 SHAP 重要性分析结果（图 4（a））显示，运行参数对模型输出的贡献度最高（38.6%），显著高于加载介质参数（30.5%）和药剂参数（23.2%），而水质参数的贡献度相对最低（7.7%）。水质参数贡献度相对较低，表明加载絮凝工艺对其进水水质波动可能不敏感，因而具备处理多种水体的潜力。这一发现与文献中报道的该工艺广泛应用于地表水、雨水和污水处理的工程实践相吻合[2,6]。



(a) 基于LightGBM模型的SHAP条形图和特征类别重要性　　(b)基于LightGBM模型的SHAP蜂群图

注：点的颜色表示特征值的大小，横坐标表示其对预测（输出）的贡献[38]。

图 4　基于 LightGBM 模型的 SHAP 可解释性分析

Fig. 4　SHAP interpretability analysis based on LightGBM model

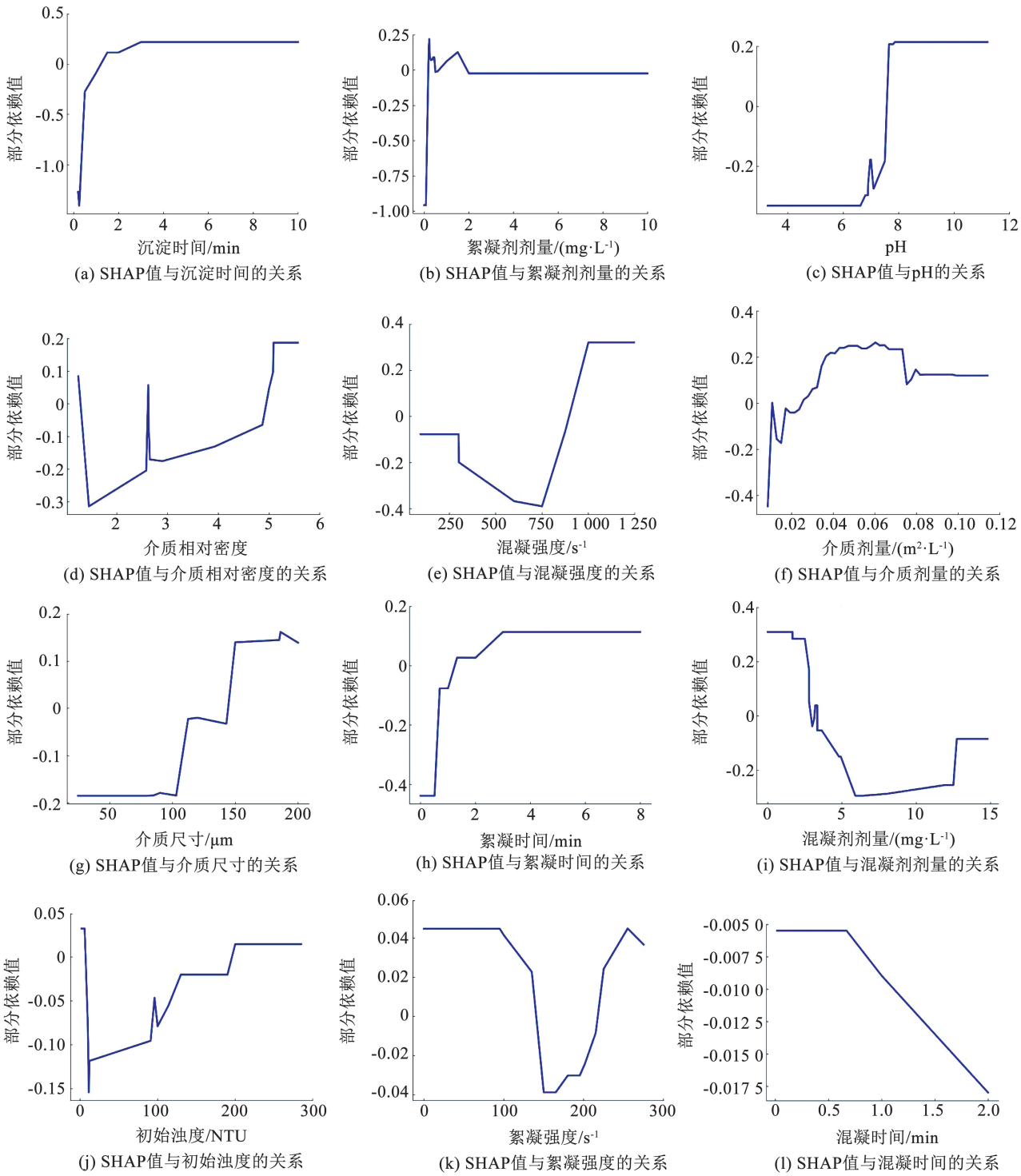
运行参数的优化对于平衡工艺的澄清效果与运行成本至关重要。从全局特征重要性(图4(a))及SHAP蜂群图(图4(b))均可看出,沉淀时间、混凝强度与絮凝时间等运行参数在所有特征中居于主导地位。其中,沉淀时间被识别为影响澄清效果的最关键因子。部分依赖图(图5(a))表明,颗粒物去除率随沉淀时间(0.2~3 min)的增加而显著上升,但当沉淀时间超过3 min后,去除率的增长趋势趋于平缓。这一规律与加载絮体特有的沉降动力学相符,即高密度絮体可在混合结束后短时间内(<3 min)完成主体沉降过程[8]。进一步地,SHAP交互分析(图6(a))揭示,沉淀时间与介质相对密度之间存在显著的协同效应。具体而言,当系统采用高相对密度介质(如磁粉,相对密度>4)时,配合超过2 min的沉淀时间可实现最优的污染物去除效能。这一协同机制可通过斯托克斯定律与"加载絮凝"的核心理念得到合理解释。颗粒的沉降末速 v 用斯托克斯公式近似描述为

$$v=\frac{(\rho_{\mathrm{p}}-\rho_{\mathrm{f}})gr^{2}}{18\mu} \tag{5}$$

式中:g 为重力加速度,ρ_{p}和ρ_{f}分别为颗粒与流体的密度,r 为颗粒粒径,μ 为流体的黏度。高相对密度介质与絮体结合后,能显著增大絮体的有效密度与粒径,从而大幅提升其沉降末速[39]。另一方面,高相对密度介质作为絮体生长的"核心",其充分结合与絮体结构的完善需要一定的时间。大于2 min的沉淀时间为高密度加载絮体提供了充分的沉降路径,并使其能与水中残留的微小颗粒发生有效的接触扫掠。因此,高相对密度介质提供了关键的沉降驱动力,充足的沉降时间则为高效分离提供了保障,二者协同作用,缺一不可。一个重要的发现是,当沉淀时间延长至3 min后,介质相对密度的影响显著减弱。此时,选用低成本介质(如活性炭、无烟煤)可在维持效能的同时降低运行成本,这为工艺的经济性优化提供了关键依据。对于混凝强度,SHAP分析进一步揭示了其与初始浊度的动态耦合关系。部分依赖分析(图5(e))表明,混凝强度为780~1 000 s^{-1}对去除率的正向贡献最为显著,该区间对应着絮体密实度与沉降性的最优平衡点[40]。这与原位显微观测结果一致,即较高的速度梯度可加速絮体初期阶段的生长速率[41]。SHAP交互作用分析(图6(b))显示,当初始浊度高于175 NTU,更高的混凝强度(>1 000 s^{-1})相较于低速度梯度(<600 s^{-1})能显著提升澄清效果。这一现象可以从胶体颗粒的碰撞动力学与絮体结构演变角度进行解释。当初始浊度高于175 NTU,水中胶体颗粒的浓度极大。根据斯莫卢霍夫斯基(Smoluchowski)碰撞理论,颗粒的碰撞频率与颗粒浓度的平方成正比[42]。在此条件下,采用高速度梯度既能提供充足能量克服颗粒间的电斥能垒、促进脱稳,又能通过强烈的流体剪切作用,极大提升单位时间内的颗粒碰撞频率,从而快速形成结构密实的微絮体。在高浊度背景下,这种由高强度搅拌带来的高碰撞频率是形成可沉降絮体的主导因素,其正面效应远大于剪切可能造成的破碎风险[40,43]。此外,图4(a)和4(b)显示,混凝强度的重要性显著高于絮凝强度。这是因为在混凝阶段的高搅拌强度直接决定了胶体颗粒的脱稳效率与初始絮体的结构形态[37],此为后续絮体生长的基础;而絮凝阶段的絮体仅需中等搅拌强度使絮体缓慢长大并保持结构稳定[7]。部分依赖图(图5(h))进一步表明,较长的絮凝时间(3~8 min)有利于颗粒物的去除。这是因为更长的絮凝时间为聚合物充分发挥桥接吸附作用提供了保障[28],从而促进絮体结构稳定与完善。综上,SHAP解释系统揭示了运行参数在澄清性能中的分层贡献机制,即沉淀时间主导物理分离速率,混凝强度决定初始絮体形态,絮凝时间影响加载絮体生长质量。基于上述机制,建议通过匹配中高相对密度介质与中等沉降时间、在高浊度水体下采用较高混凝强度(>1 000 s^{-1})等策略,以实现加载絮凝过程能效与成本的协同优化。

进一步利用SHAP方法解析了加载介质参数对工艺效能的影响机制。在介质剂量方面,图5(f)显示,剂量为0.015~0.04 m^2/L对澄清效果贡献显著,此区间内颗粒物去除率的预测值随剂量增加而快速提升,表明该范围是形成高密度絮体的关键窗口。当剂量超过0.07 m^2/L后,其特征重要性(SHAP值)呈下降趋势,说明其影响出现边际递减。这主要源于过量介质在搅拌中会增加絮体破碎的风险,构成动力学限制[12]。这一发现提示,在工程应用中需审慎平衡介质投加的经济效益与运行稳定性。此外,不同水体需要的加载介质剂量也有不同,相比饮用水,污水处理过程达到相同最终浊度应用的加载介质剂量是前者的1.5倍[7]。介质相对密度的作用机制更为复杂。图5(d)的部分依赖图显示,介质相对密度的PDP值随相对密度的增加而递增,这说明提升介质相对密度是提高澄清效率的有效途径。研究显示,采用磁粉等高相对密度介质替代石英砂,可使系统处理的总悬浮固体负荷提升两倍以上[7]。然而,当介质相对密度超过5后,其对性能的提升效益逐渐趋于饱和。Lapointe等[7]提出,使用更高相对密度的加载介质需匹配更高的速度梯度,这将同时加剧剪切效应与能耗,构成其主要约束。因此,介质相对密度的选择本质上是"沉降增益"与"剪切风险"之间的权衡。介质尺寸的影响机制揭

示了其与运行参数的紧密耦合。图 4(b) 的 SHAP 摘要图显示，增大介质尺寸对提升澄清效率具有积极影响。部分依赖结果(图 5(g))进一步表明，在 100 ~ 150 μm 内，澄清效率随加载介质尺寸增大而提升。这主要是因为更大尺寸的介质能够作为更坚实的骨架，促进形成更大、更密实的絮体，该现象与已有研究结论一致[8]。SHAP 交互作用分析(图 6(c))揭示，对于较大尺寸(>150 μm)的介质，需匹配较高的絮凝强度(>200 s^{-1})才能充分发挥其作为絮体骨架的包裹能力。有研究证实，当介质尺寸与絮凝强度实现优化匹配时，仅需 30 s 的极短沉降时间即可实现 99.5% 的颗粒物去除率[12]。上述现象源于流体动力条件对“介质-絮体”接触效率的关键调控作用。这表明对介质参数的优化不能孤立进行，必须将其与水力条件视为一个耦合系统进行协同设计与整体优化。



注：X 轴表示输入特征的范围，Y 轴表示输入特征对模型预测结果的边际效应。

图 5　加载絮凝特征的部分依赖图

Fig. 5　Partial dependence plots for ballasted flocculation features

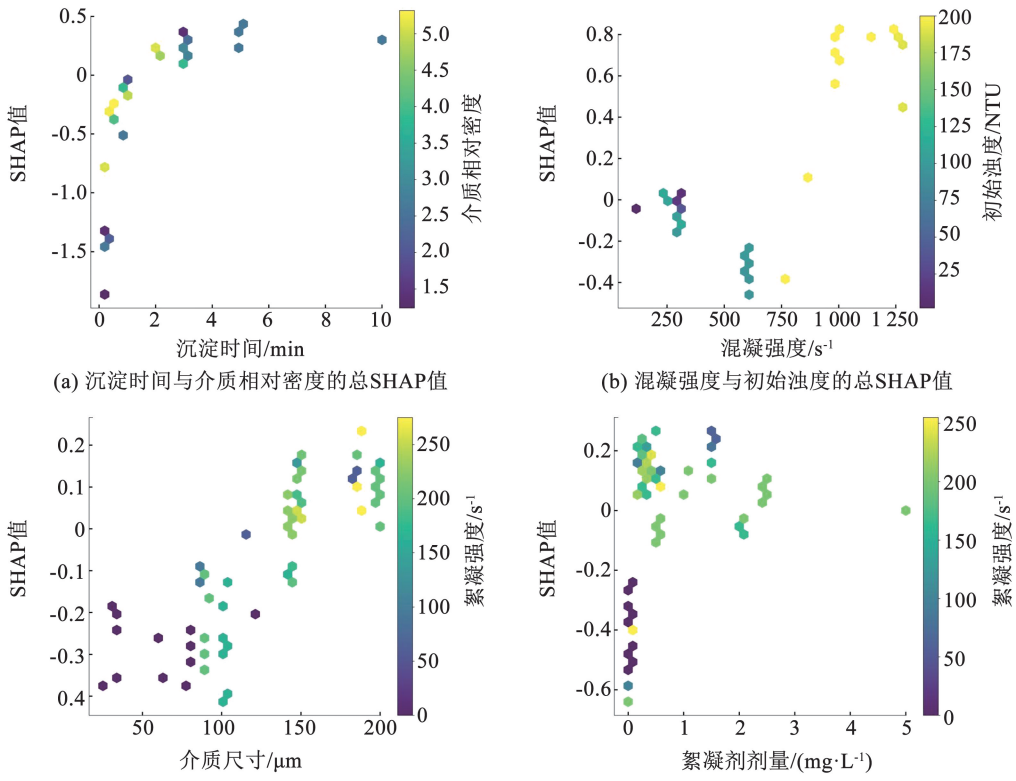


图 6　SHAP 量化分析

Fig. 6　SHAP quantitative analysis

虽然在全局重要性排序中药剂参数的贡献度相对有限,但对于调节悬浮颗粒及絮体的表面性质起到至关重要的作用。如图 4(a)所示,在全部特征中,絮凝剂剂量是重要性排名前三的特征之一,且其平均 SHAP 值显著高于其他药剂参数。部分依赖分析(图 5(b))表明,当絮凝剂剂量高于 1 mg/L 时,其对澄清效率产生显著正向影响,颗粒物去除率的预测值随之快速升高,表明在此剂量范围内可显著促进加载絮凝过程。而在 0.3 ~0.5 mg/L 的低剂量区间,PDP 值较低,表明过低剂量的聚合物不利于加载絮凝的澄清效能。此外,高于 2 mg/L 絮凝剂剂量会导致澄清效能下降,此趋势与 SHAP 值的分布特征相吻合,共同指向高剂量聚合物可能引发“桥联饱和”[23]或“胶体保护”效应[10,44],从而抑制絮体的有效聚集与生长。SHAP 交互作用分析(图 6(d))进一步揭示,当絮凝剂剂量高于 2 mg/L 时,若匹配较高的絮凝强度(>200 s^{-1}),可协同实现更优的澄清效果。充分的流体剪切有助于聚合物分子链的伸展与吸附点位的暴露,从而提升其桥联吸附能力。反之,在低搅拌强度下,高剂量絮凝剂的 SHAP 值为负值,表明体系内形成了稳定的微胶体结构,阻碍了有效絮凝。混凝剂的核心作用是促进悬浮颗粒脱稳并形成微絮体。其 SHAP 值随剂量的变化呈现显著的非线性特征:在最佳投加范围内,SHAP 值为正向且贡献显著,反映了电中和效应有效促进了微絮体的聚集;一旦剂量超过该区间,SHAP 值转为负值,导致体系澄清效率下降。这一非线性趋势可归因于两种机制:在低剂量下,电中和作用不充分,无法实现有效脱稳;在过量投加时,则会引发“电荷反转”与水解产物包覆效应,致使胶体颗粒重新稳定[37,45]。图 4(b)中的负 SHAP 分布进一步验证了这种“过度投加抑制”效应。此外,虽然混凝剂与絮凝剂的电荷类型对絮体特性存在影响[23],但其 SHAP 重要性远低于投加量参数。这表明在加载絮凝工艺中,相较于药剂的固有种类特性,其投加剂量与絮凝强度之间的协同匹配对于最终澄清效率起着更为关键的作用。

如图 4(a)所示,水质参数的全局贡献度最低(7.7%),但这从侧面反映了加载絮凝工艺对不同水质条件波动具有良好的适应性,有助于保障运行稳定性与药剂利用效率。部分依赖分析(图 5(c))表明,当 pH > 8 时,PDP 值显著升高,表明该 pH 下对颗粒物去除具有明显的促进作用。这一正向贡献主要归因于高 pH 条件,金属氢氧化物沉淀及无定型羟基络合物的快速形成,增强了对悬浮颗粒的网捕卷扫与吸附包裹效应[45-46]。此外,Zhu 等[47]对溶液 pH 的研究表明,碱性环境能够提高高分子絮凝剂在颗粒表面的吸附强度,从而改善加载絮体的

结构稳定性与沉降性能。初始浊度的部分依赖图(图 5(j))表明，当浊度从 20 NTU 增至 200 NTU，对应的 PDP 值呈持续上升趋势，意味着在此范围内澄清效率得到显著提升。该现象符合经典的絮凝动力学原理，即较高的颗粒物浓度通过增加碰撞频率，有效促进了絮体的生长。SHAP 交互作用分析(图 6(b))进一步揭示，初始浊度与混凝强度之间存在显著的协同效应。这表明对于高浊度进水，必须匹配优化后的混凝强度，才能协同促进絮体尺寸与密实度的同步增长，该发现与文献中关于高浊度条件下易形成超高速沉降絮体的报道高度吻合[24]。

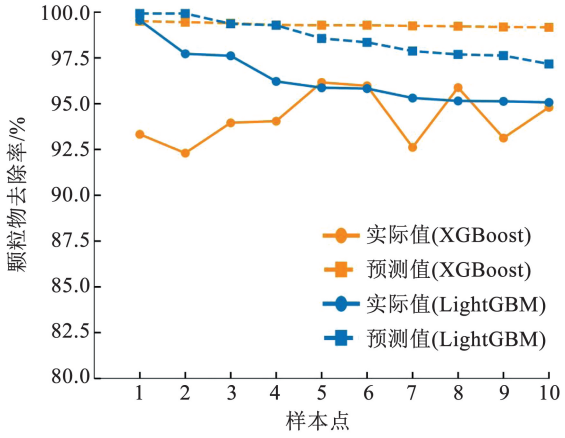
2.5　实际溢流污水验证

为评估模型在实际应用中的性能，采用独立的实际溢流污水水样，对由 LightGBM 与 XGBoost 模型所推荐的最优工艺参数组合进行了实验验证。两种模型推荐的参数组合详见表 5。实验验证结果(图 7)清晰表明，LightGBM 模型的预测准确性与可靠性显著优于 XGBoost 模型。具体而言，LightGBM 模型的预测值与实测值高度吻合，两者变化趋势一致，且绝对误差维持在较低水平。这表明 LightGBM 模型成功地学习了数据中潜在的加载絮凝规律，其推荐参数能够有效地迁移到新的实际溢流污水水样中，具备强大的泛化能力。相比之下，XGBoost 模型则表现出明显的系统性偏差，其预测值持续且稳定地高于实测值，存在明显的正偏差。这种系统性的高估，即过于“乐观”的预测，在实际工程应用中蕴含风险。若依此决策，可能导致药剂投加不足或操作条件偏离最优区间，从而使实际出水水质无法达标。上述性能差异根源在于模型学习策略的不同。XGBoost 模型在本研究中很可能对训练数据中的特定模式或噪声产生了过拟合，即过度专注于训练集的局部特征而未能充分掌握通用规律，致使其对新的实际水样适应性不足。尽管 XGBoost 在训练集上表现出色，但其泛化能力在此次外部验证中暴露了不足。综上所述，与 XGBoost 模型相比，LightGBM 模型展现出更优的预测准确性与稳健性，因而其推荐的工艺参数在实际应用中具有更高的实用价值与可靠性。

表 5　XGBoost 与 LightGBM 模型优化工艺参数

Tab. 5　Optimized process parameters of XGBoost and LightGBM models

模型	介质尺寸/μm	介质相对密度	介质剂量/($m^2 \cdot L^{-1}$)	混凝剂剂量/($mg \cdot L^{-1}$)	絮凝剂剂量/($mg \cdot L^{-1}$)	混凝强度/s^{-1}	混凝时间/min	絮凝强度/s^{-1}	絮凝时间/min	沉淀时间/min
XGBoost	100	1.79	0.197	4.133	8.948	1 272.852	0.435	61.192	6.797	6.473
	130	2.20	0.111	3.857	5.221	1 252.175	0.125	55.946	2.672	4.450
	130	4.16	0.093	0.897	1.671	1 010.997	0.818	60.842	5.272	1.777
	100	2.20	0.101	2.631	0.591	1 148.373	1.957	11.119	4.326	1.929
	160	1.79	0.289	7.858	1.721	1 173.540	1.700	152.578	3.090	5.194
	100	1.79	0.039	3.528	1.435	148.560	0.611	50.219	7.996	2.987
	160	2.20	0.320	0.269	2.969	1 093.774	0.427	35.419	6.535	9.001
	130	1.79	0.018	4.943	2.909	1 190.284	1.753	2.216	2.606	3.721
	130	1.79	0.253	2.505	5.673	967.561	0.900	12.929	3.772	2.315
	160	1.79	0.322	0.209	4.561	977.714	1.195	238.663	3.636	7.681
LightGBM	160	2.20	0.225	1.573	9.790	662.752	0.876	167.793	6.112	7.551
	130	2.20	0.151	0.145	9.206	292.515	1.649	21.082	7.169	8.097
	160	4.16	0.070	13.354	3.844	132.608	0.960	81.891	5.583	5.514
	100	2.20	0.244	1.206	4.828	1 222.485	1.526	256.043	5.314	5.510
	100	2.20	0.185	2.212	7.012	127.448	0.701	34.490	3.182	8.579
	160	2.65	0.310	2.699	8.766	247.184	0.052	94.745	3.207	6.856
	130	2.20	0.140	13.116	1.256	411.564	1.216	93.067	6.730	5.386
	100	2.20	0.320	1.916	8.183	488.560	0.983	58.294	3.264	5.585
	160	4.16	0.199	4.428	7.676	1 225.525	0.579	7.792	6.122	5.001
	160	2.20	0.183	0.740	2.070	933.594	0.428	238.933	4.456	8.938



注:线的每个拐点对应一个样本点。

图 7 实验数据集的实际值与模型预测值的折线图

Fig. 7 Line graph of actual values of experimental data set and predicted values of model

2.6 工艺参数优化的经济性分析与发展前景

基于机器学习的数据驱动优化,系统识别了加载絮凝工艺实现高效颗粒物去除的关键参数组合,其核心特征包括采用中高相对密度介质、维持较高的混凝强度并匹配相应的絮凝剂投加量。然而,该高效参数组合从实验室走向大规模工程应用,必须跨越经济可行性与长期运行稳定性的鸿沟。该参数组合的运行成本主要来源于药剂与能耗。较高的絮凝剂投加量直接推高了物料成本,尽管本研究确定的投加区间逼近技术性能的极限,但未必与全生命周期成本模型下的经济最优点相重合。此外,高混凝强度依赖于高功率混合设备,导致电耗急剧上升。由于搅拌功率与速度梯度平方正相关,能耗随强度提升呈非线性增长,这一效应在规模化处理设施中会被急剧放大,成为制约经济性的关键因素。

为提升工艺的经济可行性,提出两项核心策略:一是实施精准智能的动态调控,即基于进水水质(如浊度、流量)实施前馈-反馈控制,在低负荷时自动降低混凝强度与药剂投加,在高负荷时启动强化模式,实现处理资源的“按需分配”,从而统筹达标排放与运行能耗;二是构建生命周期成本模型,整合当地的能源单价、药剂费用及设备维护开销,在技术可行的参数空间内,识别适用于不同处理规模与水质特征的经济最优解,为工程定制化设计提供依据。展望未来,通过融合先进传感技术与智能算法,有望实现加载絮凝工艺的自主实时优化与精准控制,最终构建高效、稳定、低耗的智慧水处理单元。

3 结 论

1)本研究成功构建了一个数据驱动的研究框架,系统地将机器学习方法应用于加载絮凝工艺的澄清效率预测与运行优化,推动了该工艺的调控范式由传统经验决策向数据智能驱动的转变。

2)通过系统对比 RF、MLP、XGBoost 和 LightGBM 4 种模型,确立了 LightGBM 模型为预测加载絮凝颗粒物去除率的最优模型($E_{RMS}=7.31\%$,$R^2=0.74$)。该模型在包含溢流污水在内的多种水质条件下表现出优异的泛化稳定性。

3)基于 SHAP 可解释性分析结果发现,运行参数是决定加载絮凝澄清效能的核心类别(贡献度 38.6%)。部分依赖与特征交互分析揭示了加载介质、运行条件及药剂投加之间存在复杂的非线性耦合关系,从数据视角深化了对颗粒物生长与沉降动力学机制的理解。

4)依据模型解析结果提出的参数优化策略,经实际溢流污水实验验证后,能够稳定实现 95% 以上的颗粒物去除率,验证了预测结果的可靠性,并凸显了该数据驱动方法在实际工程应用中的可行性与价值。

5)从工程应用与经济优化角度出发,本研究进一步提出了动态投加与全生命周期成本模型的构建策略,为在实践中平衡技术效能与经济成本提供了可行路径。

参考文献

[1]徐祖信, 张竞艺, 徐晋, 等. 城市排水系统提质增效关键技术研究:以马鞍山市为例[J]. 环境工程技术学报, 2022, 12(2): 348
XU Zuxin, ZHANG Jingyi, XU Jin, et al. Research on key technologies for quality improvement and efficiency enhancement of urban drainage system: a case study of Ma'anshan City[J]. Journal of Environmental Engineering Technology, 2022, 12(2): 348. DOI: 10.12153/j.issn.1674-991X.20210842

[2]GASPERI J, LABORIE B, ROCHER V. Treatment of combined sewer overflows by ballasted flocculation: removal study of a large broad spectrum of pollutants[J]. Chemical Engineering Journal, 2012, 211: 293. DOI:10.1016/j.cej.2012.09.025

[3]徐祖信, 王卫刚, 李怀正, 等. 溢流污水中溶解性有机物的混凝去除特性[J]. 同济大学学报(自然科学版), 2012, 40(12): 1842
XU Zuxin, WANG Weigang, LI Huaizheng, et al. Coagulation removal characteristics of dissolved organic matter in overflow wastewater[J]. Journal of Tongji University (Natural Science), 2012, 40(12): 1842. DOI:10.3969/j.issn.0253-374x.2012.12.017

[4]马珍, 韩梦琪. 管道溢流的污染特征及水质管控技术与策略[J]. 给水排水, 2022, 58(9): 147
MA Zhen, HAN Mengqi. Pollution characteristics of pipeline overflow and technology and strategy for water quality management[J]. Water & Wastewater Engineering, 2022, 58(9): 147. DOI: 10.13789/j.cnki.wwe1964.2021.12.19.0005

[5]MURUJEW O, GEOFFROY J, FOURNIE E, et al. The impact of

polymer selection and dose on the incorporation of ballasting agents onto wastewater aggregates[J]. Water Research, 2020, 170: 115. DOI:10.1016/j.watres.2019.115346

[6] LAPOINTE M, BARBEAU B. Characterization of ballasted flocs in water treatment using microscopy[J]. Water Research, 2016, 90: 119. DOI:10.1016/j.watres.2015.12.018

[7] LAPOINTE M, BROSSEAU C, COMEAU Y, et al. Assessing alternative media for ballasted flocculation [J]. Journal of Environmental Engineering, 2017, 143(11): 04017071. DOI:10.1061/(ASCE)EE.1943-7870.0001271

[8] DESJARDINS C, KOUDJONOU B, DESJARDINS R. Laboratory study of ballasted flocculation[J]. Water Research, 2002, 36(3): 744. DOI:10.1016/S0043-1354(01)00256-1

[9] GHANEM A V, YOUNG J C, EDWARDS F G. Mechanisms of ballasted floc formation[J]. Journal of Environmental Engineering, 2007, 133(3): 271. DOI:10.1061/(ASCE)0733-9372(2007)133:3(271)

[10] HE Weipeng, LUO Jiacheng, HUANG Jinguo, et al. Two-stage injection of polymer and microsand during ballasted flocculation for treating kaolin waters with or without humic acid: floc evolutional characteristics, performance and mechanisms[J]. Water Research, 2024, 259: 121846. DOI:10.1016/j.watres.2024.121846

[11] GUIBELIN E, DELSALLE F, BINOT P. High speed settling of stormwater with microsand[C]//Chemical Water and Wastewater Treatment III. Berlin, Heidelberg: Springer, 1994: 307. DOI:10.1007/978-3-642-79110-9_23

[12] QASIM M, PARK S, KIM J O. The role of ballast specific gravity and velocity gradient in ballasted flocculation [J]. Journal of Hazardous Materials, 2020, 399: 122970. DOI: 10.1016/j.jhazmat.2020.122970

[13] YANG Bingqian, ZHOU Peng, TIAN Long, et al. The nanoscale explanation of metal cations differences in enhancing the Fe(III) coagulation performance [J]. Water Research, 2025, 280: 123524. DOI:10.1016/j.watres.2025.123524

[14] ZHONG Shifa, ZHANG Kai, BAGHERI M, et al. Machine learning: new ideas and tools in environmental science and engineering[J]. Environmental Science & Technology, 2021, 55(19): 12741. DOI:10.1021/acs.est.1c01339

[15] DAI Zhenhua, XU Jihong, GUAN Jian, et al. TP-Transformer: an interpretable model for predicting the transformation pathways of organic pollutants in chemical oxidation processes[J]. Environmental Science & Technology, 2025, 59(30): 15853. DOI:10.1021/acs.est.5c02701

[16] ZHANG Jing, FU Kaixing, ZHONG Shifa, et al. Data-driven insights into resin screening for targeted per-and polyfluoroalkyl substances removal using machine learning [J]. Environmental Science & Technology, 2025, 59(7): 3603. DOI:10.1021/acs.est.4c14223

[17] LIN Subin, KIM J, HUA Chuanbo, et al. Coagulant dosage determination using deep learning-based graph attention multivariate time series forecasting model[J]. Water Research, 2023, 232: 119665. DOI:10.1016/j.watres.2023.119665

[18] YOKOYAMA H, YAMASHITA T, KOJIMA Y, et al. Deep learning-based flocculation sensor for automatic control of flocculant dose in sludge dewatering processes during wastewater treatment [J]. Water Research, 2024, 260: 121890. DOI: 10.1016/j.watres.2024.121890

[19] LU Chun, XU Zuxin, DONG Bin, et al. Machine learning for the prediction of heavy metal removal by chitosan-based flocculants[J]. Carbohydrate Polymers, 2022, 285: 119240. DOI: 10.1016/j.carbpol.2022.119240

[20] BANKOLE A O, MORUZZI R, NEGRI R G, et al. Machine learning framework for modeling flocculation kinetics using non-intrusive dynamic image analysis [J]. Science of the Total Environment, 2024, 908: 168452. DOI: 10.1016/j.scitotenv.2023.168452

[21] LAPOINTE M, BARBEAU B. Selection of media for the design of ballasted flocculation processes[J]. Water Research, 2018, 147: 25. DOI:10.1016/j.watres.2018.09.041

[22] SIELIECHI J, LARTIGES B, SKALI-LAMI S, et al. Floc compaction during ballasted aggregation [J]. Water Research, 2016, 105: 361. DOI:10.1016/j.watres.2016.09.015

[23] HE Weipeng, LUO Jiacheng, WU Yujie, et al. Differently charged polyacrylamides (PAMs) significantly affect adsorption affinity and associated floc growth behaviors during ballasted flocculation: performance and mechanism [J]. Separation and Purification Technology, 2024, 329: 125279. DOI:10.1016/j.seppur.2023.125279

[24] SUZUKI Y, KAKU R, TAKAHASHI K, et al. Optimum conditions for high-speed solid-liquid separation by ballasted flocculation[J]. Water Science and Technology, 2023, 88(1): 35. DOI:10.2166/wst.2023.204

[25] QASIM M, PARK S, KIM J O. Unraveling the role of ballast surface charge at floc growth behavior in ballasted flocculation[J]. Separation and Purification Technology, 2021, 278: 119507. DOI: 10.1016/j.seppur.2021.119507

[26] HE Weipeng, XIE Zhoulong, LU Wenjing, et al. Comparative analysis on floc growth behaviors during ballasted flocculation by using aluminum sulphate (AS) and polyaluminum chloride (PACl) as coagulants[J]. Separation and Purification Technology, 2019, 213: 176. DOI:10.1016/j.seppur.2018.12.043

[27] KUMAR S, GHOSH N C, KAZMI A A. Ballasted sand flocculation for water, wastewater and CSO treatment [J]. Environmental Technology Reviews, 2016, 5(1): 57. DOI:10.1080/21622515.2016.1207715

[28] LAPOINTE M, BARBEAU B. Substituting polyacrylamide with an activated starch polymer during ballasted flocculation[J]. Journal of Water Process Engineering, 2019, 28: 129. DOI: 10.1016/j.jwpe.2019.01.011

[29] ZHONG Shifa, CHEN Yushan, LI Jibai, et al. Screening environmentally benign ionic liquids for CO_2 absorption using representation uncertainty-based machine learning[J]. Environmental Science & Technology Letters, 2024, 11(11): 1193. DOI:10.1021/acs.estlett.4c00524

[30] GAO Haiping, ZHONG Shifa, ZHANG Wenlong, et al. Revolutionizing membrane design using machine learning-bayesian optimization[J]. Environmental Science & Technology, 2022, 56(4): 2572. DOI:10.1021/acs.est.1c04373

[31] ZHONG Shifa, GUAN Xiaohong. Count-based morgan fingerprint: a more efficient and interpretable molecular representation in developing machine learning-based predictive regression models for water contaminants′ activities and properties [J]. Environmental

Science & Technology, 2023, 57(46): 18193. DOI:10.1021/acs.est.3c02198

[32] AAS K, JULLUM M, LØLAND A. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values[J]. Artificial Intelligence, 2021, 298: 103502. DOI:10.1016/j.artint.2021.103502

[33] RODRÍGUEZ-PÉREZ R, BAJORATH J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions[J]. Journal of Computer-Aided Molecular Design, 2020, 34(10): 1013. DOI: 10.1007/s10822-020-00314-0

[34] ZHANG Kai, ZHONG Shifa, ZHANG Huichun. Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning[J]. Environmental Science & Technology, 2020, 54(11): 7008. DOI: 10.1021/acs.est.0c02526

[35] ZHONG Shifa, ZHANG Yanping, ZHANG Huichun. Machine learning-assisted QSAR models on contaminant reactivity toward four oxidants: combining small data sets and knowledge transfer[J]. Environmental Science & Technology, 2022, 56(1): 681. DOI: 10.1021/acs.est.1c04883

[36] KERRIGAN D, BARR B, BERTINI E. PDPilot: exploring partial dependence plots through ranking, filtering, and clustering[J]. IEEE Transactions on Visualization and Computer Graphics, 2025: 1. DOI:10.1109/TVCG.2025.3545025

[37] DUAN Jinming, GREGORY J. Coagulation by hydrolysing metal salts[J]. Advances in Colloid and Interface Science, 2003, 100: 475. DOI:10.1016/S0001-8686(02)00067-2

[38] GAO Yidan, ZHONG Shifa, ZHANG Kai, et al. Abiotic reduction of organic and inorganic compounds by Fe(Ⅱ)-associated reductants: comprehensive data sets and machine learning modeling[J]. Environmental Science & Technology, 2023, 57(46): 18026. DOI:10.1021/acs.est.2c09724

[39] QASIM M, PARK S, LEE J, et al. Evaluation of floc settling velocity models through image analysis for ballasted flocculation[J]. Desalination and Water Treatment, 2019, 144: 37. DOI:10.5004/dwt.2019.23586

[40] WANG Zhenbei, NAN Jun, JI Xiaoyu, et al. Effect of the micro-flocculation stage on the flocculation/sedimentation process: the role of shear rate[J]. Science of the Total Environment, 2018, 633: 1183. DOI:10.1016/j.scitotenv.2018.03.286

[41] HE Weipeng, NAN Jun, LI Haoyu, et al. Characteristic analysis on temporal evolution of floc size and structure in low-shear flow[J]. Water Research, 2012, 46(2): 509. DOI: 10.1016/j.watres.2011.11.040

[42] YANG Zhen, YANG Hu, JIANG Ziwen, et al. A new method for calculation of flocculation kinetics combining Smoluchowski model with fractal theory[J]. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2013, 423: 11. DOI: 10.1016/j.colsurfa.2013.01.058

[43] SERRA T, COLOMER J, LOGAN B E. Efficiency of different shear devices on flocculation[J]. Water Research, 2008, 42(4/5): 1113. DOI:10.1016/j.watres.2007.08.027

[44] LÓPEZ-MALDONADO E A, HERNÁNDEZ-GARCÍA H, ZAMUDIO-AGUILAR M A M, et al. Chemical issues of coffee and tule lignins as ecofriendly materials for the effective removal of hazardous metal ions contained in metal finishing wastewater[J]. Chemical Engineering Journal, 2020, 397: 125384. DOI:10.1016/j.cej.2020.125384

[45] TEH C Y, BUDIMAN P M, SHAK K P Y, et al. Recent advancement of coagulation-flocculation and its application in wastewater treatment[J]. Industrial & Engineering Chemistry Research, 2016, 55(16): 4363. DOI: 10.1021/acs.iecr.5b04703

[46] CAMP T R, ROOT D A, BHOOTA B V. Effects of temperature on rate of floc formation[J]. Journal AWWA, 1940, 32(11): 1913. DOI:10.1002/j.1551-8833.1940.tb19608.x

[47] ZHU Liyi, LYU Wensheng, MAO Xiaohui, et al. Effect of solution pH and polyethylene oxide concentration on surface/interface properties, flocculation and rheology of concentrated monodisperse ultrafine synthetic tailings slurry[J]. Powder Technology, 2023, 430: 119002. DOI:10.1016/j.powtec.2023.119002

(编辑 刘 彤)