

王鑫瑜, 姜德坤, 林荣, 等. 基于多特征选择方法的 BiLSTM-GRU 融合模型在典型新污染物预测中的模拟对比[J]. 净水技术, 2026, 45(3): 25-37, 64.

Wang X Y, Jiang D K, Lin R, et al. Simulated comparison of BiLSTM-GRU fusion model based on multi-feature selection methods in prediction of typical novel pollutants[J]. Water Purification Technology, 2026, 45(3): 25-37, 64.

基于多特征选择方法的 BiLSTM-GRU 融合模型在典型新污染物预测中的模拟对比

王鑫瑜¹, 姜德坤¹, 林荣², 林绍霞³, 刘宝震^{1,*}, 王永磊^{1,*}

(1. 山东建筑大学市政与环境工程学院, 山东济南 250101; 2. 芭乐<山东>环境科技有限公司, 山东济南 251400; 3. 济南水务集团有限公司凤凰路水厂, 山东济南 250000)

摘要 【目的】针对给水处理中磺胺类抗生素[磺胺甲恶唑(SMZ)、磺胺氯吡啶(SCP)]与臭味物质[2-甲基异茨醇(2-MIB)、土臭素(GSM)]等典型新污染物浓度低、时序波动大,且与工艺参数存在复杂非线性时滞关联导致精准预测难的问题,本文旨在解决传统特征选择方法无法有效捕捉此类动态关联、预测模型性能受限的痛点,为保障饮用水安全提供精准预测技术支持,同时为高级氧化工艺动态调控提供靶向依据。【方法】本文以济南市某中试基地高级氧化系统 1 200 余组监测数据为基础,构建双向长短期记忆网络-循环门单元(BiLSTM-GRU)融合模型。BiLSTM 通过双向时序建模捕捉参数间长期依赖关系,GRU 以简化门控结构提升计算效率并强化短期波动捕捉能力,二者融合实现“长时关联挖掘-短时特征细化”的互补优势。系统对比全特征输入、皮尔逊(Pearson)相关系数法(线性特征筛选)、格兰杰因果检验法(时序因果挖掘)3 种特征选择策略的预测效果。通过平均绝对误差(MAE)、均方根误差(RMSE)、决定系数(R^2)及 95%置信/预测区间等指标综合评价。【结果】格兰杰因果检验法筛选的 12 项特征能最大程度保留关键关联信息,使模型预测精度相对较高,2-MIB、GSM、SMZ、SCP 的 R^2 值分别达 0.941、0.935、0.893、0.916,较全特征输入相对提升 7.54%、8.22%、13.18%、11.98%,MAE 较全特征输入分别降低 7.0%、8.4%、13.2%、14.0%,RMSE 分别降低 25.8%、5.8%、0、26.5%,且置信区间与预测区间最窄,泛化能力相对更强。【结论】Pearson 相关系数法因其仅能捕捉线性关联,预测性能次之;全特征输入受冗余信息干扰,表现最差。本文明确了新污染物预测的最优特征选择策略,为高级氧化工艺动态调控提供精准靶点,也为低浓度复杂污染物的时序预测提供了技术参考。**关键词** 新污染物 双向长短期记忆网络-循环门单元(BiLSTM-GRU)融合模型 特征选择 Pearson 相关系数 格兰杰因果检验

中图分类号: X703 文献标志码: A 文章编号: 1009-0177(2026)03-0025-14

DOI: 10.15890/j.cnki.jsjs.2026.03.003

Simulated Comparison of BiLSTM-GRU Fusion Model Based on Multi-Feature Selection Methods in Prediction of Typical Novel Pollutants

Wang Xinyu¹, Jiang Dekun¹, Lin Rong², Lin Shaoxia³, Liu Baozhen^{1,*}, Wang Yonglei^{1,*}

(1. School of Municipal and Environmental Engineering, Shandong Jianzhu University, Jinan 250101, China;

2. Bale <Shandong> Environmental Technology Co., Ltd., Jinan 251400, China;

3. Fenghuang Road Waterworks, Jinan Water Group Co., Ltd., Jinan 250000, China)

Abstract [Objective] Aiming at the problem that typical novel pollutants such as sulfonamide antibiotics [sulfamethoxazole (SMZ), sulfachloropyridazine (SCP)] and odorants compounds [2-methylisoborneol (2-MIB), geosmin (GSM)] in water treatment are

[收稿日期] 2025-12-16

[基金项目] 国家重点研发计划(2022YFC3203704);山东省重点研发计划(2023TZXD019)

[作者简介] 王鑫瑜(2001—),女,硕士研究生,研究方向为水处理理论与技术,E-mail:1441394203@qq.com.

[通信作者] 刘宝震(1993—),男,副教授,研究方向为水处理理论与技术、污水处理及资源化、固体废弃物高值资源化,E-mail:liubaozhen23@sdjzu.edu.cn;

王永磊(1977—),男,教授,研究方向为水污染控制与资源化理论与技术,E-mail:wyl1016@sdjzu.edu.cn.

difficult to accurately predict due to their low concentrations, significant temporal fluctuations, and complex nonlinear time-lag correlations with process parameters, this paper aims to address the pain point that traditional feature selection method cannot effectively capture such dynamic correlations, leading to limited performance of prediction models. It provides technical support for accurate prediction to ensure drinking water safety and offers targeted basis for the dynamic regulation of advanced oxidation processes.

[**Methods**] Based on more than 1 200 sets of monitoring data from an advanced oxidation system at a pilot-scale test base in Jinan City, this paper constructed a bidirectional long short-term memory-gated recurrent unit (BiLSTM-GRU) hybrid model. BiLSTM captured the long-term dependencies among parameters through bidirectional temporal modeling, while GRU improved computational efficiency and enhanced the capability of capturing short-term fluctuations with its simplified gating structure. The fusion of the two realized the complementary advantages of long-term correlation mining and short-term feature refinement. The prediction effects of three feature selection strategies, namely full-feature input, Pearson correlation coefficient method (linear feature selection), and Granger causality test method (temporal causal mining) were systematically compared. A comprehensive evaluation was performed using indices including mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and the 95% confidence/prediction intervals. [**Results**] The 12 features selected by the Granger causality test method retained key correlation information to the greatest extent, leading to the optimal model prediction accuracy. The R^2 values for 2-MIB, GSM, SMZ, and SCP reached 0.941, 0.935, 0.893, and 0.916, respectively, which were 7.54%, 8.22%, 13.18%, and 11.98% higher than those of full-feature input. MAE values and RMSE values were 7.0%, 8.4%, 13.2%, 14.0% and 25.8%, 5.8%, 0, 26.5% lower than those of full-feature input. Additionally, the corresponding confidence and prediction intervals were the narrowest, indicating stronger generalization ability. [**Conclusion**] The Pearson correlation coefficient method achieves the second-best prediction performance as it only captures linear correlations. Full-feature input performs the worst due to interference from redundant information. This paper identifies the optimal feature selection strategy for emerging pollutant prediction, provides accurate targets for the dynamic regulation of advanced oxidation processes, and offers technical references for the temporal prediction of low-concentration complex pollutants.

Keywords novel pollutant bidirectional long short-term memory-gated recurrent unit (BiLSTM-GRU) fusion model feature selection Pearson correlation coefficient Granger causality test

给水处理中的磺胺类抗生素、臭味物质等典型新污染物已成为威胁饮用水安全的重要物质。磺胺类抗生素因长期滥用使其在水环境中残留率高,在我国湖泊中检出的质量浓度为 $0 \sim 67.18 \text{ ng/L}^{[1-2]}$,地中海流域检出率更是高达 $100\%^{[3]}$,其耐药性风险对公共健康构成潜在威胁;臭味物质则因气味阈值极低,即使微量残留也易引发用户投诉,且传统混凝、沉淀工艺对其去除率仅为 $11.5\% \sim 20.7\%^{[4]}$,需依赖紫外(UV)-过氧化氢(H_2O_2)-臭氧(O_3)等高级氧化工艺调控。然而,新污染物浓度与工艺参数的关联具有显著复杂性:一方面,新污染物浓度低且易受总有机碳(TOC)、 254 nm 处UV吸光度(UV_{254})等进水水质、 O_3 投加量时滞等工艺波动影响,时序数据呈现低基数、高波动特性^[5];另一方面,参数间交互关系隐蔽,如 O_3 浓度与2-甲基异莰醇(2-MIB)降解存在非线性时滞关联, H_2O_2 过量会引发自由基淬灭效应,传统特征选择方法难以精准捕捉此类动态关联^[6]。

深度学习模型在时序预测中展现出优势,其中长短期记忆网络(LSTM)通过门控机制解决传统循

环神经网络(RNN)的梯度消失问题,能有效捕捉长时依赖关系,但存在参数冗余、计算效率偏低的问题^[7];循环门单元(GRU)简化了LSTM的门控结构,将输入门与遗忘门整合为更新门,在保持相近预测性能的同时降低了计算复杂度,提升了训练速度,但其单向传播机制难以充分利用时序数据的上下文信息^[8];双向长短期记忆网络(BiLSTM)通过正向与逆向2个LSTM单元并行建模,能同时捕捉历史与未来的时序特征,强化了对复杂依赖关系的挖掘能力,但对短期突变特征的响应不够灵敏^[9]。基于此,本文构建双向长短期记忆网络-循环门单元(BiLSTM-GRU)融合模型,通过BiLSTM层挖掘参数间的长时因果关联与非线性依赖,利用GRU层优化短期波动特征的捕捉效率,既保留了BiLSTM对复杂时序关系的挖掘能力,又发挥了GRU的轻量化优势,提升模型在低浓度、高波动新污染物预测中的精度与效率。

特征选择是提升新污染物预测模型性能的核心环节,但当前方法存在显著局限。首先,全特征输入策略虽能保留潜在信息,但水质多维数据间交互影

响非常复杂,各个参数与新污染物浓度的关联度极低,冗余特征不仅导致模型学习噪声,还会增加计算负荷,尤其新污染物样本量有限,冗余信息对预测精度的干扰更为显著。其次,皮尔逊(Pearson)相关系数法作为传统线性特征选择工具,虽广泛应用于水质数据筛选,但存在本质局限^[10]。Pearson相关系数法适用于研究2个参数间的线性关系,由-1到1进行度量,且需通过相关强度划分筛选特征^[11]。然而,新污染物降解过程中的关键关联多为非线性,Pearson相关系数仅能捕捉整体线性趋势,对于非线性关系可能存在误差,这种关联模糊性对低浓度新污染物的精准预测极为不利^[12-13]。格兰杰因果检验法作为时序数据因果挖掘工具,更适配新污染物预测的需求^[14]。该方法通常用于判断平稳时间数据序列之间是否具有因果关系,匹配新污染物与工艺参数的动态关联^[15]。该方法需严格遵循平稳性检验、协整检验、向量自回归(VAR)模型构建的流程:首先,通过单位根检验(ADF)验证序列平稳性,避免出现虚假因果;其次,构建VAR模型,通过变量滞后阶数选择(varsoc)命令确定最优滞后期,且需验证模型稳定性,确保VAR模型有效;最终,通过F统计量与p值判断因果关系。该方法能有效挖掘水质参数间潜在因果关系,这对新污染物预测至关重要^[16]。

现有研究多聚焦新污染物去除工艺优化,针对其浓度预测的特征选择对比研究较少,且未验证3种方法的适用性。因此,本文以BiLSTM-GRU融合模型为核心,重点解决2大问题:(1)验证全特征输入、Pearson相关系数法、格兰杰因果检验法对新污染物浓度预测的性能差异,明确最优特征选择策略;(2)解析特征选择如何优化新污染物与工艺参数的关联机制,为高级氧化工艺动态调控提供精准靶点,聚焦特征选择这一关键环节,为新污染物精准预测提供技术路径。

1 材料与方法

1.1 数据来源与预处理

1.1.1 数据来源

本文的数据来源于济南某中试基地高级氧化系统的试验监测结果,试验以水厂原水为进水,通过加药泵控制磺胺类抗生素[磺胺甲恶唑(SCP)、磺胺氯哒嗪(SMZ)]与臭味物质[2-MIB、土臭素(GSM)]的

初始质量浓度维持在300 ng/L左右,过程中改变UV灯打开数量、O₃投加量、H₂O₂投加量以获取不同工况下的监测数据。监测时间为2024年8月—2025年3月(连续7个月),数据粒度为每2小时1组,共采集1200余组有效样本,监测指标涵盖11项进水特征,包括温度、pH、浊度、电导率、TOC、UV₂₅₄、氨氮、硝酸盐(NO₃⁻)、亚硝酸盐(NO₂⁻)、高锰酸盐指数、初始新污染物浓度;5项工艺参数,包括O₃浓度、H₂O₂投加量、UV功率、停留时间、进水流量;4项预测目标,包括出水2-MIB、GSM、SCP、SMZ浓度。

1.1.2 数据预处理

为保障模型训练的稳定性与数据质量,对原始数据进行系统化预处理:针对缺失值,采用线性插值法填补缺失率<3%的样本,通过已知数据点间的线性关系估算缺失值,避免因数据空缺导致的模型训练偏差^[17];针对异常值,考虑到新污染物浓度可能因藻类暴发、药剂投加波动等实际工况出现突变,且样本总量有限,若剔除异常值可能丢失关键动态信息,因此为保证数据的完整性保留所有异常值,以还原新污染物浓度的真实时序特征^[18];针对数据量纲差异,采用最小-最大(Min-Max)归一化方法将所有特征数据缩放至[0,1],消除不同量级指标的干扰,满足深度学习模型对输入数据尺度一致性的要求,确保处理后的数据既保留原始分布趋势,又能加速模型训练收敛^[19]。

考虑到水质数据的时序特性,采用时间序列滚动划分法分配数据,避免数据泄露导致的模型过拟合。将样本数据按时间顺序划分为训练集、验证集、测试集,比例为7:1:2,确保测试集数据均晚于训练集和验证集,符合实际预测场景。模型训练过程中,采用5折时间序列交叉验证,每次折叠均以更早的数据作为训练集、相邻的后续数据作为验证集,避免跨时间维度的信息泄露;同时进行10次重复试验,每次试验随机初始化模型权重,以降低随机因素对模型性能的影响,确保验证结果的可靠性与稳定性。

1.2 特征选择方法设计

1.2.1 全特征输入

作为基准对照,全特征输入直接纳入11项进水特征和5项工艺参数的所有16项监测指标,不进行任何特征筛选,该方法的核心目的是完整保留潜在

的参数关联信息,同时通过与其他特征选择方法的对比,验证冗余特征对模型性能的影响。

1.2.2 Pearson 相关系数法

基于 Pearson 相关系数实现线性特征筛选,该方法适用于量化 2 个参数间的线性关联强度,相关系数(r)取值为 $-1\sim 1$,通过表 1 的相关强度划分标准,设定筛选阈值为 $|r|>0.5$ (r 可正可负,故以绝对值形式表述),即保留强相关及极强相关特征。通过计算各输入特征与新污染物浓度的 Pearson 相关系数,筛选出相关性高的特征,剔除极弱相关或弱相关特征,最终形成精简后的线性关联特征集^[20]。

表 1 相关系数

Tab. 1 Correlation Coefficient

$ r $	相关强度
0.8~1.0	极强相关
0.6~0.8	强相关
0.4~0.6	中等程度相关
0.2~0.4	弱相关
0.0~0.2	极弱相关

1.2.3 格兰杰因果检验法

格兰杰因果检验方法常常被用于判断平稳时间数据序列之间是否具有因果关系,能够利用该方法判断一个时间序列是否对另一个时间序列具有因果影响^[21]。若序列 X 是序列 Y 的因,则表明 X 的相关变化能够对 Y 产生因果影响,这成为选取该方法作为水质数据处理方法的关键原因^[22]。

(1) 平稳性检验和协整检验

格兰杰因果检验法的应用前提是时间序列数据为平稳数据,否则会出现虚假检验结果,影响特征选择的合理性^[23]。如果序列满足平稳性条件,则可以进行检验的下一步操作;但若序列是非平稳的状态,则需要进一步检验时间序列之间是否存在协整关系。如果协整关系成立,也可以在协整框架下分析变量之间的因果关系。这一流程确保了因果检验的可靠性,同时兼顾了非平稳序列的特性^[24]。

(2) ADF

在格兰杰因果检验中常利用 ADF 检验来判断序列是否平稳,该方法通过数据是否存在单位根的原理进行判断分析。如果数据包含单位根,则该序列是非平稳的,这可能导致时间序列回归分析中出现伪回归现象^[25]。

(3) 构建 VAR 模型

格兰杰因果检验需要构建 VAR 模型。VAR 模型比传统联立方程模型预测更精准,因为它不受传统模型的诸多限制,能更准确地评估内生变量间的动态关系,从而更真实地反映现实情况。

1.3 模型构建与评价

1.3.1 BiLSTM-GRU 融合模型结构

BiLSTM-GRU 融合模型的构建逻辑是:BiLSTM 通过正向与逆向双通路建模,解决了单一 LSTM 仅能利用历史数据的局限性,充分挖掘时序数据的上下文关联性,强化对新污染物降解过程中长时非线性依赖的捕捉能力;GRU 通过简化门控结构,弥补了 BiLSTM 参数冗余、计算效率偏低的不足,同时,其动态更新机制能更灵敏地捕捉新污染物浓度的短期突变特征;二者融合实现了长时关联挖掘与短时特征细化的功能互补,既保证了对复杂时序关系的建模能力,又提升了模型的计算效率与泛化性能。

构建 BiLSTM-GRU 融合模型所涉及的超参数分为 4 项,分别是隐藏层数量、神经元个数、批量大小(batch size)、迭代次数(epochs)。但 BiLSTM 与 GRU 的层数比例需通过试验调整,因此,在训练时针对不同污染物进行了超参数调优,以保证后续预测效果,其参数设置如表 2 所示。

对于磺胺类抗生素(SMZ、SCP)数据集,BiLSTM 模型搭建为 2 层结构,每层均含有 32 个神经元节点,且都设置了 0.15 的随机失活(dropout)率,同时将 batch size 定为 85,epochs 定为 100。在该 BiLSTM 模型基础上,对 GRU 模型进行参数调整,将其构建为单层结构,神经元节点数保持 32 个,但将 dropout 率略微调高至 0.3, batch size 依旧为 85,不过训练轮次缩短至 85 轮。对于臭味物质(2-MIB、GSM)数据集,BiLSTM 模型设计为 3 层结构,各层神经元节点数依次为 128、128、64,对应的 dropout 率分别为 0.15、0.2、0.2, batch size 调整为 100, epochs 同样是 100。而 GRU 模型在此数据集上采用单层结构,神经元节点数为 32 个, dropout 率设为 0.2, batch size 为 100,训练轮次也是 100 轮。2 种模型在构建时,均选用 ReLU 作为激活函数,以 Adam 作为优化器来提升模型性能。

图 1 为模型在验证集上的学习曲线,随着迭代次数的增加,损失值(MSE)逐渐下降,模型的

表 2 BiLSTM-GRU 参数设定
Tab. 2 Parameters Setting of BiLSTM-GRU

污染物数据集	模型	层数	神经元个数	dropout 率	batch size	epoch
SMZ	BiLSTM	2	[32,32]	[0.15,0.15]	85	100
	GRU	1	32	0.3	85	100
SCP	BiLSTM	2	[32,32]	[0.15,0.15]	85	100
	GRU	1	32	0.3	85	100
GSM	BiLSTM	3	[128,128,64]	[0.15,0.2,0.2]	100	100
	GRU	1	32	0.2	100	100
2-MIB	BiLSTM	3	[128,128,64]	[0.15,0.2,0.2]	100	100
	GRU	1	32	0.2	100	100

预测精度不断提高。在训练初期,损失值较高,但迅速下降。这表明模型在初始阶段快速学习到了数据中的模式和特征。在经过大约 20 次迭代后,训练损失和测试损失都趋于平稳,稳定在 0.05 以下,说明

模型的学习逐渐趋于饱和。在整个训练过程中,训练损失和测试损失曲线非常接近,这表明模型在训练集和验证集上的表现相似,没有明显的过拟合现象。

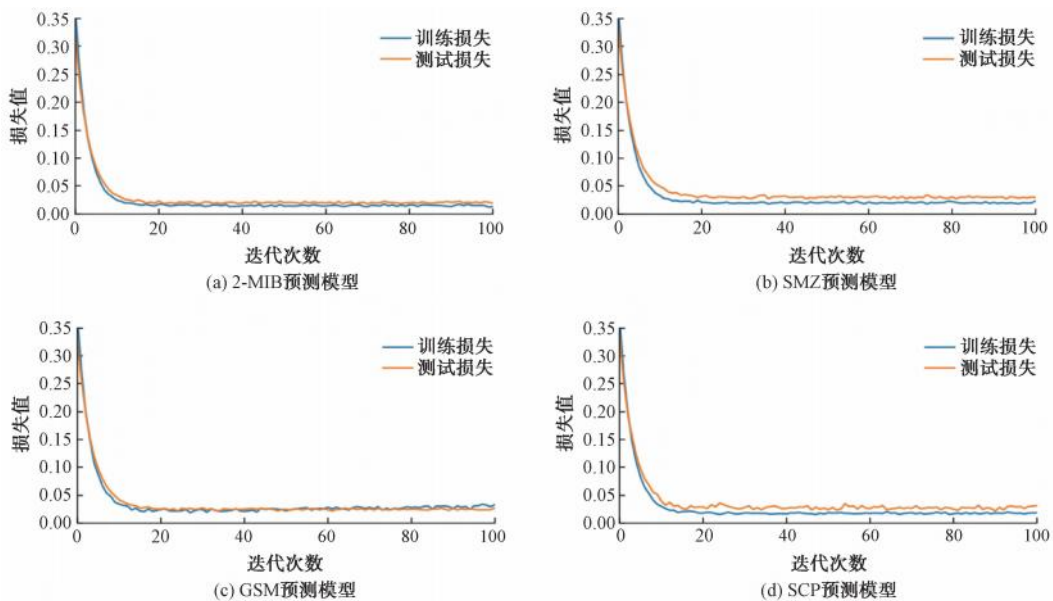


图 1 融合模型在验证集上的学习曲线图

Fig. 1 Learning Curve of Fusion Model on the Validation Set

1.3.2 模型评价指标

水中污染物预测是对未来水质的一种估计行为,任何预测方法都会存在一定的偏差,为了更有效地评估各种预测模型的预测性能,进而分析模型性能,减小预测误差,提高预测精度,下面介绍本文预测效果评价指标。

(1) 预测误差评价指标

1) 平均绝对误差 (MAE)

MAE 能够反映预测值与真实值之间的差异,值

为 $[0, +\infty]$, MAE 的值越小,表明模型的预测效果越好。

2) 均方根误差 (RMSE)

RMSE 能够从数据集整体层面体现预测效果,直观反映模型的综合预测精度。RMSE 表示的是期望值与实际值之间的距离,值为 $[0, +\infty]$, RMSE 的值越小,表明模型的预测效果越精准。

3) 决定系数 (R^2)

R^2 也称拟合优度,可用来衡量模型对数据的拟

合程度,取值为 0~1。值越接近 1,说明模型解释能力越强。

(2) 预测可靠性评价指标

使用 Origin 软件,对预测值和真实值进行了 95% 置信区间和 95% 预测区间的计算,并且进行了回归线性拟合线的绘制。

1) 95% 置信区间:使用置信区间来映射模型在未知数据的预测过程的不确定性,表示 95% 的预测点落在该区间内。置信区间越窄,模型预测越确定,预测能力越强,因为预测误差小。反之,置信区间越宽,预测越不确定,预测能力越弱。通常选择置信区间窄的模型作为更可靠的预测模型。

2) 95% 预测区间:预测区间表示在真实值下,95% 的预测点落在该区间内,反映模型预测值的不确定性范围。预测区间越窄,模型预测能力越强,预测值越准确可靠。但需注意,预测区间窄并不意味着预测准确性高,因为模型可能过拟合训练数

据,导致新数据预测表现不佳。

3) 回归线性拟合线:该拟合线利用最小二乘法的计算方法生成,常通过与正比例函数 $Y=X$ 的夹角来分析结果。正比例函数这条直线代表着预测值和真实值相同,预测无误差的情况,因此回归线性拟合线与正比例函数的角度越小,模型预测准确性越高。

2 结果与讨论

2.1 不同特征选择方法的特征差异

2.1.1 基于 Pearson 相关系数的特征选择

图 2(a)、图 2(b) 展示了 SCP 和 2-MIB 2 种新污染物数据集中 17 个参数的相关性分析热力图。当 r 为 -1 时,表示完全负相关; r 为 0 时,表示无线性关系; r 为 1 时,表示完全正相关。 $|r|$ 越接近 1,说明 2 个变量之间的线性关系越强。这种分析方法能够帮助我们识别参数之间的潜在关联性,从而为进一步的研究提供方向^[26]。

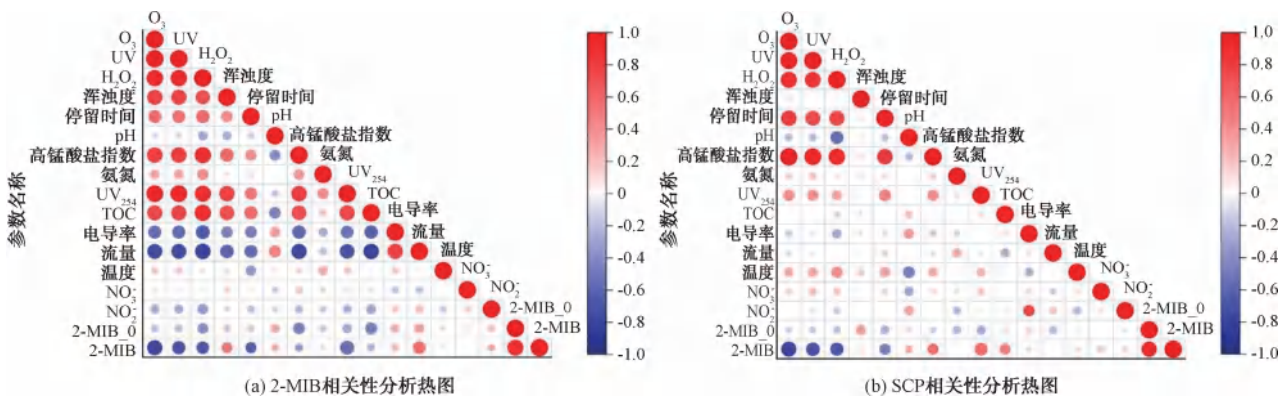


图 2 新污染物相关性分析热力图

Fig. 2 Correlation Analysis Heatmap of Novel Contaminants

在本文中,所有系统参数与出水 SCP 浓度和 2-MIB 浓度的 $|r|$ 均大于 0.02,表明所有特征值与出水水质指标存在一定的相关性。数据集中,与出水 SCP、2-MIB 浓度呈最大线性正相关的特征参数是其对应进水浓度, r 值约为 0.85;而与出水 SCP、2-MIB 浓度呈最大线性负相关的特征参数是 O_3 浓度, r 值分别为 -0.78 和 -0.82。最终经过相关系数法的选择,磺胺类抗生素数据集采用 O_3 、 H_2O_2 、UV 功率、停留时间、初始磺胺类抗生素浓度、 UV_{254} 、TOC、高锰酸盐指数 8 个特征参数作为输入,嗅味物质数据集采用 O_3 、 H_2O_2 、UV 功率、浑浊度、停留时间、初始嗅味物质浓度、 UV_{254} 、流量 8 个特征参数作为输入。

2.1.2 基于格兰杰因果检验的特征选择

通过 stata 数据分析软件对数据进行平稳性检验,检验结果如表 3 所示。

对参数估计值进行平稳性检验。如果 p 值小于 0.05,则说明原始假设不成立,意味着序列 X 是平稳序列。否则,即表明序列有单位根。若序列存在单位根则需要进一步进行趋势检验。由表 3 可知,在 5% 显著水平下,所有历史数据均通过了平稳性检验,可以对各因素之间进行格兰杰因果检验。

为直观验证时间序列数据的平稳性,绘制关键水质参数及目标污染物浓度的时间序列折线图如图 3 所示,通过趋势性与波动性分析辅助判断数据平稳性。选取浑浊度、氨氮、高锰酸盐指数、SMZ 浓度

表 3 ADF 平稳性检验表
Tab. 3 Table of ADF Stationarity Test

数据名称	ADF 检验统计量	临界值统计量		通过平稳性检验	p 值
		1%	5%		
温度	-10.849	-4.024	-3.461	√	0.000***
pH	-5.631	-4.024	-3.461	√	0.000***
浑浊度	-10.424	-4.024	-3.461	√	0.000***
TOC	-6.030	-4.024	-3.461	√	0.000***
高锰酸盐指数	-11.361	-4.024	-3.461	√	0.000***
UV ₂₅₄	-19.659	-4.024	-3.461	√	0.000***
电导率	-8.226	-4.024	-3.461	√	0.000***
氨氮	-10.648	-4.024	-3.461	√	0.000***
NO ₃ ⁻	-7.440	-4.024	-3.461	√	0.000***
NO ₂ ⁻	-6.746	-4.024	-3.461	√	0.000***
SCP_in	-10.116	-4.024	-3.461	√	0.000***
GSM_in	-6.610	-4.024	-3.461	√	0.000***
2-MIB_in	-14.040	-4.024	-3.461	√	0.000***
SMZ_in	-5.113	-4.024	-3.461	√	0.000***
O ₃	-13.281	-4.024	-3.461	√	0.000***
H ₂ O ₂	-11.368	-4.024	-3.461	√	0.000***
UV	-12.330	-4.024	-3.461	√	0.000***
停留时间	-4.336	-4.024	-3.461	√	0.000***
流量	-2.669	-4.024	-3.461	√	0.000***
SCP_out	-11.423	-4.024	-3.461	√	0.000***
GSM_out	-8.116	-4.024	-3.461	√	0.000***
2-MIB_out	-18.068	-4.024	-3.461	√	0.000***
SMZ_out	-7.737	-4.024	-3.461	√	0.000***

注:统计学中,星号是显著性水平的常用标记,*** 对应 1%、** 对应 5%、* 对应 10%,论文中所有数据的 p 值均小于 0.01,故统一标注***,明确序列平稳性检验的严格性与可靠性。

4 项核心指标。浑浊度、氨氮、高锰酸盐指数、SMZ 浓度虽因工艺调控和水质波动出现短期突变,但整体围绕均值上下波动,无长期趋势性变化,且波动幅度在合理工艺范围内,结合表 3 的 ADF 平稳性检验结果,所有指标的 ADF 检验统计量均小于 1% 临界值,且时间序列折线图无明显趋势性与周期性漂移,双重验证了数据的平稳性,为格兰杰因果检验的准确性提供可视化支撑。

平稳性检验已经证明了各个特征均为平稳状态,因此可直接进行 VAR 模型的构建。由于中试装

置的水力停留时间为 5 min, O₃ 与 H₂O₂ 协同氧化反应的关键反应周期为 10~15 min, 对应 2 h 1 组的数据采集粒度, 反应达到稳定状态需覆盖 3~5 个数据采集周期; 同时新污染物降解存在时滞效应, 如 O₃ 投加后需经过 2~3 个反应周期才能完全体现降解效果, 结合数据粒度换算, 需至少 6 个滞后阶数才能捕捉完整的时滞关联。因此, 本文选取 6 作为最大滞后阶数, 通过 varsoc 命令确定最佳滞后期。下面以 SMZ 与各参数的格兰杰因果关系结果进行展示, 检验结果如图 4 所示。

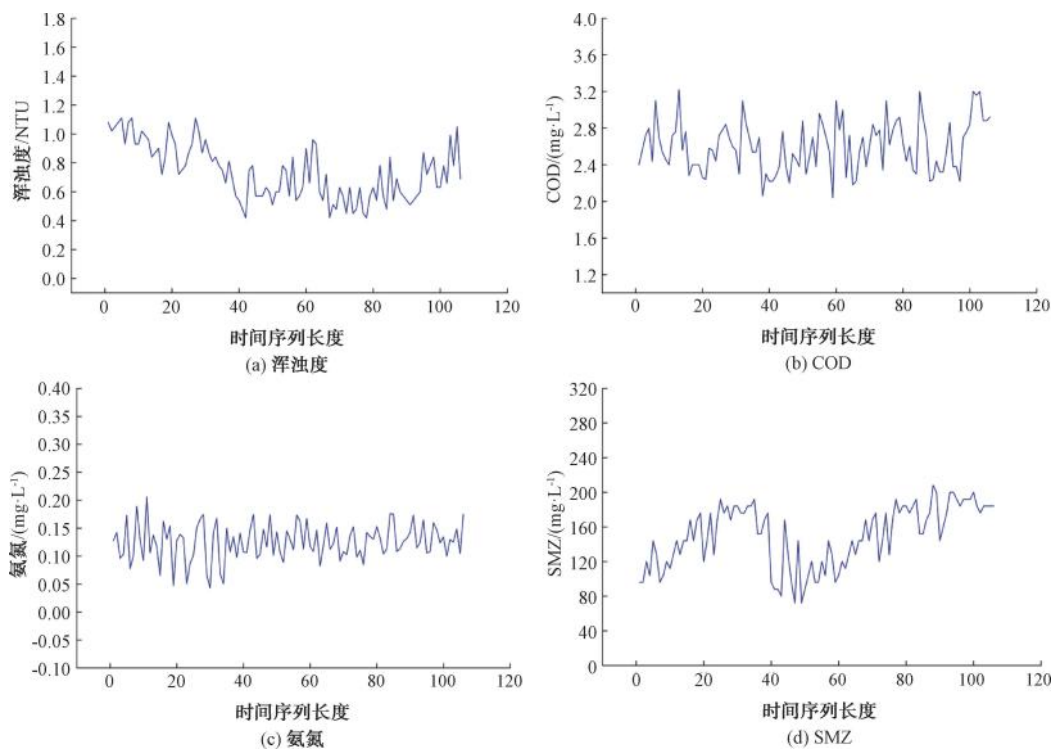


图3 关键水质参数及目标污染物浓度数据集

Fig. 3 Key Water Quality Parameters and Target Pollutants Concentration Dataset

由图4可知,对检验所有参数所有根均位于圆内,表明VAR模型通过了平稳性检验,模型有效且稳定性良好,可用于后续各因素对SMZ序列的格兰杰因果检验。

结果如表4所示,各因素与SMZ的因果检验中,在5%的临界显著水平内的因素通过检验^[27],然而,水质参数间的因果关联具有隐蔽性与弱相关性,尤其是新污染物降解过程中,部分关键影响因素与目标污染物浓度的关联并非强线性,而是呈现弱因果、强影响的特点。若采用5%严格显著性水平,易遗漏此类重要因果特征。相关研究^[28-29]证实,在水质时序数据特征选择中,由于水质参数交互复杂、时滞效应显著,将格兰杰检验显著性水平放宽至10%可有效捕捉弱因果关联,避免关键信息丢失;该做法在低浓度污染物预测、多参数交互作用显著的水处理场景中已成为常规操作,与本文中试验数据特性高度契合。因此,本文将显著水平提升到10%,以保证不错过相应的特征^[29]。最终经过格兰杰因果检验特征选择后,磺胺类抗生素确定最终选定特征为O₃投加量、H₂O₂投加量、UV功率、氨氮、UV₂₅₄、浊度、高锰

酸盐指数、pH、停留时间、TOC、初始磺胺类抗生素浓度、电导率。即这12个特征存在相关因素的输出SMZ序列的格兰杰因果关系,这些因素也可能有助于磺胺类抗生素的预测。

嗅味物质的格兰杰因果检验结果如表5所示,经过格兰杰因果检验特征选择后,嗅味物质确定最终特征为O₃、H₂O₂、UV功率、流量、UV₂₅₄、浊度、高锰酸盐指数、pH、停留时间、TOC、初始嗅味浓度、电导率这12个因素。

为进一步明确格兰杰因果检验筛选的12项特征的合理性,结合给水处理化学原理补充分析。以电导率对SMZ预测的因果贡献为例,SMZ作为极性化合物,其降解速率受水中离子强度调控。高电导率会促进O₃分解生成次氯酸等次生氧化剂,增强SMZ的脱氨基反应;而低电导率下O₃更易直接攻击SMZ的苯环结构,2种路径的降解速率差异通过电导率的时序变化体现,因此电导率成为SMZ浓度预测的重要因果特征。同理,浊度通过影响UV光利用效率和污染物吸附解吸平衡,其动态变化与2-MIB的光催化降解效果直接相关,故被纳入特征集。

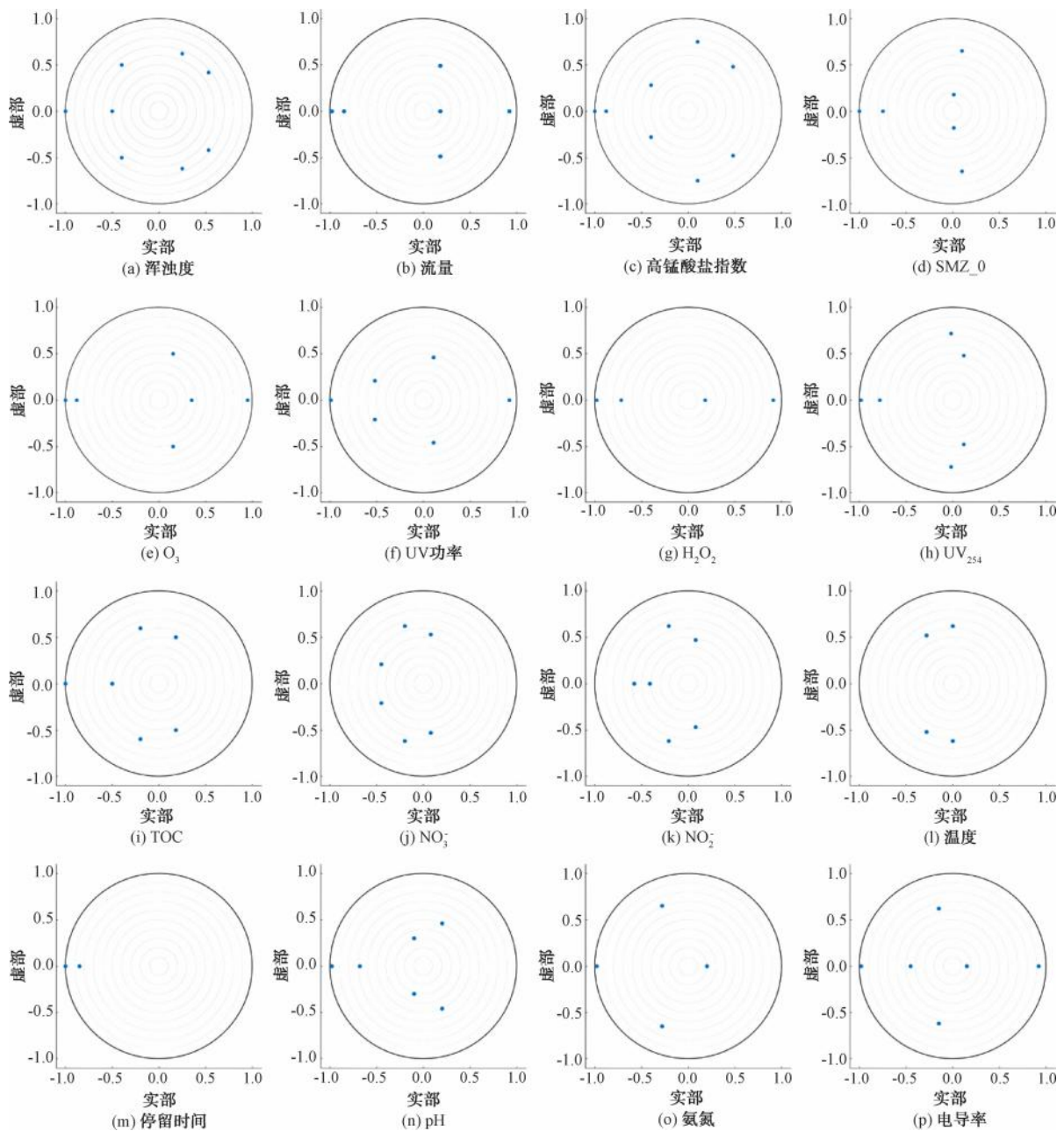


图 4 模型稳定性检验结果

Fig. 4 Results of Model Stability Test

2.2 基于不同特征选择方法的结果对比

为验证特征选择的必要性以及方法的差异对模型预测的影响,试验对比了全特征输入和两种特征选择方法后的输入对预测效果的影响,利用 BiLSTM-GRU 神经网络建立 3 种特征集的预测模型,其中模型 A 采用全特征输入,模型 B 使用 Pearson 相关系数法筛选过后的特征作为输入,模型 C 使用格兰杰因果检验选择的特征作为输入。针对 2-MIB、GSM、SCP、SMZ 建立单目标预测模型,比较

每种目标特征的预测值与真实值,误差图如图 5 所示。

通过对比图 5 中各单目标预测模型对典型嗅味物质和磺胺类抗生素的预测表现,可以发现,经过格兰杰因果检验法特征选择后的 BiLSTM-GRU 模型具有最窄的置信区间和预测区间,表明其预测精度最高,泛化能力最强。此外,该模型的预测值与真实值的线性拟合线与正比例函数的偏移角度最小,说明预测值与真实值的相关性最强,误差最小。相比之

表 4 各因素与 SMZ 的格兰杰因果检验

Tab. 4 Granger Causality Test between Various Factors and SMZ

假设	F 统计量	滞后阶数	p 值	通过格兰杰因果检验
温度	7.13	2	0.068	×
pH	8.34	3	0.039	√
浑浊度	4.53	4	0.034	√
TOC	2.02	3	0.087	√
高锰酸盐指数	14.06	4	0.070	√
UV ₂₅₄	1.74	3	0.099	√
电导率	5.13	3	0.062	√
氨氮	0.39	3	0.082	√
NO ₂ ⁻	3.05	3	0.217	×
NO ₃ ⁻	1.67	3	0.644	×
SCP ₀	6.44	3	0.100	√
SMZ ₀	5.91	3	0.026	×
O ₃	7.47	3	0.048	√
H ₂ O ₂	7.34	2	0.026	√
UV 功率	8.34	3	0.039	√
停留时间	0.37	1	0.055	√
流量	3.98	3	0.263	×

表 5 各因素与 2-MIB 的格兰杰因果检验

Tab. 5 Granger Causality Test between Various Factors and 2-MIB

假设	F 统计量	滞后阶数	p 值	通过格兰杰因果检验
浑浊度	5.03	4	0.036	√
UV ₂₅₄	1.66	3	0.095	√
TOC	2.72	3	0.068	√
氨氮	0.47	3	0.824	×
高锰酸盐指数	11.18	4	0.078	√
NO ₃ ⁻	1.38	3	0.644	×
NO ₂ ⁻	3.18	3	0.217	×
电导率	5.66	3	0.089	√
温度	6.83	2	0.713	×
pH	7.93	3	0.046	√
O ₃	9.73	3	0.026	√
H ₂ O ₂	6.11	2	0.046	√
UV 功率	9.35	3	0.041	√
停留时间	0.46	1	0.066	√
GSM ₀	8.83	3	0.293	×
2-MIB ₀	7.93	3	0.078	√
流量	3.37	3	0.069	√

下,使用相关系数法选择特征的模型置信区间和预测区间变化较小,这可能是由于水质数据的相关性较为隐蔽,而 Pearson 相关系数法仅能揭示表面关系,未能为模型提供有效帮助。因此,格兰杰因果检验法对单目标预测模型提升效果更加明显,这表明通过合适的特征选择方法能够有效应对高级氧化反应器处理臭味物质和磺胺类抗生素时,各参数间更强的非线性作用关系,通过降低高维特征对模型的学习难度,提高模型泛化能力。

不同特征选择方法在预测出水 2-MIB 和 GSM 时的误差分析如表 6 所示。BiLSTM-GRU 模型在应用相关系数法和格兰杰因果检验法预测 GSM 后,相较于全特征输入,展现了更优的预测性能,模型 B 经过 r 选择后的 MAE、RMSE 值分别低至 1.96 和 3.06, R^2 值达到 0.911,较全特征输入提高 7.68%。模型 C 经过因果检验法选择后 MAE、RMSE 值分别低至 1.84 和 2.93, R^2 值达到 0.935,较全特征输入提高 8.22%。在 2-MIB 预测中,模型 B 和 C 的 R^2 值分别达到 0.896 和 0.941,较未进行特征选择提高 2.40% 和 7.54%,且 MAE、RMSE 等误差指标均低于未进行特征选择的预测结果。相比之下,相关系数特征选择法表现欠佳,原因是相关系数法不能捕捉水质数据隐藏的内部关联。经过试验验证发现格兰杰因果检验法通过筛选相关特征显著提升了预测的准确度。同臭味物质试验结果类似,特征选择使模型对出水 SMZ 和 SCP 的预测展现出更优的效果。其中格兰杰因果检验法效果最优,SMZ 和 SCP 的 R^2 值分别达到 0.893 和 0.916,对比全特征输入提高 13.18% 和 11.98%,对比相关系数法提高 6.18% 和 6.14%,经过因果检验法选择后 MAE、RMSE 值也均为最低,MAE 分别低至 1.38 和 1.22, RMSE 值分别低至 2.10 和 1.36。

综合 3 种特征选择方法的对比表明,特征选择显著提升了 BiLSTM-GRU 模型的学习能力和预测精度与稳定性,其中格兰杰因果检验法展现了明显的优势。进一步从机制层面分析。

格兰杰因果检验筛选的特征之所以能有效减少预测误差,核心在于其精准捕捉了新污染物降解过程中的动态关联规律:一方面,因果特征剔除了 NO₃⁻、温度等与污染物去除无直接因果关系的冗余参数,降低了模型学习噪声,使网络聚焦于 O₃ 投加量、UV 功率等关键调控因子,减少了无关特征对预

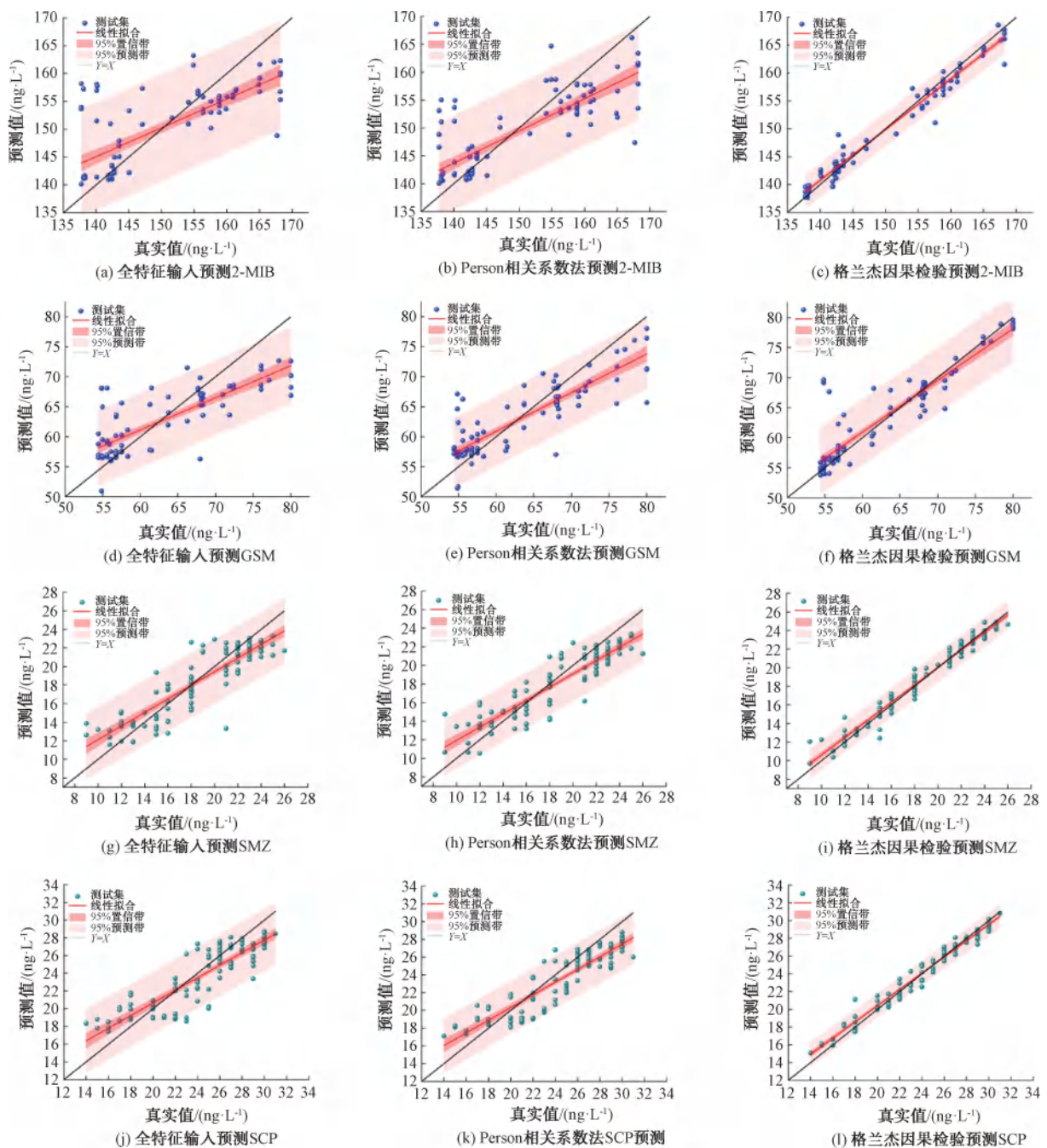


图5 典型新污染物预测值及置信区间分布

Fig. 5 Predicted Values and Confidence Interval Distributions of Typical Novel Pollutants

测结果的干扰;另一方面,因果特征挖掘了参数间的时滞关联,避免了静态特征选择对动态过程的误判,使模型能更准确地拟合污染物浓度的时序变化。

这种优势对低浓度污染物预测尤为关键。低浓度新污染物的浓度信号易被水质背景噪声掩盖,因果特征通过聚焦核心反应参数,能有效放大污染物与工艺参数的关联信号,降低噪声对低浓度信号的

干扰;低浓度污染物的降解过程对工艺参数波动更敏感,因果特征所捕捉的参数间因果关系,能更精准地量化这种敏感关联,避免因特征选择不当导致的预测偏差;低浓度污染物在高级氧化过程中易受多参数交互影响,因果特征通过 VAR 模型捕捉的多变量动态关联,能更全面地表征这种交互作用,提升模型对复杂反应过程的拟合能力。相比之下,全特征

表 6 特征选择典型新污染物评价指标

Tab. 6 Evaluation Indicators of Typical Novel Pollutants for Feature Selection

预测目标	预测模型	MAE	RMSE	R^2
2-MIB	模型 A	3.19	3.92	0.875
	模型 B	3.04	3.63	0.896
	模型 C	2.97	2.91	0.941
GSM	模型 A	2.01	3.11	0.864
	模型 B	1.96	3.06	0.911
	模型 C	1.84	2.93	0.935
SMZ	模型 A	1.59	2.10	0.789
	模型 B	1.43	2.03	0.841
	模型 C	1.38	2.10	0.893
SCP	模型 A	1.42	1.85	0.818
	模型 B	1.39	1.47	0.863
	模型 C	1.22	1.36	0.916

输入因冗余特征引入的噪声的干扰、Pearson 相关系数法因线性假设无法捕捉的非线性因果关联,均会在低浓度污染物预测中被放大,导致预测精度下降。Pearson 相关系数法表现欠佳的核心局限性源于对水质数据复杂关联特性的适配不足。新污染物降解与工艺参数的关联存在明显时间滞后,而 Pearson 相关系数仅计算变量间的静态线性相关性,无法捕捉这种动态时滞因果关系,对时滞关联不敏感,导致关键时序关联特征被遗漏。高级氧化工艺中, O_3 、

H_2O_2 、UV 功率存在协同活化效应,这种多变量交互对新污染物降解的影响是非线性的,Pearson 相关系数仅能量化变量间的线性关系,无法表征多参数协同或拮抗的复合效应,导致筛选的 8 项特征仅覆盖单一参数关联,遗漏了交互关联带来的预测信息。新污染物浓度低,易受 TOC、浑浊度波动等水质背景噪声干扰,信噪比低,Pearson 相关系数对数据噪声敏感,其线性关联计算易被噪声扭曲。

为进一步明确不同特征选择策略在实际应用中的性价比,表 7 展示了计算成本的对比分析。全特征输入预处理耗时最短,仅为 1.2 min,但因冗余特征干扰,模型训练时间最长,且预测精度最低;Pearson 相关系数法通过简单线性筛选减少冗余特征,使训练时间缩短 21.8%,平均 R^2 提升 5.41%,总计算时间与全特征输入接近,性价比中等;格兰杰因果检验法预处理耗时最长,主要源于 ADF 平稳性检验、VAR 模型构建及稳定性验证等流程,但通过精准筛选因果特征,模型训练时间缩短 33.1%,且平均 R^2 提升 9.72%,4 类污染物 R^2 均突破 0.89,其中 2-MIB 和 GSM 的 R^2 超过 0.93,精度提升显著;从性价比来看,格兰杰因果检验法增加了 3.3 min 计算时间换取了翻倍的精度提升,且在实际动态调控场景中,预处理过程可离线完成,模型训练时间的缩短更有利于实时响应,因此其计算成本的增加具有实际应用价值。

表 7 不同特征选择策略的计算成本与精度对比

Tab. 7 Comparison of Computational Cost and Accuracy among Different Feature Selection Solutions

特征选择策略	预处理耗时/min	模型训练时间/min	总计算时间/min	平均 R^2 提升(较全特征输入)
全特征输入	1.2	12.4	13.6	—
Pearson 相关系数法	3.5	9.7	13.2	5.41%
格兰杰因果检验法	8.6	8.3	16.9	9.72%

3 结论

本文针对给水处理中典型新污染物浓度预测的核心痛点,聚焦特征选择这一关键环节,以 BiLSTM-GRU 融合模型为载体,系统开展全特征输入、Pearson 相关系数法、格兰杰因果检验法的对比研究,结合中试实测数据验证了不同策略的适用性与优越性,主要结论如下。

(1) 特征选择对新污染物预测模型性能提升具有决定性作用。全特征输入因包含 NO_3^- 、温度等低关联冗余特征,不仅增加模型计算负荷,还引入噪声

干扰,导致预测精度与稳定性最差,4 类污染物 R^2 值均低于 0.88,验证了冗余特征剔除的必要性。

(2) 格兰杰因果检验法是适配济南某中试基地典型新污染物(磺胺类抗生素 SMZ、SCP;嗅味物质 2-MIB、GSM)预测的较优特征选择策略。该方法通过 ADF 平稳性检验、VAR 模型构建等规范流程,有效挖掘了工艺参数与新污染物浓度间的潜在因果关系,筛选的 12 项特征既规避了冗余信息,又完整保留了非线性时滞关联,使 BiLSTM-GRU 模型的预测误差最小、拟合度最高,其中 GSM 与 2-MIB 的 R^2 值

均超过 0.93, SMZ 与 SCP 的 R^2 值接近 0.90, 且 95% 置信区间最窄, 泛化能力相对优于其他 2 种方法。

(3) 不同特征选择方法的适配性差异源于对参数关联特性的捕捉能力。全特征输入因冗余信息干扰导致模型学习噪声增加, 预测精度受限; Pearson 相关系数法仅能捕捉线性关联, 无法适配新污染物降解过程中的非线性、时滞性关联, 特征筛选有效性不足; 而格兰杰因果检验法虽预处理耗时增加, 但通过精准挖掘时序数据中的因果关联与多变量交互效应, 使模型训练时间显著缩短, 且平均 R^2 提升 9.72%, 计算成本的增加换取了高性价比的精度提升, 更适用于实际动态调控场景。该方法弥补了线性特征选择方法的局限, 为中试规模低浓度、高波动新污染物的精准预测提供了可靠特征支撑。

参考文献

- [1] Li F F, Chen L J, Chen W D, et al. Antibiotics in coastal water and sediments of the East China Sea: Distribution, ecological risk assessment and indicators screening [J]. *Marine Pollution Bulletin*, 2020, 151: 110810. DOI: 10.1016/j.marpolbul.2019.110810.
- [2] Luo Y Y, Liu C, Wang Y, et al. Occurrence, distribution and their correlation with different parameters of antibiotics and antibiotic resistance genes in lakes of China: A review [J]. *Marine Pollution Bulletin*, 2023, 193: 115189. DOI: 10.1016/j.marpolbul.2023.115189.
- [3] Afsa S, Hamden K, Martin P A L, et al. Occurrence of 40 pharmaceutically active compounds in hospital and urban wastewaters and their contribution to Mahdia coastal seawater contamination [J]. *Environmental Science and Pollution Research*, 2020, 27(2): 1941–1955.
- [4] Li K C, Shooter D. Analysis of sulfur-containing compounds in ambient air using solid-phase microextraction and gas chromatography with pulsed flame photometric detection [J]. *International Journal of Environmental Analytical Chemistry*, 2004, 84(10): 749–760.
- [5] 肖向群. UV/O₃/H₂O₂ 深度处理炼化废水的效果与作用机制研究 [D]. 北京: 中国石油大学(北京), 2022.
Xiao X Q. Study on the effect and mechanism of advanced treatment of refinery wastewater by UV/O₃/H₂O₂ [D]. Beijing: China University of Petroleum (Beijing), 2022.
- [6] Luo W, Deng L, Hu J, et al. Degradation of metronidazole during the UV₃₆₅-LED/chlorine process: Kinetics, influencing factors, and halonitromethanes formation [J]. *Journal of Environmental Chemical Engineering*, 2023, 11(5): 110524. DOI: 10.1016/j.jece.2023.110524.
- [7] DiPietro R, Hager G D. Handbook of medical image computing and computer assisted intervention: Deep learning: RNNs and LSTM [M/OL]. Amsterdam: Elsevier, 2020: 503–519. <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>.
- [8] Aslam M, Kim J S, Jung J. Multi-step ahead wind power forecasting based on dual-attention mechanism [J]. *Energy Reports*, 2023, 9: 239–251. DOI: 10.1016/j.egyr.2022.11.167.
- [9] Siami-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series [C] // 2019 IEEE International Conference on Big Data (Big Data). December 9–12, 2019. Los Angeles, CA, USA. IEEE, 2019: 3285–3292.
- [10] Hauduc H, Neumann M B, Muschalla D, et al. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment [J]. *Environmental Modelling & Software*, 2015, 68: 196–204. DOI: 10.1016/j.envsoft.2015.02.004.
- [11] Gupta H V, Kling H, Yilmaz K K, et al. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling [J]. *Journal of Hydrology*, 2009, 377(1/2): 80–91.
- [12] Knoben W J M, Freer J E, Woods R A. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores [J]. *Hydrology and Earth System Sciences*, 2019, 23(10): 4323–4331.
- [13] Willmott C J, Robeson S M, Matsuura K. A refined index of model performance [J]. *International Journal of Climatology*, 2012, 32(13): 2088–2094.
- [14] Graf R. A multifaceted analysis of the relationship between daily temperature of river water and air [J]. *Acta Geophysica*, 2019, 67(3): 905–920.
- [15] Zolghadr-Asli B, Enayati M, Pourghasemi H R, et al. Application of Granger-causality to study the climate change impacts on depletion patterns of inland water bodies [J]. *Hydrological Sciences Journal*, 2021, 66(12): 1767–1776.
- [16] Detto M, Molini A, Katul G, et al. Causality and persistence in ecological systems: A nonparametric spectral granger causality approach [J]. *The American Naturalist*, 2012, 179(4): 524–535.
- [17] Singh D, Singh B. Investigating the impact of data normalization on classification performance [J]. *Applied Soft Computing*, 2020, 97: 105524. DOI: 10.1016/j.asoc.2019.105524.
- [18] Agyeman J K, Ameyaw B, Li Y, et al. Modeling the long-run drivers of total renewable energy consumption: Evidence from top five heavily polluted countries [J]. *Journal of Cleaner Production*, 2020, 277: 123292. DOI: 10.1016/j.jclepro.2020.123292.

(下转第 64 页)

- Li C J. Research on key technologies of heterogeneous knowledge graph embedding based on federated learning[D]. Changchun: Changchun University of Technology, 2025.
- [11] 王依科, 张东月. 基于共享潜在嵌入矩阵的联邦知识图谱嵌入[J]. 计算机技术与发展, 2025, 35(11): 154-160.
Wang Y K, Zhang D Y. Federated knowledge graph embedding based on latent embedding matrix sharing [J]. Computer Technology and Development, 2025, 35(11): 154-160.
- [12] Siriwardhana S, Weerasekera R, Wen E, et al. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering[J]. Transactions of the Association for Computational Linguistics, 2023, 11: 1-17. DOI: 10.1162/tacl_a_00530.
- [13] Edge D, Trinh H, Cheng N, et al. From local to global: A graph RAG approach to query-focused summarization [PP/OL]. V2. arXiv (2025-02-19) [2025-10-28]. <https://doi.org/10.48550/arXiv.2404.16130>.
- [14] Sun Z Q, Deng Z H, Nie J Y, et al. RotatE: Knowledge graph embedding by relational rotation in complex space [PP/OL]. V1. arXiv (2019-02-26) [2025-10-28]. <https://doi.org/10.48550/arXiv.1902.10197>.
- [15] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data: Proceedings of the 27th International Conference on Neural Information Processing Systems, Nevada, December 5-10, 2013 [C]. United States: Curran Associates Inc., 2013.
- [16] Yang B S, Yih W T, He X D, et al. Embedding entities and relations for learning and inference in knowledge bases [PP/OL]. V4. arXiv (2015-08-29) [2025-10-28]. <https://doi.org/10.48550/arXiv.1412.6575>.
- [17] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction: Proceedings of the 33th International Conference on Machine Learning, New York, June 19-24, 2016 [C]. United States: PMLR, 2016: 2071-2080.

(上接第 37 页)

- [19] Taoufik N, Boumya W, Achak M, et al. The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning [J]. Science of the Total Environment, 2022, 807: 150554. DOI: 10.1016/j.scitotenv.2021.150554.
- [20] Zounemat-Kermani M, Matta E, Cominola A, et al. Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects [J]. Journal of Hydrology, 2020, 588: 125085. DOI: 10.1016/j.jhydrol.2020.125085.
- [21] Smirnov D A, Mokhov I I. From Granger causality to long-term causality: Application to climatic data [J]. Physical Review E, 2009, 80: 016208. DOI: 10.1103/PhysRevE.80.016208.
- [22] Kodra E, Chatterjee S, Ganguly A R. Exploring Granger causality between global average observed time series of carbon dioxide and temperature [J]. Theoretical and Applied Climatology, 2011, 104(3/4): 325-335.
- [23] McGraw M C, Barnes E A. Memory matters: A case for granger causality in climate variability studies [J]. Journal of Climate, 2018, 31(8): 3289-3300.
- [24] 朱得康. 新型格兰杰因果检验法的研究及其应用 [D]. 青岛: 山东科技大学, 2020.
Zhu D K. Research and application of the new granger causality test method [D]. Qingdao: Shandong University of Science and Technology, 2020.
- [25] Tiwari A K, Adewuyi A O, Adeleke M A, et al. A time-varying Granger causality analysis between water stock and green stocks using novel approaches [J]. Energy Economics, 2023, 126: 107010. DOI: 10.1016/j.eneco.2023.107010.
- [26] Liu P, Wang J, Sangaiah A K, et al. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment [J]. Sustainability, 2019, 11(7): 2058. DOI: 10.3390/su11072058.
- [27] Ali Shah S, Ai S T, Rack W. Interconnections, trend analysis and forecasting of water-air temperature with water level dynamics in Blue Moon Lake Valley: A statistical and machine learning approach [J]. Journal of Environmental Management, 2025, 379: 124829. DOI: 10.1016/j.jenvman.2025.124829.
- [28] Tiwari A K, Adewuyi A O, Adeleke M A, et al. A time-varying Granger causality analysis between water stock and green stocks using novel approaches [J]. Energy Economics, 2023, 126: 107010. DOI: 10.1016/j.eneco.2023.107010.
- [29] 吕智慧. 格兰杰因果关系的检验方法及其应用 [D]. 长春: 东北师范大学, 2020.
Lü Z H. Granger causality tests and their applications [D]. Changchun: Northeast Normal University, 2020.