DOI:10.11918/202512179

# 基于多头自注意力机制模型的再生水厂泥线预测

高梦泽[1,2,3],郁达伟[1,2,3,4],郑利兵[1,2,3],魏源送[1,2,3]

(1. 区域环境安全全国重点实验室环境水质学重点实验室(中国科学院生态环境研究中心),北京 100085;
2. 水污染控制实验室(中国科学院生态环境研究中心),北京 100085; 3. 中国科学院大学 资源与环境学院,
北京 100085; 4. 国家黄河流域生态保护和高质量发展联合研究中心,北京 100012)

**摘 要**:再生水厂泥线在实际工程中存在参数敏感性高、漂移非线性、时滞性大等问题,给泥线的运行调度带来较大的不确定性,对减污降碳协同能力形成制约。传统机理模型预测需要大量、高质量且标注规范的输入数据,难以对泥线参数进行准确、高效的预测。为增强再生水厂泥线的减污降碳协同,以北京某再生水厂水线、泥线运行数据为基础,采用经验模型基线-残差预测的混合建模策略,构建基于多头自注意力机制(Transformer)的预测模型,先通过经验模型线性拟合生成基线,再利用Transformer模型对残差进行精准预测,以水线进出水水质指标和运行参数为输入,同步预测脱水污泥产量和污泥含水率,并与LightGBM、块状循环神经网络BlockRNN(LSTM)、时域卷积网络(TCN)模型进行性能对比。结果表明:依托多头自注意力机制的Transformer模型的全局时序关联捕捉能力,可有效适配泥线运行参数的非线性波动特征,综合性能最优,脱水污泥产量预测 $R^2$ 达0.778 8,污泥含水率预测 $R^2$ 达0.679,均显著高于各对比模型,同时回测机制可提升模型的运行稳定性,持续优化模型预测精度,可适配实际工程中长时序预测场景。该模型可实现泥线关键参数的高精度预测,为再生水厂泥线工艺实时调控、药剂精准投加及设备智能调度等场景提供可靠技术支撑,助力泥线减污降碳协同治理。

**关键词**:再生水厂;泥线;多头自注意力;预测模型;运行优化
**中图分类号**:X703 **文献标志码**:A **文章编号**:0367-6234(2026)06-0090-11

# Prediction of sludge line in wastewater reclamation plant based on multi-head self-attention mechanism model

GAO Mengze[1,2,3], YU Dawei[1,2,3,4], ZHENG Libing[1,2,3], WEI Yuansong[1,2,3]

(1. State Key Laboratory of Regional Environment and Sustainability (Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences), Beijing 100085, China; 2. Laboratory of Water Pollution Control (Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences), Beijing 100085, China; 3. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100085, China; 4. National Joint Research Center for Ecological Conservation and High Quality Development of the Yellow River Basin, Beijing 100012, China)

**Abstract**: In practical engineering, the sludge line of wastewater reclamation plants faces challenges including high parameter sensitivity, drift nonlinearity, and significant time lag, which introduce considerable uncertainty into the operation and scheduling of the sludge line and restrict the synergistic capacity of pollution and carbon reduction. Conventional mechanism-driven models require a large amount of high-quality and well-annotated input data, making it difficult to predict sludge line parameters accurately and efficiently. To enhance the synergy of pollution and carbon reduction for the sludge line in wastewater reclamation plants, this paper established a prediction model based on the multi-head self-attention mechanism (Transformer) using a hybrid modeling strategy of "empirical model baseline and residual prediction", based on the operational data of the water line and sludge line of a wastewater reclamation plant in Beijing. Specifically, this paper first generated a baseline through linear fitting of an empirical model and then utilized the Transformer model to accurately predict the residual. Taking the inlet and outlet water quality indexes and operating parameters of the water line as inputs, the model simultaneously predicted the dewatered sludge output and sludge moisture content and compared its performance with LightGBM, block recurrent neural network BlockRNN (LSTM) and temporal convolutional network (TCN) models. The results indicate that relying on the global temporal correlation capture capability of the Transformer model based on multi-

**收稿日期**:2025-12-31;**录用日期**:2026-03-23;**网络首发日期**:2026-04-16
**网络首发地址**:https://link.cnki.net/urlid/23.1235.T.20260415.1453.002
**基金项目**:国家重点研发计划(2022YFC3203102);黄河流域生态保护和高质量发展联合研究项目
**作者简介**:高梦泽(2002—),男,硕士研究生;魏源送(1969—),男,研究员,博士生导师
**通信作者**:郁达伟,dwyu@rcees.ac.cn

head self-attention mechanism, it can effectively adapt to the nonlinear fluctuation characteristics of sludge line operating parameters and achieve the optimal comprehensive performance: $R^2$ of dewatered sludge output prediction reaches 0.778 8, and that of sludge moisture content prediction reaches 0.679, both of which are significantly higher than those of the comparative models. Meanwhile, the backtesting mechanism can improve the operational stability of the model, continuously enhance its prediction accuracy, and thus adapt to the long-term time series prediction scenarios in practical engineering. This model can realize high-precision prediction of key parameters in the sludge line, provide reliable technical support for scenarios such as real-time regulation of the sludge line process, precise chemical dosing, and intelligent equipment scheduling in wastewater reclamation plants, and facilitate the synergistic governance of pollution and carbon reduction for the sludge line.

**Keywords**: wastewater reclamation plant; sludge line; Multi-Head Self-Attention; prediction model; operation optimization

再生水厂泥线的机理建模存在诸多难点，限制了其在实际工程中的应用和减污降碳协同能力。再生水厂泥线的污泥在厌氧和脱水处理过程中，具有高含水率(75% ~99%)、高有机物含量和生物活性等特性，包含复杂的微生物功能、有机无机污染组成和气-液-固多相相互作用，整体性状和作用机制非常复杂，导致泥线关键参数(如污泥产量)难以预测。由于泥线的复杂性，机理模型在实际工程中的应用存在参数敏感性高、漂移非线性、时滞性大等问题，给泥线的运行调度带来较大的不确定性，最终影响泥线集中处理处置设施的高效运行调度，导致污泥处理与处置环节的成本有时甚至超过水处理与再生环节。因此，亟需解决再生水厂泥线关键运行参数的预测问题。

泥线预测模型涵盖经验模型、机理驱动型模型、数据驱动型模型以及混合驱动型模型等。其中，经验模型以预测污泥产量的表观污泥产率经验模型作为典型，机理模型可以分为单方程动力学模型，如预测强化厌氧消化甲烷产量的 Gompertz 模型，以及多方程动力学模型等[1]，如国际水协的厌氧消化模型(anaerobic digestion model, ADM)[2]、理化模型(physical chemistry model, PCM)等特定体系的反应动力学过程专用模型[3]，也包括流体动力学(computational fluid dynamic, CFD)模型、黏弹性模型(viscoelasticity model)等更为通用模型的应用。ADM 模型能够较为准确地预测污泥的甲烷产量[4]，并被用于不同系统的对比优化设计[5]。然而，相较于已成熟应用于智能优化调控及数字孪生在线控制的水线机理模型(activated sludge models, ASMs)，泥线机理模型在实际工程中的应用较为有限[2]。

模型预测调控的关键在于预测模型的稳定性和准确度。泥线模型通过对其复杂过程进行量化和可视化，以预测污泥产生量以及最终产物特性等，从而为泥线的运行调度、集中管理提供科学、可靠的决策依据。污泥处理过程属于复杂的多相非线性系统，水动力学、反应动力学等过程强耦合，存在显著的不确定性和系统整体建模的精准预测难度。随着对污泥微生物生态机理认识的飞速发展，长期以来 ADM 机理模型都是研究的热点，但其在实际工程中的应用仍存在参数敏感性高、数据需求量大、漂移非线性等问题，限制了 ADM 机理模型的广泛应用[2]。机理模型预测依赖大量、高质量和标准规范的输入数据，而泥线通常缺乏连续或系统的监测数据[1]，因此，泥线运行的预测难度极高。泥线运行系统规模庞大、动态性强，运行过程易受多种外界因素干扰，机理模型对现场运维人员的检测能力、在线监测系统的维护水平提出了较高要求。

数据驱动模型发展迅速，逐渐被应用到泥线预测的实践中。长短期记忆递归神经网络(Long Short-Term Memory, LSTM)、随机森林(Random Forest, RF)和极端梯度提升(eXtreme Gradient Boosting, XGBoost)为当前的主流模型[6-7]，如贝叶斯优化的 XGBoost 模型针对脱水 AOP 药剂投加预测与优化，提升了脱水效率，SHAP 识别的主控因子为药剂和 pH[7]。神经网络可处理复杂的多变量、多层级数据并挖掘全局规律，但存在解释性差、对样本量依赖性强等局限，通常仅适用于“大数据驱动的纯预测”场景[8]。混合驱动模型是实际应用推理中解释性较强的方向，如集成模型精准预测碳源投加量与剩余污泥量，确保节约成本和运行可扩展性的应用目的[9]。但此类模型透明度降低，对模型解释能力的影响仍需进一步探究。

强化学习特别适用于自适应动态过程(如多工艺点协同)，已逐渐成为未来 AI 工艺控制领域的热点，如强化学习模型联控 7 项关键工艺参数，运行指标综合提升 8.3%，显著降低运行和感知成本[10-11]，值得大型工艺智能控制领域的关注探索，有望实现模型的滚动优化和持续进化。在实际应用中，需要开发同步进行多步预测的方法，以应对泥线多个指标的预测。注意力机制模型已经被证实能够

对 COD、$NH_3$ 等水线水质指标进行精确预测[12]，并能够与机理模型结合[13]，在增强模型解释性的基础上进一步提升模型的稳定性。目前，基于注意力机制模型泥线参数的多模型预测研究较少。

针对再生水厂泥线关键运行参数的预测现状与实际需求，依托经验模型预测基线、机器学习模型预测残差的混合建模策略，构建了基于多头自注意力机制的污泥产量与含水率预测模型。该模型能够学习水线运行状态数据决定泥线数据所隐含的全局规律，以进出水水质和运行参数等水线数据作为输入，预测泥线数据，提高泥线预测稳定性和准确性。此外，进一步用当日实际泥线关键运行参数滚动优化，考察了模型实际运行时的持续进化能力，并比较了多头自注意力模型与各经典模型的性能。模型将多头自注意力机制引入再生水厂泥线运行，可以对机理模型难以预测的指标进行有效预测。通过运用多目标模型，以较高的鲁棒性和较小的误差实现泥线指标参数的精准预测。模型在应用层面上可为实现泥线智能优化运行的预测调控提供支持，为污水处理厂的工艺调度以及智能投药等减污降碳协同治理提供技术支持。

## 1 材料与方法

### 1.1 再生水厂工艺和数据采集

研究所用数据来自北京某再生水厂，承接周边生活污水以及高校校园等园区的合流废水。该再生水厂水线采用厌氧-缺氧-好氧的 $A^2O$ 工艺，二沉出水经过脱氮滤池深度处理并消毒后，供城市河湖景观环境用水的生态补水，出水执行 DB11/890—2012《北京市城镇污水处理厂水污染物排放标准》B 标准。也有部分出水需经超滤膜深度处理并消毒后，作为市政杂用等再生水。由于进出水水量水质边界条件变化大，显著限制经典机理模型的预测稳定性和实际场景应用。泥线由贮泥池调蓄，剩余污泥经离心脱水、干化后外运处置。泥线的池容空间有限，采用以机械处理为主的技术路线，进一步限制了其动态适应能力，对泥量、污泥含水率等泥线运行数据的需求尤为迫切。

本研究在该再生水厂采集了水线和泥线相关数据，在进水口测量了进水流量 $Q_{inf}$（$km^3$）和水质指标 $BOD_{5inf}$、$SS_{inf}$、$COD_{inf}$、$NH_4-N_{inf}$、$TN_{inf}$、$TP_{inf}$；在出水口测量了 $COD_{eff}$、$NH_4-N_{eff}$、$TN_{eff}$、$TP_{eff}$、$pH_{eff}$，采用统一的单位 mg/L（除 pH）。在泥线工艺运行中，测量了脱水污泥产量 Sludge（t）和含水率 MC（%）。

### 1.2 数据清洗和预处理

数据预处理过程包括变量划分、缺失与非数值处理、数据平滑与降噪、时间序列排序与标准化，并按照数据相关性进行特征筛选。协变量是指可能影响目标序列但本身非预测目标的外部时间序列数据。在模型中，获取有效利用特征是提升预测精度的关键手段。根据预测目标，将数据变量划分为协变量和目标序列，其中，目标序列为脱水污泥产量 Sludge（t）和含水率（%），其余变量均为协变量。将数据中非数值字符串替换为缺失值（NaN），随后，将所有列尝试转换为浮点型（float）。

应用 2 d 窗口滚动平均的方法，以平滑日频噪声，即

$$\bar{x}_t = \frac{1}{N_t}\sum_{\tau\in W_t} x_\tau \tag{1}$$

式中：$W_t$ 为时间窗口，$N_t$ 为窗口内的有效数据点数，$x_\tau$ 为时间点 $\tau$ 对应的原始数据值，$\bar{x}_t$ 为时间点 $t$ 对应的滚动平均结果，$t$ 表示需要计算滑动平均值的目标时间点，$\tau$ 表示落入时间窗口内的原始数据时间点。

最后，按时间升序排序，设置时间索引并对目标序列与协变量数据进行 Min-Max 归一化，从而在训练和预测过程中进行尺度一致性约束，即

$$X_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

式中：$x_t$ 为时间 $t$ 的值，$x_{max}$ 和 $x_{min}$ 分别为变量的最大值和最小值，$X_t$ 为 $x_t$ 的归一化值。

建模前，对特征采用斯皮尔曼等级相关性分析方法，对水质核心指标进行两两相关性检验。为排除随机误差和多重检验假阳性，对显著性 $P$ 值采用 FDR 法校正，仅将校正后 $P<0.05$ 的相关性系数判定为具有统计学意义。分析前通过线性插值填充时序数据缺失值，确保数据集无缺失、无异常干扰。

为进一步优化模型并保证模型训练的公平性，在模型预测中优先选取对目标变量（Sludge、MC）影响显著的特征，剔除相关性极弱的无关特征。同时，为避免模型因强共线性特征出现参数估计不稳定的问题，需对强相关特征进行针对性筛选。

### 1.3 建模方法

Transformer 将自注意力机制引入了时间序列预测，采用编码器-解码器结构，每部分都由多层堆叠的相同模块构成，其核心是多头自注意力机制（Multi-Head Self-Attention），允许模型在处理每个时间点时同时对输入序列中所有其他时间点的重要性进行计算与权衡[14]。因此，Transformer 模型非常适合捕捉复杂的、非局部的长期依赖关系，其高度并行的特性使其在 GPU 上训练效率很高。然而，其通常需要大量的训练数据才能发挥最佳性能，并且对超

参数较为敏感。

1.3.1　多头自注意力

自注意力机制是 Transformer 的核心组件。自注意机制允许模型在处理序列时，动态地为每个位置分配不同的权重，从而捕捉序列中任意两个位置之间的依赖关系。输入序列中的每个词通过词嵌入（Embedding）转换为向量表示，进而通过计算查询（Query）、键（Key）和值（Value）之间的点积，得到每个词与其他词的相关性权重。最后，使用注意力权重对值（Value）进行加权求和，即

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \quad (3)$$

式中：$\boldsymbol{Q}$ 为查询矩阵；$\boldsymbol{K}$ 为键矩阵；$\boldsymbol{V}$ 为值矩阵；$d_k$ 为向量的维度，用于缩放点积，防止梯度爆炸。

1.3.2　编码器–解码器架构

编码器由 NN 层相同的模块堆叠而成，每层包含两个子层：多头自注意力机制（Multi-Head Self-Attention）计算输入序列中每个词与其他词的相关性；前馈神经网络（Feedforward Neural Network）对每个词进行独立的非线性变换。解码器根据编码器的输出生成目标序列。每个解码器层包含两个注意力机制（自注意力和编码器–解码器注意力）和一个前馈神经网络。为了稳定训练过程，每个子层（如自注意力层和前馈神经网络）后面都会接一个残差连接（Residual Connection）和层归一化（Layer Normalization）（见图 1）。

## 1.4　模型构建与评估

预测模型的输入参数包括进水的流量 $Q_{\text{inf}}$（$\text{km}^3$）和水质指标 $\text{COD}_{\text{inf}}$、$\text{TN}_{\text{inf}}$、$\text{TP}_{\text{inf}}$，出水水质指标 $\text{COD}_{\text{eff}}$、$\text{NH}_4\text{-N}_{\text{eff}}$、$\text{TN}_{\text{eff}}$、$\text{TP}_{\text{eff}}$，7 项水质指标统一单位为 mg/L。数据时间序列范围为 2013—2014 年全年，总样本量为 730 个，按照时间节点分为训练集和测试集，基于时间点将 2014 年 9 月 1 日之前的数据划为训练集，之后的数据为测试集，训练集和验证集比例为 5∶1。为提升模型预测质量，将上述指标作为特征变量，以污泥产量（Sludge）、污泥含水率（MC）为目标变量，构建经验模型进行线性拟合生成基线，再采用优化后的机器学习/深度学习模型对残差进行精准预测，并将经验模型的基线拟合值与机器学习模型的残差预测值相加，得到最终的综合预测结果。深度模型基于 PyTorch 训练，采用反向传播算法对均方误差（MSE）损失函数进行最小化优化。Transformer 模型特征维度 d_model = 64，4 头注意力，3 层编码器与 3 层解码器，前馈网络维度 512，使用 ReLU 激活，优化器为 Adam，其他相关参数如表 1 所示。Transformer 模型框架默认通过正弦–余弦位置编码（Sinusoidal Positional Encoding）为输入序列的每个时间步注入绝对位置信息。模型输入设定为 40 个时间步长，该长度覆盖了污水处理系统主要工艺环节的最大时延窗口，进一步强化位置编码对强时延性的适配性。将多头自注意力机制模型与 LightGBM 模型、块状循环神经网络（Block Recurrent Neural Network，BlockRNN（LSTM））模型、时域卷积网络（Temporal Convolutional Network，TCN）进行比较，以验证自注意力模型在泥线指标预测中的性能优势。为防止深度学习模型发生过拟合，所有深度学习模型均引入早停机制和 Dropout 正则化参数。

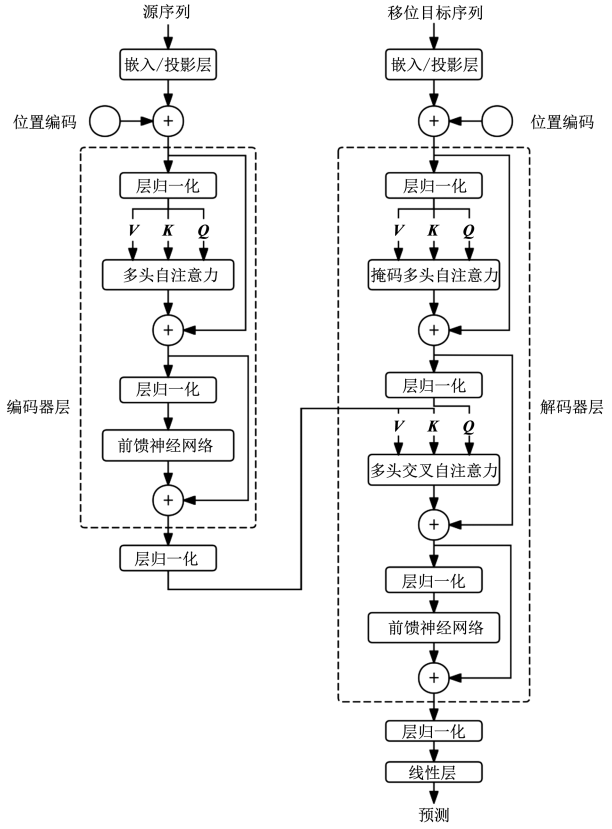



**图 1　多头自注意力机制模型示意**

Fig. 1　Schematic diagram of multi-head self-attention mechanism model

**表 1　模型详细参数**

Tab. 1　Detailed parameters of models

| 参数 | 参数解释 | 数值 |
|---|---|---|
| input_chunk_length | 深度学习模型输入的时间步长 | 40 |
| output_chunk_length | 预测的时间步长 | 10 |
| n_epochs | 深度学习模型的训练轮数 | 1 000 |
| lags | 传统机器学习模型的核心滞后阶数 | 90 |
| lags_past_covariates | 传统机器学习模型中，过去协变量的滞后阶数 | 40 |

回测(backtest)是通过移动时间窗口模拟历史预测过程的分析方法,通过模型在历史数据上做“滚动式预测”,按时间推进多次生成预测并与真实历史值比较,用于评估模型在真实时间序列上的泛化能力。预测(forecast)为模型的实际应用形式,即基于训练完成的模型直接预测时间序列的后续值,对应实际应用场景中的未来预测场景。所有模型在归一化尺度上训练与回测,预测结果在评估前通过 target_scaler. inverse_transform 还原到原始尺度。回测采用滚动时间窗口法,与直接预测的结果进行对比,并分别统计各模型的性能指标。

为量化预测性能,使用以下常用误差指标:决定系数($R^2$)、均方根误差(RMSE,$E_{\mathrm{RMS}}$)、平均绝对误差(MAE,$E_{\mathrm{MA}}$)和平均绝对百分比误差(MAPE,$E_{\mathrm{MAP}}$)。具体公式如下:

$$R^2 = 1 - \frac{\sum_{m=1}^{M}(y_m - \hat{y}_m)^2}{\sum_{m=1}^{M}(y_m - \bar{y})^2} \tag{4}$$

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y_m - \hat{y}_m)^2} \tag{5}$$

$$E_{\mathrm{MA}} = \frac{1}{M}\sum_{m=1}^{M}|y_m - \hat{y}_m| \tag{6}$$

$$E_{\mathrm{MAP}} = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{y_m - \hat{y}_m}{y_m}\right| \times 100\% \tag{7}$$

式中:$m = 1, 2, \cdots, M$ 为样本点,$\hat{y}_m$ 和 $y_m$ 分别表示模型预测值和 $m$ 处的实际值,$\bar{y}$ 表示所有样本真实值的平均值。

## 2 结果与讨论

### 2.1 数据相关性分析

对再生水厂水线和泥线数据进行相关性分析,结果如图 2 所示。总体来看,图中协变量内部存在显著的高相关集群,表明数据集存在多重共线性问题。而协变量与目标变量的相关性特征呈现差异化规律。

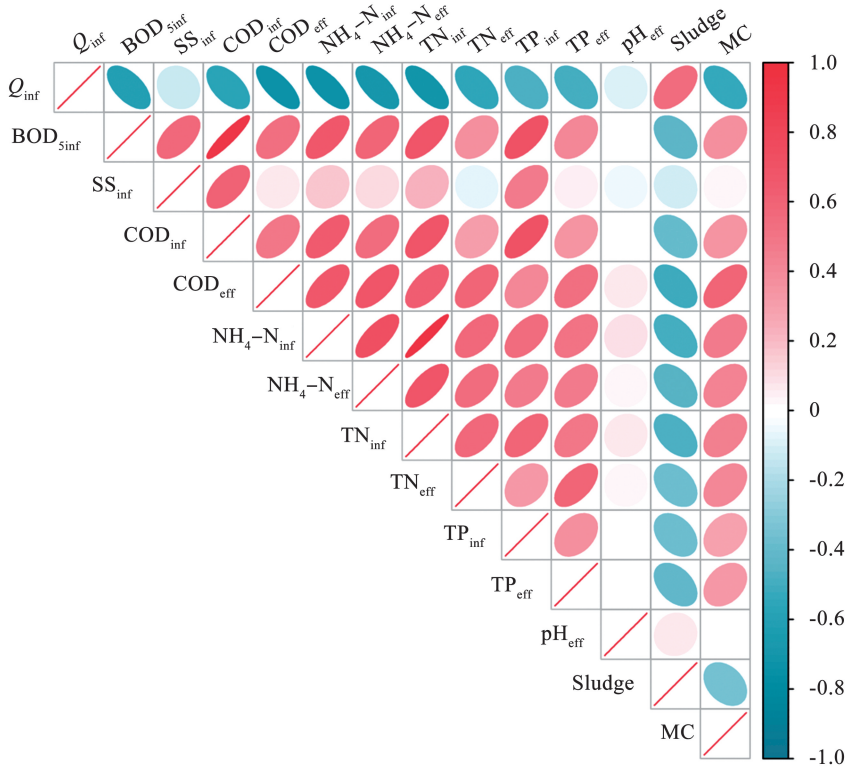



**图 2 再生水厂水线与泥线参数相关性热力图**

Fig. 2 Correlation heatmap of water line and sludge line parameters in wastewater reclamation plant

进水量 $Q_{\mathrm{inf}}$作为核心调控变量,对水线、泥线指标均存在一定影响,与进水污染物质量浓度($BOD_{5inf}$、$COD_{inf}$、$NH_4$–$N_{inf}$、$TN_{inf}$)呈负相关(相关系数范围 −0.735 ~ −0.136),体现水力稀释作用;与出水污染物质量浓度亦呈负相关,说明高流量低浓度进水更易实现达标排放。进水数据中,$BOD_{5inf}$与 $COD_{inf}$呈强正相关(0.928),表明进水可生化性稳定;$NH_4$–$N_{inf}$与 $TN_{inf}$相关性达 0.959,说明进水总氮变化主要由氨氮主导。悬浮物 $SS_{inf}$与除 COD、BOD 外的其他所有协变量相关性均较弱,波动相对独立,

这是因为 SS 中包含大量悬浮态有机物碎屑，这类悬浮有机物同时贡献了 $SS_{inf}$ 的质量浓度和 COD、BOD 的质量浓度。

出水水质类协变量间同样存在显著的相关性特征，其中，进水碳氮类变量（$BOD_{5inf}$、$COD_{inf}$、$NH_4$-$N_{inf}$、$TN_{inf}$）与对应出水水质变量（$COD_{eff}$、$NH_4$-$N_{eff}$、$TN_{eff}$）相关性较强，$TP_{inf}$ 与 $TP_{eff}$ 呈中等正相关（0.363）。出水 pH 协变量（$pH_{eff}$）与所有协变量的相关系数均接近 0，属于相对独立的协变量。

此外，协变量与目标变量相关性之间存在差异化规律，这为后续进行特征选择提供了方向。$Q_{inf}$ 同时与 Sludge 呈正相关（0.548）、与 MC 呈负相关（-0.536），表明进水量越大，系统产泥量越多且污泥含水率越低，体现水力协变量对泥线目标变量的调控作用。而除了 $pH_{eff}$、$SS_{inf}$ 外的其他协变量均与目标变量存在一定相关性，说明水质类协变量与目标变量整体呈相关关系，进一步印证协变量对目标变量的综合调控作用。

协变量之间相关性过高，即“多重共线性”会给模型训练和预测带来多方面的负面影响，如模型参数估计不稳定、特征重要性失真。在后续处理中应对其进行特征选择或降维等方式处理，减少冗余。

通过对数据相关性的研究，可以得到协变量内部相关性以及协变量-目标相关性。通过特征处理可以帮助在建模时优化相应参数，提高模型稳定性与可解释性。多头自注意力模型会根据协变量的相关性自适应分配变量权重，并借助多组注意力头充分挖掘变量间的复杂关联。考虑到在后续处理中更好地提升模型，在模型预测时优先选取进水量、碳氮指标等对目标变量（Sludge、MC）影响显著的协变量，剔除 $pH_{eff}$、$SS_{inf}$ 等相关性极弱的无关协变量，同时对 $BOD_{5inf}$ 与 $COD_{inf}$、$NH_4$-$N_{inf}$ 与 $TN_{inf}$ 两对强相关变量进行特征筛选，消除冗余特征，简化模型结构，降低建模复杂度。

### 2.2　污泥产量预测

不同模型对脱水污泥产量 Sludge（t）的预测效果见图 3 和表 2。污泥产率呈现季节性波动特征[15]，这也增加了污泥产量指标的预测难度。在此背景下，Transformer 模型在预测和回测场景中的综合性能均优于其他对比模型。

在回测场景中，因为通过滚动窗口进行模型优化，各模型性能差异相对较小。Transformer 模型拟合效果最优，$R^2$ 达 0.737 5，同时各项误差指标均最低，图 3 中模型拟合曲线表明其对历史脱水污泥产量数据的线性拟合度最高。LightGBM 模型 $R^2$ 为 0.720 9，误差指标与 Transformer 接近，符合其对结构化数据的适配特性。BlockRNN（LSTM）模型的 $R^2$ 为 0.631 4，说明其对历史时序数据的拟合精度有限。TCN 模型拟合效果最差，$R^2$ 仅为 0.431 6，各项误差指标均最高，可推测其对历史脱水污泥产量变化规律的表征能力不足，拟合误差增大。

在预测场景下，各模型的预测性能呈现显著分化。Transformer 模型仍表现出最优的预测性能，$R^2$ 超过 0.7，且各项误差指标均为最低，表明其能有效捕捉 Sludge 的时序变化规律，泛化能力极强。LightGBM 模型泛化能力稳健，$R^2$（0.710 3）与回测场景（0.720 9）差距较小，误差指标虽略有上升但整体波动不大，结合其模型特性说明其对脱水污泥产量指标的非线性关联拟合与表征稳定，具备一定的实际应用价值。BlockRNN（LSTM）与 TCN 模型的预测性能则出现显著衰减，$R^2$ 相对回测场景分别降低 69% 和 21%。BlockRNN（LSTM）模型的 $R^2$ 从回测的 0.631 4 降至 0.196 8，表明其对脱水污泥产量指标的长期时序变化规律捕捉能力薄弱，易出现预测偏差，且无法适配同一时间段内的残差波动，实际应用价值有限。TCN 模型 $R^2$（0.339 4）虽高于 BlockRNN（LSTM），但仍低于 0.5，说明其无法有效适配脱水污泥产量的时序波动特征，难以满足实际预测需求。

从误差特征来看，进一步验证了各模型的性能差异。RMSE 对异常值更敏感，BlockRNN（LSTM）和 TCN 模型在预测场景的 RMSE 呈大幅上升趋势（分别升至 44.653 3、38.027 2），而 Transformer 和 LightGBM 模型的 RMSE 在回测与预测场景波动较小。结合图 3 对脱水污泥产量实测值与预测值的对比可知，相比 BlockRNN（LSTM）和 TCN 模型，Transformer 和 LightGBM 模型对脱水污泥产量的异常波动具有更强的适配性，预测精度显著更高，这与图 4 中呈现的拟合离散程度一致。

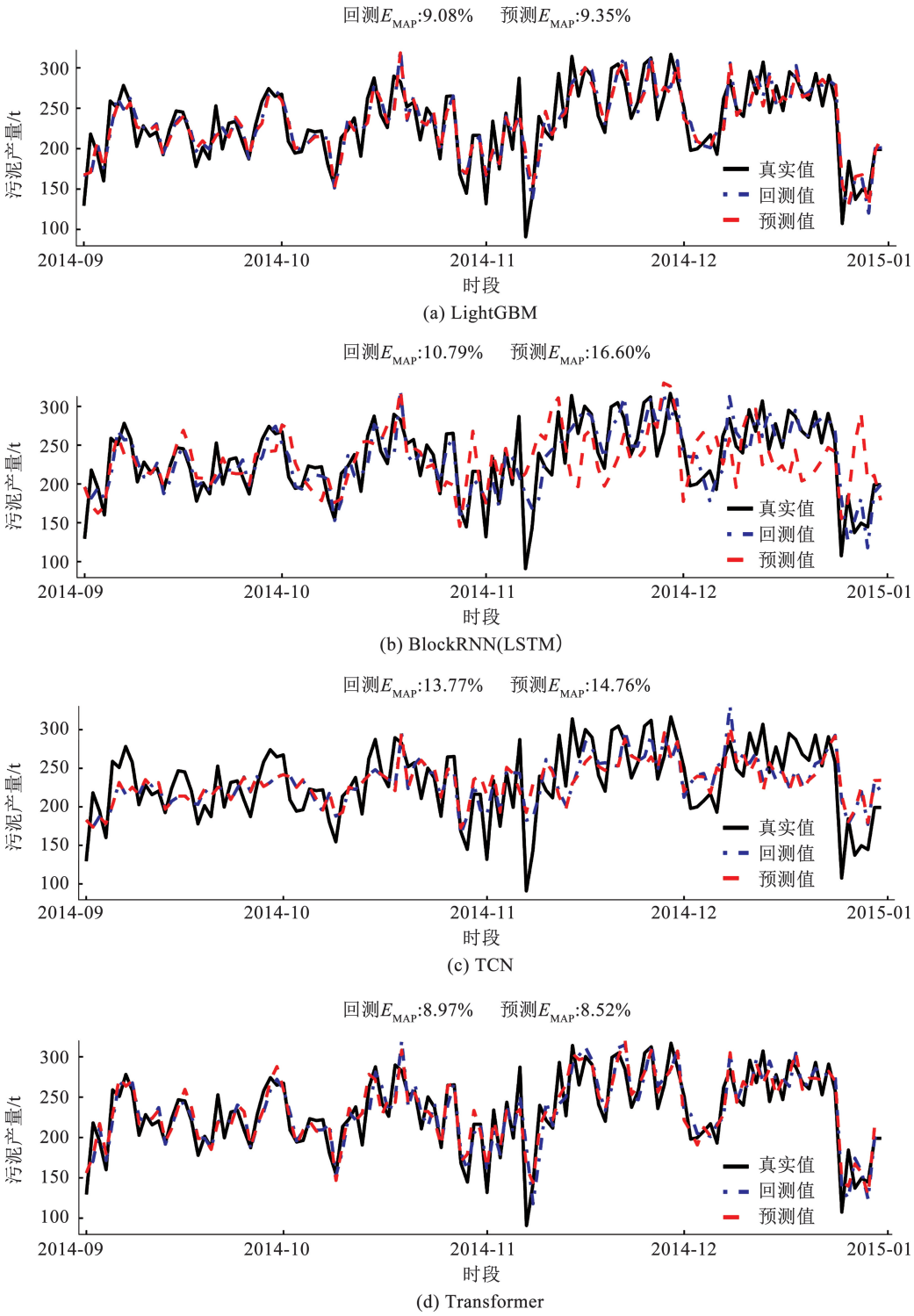



**图 3 再生水厂脱水污泥产量的模型预测效果对比**

Fig. 3 Comparison of model prediction effects for dewatered sludge output in wastewater reclamation plant

**表 2 模型脱水污泥产量预测性能参数**

Tab. 2 Prediction performance parameters of models for dewatered sludge output

| 模型 | $R^2$ | | $E_{MA}$ | | $E_{RMS}$ | | $E_{MAP}$/% | |
|---|---|---|---|---|---|---|---|---|
| | 回测 | 预测 | 回测 | 预测 | 回测 | 预测 | 回测 | 预测 |
| LightGBM | 0.720 9 | 0.710 3 | 19.193 3 | 19.634 6 | 24.783 8 | 25.181 6 | 9.081 6 | 9.348 6 |
| BlockRNN（LSTM） | 0.631 4 | 0.196 8 | 22.584 2 | 34.794 9 | 28.689 5 | 44.653 3 | 10.794 5 | 16.604 2 |
| TCN | 0.431 6 | 0.339 4 | 28.718 7 | 30.611 6 | 35.284 0 | 38.027 2 | 13.766 2 | 14.757 8 |
| Transformer | 0.737 5 | 0.778 8 | 19.054 4 | 18.140 9 | 24.249 1 | 22.096 4 | 8.966 0 | 8.517 9 |

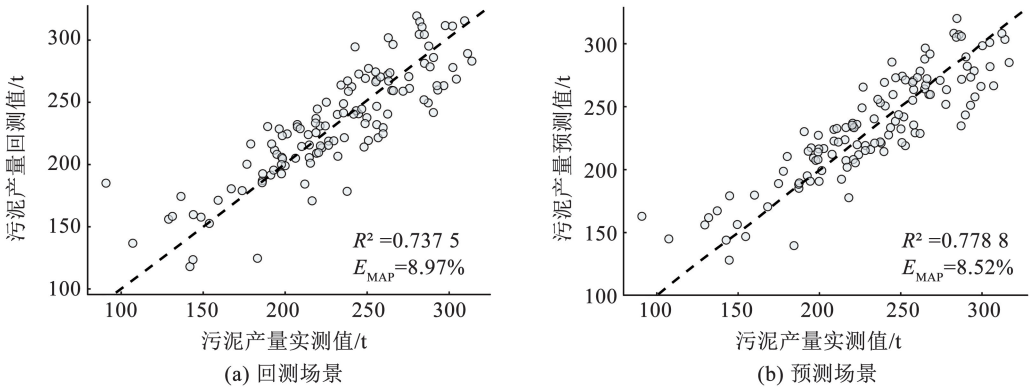



**图 4　Transformer 模型预测再生水厂脱水污泥产量拟合效果**

Fig. 4　Fitting effect of Transformer model predicting dewatered sludge output in wastewater reclamation plant

Transformer 的多头自注意力机制可以建立多变量全局关联，兼顾数据强弱相关特性，在直接预测时有效提取长时序依赖特征，通过注意力权重的动态调整特性适配未来数据的分布变化，实现了“历史拟合精度-未来外推泛化”的均衡。LightGBM 模型是基于决策树的集成模型，依赖“局部特征阈值划分”学习数据规律，因此，当污泥产量出现阶梯式上升时，模型能快速捕捉强关联特征的阈值变化，对突变的响应速度快，但其对时序依赖特征的提取能力存在局限性，需手动设计滞后特征，难以自动学习长时序依赖。而 LSTM 模型的门控机制能平滑序列波动，预测曲线更稳定，但当实测值骤升时模型需要多步数据积累才能更新记忆，导致对数据波动的响应滞后。TCN 模型因卷积窗口的局部性特征，无法适配未来数据的时序分布变化，当脱水污泥产量实测值出现骤升等异常波动时，模型无法捕捉这种全局时序变化趋势，仅能依赖局部历史模式推测，导致预测波动加剧，误差大幅升高。

### 2.3　污泥含水率预测

污泥含水率预测效果见图 5 和表 3。污泥含水率 MC 实测值相较于脱水污泥产量无明显大幅度波动，整体呈平稳变化趋势，但仍存在一定的季节性特征，且呈现夏季低、冬季高的季节特征[16]。从含水率预测整体性能排序来看，与脱水污泥产量预测趋势一致，Transformer 和 LightGBM 模型性能优于 BlockRNN(LSTM)和 TCN 模型。但不同的是，含水率预测的整体 $R^2$ 均低于脱水污泥产量预测。其核心原因在于含水率波动幅度小，对模型输出的细微误差更为敏感，进一步提升了预测难度，这与时序图呈现的含水率预测拟合离散程度略高于脱水污泥产量的特征一致。

由模型预测效果可视化分析可知，LightGBM 和 Transformer 模型的预测和回测曲线均与预测时段内的实测值高度贴合，与图 6 中呈现的拟合离散程度一致，而 BlockRNN（LSTM）和 TCN 模型的拟合效果最差，其中，TCN 对小幅度波动的响应存在滞后性。TCN 依赖固定大小的卷积窗口捕捉局部时序模式，难以建立全局时序关联，而含水率属于季节性长期波动，要求模型不仅关联前期数据的潜在规律，还需适配后期小幅波动变化。卷积窗口仅能聚焦短期局部数据，无法捕捉该类跨时段的小幅度波动规律，导致波动响应滞后。Transformer 的多头注意力，实现了全局趋势拟合与局部波动响应的双重适配，可精准契合含水率季节性长期趋势与短期小幅波动的复合特征，拟合效果更优。这也印证了深度学习模型中非循环、全局关联建模结构对弱规律、随机波动参数的适配优势。

综合来看，Transformer 模型在泥线参数预测中，尤其是未来数据外推场景下，性能显著优于传统模型与其他深度学习模型。其优势源于多头自注意力机制对多变量时序关联的捕捉能力，能更精准适配泥线参数的非线性、波动性特征，为实际工程中的工艺调度提供更可靠的预测支持。

### 2.4　模型应用

污水处理厂泥线参数（脱水污泥量、含水率）与水线指标存在复杂的非线性关联。相比 LSTM 的单维度序列建模、TCN 的局部卷积特征提取，基于多头自注意力机制模型的特性，可实现对再生水厂泥线核心参数的高精度多参数协同预测，为泥线系统的智能优化运行提供关键的数据支撑。多头自注意力机制能够同时捕捉泥线各参数之间的长短期依赖关系与复杂耦合关联，可以挖掘污泥产生量与进水水质、处理负荷的时序非线性关联特征，突破了传统单参数预测模型的局限性。污泥相关指标本身具有复杂性，如污泥含水率受人为操作因素影响显著，导致其预测难度相对更高。后续基于该模型优化工艺运行参数，可使泥线实际运行指标与模型预测值的拟合度显著提升；同时，通过滚动窗口的动态预测，可对模型进行实时修正，持续提升模型的稳定性与预测精度。

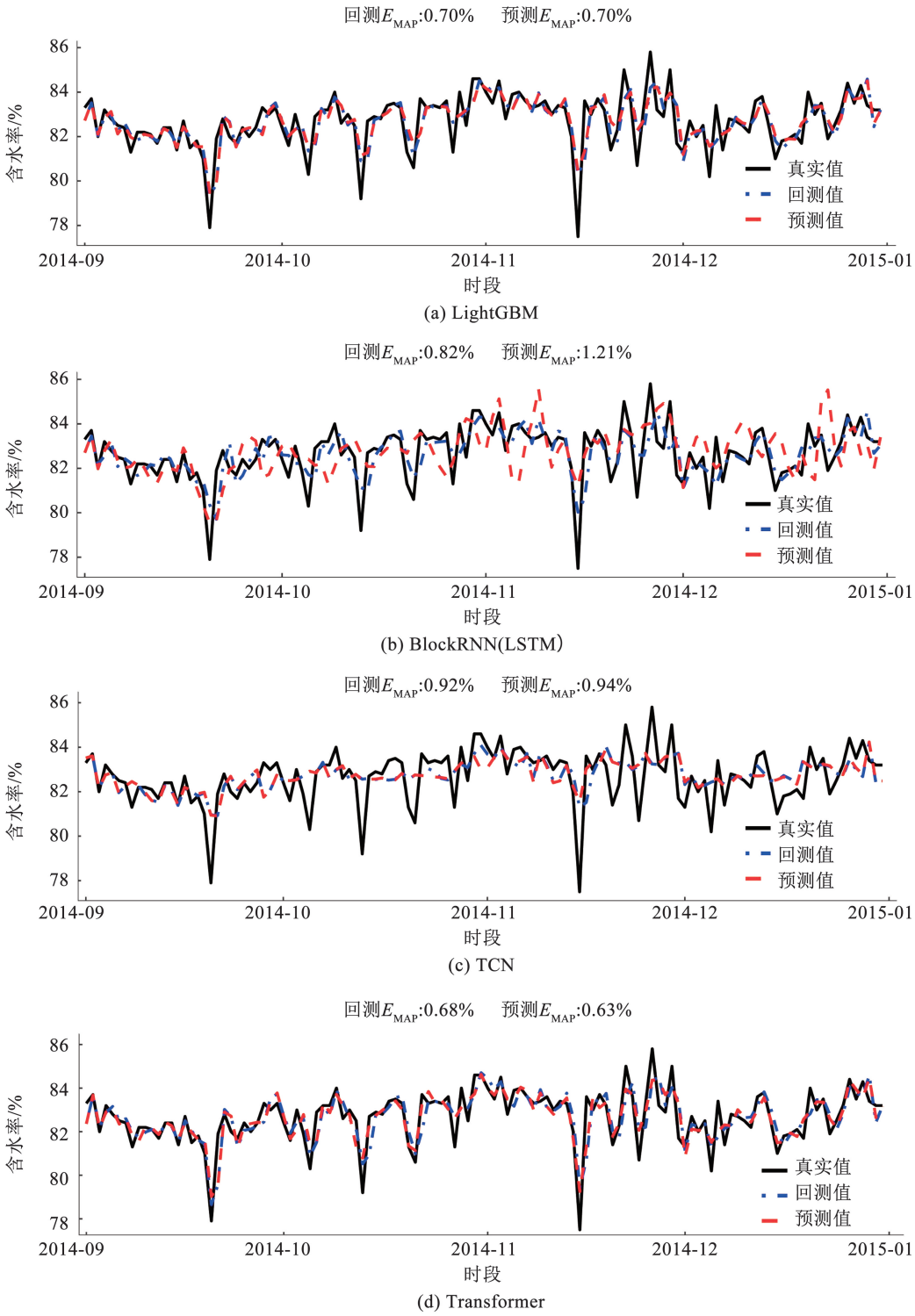



**图 5 再生水厂污泥含水率的模型预测效果对比**

Fig. 5 Comparison of model prediction effects for sludge moisture content in wastewater reclamation plant

**表 3 模型污泥含水率预测性能参数**

Tab. 3 Prediction performance parameters of models for sludge moisture content

| 模型 | $R^2$ | | $E_{MA}$ | | $E_{RMS}$ | | $E_{MAP}$/% | |
|---|---|---|---|---|---|---|---|---|
| | 回测 | 预测 | 回测 | 预测 | 回测 | 预测 | 回测 | 预测 |
| LightGBM | 0.588 5 | 0.582 8 | 0.573 5 | 0.572 3 | 0.787 3 | 0.794 4 | 0.696 4 | 0.695 6 |
| BlockRNN（LSTM） | 0.455 1 | 0.147 7 | 0.671 9 | 0.998 0 | 0.910 9 | 1.271 8 | 0.815 9 | 1.211 9 |
| TCN | 0.283 2 | 0.267 8 | 0.755 0 | 0.768 0 | 1.041 8 | 1.053 0 | 0.919 6 | 0.935 2 |
| Transformer | 0.622 5 | 0.679 0 | 0.556 9 | 0.515 0 | 0.757 6 | 0.695 1 | 0.675 0 | 0.625 1 |

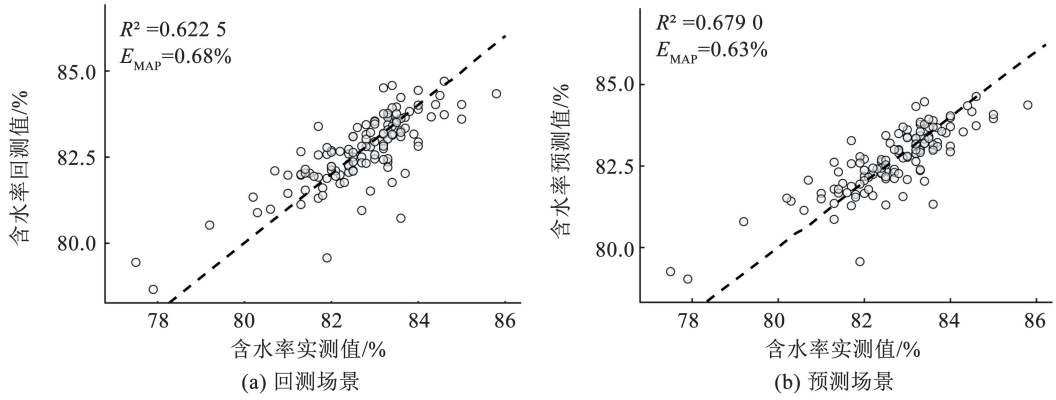



**图 6　Transformer 模型预测再生水厂污泥含水率拟合效果**

Fig. 6　Fitting effect of Transformer model predicting sludge moisture content in wastewater reclamation plant

将该模型应用于水厂实际运行场景，可实现泥线全流程工况的精细化智能调度。结合各模型预测场景(forecast)的实际性能数据，水厂可建立泥线参数预测预警机制，针对不同指标设定合理预警阈值：针对污泥产量，参考 Transformer 模型预测表现，设定 $E_{MA} \leqslant 20.0$、$E_{MAP} \leqslant 10.0\%$；针对污泥含水率，结合模型预测精度，设定 $E_{MA} \leqslant 0.7$、$E_{MAP} \leqslant 0.8\%$，当模型预测结果超出上述阈值时，及时触发预警并排查模型波动原因，保障工艺稳定运行。当预测到进水污染物浓度上升将导致污泥产量激增时，系统可自动启动备用脱水设备，并优化药剂投加策略，保障污泥脱水效果的同时避免药剂浪费；当监测到污泥含水率低于工艺要求阈值时，可及时降低脱水机运行功率，减少无效能量损耗。通过上述调度模式，不仅能够提升泥线的整体运行效率，降低污泥处理环节的电能、药剂等资源消耗，还能减少污泥处置过程中的碳排放与污染物排放，最终实现水厂泥线的高效、低碳运行。

基于多头自注意力机制的 Transformer 模型可实现泥线多参数的精准预测，为泥线系统的智能优化调控、数字孪生在线控制提供技术支撑。依托该模型对水厂泥线开展智能调度，可提升泥线运行效率、降低能耗，最终实现污水厂减污降碳的协同目标。综上，基于多头自注意力机制的泥线参数预测模型可为污水处理厂减污降碳工作提供重要的技术支撑。

## 3　结　论

以北京某再生水厂的水线与泥线运行数据为基础，采用经验模型基线-残差预测的混合建模策略，构建基于多头自注意力机制的泥线指标预测模型，通过 Transformer 模型对残差部分进行精准预测，并与多种经典模型开展性能对比分析，主要结论如下：

1) Transformer 模型依托多头自注意力机制的全局时序关联捕捉能力，可有效适配泥线参数的非线性波动特征，解决机理模型在实际预测场景中性能不足的问题，从而实现对泥线指标的高精度预测。

2) Transformer 模型依托多头自注意力机制，可并行挖掘水线-泥线多变量间的全局时序关联，动态适配泥线参数的非线性、波动性特征，脱水污泥产量预测 $R^2$ 达 0.778 8，污泥含水率预测 $R^2$ 达 0.679，均显著高于其他对比模型。同时，模型引入回测机制，借助滚动窗口的动态预测持续提升模型的稳定性与预测精度。

3) Transformer 模型可实现再生水厂泥线关键参数的可靠预测，能够支撑再生水厂泥线工艺的实时调控，为泥线协同减污降碳提供关键技术支撑。

## 参考文献

[1] EMEBU S, PECHA J, JANÁČOVÁ D. Review on anaerobic digestion models: model classification & elaboration of process phenomena[J]. Renewable and Sustainable Energy Reviews, 2022, 160: 112288. DOI: 10.1016/j.rser.2022.112288

[2] MO Rongrong, GUO Wenjie, BATSTONE D J, et al. Modifications to the anaerobic digestion model No. 1 (ADM1) for enhanced understanding and application of the anaerobic treatment processes: a comprehensive review[J]. Water Research, 2023, 244: 120504. DOI: 10.1016/j.watres.2023.120504

[3] D'BASTIANI C, KENNEDY D, REYNOLDS A. CFD simulation of anaerobic granular sludge reactors: a review[J]. Water Research, 2023, 242: 120220. DOI: 10.1016/j.watres.2023.120220

[4] OZGUN H. Anaerobic digestion model No. 1 (ADM1) for

mathematical modeling of full-scale sludge digester performance in a municipal wastewater treatment plant[J]. Biodegradation, 2019, 30(1): 27. DOI: 10.1007/s10532-018-9859-4

[5] ELMITWALLI T, FENG Y, BEHRENDT J, et al. Anaerobic digestion potential for ecological and decentralised sanitation in urban areas[J]. Water Science and Technology, 2006, 53(9): 45. DOI: 10.2166/wst.2006.276

[6] GHAZIZADE-FARD M, KOUPAIE E H. Anaerobic co-digestion of wastewater sludge and food waste: a machine learning approach to process modeling and optimization[J]. Journal of Environmental Management, 2025, 393: 126985. DOI: 10.1016/j.jenvman.2025.126985

[7] LING Zichen, WANG Jingjing, YUAN Shijie, et al. Machine learning-optimized advanced oxidation for enhanced sludge dewatering: EPS mechanistic insights and predictive modeling[J]. Water Research, 2025, 283: 123804. DOI: 10.1016/j.watres.2025.123804

[8] OULEBSIR R, LEFKIR A, SAFRI A, et al. Optimization of the energy consumption in activated sludge process using deep learning selective modeling[J]. Biomass and Bioenergy, 2020, 132: 105420. DOI: 10.1016/j.biombioe.2019.105420

[9] LI Bowen, LIU Li, XU Zikang, et al. Optimizing carbon source addition to control surplus sludge yield via machine learning-based interpretable ensemble model[J]. Environmental Research, 2025, 267: 120653. DOI: 10.1016/j.envres.2024.120653

[10] CHOI S, KIM S I, CHAIRATTANAWAT C, et al. Transformer-based multi-step time series forecasting of methane yield in full-scale anaerobic digestion[J]. Water Research, 2025, 286: 124276. DOI: 10.1016/j.watres.2025.124276

[11] CROLL H C, IKUMA K, ONG S K, et al. Unified control of diverse actions in a wastewater treatment activated sludge system using reinforcement learning for multi-objective optimization[J]. Water Research, 2024, 263: 122179. DOI: 10.1016/j.watres.2024.122179

[12] LI Junchen, LIN Sijie, ZHANG Liang, et al. Innovative multistep and synchronous soft sensing prediction of COD and $NH_3$ in WWTPs via multimodal data and multiple attention mechanisms[J]. Water Research, 2025, 278: 123405. DOI: 10.1016/j.watres.2025.123405

[13] 韩文杰, 邵先涛, 周家中, 等. 机理与 AI 协同驱动的污水智能加药工程效果评估[J]. 环境工程学报, 2026, 20(2): 340
HAN Wenjie, SHAO Xiantao, ZHOU Jiazhong, et al. Effect evaluation of sewage intelligent dosing control engineering driven by mechanism and AI collaboration[J]. Chinese Journal of Environmental Engineering, 2026, 20(2): 340

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017: 30. DOI: 10.48550/arXiv.1706.03762

[15] 王磊. 城市污水厂污泥产率季节变化与影响因素分析[J]. 净水技术, 2018, 37(6): 36
WANG Lei. Analysis of seasonal variation and influencing factors of sludge yield in municipal sewage plant[J]. Water Purification Technology, 2018, 37(6): 36. DOI: 10.15890/j.cnki.jsjs.2018.06.006

[16] 刘吉宝, 李亚明, 吕鑑, 等. 污水处理厂不同工艺的污泥脱水效能分析及其影响因素研究[J]. 环境科学, 2015, 36(10): 3794
LIU Jibao, LI Yaming, LÜ Jian, et al. Performance and factors analysis of sludge dewatering in different wastewater treatment processes[J]. Environmental Science, 2015, 36(10): 3794. DOI: 10.13227/j.hjkx.2015.10.031

(编辑 刘 彤)