

# 基于深度学习模型的中试纳滤系统膜污染预测研究

陈艳, 牛亚林, 彭兴, 郑文静, 徐小虎, 徐鹏成

(兰州交通大学 环境与市政工程学院, 兰州 730070)

**摘要:**膜污染的预测或模拟对于阐明膜污染机理和制定有效的污染控制措施具有重要意义。建立了一种门控循环单元(GRU)模型,用于预测中试纳滤系统的过滤性能,并探究膜的污染机理。采用跨膜压差(TMP)作为输出变量,考察以TMP自身为输入变量的单输入时间序列模型的预测效果,并探讨了9种水质参数:pH、水温、电导率、总溶解性固体(TDS)、总硬度(TH)、浊度、高锰酸盐指数(COD<sub>Mn</sub>)、溶解性有机碳(DOC)以及UV<sub>254</sub>的独立和耦合效应对膜污染及预测结果的影响,筛选出最优模型与膜污染预测中常用的机器学习算法——随机森林(RF)和时间序列预测中广泛使用的长短时记忆网络(LSTM)进行比较。结果表明:以TMP自身数据作为输入的LSTM以及GRU单输入时间序列预测模型的 $R^2$ 分别达到了0.9613和0.9861;水质参数对TMP的相关性顺序依次为:温度>电导率>TDS>总硬度>高锰酸盐指数>DOC>浊度>UV<sub>254</sub>>pH;以温度和电导率为输入的多变量GRU模型预测效果最佳( $R^2=0.8344$ ),预测精度优于相同时间步长的单输入GRU模型( $R^2=0.4555$ )以及相同输入参数的多变量LSTM( $R^2=0.6428$ )和RF模型( $R^2=-4.1894$ ),在此基础上增加或减少水质参数作为模型输入,模型的预测精度均有所下降。时间序列模型在膜污染预测方面展现出了较高的可靠性,GRU模型预测精度更高,在膜污染预测中具备更高的应用潜力。输入变量的特征选择对高效预测膜污染具有重要意义,对输入水质数据进行特征选择,可以显著提升模型的预测性能。此外,预测结果反向验证了水温和无机离子耦合是造成冬季纳滤膜污染的主要原因,冬季纳滤运行过程中要注意温度和无机离子污染对纳滤膜运行稳定性的影响。

**关键词:**膜污染预测;深度学习;GRU;纳滤;中试

中图分类号:TQ028.3

文献标志码:A

## Prediction of Membrane Fouling in Pilot Nanofiltration System based on Deep Learning Model

CHEN Yan, NIU Yalin, PENG Xing, ZHENG Wenjing, XU Xiaohu, XU Pengcheng

(School of Environmental and Municipal Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

**Abstract:** The prediction or simulation of membrane fouling is of great significance for elucidating the mechanism of membrane fouling and devising effective pollution control measures. In this study, a gated recurrent unit (GRU) model was established to predict the filtration performance of a pilot-scale nanofiltration system and explore the fouling mechanism of membranes. Trans membrane pressure (TMP) was adopted as the output variable, and the predictive effectiveness of a single-input time series model using TMP itself as the input variable was examined. Furthermore, the independent and coupled effects of nine water quality parameters: pH, water temperature, conductivity, total dissolved solids (TDS), total hardness (TH), turbidity, permanganate index, dissolved organic carbon (DOC), and UV<sub>254</sub> on membrane fouling and prediction results were investigated. The optimal model was selected and compared with commonly used machine learning algorithms in membrane fouling prediction, namely random forest (RF) and long short-term memory (LSTM) networks widely used in time series prediction. The results indicated that the  $R^2$  values of the

LSTM and GRU single-input time series prediction models using TMP itself as input reached 0.961 3 and 0.986 1, respectively. The order of correlation between water quality parameters and TMP was as follows: temperature>conductivity>TDS>TH>COD<sub>Mn</sub>>DOC>turbidity>UV<sub>254</sub>>pH. The multivariate GRU model with temperature and conductivity as inputs exhibited the best prediction performance ( $R^2=0.834 4$ ), with higher accuracy than the single-input GRU model with the same time step ( $R^2=0.455 5$ ) and the multivariate LSTM ( $R^2=0.642 8$ ) and RF models ( $R^2=-4.189 4$ ) with the same input parameters. Therefore, increasing or decreasing water quality parameters as model inputs will result in a decrease in the prediction accuracy of the model. Time series models have demonstrated high reliability in predicting membrane fouling, with the GRU model showing higher prediction accuracy and greater potential for application in membrane fouling prediction. Feature selection of input variables is crucial for efficiently predicting membrane fouling, and selecting features from input water quality data can significantly enhance model predictive performance. Additionally, the prediction results confirmed that the coupling of water temperature and inorganic ions is the main cause of winter nanofiltration membrane fouling. Therefore, attention should be paid to the influence of low temperature and inorganic ion pollution on membrane operation stability during winter membrane operation processes.

**Key words:** membrane fouling prediction; deep learning; GRU; nanofiltration; pilot test

膜污染是制约膜工艺发展的主要因素,有关膜污染的研究一直是膜处理领域持续研究的重点。在膜污染的相关研究中,膜污染的预测或模拟对于阐明膜污染机理和制定有效的污染控制措施具有重要意义<sup>[1-2]</sup>。然而,膜污染过程非常复杂,影响因素众多。尽管目前的研究已经积累了大量有关膜污染的数据,但这些数据往往蕴含着各种反应机理与控制要素,各数据之间存在着非常复杂的非线性关系。传统的数学建模方法难以解决这种复杂的非线性特征问题,导致预测过程中往往存在准确性低、灵敏度差等问题<sup>[3]</sup>。

随着机器学习的发展,以数据驱动的机器学习模型取代传统模型的可行性开始被广泛讨论<sup>[4-6]</sup>。机器学习通过从大量信息中提取推理并将其转换为有用的模型,从而提供处理回归任务的能力<sup>[7-8]</sup>,有助于解释理解复杂的水质参数对膜污染的贡献和影响。人工神经网络(ANN)及其衍生模型(例如,模糊逻辑(FL)<sup>[9]</sup>、支持向量机(SVM)<sup>[10]</sup>和搜索算法<sup>[11]</sup>等基本算法)已开发出来用于预测膜污染过程,其预测不需要考虑输入参数和模型输出之间关系的数学方程<sup>[12]</sup>,却显示出比现有传统膜污染数学模型更高的预测精度<sup>[13-14]</sup>。然而,由于ANN算法的特性,ANN网络架构随着神经元数量的增加而逐渐变大,在面对中试或者实际规模水厂中更复杂的数据集,ANN网络在膜污染预测方面面临巨大挑战<sup>[15]</sup>。

随着机器学习模型的进一步发展,深度学习模型成为更先进的数据驱动模型被引入<sup>[16-17]</sup>,Im等<sup>[18]</sup>基于实时膜表面图像和卷积神经网络(CNN)

模型来预测膜的过滤性能。结果显示基于图像的膜污染预测的优越性,使得能够通过3D图像生成提高污染层生长的预测精度并进行可视化。Lee等<sup>[19]</sup>使用长短期记忆网络(LSTM)预测了膜的过滤性能,该模型成功预测了两个参数:渗透通量和污染层厚度;并指出时间序列预测可能更适用于膜污染的预测。原因可能为膜污染过程是在不同的给水特性和操作条件下受到连续过滤影响,而无论是收集的水质数据还是纳滤TMP(或膜通量)数据都具有较强的时间序列特征,因此基于时间序列的纳滤膜TMP(或膜通量)深度学习预测模型具有实际意义,可以提供比使用ANN和CNN等机器学习模型更高的预测性能<sup>[20-21]</sup>。然而,尽管时间序列算法一直在发展,在基于LSTM上进一步发展出门控循环单元网络(GRU),与LSTM相比,可以减少操作量,有效加快网络的收敛速度,但使用GRU时间序列模型进行膜污染预测的案例仍然很少,且大多数研究集中在模拟实验室规模膜系统中的通量或跨膜压力变化,很少有研究证明时间序列深度学习模型在中试规模上对膜系统污染预测的有效性。此外,深度学习在训练过程中能学习到与目标变量相关的特征,但其过程未知,无法从原理方面对预测结果进行解释。因此模型输入前进行特征工程的必要性一直存在讨论空间,进一步探讨模型输入变量选择对预测结果的影响,可以提高模型的预测精度和效率,也可以进一步对实际膜污染过程做出解释。

基于上述背景,研究引入了一种新的时间序列模型,用于预测中试规模的膜过滤性能。该模型的核心算法为基于LSTM的变种——GRU(gate recurrent unit)算法。研究考察了仅依靠膜污染数据

的自相关性进行预测的效果。此外,在中试纳滤系统中跨膜压差与水质数据的相关性分析的基础上,探讨不同水质参数独立和耦合效应对膜污染及预测结果的影响。期望本研究能够为深度学习应用于膜污染预测提供一定的参考。

## 1 材料及方法

### 1.1 中试概述

中试实验地点在江苏省苏州市某水厂,陶氏公司生产的 NF270 为实验用膜。纳滤系统采用恒定通量变压模式运行,通量为  $25 \text{ L}/(\text{m}^2 \cdot \text{h})$ ,回收率为 30%。整组纳滤系统接在水厂砂滤池后,工艺流程如图 1 所示。试验期间 TMP 数据由 PLC 记录,经过数据预处理后作为深度学习模型的数据集。

### 1.2 数据来源和预处理

纳滤膜的 TMP 数据采集时间为 2021-11-05—2021-12-15,采集频率为 1 min。对数据的缺失值,采取两种处理办法:如果一次监测中仅缺失一个指标,即使用前一次的监测值来补上;如果某个时间段 TMP 数据连续缺失,为避免人为填补造成较大误差,删除这些监测时刻的数据<sup>[19]</sup>,最终得到总计 56235 组 TMP 数据,数据趋势如图 2 所示。由于 TMP 数据记录间隔较短,TMP 在一小段时间内没有变化,且数据噪点较大,这会对于模型的训练不利<sup>[20]</sup>。因此,为充分发掘时序数据之间的关联性,将原始数据分别按照 1 min、30 min、1 h 和 1 d 的时间步长重新划分数据,用作单变量时间序列模型的训练与测试。

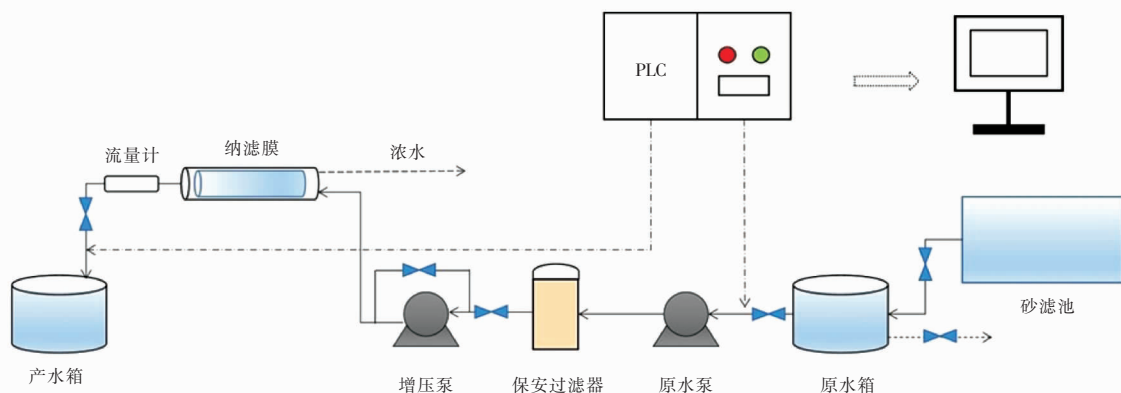


图 1 纳滤中试工艺流程图

Fig. 1 Scheme of the pilot-scale NF system in this study

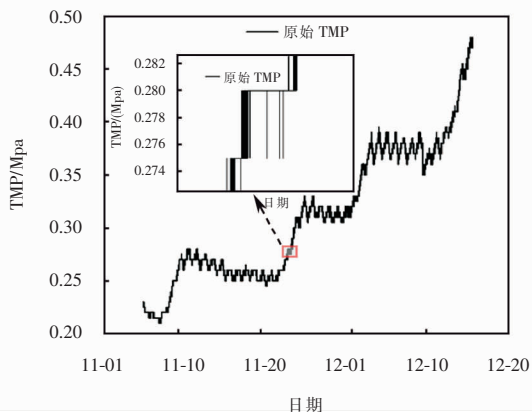


图 2 原始 TMP 数据趋势图

Fig. 2 Original trend chart of TMP data

试验同期对纳滤膜进水的 9 种水质数据进行监测,包括 pH、水温、电导率、总溶解性固体(TDS)、总硬度(TH)、浊度、高锰酸盐指数( $\text{COD}_{\text{Mn}}$ )、溶解性有机碳(DOC)以及  $\text{UV}_{254}$ 。采集方法均按照 GB/T 5750.1—2006《生活饮用水标准检验方法》进行。

采集频率为 2~3 天/次。对水质数据集进行线性插值处理,使其时间步长为 1 d 来扩充数据集,最终得到  $9 \times 41$  组水质数据,水质数据的整体描述性统计情况如表 1 所列。在考察水质特性时加入对模型预测效果的影响,为配合水质数据的步长,TMP 数据同样采用 1 d 的时间步长。水质和 TMP 数据均以 8 : 2 的比例划分为训练集和测试集,分别用于模型的训练和测试。

### 1.3 水质相关性分析

研究进水水质参数与膜污染之间的关系,考虑水质特性对膜污染的影响,以提升模型预测膜污染的能力。水质参数对膜污染的影响机理复杂,输入参数存在冗余,会使得预测模型的准确性下降或使模型产生过拟合,输入参数特征太少,神经网络对膜污染数据特征的提取和学习会有所不足,造成模型效果不佳以及缺乏对膜污染变化趋势原因的洞察。

通过特征选择可以减少模型学习时间,提升模

型效率,使模型泛化能力更强,也有助于增加模型的可解释性<sup>[21]</sup>。研究以相关性分析作为特征选择的依据,通过皮尔逊相关系数(Pearson's  $r$ )综合评价

水质参数与 TMP 之间的相关性,分析水质特性对膜污染的影响,并据此进行水质特征选择,确定模型输入变量。

表 1 水质数据的描述性统计情况

Tab. 1 Descriptive statistics of water quality parameters

	pH	水温 /°C	电导率 /( $\mu$ S/cm)	TDS /(mg/L)	总硬度 /(mg/L)	浊度 /(NTU)	COD <sub>Mn</sub> /(mg/L)	DOC /(mg/L)	UV <sub>254</sub> /( $\text{cm}^{-1}$ )
极大值	7.53	20.20	413.00	206.00	99.68	0.17	1.89	2.82	0.035
极小值	7.18	11.30	355.00	177.60	76.67	0.09	1.20	2.02	0.020
均值	7.40	14.41	400.43	199.15	93.72	0.11	1.51	2.53	0.027
标准差	0.07	2.26	10.91	6.35	4.61	0.02	0.19	0.17	0.003

#### 1.4 GRU 膜污染预测模型

将 LSTM 的输入门和遗忘门耦合为更新门,用于控制隐藏状态的更新,减少了矩阵乘法的运算,可以有效加速网络的收敛<sup>[22]</sup>。GRU 在结构上比 LSTM 更加简洁,参数数量更少,计算成本更低。其结构如图 3 所示。

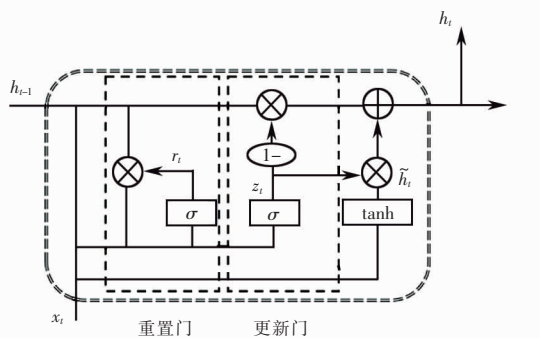


图 3 GRU 内部结构示意图

Fig. 3 Schematic diagram of the internal structure of the GRU

GRU 模型通过上一个传输下来的状态  $h_{t-1}$  和当前节点的输入  $x_t$  来确定单元的输入状态。公式如式(1)、(2),其中,式(1)通过重置门  $r_t$  决定了前一个时间步长的隐藏状态  $h_{t-1}$  在当前时间步长的记忆中的影响程度。式(2)通过更新门  $z_t$  决定了前一个时间步长的隐藏状态  $h_{t-1}$  对当前时间步长的输入  $x_t$  的影响程度。

$$r_t = \sigma(\omega_r \cdot [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(\omega_z \cdot [h_{t-1}, x_t]) \quad (2)$$

模型通过式(3)、(4)实现信息的储存,使用重置门来得到“重置”之后的数据  $h'_{t-1}$ ,将  $h'_{t-1}$  与  $x_t$  进行拼接后,再通过  $\tanh$  激活函数来生成新的数据,相当于“记忆了当前时刻的状态”。

$$r'_{t-1} = h_{t-1} \odot r_t \quad (3)$$

$$\tilde{h}_t = \tanh(\omega \cdot [h'_{t-1}, x_t]) \quad (4)$$

最后,更新旧单元信息,这个阶段相当于“更新记忆”阶段,控制新记忆和先前记忆之间的权重,同时进行选择性地遗忘和记忆两个步骤,公式如式(5),最后输出公式如式(6)。

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (5)$$

$$y_t = \sigma(\omega_0 \cdot h_t) \quad (6)$$

研究使用的 GRU 模型结构与主要参数如表 2 所列,第 1 层为 80 个神经元的 GRU 训练模型,并设置返回序列,以此捕捉输入序列中的时间相关性和序列信息,以增加模型的表达能力。第 2 层是一种正则化技术,通过随机丢弃部分神经元以减少过拟合的风险。第 3 层是另一个具有 100 个神经元的 GRU 层,未设置返回序列,该层的输出将是最后一个时间步长的隐藏状态。基于上述原理,模型将基于过往的拟合结果进行学习,利用 GRU 具有记忆的特性优化预测能力。第 4 层同第 2 层目的一致,是为了防止发生过拟合情况,让所有神经元更好的学习。连接到第 5 层,对上一层的神经元进行全连接并进行输出。

表 2 模型结构

Tab. 2 Network structure

神经元层	模型	输出张量	参数量
1	GRU	(None, 60, 80)	19 920
2	Dropout	(None, 60, 80)	0
3	GRU_1	(None, 100)	54 600
4	Dropout	(None, 100)	0
5	Dense	(None, 1)	101

#### 1.5 模型比较与定量评估方法

为了评价本文采用的 GRU 模型的可靠性以及

优越性,选择在时间序列预测中已较为成熟的长短时记忆网络(LSTM)以及常用于膜污染预测的传统机器学习算法—随机森林(RF)进行对比。模型损失函数是评估模型性能的必要条件。研究选用 4 个模型评估指标:平均绝对误差(MAE)、均方误差(MSE)、均方根误差(RMSE)和决定系数( $R^2$ )来定量评估模型预测效果,具体表达式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |T_i - P_i| \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (T_i - P_i)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_i - P_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (T_i - P_i)^2}{\sum_{i=1}^n (T_i - \bar{T}_i)^2} \quad (10)$$

式中: $P_i$ 为模型预测值; $T_i$ 为实际值; $\bar{T}_i$ 为实际值的平均值; $n$ 表示每个比较组的样本数; $MAE$ 用于评估真实值与预测值之间的偏差程度; $MSE$ 和 $RMSE$ 用于评估模型拟合值与真实值之间的差异程度; $R^2$ 用于反映因变量的全部变异能通过回归关系被变量

解释的比例。

## 2 结果和讨论

### 2.1 模型输入特征选择

输入序列的步长对预测结果会产生重要影响。在单变量时间序列预测中,为寻求最佳的输入序列步长,探究输入序列步长为 1 min、30 min、1 h、1 d 的预测效果,结果如图 4 所示。

从图 4 中可以看出,采用时间步长为 1 min,数据集噪声过大干扰了模型的学习过程,在实际调试过程中,模型频繁出现过拟合现象,尽管预测的  $R^2$  达到 0.906 7,但最终生成的模型泛化性能较差,在预测集后半段预测值与实际值差距较明显。时间步长为 30 min 时,数据集的噪声明显减少,模型的预测性能明显提升,较好反映出了实际趋势, $R^2$  达到了 0.973 3。选择 1 h 作为时间步长,模型预测性能能进一步提升( $R^2 = 0.986 1$ ),预测值与实际值之间的  $MAE$ 、 $MSE$ 、 $RMSE$  相比时间步长为 30 min 时,分别降低了 29.0%、43.3%和 25.6%。继续增加时间步长为 1 d,由于数据集缩小,模型没有足够的数据集来学习数据的特征,模型预测性能不佳( $R^2 = 0.455 5$ )。

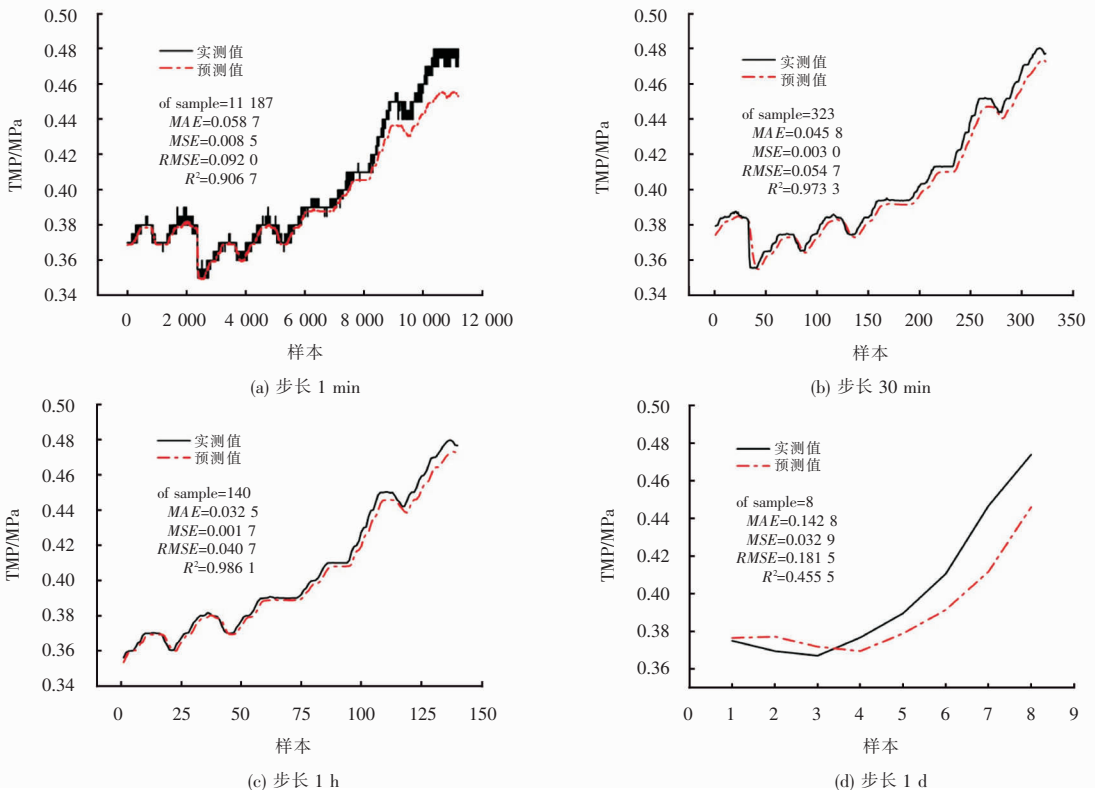


图 4 不同时间步长数据输入下的单变量模型预测结果

Fig. 4 Uni variate model prediction results under different time step data inputs

通过线性拟合的方式判断 TMP 与水质指标之间的相关性,以相关性先验不同水质参数与 TMP

之间的关系,结果如图5所示。

一般认为,皮尔逊相关系数在 $0\sim 0.3$ 为不相关或弱相关, $0.3\sim 0.6$ 为中度相关, $0.6\sim 0.8$ 之间为强相关, $0.8\sim 1$ 为极强相关<sup>[23]</sup>。从图中可以看出,温度与TMP具有高度相关性(Pearson's  $r > 0.8$ ),说明中试试验期间,进水水温是影响纳滤膜污染重要因素。温度对膜过滤性能的影响已有大量文献报道,水温下降会导致水的密度和黏度增大从而增加了过流阻力<sup>[24]</sup>。同时水温降低对也会影响膜的结构(如膜孔收缩),随着孔径的减小,在保持恒通量运行的条件下,施加在膜组件上的压力因此增加<sup>[25]</sup>。衡量水中无机离子指标的电导率、TDS与TMP呈中度相关

( $0.8 > \text{Pearson's } r > 0.6$ ),总硬度与TMP低度相关( $0.6 > \text{Pearson's } r > 0.3$ )。说明无机离子也可能是造成纳膜污染的重要因素,由于NF膜对无机离子的截留,导致膜表面及附近的离子浓度增加,可能超过了微溶盐的溶解度极限(溶度积)形成无机垢。但值得注意的是,pH、浊度以及有机物指标(高锰酸盐指数、DOC、UV<sub>254</sub>)与TMP并无相关性(Pearson's  $r < 0.3$ ),这可能与试验在冬季进行,原水中的有机物含量低,且纳滤进水经过(混凝-沉淀-砂滤)常规预处理工艺处理后,水的pH、浊度、DOC、UV<sub>254</sub>、高锰酸盐指数等被控制在相对稳定的水平,变化幅度较小,导致与膜压之间的相关性不显著。

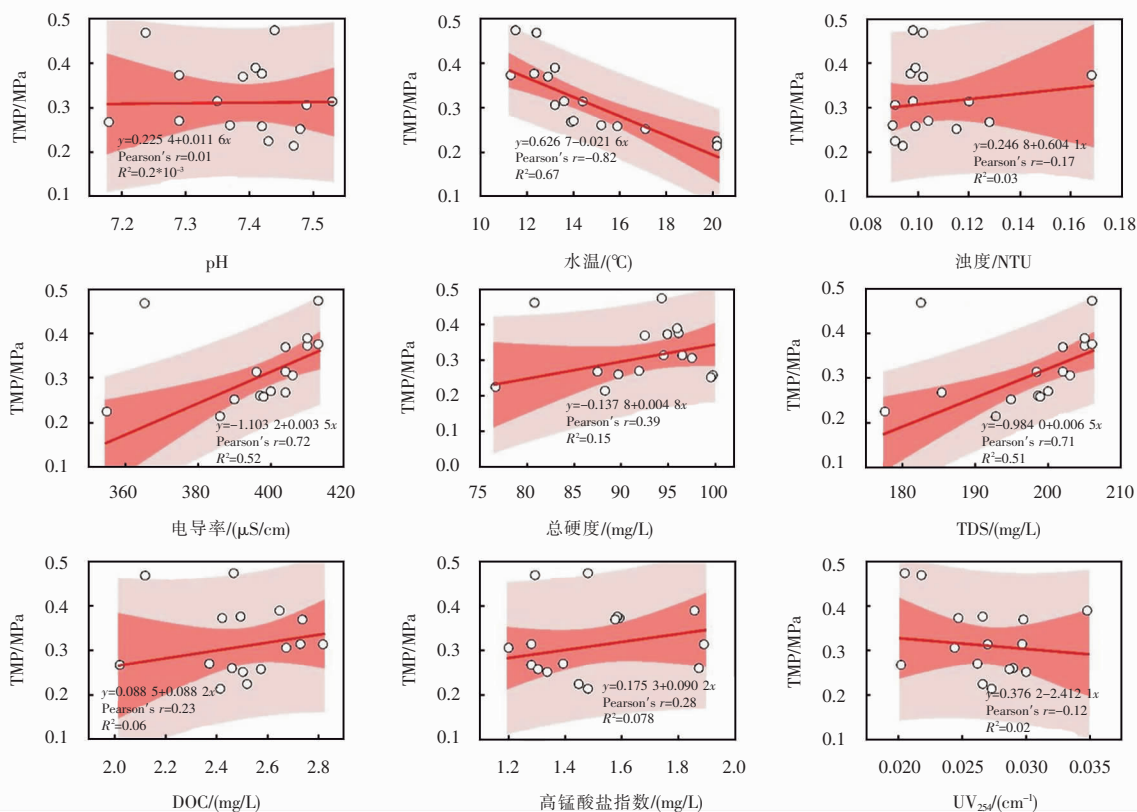


图5 9种水质参数与TMP之间的相关性

Fig. 5 Correlation between 9 water quality parameters and TMP

根据相关性分析结果,各种水质参数对TMP的影响顺序依次为:温度>电导率>TDS>总硬度>高锰酸盐指数>DOC>浊度>UV<sub>254</sub>>pH,因此,根据先验关系,选择模型的输入特征选择参数依次按此顺序进行增加,这种选择方式假设了权重较高的水质参数组合输入模型会获得更好的预测结果,大量研究表明,经过提取主要特征用作机器学习模型的输入参数,可以提升模型的预测效能,加入更多权重较低的数据反而会使模型的预测效果下降<sup>[26-27]</sup>,这样做可以用最快的速度找到最佳特征选择参数的组

合,以提高模型的预测效率和精度。此外,根据模型的预测结果,反向验证具体几个水质参数耦合对膜污染的贡献最大,进一步为膜污染的预测和膜污染的机理解释提供参考。最终选择模型输入的特征选择见表3。模型根据数据来源分为2类,一类为不考虑水质参数的单输入时间序列预测模型,另一类为考虑水质参数输入的多变量预测模型。

## 2.2 单变量时间序列预测

考察所选GRU模型的预测效果,与时间序列中最常用的LSTM模型做比较。两种模型的预测

结果如图 6 所示,表 4 为两种时间序列预测模型的 训练结果评价指标。

表 3 模型输入变量  
Tab. 3 Model input variables

模型	选择方法	输入变量	时间步长	输出变量
单变量	时间序列	TMP	1 h	
	特征选择_1	温度、TMP		
	特征选择_2	温度、电导率、TMP		
	特征选择_3	温度、电导率、TDS、TMP		
	特征选择_4	温度、电导率、TDS、总硬度、TMP		
多变量	特征选择_5	温度、电导率、TDS、总硬度、高锰酸盐指数、TMP	1 d	TMP
	特征选择_6	温度、电导率、TDS、总硬度、高锰酸盐指数、DOC、TMP		
	特征选择_7	温度、电导率、TDS、总硬度、高锰酸盐指数、DOC、浊度、TMP		
	特征选择_8	温度、电导率、TDS、总硬度、高锰酸盐指数、DOC、浊度、UV <sub>254</sub> 、TMP		
	特征选择_9	温度、电导率、TDS、总硬度、高锰酸盐指数、DOC、浊度、UV <sub>254</sub> 、pH、TMP		

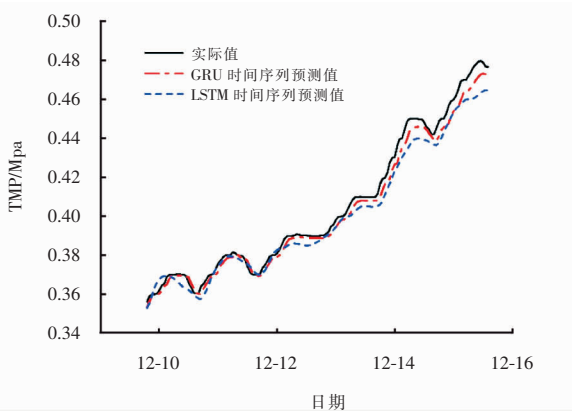


图 6 单变量时间序列预测趋势图

Fig. 6 Uni-variate time series forecast trend chart

从图 6 及表 4 中可以看出,以 TMP 自身数据作为输入的时间序列预测模型能达到很好的预测效果,GRU 以及 LSTM 模型的  $R^2$  分别达到了 0.961 3 和 0.986 1,表明了时间序列模型用于膜污染预测的潜力。与 LSTM 模型相比,GRU 模型的 MAE、MSE、RMSE 分别降低了 41.22%、65.05%、40.88%,显示出了更高的预测精度。在预测后期,两种时间序列模型都存在预测值偏小,预测效果下降的情况。推测造成这种现象的原因,一方面是因为时间序列是依靠前一段数据来预测下一个数据,由于误差传播定律,造成在预测后期误差不断增加。另一方面或是因为外界因素波动的影响,造成了这一时期的跨膜压差突然增加,但是这种趋势很难被仅以 TMP 自身作为输入的时间序列预测模型学习到,需要进一步增加输入特征以优化模型。

2.3 考虑给水特性影响的多变量预测模型

考虑水温、水质等参数影响的几种模型的预测

结果与实测值对比如图 7 所示,模型的定量评价对比结果如表 5 所列。

表 4 GRU 与 LSTM 时间序列预测模型训练评价

Tab. 4 Training evaluation indexes of GRU and LSTM time series forecasting models

模型	评价指标			
	$R^2$	MAE	MSE	RMSE
GRU 时间序列	0.986 1	0.032 5	0.001 7	0.040 7
LSTM 时间序列	0.961 3	0.055 3	0.004 7	0.068 8

结果表明,在加入水质数据的多变量预测模型中,以水温和电导率作为输入的模型(GRU 特征选择\_2)取得了最好的预测效果( $R^2 = 0.834 4$ ,  $MAE = 0.097 5$ ,  $MSE = 0.014$ ,  $RMSE = 0.118 3$ ),仅以温度为输入参数的模型(GRU 特征选择\_1)预测精度( $R^2 = 0.745 3$ ,  $MAE = 0.117 1$ ,  $MSE = 0.018 6$ ,  $RMSE = 0.136 5$ )要低于 GRU 特征选择\_2 模型。这说明水温和电导率共同作用是导致膜压变化的主要原因。尽管水温的降低也会导致电导率的升高,但通过深度学习模型的特征提取结果,这种影响显然不全是水温主导,无机离子对膜污染的影响同样至关重要。继续增加水质参数作为模型的输入参数,其预测效果不断下降。将 9 种水质参数作为输入,模型的预测效果最差( $R^2 = -4.092 8$ ,  $MAE = 0.372 6$ ,  $MSE = 0.185 4$ ,  $RMSE = 0.430 6$ ),说明这些水质数据的输入,并不能对膜压的变化做出更为准确的解释,即这些水质参数并不是造成 TMP 增加的主要原因,这些冗余的数据作为模型输入干扰了学习过程,导致模型预测精度下降。张怡婷等<sup>[28]</sup>采用 LSTM 进行河流水质预测,结果也表明,输入

的变量经过特征选择时,模型预测效果优于全部水质变量作为输入的预测结果,不进行特征选择时,输

入变量增加,模型会因为学习到噪声而使精度下降,这与本文的研究结果一致。

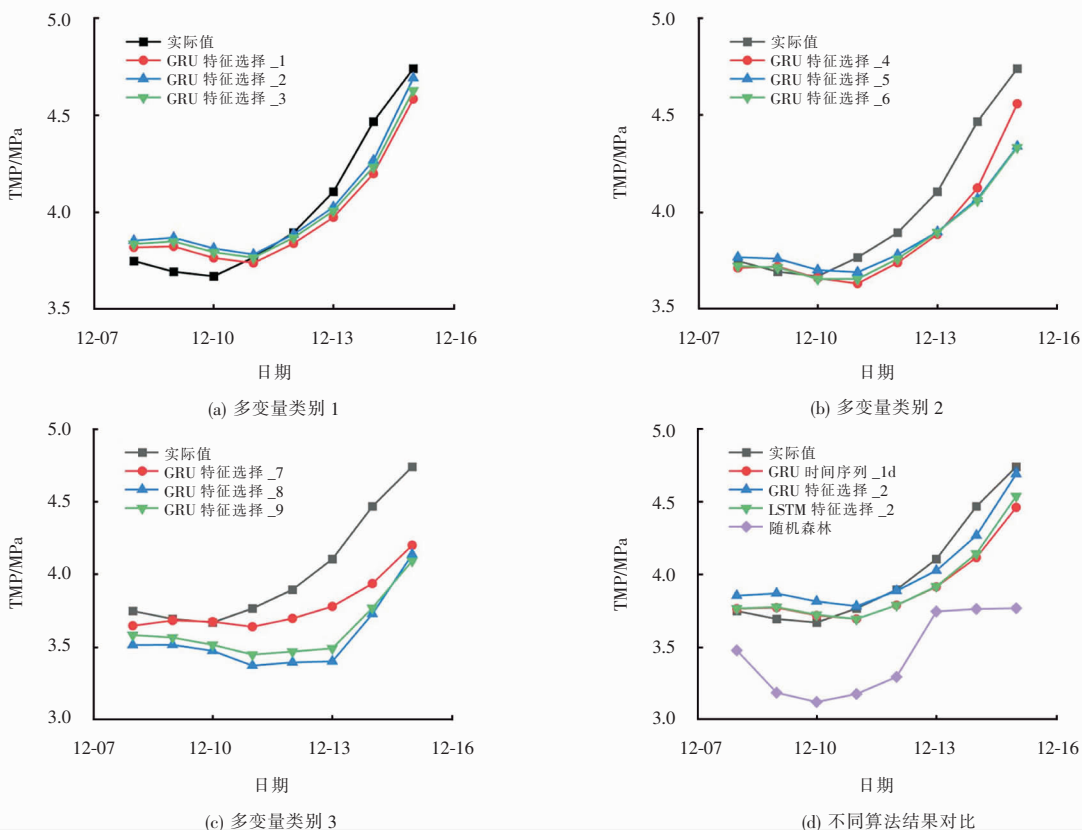


图 7 TMP 预测趋势图

Fig. 7 TMP forecast trend chart

表 5 模型特征预测效果比较

Tab. 5 Comparison of model feature prediction effects

序号	模型	评价指标			
		$R^2$	MAE	MSE	RMSE
1	GRU 时间序列_1d	0.455 5	0.142 8	0.032 9	0.181 5
2	GRU 特征选择_1	0.745 3	0.117 1	0.018 6	0.136 5
3	GRU 特征选择_2	0.834 4	0.097 5	0.014 0	0.118 3
4	LSTM 特征选择_2	0.642 8	0.131 2	0.026 0	0.161 3
5	GRU 特征选择_3	0.793 5	0.105 5	0.015 8	0.125 5
6	GRU 特征选择_4	0.653 5	0.138 3	0.030 4	0.174 4
7	GRU 特征选择_5	-0.101 4	0.164 3	0.048 3	0.219 8
8	GRU 特征选择_6	-0.018 8	0.165 9	0.050 7	0.225 1
9	GRU 特征选择_7	-1.813 2	0.229 6	0.093 1	0.305 1
10	GRU 特征选择_8	-3.219 2	0.442 8	0.241 5	0.491 4
11	GRU 特征选择_9	-4.092 8	0.372 6	0.185 4	0.430 6
12	随机森林	-4.189 4	0.568 0	0.362 5	0.602 1

通过加入水质数据的多变量预测模型预测结果,结合上述 2.1 节水质参数与 TMP 的相关性分析,可以得出该中试试验期间纳滤膜污染的行为与影响机制。中试试验在冬季进行,水温变化比较剧

烈,对膜的运行稳定性有很大影响,而无机离子结垢则可能是造成了冬季纳滤膜污染的主要原因。冬季原水中有机物含量低,微生物代谢活动缓慢,且纳滤进水经过常规工艺预处理工艺处理后,水的 pH、浊

度、UV<sub>254</sub>、高锰酸盐指数、DOC 等被控制在相对稳定的水平,胶体和有机污染对膜的影响较低。但水温降低会导致水中钙、镁等离子溶解度增加,而常规处理工艺过程为了保证冬季混凝的效果,往往需要增加混凝剂的投加量,因此,纳滤膜更容易受到无机离子污染的影响。

图 7(d)为筛选出的多输入最优模型(GRU 特征选择\_2),将其与同时间步长(1 d)的单输入时间序列预测模型(GRU 时间序列\_1 d)、LSTM 多输入模型(GRU 特征选择\_2)以及 RF 模型的预测结果对比。结果显示,与 GRU 时间序列\_1 d 相比,GRU 特征选择\_2 的预测效果更优, $R^2$  由 0.455 5 增加到 0.834 4,且在后期膜压突然增加时,更容易捕捉到这种趋势,可见筛选后的水质参数加入,显著提升了模型的预测效果。同时,GRU 特征选择\_2 的预测精度亦优于 LSTM 多输入模型(GRU 特征选择\_2)以及 RF 模型,表明了 GRU 模型在膜污染预测中的可靠性。

### 3 结论

1) 时间序列模型在膜污染预测方面展现出了较高的可靠性。以 TMP 自身数据作为输入的 LSTM 和 GRU 模型均达到很好的预测效果( $R^2 > 0.95$ )。与 LSTM 和 RF 模型相比,GRU 模型预测精度更高,在膜污染预测中具备更高的应用潜力。

2) 多输入模型的预测结果表明,对输入水质数据进行特征选择,可以显著提升模型的预测性能,而加入更多维度的数据并不能提升模型预测效能。因此输入变量的特征选择对高效预测膜污染具有重要意义。

3) 水质参数与 TMP 的相关性分析表明,试验期间,相比胶体和有机污染,无机结垢对膜污染的贡献更大,水温同样是影响膜过滤性能的重要因素。多输入模型的预测表明,以水温和电导率为输入的模型预测效果最优。因此,试验期间水温和无机离子的耦合作用是导致膜过滤性能变差的主要原因,冬季纳滤运行过程中要注意温度和无机离子污染对纳滤膜运行稳定性的影响。

### 参考文献:

[1] NIU C X, LI X S, DAI R B, et al. Artificial intelligence-incorporated membrane fouling prediction for membrane-based processes in the past 20 years: a critical review[J]. Water Research, 2022, 216: 118299.

[2] 闵芮, 马凯, 任豪, 等. 混凝-氯/紫外高级氧化与双膜工艺联用处理微污染窖水[J]. 兰州交通大学学报, 2022, 41(4): 89-95.

[3] 皇甫小留, 王晶瑞, 龙鑫隆, 等. 机器学习在水处理系统中的应用[J]. 给水排水, 2022, 48(11): 153-165.

[4] 崔晓宁, 王起才, 张戎令, 等. 基于随机森林的高性能混凝土抗压强度预测[J]. 兰州交通大学学报, 2021, 40(6): 1-6, 14.

[5] SAIF A A, TALAL B, HASAN S W, et al. Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination? [J]. Desalination, 2019, 458(1): 84-96.

[6] WANG T J, LI Y Y. Predictive modeling based on artificial neural networks for membrane fouling in a large pilot-scale anaerobic membrane bioreactor for treating real municipal wastewater[J]. Science of the Total Environment, 2024, 912: 169164.

[7] HUANG R X, MA C X, MA J, et al. Machine learning in natural and engineered water systems[J]. Water Research, 2021, 205: 117666.

[8] XU R Z, CAO J S, YE T, et al. Automated machine learning-based prediction of microplastics induced impacts on methane production in anaerobic digestion [J]. Water Research, 2022, 223: 118975.

[9] MADAENI S S, KURDIAN A R. Fuzzy modeling and hybrid genetic algorithm optimization of virus removal from water using microfiltration membrane[J]. Chemical Engineering Research & Design, 2011, 89(4): 456-470.

[10] LIU J W, KANG X Y, LUAN X R, et al. Performance and membrane fouling behaviors analysis with SVR-LibSVM model in a submerged anaerobic membrane bioreactor treating low-strength domestic sewage-science direct[J]. Environmental Technology & Innovation, 2020, 19: 100844.

[11] ZHANG Y Y, GAO X, SMITH K, et al. Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network[J]. Water Research, 2019, 144: 114888.

[12] MOGHADDAM J S H A H. Membrane permeate flux and rejection factor prediction using intelligent systems [J]. Desalination, 2012: 284.

[13] SCHMITT F, BANU R, YEOM I T, et al. Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater[J]. Biochemical Engineering Journal, 2018, 133: 47-58.

- [14] GHANDEHARI S, MONTAZER-RAHMATI M M, ASGHARI M. A comparison between semi-theoretical and empirical modeling of cross-flow microfiltration using ANN[J]. *Desalination*, 2011, 277 (1-3): 348-355.
- [15] 徐小虎. 公共供水中纳滤膜污染行为的影响因素研究[D]. 兰州: 兰州交通大学, 2023.
- [16] DALTO M, MATUSKO J, VASAK M. Deep neural networks for ultra-short-term wind forecasting[C]. New York: IEEE - Conference Publishing Services, 2015.
- [17] VESELY K, GHOSHAL A, BURGET L, et al. Sequence-discriminative training of deep neural networks[C]// Proceedings of the Conference of the International Speech Communication Association. Lyon: Interspeech Press, 2013: 2345-2349.
- [18] IM S J, VIET N D, JANG A. Real-time monitoring of forward osmosis membrane fouling in wastewater reuse process performed with a deep learning model[J]. *Chemosphere*, 2021, 275: 130047.
- [19] LEE T M, OH H, CHOUNG Y K, et al. Prediction of membrane fouling in the pilot-scale microfiltration system using genetic programming[J]. *Desalination*, 2009, 247(1): 285-294.
- [20] XU L J, YANG Y X, YANG L. A new method for instantaneous frequency rate estimation of polynomial phase signal[C]. New York: IEEE Press, 2013.
- [21] 崔鸿雁, 徐帅, 张利锋, 等. 机器学习中的特征选择方法研究及展望[J]. *北京邮电大学学报*, 2018, 41(1): 1-12.
- [22] 朱一虎, 夏虹, 杨波, 等. 基于 CEEMDAN-GRU 的主泵电机绕组温度预测[J]. *应用科技*, 2023, 50(4): 14-20.
- [23] 董立朋, 聂清浩, 孙晓坤, 等. 基于皮尔逊相关系数法的盾构掘进参数对地表沉降影响分析[J]. *施工技术(中英文)*, 2024, 53(1): 116-23.
- [24] CUI L, GOODWIN C, GAO W, et al. Effect of cold water temperature on membrane structure and properties[J]. *Journal of Membrane Science*, 2017, 540: 19-26.
- [25] TIKKA A, GAO W, LIAO B. Reversibility of membrane performance and structure changes caused by extreme cold water temperature and elevated conditioning water temperature[J]. *Water Research*, 2019, 151: 260-270.
- [26] GU C, YAO J Q, WU Z Y, et al. Membrane fouling prediction of anaerobic membrane bioreactor based on BP neural network model optimized by principal component analysis[J]. *Advances in New and Renewable Energy*, 2022, 10(2): 95-102.
- [27] LEI Y, WANG Z C. Explore application of the error propagation law[J]. *Modern Property Management*, 2014.
- [28] 张贻婷, 李天宏. 基于长短时记忆神经网络的河流水质预测研究[J]. *环境科学与技术*, 2021, 44(8): 163-169.

(责任编辑: 赵冬艳)