



Machine learning-based analysis of calcium sulfate scaling-induced permeability decline in semi-batch reverse osmosis

Jungbin Kim^{a,b,*}, Haowen Zheng^a, Yiheng Yan^a, Weijian Jia^a, Linyinxue Dong^a, Daliang Xu^c, Kiho Park^d

^a Center for Sustainable Environmental and Ecosystem Research, Department of Environmental Science, College of Science, Mathematics and Technology, Wenzhou-Kean University, 88 Daxue Road, Ouhai, Wenzhou 325060, Zhejiang Province, China

^b Department of Environmental and Sustainability Sciences, The Dorothy and George Hennings College of Science, Mathematics and Technology, Kean University, 1000 Morris Avenue, Union 07083, NJ, United States

^c State Key Laboratory of Urban-Rural Water Resource and Environment, Harbin Institute of Technology, Harbin 150090, China

^d Department of Chemical Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

ARTICLE INFO

Editor name; Dr. Junyong Zhu

Keywords:

Semi-batch reverse osmosis

Water reuse

Calcium sulfate

Permeability decline

Machine learning

Process control

ABSTRACT

Semi-batch reverse osmosis (SBRO) increases water recovery while producing high-quality permeate for water reuse. However, SBRO operation is often constrained by calcium sulfate (CaSO_4) scaling during the treatment of industrial and municipal effluents for water reuse. Because permeability decline under scaling conditions arises from complex interactions between feed and hydrodynamic conditions, its interpretation remains challenging. This study analyzes CaSO_4 scaling-induced permeability decline in SBRO using machine learning (ML). A 7307×16 dataset comprising 12 experimental sets was obtained using a pilot-scale SBRO unit equipped with a 2.5-in. spiral-wound RO element and operated at 90% water recovery. Correlation analysis indicated that feed variables exhibited strong negative correlations (up to $r = -0.87$) with normalized temperature-corrected water permeability ($A_{TC}/A_{TC,0}$). In contrast, the effects of operating variables, such as feed hydraulic pressure (P_f), feed flow rate (Q_f), and permeate flow rate (Q_p), on $A_{TC}/A_{TC,0}$ were not clearly identified by correlation analysis because these variables were controlled discretely. To identify the main variables associated with CaSO_4 scaling-induced $A_{TC}/A_{TC,0}$ decline in SBRO, several ML models were trained without temporal variables. The Matern 5/2 Gaussian process regression model ($R^2 = 0.998$), coupled with Shapley additive explanations, indicated greater contributions of operating variables than feed variables to the predicted $A_{TC}/A_{TC,0}$ within the investigated dataset. These statistical relationships are consistent with the non-ideal flow behavior of the spiral-wound RO module and the dynamic operating characteristics of SBRO. Finally, increasing Q_f and Q_p was proposed as a process strategy to limit CaSO_4 scaling-induced $A_{TC}/A_{TC,0}$ decline in SBRO.

1. Introduction

Growing freshwater scarcity has increased reliance on nontraditional water sources. Accordingly, seawater desalination and water reuse have emerged as important alternatives [1]. Seawater desalination enhances water availability by producing freshwater from seawater, whereas water reuse enhances local water availability by reclaiming treated municipal or industrial wastewater. Although the specific energy consumption (SEC) of seawater reverse osmosis (SWRO) has decreased from 3.5–4.5 kWh/m³ to approximately 2.8 kWh/m³ in recent desalination plants [2,3], its energy demand remains significantly higher than that of

surface or brackish water treatment. Water reuse also relies on reverse osmosis (RO) to produce high-quality reclaimed water while presenting relatively lower SEC and reduced environmental burden [4]. However, the water recovery of RO is still constrained by flow design limitations and the progressive reduction in effective driving pressure along the membrane [5,6]. Consequently, high-recovery RO processes are critical for enhancing water production efficiency and reducing concentrate discharge in sustainable water reuse [7].

Semi-batch reverse osmosis (SBRO) is considered an emerging approach to increase the recovery of RO under brackish conditions [8,9]. In SBRO, a portion of the concentrate is pressurized and recycled to mix with the fresh feed, allowing for a single-stage configuration that

* Corresponding author at: Center for Sustainable Environmental and Ecosystem Research, Department of Environmental Science, College of Science, Mathematics and Technology, Wenzhou-Kean University, 88 Daxue Road, Ouhai, Wenzhou 325060, Zhejiang Province, China.

E-mail address: kjungbin@kean.edu (J. Kim).

<https://doi.org/10.1016/j.seppur.2026.139835>

Received 1 May 2026; Received in revised form 18 August 2026; Accepted 19 August 2026

Available online 20 August 2026

1383-5866/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Nomenclature			
Symbol	Description [Unit]		
A	Water permeability[L/m ² h bar]	Q_p	Permeate flow rate[L/h]
A/A_0	Normalized water permeability[–]	R	Ideal gas constant[L bar/mol K]
$A_{TC,0}$	Temperature-corrected water permeability[L/m ² h bar]	SI	Saturation index (in the bulk feed)[–]
$A_{TC}/A_{TC,0}$	Normalized temperature-corrected water permeability [–]	T	Feed temperature[K] or [°C]
A_m	Membrane area[m ²]	T_m	Feed temperature at the membrane[K] or [°C]
CF	Concentration factor[–]	T_t	Feed temperature in the tank[K] or [°C]
CPF	Concentration polarization facto[–]	TDS_f	Feed total dissolved solids[mg/L]
J_w	Water flux[L/m ² h]	V	System volume[L]
K_{sp}	Solubility product[mol ² /L ²]	Y	Element recovery[%]
M	Molecular weight[g/mol]	i	Effective ionization coefficient[–]
M_{CaCl_2}	Molecular weight for CaCl ₂ [g/mol]	k	EC-TDS conversion constant[(mg/L)/(μS/cm)]
$M_{Na_2SO_4}$	Molecular weight for Na ₂ SO ₄ [g/mol]	pH_f	Feed pH[–]
N_{cycles}	Number of cycles[–]	r	Pearson correlation coefficient[–]
P	Hydraulic pressure[bar]	t_{cycle}	A cycle time[min]
P_f	Feed hydraulic pressure[bar]	t_{op}	Operation time[min]
Q	Flow rate[L/h]	w	Mass fraction[–]
Q_c	Concentrate flow rate[L/h]	w_{CaCl_2}	Mass fraction for CaCl ₂ [–]
Q_f	Feed flow rate[L/h]	$w_{Na_2SO_4}$	Mass fraction for Na ₂ SO ₄ [–]
$Q_{f,fresh}$	Fresh feed flow rate[L/h]	κ_f	Feed electrical conductivity[μS/cm]
		κ_p	Permeate electrical conductivity[μS/cm]
		π	Osmotic pressure[bar]
		π_{CaCl_2}	Osmotic pressure for CaCl ₂ [bar]
		$\pi_{Na_2SO_4}$	Osmotic pressure for Na ₂ SO ₄ [bar]

concentrates the feed while continuously producing permeate [10]. The use of RO membranes in SBRO enables the production of high-quality permeate for water reuse by effectively removing dissolved solids, organic matter, and microorganisms [6,11]. Theoretically, SBRO can achieve a higher recovery with comparable or lower SEC than conventional RO [12]. Thus, SBRO can be utilized as a leading pathway toward high-recovery RO in water reuse [13]. For a more practical approach, a two-year SBRO demonstration treating RO concentrate achieved stable operation, extended clean-in-place intervals, and increased overall plant recovery from approximately 85% to 91% without compromising permeate quality [11]. A modified SBRO configuration has also been conceptualized and tested to enhance energy efficiency and operational performance [6,10].

However, operating SBRO at high recovery remains challenging due to scaling formation during the process [14] (Fig. 1). High-recovery operation with concentrate recirculation increases the concentration of sparingly soluble ions, leading to scaling once the ion activity product exceeds the solubility limit of the salt [15]. As the concentrate becomes more enriched, local supersaturation near the membrane surface further intensifies scaling risk. In SBRO, cyclic salinity rise toward the end of each concentration phase promotes rapid nucleation and crystal growth of salts, rendering scale control a critical factor for system stability and recovery optimization [16]. Moreover, flushing strategies influence concentration overshoot and filtration duration, thereby affecting scaling propensity in SBRO systems [17]. Scaling raises hydraulic

resistance and transmembrane pressure, accelerates flux decline, and reduces membrane lifespan, ultimately constraining both recovery and energy efficiency [18,19]. Even with antiscalant dosing, scaling remained the principal constraint on recovery [20].

Calcium sulfate (CaSO₄ or gypsum) is one of the major scaling minerals in high-recovery RO. Feed sources for water reuse, including industrial or municipal effluents, typically contain elevated Ca²⁺ and SO₄^{2−}, and CaSO₄ precipitation severely impairs RO system performance by decreasing water permeability (A) [21,22]. Experimental comparisons revealed shorter CaSO₄ induction times and higher nucleation rates in SBRO compared to RO at the same recovery, as its operation reaches higher transient supersaturation before flushing [23]. Understanding CaSO₄ scaling is essential to optimize antiscalant dosage and operational strategy while enabling sustainable high-recovery RO design [24]. However, scaling involves numerous nonlinear interactions among various variables, including feed conditions, hydrodynamics, and membrane properties.

Machine learning (ML) offers a systematic approach for identifying nonlinear relationships among variables associated with membrane performance degradation. It captures complex, nonlinear interactions, integrates multivariate datasets, and reveals hidden patterns in scaling behavior that exceed the capabilities of conventional numerical models [25]. ML can predict performance degradation and scaling tendencies in RO systems [26]. Previous studies have developed empirical and data-driven models to predict membrane performance decline in

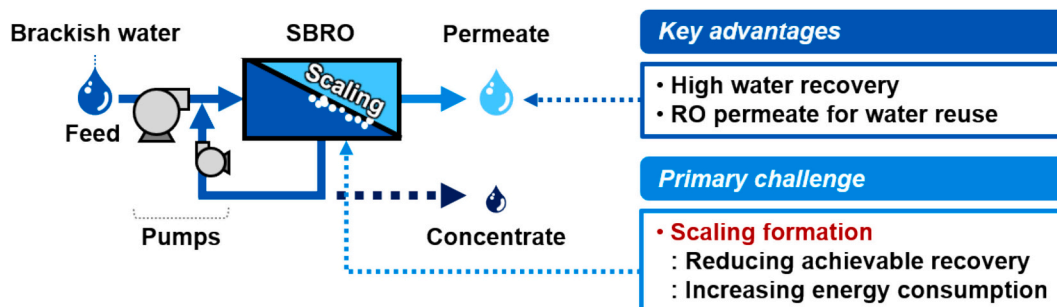


Fig. 1. Scheme of SBRO with its key advantages and principal challenge.

conventional RO systems. An empirical model was developed to predict the decline in water permeability using 10 years of operating data from a full-scale brackish water reverse osmosis (BWRO) plant [27]. A nonlinear recurrent neural network (RNN) with long short-term memory (LSTM) architecture was combined with a solution-diffusion model to predict long-term performance in a BWRO system [28]. However, previous studies focused on performance prediction under conventional RO conditions, while variable contributions under the dynamic conditions unique to SBRO remain largely unexplored. Although a recent study applied LSTM to SBRO for performance prediction and control, it did not address scaling-related interpretation [29]. To address this gap, interpretable ML using Shapley additive explanations (SHAP) provides an approach for identifying the contributions of individual variables to model predictions [30]. Accordingly, SHAP provides a suitable approach for interpreting these contributions under dynamic SBRO conditions.

This study analyzes CaSO_4 scaling-induced permeability decline in SBRO for high-recovery water reuse through ML interpretation. First, SBRO operational data are analyzed to examine the interactions among variables and clarify their mutual effects on normalized water permeability (A/A_0) and normalized temperature-corrected water permeability ($A_{TC}/A_{TC,0}$). The response of each variable to $A_{TC}/A_{TC,0}$ is identified, followed by a case analysis to examine the dominant variables. To address the nonlinear complexity and interdependence among variables, various ML models are evaluated for subsequent interpretation. SHAP analysis is then applied to identify the contributions of individual variables, and the effects of the most influential variables on $A_{TC}/A_{TC,0}$ are further investigated. Based on the results, potential SBRO process control strategies for limiting scaling-induced $A_{TC}/A_{TC,0}$ decline are proposed.

2. Materials and methods

2.1. SBRO experiment

2.1.1. SBRO system

The SBRO experiments were conducted using a pilot-scale RO system (Fig. 2) manufactured by Shanghai Langji Membrane Separation Equipment Engineering Co., Ltd. (Shanghai, China). The capacity of the system can reach up to 100 L/h (2.4 m³/d) depending on the permeate

flow rate (Q_p). The system enabled flexible control of SBRO operating conditions, facilitating its use for scaling studies. The system incorporated a programmable logic controller (PLC) for frequency regulation and data acquisition. All operating variables were automatically recorded at 11 s intervals and exported for subsequent ML interpretation. It was equipped with an integrated monitoring platform that continuously records pH, electrical conductivity (EC, κ), feed temperature (T), hydraulic pressure (P), and flow rate (Q). A high-pressure pump (HPP) with a variable frequency drive (VFD) ensured stable operation at constant permeate production under a maximum pressure of 50 bar. A single 2.5-in. spiral-wound RO membrane element (SW30-2540, Dupont) was installed. This membrane is suitable for both seawater and brackish water desalination, depending on the operational requirements. During SBRO operation, the fresh feed flow rate ($Q_{f,\text{fresh}}$) was supplied by a booster pump (BP) at the same rate as Q_p , thereby sustaining a system volume (V) of 4 L to represent the closed-circuit (CC) phase. The concentrate flow rate (Q_c) was recirculated to the feed tank without boosting P .

2.1.2. SBRO operation

The simulated ionic composition conditions were designed to approximate environments that promote CaSO_4 scaling under industrial or groundwater reuse conditions [31,32]. Calcium chloride anhydrous (CaCl_2 , $M_{\text{CaCl}_2} = 110.98$ g/mol) was dissolved to achieve a concentration of 138.5 mg/L, corresponding to 50 mg/L of Ca^{2+} , and sodium sulfate (Na_2SO_4 , $M_{\text{Na}_2\text{SO}_4} = 142.04$ g/mol) was added at 148 mg/L, yielding 100 mg/L of SO_4^{2-} . The solution was concentrated from an initial conductivity of approximately 610 $\mu\text{S}/\text{cm}$ to over 6000 $\mu\text{S}/\text{cm}$ using the SBRO system, corresponding to approximately 90% recovery. The pH of the feed solution was continuously adjusted to the desired value using sodium hydroxide (NaOH) throughout the experiments.

The SBRO experimental conditions were designed using Latin hypercube sampling (LHS) to achieve broad coverage of the variable space while minimizing sample clustering. This balanced design enabled systematic evaluation of operating variables on system performance. 12 experimental sets were generated to investigate the influence of multiple variables on system performance (Table 1). Q_p was controlled between 35 and 60 L/h, corresponding to a specific flux of 12.50–21.43 L/m² h. Higher Q_p led to fewer data samples because the target concentration was reached more rapidly [33]. The HPP frequency was adjusted

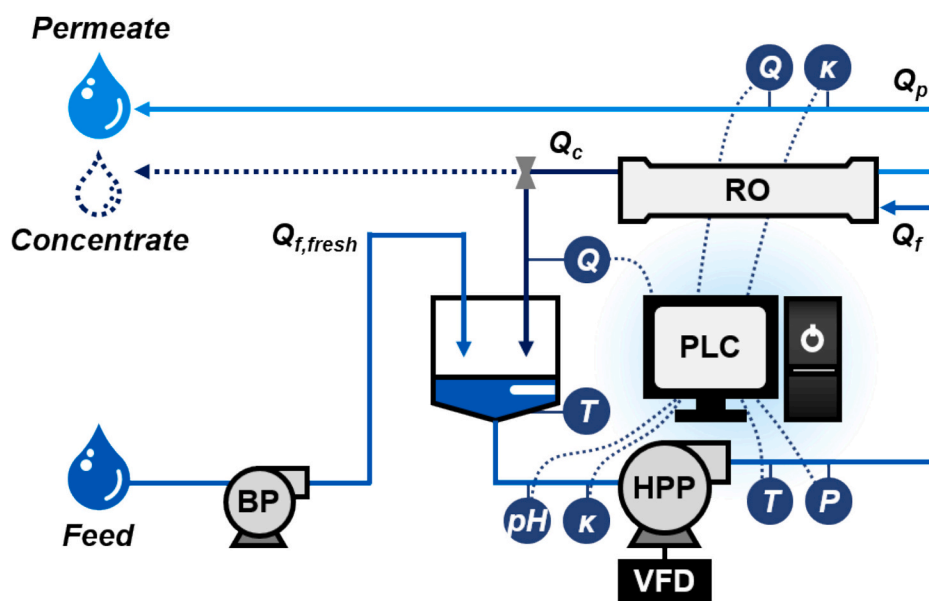


Fig. 2. A pilot-scale RO system used for SBRO operation. Q_c was recirculated within the SBRO system during the CC phases (solid line) and discharged once the target concentration was reached (dotted line).

Table 1
Experimental sets of SBRO with controllable input variables by LHS.

Experimental set	Q_p [L/h]	Frequency of HPP [Hz]	pH_f [–]	Number of data samples
1	36	40	8.9	849
2	38	39	9.8	869
3	41	33	7.9	816
4	51	30	9.2	583
5	47	36	9.3	549
6	43	36	7.8	589
7	55	34	9.7	460
8	48	32	8.8	524
9	42	38	8.3	805
10	53	35	8.1	460
11	57	31	8.6	465
12	59	37	9.5	338
Total				7307

between 30 and 40 Hz, affecting the feed flow rate (Q_f) into the RO module. The feed pH (pH_f) [–] was regulated between 7.6 and 10.0, in addition to studying the role of pH in CaSO_4 scaling development. The SBRO system automatically adjusted the feed hydraulic pressure (P_f) through the PLC to maintain a constant permeate production rate. It should be noted that the system was cleaned before and after each experiment to maintain consistent performance and to prevent any influence from residual solutes. Consequently, the temperature-corrected water permeability (A_{TC}) was maintained at $1.73 \pm 0.06 \text{ L/m}^2 \text{ h bar}$.

2.2. Variable calculation

2.2.1. Osmotic pressure

EC reflects the ionic strength of an aqueous solution and is commonly used as an indicator of total dissolved solids (TDS). However, a conversion process is required to estimate TDS and osmotic pressure (π) from EC values. Although the conversion constant may vary depending on concentration, the correlation between the feed TDS (TDS_f) [mg/L] and feed EC (κ_f) [$\mu\text{S/cm}$] obtained during solution preparation yielded a constant k of 0.47 [(mg/L)/($\mu\text{S/cm}$)] (Eq. (1)).

$$TDS_f = k \times \kappa_f \quad (1)$$

π [bar] of each solute was determined using the van't Hoff equation ($\pi = iCRT$), where the effective ionization coefficient (i) [–] was taken as 2.7, the ideal gas constant (R) as $8.314 \times 10^{-2} \text{ L bar/mol K}$, and T as 298 K. Given that the mass fractions (w) [–] contributing to TDS were 0.483 for CaCl_2 and 0.517 for Na_2SO_4 , π of CaCl_2 and Na_2SO_4 were calculated according to Eq. (2) and Eq. (3), respectively.

$$\pi_{\text{CaCl}_2} = i \times \frac{TDS_f \times w_{\text{CaCl}_2}}{10^3 \times M_{\text{CaCl}_2}} \times R \times T \approx 2.91 \times 10^{-4} \times TDS_f \quad (2)$$

$$\pi_{\text{Na}_2\text{SO}_4} = i \times \frac{TDS_f \times w_{\text{Na}_2\text{SO}_4}}{10^3 \times M_{\text{Na}_2\text{SO}_4}} \times R \times T \approx 2.43 \times 10^{-4} \times TDS_f \quad (3)$$

The feed π (π_f) [bar] was obtained by summing the individual π of the solutes, which enabled the estimation of π directly from κ_f as described in Eq. (4).

$$\pi_f = \pi_{\text{CaCl}_2} + \pi_{\text{Na}_2\text{SO}_4} = 5.34 \times 10^{-4} \times TDS_f = 2.51 \times 10^{-4} \times \kappa_f \quad (4)$$

2.2.2. Saturation index

The saturation index (SI) in the bulk feed was considered the primary indicator for scaling formation. SI for CaSO_4 can be calculated using Eq. (5), where K_{sp} [mol^2/L^2] is the solubility product of CaSO_4 .

$$SI = \log \left(\frac{[\text{Ca}^{2+}][\text{SO}_4^{2-}]}{K_{sp}} \right) \quad (5)$$

The concentration of Ca^{2+} and SO_4^{2-} can be expressed as a function of κ_f in Eqs. (6)–(7). The logarithm of the ion product is expressed in Eq.

(8).

$$[\text{Ca}^{2+}] = \frac{TDS_f \times w_{\text{CaCl}_2}}{10^3 \times M_{\text{CaCl}_2}} = 2.05 \times 10^{-6} \times \kappa_f \quad (6)$$

$$[\text{SO}_4^{2-}] = \frac{TDS_f \times w_{\text{Na}_2\text{SO}_4}}{10^3 \times M_{\text{Na}_2\text{SO}_4}} = 1.71 \times 10^{-6} \times \kappa_f \quad (7)$$

$$\log([\text{Ca}^{2+}][\text{SO}_4^{2-}]) = \log(3.51 \times 10^{-12} \times \kappa_f^2) = 2\log\kappa_f - 11.45 \quad (8)$$

Although K_{sp} varies with T , the change in $\log K_{sp}$ within the SBRO operational range is negligible. Therefore, the K_{sp} of CaSO_4 at 25 °C, taken as $2.4 \times 10^{-5} \text{ mol}^2/\text{L}^2$, was used without accounting for T [34]. By combining Eqs. (5) and (8), the SI can be expressed as shown in Eq. (9).

$$SI = \log([\text{Ca}^{2+}][\text{SO}_4^{2-}]) - \log K_{sp} = 2\log\kappa_f - 6.83 \quad (9)$$

2.2.3. Derived variables

While several variables were directly recorded from the PLC system, others were derived through subsequent calculations. These calculated variables offered additional insight into the ML interpretation. Concentration factor (CF) [–] was defined as the ratio of κ_f to its initial value, $\kappa_{f,0}$ (Eq. (10)).

$$CF = \frac{\kappa_f}{\kappa_{f,0}} \quad (10)$$

A cycle time (t_{cycle}) [min] was determined based on V [L] and Q_c [L/h] in Eq. (11). Also, the number of cycles (N_{cycles}) [–] was determined by dividing the operation time (t_{op}) [min] by t_{cycle} (Eq. (12)) and rounding the result up to the nearest integer.

$$t_{\text{cycle}} = \frac{60V}{Q_c} \quad (11)$$

$$N_{\text{cycles}} = \frac{t_{\text{op}}}{t_{\text{cycle}}} = \frac{t_{\text{op}}Q_c}{V} \quad (12)$$

Q_f [L/h] was calculated as the sum of Q_p [L/h] and Q_c (Eq. (13)). It should be noted that $Q_{f,\text{fresh}}$ [L/h] was equal to Q_p in SBRO system (Eq. (14)). Water flux (J_w) [$\text{L/m}^2 \text{ h}$] was calculated by dividing Q_p by the membrane area (A_m) [m^2] (Eq. (15)), where A_m is 2.8 m^2 for a single 2.5-in. RO element. Using the element recovery (Y) [%] in Eq. (16), the concentration polarization factor (CPF) [–] was calculated using Eq. (17) [35]. A was calculated using Eq. (18) based on the solution-diffusion model, where P and π of the permeate are neglected. The calculated A was used for the A/A_0 [–] calculation, where A_0 is the initial A value for each experiment. In addition, pressure drop across the membrane module was not considered because a single RO element was used. However, pressure drop can become significant in SBRO systems when multiple RO membrane modules are installed in series within a pressure vessel.

$$Q_f = Q_p + Q_c \quad (13)$$

$$Q_{f,\text{fresh}} = Q_p \quad (14)$$

$$J_w = \frac{Q_p}{A_m} \quad (15)$$

$$Y = \frac{Q_p}{Q_f} \times 100 \quad (16)$$

$$CPF = \exp \left(0.7 \times \frac{Y}{100} \right) \quad (17)$$

$$A = \frac{J_w}{\Delta P - CPF \times \Delta \pi} \approx \frac{J_w}{P_f - 2.5 \times 10^{-4} \times \exp\left(0.7 \times \frac{Y}{100}\right) \times \kappa_f} \quad (18)$$

However, the temperature of the SBRO system increased during operation due to heat generated by the HPP, which could lead to misinterpretation of A . Therefore, a temperature correction factor (TCF) [–] was employed as defined in Eq. (19), using the temperature at the RO membrane (T_m) rather than that of the tank (T_i). The ratio of membrane activation energy to the gas constant (e/R) was assumed to be 2640 K for $T_m > 25$ °C, and 3020 K for $T_m \leq 25$ °C [29]. A_{TC} was then calculated from A and TCF using Eq. (20).

$$TCF = \exp\left[\frac{e}{R} \left(\frac{1}{298} - \frac{1}{273 + T_m}\right)\right] \quad (19)$$

$$A_{TC} = A/TCF \quad (20)$$

2.3. ML interpretation

2.3.1. Regression models

With a 7307×16 dataset prepared (Appendix A), ML-based regression analysis and related statistical evaluations were conducted using the Regression Learner App in MATLAB 2025a. This approach can identify nonlinear correlations between input and output variables and assess model performance in terms of predictive accuracy and interpretability.

Since the most suitable model for interpreting the given dataset was unknown, several regression models were applied for comparative analysis. The models included basic linear regression, fine tree, and support vector machine (SVM) variants, such as quadratic SVM (QSVM), medium Gaussian SVM (MGSM), and kernel-approximation methods applied to SVM kernel. In addition, neural network (NN) models (e.g., medium neural network (MNN)) and Gaussian process regression (GPR) variants, including Matern 5/2 GPR (M5/2GPR) and rational quadratic GPR (RQGPR), were employed to capture complex nonlinear relationships.

Tree models approximate nonlinear relationships through recursive partitioning of the input space into regions fitted with simple local models (e.g., piecewise constant or linear), with complexity controlled by tree depth, minimum leaf size, and pruning [36]. SVM learns a real-valued function in a high-dimensional feature space using an ϵ -insensitive loss function and kernel mapping. GPR captures smooth probabilistic relationships with quantifiable uncertainty, whereas NN approximates nonlinear mappings through the optimization of layered weights [37]. While several input variables are interdependent, SHAP analysis was applied to quantify the relative contributions of individual variables to model predictions [38].

2.3.2. ML model evaluation and selection

Selecting an appropriate ML model is critical for reliable SHAP and partial dependence plot (PDP) analyses because different algorithms vary in their ability to capture nonlinear relationships and handle noisy or limited datasets [39]. Therefore, comparing multiple ML models enables the selection of an algorithm that provides the best balance between predictive accuracy, generalization, and interpretability [40].

As the dataset consisted of temporally correlated operational data for $A_{TC}/A_{TC,0}$, random 5-fold cross-validation could overestimate predictive performance. To address this limitation, leave-one-dataset-out (LODO) evaluation was performed by excluding one complete experimental dataset from model training and reserving it exclusively for testing. This strategy provides a more stringent assessment of model generalization to unseen operating conditions [41]. For example, experimental datasets 1–11 were used for training, whereas experimental dataset 12 was reserved for testing (Appendix A; Table S1–S10). The LODO evaluation was also conducted without temporal variables (e.g., t_{cycle} and N_{cycles}),

thereby reducing the influence of time-dependent information (Appendix A; Table S11–S19). This further minimized potential information leakage and enabled a more reliable evaluation of model robustness and interpretability.

Table 2 summarizes the predictive performance of the regression ML models for $A_{TC}/A_{TC,0}$ under both 5-fold cross-validation and LODO evaluation without temporal variables. Model performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Despite the removal of temporal variables, the two GPR variants, M5/2GPR and RQGPR, achieved the highest predictive performance ($R^2 = 0.998 \pm 0.000$) under 5-fold cross-validation. In contrast, LODO evaluation yielded lower predictive performance and greater variability (M5/2GPR: $R^2 = 0.539 \pm 0.431$; RQGPR: $R^2 = 0.536 \pm 0.440$). Lower predictive performance was occasionally observed when the held-out dataset contained operating-condition combinations absent from the training datasets (Appendix A; Tables S11 and S12). In particular, the relatively large standard deviations reflected poor predictions for experimental datasets 3 and 5, which strongly affected the variability across the 12 LODO iterations.

These results indicate that the developed models accurately predicted SBRO performance within the investigated dataset but exhibited reduced generalization to unseen experimental datasets with different operating-condition combinations. Given that the objective of the SHAP and PDP analyses was to interpret variable effects within the investigated dataset rather than to evaluate model generalization, the ML models trained on the complete dataset without temporal variables were used (Appendix A; Table S19). Accordingly, the SHAP and PDP results can be interpreted within the investigated data range.

3. Results and discussion

3.1. Variable correlation

Unlike the steady-state operation of RO, the variables in SBRO are inherently dynamic. The constantly changing interdependence among feed conditions, hydrodynamics, and thermodynamic driving forces ultimately affects A/A_0 . Identifying these relationships is necessary to understand the variables associated with permeability decline under scaling conditions. The Pearson correlation matrix captured pairwise linear associations and highlighted the dominant linkages, as indicated by the Pearson correlation coefficients (r) (Fig. 3). The variables exhibited higher correlations with $A_{TC}/A_{TC,0}$ than with A/A_0 because the TCF corrected temperature-induced changes in membrane permeability.

Table 2

Summary of 5-fold cross-validation and LODO evaluation performance of the regression ML models without temporal variables. Values are presented as the mean $\pm$ SD across the 12 LODO iterations.

ML models	5-fold cross-validation			LODO evaluation		
	R^2 [–]	RMSE [–]	MAE [–]	R^2 [–]	RMSE [–]	MAE [–]
M5/2GPR	0.998 $\pm$ 0.000	0.003 $\pm$ 0.000	0.002 $\pm$ 0.000	0.539 $\pm$ 0.431	0.038 $\pm$ 0.021	0.033 $\pm$ 0.018
RQGPR	0.998 $\pm$ 0.000	0.003 $\pm$ 0.000	0.002 $\pm$ 0.000	0.536 $\pm$ 0.440	0.038 $\pm$ 0.021	0.033 $\pm$ 0.018
MNN	0.996 $\pm$ 0.000	0.004 $\pm$ 0.000	0.003 $\pm$ 0.000	–0.067 $\pm$ 1.047	0.056 $\pm$ 0.031	0.051 $\pm$ 0.032
Fine tree	0.990 $\pm$ 0.001	0.007 $\pm$ 0.001	0.004 $\pm$ 0.000	0.468 $\pm$ 0.454	0.041 $\pm$ 0.018	0.036 $\pm$ 0.016
MGSM	0.991 $\pm$ 0.001	0.006 $\pm$ 0.001	0.005 $\pm$ 0.000	0.464 $\pm$ 0.401	0.043 $\pm$ 0.021	0.038 $\pm$ 0.017
SVM	0.982 $\pm$ 0.004	0.009 $\pm$ 0.001	0.006 $\pm$ 0.000	–1.500 $\pm$ 1.542	0.092 $\pm$ 0.031	0.076 $\pm$ 0.024
QSVM	0.908 $\pm$ 0.029	0.020 $\pm$ 0.003	0.015 $\pm$ 0.003	–0.792 $\pm$ 2.633	0.076 $\pm$ 0.071	0.071 $\pm$ 0.071
Linear	0.879 $\pm$ 0.014	0.023 $\pm$ 0.002	0.018 $\pm$ 0.001	0.580 $\pm$ 0.440	0.035 $\pm$ 0.020	0.032 $\pm$ 0.019

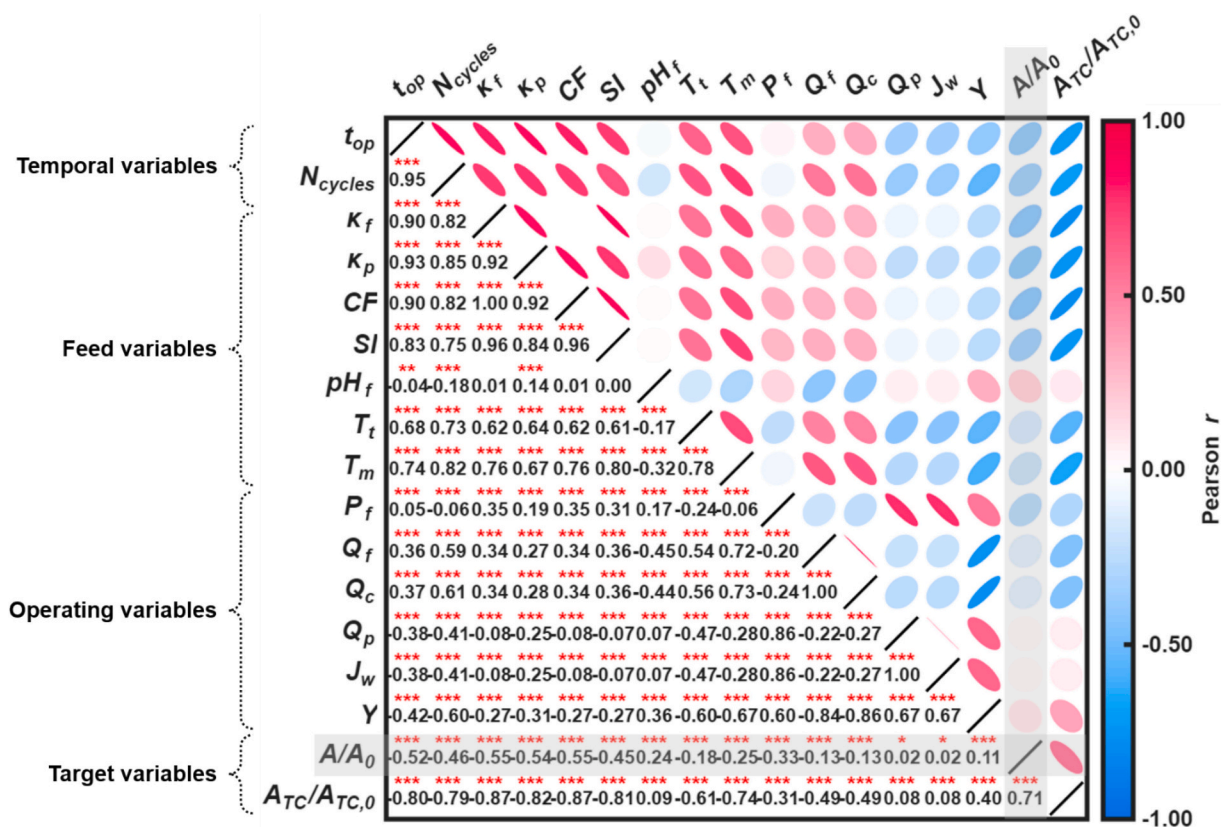


Fig. 3. Pearson correlation matrix of the variables. The correlations of variables with A/A_0 and $A_{TC}/A_{TC,0}$ were compared to identify the more appropriate indicator.

Higher temperature enhances A_{TC} by reducing water viscosity and increasing water diffusivity across the membrane [42]. Therefore, $A_{TC}/A_{TC,0}$ was used as a more appropriate indicator of scaling behavior.

The temporal variables were strongly positively correlated with one another. The correlation between t_{op} and N_{cycles} was 0.95 because N_{cycles} is a linear function of t_{op} , as defined in Eq. (12). As t_{op} increased, the feed variables also exhibited strong positive correlations, including κ_f ($r = 0.90$), permeate EC (κ_p ; $r = 0.93$), CF ($r = 0.90$), and SI ($r = 0.83$). Despite the strong correlations among the temporal variables, their correlations with $A_{TC}/A_{TC,0}$ were slightly lower, with r values of approximately -0.80 .

Among the feed variables, the strongest correlations with $A_{TC}/A_{TC,0}$ were observed for κ_f ($r = -0.87$) and CF ($r = -0.87$). SI also exhibited a high negative correlation with $A_{TC}/A_{TC,0}$ ($r = -0.81$), consistent with the widely accepted role of SI in predicting scaling [36], although this correlation was slightly weaker than those of κ_f or CF . This correlation could be further improved by considering the SI at the membrane surface rather than in the bulk feed. T exhibited moderate correlations with $A_{TC}/A_{TC,0}$ ($r = -0.61$ and -0.74 for T_t and T_m , respectively). T increased during the initial stage due to heat input from the HPP and stabilized as thermal equilibrium was reached. In contrast, pH_f exhibited little correlation with $A_{TC}/A_{TC,0}$, because $CaSO_4$ scaling is relatively insensitive to pH_f within the investigated range, unlike calcium carbonate ($CaCO_3$) scaling [43].

Notably, the operating variables exhibited relatively weak correlations with $A_{TC}/A_{TC,0}$. P_f exhibited a strong positive correlation with Q_p and J_w ($r = 0.86$), as P_f governs permeate transport, but only a weak negative correlation with $A_{TC}/A_{TC,0}$ ($r = -0.31$). In contrast, flow rates exhibited moderate to negligible correlations with $A_{TC}/A_{TC,0}$. Q_f and Q_c were moderately correlated ($r = -0.49$), whereas Q_p and J_w displayed negligible correlations ($r = 0.08$). Although P_f and J_w were strongly correlated ($r = 0.86$), only P_f was weakly correlated with $A_{TC}/A_{TC,0}$, whereas the correlation for J_w was negligible.

Overall, feed variables exhibited a strong inverse relationship with $A_{TC}/A_{TC,0}$ when TCF was applied, whereas operating variables exhibited moderate to negligible correlations. However, $A_{TC}/A_{TC,0}$ reflects complex interactions among osmotic-pressure buildup, scaling kinetics, and hydrodynamic factors, suggesting that pairwise correlations alone are insufficient to characterize the nonlinear relationships underlying the decline of $A_{TC}/A_{TC,0}$.

3.2. Variable response

The decline in $A_{TC}/A_{TC,0}$ would result from complex and interdependent interactions among multiple input variables. To gain deeper insight into these relationships, a detailed trend analysis of each variable with respect to $A_{TC}/A_{TC,0}$ was conducted. The plots displaying monotonic and parabolic patterns support the view that multiple competing effects develop and evolve during operation (Fig. 4).

While A/A_0 exhibited a convex response to CF due to the initial temperature increase during SBRO operation, $A_{TC}/A_{TC,0}$ exhibited a clear declining trend (Fig. 4a). In the early stage of SBRO, heat generated from the HPP increased T_m and significantly increased A . As the operation time was extended to achieve higher CF , the system approached a stabilized temperature. In other words, A/A_0 did not increase monotonically with CF because of the concentration process and temperature variation. Consequently, temperature correction enabled a clearer representation of the declining trend in $A_{TC}/A_{TC,0}$ ($r = -0.87$). Moreover, at higher CF , osmotic pressure buildup reduced the effective driving pressure and enhanced boundary-layer supersaturation, which promoted the nucleation of sparingly soluble salts [44].

The response of $A_{TC}/A_{TC,0}$ to SI formed a slight parabolic pattern and exhibited an apparent decline when SI was positive (Fig. 4b). At undersaturated conditions ($SI < 0$), $CaSO_4$ precipitation is thermodynamically unfavorable, and kinetics remain inhibited. Although SI became positive, precipitation did not begin immediately because of the

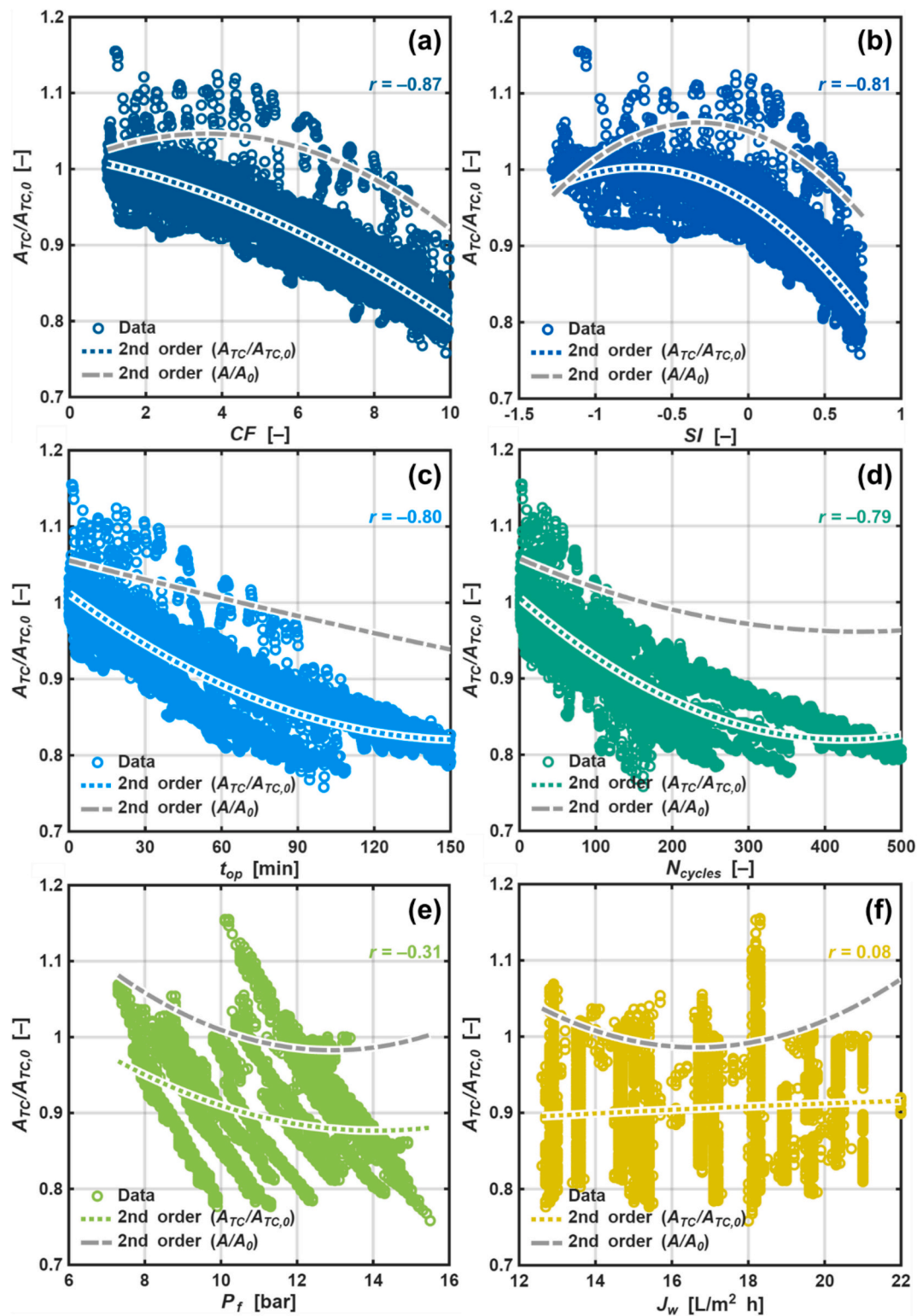


Fig. 4. Response of $A_{TC}/A_{TC,0}$ to variables in SBRO operation: (a) CF , (b) SI , (c) t_{op} , (d) N_{cycles} , (e) P_f and (f) J_w . The second-order decline in A/A_0 became more pronounced after temperature correction, as reflected by $A_{TC}/A_{TC,0}$.

induction time, and $A_{TC}/A_{TC,0}$ did not decline dramatically. Once SI increased further and crystallization kinetics became dominant, heterogeneous nucleation developed on the membrane surface, resulting in a rapid decrease in $A_{TC}/A_{TC,0}$ [44]. The current trend ($r = -0.81$) would be expected to become stronger if the SI at the membrane surface were considered.

$A_{TC}/A_{TC,0}$ progressively decreased for both t_{op} ($r = -0.80$) and N_{cycles}

($r = -0.79$) due to the cumulative effects of concentration polarization (CP) and surface scaling during extended operation (Fig. 4c and d). Since N_{cycles} is inherently linked to t_{op} via Eq. (12), both variables exhibited similar responses. This observation is consistent with previous reports that prolonged operation promotes RO scaling [44]. Also, $A_{TC}/A_{TC,0}$ exhibited a concave downward trend with respect to both t_{op} and N_{cycles} (Fig. 4c and d), unlike CF and SI (Fig. 4a and b). This behavior arises

because t_{op} and N_{cycles} are temporal variables, whereas concentration dynamics govern CF and SI . At low CF or SI , the SBRO system requires a longer time to reach highly concentrated conditions, whereas the residence time under such conditions is relatively short.

Operating variables such as P_f and J_w presented weak correlations with $A_{TC}/A_{TC,0}$, with $r = -0.31$ and $r = 0.08$, respectively (Fig. 4e and f). Unlike feed variables, operating variables were discretely controlled during SBRO. Consequently, their individual effects were not fully reflected in the correlation analysis. Therefore, further analysis was required to isolate variable interactions and clarify their contributions to $A_{TC}/A_{TC,0}$.

To gain a more comprehensive understanding of the influences on $A_{TC}/A_{TC,0}$ in SBRO, 200 representative data points corresponding to high and low $A_{TC}/A_{TC,0}$ values were analyzed. This case-based approach identified variables that enhance or diminish $A_{TC}/A_{TC,0}$. Higher or lower $A_{TC}/A_{TC,0}$ values were associated with specific ranges of the variables (Fig. 5). Feed variables such as κ_f , CF , and SI exhibited more distinct ranges, whereas operating variables displayed overlapping ranges across both $A_{TC}/A_{TC,0}$ groups. Although variables within specific ranges appeared to dominate SBRO operational sensitivity, further interpretability analyses are required to clarify their contributions.

3.3. SHAP analysis

Data-driven interpretations should be supported by physical reasoning and established physicochemical knowledge. Interpretable ML techniques, such as SHAP, can complement traditional case-based analyses by quantifying the marginal contribution of each variable to $A_{TC}/A_{TC,0}$, while accounting for nonlinear interactions and coupled dependencies. This approach enables a more systematic and objective evaluation of variable importance within the investigated dataset, clarifying overlapping influences that cannot be discerned through conventional descriptive analyses. Among the ML models trained on the complete dataset without temporal variables (Section 2.3.2), M5/2GPR and MNN exhibited the highest R^2 values (Appendix A; Table S19) and were therefore selected for comparison.

Operating variables exhibited greater contributions to the predicted $A_{TC}/A_{TC,0}$ than feed variables in the SHAP analysis. In the M5/2GPR model, P_f and Y emerged as the most influential contributors to the predicted $A_{TC}/A_{TC,0}$ (Fig. 6a). Higher P_f and lower Y were associated with lower SHAP values, indicating negative contributions to $A_{TC}/A_{TC,0}$. Since Y is a function of Q_f and Q_p , these flow rates also emerged as important contributors. Because Q_p is proportional to J_w , its high SHAP importance reflects the corresponding contribution of J_w to the model prediction. Furthermore, operating variables accounted for 85.5% of the

total SHAP importance, whereas feed variables accounted for only 14.5%. These results indicate greater contributions of operating variables than feed variables to the predicted $A_{TC}/A_{TC,0}$ within the investigated dataset [45,46].

The greater SHAP contributions of operating variables are consistent with the non-ideal flow behavior associated with the spiral-wound RO module configuration (Fig. 6a). In spiral-wound modules, surface crystallization is favored under high-pressure and low-crossflow-velocity conditions [47,48]. Higher P_f and lower Q_f were associated with lower SHAP values, following trends reported for scaling behavior. High P_f may also reflect cumulative transport effects, including residual solute retention and incomplete flushing [49,50]. Moreover, axial dispersion redistributes solute along the flow direction, moderating localized supersaturation and influencing scaling propensity [50,51]. In contrast, feed variables were derived from bulk feed conditions and therefore may not accurately represent membrane-surface concentrations [52]. These reported physicochemical phenomena provide context for interpreting the statistical relationships identified by SHAP, although processes such as nucleation and crystal growth cannot be directly determined from the analysis.

The MNN model yielded a similar result, although its interpretation was less consistent with the physical behavior of the SBRO system (Fig. 6b). P_f remained the most influential variable, followed by Q_p , Y , and Q_c . However, several SHAP relationships were not physically consistent. In particular, CF exhibited SHAP effects opposite to those of κ_f , despite both variables representing the progression of concentration. Moreover, the MNN model exhibited negative R^2 values in the LODO evaluation (Table 2), indicating poor generalization to unseen experimental datasets. Accordingly, SHAP results should be interpreted alongside domain knowledge and within the range of the investigated dataset [53].

3.4. SHAP dependence

The SHAP dependence plot from the M5/2GPR model indicated consistent relationships of P_f and Y with SHAP values, regardless of the SI . The SHAP value for P_f decreased from 0.26 to -0.23 as P_f increased from 7.5 to 14.3 bar (Fig. 7a). The association of higher P_f with lower SHAP values may be related to enhanced CP and local solute supersaturation under elevated pressure conditions. However, the SI in the bulk feed exhibited limited relevance, as high SI at low P_f did not correspond to a reduction in SHAP value. In contrast, the SHAP value increased from -0.18 to 0.26 as Y rose from 4.2% to 12.0% (Fig. 7b). Higher Y corresponded to greater $Q_{f, fresh}$, which delayed the rise in solute concentration due to feed dilution.

Both Q_f and Q_p exhibited positive relationships with SHAP values, while the effect of SI was negligible. As Q_f increased from 425.1 to 862.5 L/h, the SHAP value rose from -0.10 to 0.07 (Fig. 7c). This trend may be associated with increased crossflow velocity at higher Q_f , which can alleviate CP and local supersaturation near the membrane surface [54,55]. Similarly, Q_p showed a positive correlation, with SHAP values increasing from -0.06 to 0.05 as Q_p rose from 35.7 to 59.1 L/h (Fig. 7d). This association was also consistent with the operational characteristics of SBRO observed in the experimental data.

Overall, P_f and Y were identified as the most influential variables contributing to the predicted $A_{TC}/A_{TC,0}$. Lowering P_f while increasing Y may help limit $A_{TC}/A_{TC,0}$ decline under the investigated conditions. Moreover, since Q_f and Q_p are linked to Y , flow control should be further evaluated as a potential strategy for limiting permeability decline. At the same time, because a reduction in Q_c during the CC phase has been reported to lower SEC, increasing Y is also beneficial for energy savings [33].

3.5. SBRO process control strategies

Based on the SHAP results, lowering P_f and increasing Y were

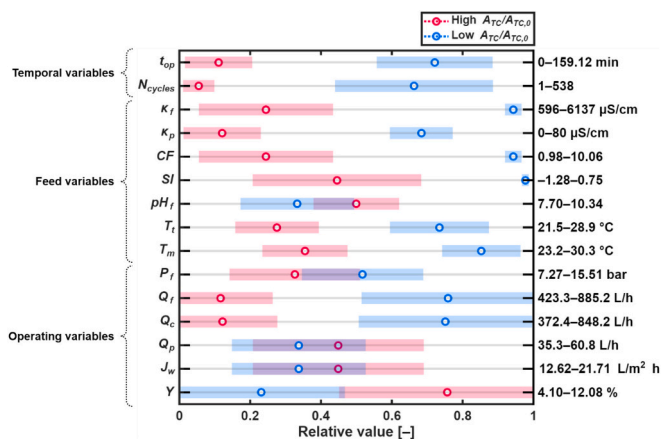


Fig. 5. Range of the relative values of each input variable under high and low $A_{TC}/A_{TC,0}$ conditions. Means are presented, with standard deviations (SD) indicating variability. Actual value ranges are provided on the right y-axis.

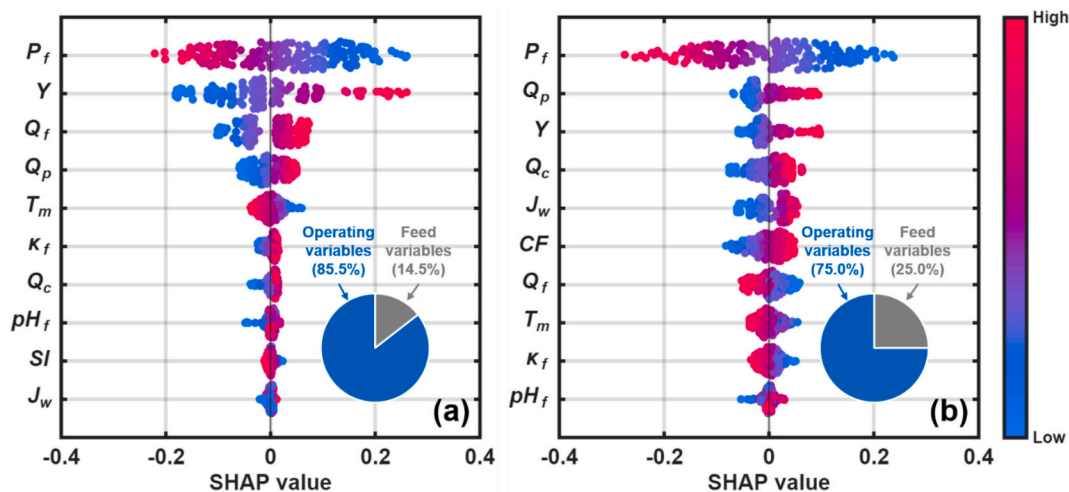


Fig. 6. SHAP analysis of feature importance in system operation toward $A_{TC}/A_{TC,0}$ prediction using different ML models: (a) M5/2GPR and (b) MNN. The operating variables contributed significantly more to the total SHAP importance than the feed variables.

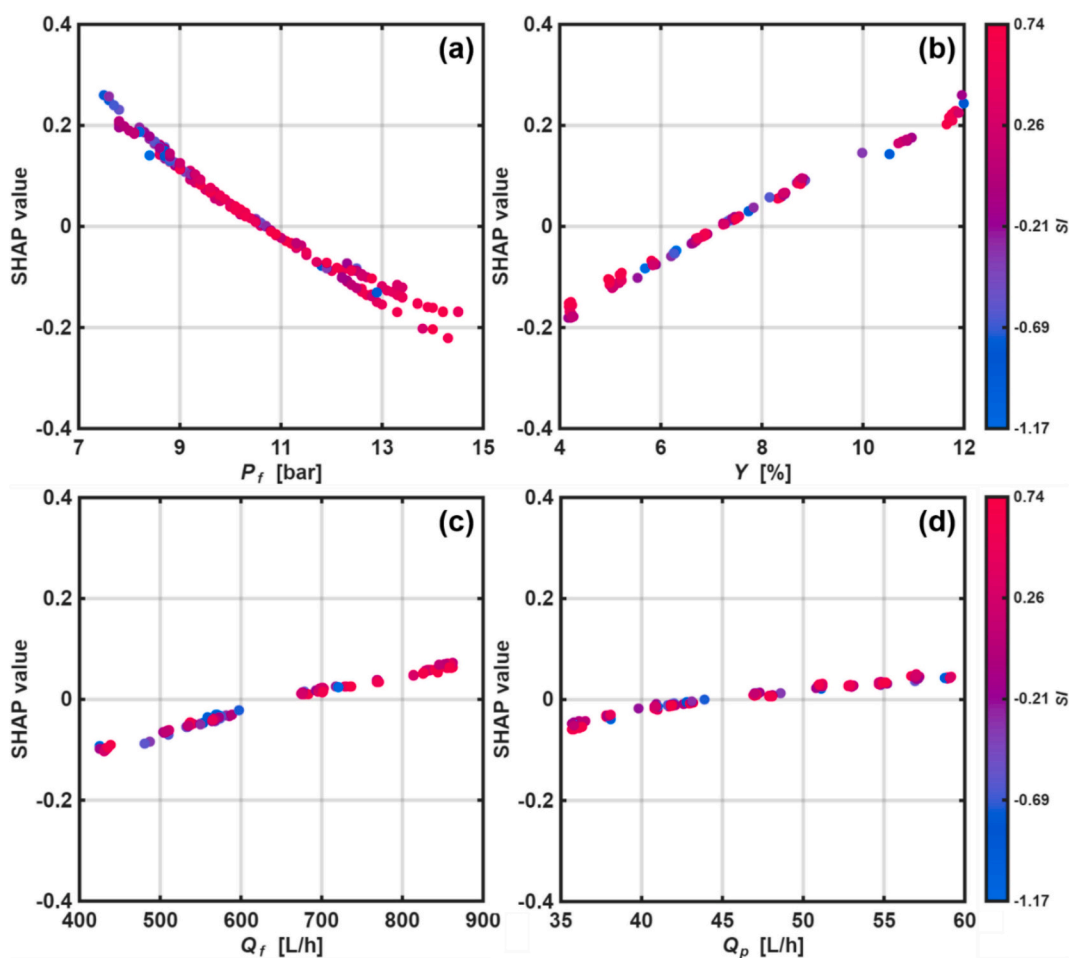


Fig. 7. SHAP dependence plots of four key operating variables on $A_{TC}/A_{TC,0}$ based on the M5/2GPR model. The color bar indicates the intensity of SI.

considered potential operational strategies to limit the scaling-induced decline in membrane permeability. This can be achieved by controlling the operating pressure or applying RO membranes with higher A . However, in practice, operating pressure is typically adjusted to maintain target recovery and flow conditions, whereas membrane replacement is infrequent due to system downtime and replacement costs. On the other hand, the flow rate can be directly adjusted during operation.

Accordingly, non-redundant and controllable variables were selected to evaluate their influence through a PDP analysis.

Increasing Q_f can contribute to maintaining $A_{TC}/A_{TC,0}$ during the rise of P_f (Fig. 8a). In the SBRO system, P_f inevitably increases to counteract the rise in π_f and sustain the net driving pressure. Although P_f is a key variable in the decline of $A_{TC}/A_{TC,0}$, reducing it during SBRO operation remains challenging. An increase in Q_f enhances crossflow velocity,

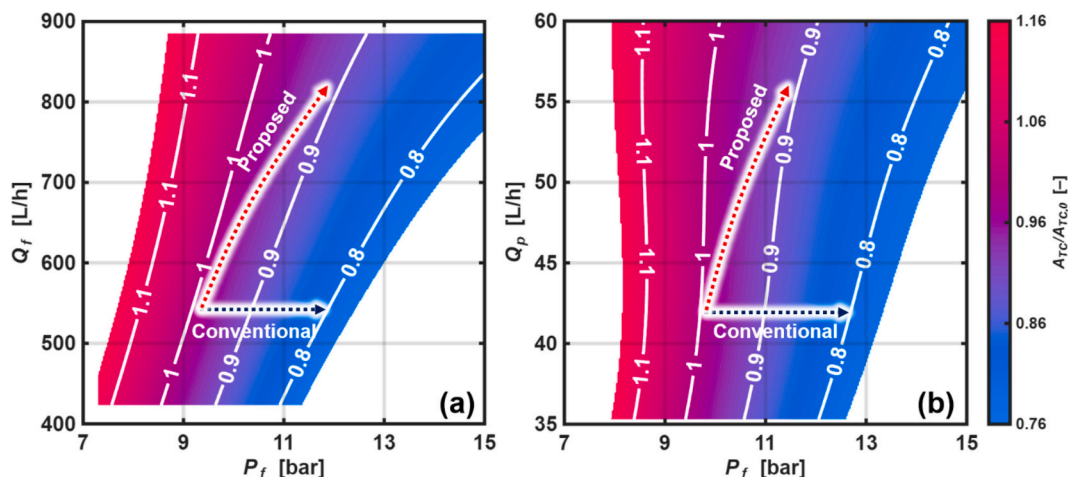


Fig. 8. PDP on $A_{TC}/A_{TC,0}$: (a) P_f and Q_f ; (b) P_f and Q_p . All variables except the PDP variable were fixed at their mean values, and only values within the operational range were evaluated. Conventional and proposed strategies are indicated.

which can reduce scaling-induced $A_{TC}/A_{TC,0}$ decline [46,47]. Therefore, increasing Q_f could be considered when $A_{TC}/A_{TC,0}$ declines as P_f rises. However, N_{cycles} may increase as Q_c increases with Q_f , while Q_p is held fixed (Eq. (13)), thereby elevating the pressure drop during the concentration phase and increasing the SEC.

With increasing Q_f , elevating Q_p during the concentration phase can be an effective strategy (Fig. 8b). A higher Q_p improves $A_{TC}/A_{TC,0}$ by enhancing $Q_{f,fresh}$ according to the operating principle of SBRO, which promotes internal dilution and moderates the rate of concentration increase [33]. Although a rise in Q_p necessitates a corresponding increase in P_f , this operational adjustment shifts the trajectory diagonally upward rather than horizontally toward lower $A_{TC}/A_{TC,0}$, thereby reducing degradation of membrane performance. Under these conditions, concentration can progress while limiting the predicted decline in $A_{TC}/A_{TC,0}$.

This flow-control strategy could limit scaling-induced $A_{TC}/A_{TC,0}$ decline in SBRO. Initially, Q_f should be increased, followed by a further increase in Q_p . The rate of increase in Q_p must exceed that of Q_f to achieve a higher Y during the CC phase. This model-derived operational strategy could provide a basis for future optimization of SBRO operation under scaling-prone conditions. Furthermore, greater scaling accumulation under high-recovery or prolonged operation may require higher Q_f and Q_p . However, such adjustments should be optimized within the hydraulic and flow design constraints of the SBRO system. Nevertheless, these strategies do not fully resolve fundamental issues such as residual solute on the membrane and CP. Additional mitigation may be achieved through pretreatment, antiscalant, and membrane surface modification [56–58]. However, the optimal mitigation strategy depends on the type of membrane scaling.

4. Conclusions

To enable high-recovery SBRO for water reuse, this study employed ML to analyze CaSO_4 scaling-induced permeability decline. Operational data were systematically collected during the pilot-scale SBRO operation under brackish water conditions. Variable interactions were analyzed, ML-based interpretation was performed, and operational strategies for SBRO were proposed. The main findings are outlined below:

- With temperature correction, feed variables exhibited strong negative correlations (up to $r = -0.87$) with $A_{TC}/A_{TC,0}$. On the other hand, operating variables displayed moderate or weak associations ($|r| < 0.5$). The relatively low $A_{TC}/A_{TC,0}$ was associated with complex hydrodynamic interactions that were not captured by correlation

analysis. These findings indicate that correlation analysis alone was insufficient to explain the nonlinear behavior of $A_{TC}/A_{TC,0}$.

- Operating variables exhibited greater SHAP contributions to the predicted $A_{TC}/A_{TC,0}$ than feed variables under the investigated conditions. The M5/2GPR model identified P_f and Y as the most influential contributors to the predicted $A_{TC}/A_{TC,0}$, followed by Q_f and Q_p . The MNN model also selected P_f , Q_p , Y , and Q_c as important variables, although several SHAP relationships were not physically consistent. Using the M5/2GPR model, SHAP dependence analysis revealed that increasing P_f decreased SHAP values, whereas increasing Y , Q_p , and Q_f increased them.
- As P_f is determined by feed and operating conditions, flow rate control can help limit scaling-induced $A_{TC}/A_{TC,0}$ decline in SBRO. Under elevated pressure conditions, increasing Q_f can reduce the decline in $A_{TC}/A_{TC,0}$. Likewise, increasing Q_p can help maintain $A_{TC}/A_{TC,0}$. Achieving a higher Y requires that the rate of increase in Q_p exceed that of Q_f .

Although the analysis focused on CaSO_4 scaling under specific system conditions, it provides insights into the relative statistical contributions of operating and feed variables within the investigated dataset. Further validation across broader SBRO configurations and operating conditions is required to assess the applicability of these relationships and the practical effectiveness of the proposed operational strategies. With larger, continuous operational datasets, temporal deep learning models, such as LSTMs, could further improve model generalization.

CRedit author statement

Jungbin Kim: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition. **Haowen Zheng:** Investigation, Data Curation, Writing - Original Draft. **Yiheng Yan:** Data Curation, Writing - Original Draft. **Weijian Jia:** Investigation, Project administration. **Linyinxue Dong:** Methodology, Project administration. **Daliang Xu:** Data Curation, Writing - Review & Editing. **Kiho Park:** Data Curation, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was financially supported by the National Natural Science Foundation of China (W2433147).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.seppur.2026.139835>.

Data availability

The experimental data are available in Appendix A.

References

- [1] R.J. Ohlund, B.H. Dahdah, G.R. Guillen, A.E. Childress, Augmenting Ocean water desalination with potable reuse: concept feasibility in terms of cost and environmental impacts, *Desalination* 569 (2024) 116941.
- [2] J. Kim, K. Park, D.R. Yang, S. Hong, A comprehensive review of energy consumption of seawater reverse osmosis desalination plants, *Appl. Energy* 254 (2019) 113652.
- [3] C. Wang, L. Wang, L. Dong, H.K. Shon, J. Kim, Specific energy consumption of seawater reverse osmosis desalination plants using machine learning, *Desalination* 602 (2025) 118654.
- [4] J. Heffron, M. Samsami, S. Juedemann, J. Lavin, S. Tavakoli Nick, B.A. Kieke, B. K. Mayer, Mitigation of viruses of concern and bacteriophage surrogates via common unit processes for water reuse: a meta-analysis, *Water Res.* 252 (2024) 121242.
- [5] A. Ruiz-García, I. Nuez, M.D. Carrascosa-Chisvert, J.J. Santana, Simulations of BWRO systems under different feedwater characteristics. Analysis of operation windows and optimal operating points, *Desalination* 491 (2020) 114582.
- [6] Y. Lee, W. Kang, W. Lee, J.H. Lee, J. Kim, Improving water flux distribution and energy efficiency in reverse osmosis for high-recovery water reuse, *Desalination* 585 (2024) 117779.
- [7] J. Li, W. Cheng, H. Wang, Y. Luo, Q. Liu, X. Wang, L. Wang, T. Zhang, Reverse osmosis and nanofiltration processes in industrial wastewater treatment: the recent progress, challenge, and future opportunity, *Sep. Purif. Technol.* 362 (2025) 131687.
- [8] Z. Mo, D. Li, Q. She, Semi-closed reverse osmosis (SCRO): a concise, flexible, and energy-efficient desalination process, *Desalination* 544 (2022) 116147.
- [9] B. Sutariya, H. Raval, Analytical study of optimum operating conditions in semi-batch closed-circuit reverse osmosis (CCRO), *Sep. Purif. Technol.* 264 (2021) 118421.
- [10] Y. Lee, W. Kang, W. Lee, M.J. Park, J. Kim, Pilot study of circle sequence reverse osmosis for high-recovery water reuse in industry using tertiary effluent, *Desalination* 600 (2025) 118480.
- [11] H. Gu, M.H. Plumlee, M. Boyd, M. Hwang, J.C. Lozier, Operational optimization of closed-circuit reverse osmosis (CCRO) pilot to recover concentrate at an advanced water purification facility for potable reuse, *Desalination* 518 (2021) 115300.
- [12] D.M. Warsinger, E.W. Tow, K.G. Nayar, L.A. Maswadeh, Energy efficiency of batch and semi-batch (CCRO) reverse osmosis desalination, *Water Res.* 106 (2016) 272–282.
- [13] J. Kim, L. Dong, H.K. Shon, K. Park, Current progress in semi-batch reverse osmosis for brackish water desalination, *Desalination* 578 (2024) 117434.
- [14] J.B. Roman, A.J.B. Kemperman, W.G.J. van der Meer, J.A. Wood, Closed circuit reverse osmosis and adsorption: PFAS removal and impact on scaling behaviour, *Sep. Purif. Technol.* 393 (2026) 137295.
- [15] W. Yu, D. Song, W. Chen, H. Yang, Antiscalants in RO membrane scaling control, *Water Res.* 183 (2020) 115985.
- [16] L. Burlace, P.A. Davies, Fouling and fouling mitigation in batch reverse osmosis: review and outlook, *Desalin. Water Treat.* 249 (2022) 1–22.
- [17] M. Li, Role of flushing in closed-circuit RO: more than membrane regeneration, *Desalination* 634 (2026) 120285.
- [18] A. Matin, F. Rahman, H.Z. Shafi, S.M. Zubair, Scaling of reverse osmosis membranes used in water desalination: phenomena, impact, and control; future directions, *Desalination* 455 (2019) 135–157.
- [19] N. Franco-Clavijo, S. Cespedes, A. Farinha, G.-J. Witkamp, C. Picioreanu, J. S. Vrouwenvelder, B. Blankert, Optical coherence tomography for early detection, visualization and characterization of growth and deposition-driven scaling in reverse osmosis, *Sep. Purif. Technol.* 378 (2025) 134664.
- [20] M. Futterlieb, I.M.A. ElSherbiny, M. Tuczinski, J. Lipnizki, S. Panglisch, Limits of high recovery inland desalination: closed-circuit reverse osmosis – a viable option? *Chem. Ing. Tech.* 93 (2021) 1359–1368.
- [21] W. Yu, D. Song, A. Li, H. Yang, Control of gypsum-dominated scaling in reverse osmosis system using carboxymethyl cellulose, *J. Membr. Sci.* 577 (2019) 20–30.
- [22] Z. Dai, Y. Zhao, S. Paudyal, X. Wang, C. Dai, S. Ko, W. Li, A.T. Kan, M.B. Tomson, Gypsum scale formation and inhibition kinetics with implications in membrane system, *Water Res.* 225 (2022) 119166.
- [23] T. Lee, J.Y. Choi, Y. Cohen, Gypsum scaling propensity in semi-batch RO (SBRO) and steady-state RO with partial recycle (SSRO-PR), *J. Membr. Sci.* 588 (2019) 117106.
- [24] Y. Yin, S. Kalam, J.L. Livingston, R. Minjarez, J. Lee, S. Lin, T. Tong, The use of anti-scalants in gypsum scaling mitigation: comparison with membrane surface modification and efficiency in combined reverse osmosis and membrane distillation, *J. Membr. Sci.* 643 (2022) 120077.
- [25] S. Park, S.S. Baek, J. Pyo, Y. Pachepsky, J. Park, K.H. Cho, Deep neural networks for modeling fouling growth and flux decline during NF/RO membrane filtration, *J. Membr. Sci.* 587 (2019) 117164.
- [26] Y. Choi, Y. Lee, K. Shin, Y. Park, S. Lee, Analysis of long-term performance of full-scale reverse osmosis desalination plant using artificial neural network and tree model, *Environ. Eng. Res.* 25 (2020) 763–770.
- [27] A. Ruiz-García, I. Nuez, Long-term performance decline in a brackish water reverse osmosis desalination plant. Predictive model for the water permeability coefficient, *Desalination* 397 (2016) 101–107.
- [28] D. Gaubomme, W. Quaghebeur, A. Van Droogenbroeck, M. Vanoppen, B. De Gussemme, A. Verliefde, I. Nopens, E. Torfs, A hybrid modelling approach for reverse osmosis processes including fouling, *Desalination* 564 (2023) 116756.
- [29] F. Abdullah, K. Villez, S. Allu, Machine learning modeling and model predictive control of a closed-circuit reverse osmosis system, *Chem. Eng. Res. Des.* 222 (2025) 29–47.
- [30] Y.Y. Liang, M. Li, Unsteady-state dynamics and AI in membrane desalination: challenges and emerging opportunities, *Desalination* 618 (2026) 119507.
- [31] J. Rolf, T. Cao, X. Huang, C. Boo, Q. Li, M. Elimelech, Inorganic scaling in membrane desalination: models, mechanisms, and characterization methods, *Environ. Sci. Technol.* 56 (12) (2022) 7484–7511.
- [32] E. Alhseinat, M. Abi Jaoude, A. Alkathheeri, I. Darawshah, A. Safieh, Insights into the composite scale formation and coprecipitation behavior of CaSO₄ and SrSO₄ at different salinity level, *Surf. Interfaces* 22 (2021) 100875.
- [33] L. Dong, Y. Song, Y. Liu, M. Xie, D. Mu, J. Kim, Sensitivity analysis of design variables in semi-batch reverse osmosis for energy-efficient water treatment, *J. Clean. Prod.* 541 (2026) 147553.
- [34] L. Shen, H. Sippola, X. Li, D. Lindberg, P. Taskinen, Thermodynamic modeling of calcium sulfate hydrates in the CaSO₄-H₂O system from 273.15 to 473.15 K with extension to 548.15 K, *J. Chem. Eng. Data* 64 (2019) 2697–2709.
- [35] M. Li, J. Li, A closed-form analytical model for estimating permeate salinity in reverse osmosis systems, *J. Membr. Sci.* 735 (2025) 124526.
- [36] L. Breiman, J. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Chapman and Hall/CRC, 2017.
- [37] V.L. Deringer, A.P. Bartók, N. Bernstein, D.M. Wilkins, M. Ceriotti, G. Csányi, Gaussian process regression for materials and molecules, *Chem. Rev.* 121 (16) (2021) 10073–10141.
- [38] Z. Zhao, M. Wu, D. Lu, Z. Kong, Z. Ye, L. Liang, L. Zhang, A physics-informed machine learning framework for understanding the rejection mechanism of hydrophilic micropollutants in polyamide membranes, *J. Membr. Sci.* 740 (2026) 124925.
- [39] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (2020) 56–67.
- [40] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nat. Mach. Intell.* 1 (2019) 206–215.
- [41] Y. Okazaki, S. Okazaki, S. Asamoto, T. Yamaji, M. Ishige, Estimator for generalization performance of machine learning model trained by biased data collected from multiple references, *Comput. Aided Civ. Infrastruct. Eng.* 38 (2023) 2145–2162.
- [42] C.P. Koutsou, E.M. Kritikos, A.J. Karabelas, M. Kostoglou, Analysis of temperature effects on the specific energy consumption in reverse osmosis desalination processes, *Desalination* 476 (2020) 114213.
- [43] M. Futterlieb, S. Panglisch, Incipient CaCO₃-scaling detection for scaling mitigation, *Desalination* 613 (2025) 119023.
- [44] M. Philibert, L.O. Villacorte, Y. Ekowati, A. Abushaban, S.G. Salinas-Rodriguez, Fouling and scaling in reverse osmosis desalination plants: a critical review of membrane autopsies, feedwater quality guidelines and assessment methods, *Desalination* 592 (2024) 118188.
- [45] M. Li, Do membrane transport properties vary under dynamic and cyclic RO? Evidence for constant permeability and hysteresis from channel history, *Desalination* 620 (2026) 119644.
- [46] M. Nassr, S.M. Dischinger, J.Y. Lee, K.L. Gleason, S. Molins, N. Spycher, P. Bigdelou, J. Johnston, N.A. Lynd, N. Tilton, W.T. Stringfellow, B.D. Freeman, D. J. Miller, Mineral scale formation during crossflow reverse osmosis at constant flux and constant transmembrane pressure conditions, *Ind. Eng. Chem. Res.* 64 (2025) 1295–1308.
- [47] A. Antony, J.H. Low, S. Gray, A.E. Childress, P. Le-Clech, G. Leslie, Scale formation and control in high pressure membrane water treatment systems: a review, *J. Membr. Sci.* 383 (2011) 1–16.
- [48] S. Lee, J. Kim, C.H. Lee, Analysis of CaSO₄ scale formation mechanism in various nanofiltration modules, *J. Membr. Sci.* 163 (1999) 63–74.
- [49] M. Li, Cyclic simulation and energy assessment of closed-circuit RO (CCRO) of brackish water, *Desalination* 545 (2023) 116149.
- [50] T. Lee, A. Rahardianto, Y. Cohen, Multi-cycle operation of semi-batch reverse osmosis (SBRO) desalination, *J. Membr. Sci.* 588 (2019) 117090.
- [51] M. Li, An improved closed-circuit RO (CCRO) system: design and cyclic simulation, *Desalination* 554 (2023) 116519.

- [52] W. Bai, L. Samineni, P. Chirontoni, I. Krupa, P. Kasak, A. Popelka, N.B. Saleh, M. Kumar, Quantifying and reducing concentration polarization in reverse osmosis systems, *Desalination* 554 (2023) 116480.
- [53] Y. Song, J. Moon, K. Park, Explainable AI for permeate flux prediction in forward osmosis: SHAP interpretability and theoretical validation for enhanced predictive reliability, *Desalination* 601 (2025) 118551.
- [54] E.W. Tow, Q.J. Wei, A.R. Abraham, K.L. Chua, M.J. Plumley, J.H. Lienhard, Piloting batch reverse osmosis with a flexible bladder for water recovery from scaling-prone brine, *npj clean Water* 8 (2025) 30.
- [55] Q. Liu, G.-R. Xu, R. Das, Inorganic scaling in reverse osmosis (RO) desalination: mechanisms, monitoring, and inhibition strategies, *Desalination* 468 (2019) 114065.
- [56] Z. Cao, Y. Hu, H. Zhao, B. Cao, P. Zhang, Sulfate mineral scaling: from fundamental mechanisms to control strategies, *Water Res.* 222 (2022) 118945.
- [57] Y. Zhao, L. Gao, N. Chang, J. Hou, Z. Zhang, Y. Dong, Y. Wang, H. Wang, Advancing antifouling strategies for reverse osmosis membranes: insights into fouling mechanisms and innovative fabrication approaches for membranes, *Sep. Purif. Technol.* 390 (2026) 137015.
- [58] W. Yuan, X. Chen, Z. Yu, Y. Wan, J. Lin, W. Ye, Critical review of membrane fouling in reverse osmosis treatment: characterizations, models, mechanisms, and controls, *Sep. Purif. Technol.* 363 (2025) 132119.