

Supplementary material:

An automatic machine learning model-based approach for predicting reverse osmosis membrane flux in the context of scale inhibitors: Coupling accurate classification of influent and fast screening of the model

Support material

In this experiment, the RO feed water pretreatment is shown in flow chart S1. It can be seen that the RO pretreatment process has a heat exchanger. Therefore, the effect of influent water temperature on flux was not further analyzed in this experiment. Meanwhile, the influent water is pre-treated with ultrafiltration, so it is also free of suspended particles. The scale inhibitor dosing point is set before the security filter and the inlet water sampling point is chosen after the scale inhibitor dosing, so the concentration of scale inhibitor dosing can be effectively monitored. During the data selection process, each data set was selected on the basis of stable operation of the flux for more than 3 days under the corresponding operating parameters.

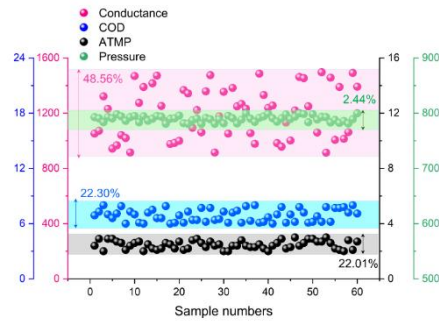
The description of the sampling locations and the distribution of water quality parameters for the four influent waters are shown in Table S1:

Table S1 Influent sampling points and parameter distribution

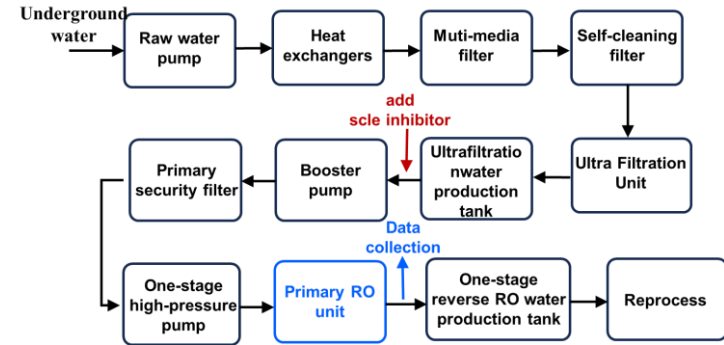
Influent type	Parameter distribution	Fluctuation dimension	Process
---------------	------------------------	-----------------------	---------



Underground water



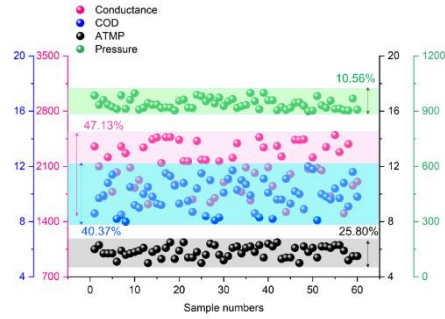
1



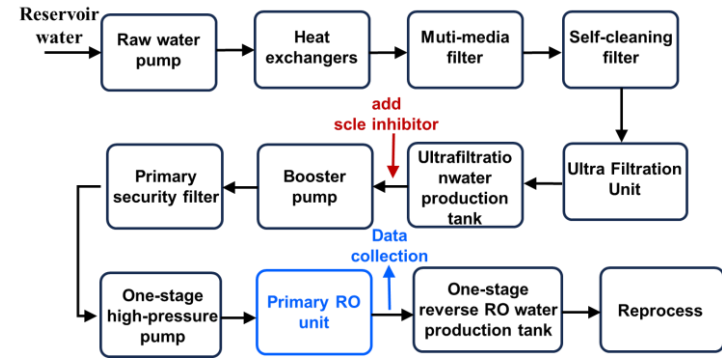
Underground water: The water sample of this experiment is taken from a well in Shandong, China, where the underground water quality is relatively stable. The distribution of different parameters is shown in the figure. Through calculation, the FC values of conductivity, COD, pressure and scale inhibitor concentration are 48.56%, 22.30%, 22.01% and 2.44%, respectively. It can be seen that the deep underground water fluctuation here is one-dimensional parameter fluctuation.



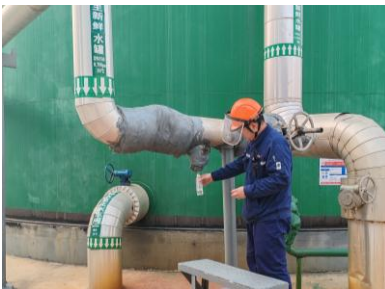
Reservoir water



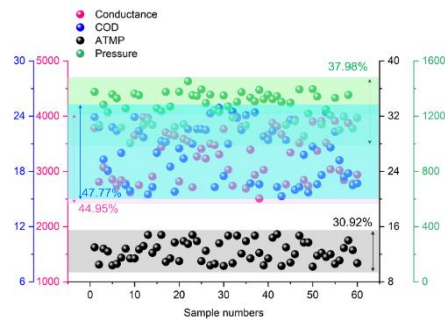
2



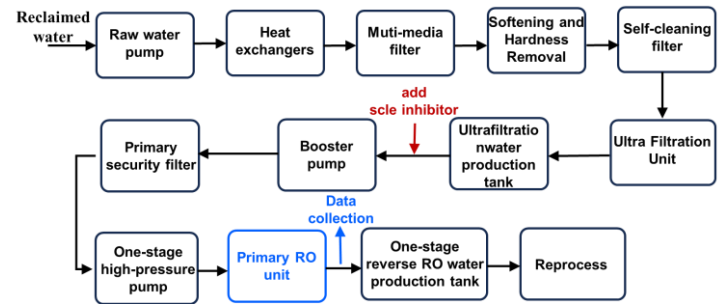
Reservoir water: The water sample is from a reservoir in Shandong, China. The distribution of different parameters is shown in the figure. Through calculation, the FC values of conductivity, COD, pressure and scale inhibitor concentration are 47.13%, 40.37%, 25.80% and 10.56%, respectively. Compared with the groundwater quality, the fluctuation range of water quality at this place is significantly improved. The fluctuation amplitude of conductivity and COD exceeds the set value, so the water quality fluctuation at the water intake point is a two-dimensional parameter fluctuation.



Reclaimed water



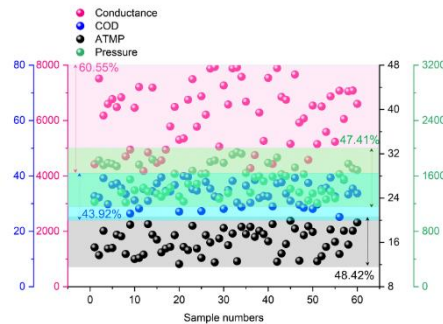
3



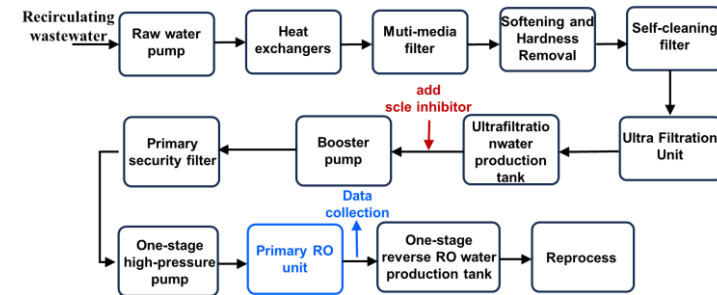
Reclaimed water: The water sample is taken from the $2 \times 100 \text{ m}^3/\text{h}$ reclaimed water reuse station of Zhonggu mining in Inner Mongolia, China. The distribution of different parameters is shown in the figure. The calculation shows that the FC values of conductivity, COD, pressure and scale inhibitor concentration are 44.95%, 47.77%, 37.98% and 30.92%, respectively. The fluctuation ranges of conductivity, COD and scale inhibitor dosing concentration in the water quality parameters at this location exceed the set values, so the water quality fluctuation at this location is a three-dimensional parameter fluctuation.



Recirculating wastewater



4

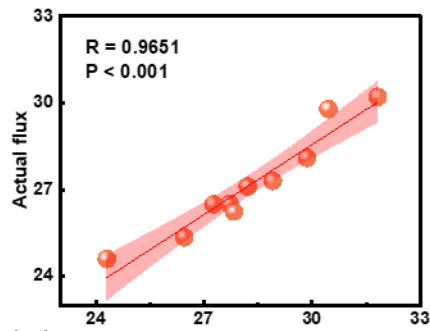


Recirculating wastewater: The water sample is taken from Shaanxi coal to hydrogen recycling station in China. The distribution of different parameters is shown in the figure. The calculation shows that the FC values of conductivity, COD, pressure and scale inhibitor concentration are 60.55%, 43.92%, 48.42% and 47.41%, respectively. Due to the complexity and fluctuation of the components in the circulating water blowdown water, all the four parameters under this condition exceed the range setting limit, so this type of influent is summarized as multi-dimensional parameter fluctuation in this experiment.

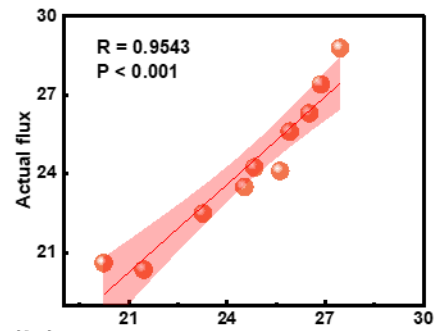
Table S2 Performance Metrics of the Selected Model Under 10-Fold Cross-Validation

Model	Fold	R2	RMSE
GBM	1	0.9775	0.0442
	2	0.9789	0.0435
	3	0.9802	0.0431
	4	0.9768	0.0448
	5	0.9781	0.0439
	6	0.9773	0.0443
	7	0.9785	0.0437
	8	0.9796	0.0432
	9	0.9769	0.0450
	10	0.9788	0.0436
	Mean		0.9783
Std		0.0011	0.0006
ANN	1	0.9541	0.0630
	2	0.9562	0.0618
	3	0.9549	0.0625
	4	0.9537	0.0634
	5	0.9560	0.0619
	6	0.9551	0.0624
	7	0.9557	0.0620
	8	0.9543	0.0628
	9	0.9564	0.0617
	10	0.9556	0.0621
	Mean		0.9552
Std		0.0009	0.0005
RF	1	0.9692	0.0528
	2	0.9711	0.0519
	3	0.9698	0.0525
	4	0.9686	0.0530
	5	0.9705	0.0522
	6	0.9700	0.0526
	7	0.9694	0.0527
	8	0.9709	0.0520
	9	0.9713	0.0517
	10	0.9701	0.0523
	Mean		0.9701
Std		0.0009	0.0004
SVR	1	0.9463	0.0578
	2	0.9501	0.0562
	3	0.9487	0.0583
	4	0.9515	0.0559
	5	0.9498	0.0565

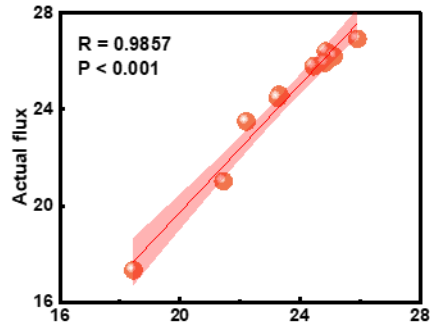
6	0.9472	0.0576
7	0.9480	0.0572
8	0.9510	0.0560
9	0.9491	0.0567
10	0.9504	0.0561
Mean	0.9492	0.0568
Std	0.0015	0.0007



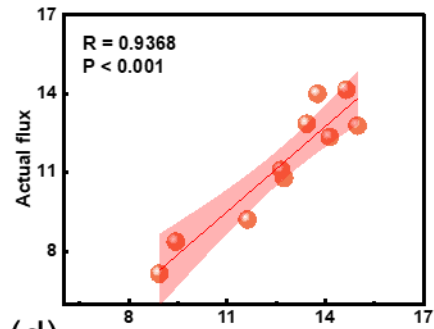
(a)



(b)



(c)



(d)

Fig.S1 Predicted values and actual values fitting curve

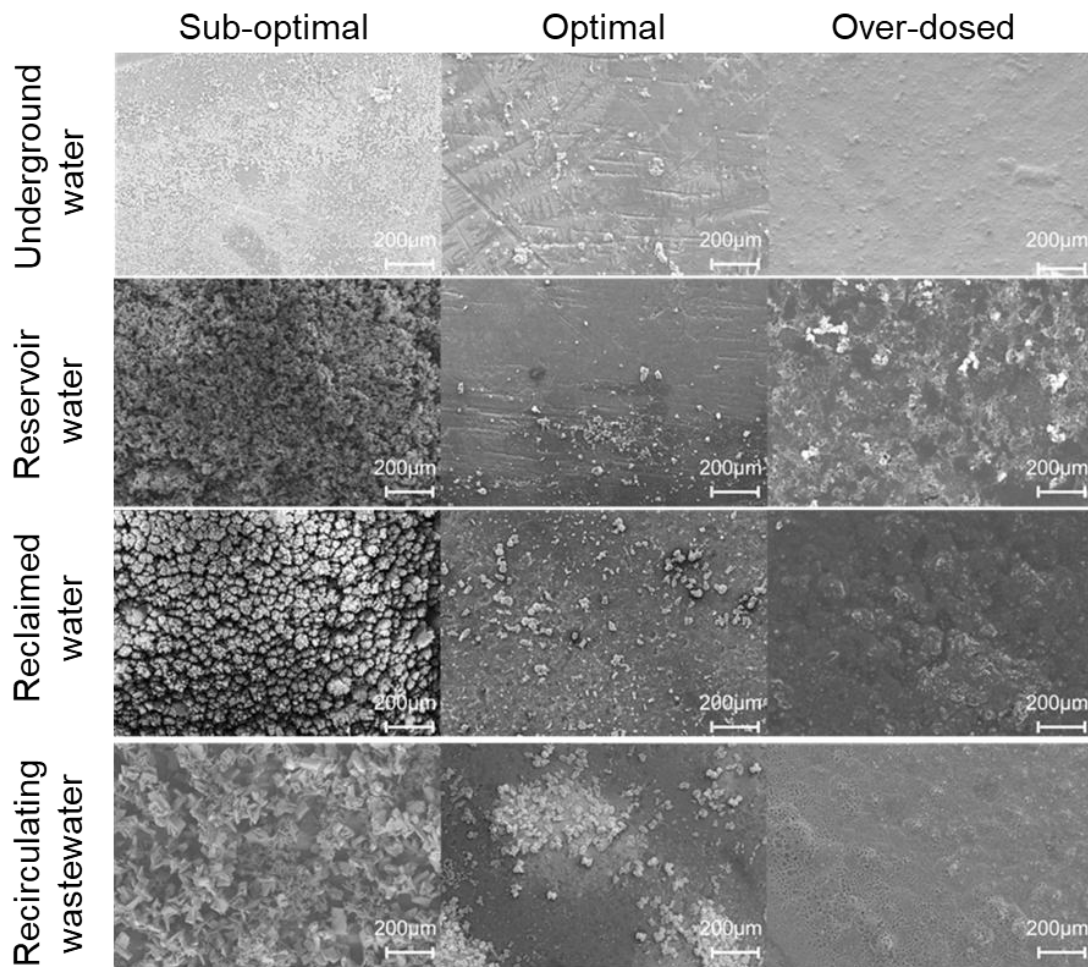


Fig.S2 SEM images of membrane surfaces under sub-optimal, optimal, and over-dosed ATMP concentrations across four influent types.