Artificial Intelligence Research and Development
T. Alsinet et al. (Eds.)

doi:10.3233/FAIA240403

Enhancing Seawater Reverse Osmosis Desalination Efficiency Using Digital Twins and Machine Learning

Aissam DAABOUB[a,b,1], Lluís ECHEVERRIA ROVIRA[a] and Edgar RUBION SOLER[a]
[a] *Eurecat, Centre Tecnològic de Catalunya, Unit of Applied Artificial Intelligence, Science and Technology Park of Lleida, Building H3, 25003 Lleida, Spain*
[b] *Agronomic Institute of Zaragoza (IAMZ), International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM), 50059 Zaragoza, Spain*

Abstract. The future faces escalating water scarcity due to population growth, climate change, and inefficient resource management. Therefore, innovative solutions for sustainable access and usage are needed. Seawater Reverse Osmosis (SWRO) desalination stands out as a key technology in tackling this dilemma. However, SWRO is energy-intensive, primarily due to the need to pressurize seawater to overcome the osmotic pressure to produce fresh water. In this regard, real-time management of operating parameters in SWRO plants enables minimizing energy consumption and chemical usage and adjusting water production in response to demand and water conditions, highlighting the need for real-time monitoring and advanced simulation tools such as digital twins. In response, this study explores the potential of eleven machine learning algorithms to simulate the SWRO process using a vast dataset of 18.816 scenarios generated through a solution diffusion transport model. Our investigation covers both non-ensemble and ensemble models. Additionally, a Shapley additive explanation analysis was carried out to gain insights into the most influential predictors and confirm the model's ability to comprehend the Reverse Osmosis (RO) process. The findings underscore the high accuracy of the algorithms, particularly XGBoost, CatBoost and ANN, in predicting key parameters such as permeate flow, permeate salinity and specific energy consumption. Furthermore, Support Vector Machine regression model shows promising in predicting permeate flow. These findings highlight the potential of data-driven models, particularly ensemble-based algorithms, in simulating SWRO behavior, laying the groundwork for future process optimization.

Keywords. Desalination, Reverse Osmosis, Machine Learning, Energy Consumption, Digital Twin

1. Introduction

Water scarcity is becoming increasingly critical because of accelerated population growth [1], exacerbated effect of climate change, and inadequate resource management. Therefore, finding innovative solutions for sustainable access and usage is imperative. Seawater desalination is one of the most attractive alternatives to supply clean and safe drinking water worldwide. Among the various desalination technologies, Reverse Osmosis (RO) stands out as the preeminent technology utilized in the process of seawater

1 Corresponding Author: Aissam Daaboub, E-mail: aissam.daaboub@eurecat.org

desalination. Desalination processes are known as energy-intensive, creating an energy-water nexus, with the two commodities essential for each other [2]. Roughly 71% of the total electricity used in RO-based desalination plants is attributed to the RO process, around 11% is consumed by pre-treatment and the remainder is used for seawater collection and distribution [3]. Moreover, the efficiency of RO plants is significantly influenced by both the feed water quality and the plant's operational parameters. This underscores the necessity for continuous monitoring and the adoption of sophisticated simulation tools, such as digital twins, to enhance performance and ultimately enable optimal control. Lately, there has been increasing interest in Artificial Intelligence (AI) and Machine Learning (ML) for tackling complex problems related to RO processes, compared to mathematical models, due to their flexibility and adaptability in managing high dynamic non-linearity and uncertainties, including fouling and fluctuations in feed water quality [4,5]. Many studies explore using Artificial Neural Networks (ANN), Multiple Linear Regression (MLR) and Support Vector Machine (SVM) models to predict the performance of RO process [6]. Research trends are shifting towards integrating more sophisticated ML prediction models, like tree-based and boosting models [4]. Encouraged by these advancements, this study aims exploring the potential of ML models to simulate and optimize the SWRO desalination process. The main goals include identifying the most accurate predictive models and gaining insights into the factors affecting the RO process performance in terms of permeate flow, permeate salinity, and specific energy consumption (SEC).

2. Methodology

The dataset comprises 18,816 instances, where each data instance represents a specific process step simulation (i.e., a set of input conditions and the corresponding process outputs). The simulator is based on a solution-diffusion transport model. For the simulation, a pressure vessel was utilized, housing seven sequentially arranged commercial Filmtec™ SW30XHR-440 spiral wound RO membrane elements. The simulations were conducted considering a grid with the following operating ranges as inputs: feed flow ranging from 5 to 17 m3/h; feed temperature ranging from 10 to 40 ℃; feed salinity ranging from 30 to 44 g/L; and feed pressure ranging from 40 to 80 bar, all in increments of 2 (see Table 1). The output parameters include permeate salinity, permeate flow and SEC.

Table 1. Summary statistics of process parameters used for predictive models' development.

	Parameter	Range	Mean	Standard deviation
Input	Feed Salinity (g/L)	30 - 44	37	4.58
	Feed Temperature (℃)	10 - 40	25	9.21
	Feed Flow(m^3/h)	5 - 17	11	4
	Feed Pressure (bar)	40 - 80	60	12.11
Output	Permeate Flow (m^3/h)	0.398 - 11.02	4.02	1.88
	Permeate Salinity (g/L)	0.057 - 2.087	0.22	0.129
	SEC (kWh/m^3)	2.68 - 46.95	4.91	2.21

For computational analyses, Python programming language, version 3.12.0, was employed for both data preprocessing and modeling. When necessary, the features were standardized. The entire dataset was randomly shuffled and split into a training set (70%) and a testing set (30%). Hyperparameters of various ML models were optimized by

cross-validated grid-search over a parameter grid, covering a diverse range of hyperparameters for each model. To estimate model uncertainty, 10-fold cross-validation was performed. In addition, the SHapley Additive exPlanations (SHAP) technique was employed to enhance the interpretability of the best-performing models. In total, eleven distinct ML regression models were implemented starting from simple non ensemble models (ANN, SVR, Kernel Ridge, Linear Regression, and Decision Tree Regressor) to more sophisticated black box ensemble models (XGBoost, CatBoost, LightGBM, AdaBoost, Hist Gradient Boosting Regressor, and Random Forest) as a screening step to come up with the model that has the highest predictive accuracy for each output parameter. For this purpose, three of the most common accuracy metrics of regression models were used to compare the predicted values against the test targets. These metrics include the coefficient of determination (R^2), the Root Mean Square Error (RMSE), and the Kling-Gupta efficiency (KGE) [7].

3. Results and discussion

Permeate Flow: Table 2 presents the performance metrics of the three highest-performing models out of eleven implemented algorithms for each output parameter. The results indicate that the CatBoost regression model exhibited superior performance in predicting permeate flow on the test dataset, achieving an RMSE of 0.0053, an R^2 value of 99.99%, and a KGE score 99.99%. Additionally, both the SVR and XGBoost regressor demonstrate high accuracy in predicting permeate flow. According to the SHAP analysis (see Fig 1 (a)), CatBoost shows that the feed pressure is the most critical factor impacting the permeate flow. Positive relationships exist between feed pressure, feed flow rate, and feed temperature with permeate flow, indicating that increases in these variables result in higher permeate flow.

Table 2: The performance metrics results for the three highest-performing models for each output (Test set).

	Permeate Flow			Permeate salinity			Specific energy consumption		
	SVR	XG Boost	Cat Boost	ANN	XG boost	Cat Boost	ANN	XG Boost	Cat Boost
RMSE	0.0063	0.0158	0.0053	0.0068	0.0068	0.0047	0.1072	0.1040	0.1029
R2	99.99	99.99	99.99	99.69	99.71	99.86	99.74	99.76	99.76
KGE	99.98	99.99	99.99	98.95	99.57	99.90	99.04	99.36	99.31

Permeate salinity: The CatBoost model demonstrated outstanding performance in predicting permeate salinity, with an RMSE of 0.0047, an R^2 of 99.86%, and a KGE of 99.90%. Both ANN and XGBoost also exhibited notable performance. SHAP analysis revealed that permeate salinity is most significantly influenced by feed flow rate, followed by feed salinity, feed pressure, and feed temperature (see Fig 1 (b)). This suggests that higher feed pressure, lower temperature, and increased feed flow are desirable to obtain low permeate salinity aligning with Mohammed et al. findings [4].

Specific Energy Consumption: Both CatBoost and XGBoost were ranked as the best predictive performance of SEC on test data prediction among other methods, proved by the lowest RMSE ranging from 0.1029 to 0.1040, identical highest R^2 (99.76%) and KGE ranging from 99.31% to 99.36%, respectively. Following closely, the ANN algorithm also demonstrates a high prediction capacity. SHAP analysis (see Fig. 1 (c)) indicated that feed pressure is the most critical factor impacting SEC. Thus, increasing the feed pressure results in reduced SEC due to producing a high permeate flow rate.

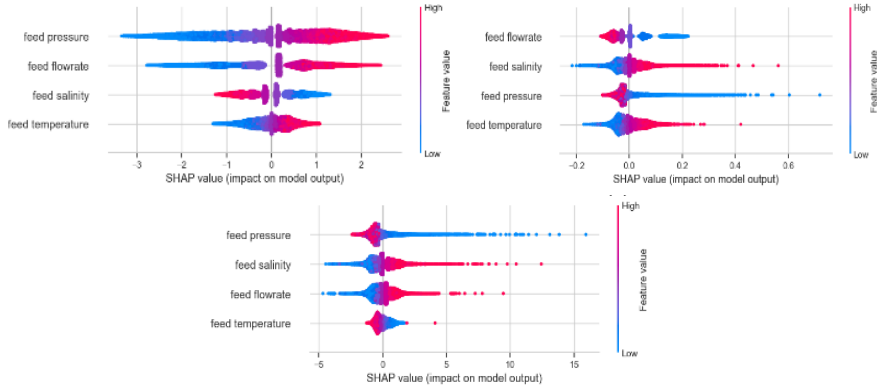

Figure 1. SHAP analysis. (a) permeate flow and (b) permeate salinity using CatBoost, (c) SEC using XGBoost

4. Conclusion

In summary, our findings demonstrated that ML regression models such as Catboost and XGBoost could effectively capture the mechanisms of the RO desalination process. Additionally, the insights provided by these models can play a crucial role in optimizing processes, understanding system behaviors, and enhancing the overall performance of the RO desalination process. They can serve as a rapid process simulation tool for a subsequent process optimization stage. However, further validation with real data from various desalination plants is necessary to confirm their reliability and applicability.

Acknowledgements

This work was financially supported by the Catalan Government through the funding grant ACCIÓ-Eurecat (Project FLAGSHIP 2023-CIRCLE).

References

[1] Zubair MM, Saleem H, Zaidi SJ. Recent progress in reverse osmosis modeling: An overview. Desalination 2023;564:116705. https://doi.org/10.1016/J.DESAL.2023.116705.

[2] Liu SY, Wang ZY, Han MY, Wang GD, Hayat T, Chen GQ. Energy-water nexus in seawater desalination project: A typical water production system in China. J Clean Prod 2021;279:123412. https://doi.org/10.1016/J.JCLEPRO.2020.123412.

[3] Voutchkov N. Energy use for membrane seawater desalination – current status and trends. Desalination 2018;431:2–14. https://doi.org/10.1016/J.DESAL.2017.10.033.

[4] Mohammed A, Alshraideh H, Alsuwaidi F. A holistic framework for improving the prediction of reverse osmosis membrane performance using machine learning. Desalination 2024;574:117253. https://doi.org/10.1016/J.DESAL.2023.117253.

[5] Golabi A, Erradi A, Qiblawey H, Tantawy A, Bensaid A, Shaban K. Optimal operation of reverse osmosis desalination process with deep reinforcement learning methods. Applied Intelligence 2024:1–21. https://doi.org/10.1007/S10489-024-05452-8/TABLES/4.

[6] Behnam P, Faegh M, Khiadani M. A review on state-of-the-art applications of data-driven methods in desalination systems. Desalination 2022;532:115744. https://doi.org/10.1016/J.DESAL.2022.115744.

[7] Kling H, Fuchs M, Paulin M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J Hydrol (Amst) 2012;424–425:264–77. https://doi.org/10.1016/J.JHYDROL.2012.01.011