ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Predicting Initial Trans-Membrane Pressure for Optimized Operations in UF Unit Using Random Forest

S Mukherjee*, K Villez*, A Melin*, S Ghanem*, T Cath** and A Hering***

* mukherjees2@ornl.gov, * villezk@ornl.gov, * melina@ornl.gov , * ghanemss@ornl.gov, **tcath@mines.edu, ***hering@baylor.edu

**Abstract:** With the growing scarcity of freshwater, innovative process design mechanisms like Reverse Osmosis (RO) are increasingly gaining attention among water treatment utilities to address the rising demand. Ensuring reliable water production necessitates efficient resource utilization, minimizing downtime in (ultra-filtration) UF systems. Recent advancements in machine learning (ML) have enabled the development of accurate data-driven models for Model Predictive Control (MPC), often requiring minimal prior knowledge of underlying physical processes. In this study, we present predictive regression models based on Random Forest (RF) and Auto-Regressive (AR) approaches to forecast the initial Trans-Membrane Pressure (TMP) for each filtration cycle in data generated by Direct Potable Reuse (DPR) systems. The proposed RF-based model demonstrates superior performance compared to baseline methods, including historical mean, Last Observation Carried Forward (LOCF), and naïve AR models, across various forecasting horizons in terms of root mean square error (RMSE) metric. To evaluate how different classes of process variables contribute to TMP dynamics over time, we examine the feature importance of independent covariates across multiple forecast horizons. This analysis provides insight into the temporal relevance of operational and sensor-derived features, guiding control and monitoring strategies. Additionally, the impact of hyperparameter tuning on TMP prediction performance is studied for both direct and recursive RF modelling approaches across increasing forecast horizons. Accurate prediction of initial TMP is critical for optimizing RO operations, as it enables the development of robust modelling frameworks by accurately estimating membrane fouling trends, thereby enhancing process efficiency and long-term reliability. The demonstrated efficacy of the RF-based approach highlights its potential as a tool for real-time decision-making in water treatment systems, paving the way for advanced process optimization and sustainable water resource management.

**Keywords**: trans-membrane pressure; direct potable reuse; ultra-filtration; regressive models.

**Introduction:** The growing scarcity of conventional drinking water has driven the adoption of DPR systems which treat wastewater to a quality suitable for direct human consumption [1-3]. DPR systems implement rigorous treatment protocols to ensure the effective removal of contaminants, pathogens, and impurities while maintaining compliance with stringent water quality standards. In addition to enabling high-level treatment, DPR offers logistical and economic advantages by reducing energy demands associated with water transport and facilitating near-complete reuse of purified water through integration with existing urban distribution infrastructure [4-7]. UF functions as a primary physical barrier in the context of DPR, ensuring the removal of contaminants that pose a risk to public health and facilitate the efficient operation of downstream processes. However, membrane fouling in UF units presents an operational challenge, necessitating regular maintenance to sustain performance and reliability [8-11]. Advances in machine learning have enabled accurate TMP forecasting, facilitating the development of robust MPC frameworks for water treatment [12-20]. The primary contribution of this work are as follows:

- Pilot-Scale Data Acquisition from a Mobile DPR Demonstration System: This study utilizes experimental data acquired from a mobile DPR trailer installed at the Colorado School of Mines (Denver, CO, USA). The system integrates six advanced unit processes—ozonation, biologically active carbon (BAC), ultrafiltration (UF), granulated activated carbon (GAC), ultraviolet advanced oxidation process (UV-AOP), and chlorination—designed to treat reclaimed water to meet potable quality standards (see Figure 1.1).
- Predictive Models Using Random Forests and Auto-Regressive Methods: We construct data-driven models including Random Forest (RF)—employing both direct and recursive forecasting architectures [21,22] —as well as Auto-Regressive (AR) and baseline predictors such as Last Observation Carried Forward (LOCF) and historical mean [23-25].

These models are trained to forecast initial TMP values at the start of each filtration cycle using a combination of set-point inputs, measured sensor data, and lagged TMP values. Additionally, the feature importance of covariates across multiple forecast horizons [26] is examined to understand the temporal influence of different operational variables on TMP dynamics.

- Comparative Performance Evaluation Across Forecast Horizons: The RF models are rigorously evaluated against baseline approaches across a range of k-step ahead forecast horizons using root mean square error (RMSE) as the primary performance metric. The direct RF model demonstrates strong predictive capability for TMP across both short- and long-term horizons, even in the presence of upstream setpoint variations. The impact of hyperparameter tuning on model performance is systematically analyzed, and comparative plots between experimental and predicted TMP values are provided for all forecast horizons.

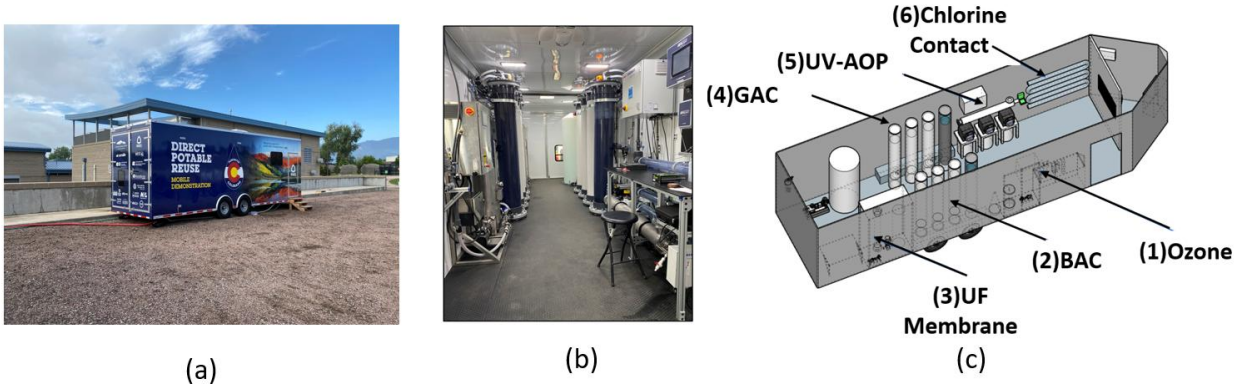
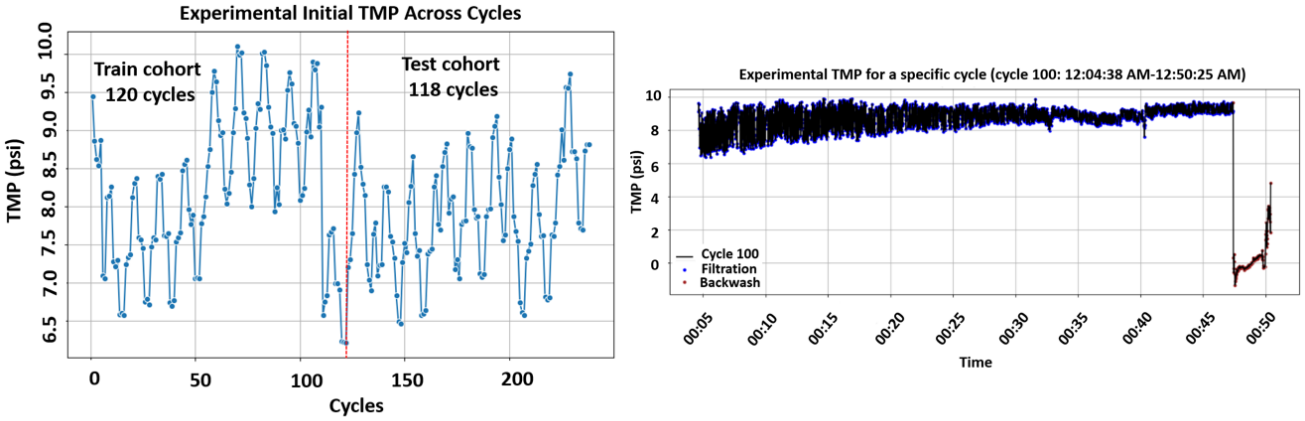



**Figure 1.1** (a) DPR mobile demonstration fitted in a trailer, (b) different components of the DPR system used to generate data, (c) schematic of the DPR pilot comprising of six units.

**II. Experimental Data:** In this study, each treatment cycle under investigation comprises a filtration phase followed by a backwash phase. The dataset includes 238 cycles, with the first 120 cycles designated for training and the remaining cycles reserved for testing (see figure 2.1 (a)). The Trans-Membrane Pressure (TMP) is experimentally calculated as the difference between the UF Feed Pressure and UF Permeate Pressure variables. Figure 2.1(a) presents the distribution of initial TMP values across all cycles, providing insight into the variability of the target variable. Due to noise in data, the mean of initial 5 minutes of data from each of the filtration cycle is used as the representative initial TMP for each cycle. This initial TMP serves as the target variable for predictive modelling. Figure 2.1(b) illustrates the distribution of TMP values across data points for one such representative cycle, highlighting the delineation of the filtration phase (blue) and backwash phase (brown). The figure shows the characteristic decrease in TMP values during the transition from filtration to backwash.

For predictive analysis, the selected covariates are listed in Table 2.1, comprising five user-controlled set-point variables, seven measured variables, and lagged TMP values from preceding cycles. These features were identified as critical for modelling the initial TMP and forecasting its values over multiple cycles. To better understand the temporal variability and patterns of critical covariates across training cycles, Figure 2.2 presents representative trends for three measured variables—UF Feed Flow (GPM), Ozone Dose (mg/L), and BAF pH (Panel A)—and three user-defined setpoint variables—UF Backwash Duration (s), UF Air Scour Duration (s), and UF Difference Flowrate (GPM) (Panel B).




(a) (b)

**Figure 2.1** (a) Initial experimental TMP of all 238 cycles, (b) initial experimental TMP across filtration and drain phase for a specific cycle (cycle 100).

**Table 2.1**: Independent covariates considered in Model analysis

| **Independent covariates involved in study** | | |
|---|---|---|
| **Measured Variables** | **Set point Variables** | **Lagged Previous TMPs** |
| UF-Feed Flow (GPM), BAF-NOX (ppmv), BAF-pH, BAF-Dissolved Oxygen (mg/L), O3:(TOC+NO2) Ratio, BAF-Turbidity [2] (NTU), Ozone Dose (mg/L), BAF-TOC (ppm) | Set BAF Filter Sequence Time (hr), Set UF Backwash Duration (s), Set UF Air Scour Duration (s), Set difference UF Flowrate (GPM), O3-Ozone (%) | Number of lag variables to consider is a tuning parameter in the model |

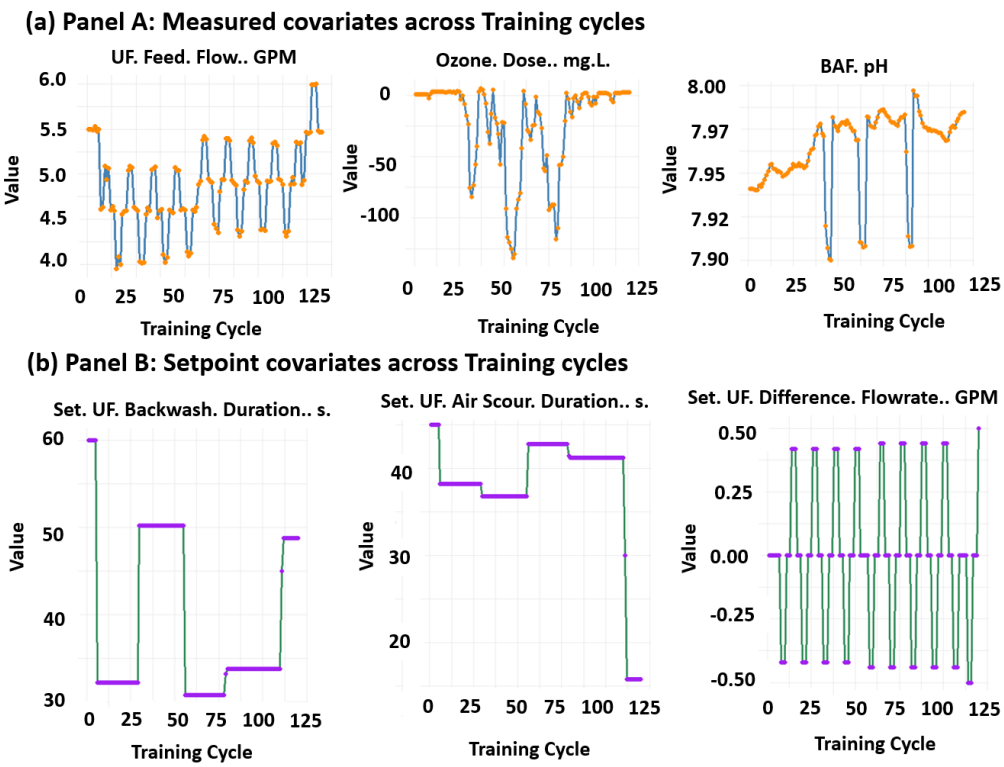



**Figure 3.2** Representative trends of three measured variables (a: panel A), three user defined set point variables (b: panel B) across training cycles

**III. Different Regressive Models- Methodologies:** At first an AR model [23,24] is fitted on the training cohort with no covariates consisting solely of lagged TMP values (where the lag L is a tunable hyper parameter) from previous cycles. The performance of the fitted AR model is evaluated on the test cohort using RMSE for k- step ahead prediction, with k ranging from 1 to 20. In AR model, the k-step ahead prediction is performed recursively: For 1-Step prediction, the last L observed TMPs from the training data are used as follows: $\hat{y}_{n+1} = \beta_1 y_n + \beta_2 y_{n-1} + \cdots + \beta_L y_{n-L+1}$, where $\beta_1, \ldots, \beta_L$ are the regression coefficients, $\hat{y}_{n+1}$ is the predicted TMP of the (n+1) cycle, $y_n, \ldots, y_{n-l+1}$ are L lagged observed TMPs from the training data .
For subsequent steps, the prediction is performed recursively by substituting the previously predicted values as follows: $\hat{y}_{n+2} = \beta_1 \hat{y}_{n+1} + \beta_2 y_n + \cdots + \beta_l y_{n-L+2}$ , $\hat{y}_{n+k} = \beta_1 \hat{y}_{n+k-1} + \beta_2 \hat{y}_{n+k-2} + \cdots + \beta_k \hat{y}_{n+k-L}$.

A non-linear random forest based regressor model [21,22] is next fitted to the training data and k-step ahead prediction is performed on test cohort. For one step ahead prediction, the initial TMP of cycle (n+1) is predicted using a RF regressor $f_{RF}$ trained on the covariates comprising set-point and measured variables, as well as lagged TMP values from previous cycles. The one-step-ahead prediction can be expressed as: $\hat{y}_{n+1} = f_{RF}\,(S'_{n+1}, M'_n, y_n, \ldots, y_{n-L+1})$ where $S'_{n+1} \epsilon\, R^{d_s}$ represents the feature vector of $d_s$ set point covariates, $m'_n \epsilon\, R^{d_m}$ represents the feature vector of $d_m$ measured covariates and $y_n, \ldots, y_{n-l+1}$ are L lagged observed TMPs from previous cycles. For k-step ahead prediction the RF model adopts a direct forecasting strategy which predicts the output $\hat{y}_{n+k}$ directly without depending on intermediate predictions $(\hat{y}_{n+1}, \hat{y}_{n+2}, ..)$ : $\hat{y}_{n+k} = f_{RF}\,(S'_{n+k}, M'_{n+k-1}, y_n, \ldots, y_{n-L+1})$. This direct formulation can avoid recursive substitution of earlier predictions, thereby reducing error propagation across long forecast horizons.

In addition to the direct RF approach, a recursive RF model is implemented to emulate autoregressive forecasting by iteratively updating lagged TMP values with previously predicted outputs. For k-step ahead prediction $(k > 1)$, recursive forecasting is performed by sequentially substituting predicted TMP to the lag structure. Thus, the prediction at step $(n + k)$ is given by:
$\hat{y}_{n+k} = f_{RF}\,(S'_{n+k}, M'_{n+k-1}, \hat{y}_{n+k-1}, \ldots, \hat{y}_{n+k-L})$.

**Feature Importance**: To evaluate the relative influence of controllable process variables on TMP dynamics across time, we conducted a feature importance analysis over all forecast horizons using the Random Forest (RF) model trained on the 120-cycle training cohort. Both set-point and measured covariates were considered in this analysis. The RF model was independently trained for each k-step ahead forecast task $(k \epsilon \{1,..,20\})$, and the feature importance was computed using the Mean Decrease in Accuracy (MDA) metric.

Given a trained model $f$, the MDA score for a feature quantifies the expected increase in prediction error when the values of $x_j$ are randomly permuted while holding all other features constant. Mathematically, the MDA score is defined as: $I_{MDA}(x_j) = E[l(\hat{y}, y) - l(\hat{y}^{perm(x_j)}, y)]$, where $\hat{y}$ are the model predictions under the original training data, $\hat{y}^{perm(x_j)}$ are the predictions after permuting values of feature $x_j$, and $l(.)$ denotes the prediction loss function based on RMSE. A higher $I_{MDA}$ value indicates greater sensitivity of the model performance to perturbations in that feature, implying higher predictive importance.

To capture the global relevance of features across multiple temporal horizons, we define the Cumulative Importance (CI) of feature $(x_j)$ as the aggregated MDA scores across all k-step ahead forecasts: $CI(x_j) = \sum_{k=1}^{20} I_{MDA}^{(k)}(x_j)$.

Figure 3.1 presents the cumulative feature importance scores for all covariates across the 20 forecast horizons, as computed on the training data. Each bar corresponds to the total MDA-based importance of a specific feature summed over k=1 to 20. Notably, Set UF Difference

Flowrate, Set UF Air Scour Duration, and Set UF Backwash Duration emerge as the three most influential features, indicating that TMP dynamics are strongly governed by user-defined operational controls in the upstream UF configuration. Among measured covariates, UF Feed Flow, Ozone Dose, and BAF Dissolved Oxygen show moderate importance.

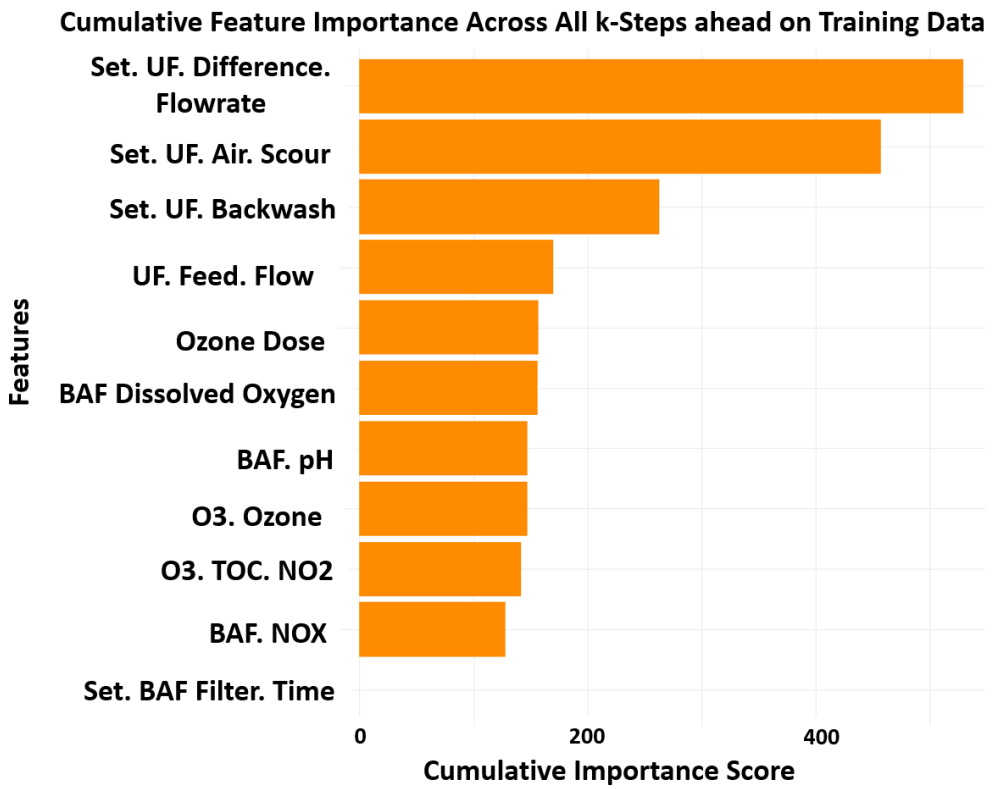



**Figure 3.1** Bar plot showing the cumulative importance score of the independent covariates across all the k steps based on training data.

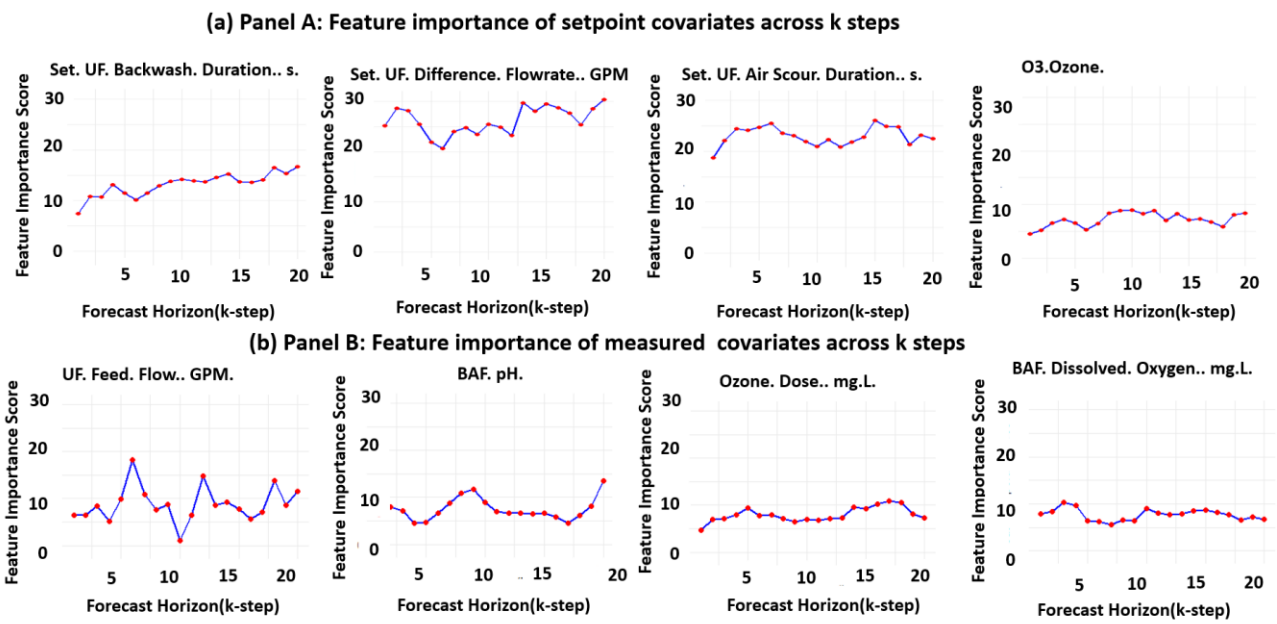



**Figure 3.2:** (a) Panel A displays the feature importance of set-point covariates obtained from RF models across k-step forecast horizons on the training dataset, (b) panel B presents the feature importance of measured covariates across the same k-step horizons.

Figure 3.2 illustrates the feature importance trajectories across forecast horizons for the top four covariates in each category: set-point variables (Panel A) and measured variables (Panel B). UF air scour duration, UF backwash duration, and UF feed–permeate difference flowrate—exhibited persistently high importance across all k, suggesting that membrane cleaning and hydraulic loading are dominant short and long-term drivers of TMP.

**IV. Results:** In Figure 4.1(a) we present a comparative analysis of the RMSE for k-step ahead predictions across four models: the Auto-Regressive (AR) model, the Random Forest (RF) model (direct and recursive), and two baseline approaches. The first baseline, the Last Observation Carried Forward (LOCF) model, assumes that the most recent TMP observation k-steps prior to the current cycle is the best predictor for k-step ahead prediction. In this approach, the predicted TMP for cycle $(n+k)$ is set as: $\hat{y}_{n+k} = y_n$ with corresponding RMSE as follows: $RMSE_k = \sqrt{\frac{1}{N-k}\sum_{n=k+1}^{N}(y(n) - y(n-k))^2}$ . This model is represented by blue curve in figure 3.2 (a). The second baseline, the historical mean model, predicts future TMP values based on the average of all TMP observations up to the n$^{th}$ cycle. The predicted TMP is given by: $\hat{y}(n+k) = \frac{1}{n}\sum_{i=1}^{n} y(i)$, and the RMSE is calculated as: $RMSE_k = \sqrt{\frac{1}{N-k}\sum_{n=k+1}^{N}(y(n) - \frac{1}{n}\sum_{i=1}^{n} y(i))^2}$ . This model is shown in red in figure 3.2 (a). The results indicate that the RF model (purple curve) achieves the lowest RMSE for k-step ahead predictions across the range of k (1 to 20), outperforming the AR, LOCF, and historical mean models. Figure 4.1(b) quantifies the improvement in prediction accuracy of each model relative to the LOCF baseline, expressed as percentage reduction in RMSE. Both RF variants show consistent improvements of over 50% across the horizon range. The recursive RF slightly outperforms the direct RF for intermediate horizons, but their performances converge at larger k. The results demonstrate that the RF model consistently outperforms all competitors, delivering significant improvements in TMP prediction accuracy for all forecast horizons.

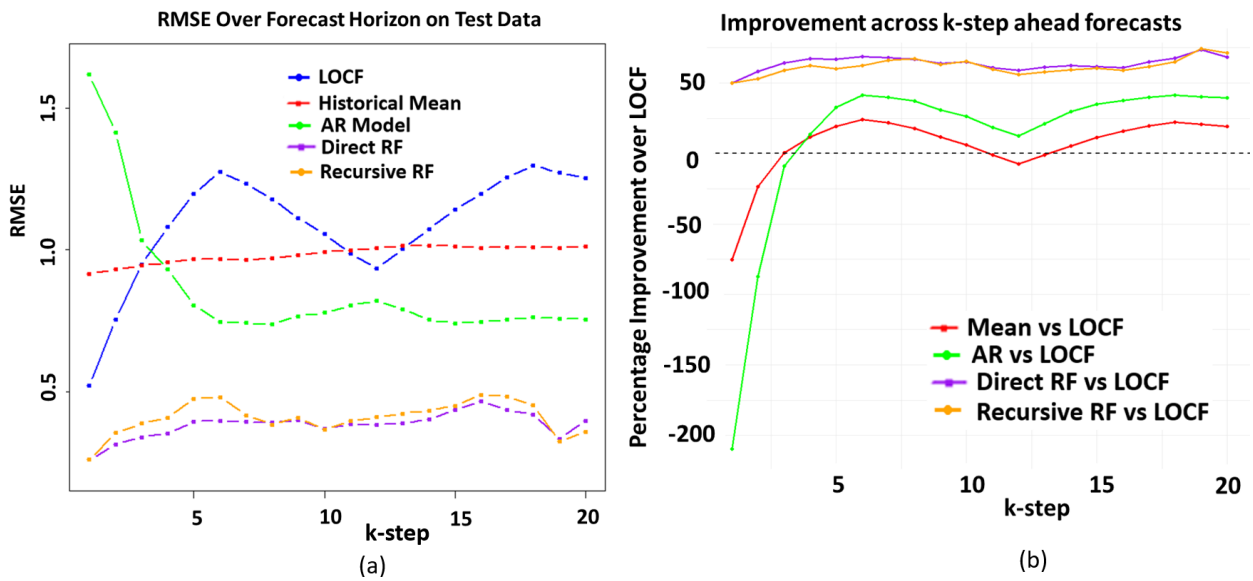



**Figure 4.1:** (a) RMSE across k-step ahead forecasts on the test dataset, comparing models: direct RF (purple), recursive RF (yellow), AR (green), LOCF ( blue), and historical mean (red), (b) percentage improvement in RMSE of all models relative to the LOCF baseline, illustrating that both RF models consistently outperform all others across the forecast horizon.

To evaluate the impact of both covariate selection and model complexity on TMP forecasting performance, we conducted a comparative analysis across several configurations of the Random Forest (RF) model, for both direct and recursive forecasting approaches. The results are summarized in Figure 4.2.

Three RF variants were examined for each forecast horizon k: (a) default RF model with 500 trees (blue), (b) extended RF model with 1000 trees (red), and (c) custom-tuned RF model (green), trained using time-series-aware validation. For the custom RF, the model is trained on a rolling window of 120 filtration cycles. Hyperparameter tuning is performed via time-slice cross-validation, implemented using the caret package. Each training window consists of the first 70% of the data (minimum 20 samples), and the remaining portion (up to 10 samples) is used for validation. This method preserves temporal causality and avoids data leakage. Hyperparameters - specifically the number of features sampled per split (mtry), minimum node size (node size), and maximum tree size (max nodes) - are selected to minimize validation RMSE. The optimal model is then evaluated on the test cohort. Each of these three RF configurations was evaluated under two covariate settings: (1) using all covariates (both set-point and measured), (2) using only set-point covariates. The results demonstrate that in both settings, the custom RF model achieves the lowest RMSE across all forecast horizons. When using all covariates, the custom RF (green) consistently outperforms both the 500-tree (blue) and 1000-tree (red) models. A similar trend is observed when using only set-point covariates, with the custom model (yellow) outperforming the 500-tree (purple) and 1000-tree (orange) models.

While models using both measured and set-point covariates slightly outperform those using set-point covariates alone at shorter forecast horizons ($k \leq 5$), this trend reverses for larger k. This is likely due to the reduction in effective training samples with increasing k, combined with the relatively high-dimensional input space comprising measured variables, set-points, and lagged TMP values. As the number of covariates approaches the number of samples, the model becomes more susceptible to overfitting. Under such conditions, restricting the model to set-point covariates offer a more parsimonious representation, resulting in improved generalization at longer forecast horizons. A parallel set of experiments is conducted using the recursive RF framework. As shown in Figure 4.2 (b), the same patterns hold: the custom RF model (green/yellow) consistently yields lower RMSEs compared to fixed parameter RFs across all k-step forecasts.

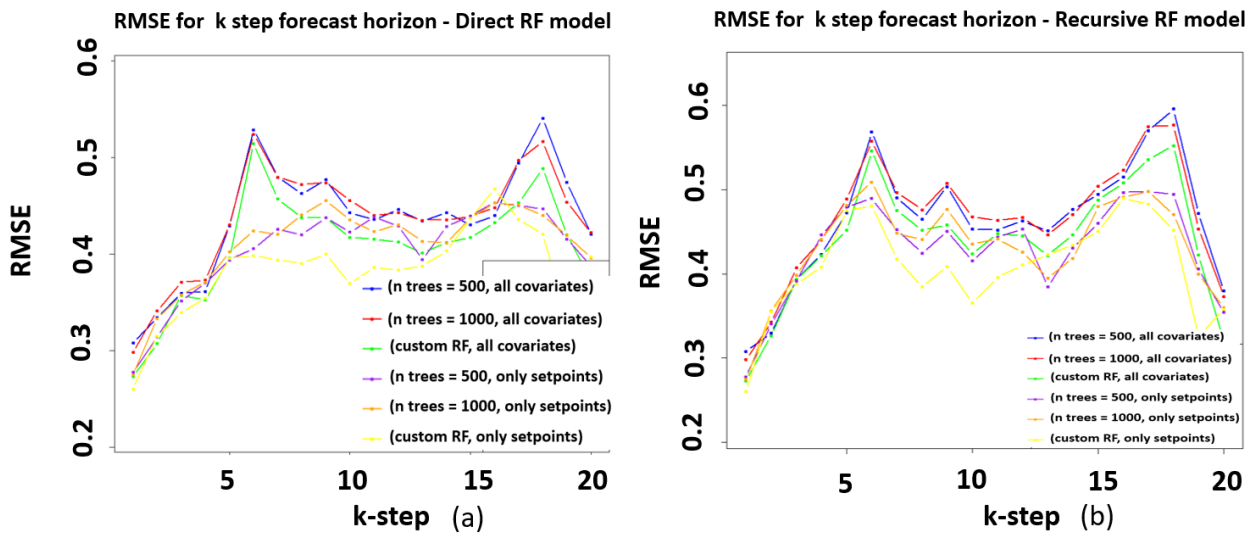



**Figure 4.2:** RMSE versus k-step ahead prediction for various Random Forest model configurations. (a) Direct RF models: Comparison of RMSE across forecast horizons using: all covariates (set-point + measured): 500 trees (blue), 1000 trees (red), custom RF (green); only set-point covariates: 500 trees (purple), 1000 trees (orange), custom RF (yellow), (b) recursive RF models: same comparison under recursive forecasting, with identical configurations and

*colour mapping. Custom RF models are tuned using time-series-aware cross-validation on a rolling 120-cycle training window.*

To illustrate the predictive fidelity of the proposed direct Random Forest (RF) model, Figure 4.3 presents a comparison between the experimentally observed initial TMP values (blue) and the TMP values predicted by the direct RF model (green) on the test dataset. Figure 4.3(a) shows the predictions for the 1-step ahead forecast horizon, while Figure 4.3(b) depicts the results for a 19-step ahead forecast. In Figure 4.3(a), the predicted TMP values exhibit strong agreement with the experimental data at $k = 1$, demonstrating the model's high accuracy under short-horizon forecasting. At $k = 19$, shown in Figure 4.3(b), prediction accuracy moderately degrades, due to the increase in temporal distance. Nevertheless, the direct RF model maintains a reasonably close match to the ground truth, effectively capturing both the level and trend of the TMP dynamics over extended horizons. This supports the robustness of the model in both short- and long-term forecasting scenarios.

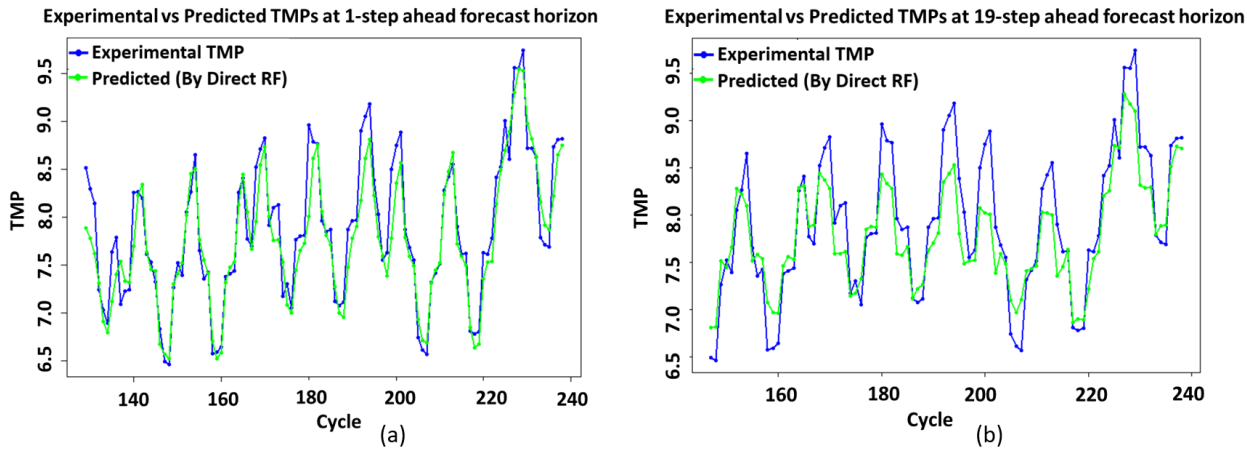



***Figure 4.3:*** *Experimental vs. predicted initial TMP values on the test dataset using the direct Random Forest model. (a) Prediction for k=1 (1-step ahead), showing near-perfect alignment with experimental TMP, (b) prediction for k=19 (19-step ahead), where accuracy slightly degrades but the predicted trend still follows the experimental profile closely.*

**V. Conclusion & Future Work:** The comparative analysis confirms that the proposed Random Forest (RF) model consistently outperforms benchmark approaches-including the Auto-Regressive (AR) model, Last Observation Carried Forward (LOCF), and Historical Mean models—achieving the lowest RMSE across all k-step ahead prediction horizons. The RF model maintains high predictive accuracy not only for near-term forecasts but also over multiple filtration–backwash cycles, despite variability in upstream pretreatment setpoints. This robustness enables improved control of transmembrane pressure (TMP), offering the potential to mitigate membrane fouling and reduce unnecessary backwashing operations. Moreover, feature importance analysis across forecast horizons highlights the dominant influence of both measured covariates-such as UF Feed Flow and Ozone Dose and user-controlled set-point variables like UF Backwash Duration and Air Scour Duration on TMP dynamics. These insights can inform real-time process monitoring and optimization strategies, supporting the development of intelligent automation frameworks for water treatment systems. This work underscores the growing potential of machine learning for advanced decision support in membrane-based water reuse, particularly in Direct Potable Reuse (DPR) systems. By identifying critical operational features and delivering accurate TMP forecasts, the RF model provides a data-driven foundation for enhancing membrane longevity and system efficiency.

As a future work we plan to extend the prediction horizon beyond 20 cycles by incorporating larger training datasets and improving generalizability. Additionally, integrating the model outputs into

a real-time control framework will enable dynamic setpoint optimization across the full treatment train, paving the way toward fully autonomous DPR systems with plant-wide operational intelligence.

## ACKNOWLEDGEMENT

This manuscript was authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government and the publisher, by accepting the article for publication, acknowledge that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, and to allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doepublic-access-plan accessed on 1 June 2024).

## REFERENCES

[1] Stover, Richard L. "Industrial and brackish water treatment with closed circuit reverse osmosis." Desalination and Water Treatment 51, no. 4-6 (2013): 1124-1130.
[2] Kucera, Jane. Reverse osmosis. John Wiley & Sons, 2023.
[3] Warsinger, David M., Emily W. Tow, Kishor G. Nayar, and Laith A. Maswadeh. "Energy efficiency of batch and semi-batch (CCRO) reverse osmosis desalination." Water research 106 (2016): 272-282.
[4] Schimmoller, L. J., M. J. Kealy, and S. K. Foster. "Triple bottom line costs for multiple potable reuse treatment schemes." Environmental Science: Water Research & Technology 1, no. 5 (2015): 644-658.
[5] Keller, Arturo A., Yiming Su, and David Jassby. "Direct potable reuse: are we ready? A review of technological, economic, and environmental considerations." ACS ES&T Engineering 2, no. 3 (2021): 273-291.
[6] Leverenz, Harold L., George Tchobanoglous, and Takashi Asano. "Direct potable reuse: a future imperative." Journal of Water Reuse and Desalination 1, no. 1 (2011): 2-10.
[7] Sanchez-Flores, Rosario, Adam Conner, and Ronald A. Kaiser. "The regulatory framework of reclaimed wastewater for potable reuse in the United States." In Water Reuse Policies for Potable Use, pp. 37-59. Routledge, 2018.
[8] Chen, Kai Loon, Lianfa Song, Say Leong Ong, and Wun Jern Ng. "The development of membrane fouling in full-scale RO processes." Journal of Membrane Science 232, no. 1-2 (2004): 63-72
[9] Chowdhury, Dhrubajit, Aurora Kuras, Derek Weix, Tani Cath, Alexander Melin, Yarom Polsky, Amanda S. Hering, Tzahi Cath, and Kris Villez. "Closing the loop: model-predictive control for a closed-circuit reverse osmosis system." Water Supply 25, no. 4 (2025): 727-748
[10] **Journal:** Peleato, NicolÁS M., Raymond L. Legge, and Robert C. Andrews. "Continuous organic characterization for biological and membrane filter performance monitoring." Journal-American Water Works Association 109, no. 4 (2017): E86-E98
[11] Niu, Chengxin, Yang Pan, Xueqin Lu, Shasha Wang, Zhongyi Zhang, Chaoting Zheng, Yujie Tan, Guangyin Zhen, Youcai Zhao, and Yu-You Li. "Mesophilic anaerobic digestion of thermally hydrolyzed sludge in anaerobic membrane bioreactor: Long-term performance, microbial community dynamics and membrane fouling mitigation." Journal of Membrane Science 612 (2020): 118264.
[12] Teng, Yun, and How Yong Ng. "Prediction of reverse osmosis membrane fouling in water reuse by integrated adsorption and data-driven models." Desalination 576 (2024): 117353.
[13] Zhao, Bin, Hao Chen, Diankui Gao, Lizhi Xu, and Yuanyuan Zhang. "Cleaning decision model of MBR membrane based on Bandelet neural network optimized by improved Bat algorithm." Applied Soft Computing 91 (2020): 106211.

[14] Shim, Jaegyu, Sanghun Park, and Kyung Hwa Cho. "Deep learning model for simulating influence of natural organic matter in nanofiltration." Water research 197 (2021): 117070.
[15] Kovacs, David J., Zhong Li, Brian W. Baetz, Youngseck Hong, Sylvain Donnaz, Xiaokun Zhao, Pengxiao Zhou, Huihuang Ding, and Qirong Dong. "Membrane fouling prediction and uncertainty analysis using machine learning: A wastewater treatment plant case study." Journal of Membrane Science 660 (2022): 120817.
[16] Bagheri, Majid, Ali Akbari, and Sayed Ahmad Mirbagheri. "Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review." Process Safety and Environmental Protection 123 (2019): 229-252.
[17] Schmitt, Félix, and Khac-Uan Do. "Prediction of membrane fouling using artificial neural networks for wastewater treated by membrane bioreactor technologies: bottlenecks and possibilities." Environmental Science and Pollution Research 24, no. 29 (2017): 22885-22913.
[18] Shahouni, Reza, Mohsen Abbasi, Mahdieh Dibaj, and Mohammad Akrami. "Utilising Artificial Intelligence to Predict Membrane Behaviour in Water Purification and Desalination." Water 16, no. 20 (2024): 2940.
[19] Guo, Hong, Kwanho Jeong, Jiyeon Lim, Jeongwon Jo, Young Mo Kim, Jong-pyo Park, Joon Ha Kim, and Kyung Hwa Cho. "Prediction of effluent concentration in a wastewater treatment plant using machine learning models." Journal of Environmental Sciences 32 (2015): 90-101.
[20] Dologlu, Pelin, and Hasan Sildir. "Data driven identification of industrial reverse osmosis membrane process." Computers & Chemical Engineering 161 (2022): 107782.
[21] Breiman, Leo. "Random forests." Machine learning 45 (2001): 5-32.
[22] Kane, Michael J., Natalie Price, Matthew Scotch, and Peter Rabinowitz. "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks." BMC bioinformatics 15 (2014): 1-9.
[23] Wu, Xinhe. "Water quality prediction based on AR and LSTM model." In Journal of Physics: Conference Series, vol. 2580, no. 1, p. 012019. IOP Publishing, 2023.
[24] Newbold, Paul. "ARIMA model building and the time series analysis approach to forecasting." Journal of forecasting 2, no. 1 (1983): 23-35.
[25] Saboia, Joao Luiz Maurity. "Autoregressive integrated moving average (ARIMA) models for birth forecasting." Journal of the American Statistical Association 72, no. 358 (1977): 264-270.
[26] Scornet, Erwan. "Trees, forests, and impurity-based variable importance in regression." In Annales de l'Institut Henri Poincare (B) Probabilites et statistiques, vol. 59, no. 1, pp. 21-52. Institut Henri Poincaré, 2023.