TECH BRIEF

Data Science Summer Fellows Program

Summer 2021

Modernizing Water and Wastewater Treatment through Data Science Education & Research

Orange County Water District – Reverse Osmosis Membrane Fouling

Caroline Anderson, Baylor University
Samuel Koller, Colorado School of Mines
Victor Martinez-Gil, McLennan Community College

SUMMARY

California is susceptible to drought, causing water shortages and mandated water restrictions. To alleviate water scarcity, technology and modern infrastructures must be used. Groundwater usage jumps from 40% to 60% of the water supply during drought years [1]. For this reason, it is important to be able to clean and reuse water efficiently. The Orange County Water District leads the world in recycled water technology, adding resilience against future drought conditions [2]. Our research focuses on understanding the measures used by the Orange County Water District to remove contaminants using reverse osmosis technology. Additionally, this paper summarizes our investigation regarding a Clean-In-Place projection model in order to analyze the rate and distinguish the influential factors of membrane fouling.

INTRODUCTION

Orange County Water District (OCWD) desired to gain insight on the relationship between influent water quality parameters and the rate of membrane fouling in the third stage of reverse osmosis across the twenty-one different filters in the OCWD Water Treatment Facility. An understanding of these relationships could help better predict when the membranes experience sufficient fouling to require maintenance in the form of a cleaning-in-place (CIP). Additionally, OCWD is curious about the effect of seasonality and temperature on the frequency of cleanings, which is dependent on the fouling rate on the membranes.

In this tech brief, we will discuss the CIP process of the Reverse Osmosis (RO) membranes. Stage 3 specific flux (S3SF) was identified as a key variable in the data and defined as the flow rate through the area of the membrane in the third stage. Specific flux decreases over time as the membranes experience fouling. Fouling occurs as particles adhere to the surface of the membrane, which causes an increase in concentrate (water concentrated with contaminants) pressure and reduces the efficiency of the RO process. CIP processes temporarily relieve membrane fouling and increase the specific flux of the RO membranes. The feed Total Organic Carbon (TOC) was identified as another fundamental influent water parameter as it provides insight on the organic loading rate on the membranes.

FACILITY SYSTEM DESCRIPTION

The RO systems found at this facility follow a pyramid format. The feed water flows into 77 stage 1 membranes. RO membranes separate the water into permeate and concentrate; the concentrate proceeds through stage 2 (49 membranes) and repeats in stage 3 (24 membranes). The permeate is used as potable water for the area served by OCWD. An overview of this process is shown in Figure 1.

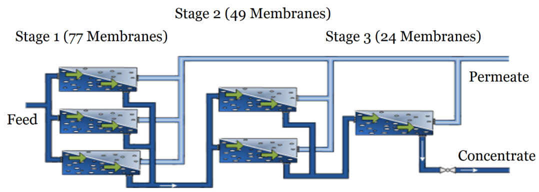


Figure 1: An overview of the Reverse Osmosis Process found at the OCWD Water Treatment Facility.

DATA DESCRIPTION

OCWD provided a very complete data set. Measurements were recorded on a daily basis with very few missing data values. The data used for this analysis was provided in twenty-five excel sheets, one for each of the twenty-one filters, two feed reports, and additional sheets containing information on each filter i.e. membrane type and age of each filter. The values in the feed reports for the A-E and F-G filters were recorded using the same influent water, but the report associated with the F-G filters contained more information such as TOC than the report associated with the A-E filters.

In addition to the variables given to us by the stakeholder, our team also created some of our own to use in our analysis. We first created a Days since CIP variable by looking at the drops in concentrate pressure that indicate when a CIP has occurred. Later on, we created a membrane age variable by subtracting the current date from the date of the membrane installation.

EXPLORATORY DATA ANALYSIS

To better visualize the data, we created various plots to identify relationships between water quality parameters, specific flux, and concentrate pressure. We developed a shiny application, which allows a user to interact with the data through a web application. This app shows a scatterplot between two chosen variables as well as a time series of a chosen variable for any of the A-E filters.

Additionally, our team conducted an investigation of the effect of seasonality on the frequency of CIPs. OCWD hypothesized CIP occurred more often in winter months than summer months. We utilized concentrate pressure which is an indicator of fouling. We combined filters' data based on their train and created boxplots to compare concentrate pressure between summer and winter seasons. From this boxplot, we saw that the concentrate pressure was higher during the winter in all trains, explaining why CIPs are needed more frequently.

Lastly, we wanted to understand how temperature and pressure behaved in relation with CIP cycles. We started by looking at each train of filters independently and then we combined all of them into one plot. In Figure 2, we can see there is not a strong relationship between the CIP cycles and the temperature or concentrate pressure.

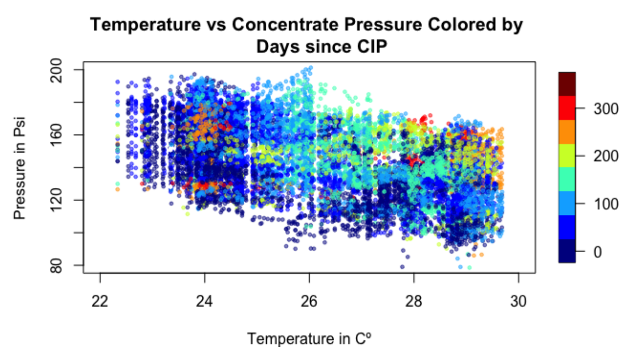


Figure 2: Temperature vs. Concentrate Pressure scatterplot, points are color-coded by the number of days since the last CIP has been done.

CIP PROJECTION MODEL

Our main goal was to project when the next CIP should be conducted on stage 3 of each filter in the facility. We developed a linear model that estimates when the next CIP should be performed on the membranes. These estimations are generated 30 days after the previous CIP. We created this linear model using the first 30-day averages after the filter is cleaned for the following variables: temperature, TOC, feed flow, and permeate flow. This model includes the age of the membrane and the first 30-day average S3SF. Our team utilized a log transformation for the S3SF and age of the membrane since this made the model more linear.

However, there were only 50 complete cycles of CIP in the data, and the model was using six covariates, which is too many. Following the "one in ten rule", the proportion of observations to covariates should be at least ten observations per covariate. We applied backwards-stepwise regression and lasso regression as methods of variable selection. Backwards-stepwise regression reduced six covariates to three covariates. The lasso regression model eliminated one covariate reducing the six covariates to five covariates. Both models contained an intercept. Figure 3 shows the true versus predicted days until the next CIP should be done for each model.

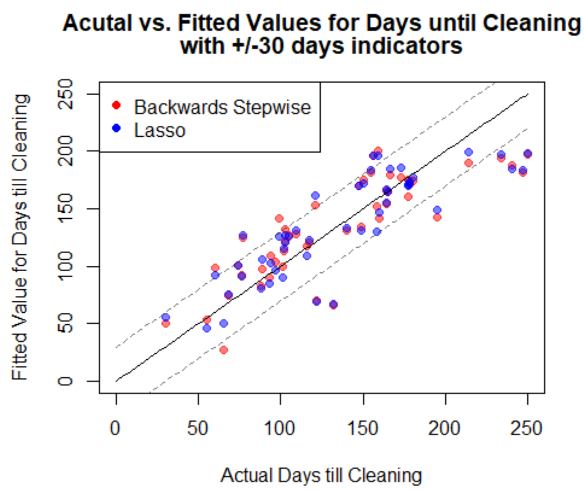


Figure 3: Scatterplot of the Actual Days until the next CIP against the predictions (fitted values) made by the backwards-stepwise and lasso regression models. The solid black line represents a perfect prediction and the dashed black lines show a buffer of +/- 30 days.

RESULTS

The backwards-stepwise regression model had an RMSE value of 28.4, whereas the lasso regression model had an RMSE value of 28.2, an insignificant difference. This model has an R^2 value of 0.68. We recommend the backwards-stepwise model because it is simpler and is not significantly different from the lasso regression model. A summary of this model can be found in Table 1.

Variable	Coefficient	Significance
Intercept	2280	3.2e-12
S3SF	139.1	2.8e-09
Temperature	-14.27	1.5e-09
Feed TOC	-88.54	4.6e-08

Table 1: Summary of the CIP projection model. Variables are more significant as the significance value decreases. Note: S3SF has a logarithmic transformation and all variables are the first 30-day averages after a CIP has been done.

From Table 1, we can see that the first 30-day average S3SF has a positive relationship with the time until the next CIP, meaning that as the starting S3SF is larger after a CIP, the time needed until the next CIP increases. Additionally, we can see that the temperature and the feed TOC have a negative relationship with the days until the next CIP, indicating that higher temperatures and concentrations of TOC reduce the number of days until the next CIP is needed.

Finally, we developed a second shiny app for the CIP projection model. Here, an employee of OCWD can enter the first 30-day averages for S3SF, temperature, and Feed TOC to receive an estimation of when the next CIP should be done. Additionally, a summary of the model is displayed here including the coefficients, significance levels, and the R^2 value.

CONCLUSIONS

Our primary focus of this project was to accurately project when the next membrane cleaning should be done. Our analysis showed that the most significant variables for this are the first 30-day averages (after the previous CIP has been done) of S3SF, temperature, and feed TOC, accompanied by an intercept.

The secondary goal of this project was to understand how the influent water parameters affected the rate of third stage fouling. We found that S3SF has a direct relationship with the time needed until the next CIP and that temperature and feed TOC had an indirect relationship with the time needed until the next CIP.

Finally, the stakeholder requested for insight into how seasonality affects the rate of cleanings at the OCWD Water Treatment Facility. We found out that trains of filters show different trends, and that these trends do not necessarily signify a cleaning cycle. After combining all filters, we saw that there was no clear relationship between CIP cycles and temperature or between CIP cycles and pressure.

Thus, we recommend the stakeholder to use the CIP projection shiny 30 days after the previous cleaning to get an idea of when the next CIP should be done. From our analysis, it seems that the membranes foul quicker at higher temperatures and concentrations of TOC, so we recommend OCWD to attempt to minimize the influent water temperature for the best results in their facility.

We recommend further analyzing the effect of seasonality on the need for CIPs and attempt to find an optimal temperature for the influent water. In the future, we would recommend looking into the one minute data and see how this shows the effect of influent water quality parameters.

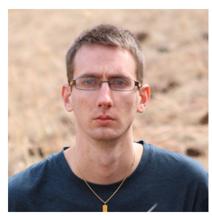
REFERENCES

[1] Becker, Rachel. "Water Shortages: Why Some Californians Are Running out in 2021 and Others Aren't." *CalMatters*, 23 June 2021, calmatters.org/environment/2021/06/california-water-shortage/.

[2] "California Water Issues Overview." *Water Education Foundation*, www.watereducation.org/aquapedia/california-water-issues-overview.

AUTHORS

Caroline is an Economics major at Baylor who enjoys growing plants and home brewing Kombucha.



Samuel is currently pursuing a Master's Degree in Statistics at Mines who likes sports and animals.

Victor is a Computer Science major at McLennan Community College who likes playing guitar and traveling.

ACKNOWLEDGEMENTS

We would like to give Han Gu and the rest of the Orange County Water District a special thank you for taking the time to meet with us and clarify questions we had about the facility.

Additionally, we would like to extend our gratitude to Dr. Hering, Dr. Nychka, Dr. Cath, Sweta Rai, Hattie Means, Aurora Waclawski, Tayler Elwell, and Luke Durell for their guidance throughout this process.