Modernizing Water and Wastewater Treatment through Data Science Education & Research

TECH BRIEF

Data Science Summer Fellows Program

Summer 2022

Direct Potable Reuse Trailer – Wastewater Treatment

Courtney Hodge, Baylor University
Kaylee Pascarella, Baylor University
Brendan Stewart, Colorado School of Mines

SUMMARY

The Direct Potable Reuse trailer treats wastewater to drinkable quality and educates the community about local and sustainable water treatment. It is important for the operators of the trailer to be able to identify when faults occur in the system, specifically using the data from the Biologically Active Filtration and Microfiltration Membrane processes. This brief focuses on determining the normal operating conditions of the trailer through data visualization and suggests methods for fault detection through functional data analysis.

INTRODUCTION

The Direct Potable Reuse (DPR) trailer is a system that accepts as input treated wastewater from a wastewater treatment facility with the goal of treating this input until it is potable, i.e., drinkable. The wastewater travels through six different processes before it is potable, but the main processes of interest in the trailer for fault detection are the Biologically Active Filtration (BAF) and the Microfiltration Membrane (MF). Since data is recorded as water flows through the DPR system, a fault in the system can be identified through behaviors of recorded variables such as water flow and pressure within the BAF and MF processes. It is of interest to first identify the normal operating conditions within the trailer, and secondly to determine where behavior deviates from what is considered normal, therefore identifying when a potential fault has occurred.

FACILITY SYSTEM DESCRIPTION

The DPR Trailer is a mobile water treatment lab designed to produce 7,000 gallons of drinkable water per day.

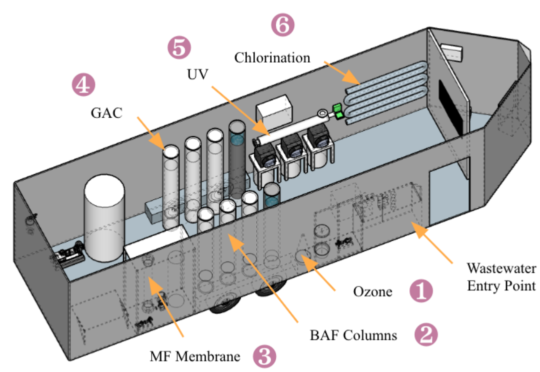


Figure 1: The DPR trailer interior. Image provided by Tzahi Cath and edited by our team

Inside, the DPR Trailer holds a treatment train of six processes: ozone, BAF, MF, granular activated carbon (GAC), ultraviolet disinfection/advanced oxidation process, and chlorination. These processes are monitored and controlled with a network of sensors.

The project's primary concern is with the BAF and MF processes. The BAF process consists of four vertical tanks that use microorganisms to reduce dissolved pollutants in wastewater effluent. The BAF system is designed to treat water using three columns at a time, leaving the fourth in standby. The MF process consists of a ceramic membrane that filters impurities and allows for maximum flow of permeate water to downstream processes.

Along with the other processes of the DPR Trailer, BAF and MF experience periods of automatic and manual operation. Automatic operation is when the DPR Trailer is actively producing potable water: this operation is where unusual process behavior, or faults, will exist. Manual operation is the opposite: this is when water is not being treated and irregular sensor readings are anticipated due to sensor drifts. Also, both processes experience regular cleanings called backwashes during automatic operation. In a backwash, water flow is reversed through BAF and MF filters to clear debris and prevent clogging. MF backwashes are hourly, whereas BAF backwashes are performed either on a longer timer, or manually.

DATA DESCRIPTION

The data are observed over 71 sessions, or independent periods of operation, with duration varying from one hour to 19 days. The data are composed entirely of readings from 165 up to 193 sensors, with the number of active sensors varying between sessions. These sensors measure pressure, flow, power, setpoint, and other parameters. All sensor data is captured every 30 seconds during a session. The dataset is generally free of missing data and large errors; as new sensors are introduced, they can spend several sessions inactive before beginning to record data.

EXPLORATORY DATA ANALYSIS

Initial exploratory data analysis (EDA) consisted of exploring variables within the MF and BAF systems to identify faulty operating conditions of system variables across sessions.

BAF Process: To explore the trends within the BAF process, we examined the BAF flow variable and the BAF valve positions across time. Visually inspecting these plots across all sessions revealed what normal operating conditions look like and what a fault could look like in the data. When the system was in manual mode, all four BAF columns were active and all BAF valves were held at a flat, constant value with no variation. When the system was in automatic mode with no fault occurring, the BAF flow stayed around the same rate throughout the session, and the valves varied slightly in order to maintain this flow rate. The flow rate fluctuated slightly between sessions, but it was determined that the flow usually remained at a point somewhere between 1.5 and 3 gallons per minute. The valves appeared to open as the session progressed, likely due to solids accumulation. Behavior that deviated from these normal operating conditions indicated a likely fault; for example, a common fault could be seen where the BAF valves increased until they were 100% open, but the flow began to decrease regardless. Figure 3 below shows an example of a session where behavior started off normal, then manual mode was activated, and then a fault can be observed near the end of the session where the BAF flow begins to decrease below optimal levels even though the BAF valves increase to 100% open.

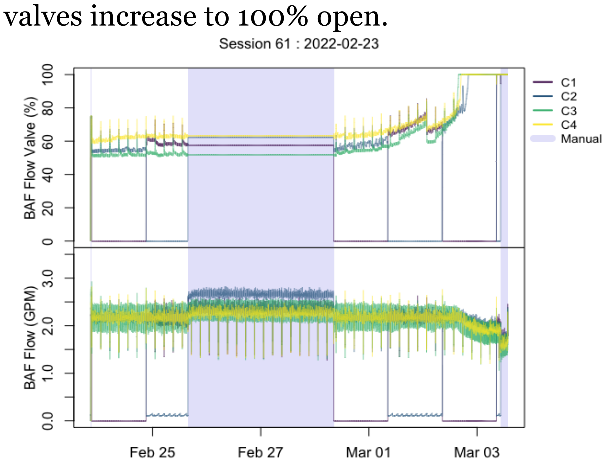


Figure 2: BAF flow and valves plotted against time for session 61 beginning on 02/23/2022

MF Process: When exploring the MF process, MF variables, such as feed pressure and feed volume, were explored for normal and faulty operational behavior. Time series plots were created for these variables across all sessions and plots were stacked for comparison. During a manual period, feed pressure and volume would flatline to 0 psi and 0 gallons respectively. During an automatic period, it was determined that feed pressure should remain between 5 and 15 psi and feed volume between 10 and 25 gallons. Anything outside of these ranges would indicate fault-like behavior. In the time series plots, backwashes are visible when both variables drop to zero for a brief period and recover to normal operation.

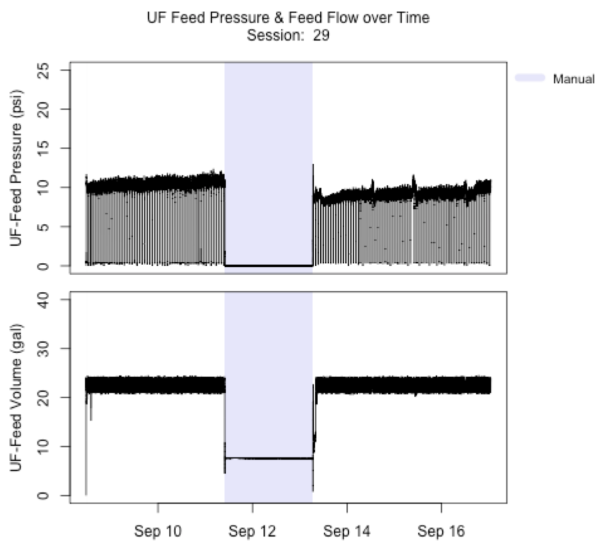


Figure 3: MF feed pressure and feed volume for session 29 beginning on 09/08/2021

DATA WRANGLING

The data was first split into subsessions, discarding long periods of manual operation identified through the filtration pump state variable. In the resulting 78 sessions, a subset of variables was selected, backwash periods were identified, and the data were partitioned such that each observation ends with the backwash pump shutting off. Each period of operation between backwashes is referred to as an observation. Ultimately, any observation containing any data made while the system was in manual operation was discarded, resulting in 62 meaningful subsessions.

STATISTICAL ANALYSIS & RESULTS

As a backwash in the filtration system may be triggered by a number of sensors, an assumption was made that normal behavior during an observation was independent of its duration. Thus, the observations within each session were registered to an equally spaced domain via cubic spline smoothing. Feed pressure data are selected from this smoothed data and magnitude outlyingness (MO) and variance outlyingness (VO) metrics are used to quantify the data. The MO metric is a measure of the mean within each normalized function, and the VO is a metric of each normalized function's variance. [1].

$$o(X_i(t)) = \frac{X_i(t) - med\{X_1(t), \ldots, X_n(t)\}}{MAD\{X_1(t), \ldots, X_n(t)\}}$$

$$MO(X_i) = \frac{1}{500}\sum_{k=1}^{500} o(X_i(t_k))$$

$$VO(X_i) = \frac{1}{500}\sum_{k=1}^{500} (o(X_i(t_k)) - MO(X_i))^2$$

Application of these metrics to a subsession with $n-1$ backwashes provides n pairs of MO and VO metrics

The MO/VO pairs are plotted for each observation in a session. Bagplots are then created for each session to identify outlier functions, which are assumed to be possible faults. A bagplot is a bivariate generalization of the traditional boxplot; it allows one to visualize the data's location, spread, and outliers. This method allows the outlying functions within each session to be identified and isolated, which the DPR trailer operators could use to catch faults before they cause extreme disruption to the system.

Below, figure 4 provides an example of the application of the bagplot to MF feed pressure over a session with outliers. VO is plotted on the log scale in order to spread the points out for a better visual representation. The dashed line represents the outlier fence. The fence is calculated by multiplying the bag by a factor of four, a value suggested by explorations into identifying known faults, where the bag contains 50% of the innermost observations. The smoothed data marks the MO/VO bagplot outliers in red. Visually, these outliers exhibit characteristics outside nominal observations, as they have larger or smaller MO or VO values. Bagplots can thus be applied to a combination of MF variables to identify observations of higher concern across other sessions.

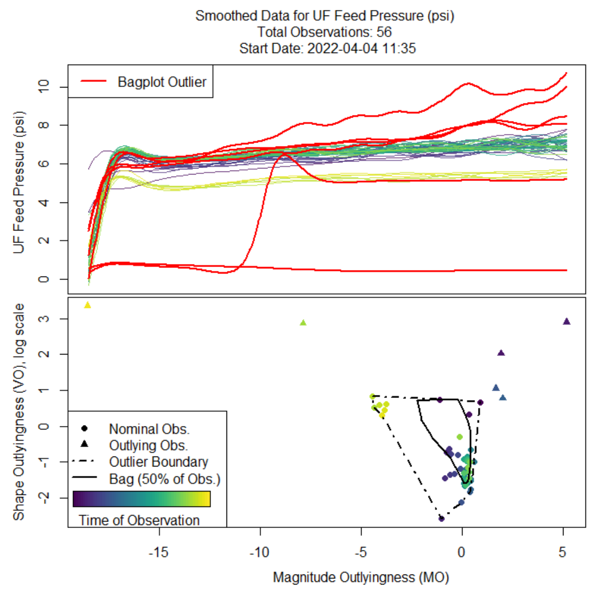


Figure 4: Smoothed Data and Bagplot of MF feed pressure for session 69 beginning on 04/04/2022

CONCLUSIONS

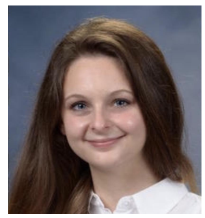
The main goal of this project was to identify system faults inside the BAF and MF processes of the DPR Trailer. This goal was accomplished by determining normal operating conditions across sessions of specific BAF and MF variables. Variables with a high correspondence to faults were identified using time series plots in the initial EDA. The BAF variables flow rate and flow valve helped to identify observed system faults in BAF. Observed system faults of MF were identified using the functional data analysis methods of outlyingness metrics and bagplots.

For further analysis, a clustering metric approach is suggested to better identify outliers. Analysis using the Calinski-Harabasz criterion, a tool for selecting the number of clusters, showed promise in improving fault identification and avoiding false negatives. Exploration of clustering methods is suggested in continuations of this project.

REFERENCES

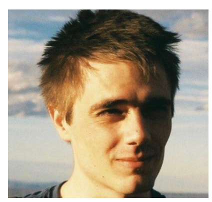
[1] Kuras, Aurora, et al. “Functional Data Analysis Approach for Detecting Faults in Cyclic Water and Wastewater Treatment Processes.” In Preparation, May 2022.

AUTHORS



Kaylee Pascarella recently graduated from Baylor University with a B.S. in Statistics. She is entering the Baylor University Statistics Ph.D. program in the fall.

Courtney Hodge is currently an undergraduate Data Science Major, Secondary Spanish Major and History Minor at Baylor University.



Brendan Stewart is currently an undergraduate student. He is seeking an M.S. in Applied Mathematics from Colorado School of Mines.

ACKNOWLEDGEMENTS

Tzahi Cath PhD., Colorado School of Mines
Amanda Hering PhD., Baylor University
Doug Nychka PhD., Colorado School of Mines
Luke Durell PhD., Baylor University
Natalie deBonoPaula, Baylor University
Mason Manross, Colorado School of Mines