An Exploratory Analysis of a Mines Park Wastewater Dataset

MATH 398 Final Project

COLORADO SCHOOL OF MINES

Laura Wiesner
Haley Vinton

23rd April 2022

1 Abstract

A major fault in the Mines Park wastewater treatment facility occurred on April 24th, 2010. It took the wastewater facility over two months to recover. Considering water scarcity on the Western slope, faults such as this one on a larger scale system could be detrimental to communities in need of water. It is our goal to be able to make recommendations to the operators of the system to potentially prevent such a fault from occurring in the future. In order to do this, an exploratory data analysis (EDA) was performed on data collected from the Mines Park treatment facility from April-May 2010. The goal of this EDA was to identify any contributing factors to the fault. Through our data analysis, we found that, in our professional opinion, that the fault was caused by two major factors; one being a slow increase in sewage flow into the sequencing batch reactor (SBR) tanks (where regular oscillations are expected by the biological community) for a noticeably long period of time, as well as the level of the MBR tanks staying too low for too long. We believe the combination of these two events contributed to the severity of the fault, and also to why it took so long for the system to recover.

Table of Contents

2 Introduction

2.1 Water Scarcity

Water scarcity is a relevant problem which occurs across the United States in multiple counties. It is most efficient to be able to recycle and reuse as much waste water as possible by being able to treat and return wastewater which is clean for use in the community [1]. Although it is most efficient to be able to return recycled water to the community, it also requires diligent monitoring and recording. This is due to sensitivities within the system, as well as safety requirements for human use and consumption of the recycled water. The Mines Park Wastewater dataset gives a good example of a smaller scale wastewater treatment, based on samples taken from Mines Park. This sort of smaller and localized wastewater treatment is more efficient in terms of water scarcity than the more common, larger wastewater treatments, which are currently only capable of releasing the recycled water back into the environment as opposed to directly into communities [1], denying communities of an extra source of clean water.

2.2 A Small Scale Example

Mines Park is an excellent example of a small-scale local wastewater treatment facility, and is overseen by a Mines faculty member. This data set has over 40,000 observations and 42 different kinds of measurements which span over a period of approximately one month, starting April 10, 2011 at 6:10 AM and ending May 5, 2011 at 5:00 AM where measurements were taken every 10 minutes. Within this wastewater treatment facility, there is a wastewater inlet which first pumps wastewater into two SBR tanks where a family of organism kills off unwanted constituents, and blowers blow oxygen into the SBR as the biological communities require oxygen. Once the wastewater completes its time in the SBR, it is pumped to two membrane bioreactors (MBR), where there is a membrane designed to separate the water from the solids. These solids are called 'return activated sludge' (RAS), which are then returned to the SBR. Finally, the water is considered to be clean or 'effluent' and can be placed back into the community water supply. Consider Figure 1 which shows the different parts of the system:

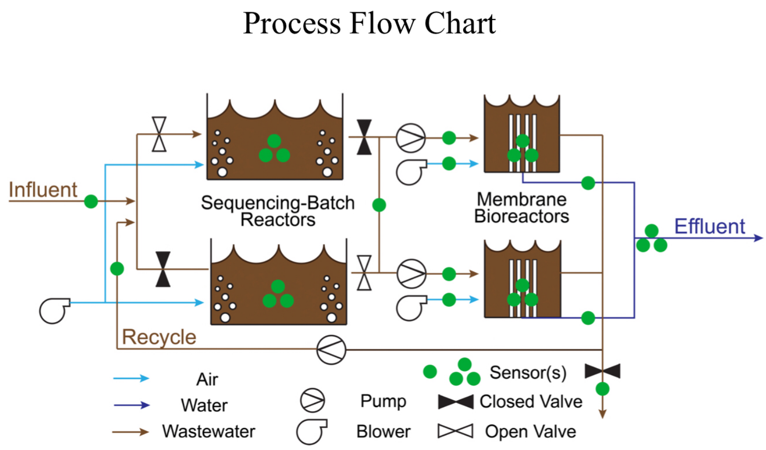


Figure 1: Sequencing Batch and Membrane Bioreactor Process Flow Chart. Shows the SBR, MBR, blower locations, pump locations, direction of airflow and water flow, and location of sensors, [1]

2.3 System Fault

In order for the biological community to stay alive, the pH of the system (which is measured from the RAS) must remain above 6.5. On April 24th, 2010 at 10:00 AM, a major fault was detected by the operators of the system. This fault caused the pH to drop below 6.5 for a period of time, and although measures were quickly taken, the system did not fully recover for approximately 2 months. Faults such as this one can be detrimental to a small scale water treatment facility.

3 Methods

3.1 General Exploratory Data Analysis

The dataset originally contained 43552 rows of data, and 42 different columns (or 42 different specific measurements). The dataset contains a categorical qualitative variable (given by the Date/Time of measurements), continuous quantitative variables (such as temperature of the RAS, taken to a decimal place) and discrete qualitative data such as the phase of the bioreactor (given by true and false values only). There are no missing variables in this dataset, but there are some small but negative values for the variables that represent flow rates. It is believed these minus signs have some sort of meaning to a system operator (such as direction of flow), so it was decided to keep them in the set.

The main goal of the data analysis was to identify any outliers in response or control variables before the fault that may have caused the system to malfunction. In order to do this, graphical representations of all the variables, in the form of time series plots, were necessary. Once plots were created, a red vertical line was overlaid on each plot, representing the day and time at which the fault occurred. The purpose of this red line was to visually represent what portion of the data was taken before and after the fault. Then, each plot was analyzed for unusual values or trend that occurred to the left of the red vertical line, which represents data taken before the fault occurred.

The first anomaly that was noticed corresponded to a variable that represented the sewage flow into the water treatment system. For a certain time period, there was non cyclical behavior observed; specifically, the sewage flow slightly and constantly increased. The time period was noted and all of the other plots were analyzed again to see if any other irregular behavior occurred at this time frame. Other unusual behavior in different measured variables emerged, and further exploration of these variables began, as is described below.

Then, a new dataset, only consisting of measurements taken during these times of odd behavior was created. Then, plots of the same type, but with smaller datasets, are created to better understand the behavior of a variable at a certain time. In total, four time series plots were created and summary statistics for three variables were calculated for the final portion of data analysis.

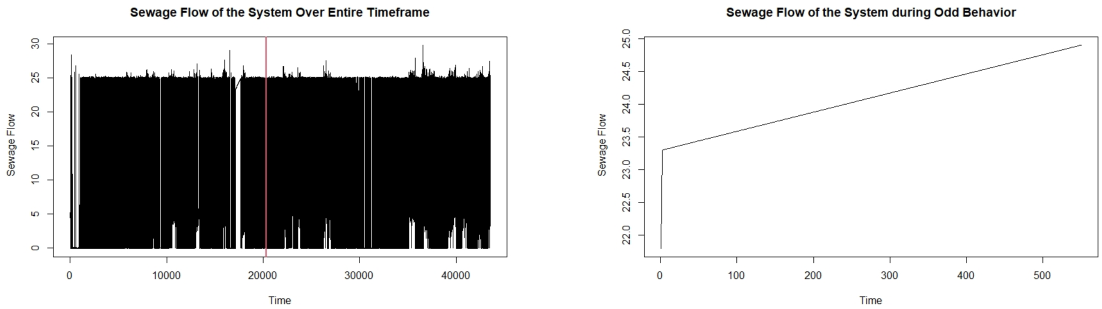
Sample Statistics

The following table shows some sample statistics for *some* of the key variables:

Variable	Mean	Standard Deviation	min	max
RAS pH	7.10	0.48	5.5	7.9
MBR Level Tank 1	9.89	0.91	0	3.5
MBR Level Tank 2	5.81	9.71	0	5

Pre/Post fault

Based on our EDA, the first abnormality that was noticed was a departure from cyclical behavior in the sewage flow graph for the whole dataset. For a subset of observations, a slight, but steady increase was observed. This behavior occurred from observations 17132 to 17681. The first set of graphs shown represent the sewage flow of the system. The graph on the left displays the sewage flow of the system before, during, and after the fault. The graph on the right displays the odd increasing behavior observed before the fault in greater detail.

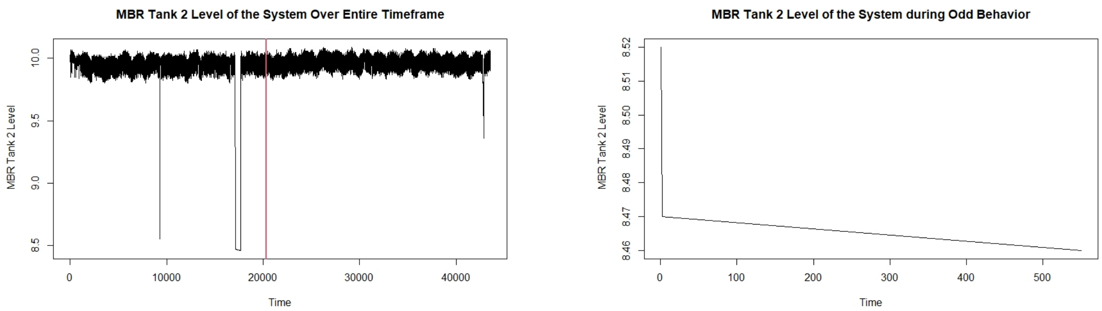


(a) Sewage Flow for all data

(b) Sewage Flow for the subset of data demonstrating odd behavior

Figure 2: Graphs of Sewage Flow

Additionally, we wish to further investigate the levels in the MBR tanks to see if there is also abnormal behavior occurring in the same set of observations denoted above. The next set of graphs shown represent the levels of the 2nd MBR tank in the system. The graph on the left displays the level of the tank before, during, and after the fault. The graph on the right displays an outlier detected before the fault and details the behavior of the level of the tank.



(a) MBR level for all data

(b) MBR level for the subset of data demonstrating odd behavior

Figure 3: Graphs of Level of MBR Tank 2

It is important to note that both of these variables did not exhibit their normal behavior during the small time frame defined above.

4 Results

4.1 Findings

It was first noted that there was a time period which began around 4:00 AM on the 22nd of April and ended on April 22nd around 2:00 in the afternoon where the sewage increased in a slow and controlled manner instead of an extremely oscillatory manner, which is clear in Figure 2a and in Figure 2b. In this Figure 2a, it can be seen that the oscillations are so frequent that they appear as a solid black box until there is a clear white space in the graph showing this period where we believe the system was shocked. Simultaneously, during this same period, we see the level of the MBR tanks decrease in a similar fashion with the increase of sewage flow, but then the level massively drops to a value of 9.12, where the average value is approximately 9.9.

By performing EDA, and by using Figures 2a, 2b and Figures 3a, 3b, it is believed that the major fault occurred due to the system being put into a state of shock which caused the delayed and major fault (which occurred within two days of the odd behavior seen in Figure 2a, 2b and Figure 3a, 3b). It is believed two major events caused this shock to the system, the first being that the biological community expects an oscillation of sewage flow which it did not have, and also expects to have reasonable levels in the MBR tanks. As seen in Figure 2b and Figure 3b, the sewage level increased constantly and simultaneously as the MBR tank level dropped dangerously low, and this combination of events likely caused the pH to drop below a threshold of 6.5 for too long (549 observations) for the biological community to survive.

An interesting fact which was noted was that there was a period earlier in the dataset where the pH dropped below 6.5 momentarily, but the system was able to immediately recover unlike with the later, major fault. This is likely because the sewage flow during this time was still quite oscillatory and the MBR levels dropped lower than normal, but not nearly as low as it did just before the major fault.

5 Discussion

5.1 Recommendations

It is important to point out that it is believed that the constant and slow inlet of sewage flow (as seen in Figure 2b) is likely due to some system process (for example, this may represent a transfer between the tanks that happens during this period). However, this slow inlet of wastewater combined with the low levels in the MBR tank are what is believed to have caused the major fault. For this reason, we have a few recommendations.

The first recommendation is to put an alarm system on the MBR level sensor, so that when the water drops below approximately 9.9, the operators are notified so they can replenish the levels quickly. This will prevent the level of the tank from dipping so low and in turn affecting the pH of the system. Another recommendation is to have an employee present at the wastewater plant if possible at all times of the day. This way, the employee could likely try to mitigate the situation of low MBR tanks or too slow of sewage inlet before the system has a chance to kill off its own biological community.

5.2 Implications and Further Questions

One of the implications which arose from this work is that it shows that a small scale wastewater treatment seems reasonable to implement on larger scales, but will require much diligence from the system operators to ensure major faults do not occur. This type of system is quite sensitive, but would be worth the extra work to help with water scarcity where needed.

More investigation could be done with this dataset. Using machine learning concepts such as linear regression could likely be useful and interesting to use on this dataset, as this could provide connections between variables which we cannot see by just visual inspection. It would also be interesting to do a similar type of EDA using a larger dataset to see if there are similar events and results.

References

[1] Professor Hering. Project 03: Mines Park Fault. Accessed: 4-18-2022. 2021. URL: `Youtube:https://www.youtube.com/watch?v=e7dpQ-OLbXE`.