TECH BRIEF

Data Science Summer Fellows Program

Summer 2020

Mo(Wa)²TER

City of Boulder –

Justin Nichols, Affiliation
Ocean Wu, Affiliation
Arlyn Alcid, Affiliation

Modernizing Water and Wastewater Treatment through Data Science Education & Research

SUMMARY

In the City of Boulder, a wastewater treatment system filtered out phosphorus in the form of orthophosphate before filtering back into Boulder Creek. The data collected from the facility monitored phosphorus levels in influent and effluent water as chemicals such as Ferric and Alum were introduced. This raw data was closely examined, cleaned, merged, and split by the date, hour, and coagulant type variables. After performing an Analysis of Covariance (ANCOVA) and Estimated Marginal Means tests, there was evidence to indicate that one chemical had a greater change in the phosphorus concentration (mg/L).

INTRODUCTION

At the 75th St. Wastewater Treatment Facility in Boulder, Colorado, two chemicals, Alum and Ferric are added to aid the removal of phosphorus. High concentrations of phosphorus are extremely dangerous as they create rapid formations of algae blooms which eat up all available oxygen needed for other aquatic organisms to survive. Consequently, over time, the diversity of aquatic ecosystems are greatly decreased.

FACILITY SYSTEM DESCRIPTION

(Fig. 1)

(Credit: The City of Boulder 75th Street Wastewater Treatment Facility Brochure)

1) Headworks, where water enters.
2) Primary Clarifiers, removal of larger rocks and pollutants in water.
3) Secondary treatment, nitrogen (in the form of ammonia) and some phosphorus are removed.
4) Aeration Basin, allows for aerobic degradation of water contaminants.
5) Mixed Liquor Waste sludge, where chemical dosing of Ferric and/or Alum occurs.
    a) Leads into the Solid Contact station where extra activated sludge occurs and phosphorus is solidified.
6) Secondary Clarifier, removes solid phosphorus.

(Fig. 2)

(Credit: The City of Boulder 75th Street Wastewater Treatment Facility Brochure)

7) UV Disinfection, neutralizes potentially harmful bacteria and microorganisms.
8) Effluent Pipeline, water is released into Boulder Creek.

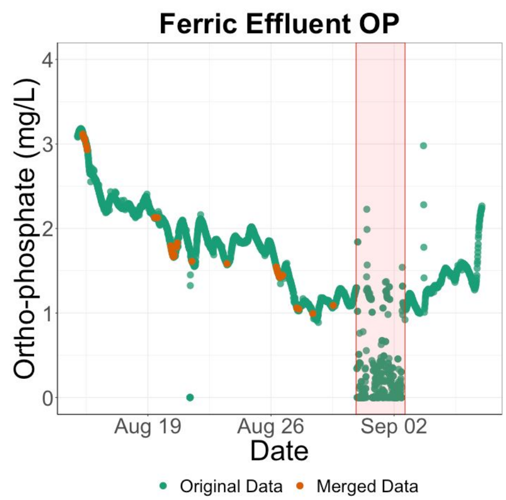
DATA DESCRIPTION

The raw data was collected from July 1, 2019, to January 31, 2020, and consisted of 7 separate datasets, containing important variables such as coagulant type, inflow water rates, chemical dosing amounts, phosphorus levels (mg/L). etc., at each major checkpoint location of the facility.

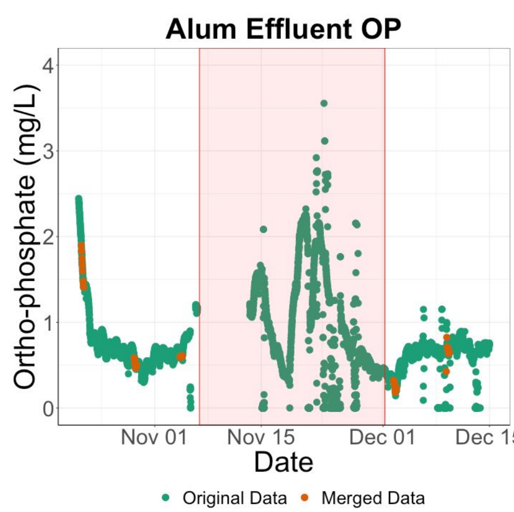
- The experiment was specifically designed to monitor the individual effect of coagulant type on the removal of phosphorus.
    - o July 1 - August 14: Control period, no chemicals added.
    - o August 15 - September 7: Ferric dosing period.
    - o September 8 - October 20: Second control period, no chemicals added.
    - o October 21 - December 15: Alum dosing period.
    - o December 16 - January 31: Final control period, no chemicals added.
- Focusing on the coagulant groups (Alum, Ferric, and None) it was clear that there were many missing observations for the "None" group.
- Data was collected from both laboratory grab samples and online sensors.

EXPLORATORY DATA ANALYSIS

A single merged dataset was created by only keeping variables in the dosing daily, flow hourly, mixed liquor hourly, and phosfax 10m datasets. Repeated variables were removed and datasets were inter joined by their date variables.



(Fig. 3)



(Fig. 4)

Figures 3 and 4 show the Ferric and Alum dosing periods, respectively, with the date on the x-axis and the effluent phosphorus measured in mg/L on the y-axis. The green points were pulled from the original phosfax 10m dataset and the orange points were pulled from the single merged dataset. Though this process, data containing only complete cases were kept for final analysis.

The large red boxes indicate a gap in the dosing period, where sensor system failures occurred. In Ferric dosing, observations between August 30 to September 2nd had unusually low values, including numerous repeated values of 0.0004, which are assumed to be the default error value. In Alum dosing, observations between November 6 to December 1st show a gap in observation values, followed by an unusual trend and noise into December. It is noted in the raw data on November 11th the phosphorus dosing system went down for three hours and was later recovered, but was down again from November 18th onward. Many factors may have caused the system to temporarily fail, including voltage failures and uncontrollable seasonal conditions. For this reason, observations falling within the red highlighted regions were omitted from the merged set.

At the beginning of each dosing period, there was a portion of data where the effluent amount of phosphorus drops dramatically. During this time, the chemical treatments are not yet in full effect. To prevent this from altering the results of the final coagulant analysis, a subset of the merged dataset, referred to as the partial dataset, was created to include only observations once effluent phosphorus concentrations had leveled off.

STATISTICAL ANALYSIS and RESULTS

In general, a t-test is used to determine if there is a difference between the means of two groups. A t-test was performed to compare the mean percent change of phosphorus after alum and ferric dosing, with the following hypotheses:

- Ho: The average percent change of phosphorus after Alum dosing = The average percent change of phosphorus after Ferric dosing.
- Ha: The average percent change of phosphorus after Alum dosing ≠ The average percent change of phosphorus after Ferric dosing.

The t-test returns a *p*-value of 3.8e-9 which is less than 0.05. As a result there is evidence to reject the null, therefore, there is evidence the average percent change of phosphorus in Alum is not equal to the average percent change of phosphorus in Ferric. The mean percent decrease of phosphorus with Alum = 56.5 and the mean percent decrease of phosphorus with Ferric = 33.9. In this case, Alum has a higher mean in the change of phosphorus.

An Analysis of Covariates (ANCOVA) is used to determine the mean of a dependent variable, change of phosphorus (mg/L), based on an independent variable, the chemical coagulants (Ferric and Alum), and the addition of a covariate, moles of metal dosed daily (Kmol).

Our ANCOVA found that there was a significant relationship (*p*= 1.92e-8 for moles of metal, *p* = 2.83e-5 for coagulant type) between the change in phosphorus. Therefore, we performed an Estimated Marginal Mean Test to find the estimated mean response for percent change in phosphorus given each coagulant type, adjusted for any other variables in the model.

- Ho: The mean change of phosphorus under the treatment of Alum after accounting for the effect of moles of metal = The mean change of phosphorus under the treatment of Ferric after accounting for the effect of moles of metal.
- Ho: The mean change of phosphorus under the treatment of Alum after accounting for the effect of moles of metal ≠ The mean change of phosphorus under the treatment of Ferric after accounting for the effect of moles of metal.

The p-value is 3e-5, which is less than 0.05, therefore, there is evidence to reject the null hypothesis and evidence to conclude that a difference exists between the mean change of phosphorus between coagulant groups with the addition of the daily added moles of metal. The 95% confidence interval strengthens this finding. The change of phosphorus in (mg/L) within Alum dosing is between 0.59 and 0.79, and the change of phosphorus in (mg/L) within Ferric is between 0.95 and 1.19. Since these intervals do not overlap, and because the change of phosphorus is much higher in the Ferric group, it appears that Ferric is outperforming Alum.

CONCLUSIONS

The goal of the Boulder Wastewater Facility was to decrease the amount of effluent phosphorus flowing back into Boulder Creek with the addition of two chemicals, Alum and Ferric. Although the t-test showed Alum had a higher percent removal of phosphorus, it did not take into account any other variables that may have affected the amount of wastewater flowing through the system. The ANCOVA tests give a more accurate conclusion that Ferric was likely the outperforming chemical, due to the addition of a covariate.

REFERENCES

Field,Andy. "Analysis of Covariance."

*https://Www.discoveringstatistics.com/Docs/Ancova.pdf,2016,*Www.discoveringstatistics.com/Docs/Ancova.pdf.

"The City of Boulder 75th Street Wastewater Treatment Facility." *City of Boulder Cororado*, 2008,bouldercolorado.gov/water?utm_source=godaddy&utm_medium=redirect&utm_campaign=boulderwater.net.

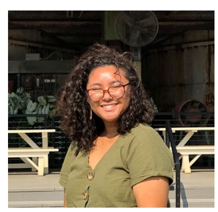
AUTHORS

Justin Nichols

jnichols2@mymail.mines.edu.

Ocean Wu

haosheng_wu1@baylor.edu



Arlyn Alcid

arle_alcid1@baylor.edu

ACKNOWLEDGEMENTS

A big thank you to Chris Marks, Dr. Hering, Dr. Nychka, Kate Newhart, and Luke Durell for their insight throughout the project, and to everyone involved in the gathering of raw data!