Mo(Wa)^{2}TER

Modernizing Water and Wastewater Treatment through Data Science Education & Research

TECH BRIEF

Data Science Summer Fellows Program

Summer 2020

City of Superior

Rahul Banka, Baylor University
Alina Gavrilov, Baylor University
Hunter Privett, Baylor University

SUMMARY

The City of Superior Wastewater Treatment Plant is responsible for treating municipal wastewater to effectively comply with state regulations before releasing it into Rock Creek. Unique to this plant is a two-track secondary treatment that splits the influent wastewater into an East and a West aeration and clarification system before joining the two back together again. Data shows that these two tracks operate as independent systems, allowing for valuable experimentation and comparative analysis in either or both tracks. Currently, the oxygen diffusers in the two aeration basins are different and information is sought on which one optimizes the necessary biological processes occurring within the basin. The stakeholders are interested in optimizing ammonia and nitrate removal to prepare for increasingly stringent Colorado State regulations, and they are continuously updating systems in the plant to enhance the precision and control of operations.

INTRODUCTION

The goals of the project were 1) to analyze whether the two tracks have a total difference in mass signifying whether one diffuser performs better than the other and 2) to determine which variables have a high impact on effluent ammonia concentrations of the clarifiers. Answering the first question statistically confirmed the operations-based hypothesis that the East and West tracks operate as independent systems. Answering the second question gave the operations team tangible variables to experiment within their greater trajectory of optimizing ammonia and nitrate removal.

FACILITY SYSTEM DESCRIPTION

In the City of Superior wastewater treatment plant, influent wastewater goes through primary and secondary treatment. The primary treatment consists of physically filtering out debris from the influent with bar racks and screens. After the primary treatment, the influent is split off into an East and West track to undergo secondary treatment which consists of an aeration basin and a clarifier. The aeration basin is responsible for the biological removal of toxic compounds in the wastewater, such as ammonia and nitrate. Together, these combined reactions create activated sludge. Mixed Liquor Suspended Solids (MLSS) measures the concentration of all suspended solids inside the aeration basin. It is important to aerate these basins accurately, otherwise, the necessary bacteria could die from malnourishment if there is a shortage of DO or lose to bacteria competing for oxygen if there is an excess amount of DO. Therefore, it is crucial to the success of the biological treatment for MLSS measures to be controlled. One way to troubleshoot an excess of activated sludge is to remove the surplus by way of a sludge wasting feed, sending it to a treatment specifically designed to neutralize harmful bacteria. A shortage of sludge is dealt with by initiating RAS (Recycled Activated Sludge) pumps to recycle some of the activated sludge from the clarifiers back into the aeration basins. Another noteworthy

variable is "percent influent return" which measures the percentage of the total influent wastewater that is returned into the aeration basins by way of RAS pumps. After aeration, each track moves on to the East and West clarifiers where the dense activated sludge settles and separates from the remaining liquid in the water. The remaining liquids and sludge recombine to form one track and move on to the last treatment phase of the wastewater plant. This last treatment is not described since the project does not explore it.

DATA DESCRIPTION

The data this team was provided with is split into lab data and online data. The lab data is collected by analyzing samples both in a rudimentary, in-house lab that provides enough accuracy and speed for operations adjustment and also a standard lab that provides heightened accuracy to pass state regulations. The lab data contains two datasets: Process Control and Overview. Process Control has 107 variables with 871 daily observations for each, starting from January 1, 2018, and ending on May 20, 2020. Overview has 68 variables and 867 daily observations for each day, starting from December 1, 2017, and ending May 20, 2020. The online data measures dissolved oxygen (DO) and oxidation reduction potential (ORP) levels using two sensors in each aeration basin. Therefore the online data has 5 variables with 400,883 observations recorded by minute from August 27, 2019, to May 31, 2020.

EXPLORATORY DATA ANALYSIS

The analysis began by exploring the behavior of various variables and relationships between them. First, the behavior of MLSS in the east and west aeration basin was explored by creating a histogram, which showed that the East track had a rightward skew and that the West track had a normal spread (Appendix Figure 1). MLSS from each of the tracks was also compared to the daily averages of DO and ORP from their respective tracks, and a linear regression line was fit on top of these plots to better see the relationship between the two (Appendix Figure 2,3 respectively). These plots showed that East MLSS has a positive linear relationship with DO and a negative linear relationship with ORP, while West MLSS has a negative linear relationship with DO and a positive linear relationship with ORP.

Next, the behavior of the clarifier ammonia was explored. Because the histogram of ammonia had an extreme right skew, a log transformation of effluent ammonia was used to achieve normality (Appendix Figure 4, 5 respectively). To further explore the differences between the two tracks, the log values were then plotted as boxplots with notches around the mean (Appendix Figure 6). The notches helped visualize the 95% confidence interval around the mean; if these notches do not overlap then it can be concluded that the means of the two boxplots are significantly different. The notches did not overlap in this case. The time series of daily and hourly averages of DO and ORP were also plotted, revealing periods of track shutdowns, which were subsequently removed from the analysis (Appendix Figure 7, 8 respectively).

STATISTICAL ANALYSIS and RESULTS

ANCOVA for Goal 1

The first goal was fulfilled by building an ANCOVA model to compare the mean MLSS values of each aeration basin after controlling for covariates. The ANCOVA tested the following hypotheses: the null hypothesis that the two means are the same; the alternative hypothesis that the two are different. The application of the first goal was predicting which basin was operating more efficiently

and effectively to formulate a conclusion on the productivity of the different diffusers in the basins. Therefore, all other variables that could impact the mean mass needed to be accounted for. If the influent flow for one basin was larger than the other, for example, the mean MLSS of that basin would be biased. One limiting factor of the ANCOVA is that to control for a covariate, it needed data for east and west individual measurements rather than total influent or effluent values. With this in mind, three covariates were identified: RAS, wasting sludge, and percent total influent return. However, the data did not include the influent flow into each basin, so an "assumed flow" variable, calculating 9 different potential pairings of influent flow amounts into the East and West basins, was created. The first pair had 10% in the East and 90% in the West, with each pair changing in increments of 10% until there was 90% in the East and 10% in the West for the 9th pair. Then, 10 different ANCOVA tests were completed, accounting for the first three covariates as well as the "assumed flow" covariate. The results are displayed in Appendix Figure 9. This plot proves that side-specific flow data is vital to making a correct prediction as to which oxygen diffuser is most effective, and the west side has a higher MLSS mean value on average than the east side regardless of flow.

Adaptive Lasso for Goal 2

To determine which operational and water quality variables have a high impact on effluent ammonia, an adaptive lasso model was constructed to regress on total effluent ammonia. The lasso model shrinks the coefficients of the less significant variables to 0, so it was utilized as a tool for variable selection. To this end, a dataset was constructed using both lab data and daily averages of online data, which then removed observations with a high proportion of missing values. Since a lasso model requires complete cases, any remaining missing values were subsequently imputed by an average of the two measurements on either side of the missing value. A transformation of $\log_{10}$ was applied to ammonia, which required the removal of observations with an ammonia concentration of 0 or less mg/L. The resulting lasso model was then analyzed for several different lambda values (a parameter used to penalize large coefficients in the model), and a model with a near-minimum mean squared error was chosen, which corresponded to seven variables. After analyzing these variables for relevance, the remaining five variables were: influent pH, drinking water sample chlorine concentration, reuse flow, and the RAS flow for the east and west tracks. However, the results found by the lasso do suffer as a consequence of the autocorrelation within the effluent ammonia, which inflated the significance of the variables. A dataset composed of the complete cases of every tenth observation was used to decrease the autocorrelation to about 20% for effluent ammonia between adjacent observations, with the lowest observed autocorrelation maintaining at least 60 observations. A linear regression model using this decreased dataset of six variables and 68 observations, combined from Overview and Process Control, indicated that RAS west, RAS east, and chlorine concentration were significant with p-values of 0.0151, 0.0182, and 0.0588, respectively. Reuse flow had a p-value of 0.2823 and influent pH had a p-value of 0.9052. This evidence supported the conclusions that RAS West had a negative relationship with effluent ammonia and that RAS East and chlorine concentration had positive relationships with effluent ammonia. The remaining variables had little indication of confidence in the nature of their relationship with ammonia, and plots comparing them to ammonia yielded little

new information. Subsequently, three of the five variables, RAS East, RAS West, and chlorine concentration, can be reasonably stated to be significant, but the other two variables, influent pH and reuse flow did not have the same amount of confidence in their significance.

CONCLUSIONS

The first goal was to prove the mean masses of the East and West tracks were different. The plant is currently undergoing updates that will add more DO and ORP sensors and increase control of influent flow splitting. Once these are complete, actual rather than assumed flow can be controlled for, and the change in mass in both tracks can be noted, allowing a conclusion on the oxygen diffuser productivity to be made. The second goal was to determine operational and water quality variables which are significant in impacting ammonia concentrations in the clarifier effluents. By analyzing the total effluent ammonia concentration, three variables were found to have significance: RAS flow in the east and west and chlorine concentration in drinking water. In addition, there were two variables that may have had a significant impact on ammonia but for which there is little statistical confidence: influent pH and reuse flow. Further research is needed to determine with increased accuracy the exact relationship these variables have with ammonia.

AUTHORS

Alina Gavrilov, senior
Humanitarian Engineering major
at Baylor University.

Rahul Banka, Sophomore

University Scholars Major at Baylor University, concentrating in Physics and Mathematics

Hunter Privett, senior
Applied Mathematics Major
at Baylor University

ACKNOWLEDGEMENTS

A special thank you to Wayne Ramey, our stakeholder, for his time and effort in providing the data and meeting with the team, Dr. Hering and Dr. Nychka for instruction on relevant statistical methods and organization of this fellowship, Kate Newhart for her indispensable subject matter expertise, and last but not least, Luke Durell for his constant support and help through every part of this project.

Appendix

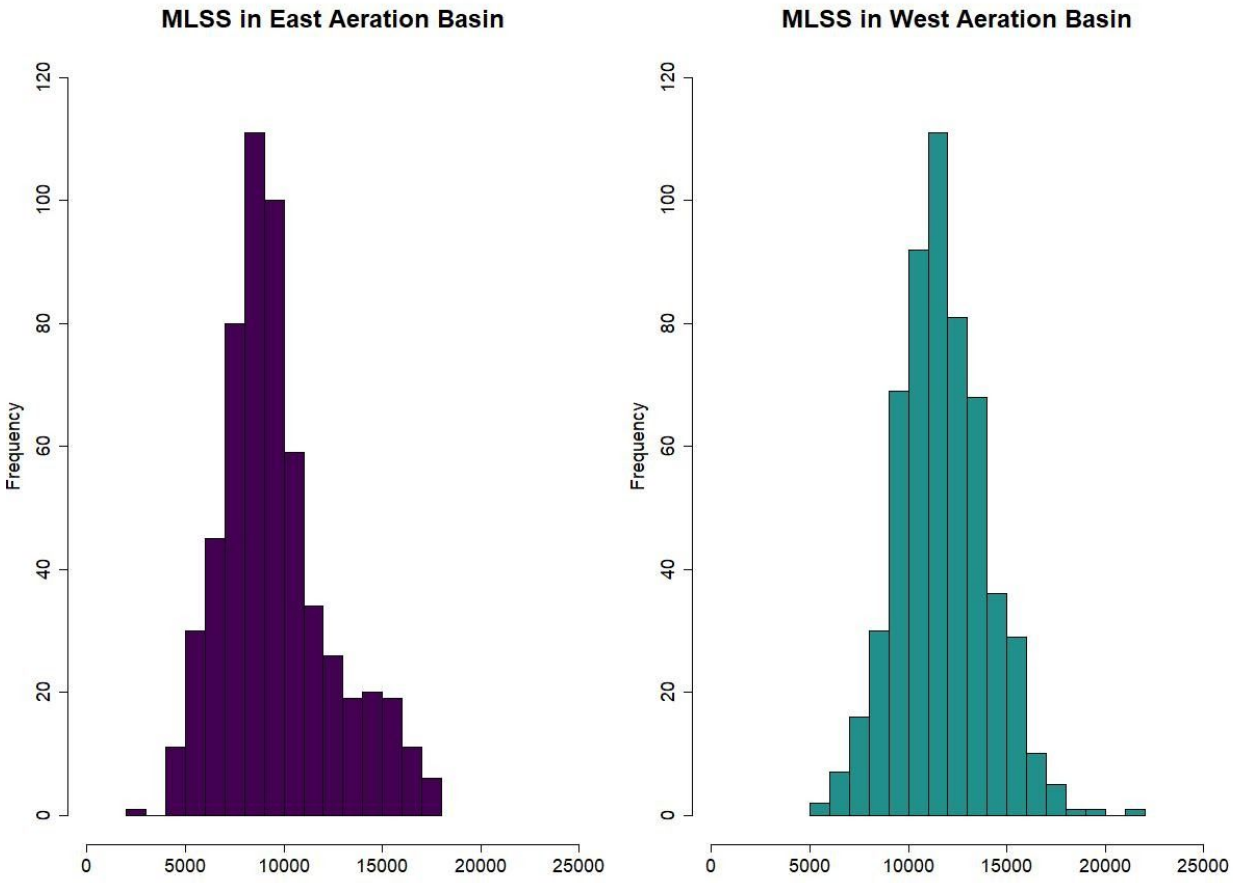

Figure 1. Histograms of East and West aeration basin MLSS.

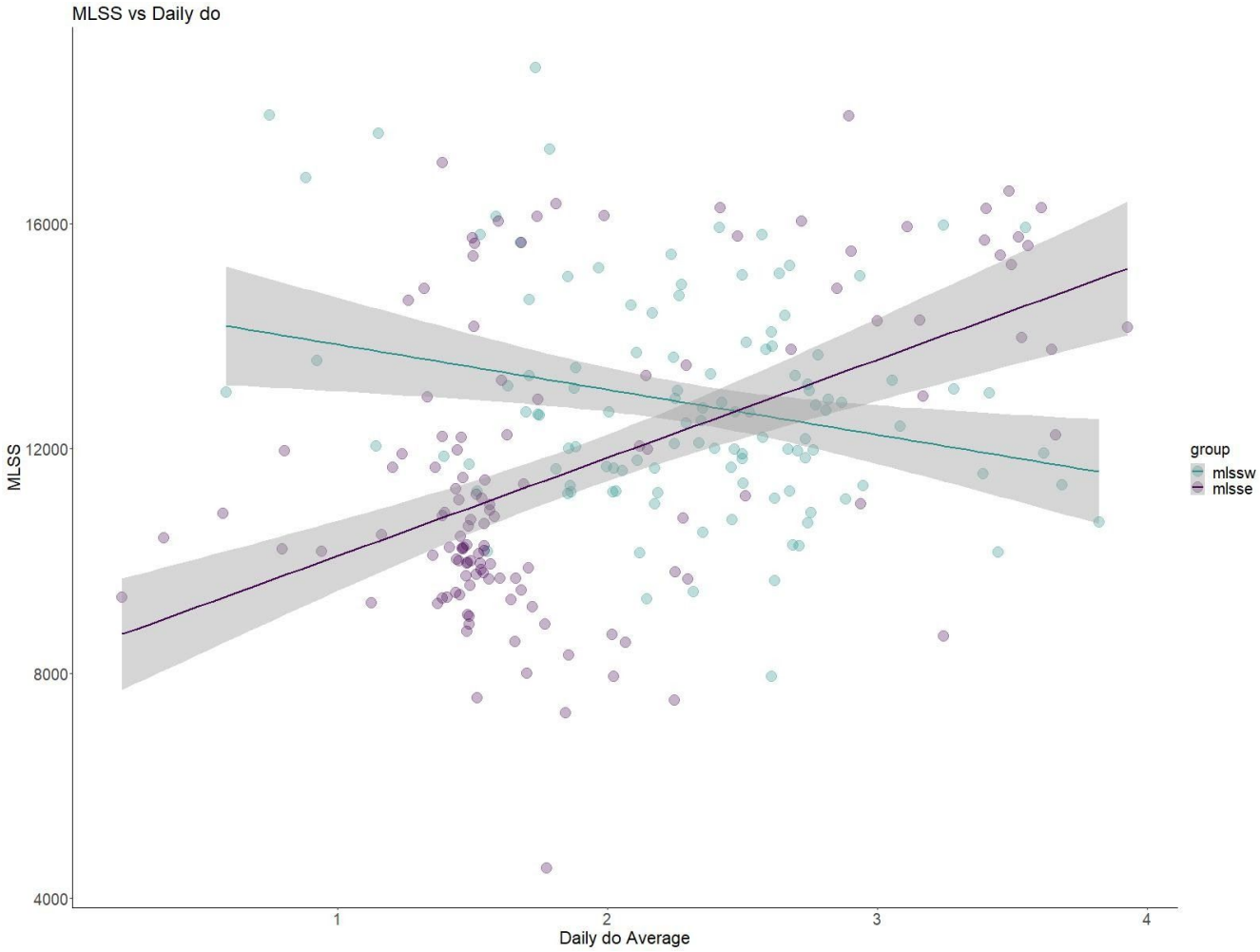

Figure 2. MLSS as a function of daily DO averages with a linear regression fitted to represent the linear relationships.

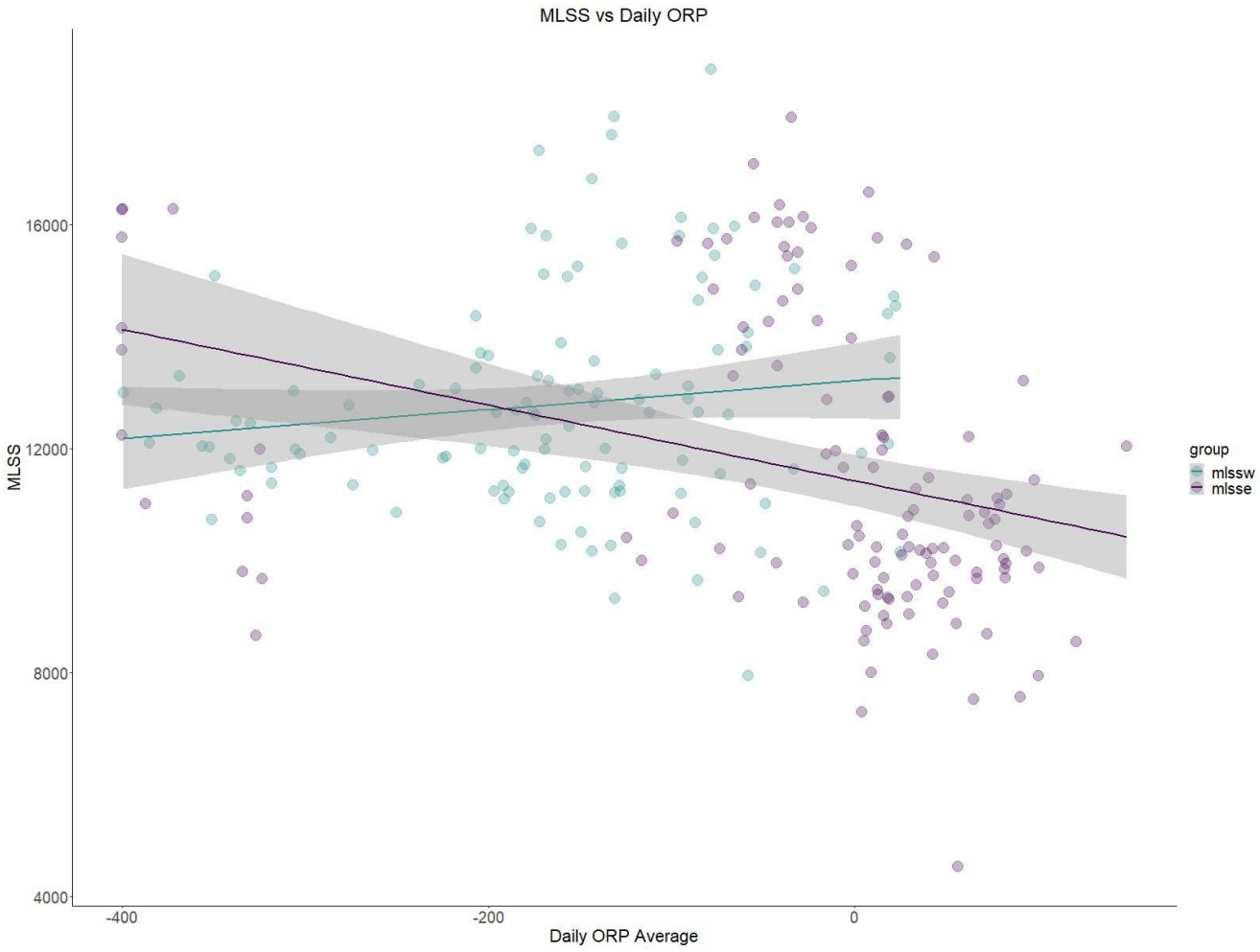

Figure 3. MLSS as a function of daily ORP averages with linear regression lines fitted to represent the relationships.

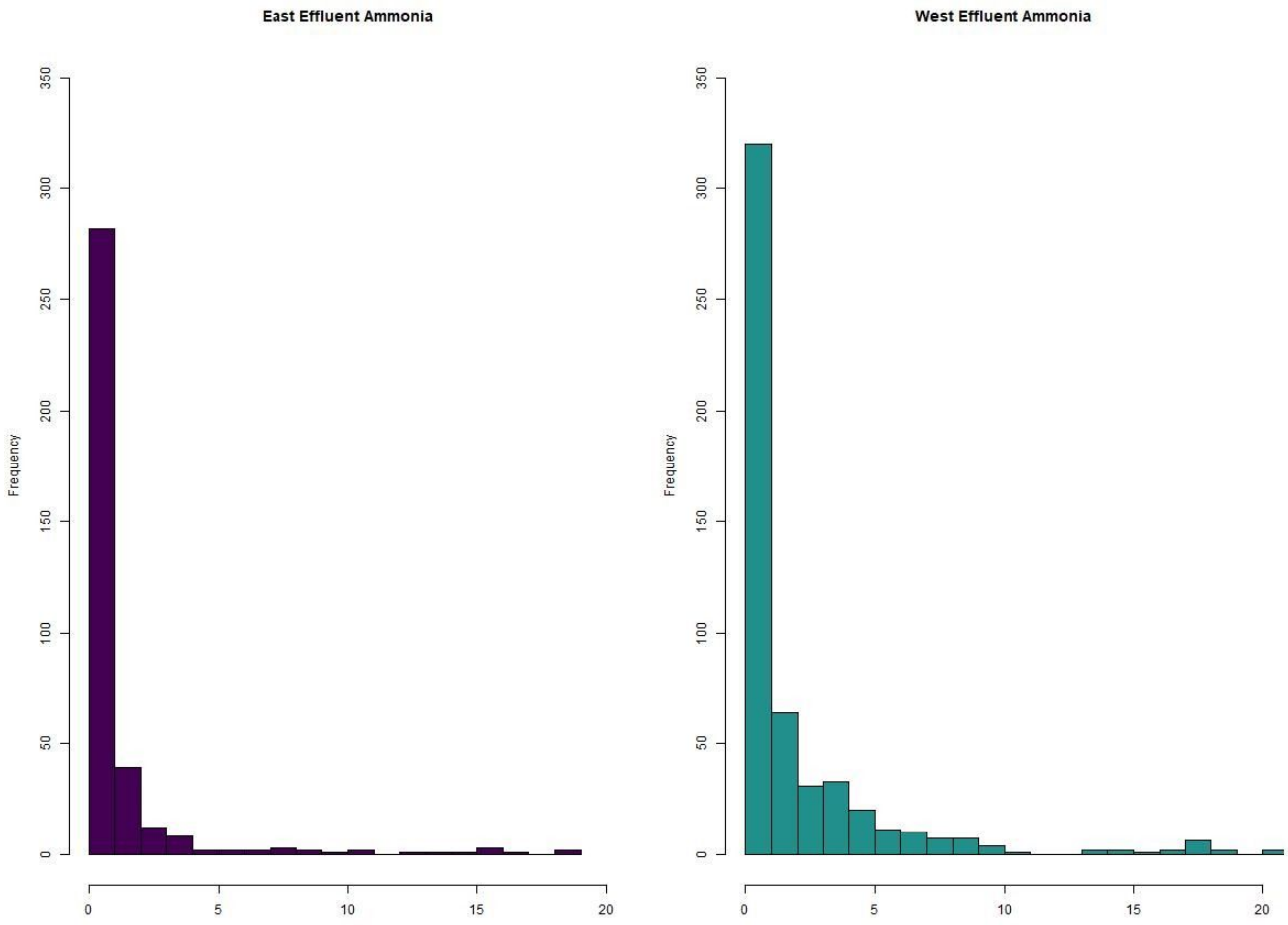

Figure 4. Histograms of East and West effluent ammonia.

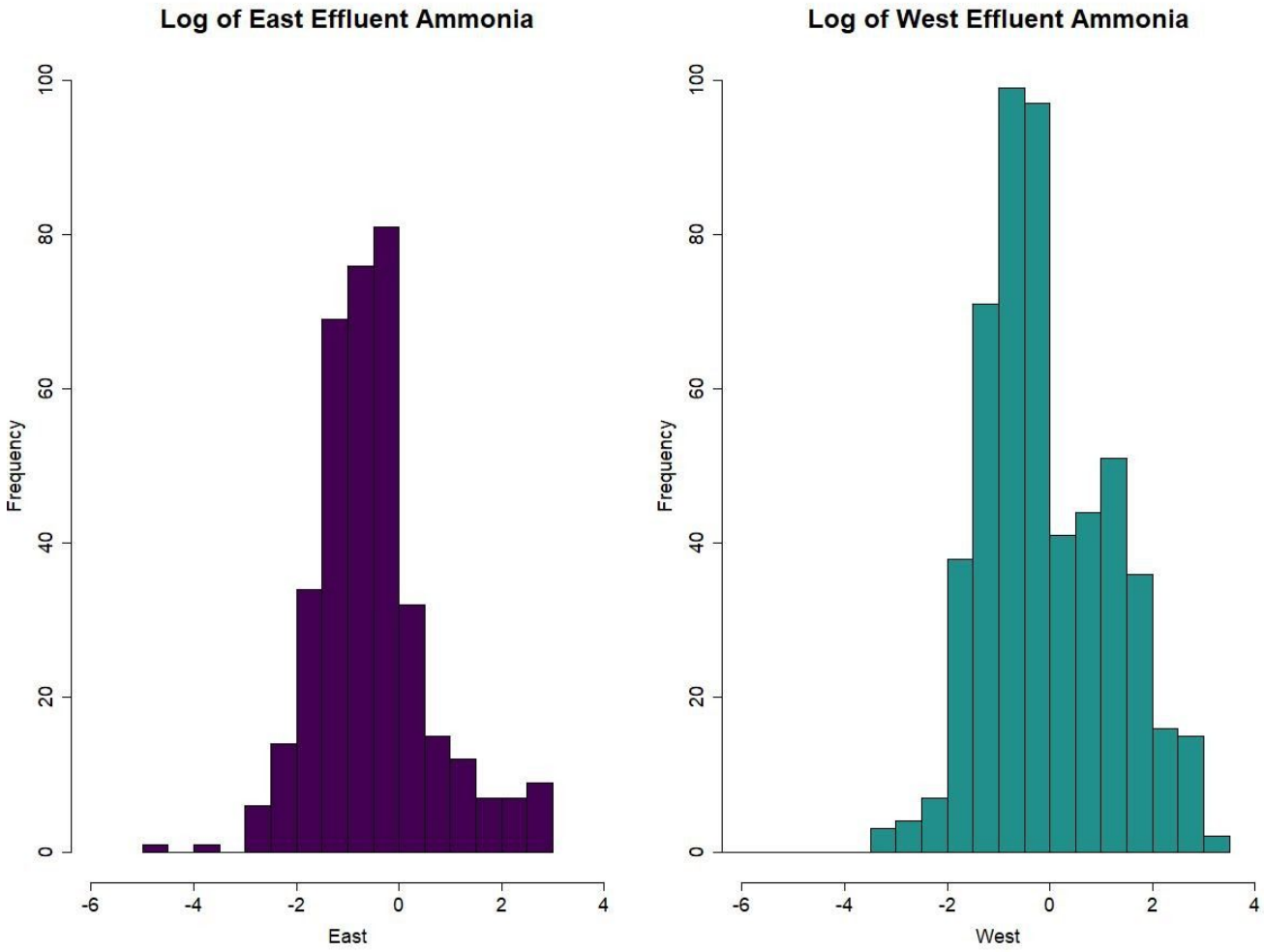

Figure 5. Histograms of the log transformation of East and West effluent ammonia.

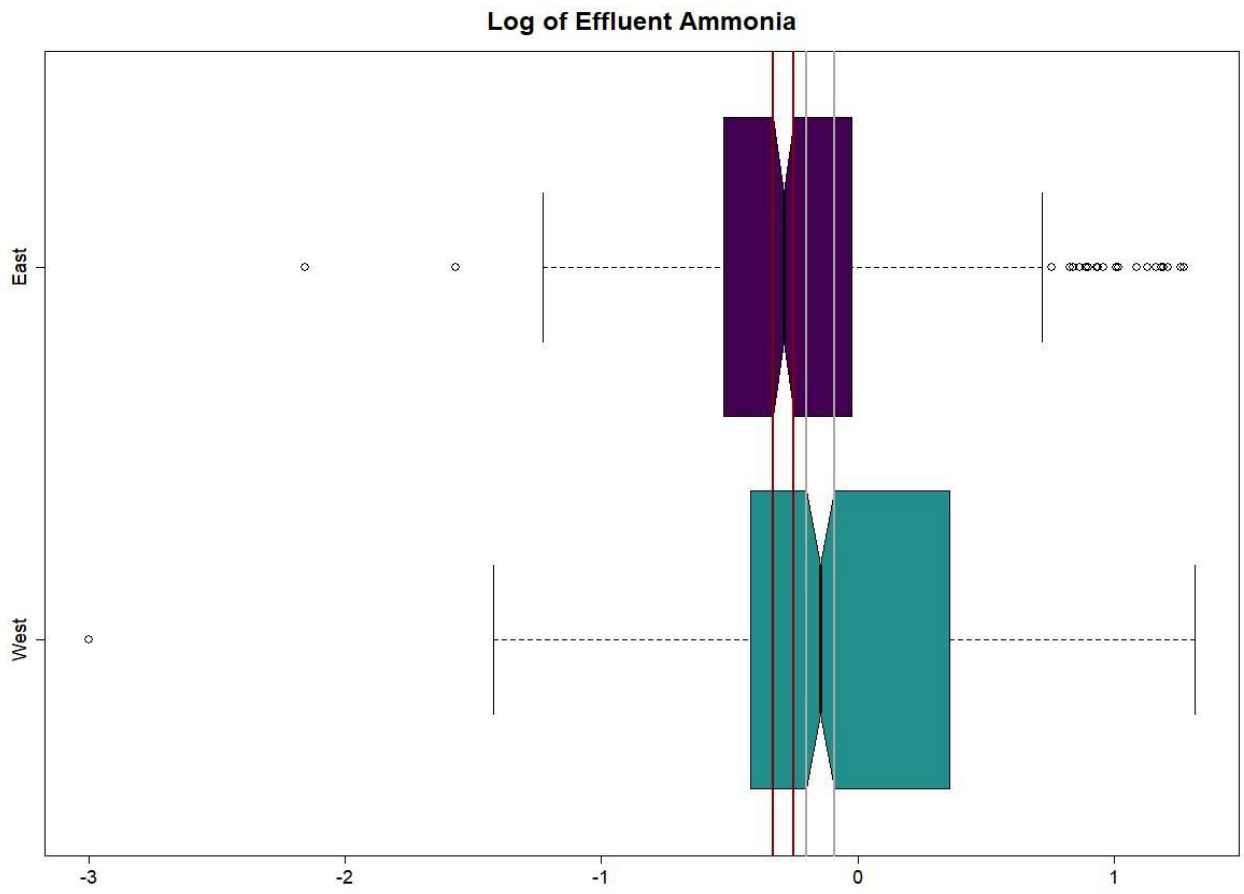

Figure 6. Boxplots of the log transformation of East and West effluent ammonia, with vertical lines to compare the confidence intervals around the mean.

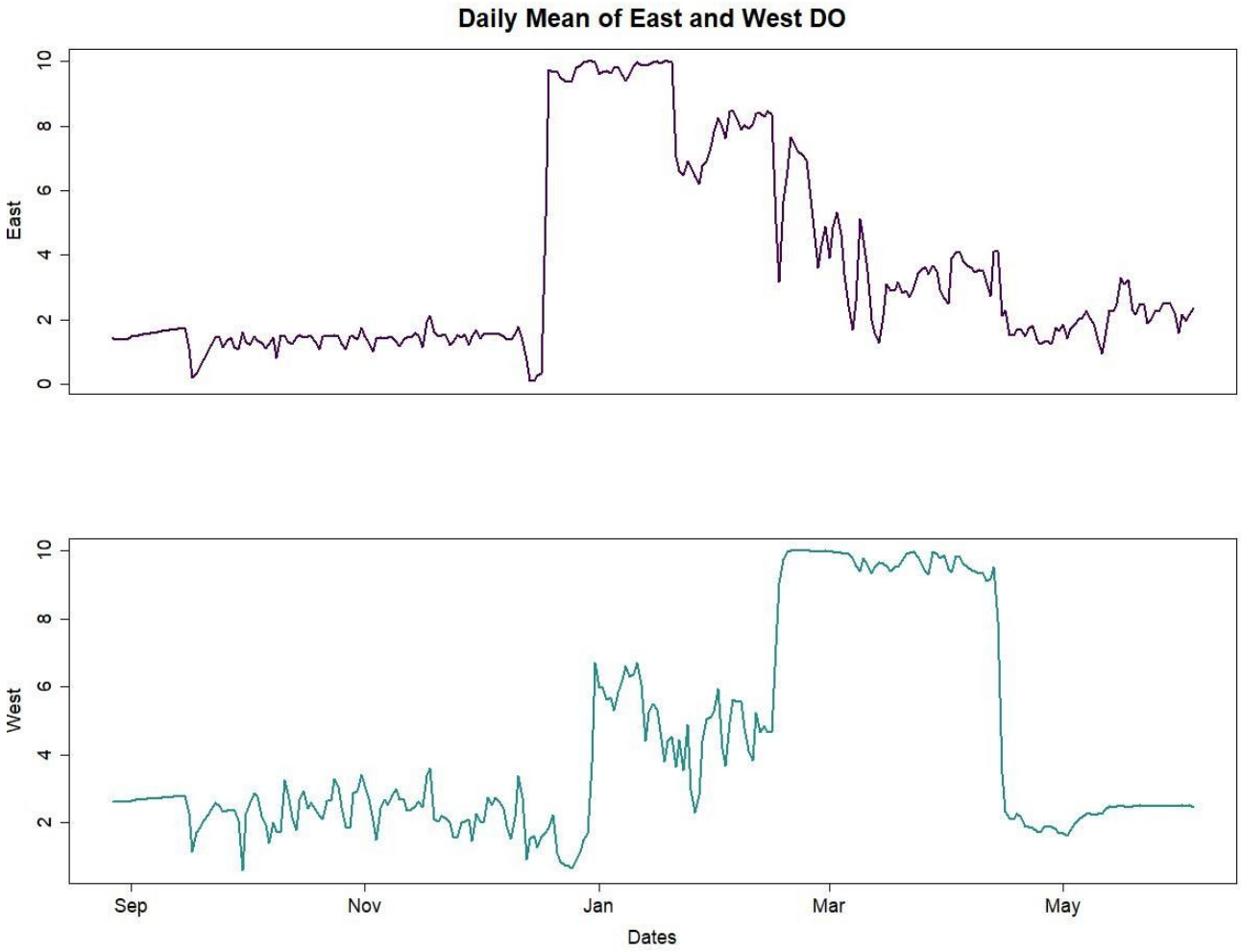

Figure 7. Time series plots of the daily averages of East and West DO.

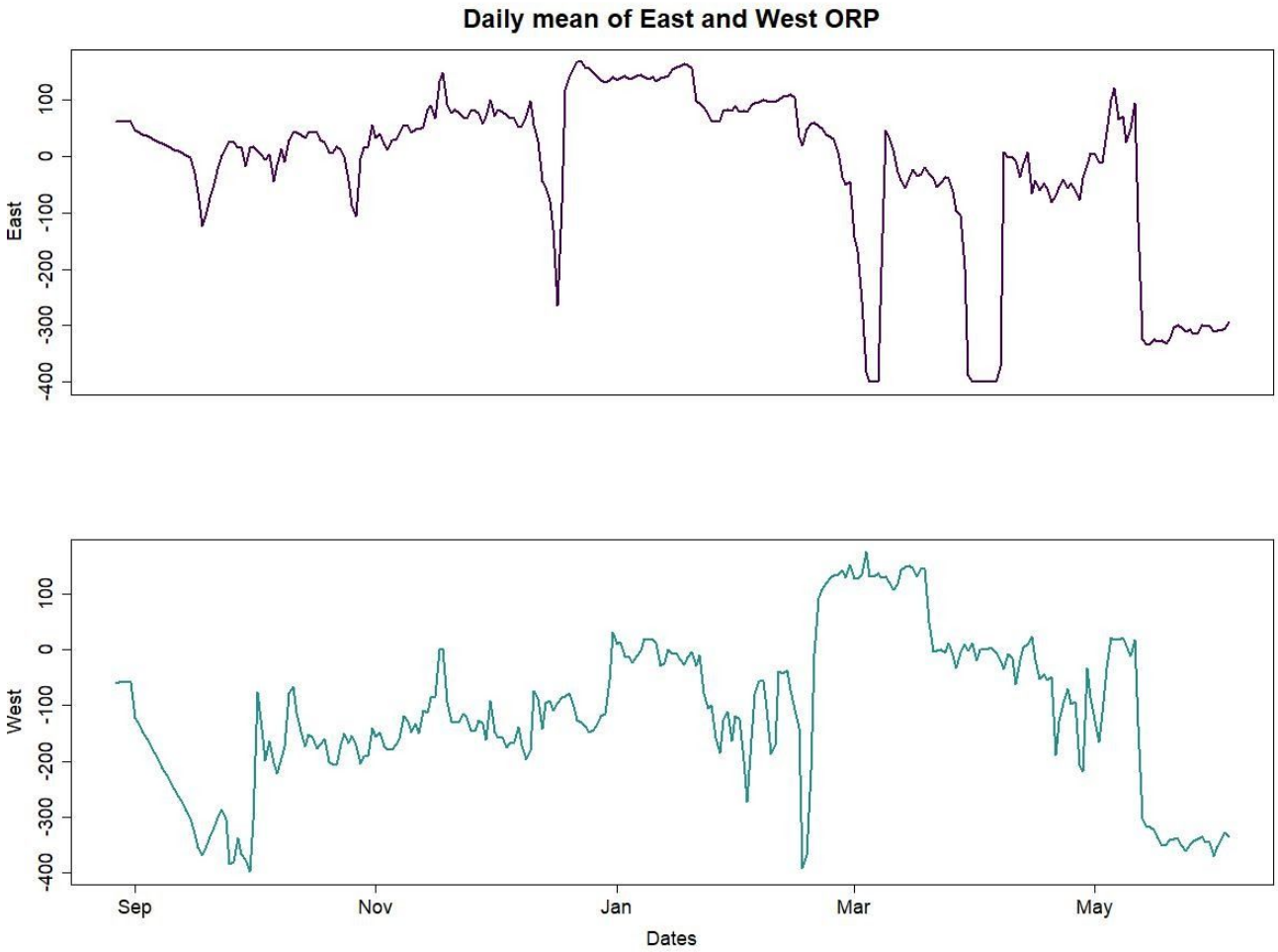

Figure 8. Time series plots of the daily averages of East and West ORP levels.

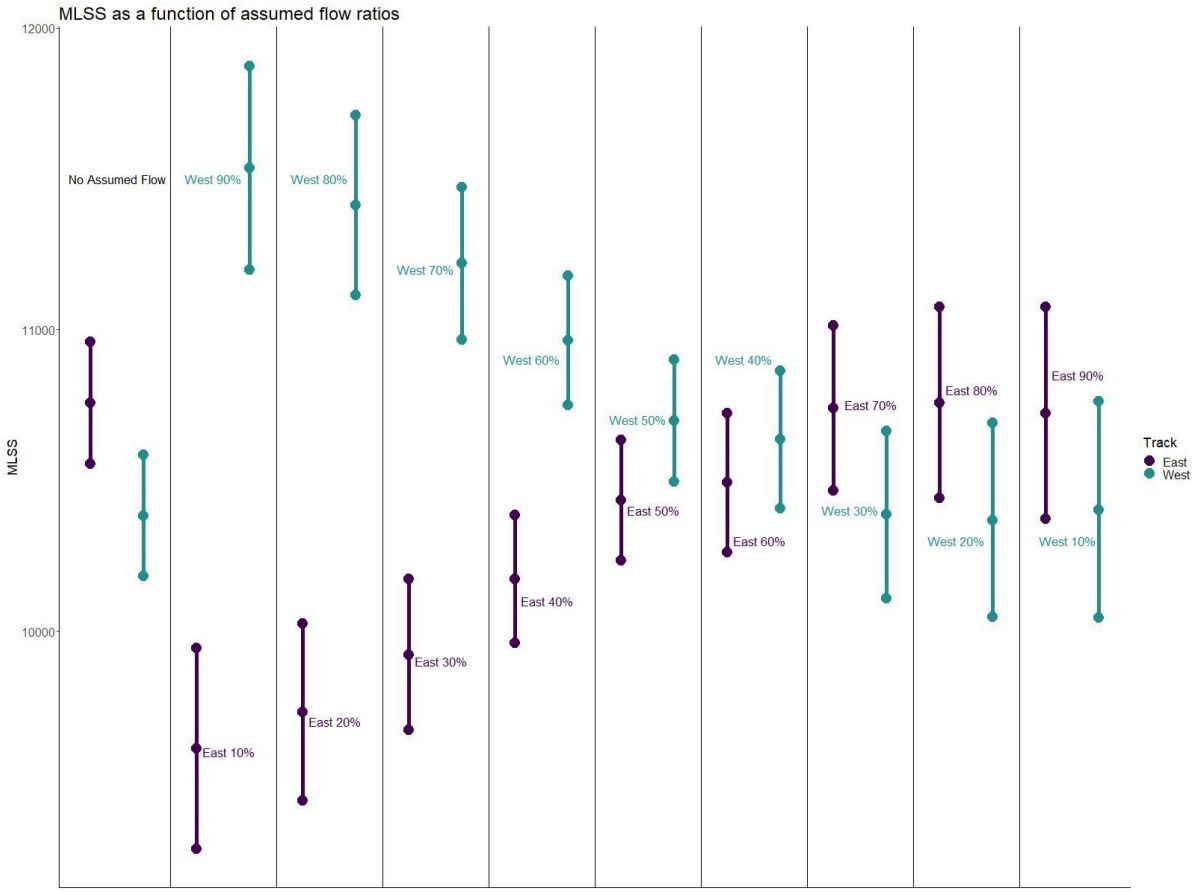

Figure 9. The results of the ANCOVA for mean MLSS values of the East and West without accounting for "assumed flow" is on the left, followed by ANCOVA results of mean values ordered by "assumed flow" variable pairs.