Title:

Use of Advanced Data Analytics and Model Predictive Control to Stabilize Ammonia-Based Aeration Control

Intelligent Water System Challenge - LIFT

Team 4:

City of Boulder, CO

Colorado School of Mines, CO

Baylor University, TX

Carollo Engineers, Inc

0 Background and Relevance

The need to remove nitrogen and phosphorus from wastewater to meet increasingly stringent regulatory limits is the most common treatment challenge for water resource recovery facilities (WRRFs) in the U.S. today. A majority of WRRFs have responded by implementing biological nutrient removal (BNR) processes and are clearly motivated to minimize operational costs of chemical addition and aeration energy and allow market-ready options for phosphorus resource recovery. While BNR processes have been extensively studied, employed, and optimized for over a century in practice and in academia, recent advances in microbiology, genetics, instrumentation, and process control technologies have enabled our industry to achieve new heights in BNR process efficiency and stability. This allows WRRFs to remove nutrients in smaller systems, with less energy input, to lower effluent standards. Through work with BNR utilities throughout the U.S., we recognize that BNR process improvements at WRRFs hinge on the ability to design systems that allow operators to maintain more consistent and accurate aeration control.

While every WRRF operator and engineer understands the importance of dissolved oxygen (DO) control for BNR processes performance, aeration remains one of the most complex and challenging systems in BNR facilities. Typical proportional-integral-derivative (PID) DO aeration control systems at WRRFs:

1. largely ignore real-time nutrient process conditions;
2. experience large fluctuations in air demand to maintain a single DO setpoint, which mechanically stresses the air blowers; and
3. stresses BNR biology with the resulting large and rapid DO fluctuations.

Model predictive aeration control (MPAC) is an alternative approach that can overcome PID control limitations. Incorporating real-time process data, MPAC can continuously fit model parameters to predict process conditions in the near future to allow for proactive control adjustments.

There are two main reasons why MPAC systems are not used more broadly by WRRFs:

1. **Lack of Understanding:** Engineers and utilities are skeptical and hesitant to hand control over to "statistical" or “data-driven” tools that are not based on mechanistic linear process logic. Furthermore, the mathematics of the methods are often unfamiliar and/or challenging to understand.

2. **Lack of Experience:** To date, there are only a handful of WRRFs in and outside of the U.S. who have implemented MPCA systems in their facilities. To our knowledge, there do not exist any side-by-side, full-scale comparisons of the efficiency and stability of MPCA with traditional PID control approaches.

Our IWS Challenge team set out to address these two limitations with this project.

1 Challenge Team Members

The members of Team 4 of the 2019 LIFT IWS challenge are listed in Table 1, and each covers a critical perspective and expertise related to the goal of this challenge.

Table 1. Team 4 Participants

Name	*Title*	*Organization*	*Contact*
Christopher Marks	Treatment Process Engineer	City of Boulder Public Works	marksc@bouldercolorado.gov
Kathryn Newhart	PhD Candidate Dept of Civil & Environmental Engineering	Colorado School of Mines	kbnewhart@mines.edu
Tzahi Cath	Professor Dept of Civil & Environmental Engineering	Colorado School of Mines	tcatch@mines.edu
Amanda Hering	Associate Professor Dept of Statistical Science	Baylor University	Mandy_Hering@baylor.edu
Tanja Rauch-Williams	Carollo's Wastewater Innovation Lead and Principal Technologist	Carollo Engineers	TRauch-Williams@carollo.com

The participating utility was the City of Boulder (City) Water Resource Recovery Facility (BWRRF, Boulder, CO). The City invested in major BNR process improvements by upgrading the facility to an activated sludge (AS) process in 2008 and a Four-Stage Bardenpho process in 2017. BWRRF operations staff recognizes that AS aeration control is vital to meeting energy efficiency goals as well as being a key factor in enhancing internal carbon management, maximizing nitrogen removal, and maintaining sludge settling quality. The City was interested in systematically evaluating the performance of the new ammonia-based aeration control (ABAC) system and evaluating the path of implementing MPAC as a means to gain more advanced control and further process goals.

Carollo Engineers (Carollo) has planned and designed the recent BNR process improvements for the City of Boulder and developed the PID control description for the ABAC system implemented at the BWRRF in 2017. Through BNR process optimization with the City and other utilities, Carollo has witnessed the need for advanced control systems in our industry. In 2017, Carollo therefore decided to dedicate a portion of its strategic annual research funds towards developing advanced automated process control solutions.

The Civil and Environmental Engineering Department at Colorado School of Mines (Mines, Golden, CO) and the Department of Statistical Sciences at Baylor University (BU, Waco, TX) have worked as university partners in recent years on developing novel multivariate statistical process monitoring (MSPM) solutions to enhance BNR performance and stability (Kazor et al., 2016; Odom et al., 2018). This project allowed CSM and BU to apply their experience with MSPM and their history of collaborative, interdisciplinary work to MPAC at the BWRRF.

2 Problem Statement

The Boulder Water Resource Recovery Facility (BWRRF) is a 25 MGD facility that utilizes an activated sludge (AS) process configured as a Four-Stage Bardenpho for secondary treatment. Aeration control previously used DO sensors and supervisory control and data acquisition (SCADA) setpoints to manage demands on the blower system. The current DO control mode frequently over-aerates causing conditions that inhibit denitrification in downstream anoxic zones, resulting in unnecessary high chemical carbon addition demands and in a poor allocation of blower demand that leads to energy inefficiencies.

To help address these problems, a non-proprietary ABAC scheme was programed into SCADA as part of the last BNR improvements in 2017. A control error during initial testing of the original ABAC system led the BWRRF operations team to continue conventional DO control instead of ABAC. A patch was created to fix the error, but operations never fully exercised the ABAC system and continued to regard DO control as a more reliable alternative.

The ABAC system itself is a four-level cascade control logic as depicted in Figure 1. Blue boxes indicate manually entered setpoints that determine the desired ammonia concentration and upper/lower DO setpoint limits. Black boxes indicate actual field measurements from ammonia sensors, DO sensors, flow meters, and valve position sensors.

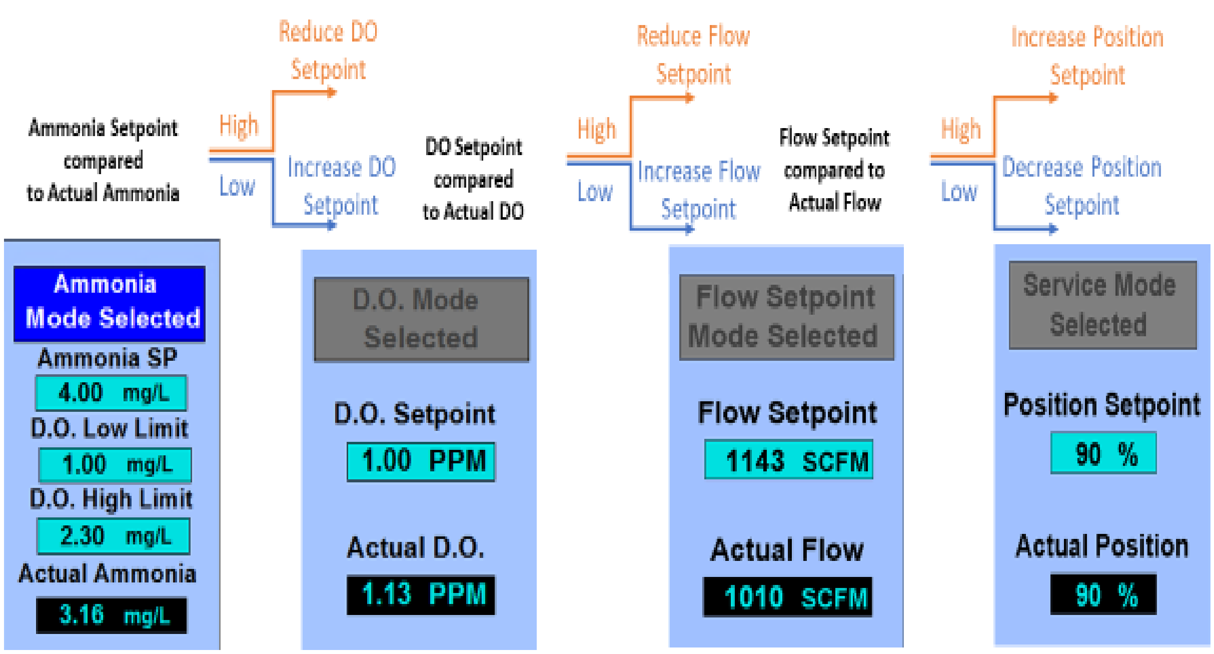

Figure 1. Four-level cascade control logic for ABAC. DO control removes the first level of control and functions without ammonia readings in the same fashion.

This IWS Challenge project set out to address two problems:

1. The PID logic can result in significant DO fluctuations when ammonia measurements exceed the ammonia concentration setpoint and the controller calculates DO setpoint adjustments that are the process must respond to rapidly. PID tuning can reduce these DO setpoint fluctuations, but the PID parameters perform best only within a narrow range of operating conditions. The first objective of this project was to determine a systematic method for testing and comparing ABAC tuning parameters by characterizing the stability of the overall process under different testing conditions.

2. If the ammonia concentration in each aeration basin used in ABAC for the DO setpoint definition could be forecast in advance of the required system changes, this could overcome the inherent response lag time of the PID controller that can lead to excessive changes in blower demands. Our team employed statistical modeling tools and online data collected from the process (ammonia, DO, airflow, etc.) to attempt to forecast the ammonia concentration. Forecast ammonia values (if sufficiently accurate) can then be used in lieu of the sensor ammonia values for DO setpoint adjustments. (i.e., as a software or 'soft' sensor). This could help stabilize the changes in aeration demand.

The questions being addressed for the Challenge are:

1. Can ABAC be a stable operating mode for BWRRF's activated sludge system compared to the traditional DO control?
2. Can ammonia concentrations in the aeration basins be forecast with sufficient accuracy to be a useful soft-sensor within the ABAC control logic to further stabilize aeration control?

The desired outcomes of this project and testing phase are multi-tiered. The primary goals directly relevant for the City are to:

a) improve energy efficiency by reducing blower system demand;
b) eliminate DO-poisoning of denitrification zones from over-aeration; and
c) gain trust of BWRRF operations staff for long-term use of ABAC system.

The secondary goals are generally of broader interest for WRRFs to:

a) develop appropriate mathematical metrics to characterize and compare process variability between different aeration control conditions (i.e., DO, ABAC, tuning parameters, etc.);
b) use real-time process data to model and forecast ammonia concentrations in the aeration basins that can then be used to improve DO setpoint selection in the ABAC logic; and
c) outline the path beyond this project for implementing MPAC feedforward control at the BWRRF.

3 Characterization of the Intelligent Water System

3.1 Boulder Water Resource and Recovery Facility

The BWRRF utilizes three parallel aeration basins depicted in Figure 2 that operate in a Four-Stage Bardenpho process configuration. The effluent from these basins combines in the mixed liquor channel before entering a final polishing aeration basin. Primary clarifier effluent is combined with return activated sludge (RAS) and internal mixed liquor return in a chimney baffle located at the inlet of zone 1 in each aeration basin. Each basin contains three influent anoxic zones, five aerated zones, and a second anoxic zone in a three-pass layout. Zone 8 is the last aerated zone in the basins from where mixed liquor is recycled to the chimney baffle at the head of each respective basin. Zone 9 is typically operated as the second anoxic zone in the Bardenpho configuration. The solids retention time (SRT) is controlled using mixed liquor wasting from the mixed liquor channel. Effluent from the final polishing aeration basin is routed to the secondary clarifiers, and all RAS from the clarifiers is combined and then split evenly among the chimney baffles of the aeration basins in service.

The ABAC uses ammonia concentrations measured in the second pass of each aeration basin for control of the DO setpoints in each aerated zone. The four setpoints in the ABAC control logic are: ammonia, DO, airflow, and valve position (see Figure 1). The ABAC uses five operator defined parameters for each of the three aeration basins:

1. ammonia setpoint;
2. high and low DO setpoint;
3. time delay for control setpoint adjustments;
4. small and large step ammonia measurement deviation; and
5. DO setpoint in each aerated zone.

The ABAC logic compares the actual ammonia concentration with the ammonia setpoint (typically set for Zone 7). If the measured concentration deviation is larger than the defined "large step deviation" (LSD), then the DO setpoint for each aerated zone changes according to the defined associated step value. If the measured deviation is between the small and larger step deviation, then the DO setpoint is adjusted using the small step. Finally, the system makes no changes if the ammonia measurement deviation is smaller than the defined the small step deviation (SSD).

Figure 2. Activated sludge process overview at the BWRRF. The purple box in upper left indicates the RAS splitting structure. One of the three aeration basins is outlined in orange on the right and shows the direction of flow through the three-pass system. The green boxes with yellow text overlain in the middle basin indicate the different zones within each of the three basins. Red boxes indicate where ammonia (Zone 7 or Z7) and nitrate (Z3 & Z9) instrumentation is located in each basin, and blue boxes indicate the location of DO sensors in each basin.

The DO setpoint values are restricted by maximum and minimum DO setpoints set by the operator, however Zone 8 has only a maximum DO setpoint. Each aerated zone also has a minimum air flow rate to satisfy minimum mixing requirements and airflows per diffuser.

3.2 Data Collection and QA/QC

The DO concentrations in Zones 6, 7, 8, and 9 of the aeration basins were continuously monitored using Endress Hauser (Reinach, Switzerland) COS61D optical DO sensors. In these zones, blower flowrate and valve position are also monitored and recorded in SCADA. AmmoLyt® Plus 700 ion-selective ammonia sensors from YSI (Yellow Springs, OH) are located in the aeration basin influent channel and in all three aeration basins in Zone 7. Ion-selective nitrate/nitrite sensors from YSI are located in Zones 3 and 9 of each basin; and the effluent channel after the final polishing aeration basin has an AmTax and NitraTax online analyzer (HACH, Loveland, CO). Aeration basin influent flow rates, wastewater temperature, and pH of the plant influent were also monitored. Online sensors were regularly maintained and calibrated by operations staff and readings periodically compared to laboratory results.

Data are collected and managed in the GE Proficy® system. For analysis, data was exported in 5-minute intervals into Microsoft Excel and imported to the statistical platform, *R*, for analysis. Observations that were identified as “Bad” within the Proficy system (i.e., due to sensor calibration or during power loss) were not used. For QA/QC, all Proficy exports were kept in their native format, and the same *R* functions were used to analyze all datasets to ensure repeatability.

4 Testing and Analytical Plan

The project was executed in the following three phases:

Phase 1: Implementation and Functional Testing of ABAC

Initially, the ABAC control scheme was operated only in one of the aeration basins during the day when supervised for functional testing. Then, operation was extended to overnight and on weekends. An initial ammonia setpoint of 3.5 mg-N/L was used for the middle of Zone 7 based on previous ammonia profiles in the aeration basins that indicated that this concentration would result in near complete nitrification by the end of Zone 8. Other ABAC setpoints were maintained throughout Phase 1 as follows:

LSD: 0.50 mg-N/L LSD Gain: 0.10 mg/L as DO

SSD: 0.15 mg-N/L SSD Gain: 0.30 mg/L as DO

Time delay for control setpoint adjustments: 90 s

Phase 1 lasted from April 2 – April 17, 2019.

Phase 2: ABAC Setpoint and Performance Testing

Once operators had gained confidence that the ABAC system could be run continuously, various tuning parameters were tested, and performance data was collected for each test run (Table 2).

Each of the three test runs was conducted for at least 8 days, and online process data were collected for statistical process evaluation and modeling. Test Run 1 was conducted with the default ABAC tuning parameter. In Test Runs 2 and 3, the Ammonia Setpoint and Time Delay parameters were increased to assess whether this would positively impact the aeration variability and process stability. The parameters for the LSD and SSD were kept as in Phase 1. During Test Run 0, process data for DO Control were collected from one of the three aeration basins.

Table 2. Phase 2 ABAC Setpoint and Performance Test Runs

Test Run	*Date (2019)*	*Test Condition*	*ABAC Setpoints*
0	1/1-4/1	DO Control	NA
1	4/18-4/28	Default ABAC Tuning Parameters	Ammonia Setpoint: 3.5 mg/L Time delay: 90 s
2	4/30-5/13	Increased Ammonia Setpoint	Ammonia Setpoint: 4.0 mg/L Time delay: 90 s
3	5/14-5/21	Increased Time Delay	Ammonia Setpoint: 4.0 mg/L Time delay: 300 s

Phase 3: Analysis and Forecasting

In Phase 3, our team used the data collected during the Test Runs in Phase 2 to quantify and compare the variability and stability of the DO and ABAC control conditions. We also quantified the operating parameters that had the greatest impact on control variability. Secondly, our team developed a data-driven model using a combination of multiple regression models in order to forecast ammonia concentrations in Zone 7 of the aeration basin.

5 Testing and Analysis Results

5.1 Implementation and Functional Testing of ABAC

Initial deployment of ABAC at the BWRRF supervised by operations confirmed that DO concentrations were maintained and ammonia concentrations did not increase in the effluent of the secondary process. These two conditions were met in each of the three aeration basins for at least three consecutive days between the hours of 07:00 and 18:00 (while the BWRRF is staffed). Based on this, ABAC was operated continuously to collect data and refine ABAC setpoints in Phase 2.

5.2 Comparison of Aeration Control Logic Process Performance

Air flow to aeration basin 3 (AB3) is depicted in Figure 3 for all Test Runs (see Table 2). The data collected in April and May 2019 represent both DO and ABAC control systems, and for ABAC two different ammonia setpoints for Zone 7, and two different time delay setpoints.

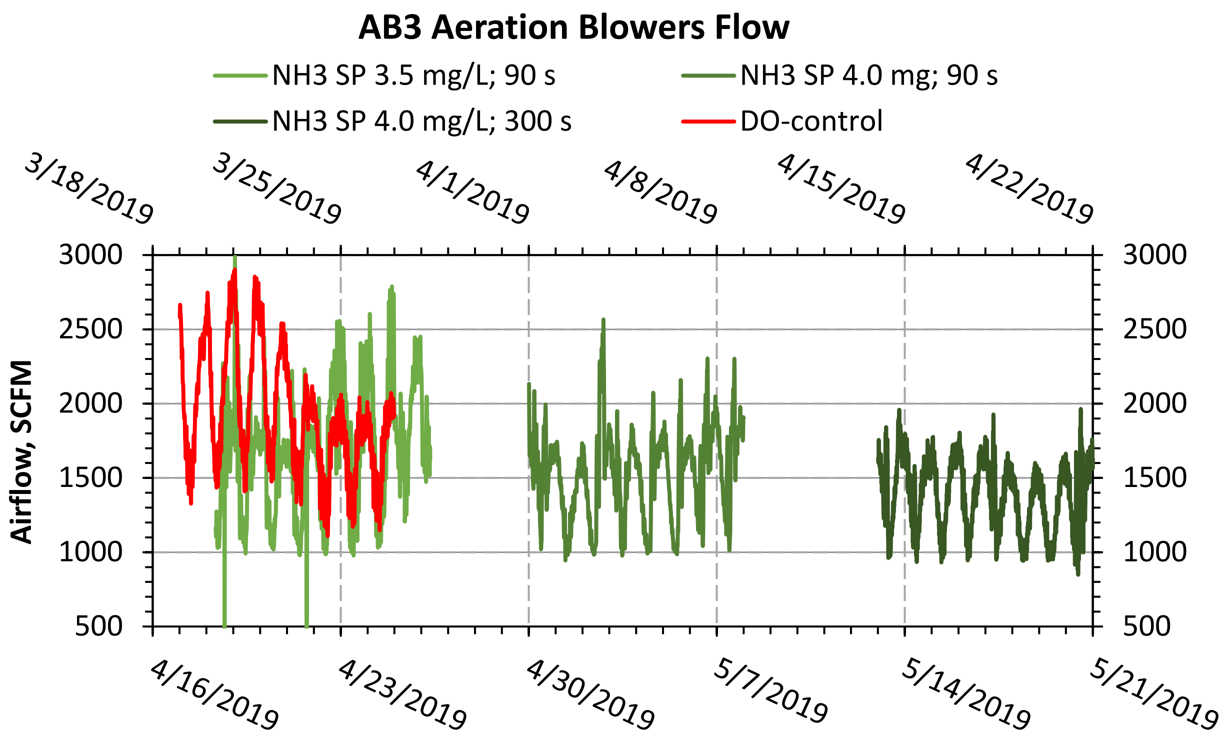

Figure 3. Time series plot of air flow to AB3 (sum of airflow to all aeration zones in AB3) in standard cubic feet per minute (SCFM) during the four Test Runs in Phase 2.

The maximum airflow was the same between Test Run 0 and 1 (DO and the first iteration of ABAC at 3.5 mg/L N and 90 s time delay) but the maximum daily airflow did decline as the ABAC parameters changed (i.e., increase ammonia setpoint and increase time delay). The minimum mixing air flow required in AB3 (approx. 1000 SCFM) was never achieved by DO control, but ABAC control reached the minimum mixing air demand every night during low flow conditions.

The daily average air flow for all test runs is depicted in Figure 4 by day of the week and as weekly averages. Average concentrations of DO within Zones 6 to 9 are depicted in Figure 5 for each aeration control logic tested. From the mean weekly air demand, ABAC could achieve an average reduction of 200 to 840 standard cubic feet per minute (SCFM) of air flow to the aeration basins. From an operations perspective, this air flow reduction to the secondary treatment process allows the utility to operate on a single blower during a wider range of treatment scenarios, and ultimately reduces energy demand and the cost of treating water.

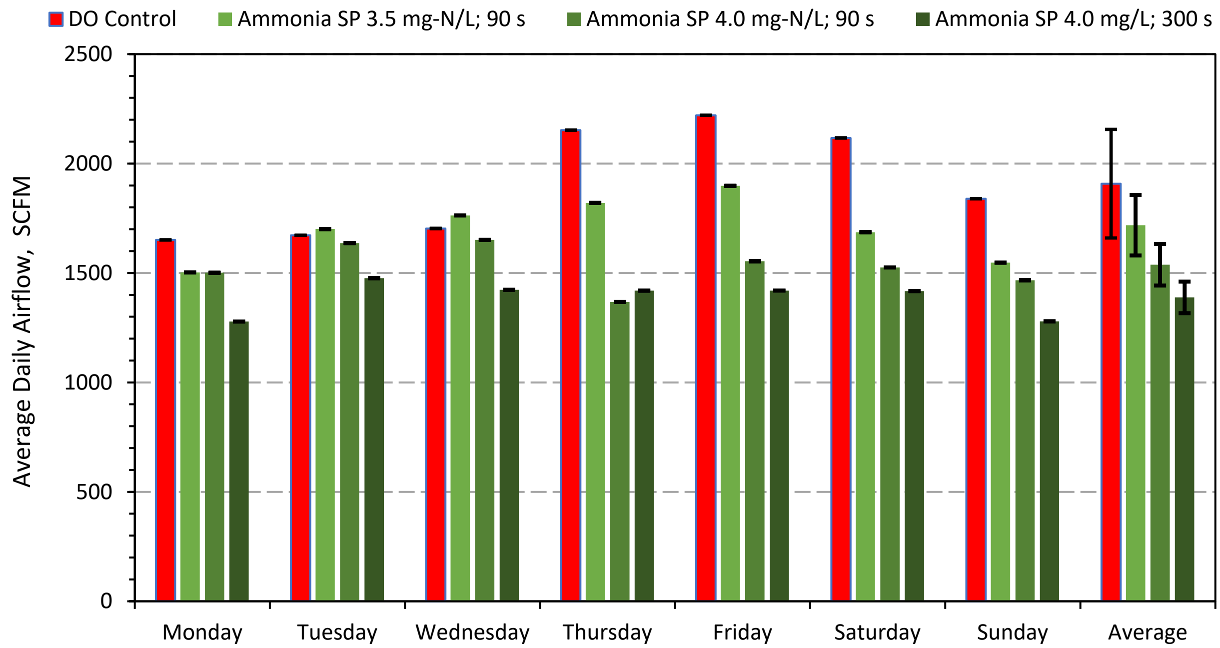

Figure 4. Average daily aeration blower flow, and average weekly demand during Phase 2 testing.

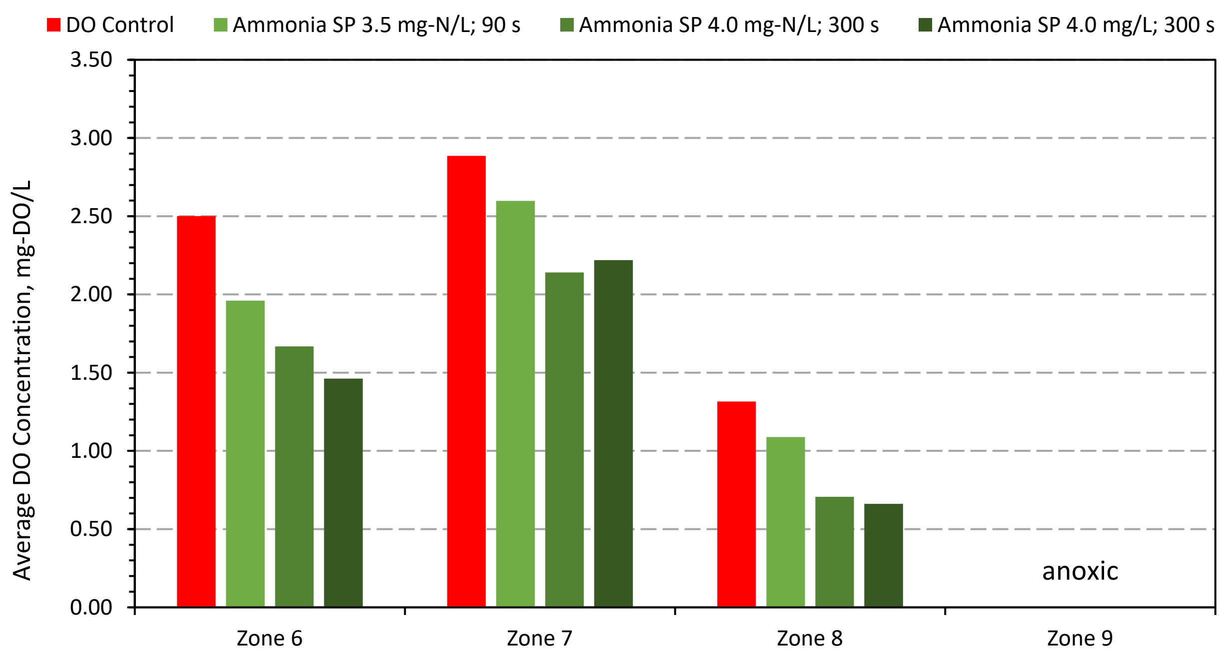

Figure 5. DO concentrations in Zones 6 - 9 during all Test Runs in Phase 2.

In all ABAC Test Runs, the DO concentrations in the aerated zones continuously decreased. Test Run 3 (ammonia setpoint of 4.0 mg-N/L, time delay: 300-s) resulted in a DO reduction entering Zone 9 of 50%, thereby significantly reducing DO poisoning of the final anoxic zone and reducing the amount of external carbon addition. At no time during Phase 2 testing was ammonia detected in the final plant effluent indicating incomplete nitrification. Also, the low DO filaments that commonly overwhelm secondary settling when aeration is reduced too much have not been a problem, nor have increases been apparent in weekly biological assessments by operations.

These test results have been well received by operations staff due to an observable reduction in aeration demand and the lack of process upsets or failures during ABAC testing. After the successful deployment and operation of ABAC in AB3, this control has now been completely adopted by operations staff at the BWRRF in all aeration basins.

5.3 Development of Metrics to Characterize and Compare Process Variability

Process variability is often "eyeballed" based on time series plots and simple statistics shown in Figures 3, 4, and 5. However, these methods only allow a cause-effect comparison of a very limited number of process parameters (e.g. ammonia setpoint and air flow).

Our team was interested in identifying metrics that allow a quantitative assessment of the variability of the aeration basin system as a whole, rather than monitoring each online process variable individually (e.g., air flow, nitrate, ammonia, DO, valve position). This approach has clear advantages to assess the overall performance of complex systems, such as activated sludge that include mechanical components, process, and water quality variables.

Two statistical methods and two metrics for each method were compared to assess the system variability in response to the four Test Run conditions in Phase 2 (Table 3). Total sample variation (TSV) incorporates the individual variances of process variables. Whereas generalized sample variance (GSV) incorporates the individual variances of process variables as well as the pairwise correlations between process variables. To compare the variation (i.e., stability) between two Test Run conditions, the TSV and GSV values for each Test Run were compared.

Table 3. Variability Assessment Methods and Metrics

Method	*Explanation*	*Metric*	*Hypothesis Test*
Total sample variation (TSV)	The trace (sum of the diagonal elements) of the sample variance-covariance matrix	*Difference* between TSVs of two data sets.	H_0: $TSV_i - TSV_j = 0$ H_1: $TSV_i - TSV_j \neq 0$
		Ratio of TSVs of two data sets.	H_0: $TSV_i / TSV_j = 1$ H_1: $TSV_i / TSV_j \neq 1$
Generalized sample variance (GSV)		*Difference* between GSVs of two data sets.	H_0: $GSV_i - GSV_j = 0$

	The determinant of the sample variance-covariance matrix		H_1: $GSV_i - GSV_j \neq 0$
		Ratio of GSVs of two data sets.	H_0: $GSV_i / GSV_j = 1$ H_1: $GSV_i / GSV_j \neq 1$

The TSV and GSV were calculated using all process variables identified in Section 3.2 in 5-minute intervals. Figure 6 shows the TSV for each Test Run conducted under Phase 2 (Table 2). It is hypothesized that if the ABAC system is more stable, then the ABAC TSV will be smaller than the TSV of data collected under the Test Run 0 (DO control). This is demonstrated for the ABAC Test Run 2 and 3 conditions that each have a smaller TSV than the Test Run 0 (DO Control).

However, the TSV for Test Run 1 (ABAC with ammonia set point 3.5 mg/L) was greater than for Test Run 0 (DO control). In fact, when plotting ammonia concentrations in Zone 7 over time for each Test Run, it becomes evident that Test Run 1 resulted in different environmental conditions than the other three Test Runs (Figure 7). The first half of the Test Run 1 (up until about 3.5 days) did not exhibit the same dual-peak daily pattern as the second half of the Test Run or in any of the other Test Run conditions. Secondly, the daily ammonia peak concentration was significantly greater than in the other test conditions (i.e., 6-10 mg/L). Both factors contributed to a higher TSV for Test Run 1 than anticipated. It is therefore recommended that the ABAC Test Run 1 condition be either repeated or excluded from the interpretation of the results.

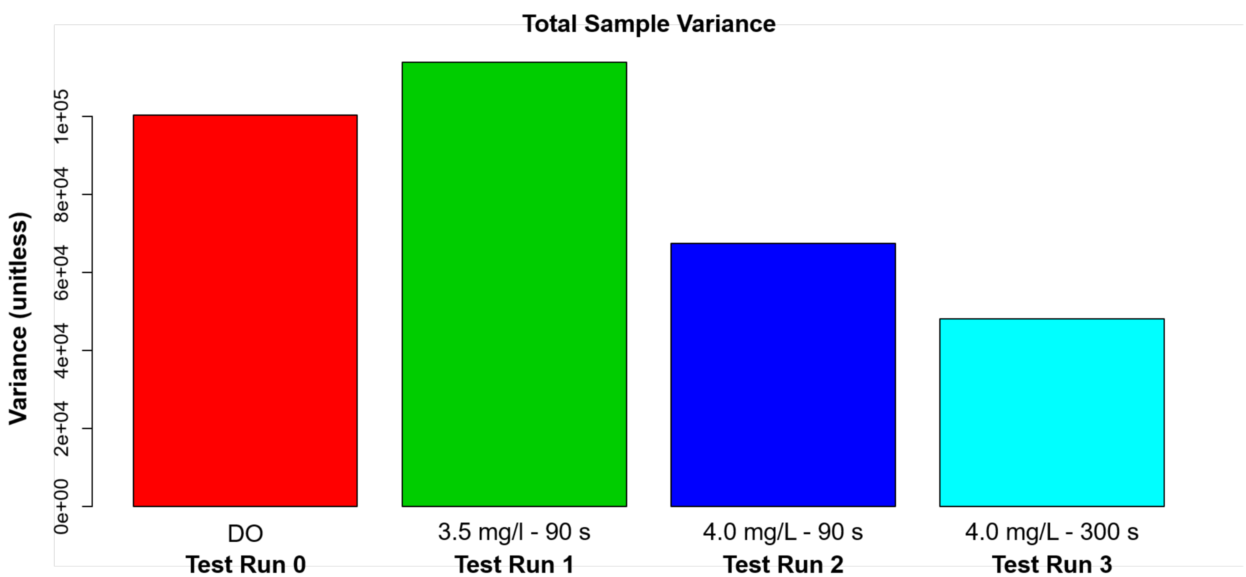

Figure 6. TSV for 7 days of process data under different control logic Test Run conditions in Phase 2.

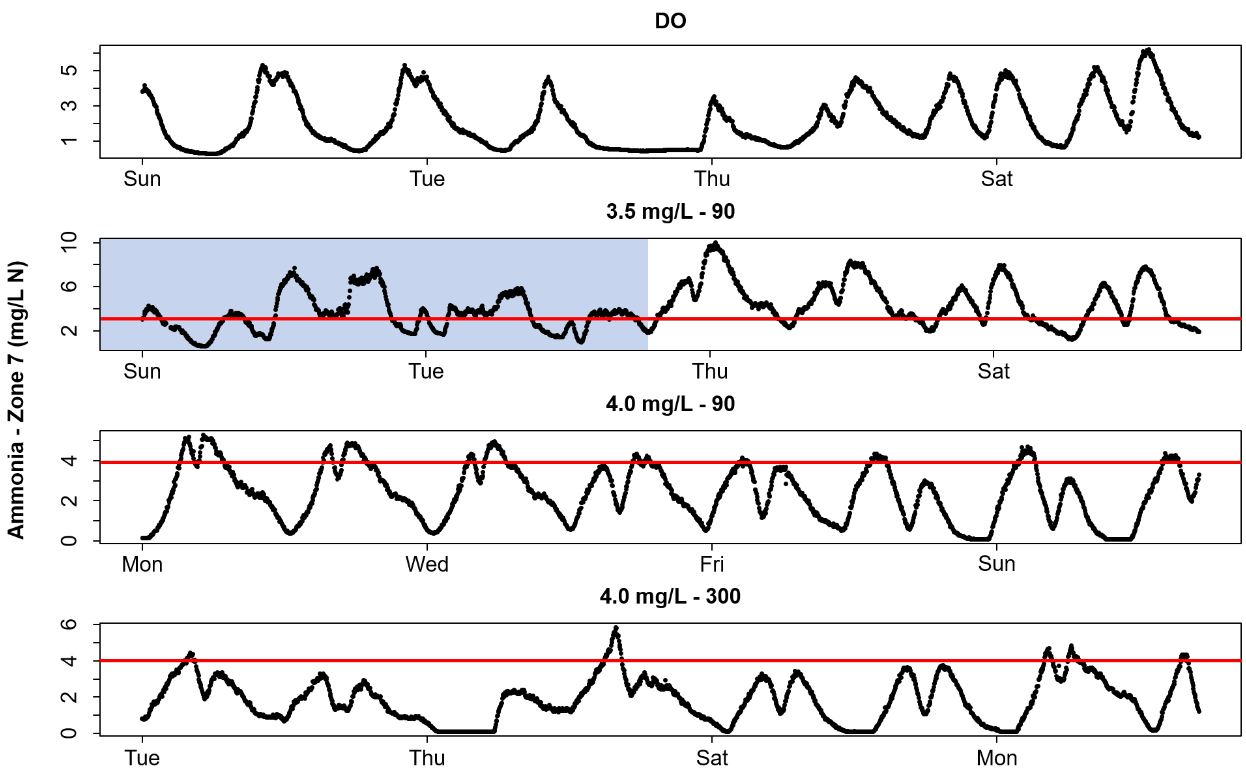

Figure 7. Time series plots of ammonia concentrations in Zone 7 for Test Runs in Phase 2. Red lines indicate ABAC ammonia setpoint. The highlighted portion of the Test Run 1 test condition deviates from the distinct patterns of the second half of the test condition as well as the other control conditions.

The smaller TSV of Test Run 2 compared to Test Run 3 (ABAC with ammonia set point 4.0 mg/L) indicates that increasing the time delay in the PID control from 90 seconds to 300 seconds led to an overall more stable process system (Figure 6). This is illustrated by the number of instances when the ammonia concentration measured in Zone 7 exceeded the 4.0 mg/L threshold (see Figure 7) (i.e., Test Run 3 resulted in more exceedances than Test Run 4).

To demonstrate that the difference in TSV or GSV is statistically significant between two Test Run conditions (*A* and *B*, to illustrate), we test the *difference* between (the TSV or GSV) condition *A* and *B* or the *ratio* between (the TSV or GSV) *A* and *B*. First, a Monte Carlo simulation is used to simulate the probability distribution of a homogenous population. In the Monte Carlo analysis, all data from two Test Run conditions are mixed and randomly split into two new sample datasets. The TSV or GSV of each of these sample datasets is calculated, and their difference and ratio are recorded. This process is repeated 10,000 times. In this way, we can obtain the distribution of the difference or ratio of GSV's and TSV's under the assumption that there is no difference between the two Test Run conditions (null hypothesis, H_0, in Table 3). If the observed difference or ratio in TSV or GSV of the actual Test Run conditions is substantially different than those calculated for the mixed datasets, then we conclude that the variability was significantly different between the two operating conditions.

The difference and ratio of the GSV for all Test Run comparisons were found to be unstable. That is, when the difference and ratio of the GSV were calculated for the mixed populations, the results were not centered about 0 or 1, respectively. Thus, GSV was not used to compare Test Run conditions. However, TSV was found to be stable and thus used. The Monte Carlo analysis showed that all Test Run conditions had TSV's that were significantly greater than the distribution of TSVs developed from randomly selected subsamples.

Figure 8 shows a comparison of the Test Run 0 (DO Control) and Test Run 2 (ABAC, ammonia set point 4.0 mg/L, time delay 90 seconds). The left figure shows the variability metric as differences in TSVs and the right figure as the ratio in TSVs. The histograms represent the distribution of the difference or ratio of TSV under the null hypothesis (no statistical difference between both Test Runs). The difference and ratio of TSV's for the full Test Run data sets (called "observed") is denoted with the red dot in both figures. Since the observed difference and ratio in TSV was much larger than would be expected under the null hypothesis, we can conclude that the variability in the data collected under Test Run 0 (DO Control) is significantly greater than the variability of the data collected under Test Run 3 (ABAC, ammonia set point 4.0 mg/L, time delay 90 seconds).

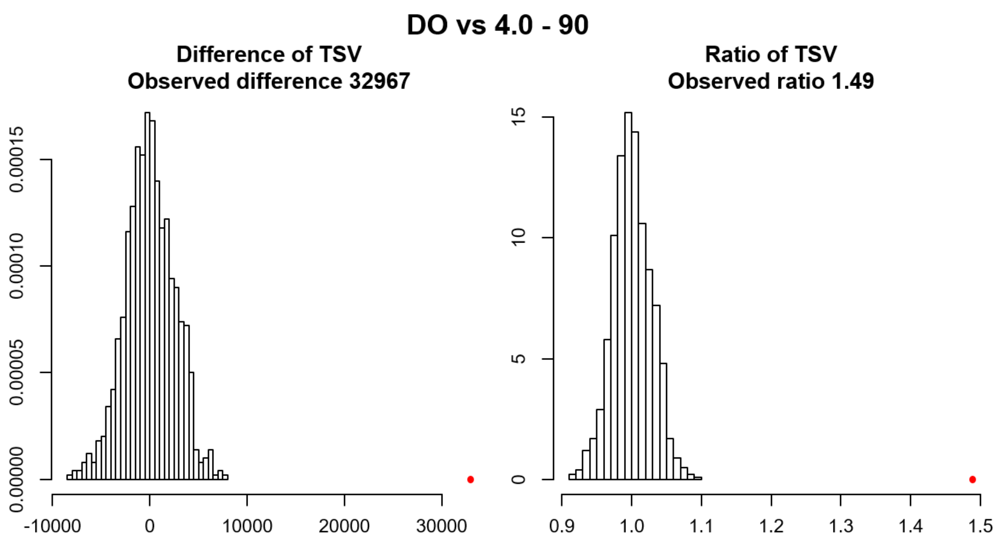

Figure 8. TSV comparison of Test Run 0 (DO Control) and Test Run 2 (ABAC, ammonia set point 4.0 mg/L, time delay 90 seconds) using the difference of TSV ($TSV_{DO} - TSV_{4.0\text{-}90}$) and the ratio of TSV ($TSV_{DO}/TSV_{4.0\text{-}90}$). The red dot (far right on each x-axis) is the observed difference or ratio in TSV whereas the vertical bars are a histogram of the observed TSV under the assumption that both datasets were mixed indiscriminately. Due to the substantial difference between the observed and mixed TSV, we conclude that the TSV for DO is significantly greater than the TSV for the 4.0 mg/L 90 s condition.

Visual inspection and simple statistical features (e.g., mean, confidence intervals) are intuitive methods of determining if a process change at a WRRF improves stability. However, for a quantitative metric that incorporates multiple variables, it is recommended that the TSV is used. By selecting a multivariate assessment method, it can be said that process instabilities or fluctuations *overall* have been reduced rather than that of a single process variable. This is important when balancing reducing excessive wear-and-tear on equipment and consistency in environmental conditions for nitrifying and denitrifying microbial communities. To compare the TSV values of two different conditions, either the difference or the ratio of the TSV can be used.

While in this work the difference and ratio showed similar results, it should be confirmed that the mixed population metrics are centered at 0 or 1, respectively. To determine if the difference between two conditions is significant, a *p*-value can be calculated by dividing the number of observations from the mixed population that exceed the observed TSV by the total number of observations from the mixed population. In this work, the *p*-value for all TSV tests were 0, but 0.01 is a common threshold for determining significance.

5.4 Ammonia Forecasting in the Aeration Basins and Soft Sensor Development

To proactively induce DO setpoint changes to eliminate the PID system response time, the ammonia concentration in Zone 7 is forecast. This response time is comprised of the time it takes for the ammonia sensor to measure a reliable concentration, for the PID loop to make setpoint adjustments when necessary, for the mechanical aeration system to make the necessary air flow changes, for the process to equilibrate to the new DO conditions, and for the biological system to adapt to the new DO conditions and reduce ammonia concentrations. Our team estimated that the response time for ammonia lasts between 5 and 20 minutes. Adding the average hydraulic retention time between the aeration basin inlet and Zone 7 where ammonia is measured, the goal of this task was to develop a model that would forecast ammonia concentrations by 50 minutes. We further defined that the required accuracy in ammonia concentration prediction should be at least within 1 mg/L to limit the required PID control response.

The MPAC model was developed by combining diurnal and linear model components to predict the ammonia concentration in Zone 7. To adapt model predictions to the continuously changing operating and environmental conditions in the aeration basins, a "moving-window training approach" was used. First, *m* days of data (observations 1 through *n*) were used to fit the diurnal-linear model. Then, observation n+1 (or the first observation of day $m+1$) is forecast by inputting current values of variables into the fitted model. The model was then retrained to include observations 2 though $n+1$, and then observation $n+2$ was forecast, and so forth. The moving window kept always the same size (*n*) and included only the most recent observations. Once all forecasts have been made, we compare the model predicted values to the actual values using root mean squared error (RMSE). We can also assess the fit of each window's model by computing the coefficient of variation, R^2.

The diurnal model (Figure 9) was developed by fitting a second order sine-cosine curve to a 24-hour (1440-minute) cycle of ammonia concentrations in 5-minute increments as follows:

$$D_t = d_0 + d_1 \sin\left(\frac{2\pi t}{1440}\right) + d_2 \cos\left(\frac{2\pi t}{1440}\right) + d_3 \sin\left(\frac{4\pi t}{1440}\right) + d_4 \cos\left(\frac{4\pi t}{1440}\right)$$

for t = 0, 5, 10, …, 1440, where d_i are model parameters that are estimated using the training window data. The differences between the fitted diurnal model ammonia predictions and the actual ammonia measurements are referred to as the ammonia residuals. These represent the portion of the ammonia concentration not captured (predicted) by the diurnal model.

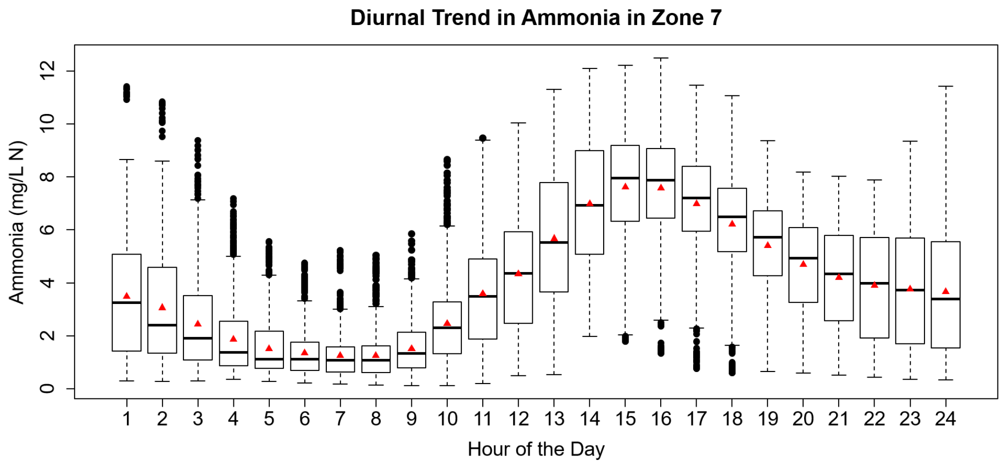

Figure 9. Boxplot of ammonia concentrations in Zone 7 binned by hour of the day. Red triangles indicate the mean of each hour. The diurnal trend is distinct with low values in the early morning and peaks occurring in early to mid-afternoon.

A multiple linear model was then fit to the residuals of the diurnal model at time *t* to forecast ammonia δ minutes ahead:

$$Y_{t+\delta} = b + m_1 Y_t + m_2 X_{1_t} + \cdots + m_{p+1} X_{p_t}$$

The predictors in the model (X_i) include the residual ammonia value at the current time, Y_t, and the current values of the process variables identified in Section 3.2.

To achieve both variable selection and parameter estimation in this model simultaneously, we use *adaptive lasso* (Zou and Qiu, 2009; Zou, 2006). Adaptive lasso is a method to estimate model parameters such that the sum of squared errors is minimized along with a penalty on the size of the regression coefficients (m_i). A preliminary estimate of the coefficients is needed, and we use ordinary least squares estimates. A required tuning parameter, λ, is often chosen through cross validation in which multiple iterations of training and testing are performed, and the results are combined to reduce variability. With increasing values of λ, more and more coefficient estimates are driven to zero (i.e., variable selection is performed by removing extraneous process variables). We used the largest value of λ such that the model prediction error is within one standard deviation of the minimum. We find that this value works very well in practice.

When selecting the size of the window used to train a model, too few observations could limit the predictions to a narrow range of operating and environmental conditions. Too many observations could unintendedly incorporate error and train to noise into the model. Various training window sizes (1 – 7 days) were tested to determine the optimum number of training observations to predict ammonia at time *t*. The model was updated (i.e., re-fit) at every timestep by removing the oldest observation and including the most recent observation to keep a constant training window size and adapt to new conditions, also known as a rolling window. The 3-day training window was found to have the lowest RMSE and was selected for further investigation (Table 4).

Table 4. Average root mean squared error (RMSE) of forecasts from Test Run 0 (DO Control) and Test Run 3 (ABAC, ammonia set point 4.0 mg-N/L, time delay 300 seconds)

Training Window Size/Dataset	*RMSE*	
	Test Run 0 (DO Control)	*Test Run 3 (ABAC)*
1 day	0.5656	0.5699
2 day	0.4638	0.4289
3 day	**0.4539**	**0.4263**
4 day	0.4647	0.4343
5 day	0.4829	0.4391
6 day	0.4891	0.4389
7 day	0.4851	0.4392

The 3-day diurnal-linear ammonia forecasting model can be further examined by forecasting horizon and by model component (i.e., diurnal or linear). In general, the average R^2 values decrease as the forecast horizon increases (Figure 10). However, the model is still fairly accurate at forecasting ammonia (R^2 0.87 at 75 minutes ahead and 0.99 at 5 minutes ahead). When R^2 values are plotted for the diurnal and linear model components in real-time, it is observed that the R^2 varies for each component with operating and environmental conditions (Figure 11). During the week, when the diurnal trend is more predictive of the ammonia concentrations, the R^2 of the diurnal model increases. During the weekend, the linear model compensates for the unusual ammonia loading patterns. This demonstrates that the dual-component model is able to achieve high levels of accuracy and flexibility.

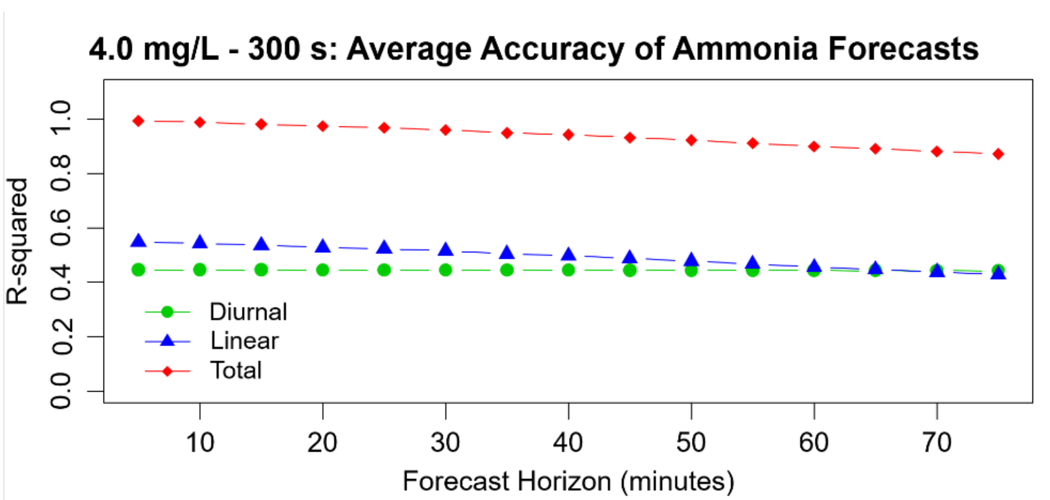

Figure 10. The proportion of the variability in actual ammonia (R^2) in Zone. 7 explained by the diurnal and linear forecasting model components. Data is fit to three days of observations

for the Test Run 3 condition, updated at 5-minute intervals, and tested on seven days of data. Red diamonds are the total R^2 for the combined model, which are the sum of the diurnal (green circle) and linear (blue triangle) R^2 values. As expected, as the forecast horizon increases, the accuracy of the linear model decreases.

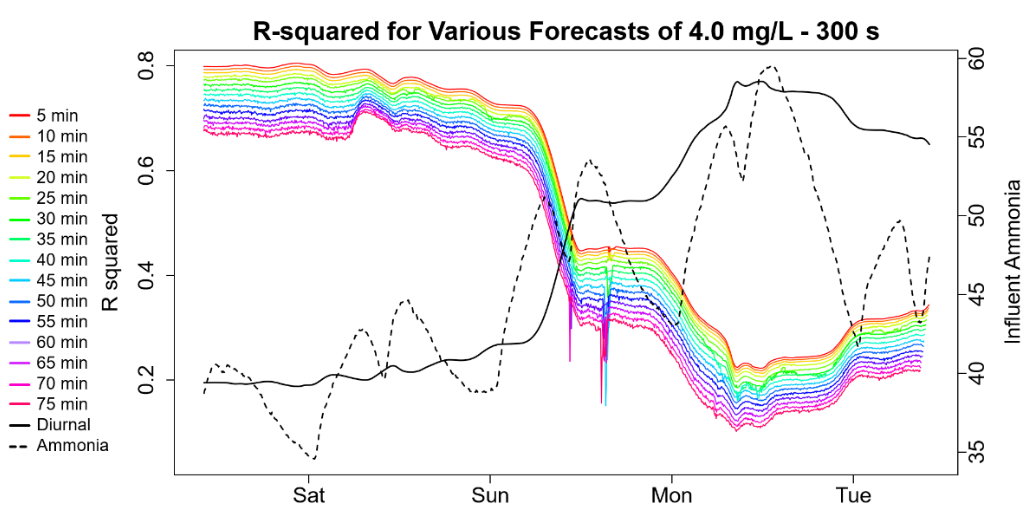

Figure 11. Instantaneous R^2 of the combination diurnal-linear forecasting model fit to three days of observations for the Test Run 3 conditions. Data updated at 5-minute intervals (left y-axis, solid lines) and the influent ammonia concentration (right y-axis, dashed line). During the transition between weekend to weekday, the dual-model structure is able to maintain a high total R^2 by shifting between the diurnal and linear model.

Figure 12(a) compares the actual and 50-minute forecast ammonia concentration in Zone 7, and Figure 12(b) plots the difference between both series as the ammonia forecasting error. Ammonia concentrations were forecast 50 minutes in advance to provide sufficient time between the forecast of ammonia concentrations and (1) the average hydraulic retention time between the inlet of the aeration basin and where ammonia is being measured in Zone 7 and (2) the response time for ammonia in Zone 7 (see discussion in Section 5.3).

Overall, the forecasts are very close to the actual ammonia concentrations in Zone 7. The model has difficulties forecasting ammonia when the actual ammonia concentration is recorded as 0 mg/L in Zone 7. However, the forecast ammonia value is able to capture the increasing and decreasing ammonia trends throughout the weekend and weekdays as well as the magnitude of daily peaks.

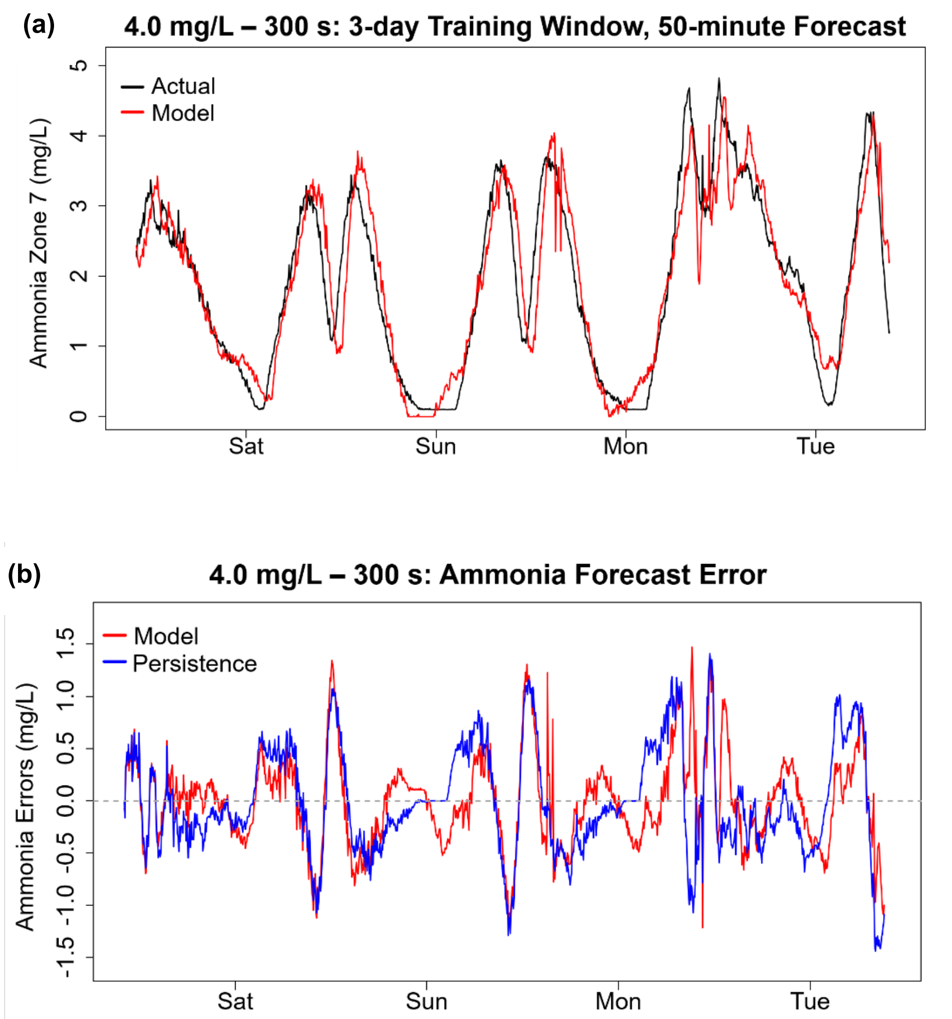

Figure 12. (a) Diurnal-linear model forecast of ammonia in Zone 7 of AB3 for the Test Run 3 (ABAC, ammonia set point 4.0 mg-N/L, time delay 300 seconds) (red) compared to the actual concentration at the forecast time (black). **(b) Forecast error and persistence forecast of ammonia in Zone 7.** The performance of the combined diurnal-linear model forecast is compared to the current measured value of ammonia, which is often termed the *persistence forecast*.

To assess the usefulness of the forecast ammonia concentration in DO and ABAC control, respectively, ammonia concentrations forecast by the diurnal-linear model were compared to the so-called *persistence* forecast. The persistence forecast assumes that the current value of ammonia measured in Zone 7 will remain the same value 50 minutes into the future. In Figure 12(b) and Table 5, the error of the diurnal-linear model forecast and the persistence forecast (essentially the actual ammonia concentration 50 min in the past) are compared. The comparison shows that for all but the 5-minute forecast horizon, the diurnal-linear model is a more accurate forecast of ammonia conditions than the use of the current ammonia measurement.

Table 5. Root mean squared error (RMSE) by duration of forecast horizon for Test Run 0 (DO Control) and Test Run 3 (ABAC, ammonia set point 4.0 mg-N/L, time delay 300 seconds) (3-day training window).

	RMSE			
	Test Run 0 (DO Control)		*Test Run 3 (ABAC)*	
Horizon / Method	*Persistence*	*Model*	*Persistence*	*Model*
5 minute	**0.1109**	0.1127	**0.1047**	0.1107
10 minute	0.1584	**0.1540**	0.1450	**0.1380**
15 minute	0.2066	**0.1931**	0.1893	**0.1699**
20 minute	0.2563	**0.2343**	0.2378	**0.2097**
25 minute	0.3065	**0.2694**	0.2863	**0.2499**
30 minute	0.3593	**0.3105**	0.3357	**0.2951**
35 minute	0.4129	**0.3574**	0.3857	**0.3386**
40 minute	0.4674	**0.4012**	0.4348	**0.3823**
45 minute	0.5207	**0.4502**	0.4831	**0.4243**
50 minute	0.5731	**0.5015**	0.5303	**0.4696**
55 minute	0.6243	**0.5518**	0.5769	**0.5134**
60 minute	0.6755	**0.5980**	0.6234	**0.5576**
65 minute	0.7273	**0.6416**	0.6690	**0.5997**
70 minute	0.7779	**0.6804**	0.7132	**0.6453**
75 minute	0.8273	**0.7238**	0.7568	**0.6842**

6 Next Steps

Using advanced data analysis and modeling, our team demonstrated that the model predicted ammonia concentrations in Zone 7 may result in a more accurate process input parameter for the PID ABAC logic than the measured concentrations given the response time delay of the system. As a next step, the City is therefore interested in using the forecasted ammonia concentrations of the diurnal-linear model within the existing ABAC control scheme to test whether this approach can further reduce process variability and unnecessary ramping up or down of the aeration blowers. Our team and the BWRRF have initiated first steps to perform a controlled test of this ammonia-forecasted ABAC in the coming weeks to demonstrate the benefits of this approach.

7 Summary and Conclusions

The need to remove nitrogen and phosphorus from wastewater to increasingly stringent regulatory limits is the most common treatment challenge of water resource recovery facilities (WRRFs) in the U.S. today. The widest and largest potential for efficient biological nutrient removal hinges on our ability to maintain tight and more accurate aeration control.

Aeration remains one of the most complex and challenging systems. PID aeration control systems are very common but come with inherent limitations. Model predictive aeration control (MPAC) is an alternative approach that uses real-time process data to continuously predict process conditions in the near future based on past conditions to allow for proactive control adjustments.

In this project, we formed an interdisciplinary team to investigate side-by-side DO and ABAC PID control systems at the BWRRF using advanced statistical methods to assess the overall process variability in response to various control logics and set points. The proposed metrics to conduct this analysis are useful to assess the entire response of many process parameters simultaneously rather than one or two parameters at a time. The analysis was instrumental in convincing operations staff of the functionality and superior process stability of ABAC compared to traditional DO control. Energy reduction was apparent in the field as the facility was able to operate largely on a single blower using ABAC.

Next, our team aimed to enhance ABAC operation further, by addressing one main limitation of the system, that is relying on past ammonia measurements as control loop inputs. We developed an open-source predictive model code using diurnal and linear model components that achieved over 90% accuracy in predicting ammonia concentrations at the ABAC control location in the aeration basins about 50 minutes into the future. The model predicted ammonia concentrations were shown to be more accurate inputs for the ABAC system than ammonia measurements that were taken 50 minutes in the past.

Through this project, the BWRRF's utility staff created a new level of understanding and comfort with statistical and data-driven process control approaches. This collaboration led to the City's willingness to conduct a side-by-side full-scale test to demonstrate the efficiency and stability of MPAC using forecast ammonia concentrations as a soft-sensor ABAC input and compare results to the current ABAC control.

Specific to the questions and goals that our team set out to answer as part of this IWS Challenge, we were able to draw the following conclusions:

1. After preliminary tuning, ABAC was a more stable aeration control mode for BWRRF's activated sludge system compared to the traditional DO control. Additional PID loop tuning may further improve the process stability.
2. ABAC operation improved the energy efficiency by reducing blower system demand. This allowed the facility to operate mostly with a single blower. Additional energy savings are possible when
    a. All three aeration basins are operated consistently in ABAC mode.
    b. ABAC tuning is further optimized.

   c. The ammonia concentration setpoint is further increased (and the aeration safety factor decreased). Throughout all Test Runs in this project, the final effluent ammonia concentration was essentially non-detect. At this time, the City does not feel comfortable to decrease aeration further but would consider doing so with an additional ammonia sensor located in the aeration basin effluent channel to better monitor the ammonia profile through the secondary process.

3. The variation was statistically significantly reduced under ABAC operation, primarily influenced by the reduction in air supply to the aeration basin as was the reduction in DO carry over into the second anoxic zone. This improved denitrification and reduced the need for external carbon addition.
4. The difference and ratio of the *Total sample variation* (TSV) of two operational datasets are proposed as useful metrics to characterize and compare total process variability between different (aeration) control conditions.
   a. These two metrics allow an assessment of the variability of the entire aeration basin system simultaneously, rather than one parameter at a time.
5. The ammonia concentrations in Zone 7 of the aeration basins can be forecast with sufficient accuracy to become a useful soft-sensor within the ABAC control logic.
   a. The forecast accuracy is about 90% for a 50-minute forecast window, keeping ammonia prediction errors typically to less than 0.5 mg/L.
   b. The facility is interested in testing the forecast ammonia concentrations as a soft-sensor input to the ABAC PID controller. By preemptively triggering a small-change DO setpoint in BWRRF's existing ABAC configuration, differences between the actual and setpoint ammonia and DO are projected to be smaller. The blower can then ramp up slowly as opposed to increasing speed over a shorter period of time.
6. The systematic evaluation of the DO and ABAC aeration control gained the trust of BWRRF operations staff for long-term use of the ABAC system.

References

Kazor, K., Holloway, R.W., Cath, T.Y., Hering, A.S., 2016. Comparison of linear and nonlinear dimension reduction techniques for automated process monitoring of a decentralized wastewater treatment facility. Stochastic Environmental Research and Risk Assessment 30, 1527–1544. https://doi.org/10.1007/s00477-016-1246-2

Odom, G.J., Newhart, K.B., Cath, T.Y., Hering, A.S., 2018. Multistate multivariate statistical process control. Applied Stochastic Models in Business and Industry 34, 880–892. https://doi.org/10.1002/asmb.2333

Zou, C.L., Qiu, P.H., 2009. Multivariate Statistical Process Control Using LASSO. J Am Stat Assoc 104, 1586–1596. https://doi.org/10.1198/jasa.2009.tm08128

Zou, H., 2006. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association 101, 1418–1429. https://doi.org/10.1198/016214506000000735