

Hybrid statistical-machine learning ammonia forecasting in continuous activated sludge treatment for improved process control

Kathryn B. Newhart^a, Christopher A. Marks^b, Tanja Rauch-Williams^c, Tzahi Y. Cath^a,
Amanda S. Hering^{d,*}

^a Civil and Environmental Engineering, Colorado School of Mines, Golden, Colorado, USA

^b City of Boulder Water Resource Recovery Facility, Boulder, Colorado, USA

^c Carollo Engineering, Inc., Broomfield, CO, USA

^d Statistical Science, Baylor University, Waco, Texas, USA



ARTICLE INFO

Keywords:

Ammonia-based aeration control

Artificial neural network

Cascade control

Stability assessment

ABSTRACT

In this work, a statistical stability metric and novel hybrid statistical-machine learning ammonia forecasting model are developed to improve the accuracy and precision of municipal wastewater treatment. Aeration for biological nutrient removal is typically the largest energy expense for municipal wastewater treatment plants (WWTP). Ammonia-based aeration control (ABAC) is one approach designed to minimize excessive aeration by adjusting air blower output from online ammonia measurements rather than from a dissolved oxygen (DO) sensor, which is the conventional aeration control approach. We propose a quantitative stability metric, *Total Sample Variance*, to compare system-wide variability of competing aeration control strategies. Using this metric, the performance of traditional DO and ABAC control strategies with varying setpoints and control parameters were compared in a medium-sized WWTP, and the most stable strategy was identified and implemented at the facility. To further improve ABAC performance, ammonia forecasting models were constructed using both statistical and machine learning to improve the accuracy of the aeration control system. Diurnal, diurnal-linear, artificial neural network (ANN), and hybrid diurnal-linear-ANN forecasting models were trained on real-time plant-wide process data. The diurnal-linear and diurnal-linear-ANN forecasts were found to most accurately forecast ammonia; improving upon the existing ammonia measurement by up to 32% and 46%, respectively, whereas the ANN model forecast was only able to improve by up to 8%. This work demonstrates the ease and flexibility of integrating statistics and machine learning methods for developing new treatment models in conventional WWTP for features in full-scale conventional activated sludge systems.

1. Introduction

One of the greatest threats to surface waterways in the United States is nutrient pollution [1]. The removal of nitrogen, specifically ammonia, from domestic wastewaters at municipal wastewater treatment plants (WWTP) is important to limiting nutrient loading into the environment and protecting aquatic life from toxic effects [2]. As states' regulatory agencies implement increasingly stringent nutrient discharge limits on WWTP, the cost of wastewater treatment will substantially increase with the implementation of new treatment technol-

ogies to achieve higher nitrogen removal [3]. The most widely-used strategy for ammonia removal is *nitrification*, which introduces dissolved oxygen (DO) via industrial air blowers to facilitate microbial uptake and conversion of organic nitrogen and ammonia in wastewater to nitrate. Consequently, aeration is one of the largest operating expenses for most WWTP [4].

In conventional biological treatment, air blower output is increased or decreased proportional to the difference between the measured concentration of DO and a target DO concentration using a proportional-integral-derivative (PID) controller. The DO concentration set-

* Corresponding author.

E-mail addresses: tcath@mines.edu (T.Y. Cath), mandy_hering@baylor.edu (A.S. Hering).

<https://doi.org/10.1016/j.jwpe.2020.101389>

Received 3 April 2020; Received in revised form 18 May 2020; Accepted 25 May 2020

2214-7144/ © 2020 Elsevier Ltd. All rights reserved.

point for nitrification is usually between 0.5 and 3 mg/L, but is contingent on site-specific factors such as the maximum ammonia loading rate [5]. While DO control can achieve complete ammonia removal, the strategy has multiple drawbacks; most notably, the DO setpoint is a poor surrogate of the immediate process demand, and thus, air is often supplied in excess of what is required for ammonia removal. To save energy and continuously achieve complete ammonia removal at WWTP, a new aeration paradigm is needed.

Ammonia-based aeration control (ABAC) is a control strategy that adjusts air flow to meet an ammonia concentration setpoint [6]. ABAC is typically implemented as feedforward, feedback, or hybrid systems. Feedforward ABAC utilizes microbial kinetic process models to calculate the aeration demand given a variety of operating parameters such as solids retention time and influent ammonia concentration [7]. This approach can reduce aeration demand but is costly to implement; requiring extensive sampling and specialized knowledge for accurate model calibration and implementation. Feedback ABAC does not require kinetic process models, but instead adjusts aeration based on the measured ammonia concentration within the process using multiple nested PID controllers in a *cascade control* configuration. In the case of ABAC cascade control, an ammonia sensor and setpoint define the outer/master control loop while inner/slave control loops define control variables such as DO, air blower flow, and air blower speed. In feedback controller design, actuator and sensor dynamics within a process and wear-and-tear on the equipment that controls the process are frequently ignored [8]. For aeration at WWTP, this can lead to a large delay between a change in demand and the aeration provided. When the aeration required to treat ammonia is not immediately met, the difference between the measured ammonia value and the setpoint becomes large, and the air blower speed is increased proportional to the difference. This can result in large and short-lived air flow peaks, causing blowers to ramp up and down excessively, thereby contributing to peak energy demand. As the daily peak energy demand is used to calculate a WWTP's electricity bill, a reduction of these aeration peaks will ultimately reduce energy and operating costs. Hybrid control approaches combine elements of feedforward and feedback control; however, these approaches have two universal limitations:

- Ammonia sensors are generally prone to drifting, noisy measurements, and require frequent field calibration.
- PID aeration control is inherently based on past measured process values due to the time delay between a measurement and a corrective control action. The resulting measurement delay is unable to accurately anticipate current and future process conditions due to the highly dynamic nature of biological wastewater treatment processes.

To address both drawbacks, this study (1) proposes a stability metric to measure abnormally varied conditions and to compare control strategies. To specifically address the second drawback, this study also (2) develops a forecasting model for ammonia that could be incorporated into a feedback ABAC system. For the latter objective, statistical and machine learning models were used to forecast ammonia concentrations, which would replace real-time ammonia measurements. The advantage of forecasting using statistical and machine learning models is that no additional sampling, microbiological analysis, or proprietary software is required to build a process model. Additionally, a forecast

can overcome the delay of feedback control and easily replace the current measured value of ammonia in the supervisory control and data acquisition (SCADA) system of a WWTP.

To date, there are only a small number of WWTP worldwide that have implemented full-scale data-driven control systems. Forecasting models have been developed for predicting influent quality [9–14], but few models exist that are designed to adjust downstream operating conditions for more efficient treatment. Additionally, much of the modeling performed in the literature is limited to water quality inputs that require hours to measure and cannot be used for real-time control [14–16]. In contrast, our work directly models the actual biological treatment response to influent water quality changes and can be done in real-time. Furthermore, model errors reported in the existing literature are too large to improve real-time control (20–40%) [17,18,14]. Specific to ammonia and ABAC modeling, Ekster et al. [19] developed a proprietary model-driven ABAC system that relies on specific mechanistic (air flow), mass transfer (DO), and mass loading models to calculate ammonia removal, but requires site-specific modeling and calibration. In contrast, the tools developed in this work are designed for full-scale implementation that do not need such expertise and specifications. Vezzaro et al. [20] developed a phenomenological model to forecast influent ammonia based on a diurnal model and historical flow and concentration measurements. Our approach further extends the diurnal model approach by (1) including more than two harmonic pairs, (2) including a linear model of operational predictor variables, and (3) pairing this diurnal-linear forecast with a machine learning method to reduce error at longer forecast horizons. Because we forecast ammonia concentrations *within* the WWTP, incorporating both changing influent quality and current operational conditions rather than solely forecasting influent water quality, this represents an improvement upon previous water quality forecasting approaches.

Another unique feature of this work is the “small” dataset used. Most machine learning approaches for water quality prediction have required 1–10 years of historical data [14,18]. Given that most wastewater facilities make domestic and industrial additions to their collection system and/or upgrades to existing infrastructure, historical data often do not accurately represent current conditions, and some variables may be sparsely sampled (e.g., intermittent sampling during optimization efforts). Other machine learning approaches use pilot scale data, which were collected under steady state conditions and do not typically account for the variability experienced at full-scale (i.e., all possible environmental and operational conditions) [18]. Our work also shows how machine and statistical learning can be applied to WWTPs using smaller windows of time of 3–7 days that allow the models to adapt quickly to infrastructure and operational changes. Additionally, previous machine learning modeling efforts have focused on amending artificial neural networks (ANN) with a mechanistic approach (e.g., the Activated Sludge Model, flow and mass balances) [21,22], but the combination of machine learning and the mechanistic models fails to substantially improve prediction accuracy.

Statistical-machine learning hybrid models have been shown to capture the real-time dynamics of an individual, full-scale WWTP better than a standard mechanistic model [23]. Zhang [24] presented a hybrid statistical (autoregressive integrated moving average or ARIMA)-ANN model for timeseries data, and similar models utilizing ARIMA-ANN variations have also been proposed for large sets of lab or flow data (Ömer [21,25]). Lotfi et al. [26] similarly proposed an

ARIMA-ANN variation that did not include operational parameters in their prediction (influent and effluent biochemical oxygen demand, chemical oxygen demand, total dissolved solids, and total suspended solids). Rather, Lotfi et al. [26] utilized the values of water quality variables not actively used in an individual model construction to predict the response, which is not ideal for real-time forecasting. By ignoring operational conditions in the forecasting models, the applications are limited to offline analysis and manual operational changes. The goal of this work is to incorporate operational and environmental conditions to adjust the forecast in real-time, using short training windows to quickly adapt to changes to the system, so that this approach could be easily incorporated into conventional WWTP's existing control strategies.

The contributions of this work are (1) to quantify the stability of DO control and ABAC strategies for the Boulder Water Resource Recovery Facility (BWRRF), a medium-sized WWTP located in Boulder, Colorado (USA), and (2) to develop data-driven methods of forecasting ammonia at BWRRF for the purpose of improving accuracy, reducing energy consumption, and reducing mechanical wear of aeration systems. DO and ABAC feedback aeration strategies were tested and compared at the BWRRF to identify the most stable aeration control scheme, which requires a balance of reliably removing ammonia to a desired concentration while minimizing operational variability and energy use. The manuscript is organized as follows: a description of the BWRRF (Section 2.1); the method and approach for quantifying variation (Section 2.2); summary of statistical and machine learning methods used to build the ammonia forecasting models (Section 2.3); a comparison of the stability metric to actual control strategy performance (Section 3.1); and an assessment of the performance of the forecasting models (Section 3.2). Additional guidance on full-scale implication of this work, including code, is included in the Supplementary Information.

2. Materials and methods

2.1. The Boulder Water Resource Recovery Facility

The BWRRF is a 25 million gallon per day (MGD) municipal WWTP, currently operating a Four-Stage Bardenpho biological nutrient removal (BNR) process at an average of 12 MGD. Given the high altitude of the facility (5115 ft above mean sea level) and low daily ammonia discharge limits (1.9 mg/L NH_4 as N), oxygen transfer efficiency is relatively low and results in high aeration demand [27]. DO control is currently implemented using a three-layer PID cascade control strategy that frequently over-aerates to ensure treatment during peak ammonia loading conditions (on average, 1–2 hours per day). This approach is energy-inefficient and causes high DO concentrations in downstream and recycle anoxic zones, which reduces total nitrogen removal (i.e.,

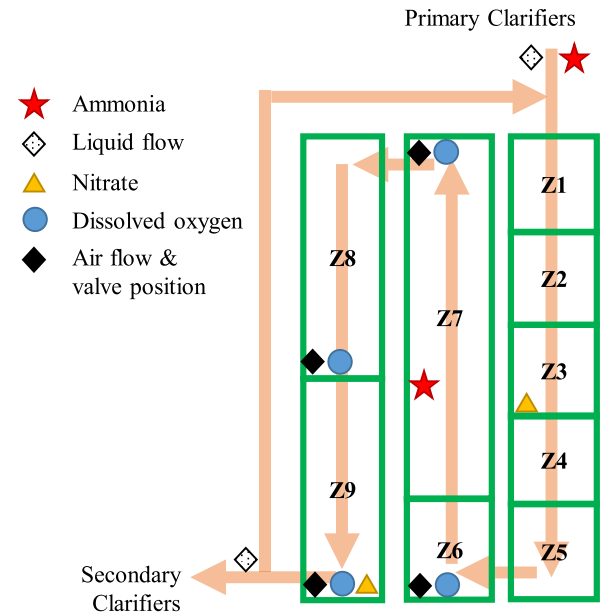


Fig. 2. Flow, zone, and sensor diagram of one of the three activated sludge aeration basins at BWRRF. Arrows indicate the direction of flow through the basin starting in Zone 1 (Z1) and ending in Zone 9 (Z9) where the majority of flow is recirculated to Zone 1. Zones 4–8 (Z4–Z8) are aerated, and Zones 6–8 (Z6–Z8) have individual DO setpoints for each zone. Triangles indicate nitrate sensors, circles indicate DO sensors used in DO control, and the star in Z7 indicates the ammonia sensor used in ABAC control.

inhibits denitrification). Consequently, aeration accounts for 35–50% of BWRRF's energy consumption. There are multiple aeration control methods programmed into BWRRF's SCADA system, all of which rely on feedback cascade control: airflow, DO, and ABAC. Airflow setpoint control adjusts valve position at the inlet of the aeration grid to produce a desired volumetric flow of air, regardless of oxygen demand; DO concentration setpoints adjust the airflow setpoint to achieve a DO concentration; and ABAC setpoints adjust DO concentration setpoints to achieve an ammonia concentration setpoint (Fig. 1). The process variables included in the control logic (y), stability analysis, and forecasting model are located in one of BWRRF's three aeration basins (Fig. 2).

The DO concentrations exiting Zones 6, 7, 8, and 9 of the aeration basins are continuously monitored using Endress Hauser COS61D optical DO sensors (Reinach, Switzerland). In the aerated zones (Zones 4–8), air flowrate is monitored and controlled by air valve position and air blower speed, all of which are recorded in the SCADA system. YSI AmmoLyt® Plus 700 ion-selective ammonia sensors (Yellow Springs,

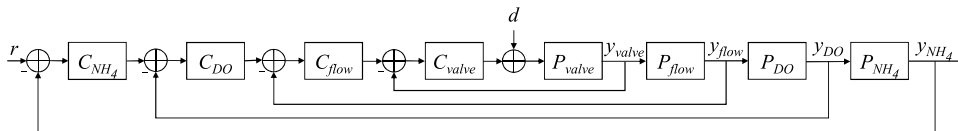


Fig. 1. Cascade control logic for ABAC at BWRRF. The reference signal r is the ammonia concentration setpoint. The difference between r and the actual ammonia measurement y_{NH_4} is input to the ammonia PID controller C_{NH_4} , which adjusts the DO concentration setpoint.

The difference between the DO setpoint and the DO measurement, y_{DO} , is input to the DO PID controller C_{DO} , which adjusts the air flow setpoint. The same logic flows for the air flow controller, C_{flow} , and air valve controller, C_{valve} , which ultimately provide a certain volume of air to the aeration basins. A disturbance d , such as a change in water quality, will impact the processes P within the basin as measured by y . The difference between the setpoints and y values will force the PID controllers to continuously provide adjustments within their loop until all setpoints are met.

Table 1

Control strategies tested vary by which is the “control” variable (i.e., outer/master loop in cascade control), the master setpoint for the control variable, and the delay term for the master PID loop.

	Control Variable	Setpoint (mg/L)	PID delay
Control	DO	2.5	90
Test Condition 1	Ammonia	3.5	90
Test Condition 2	Ammonia	4.0	90
Test Condition 3	Ammonia	4.0	300

OH) are located in the aeration basin influent channel (not shown in Fig. 2) and in Zone 7. YSI ion-selective nitrate/nitrite sensors are located in Zones 3 and 9. Aeration basin influent flow rates, wastewater temperature, and pH of the plant influent are also monitored. Online sensors are regularly maintained and calibrated by BWRRF operations staff, and sensor measurements are periodically compared to laboratory results.

Data at BWRRF are stored and managed in the GE Proficy® system. Data were exported in 5-minute intervals from Proficy 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 power loss) were replaced with NA. For real time analysis, process data were exported from Proficy using a Visual Basic script, processed using R, and forecasts were exported to Proficy using a second Visual Basic script for use within BWRRF’s SCADA system (see Supplementary Information for Visual Basic and R scripts).

The control strategies analyzed in this study are summarized in Table 1. A conventional DO aeration control method with DO setpoints for Zones 4–6, 7, and 8 of 2.5, 2.0, and 1.0 mg/L, respectively, is considered the experimental control. For ABAC, if the measured concentration of ammonia exceeds the setpoint, the DO setpoint is incrementally increased up to 2.3, 2.3, and 1.0 mg/L for Zones 4–6, 7, and 8, respectively. If the measured concentration of ammonia is below the setpoint, the DO setpoint is incrementally decreased to no less than 1.0, 1.0, and 0.1 mg/L for the respective zones. Three ABAC strategies were tested, including 3.5 mg/L and 4.0 mg/L ammonia setpoints (at the middle of the aeration basin, Zone 7 in Fig. 2) and 90 s and 300 s time delays between PID control actions. BWRRF operators initially increased the time between control actions to allow time for the DO and ammonia sensors to stabilize. Artificially low or high sensor measurements can be caused by noise or water quality changes that occur faster than the sensors can reliably measure. Incorrect sensor measurements cause a PID control loop to repeatedly change process conditions, which further changes water quality and compounds the problem of sensor instability.

2.2. Stability assessment

The goal of the stability assessment is to provide WWTP engineers and operators with a quantitative metric to decide between two control strategies on the basis of variability. To measure the stability of each operating strategy holistically, we investigated Total Sample Variance (TSV). TSV is the trace of the variance-covariance matrix of p variables ($\text{trace}(\Sigma) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2$). A large TSV is indicative of increased

Table 2

Number of observations used for training and testing each model using training window sizes from 1 to 6 days, limited to the total number of observations for each control strategy.

Window Size	1 Days	2 Days	3 Days	4 Days	5 Days	6 Days
Training	288	576	864	1,152	1,440	1,728
Testing	1,850	1,562	1,274	986	698	410

variability within the process and therefore less stability. Consequently, the most stable operating condition has the relatively lowest TSV.

To demonstrate if one operating condition is more stable than another, the difference and ratio of TSV between two conditions (i.e., Monte Carlo simulation of TSV metrics). First, data from the control and test condition are combined into a single dataset, and then randomly split into two datasets (i, j). This random split is repeated 1000 times. For each random reassignment, estimates of TSV are computed for each split dataset (TSV_i and TSV_j). Then, a distribution of the ratio or difference between TSV_i and TSV_j under the null hypothesis can be obtained, and the following sets of hypotheses can be tested:

$$H_0: D_{i,j} \leq 0 \text{ versus } H_1: D_{i,j} > 0 \text{ where } D_{i,j} = TSV_i - TSV_j \quad (1)$$

$$H_0: R_{i,j} \leq 1 \text{ versus } H_1: R_{i,j} > 1 \text{ where } R_{i,j} = TSV_i/TSV_j \quad (2)$$

To test the null hypothesis, the number of observations from the random reassignments whose calculated $D_{i,j}$ or $R_{i,j}$ are greater than the observed TSV metric is used to approximate the p -value:

$$p = \frac{1}{1000} \sum_{k=1}^{1000} I(R_k > R_{i,j}) \quad (3)$$

where R_k is the ratio of TSVs from the randomly selected subset of observations from individual control strategies; $R_{i,j}$ is the ratio of TSVs from the observed data; and $I(\bullet)$ is a function whose value is 1 if its argument is true and 0 otherwise. Thus, Eq. (3) computes the proportion of TSV ratios from the random reassignments that exceed the observed TSV ratio. A similar computation can be constructed for the difference in TSVs. A sample conclusion would be as follows: if $TSV_i = TSV_{DO}$, $TSV_j = TSV_{test}$, and $p < 0.01$, then the test control scheme is said to be significantly less variable and therefore more stable than the conventional DO control.

2.3. Ammonia forecast

To train and test an ammonia forecasting model, observations are aligned to simulate real-time prediction using past process data and current ammonia data. For an original dataset with n observations, ammonia measurement (Y), process variables ($X_1, X_2, \dots, X_p$) at the current time (t), and forecast horizon (f), an $(n - f) \times (p + 1)$ matrix for training and testing is created. Given that the shortest dataset provided by BWRRF was 7.4 days, all datasets were truncated to only include 7.4 days of online process data. Models are built using 1–6 days of process data (in intervals of 1 day or 288 observations) and are tested on the remaining observations in sequence, updating the model for each new observation while maintaining a constant training window size (Table 2). Forecast horizons of 5–75 min (1–15 observations) were

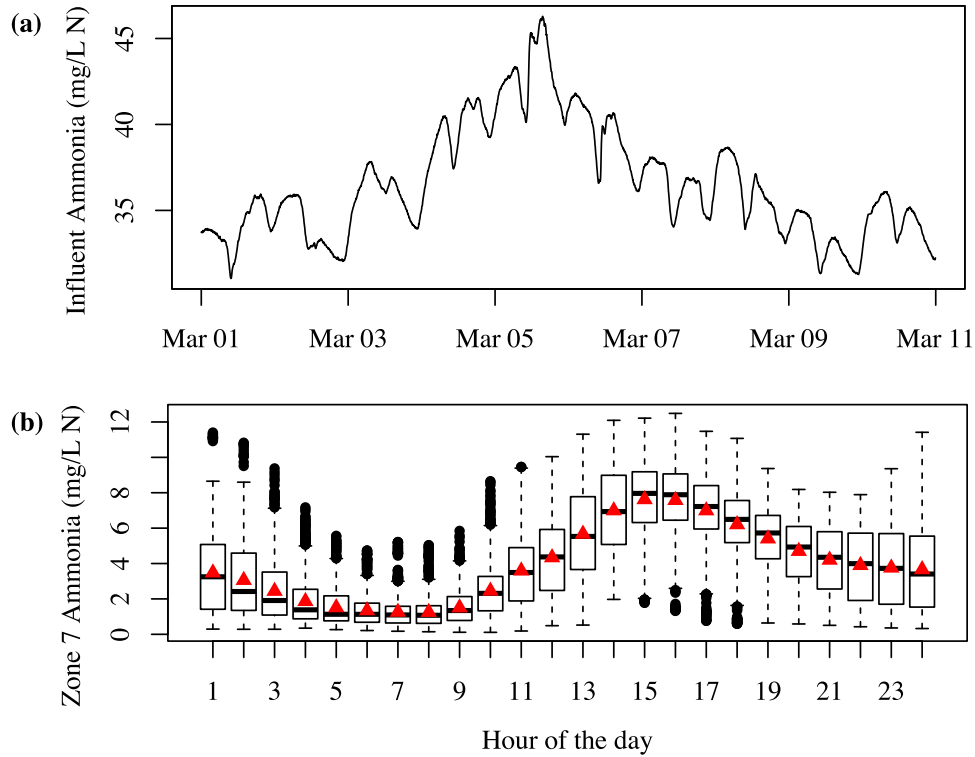


Fig. 3. (a) Timeseries plot of Zone 7 ammonia at BWRRF from 2019-03-01 to 2019-03-10 and (b) boxplot of the same influent ammonia binned by hour of the day with the hourly mean indicated by red triangles. The diurnal trend illustrated in (b) can be modeled, although it cannot explain all variation in ammonia.

evaluated in increments of 5 min (1 observation). Timeseries plots of all monitored process variables for the following analysis can be found in the Supplementary Information (Figures S1–S4). All variables were scaled to zero mean and unit variance using the training dataset for each iteration.

To quantify model fit and forecast accuracy, the coefficient of determination (R^2) and root mean squared error (RMSE), respectively, are used. R^2 is a measure of how well the model fits the *training* data. An $R^2 = 1$ indicates that the model predicts the values used to train the model exactly, while an $R^2 = 0$ indicates that the model does not explain any linear variation in the training data. Model *testing* performance is evaluated using RMSE, which is a measure of average squared error between model forecasts and the observed values. The better the model's forecasts, the lower its RMSE. The existing feedback control strategy is effectively using the current ammonia value as the forecast and is named *persistence*. A comparison of the forecasting models to persistence using RMSEs is calculated as:

$$(RMSE_{persistence} - RMSE_{model})/RMSE_{persistence} \cdot 100\% \quad (4)$$

2.3.1. Diurnal-linear model

Variation of ammonia in BNR is affected by environmental and operational conditions. The environmental component is determined by changes in the influent water quality, which follows strong daily trends (Fig. 3). To model this time-dependent component, a diurnal model is fit to forecasted ammonia. The predictors are various degrees ($d = 1, 2, \dots, 200$) of sine and cosine pairs ($\sin(d \cdot t) + \cos(d \cdot t)$), where t is the

minute of the day expressed in radians from 0 to 2π .

To model the operational component of the ammonia forecast, a multiple linear regression model is fit to ammonia, such that the combined fitted diurnal-linear model is:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 \sin(t) + \hat{\beta}_2 \cos(t) + \dots + \hat{\beta}_{2d-1} \sin(d \cdot t) + \hat{\beta}_{2d} \cos(d \cdot t) + \hat{\beta}_{2d+1} y_{t-f} + \hat{\beta}_{2d+2} x_{1t-f} + \dots + \hat{\beta}_{2d+p+1} x_{pt-f} \quad (5)$$

where p is the number of process variables; x_{pt-f} is the p^{th} process variable lagged back some forecast horizon f ; y_{t-f} is the ammonia at time $t - f$; and $\hat{\beta}_i$ are estimated linear model coefficients for $i = 0, 1, \dots, (2d + p + 1)$.

To eliminate noise and reduce variability, the final diurnal-linear model only includes the most pertinent process variables and the optimum number of degrees of sine/cosine pairs by using a modified model fit method that simultaneously achieves variable selection and coefficient estimation. The smallest d with the relatively largest R^2 is selected as the optimum d from the 1–200 d 's tested. The most pertinent process variables are selected using a multivariate linear model fitting method called adaptive lasso [28].

Multiple linear regression models have historically been fit using ordinary least squares (OLS) where model parameters are estimated such that the sum of squares of the residuals are minimized. The bias of this approach is low, but it is heavily reliant on the quality of the training dataset and can have high variance. A small change in the training data can substantially change OLS parameter estimates [29]. Thus, an alternative fitting method is selected to reduce prediction

variability and improve model interpretability.

Lasso (Least Absolute Shrinkage and Selection Operator) is one such regularization method that can decrease model complexity while maintaining or improving accuracy [30]. In lasso, the sum of squares of the residuals are minimized along with a loss function that additionally penalizes the size of model parameter estimates ($\hat{\beta}_j$), shrinking insignificant predictors' coefficients to zero. The 'shrinkage' term responsible for driving the coefficients to zero is controlled by λ . When $\lambda = 0$, lasso returns the same linear model coefficients as OLS. The larger λ is, the fewer predictors will be included in the model, which decreases variance in the prediction. However, in some cases lasso estimates are known to be biased and inconsistent. A solution proposed by Zou [28] uses adaptive weights (w_j) to overcome the limitation of traditional lasso, called adaptive lasso. The adaptive lasso loss function is included in the objective function, which is minimized to identify the best parameter selection and estimates:

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^n \left(y_i - \sum_{j=1}^{2d+p+1} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{2d+p+1} w_j |\beta_j|, \quad (6)$$

where λ is a regularization parameter, and w_j are coefficient-specific weights.

Both the diurnal model and diurnal-linear model were trained and tested using the glmnet package in R. Initially, 10-fold cross-validation is used to fit a generalized linear model using ridge regression to obtain an initial estimate of β . This method shrinks the linear model parameters towards zero, but not exactly zero. Then, the ridge regression regularization parameter is selected such that the model has the smallest cross-validation error. The parameter estimates from this model ($\tilde{\beta}_j$) are used to calculate the weights of the adaptive lasso penalty term, so we set $w_j = 1/|\tilde{\beta}_j|$. This allows different predictors to be penalized

differently based on their initial estimates and drives coefficients that are already small to zero even faster. The final adaptive lasso model is selected with λ within 1 standard error of the minimum cross-validation error.

The advantages of using adaptive lasso are two-fold: variance reduction and variable selection. The goal of the model is to forecast ammonia concentrations under environmental and operating conditions that may not have been included in the training data. Reducing the variance in model predictions prevents overfitting the model. Variable selection is an additionally appealing trait of adaptive lasso that (1) gives engineers and operators a better understanding of the most important operating process variables within a system and (2) eliminates unnecessary or noisy predictor variables. An investigation of the variables included in each forecasting model as a function of training window and forecast horizon provides further insight not available using traditional machine learning techniques, like ANN, described below.

2.3.2. Neural networks

In the event that an interpretable, linear process model cannot forecast ammonia with the desired accuracy, machine learning techniques may be a more effective substitute or addition. Conventional mechanistic models use complex formulations that are connected in mathematically simple ways. For example, a nutrient removal model based on microbial kinetics can be summed for a WWTP-wide mass balance. Machine learning methods, including ANN, use the converse approach: simple mathematical expressions are connected by complex, nonlinear functions. The detriment and advantage of ANN are that no prior knowledge is needed to predict a process output, but correspondingly, no information is produced about the underlying mechanisms defining the input-output relationship [31–33].

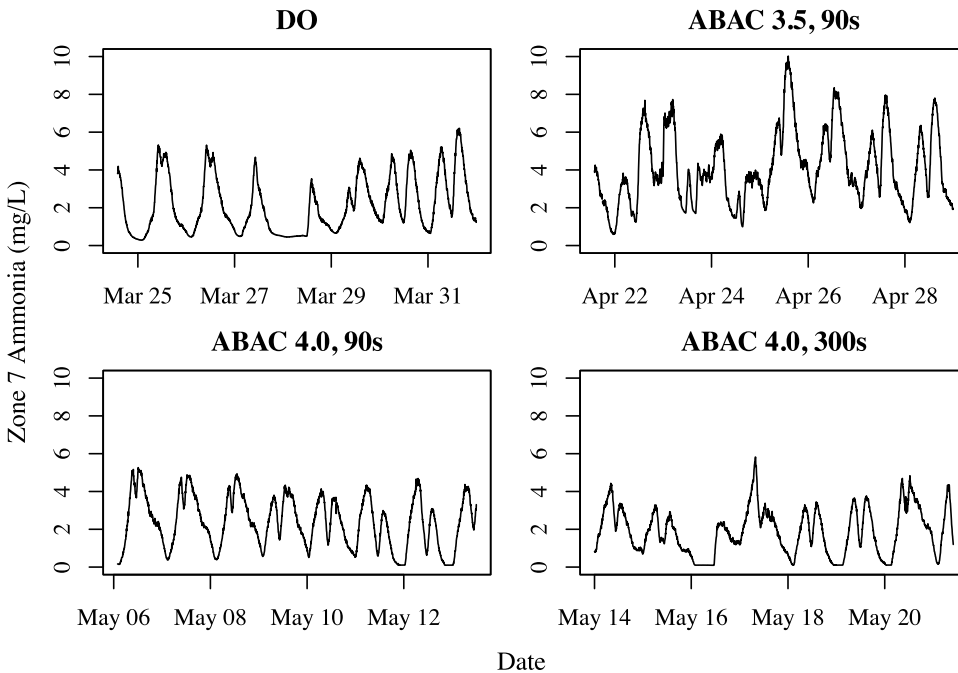


Fig. 4. Timeseries plot of Zone 7 ammonia at the BWRRF for all control strategies tested. While DO and both ABAC 4.0 mg/L conditions were able to reasonably limit Zone 7 ammonia to 4.0 mg/L, the ABAC 3.5 mg/L test condition experienced multiple exceedances greater than 4.0 mg/L and was unable to reduce variability. The variability of ammonia during the ABAC 3.5 mg/L test condition suggests a faulty sensor or abnormal environmental conditions. ABAC 4.0 at 90 s appears to control ammonia with more consistency (when only ammonia is considered), but ABAC 4.0 at 300 s reduced variation the most system-wide (when all variables are considered).

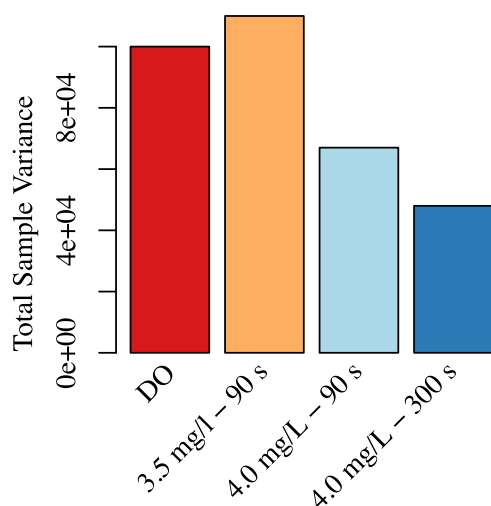


Fig. 5. Observed TSV for each control configuration. The least variable and most stable operating condition appears to be ABAC at 4.0 mg/L ammonia setpoint and 300 s PID update frequency. The most variable and least stable operating condition appears to be ABAC at 3.5 mg/L ammonia setpoint and 90 s PID update frequency. Visual inspection of the timeseries plots found in the Supplemental Information (Figure S2) and Fig. 4 by operators confirms this conclusion.

The most common ANN configuration is a feedforward ANN, which uses an input layer representative of the process variables, a single “hidden” layer of analyses, and an output layer containing the response variable. The hidden layer is comprised of nodes, and at each node, a sum of weighted values from the input layer is used by the node’s activation function to produce an output. Weights for each input variable-node connection are determined using backpropagation algorithms to minimize the error between the predicted output and the actual output. The feedforward ANN is one-directional and not computationally intensive. The lack of model “memory” means that the feedforward ANN neglects causality and order of observations. To consider temporally ordered observations, recurrent neural networks (RNN) can be used; however, while RNN were initially explored for this work, the computation time was not amenable to real-time prediction, so RNNs were excluded from this study.

Two ANN models are constructed to forecast ammonia: (1) using all lagged process variables as inputs (referred to as the “ANN model”) and (2) using all lagged process variables and the diurnal-linear forecast as inputs (referred to as the “diurnal-linear-ANN model” or “hybrid model”). The ANN training and forecasts are performed using the Python-based TensorFlow package in R. The first hidden layer includes 10 or 11 neurons (2/3 the number of inputs for the ANN and diurnal-linear-ANN, respectively), and the second hidden layer includes 15 or 16 neurons (the number of inputs) using a linear activation function for all nodes, RMSprop gradient descent optimization algorithm [34], and mean-squared-error as the objective loss function for backpropagation. The forecast accuracy of the ANN approach is compared to the diurnal-linear model to determine if there is a substantial improvement in using nonlinear machine learning in forecasting in lieu of the more straightforward and accessible statistical model. The hybrid diurnal-linear

model with the ANN model is also considered in order to maximize prediction accuracy. However, most work in the water quality prediction literature compares ANN model predictions to linear model predictions rather than combining them, or in the case of Zhu et al. [14], applies the ANN and linear model in parallel rather than series. Additional work in the literature combining mechanistic and machine learning models have not yet performed better than a machine learning model alone [21,22].

3. Results and discussion

3.1. Stability assessment

The initial implementation of ABAC at BWRRF (Test condition 1: ABAC 3.5 mg/L at 90 s) experienced a wide range and high frequency of ammonia fluctuations, as shown in Fig. 4. A follow-up with BWRRF operations staff revealed that the ammonia sensor developed a thick biofilm between an initial cleaning and calibration on April 19 and a second cleaning and calibration on April 30. The biofilm growth could have prevented the transfer of ammonia ions away from the sensor at low concentrations, which may explain why the ammonia values do not approach zero in ABAC 3.5 mg/L at 90 s, unlike the other test conditions. Once the sensor was serviced and the ammonia setpoint increased, ABAC 4.0 mg/L at 90 s demonstrated substantially more process stability. When the PID update frequency was increased from 90 s to 300 s, the tendency for the system to “overreact” was further reduced, thus leading to the least variable operating condition. Correspondingly, the TSV of ABAC 3.5 mg/L at 90 s is larger than all other test conditions and serves as an indicator of imprecise instrumentation and control. As illustrated in Fig. 5, TSV is found to decrease as the control strategies become more tuned and stable.

To determine if the TSV for the most *qualitatively* stable strategy (ABAC 4.0 mg/L at 300 s) is significantly less than the other control strategies tested, two metrics were examined: the difference in TSV’s (Eq. 1) and the ratio of TSV’s (Eq. 2). As shown in Fig. 6, a Monte Carlo simulation approximated the distribution of a mixed population (i.e., under the null hypothesis), and from this population, a *p* value was calculated (Eq. 3). The histograms are approximately normally distributed, so both the difference and ratio of TSV values are considered stable metrics. In all cases except ABAC 3.5 mg/L at 90 s, the observed difference and ratio of TSV values are significantly smaller than the DO control strategy. The increased TSV for ABAC 3.5 mg/L at 90 s is expected due to the ammonia sensor failure during that time.

The method utilized here (calculating TSV of two control strategies and using a Monte Carlo simulation to determine if the difference is significant) is not limited to aeration control or BNR and could be applied to a variety of WWTP unit processes such as anaerobic digestion and solids settling processes. The development of such a tool to quantitatively assess performance will allow utilities to systematically approach system-wide optimization rather than rely on visual interpretation of timeseries and single variable monitoring. In this case, the tool assisted operations at BWRRF in deciding which ABAC strategy to use, namely ABAC 4.0 mg/L 300 s. Additionally, TSV could be used as a monitoring tool to detect increased variability due to sensor instability, as shown in Test condition 1.

3.2. Ammonia forecast

To further improve the performance of ABAC, forecasting models were designed such that the model forecast could be used as the input for the feedback cascade control system at BWRRF. The diurnal-linear model and ANN models were fit, tested, and compared to determine which modeling approach would provide the greatest performance improvement over persistence for predicting future ammonia concentrations.

3.2.1. Diurnal-linear model

The diurnal component of WWTP influent flows and loads is a well-known phenomenon (Fig. 3) but is rarely accounted for in daily operation. To begin modeling the diurnal variation, an initial model fit

used a single sine/cosine pair. However, this approach did not capture all cyclic patterns visible in the timeseries and autocorrelation plots. Further testing evaluated model fit using between 1 and 200 sine/cosine pairs to forecast Zone 7 ammonia concentration. Fig. 7 plots model fit as a function of the number of sine/cosine pairs up to 10 pairs. The best diurnal model was effectively achieved using a 6-degree diurnal model. The relatively low R^2 value of the ABAC 3.5 mg/L at 90 s control configuration (Test condition #1) is likely due to the abnormal process variation caused by a sensor failure, evident in the minimum and maximum ammonia values discussed in the previous section (Fig. 4). The diurnal model is independent of the forecast horizon but dependent on the training window size.

To accurately forecast ammonia, the environmental and operational variations not described by diurnal effects are modeled by

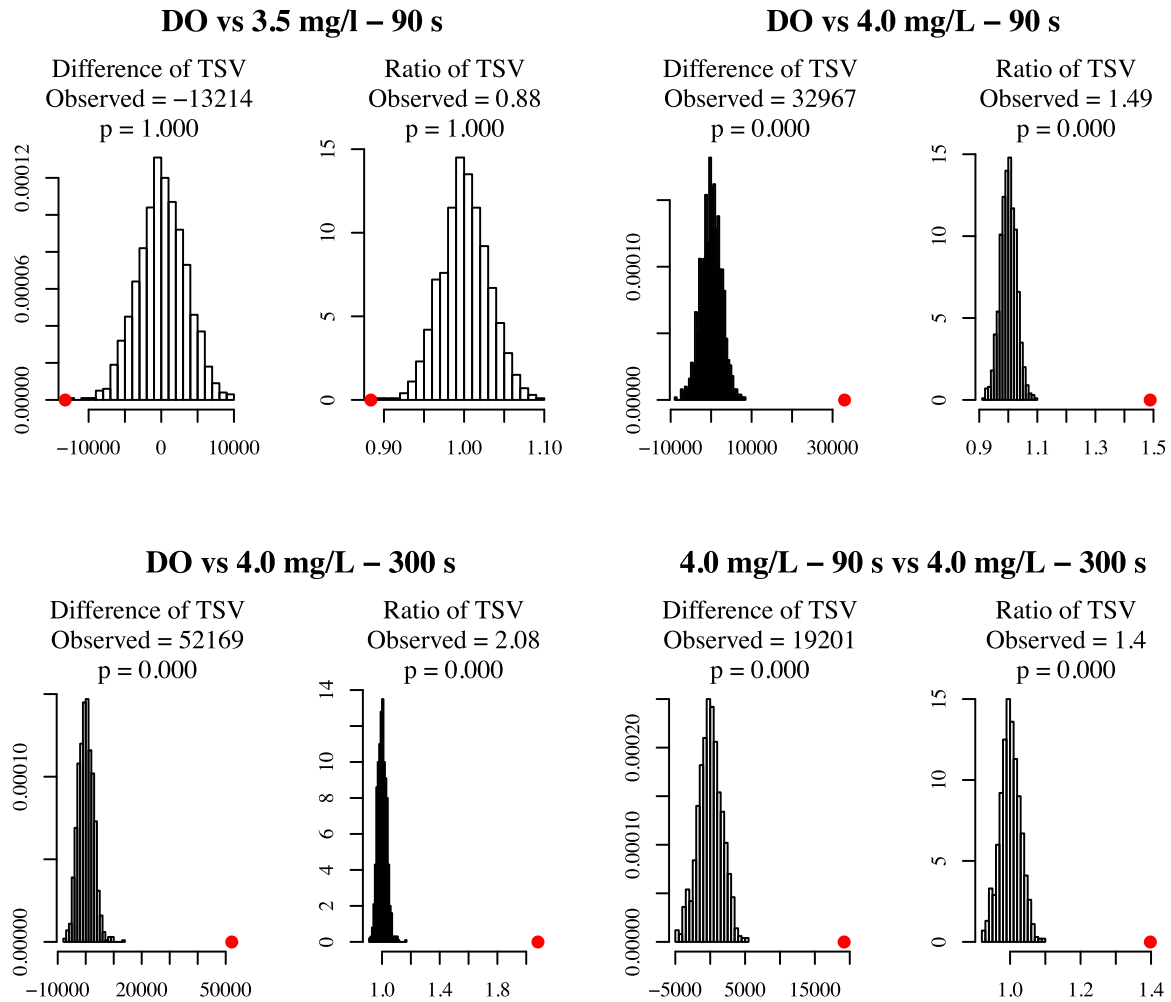


Fig. 6. Histograms of the TSV difference or ratio for a randomized sample of two control strategies. Red dots indicate the observed TSV difference or ratio for a given comparison. For example, “DO vs 4.0 mg/L - 300 s” illustrates the difference of TSV ($TSV_{DO} - TSV_{4.0,300s}$) and the ratio of TSV ($TSV_{DO} / TSV_{4.0,300s}$). If TSV_{DO} is significantly larger than $TSV_{4.0,300s}$, then the difference would be positive, and the ratio would be greater than 1. The p values are calculated by the percentage of differenced or ratio of TSV values from the mixed population that are above the observed differences or ratios. If $p < 0.01$, then the difference is said to be significant.

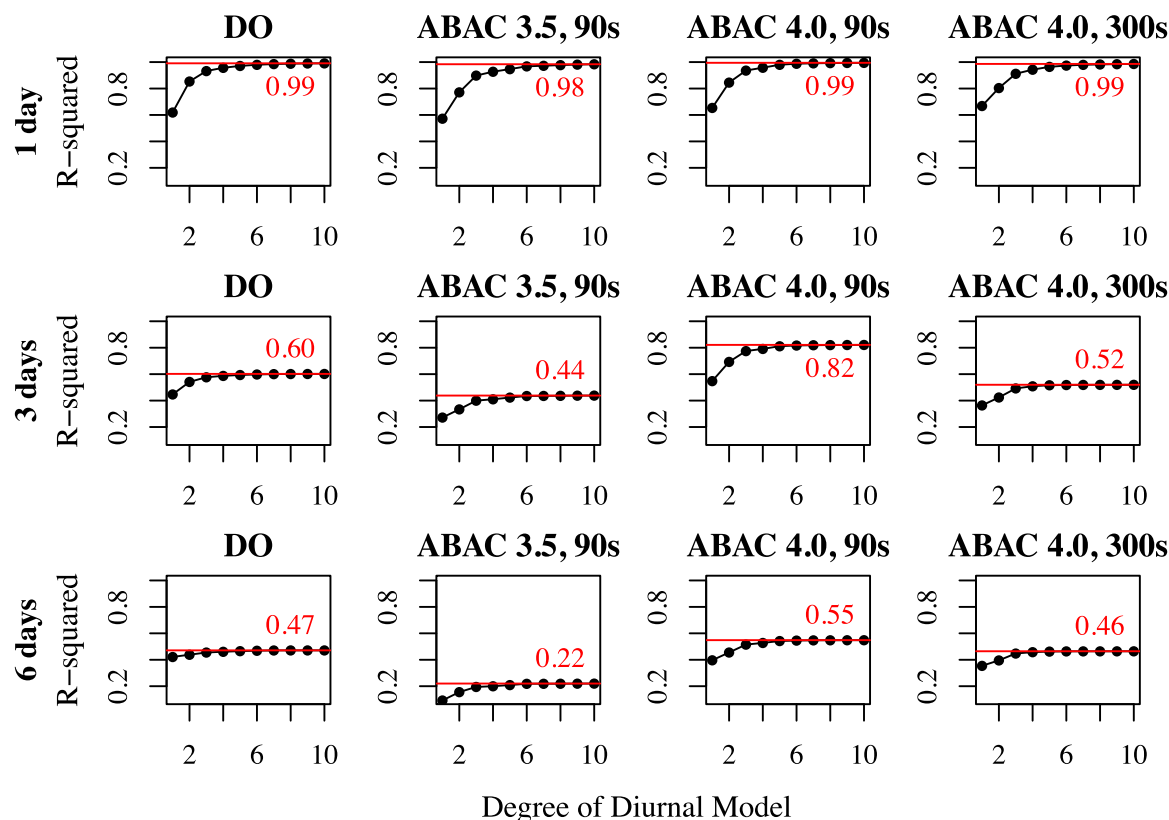


Fig. 7. Diurnal model fit as function of degree for each control configuration using 1 day (top), 3 days (middle), and 6 days (bottom) of training data. The red line indicates the R^2 for a 10th degree diurnal model, which is effectively achieved using 6 degrees or fewer. The high R^2 of the 1 day model compared to the 6 day is a consequence of the 1 day window containing less variation.

multiple linear regression, which is fit using adaptive lasso with the time-dependent diurnal terms and monitored process variables (Fig. 2). In Fig. 8, the multiple linear regression model fit is plotted versus time for each training window (1–6 days) and illustrates multiple trends. Model fit is found to be a function of forecast horizon, size of the training window, and day of the week. As the forecast horizon increases, variation unaccounted for by the linear model emerges, and the model fit deteriorates. Thus, the diurnal-linear model performs best with shorter forecast horizons (highest R^2). Shorter training windows may also have comparatively higher training R^2 than longer training windows because the model encompasses a smaller range of environmental and operating conditions. A shorter training window may be needed when adapting to weekend conditions as opposed to weekdays, which tend to exhibit more nonlinear patterns regarding time, flow, and load. However, shorter training windows produce model parameter estimates that may be unstable and exhibit poor forecast accuracy in the testing set (i.e., parameter estimates may be unrepresentative); this is examined next.

An analysis of model forecast errors on the withheld testing set (i.e., model accuracy given data not used in fitting the model itself) demonstrates a common trend in statistical learning: a high training R^2 does not always correlate to a low testing RMSE. Fig. 9 plots the average training R^2 and testing RMSE for the DO and ABAC 4.0, 90 s

control strategies as a function of forecast horizon. Similar to Fig. 8, the shorter the training window the larger the R^2 , averaged across all testing days. However, the actual performance of the forecast in a control loop is measured by RMSE. RMSE also appears to be a function of training window size, but the direction of the relationship is entirely dependent on the control strategy. For example, shorter training windows performed best across all forecast horizons for the DO control strategy, but the converse is true for the ABAC strategies. Compared to persistence, the diurnal-linear model performs better than persistence for forecast horizons greater than 15-minutes. To implement the diurnal-linear model, the variability of the control system within the training window will impact whether a long or short window of time is necessary. Additionally, R^2 alone should not be used as a selection criteria for such a forecasting model. RMSE or another metric of testing error (see [29]) should be used to identify the best choices to tune a model.

To demonstrate the potential impact of supplementing the diurnal-linear model forecast in lieu of the current ammonia sensor value, the percent improvement between the persistence forecast and the model forecast is calculated (Equation 4). Only the 5, 25, 50, and 75-minute forecast horizons are shown in Table 3 for brevity, but all forecast horizons can be found in the Supplementary Information (Table S1). A forecast horizon should be selected based on the time required to

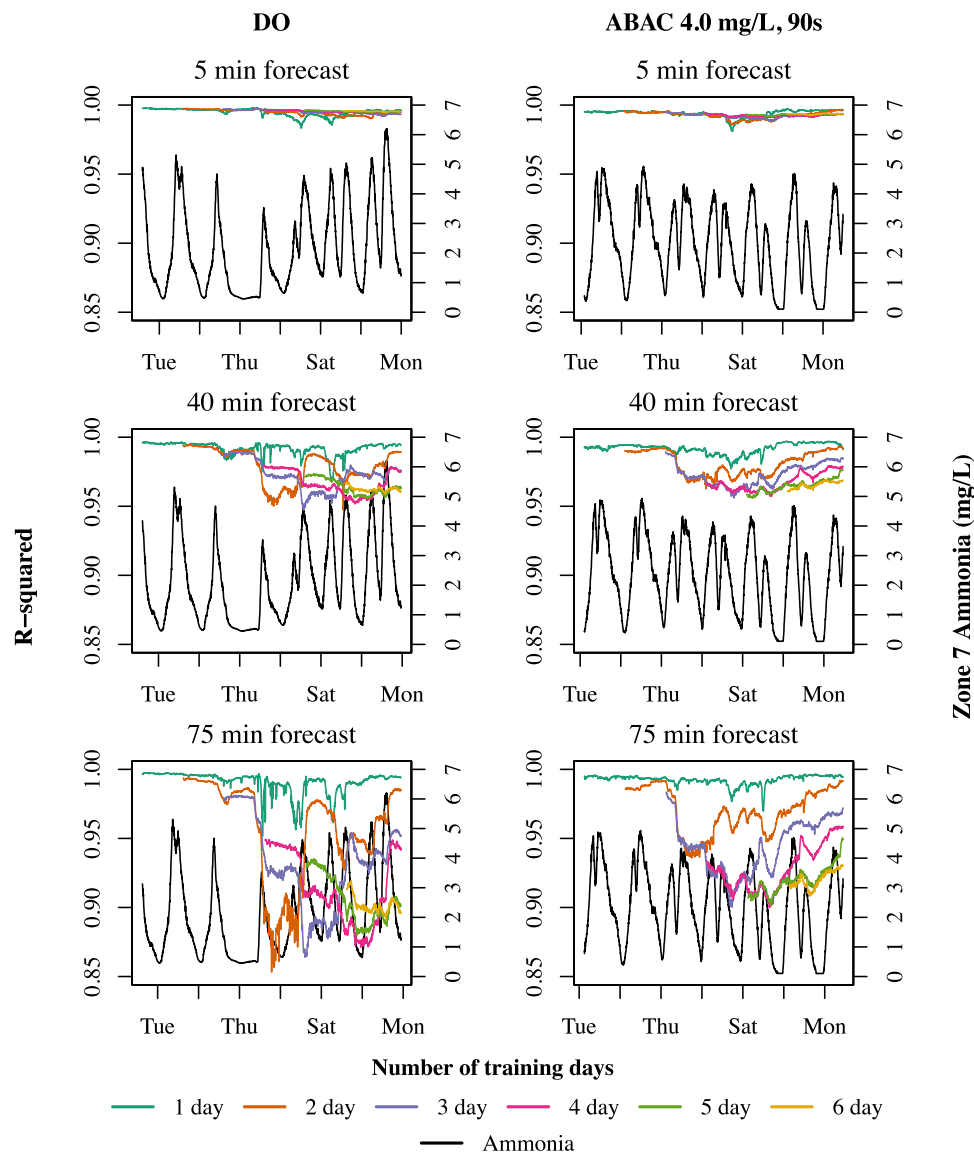


Fig. 8. Fit of diurnal-linear model as a function of training window size as measured by R^2 . Due to the difference in flow and ammonia loading patterns between weekdays and weekends, the model fit declines over the weekend. Shorter training windows generally produce a higher R^2 values, demonstrating the large daily variation in operating and environmental conditions experienced at WWTP.

trigger and implement a single cascade control action. The delay is a function of hydraulic retention time (HRT), which is the travel time between the inlet to the basin and the sensor measurement for the master PID control loop. Depending on the required forecast horizon, a training window can then be selected with the largest percent improvement (i.e., reduction in error by RMSE over persistence). At BWRRF, the average HRT between the beginning of the aeration basin and the ammonia sensor location is 25 min. An additional 15–25 minutes is required for the ammonia ion-selective electrodes and DO sensors to stabilize, so a forecast horizon of 50 min is needed. Across all control strategies listed in Table 1, the 50-minute diurnal-linear

forecasts improved over persistence between 20% and 52%. In general, a short forecast horizon performs best when trained on short windows of time. Conversely, large forecast horizons (>15–25 min, depending on the control strategy) require larger training windows. Specifically for the ABAC 4.0 mg/L at 300 s (the ABAC configuration selected by BWRRF for implementation), a 6-day training window can improve the 50-minute forecast performance by 30%. Functionally, this means that the PID loop will be triggered 30% sooner than under the current control strategy. For ABAC, this additional treatment time will primarily address the daily morning peak flow event in which the influent flow increases from ~12 MGD to ~20 MGD over the course of 60–120

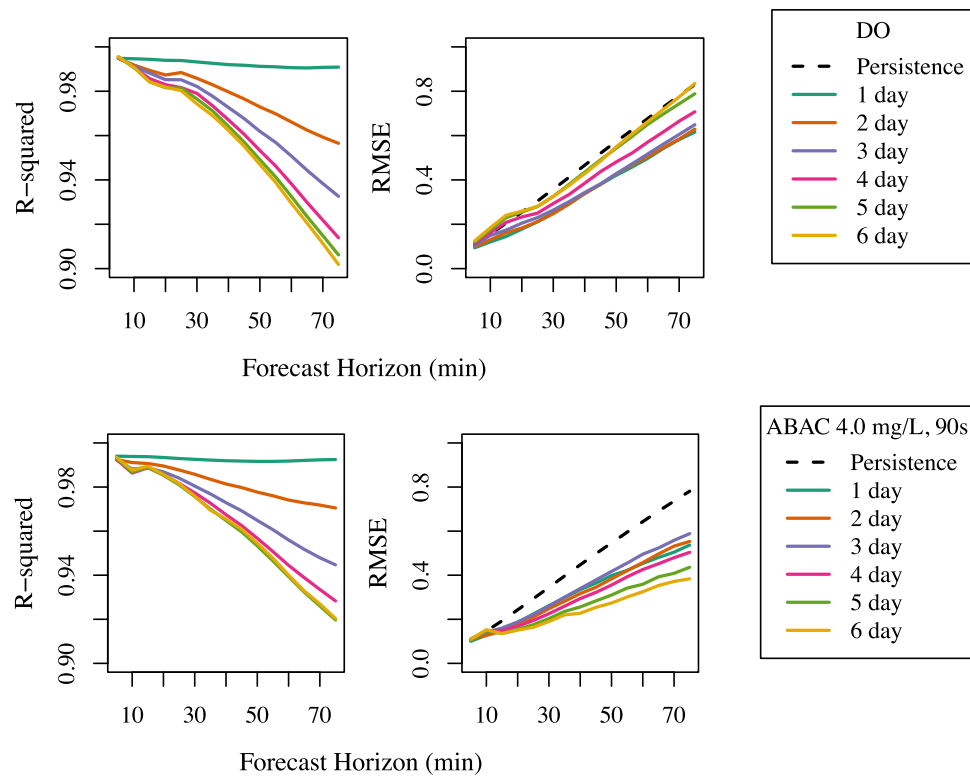


Fig. 9. Training R^2 (left) and testing RMSE (right) for the models tested on the DO and ABAC 4.0 mg/L, 90 s control configurations. The black dashed line is the RMSE of the *persistence* forecast, which corresponds to the ammonia sensor's current measured value as the forecast.

Table 3

Average percent improvement for all control strategies (DO and ABAC) by training window length and forecast horizon for the diurnal-linear model. Positive values indicate that the model forecast was more accurate than the persistence forecast. Bolded values are the best performing training window for a given forecast horizon and control strategy.

Control Configuration	Training Window	5 Min	25 Min	50 Min	75 Min
DO	1 Day	2.2	23.0	18.2	16.7
	2 Day	-1.1	23.5	17.4	15.5
	3 Day	-4.1	19.0	19.6	15.3
	4 Day	-2.7	18.3	16.1	14.6
	5 Day	-2.4	16.4	14.2	14.5
	6 Day	-2.5	20.2	17.8	14.9
ABAC 3.5 mg/L, 90 s	1 Day	1.2	-17.8	-20.8	-13.9
	2 Day	-1.4	16.2	11.3	5.1
	3 Day	-3.4	15.6	17.5	14.9
	4 Day	-1.9	20.8	20.8	22.7
	5 Day	-2.4	17.4	18.3	19.0
	6 Day	-2.8	23.2	24.7	24.2
ABAC 4.0 mg/L, 90 s	1 Day	2.5	21.6	24.7	28.7
	2 Day	-3.5	27.2	28.4	27.6
	3 Day	-3.9	24.5	23.6	24.5
	4 Day	-3.7	34.0	35.3	36.1
	5 Day	-4.1	39.9	43.7	44.9
	6 Day	-2.7	46.7	51.9	52.7
ABAC 4.0 mg/L, 300 s	1 Day	0.4	2.0	-11.5	-18.5
	2 Day	-3.3	2.4	-49.5	-104.8
	3 Day	-3.8	20.2	21.7	24.1
	4 Day	-3.2	24.9	22.1	22.0
	5 Day	-3.1	24.8	26.7	25.3
	6 Day	-3.0	23.8	30.0	31.7

min. The preemptive response with a forecasted ammonia PID control strategy will appropriately increase air flow to prevent exceeding the ammonia concentration setpoint. Additionally, the 50-minute forecast will trigger a reduction in air flow when the peak event wanes. However, the current control strategy uses minimum air flow and DO concentration setpoints to ensure treatment under all conditions, which inevitably leads to over-treatment and excessive energy use. Given the knowledge that the system will proactively adjust to peak flow conditions, more precise process control with forecasted master control variables may provide BWRRF operations staff the flexibility to lower the minimum air flow and DO concentration setpoints, reduce over-treatment, and ultimately save energy.

Further investigation into the structure of the diurnal-linear models fit using adaptive lasso provides insight as to which process variables are most important in forecasting and controlling ammonia in the DO and ABAC systems. The absolute values of the model coefficients are averaged across each training window size and forecast horizon and are plotted in Fig. 10. Variable selection for each model is also found to be a function of training window size and forecast horizon. For most models, the current values of influent ammonia, Zone 7 ammonia, and Zone 9 DO are the most important process variables to forecast future Zone 7 ammonia. Interestingly, non-ammonia process variables become more important as the forecast horizon increases while the importance of the current ammonia sensor measurement decreases. This may suggest that the future value of ammonia is more dependent on the quality of the recirculated activated sludge (i.e., Zone 9 DO, Zone 3 nitrate/NO₃). This reinforces our premise that the current measured value of ammonia is not the best representation of future ammonia values and

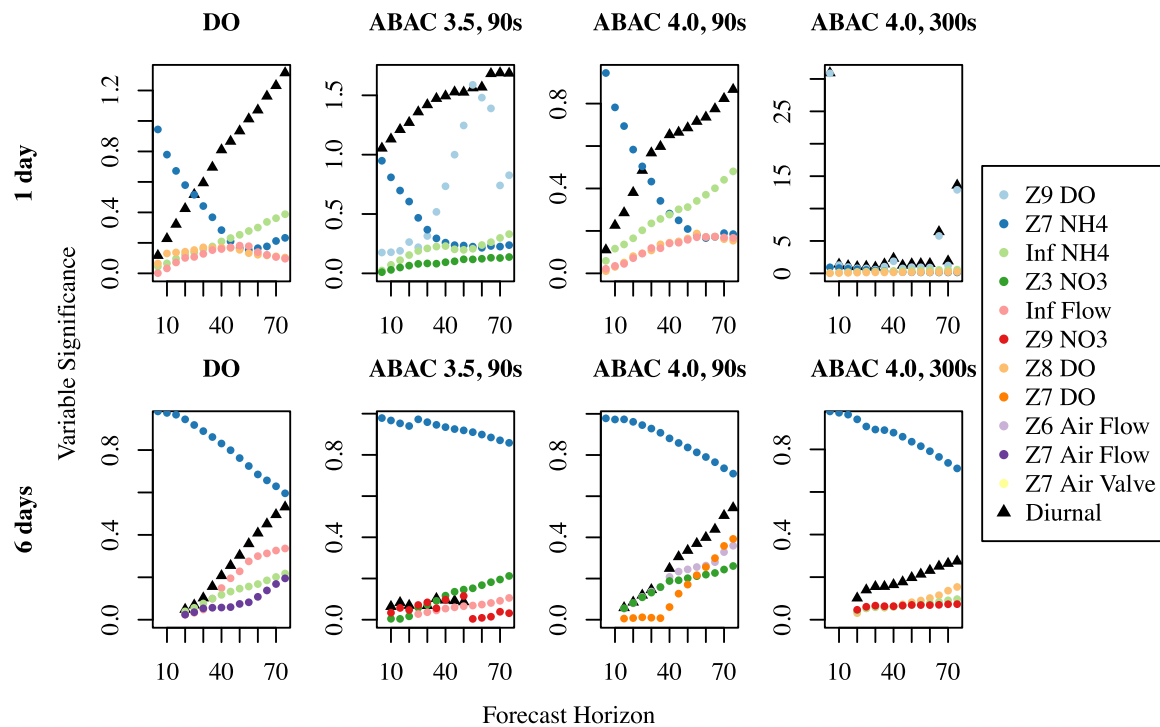


Fig. 10. Variable significance for each diurnal-linear model is calculated by the mean absolute value of the model coefficients for a particular training window, forecast horizon, and control configuration. The top 5 process variables for each control strategy and window size are plotted. The diurnal terms are included by summing the selected sine and cosine coefficients, represented as “Diurnal” in the figure. The strong presence of diurnal terms in all models demonstrates the importance of incorporating process values as well as time-dependent components when modeling WWTP.

illustrates the potential for a forecasting model to substantially improve upon BWRRF’s existing control strategy. Here, the use of adaptive lasso for variable selection as opposed to knowledge-based selection is found to be critical. For example, the conventional feedforward ABAC models only include influent ammonia to calculate air demand and would neglect the impact of oxygen bleed through into anoxic zones, thereby preventing complete nitrogen removal from the system.

3.2.2. Neural networks

The use of ANN to forecast ammonia is intended to capture the nonlinear environmental and operational process relationships that cannot be represented in the simpler diurnal-linear model. A comparison of the diurnal, diurnal-linear, ANN, and diurnal-linear-ANN model forecasts for the ABAC 4.0 mg/L at 300 s control strategy are shown in Fig. 11. The results demonstrate that the ANN overfits to the training data and does not accurately forecast ammonia in the testing dataset, producing forecasts that are worse than persistence. The lack of variable selection in ANN could contribute to the poor forecast accuracy by overtraining to noise present in the training dataset and missing the underlying patterns controlling ammonia concentration. A more complex ANN model may be required to forecast more accurately; however, the computation time required may make this approach impossible to integrate into the existing BWRRF control framework. Furthermore, such “black-box” models tend to be unappealing to process engineers because the important drivers of variation in ammonia cannot be identified.

Nonlinearities in the system that are unaccounted for in the diurnal-linear model begin to manifest themselves as the forecast horizon gets larger, and this is when the diurnal-linear-ANN model performs best. By incorporating the diurnal-linear forecasts as an input to the ANN, the testing accuracy improves over the diurnal-linear model for larger forecast horizons (> 40 min). However, the nonlinear component captured by the ANN does not result in a smooth real-time forecast. The high-frequency variation in the forecasts could be viewed negatively if bulk metrics, such as RMSE, are not used (Fig. 12). Nonlinearities may also be partially modeled by the diurnal component, resulting in a high R^2 but low RMSE due to overfitting. Model performance is also affected by error in the training window, which for full-scale implementation is unavoidable. Thus, a model should be tested on a large window of time to determine how well the model responds to sensor error or changing environmental or operational conditions. By adjusting the number of sine/cosine pairs and including the ANN, the forecast performance generally improves. The full-scale tuned model in Fig. 12 was achieved by iteratively adjusting the number of diurnal pairs, training window length, and observation frequency, and comparing the $RMSE_{model}$ to $RMSE_{persistence}$. Overall, the use of either the diurnal-linear model or the diurnal-linear-ANN model outperforms the current control strategy for all forecast horizons; the best performing combination of method and training window size (i.e., smallest RMSE) for the ABAC 4.0 mg/L at 300 s control strategy is given in Table 4.

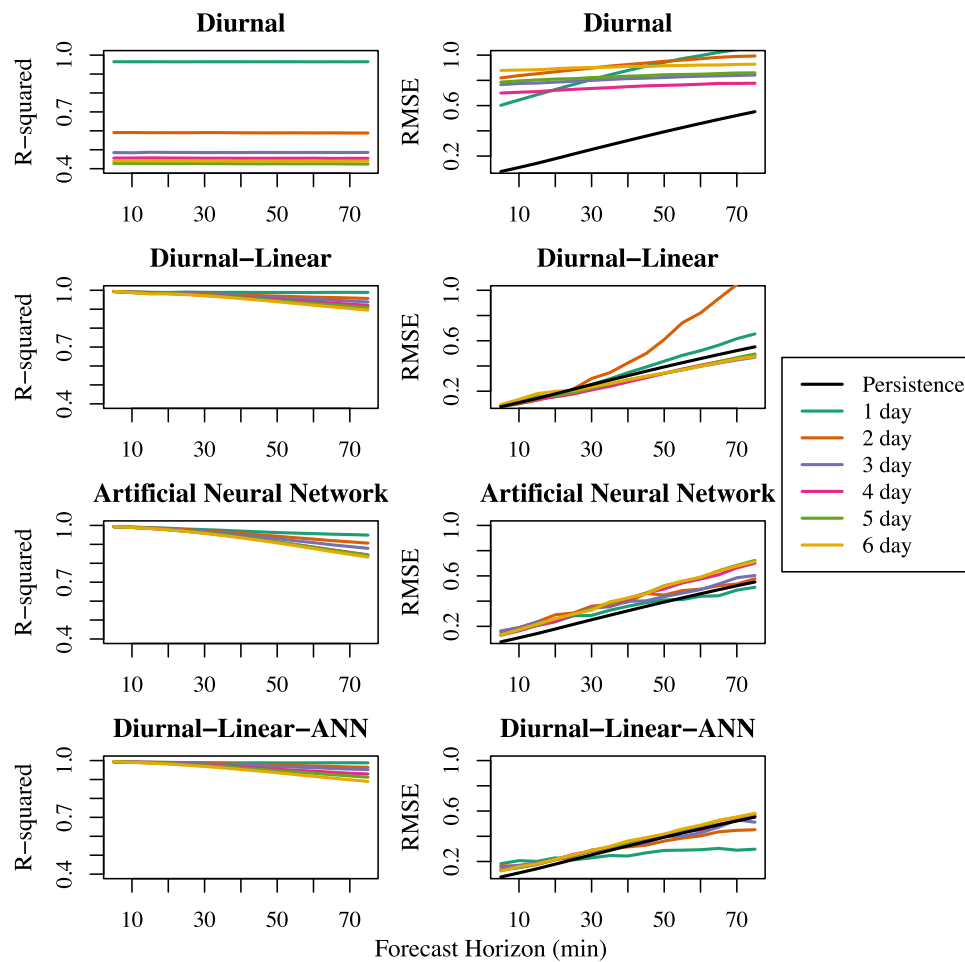


Fig. 11. A comparison of the training R^2 (left) and testing RMSE (right) of the ABAC 4.0 mg/L at 300 s delay control strategy for diurnal, diurnal-linear, ANN, and diurnal-linear-ANN models. For this control configuration, the diurnal-linear model with longer training windows performed the best for forecasting ammonia up through 70-minute horizons. The diurnal-linear-ANN model makes some improvement in forecasting with short training windows, particularly for large forecast horizons.

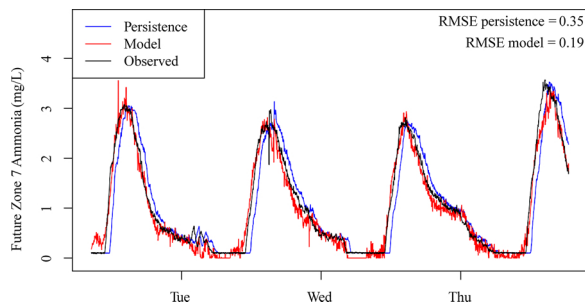


Fig. 12. Real-time ammonia forecast in the BWRRF SCADA system during full-scale tuning. The black line is the true concentration of ammonia in Zone 7 of the aeration basin (“Observed”). The colored lines represent the corresponding 50-minute forecasts. The blue line is the ammonia sensor measurement used for conventional cascade ABAC (“Persistence”). The red line is the diurnal-linear-ANN ammonia forecast trained on 4-days of process data, 2-minute observation frequency, and 3 diurnal pairs (“Model”). In spite of the high-frequency oscillations in the diurnal-linear-ANN forecast, it is still a better representation of anticipated conditions in the basin than the measured value of ammonia (e.g., $RMSE_{\text{model}} < RMSE_{\text{persistence}}$). Error below 0.5 mg/L could potentially be avoided by moving the ammonia sensor upstream in the aeration basin where the ammonia concentrations are higher, and the sensor’s measurements would be within the calibrated range (>1 mg/L).

Table 4

The combination of model and training window with the lowest RMSE for each forecast horizon is listed for the ABAC 4.0 mg/L at 300 s control strategy. For short forecast horizons, the diurnal-linear model performs best with a short to moderate size training window. For longer forecast horizons, the hybrid diurnal-linear-ANN model performs best with a one-day training window.

Forecast Horizon (min)	Best Method / Window (days)	Improvement Over Persistence (%)
5	Diurnal-Linear / 1	0.4
10	Diurnal-Linear / 1	8.6
15	Diurnal-Linear / 2	12.9
20	Diurnal-Linear / 4	20.6
25	Diurnal-Linear / 4	24.9
30	Diurnal-Linear / 4	24.4
35	Diurnal-Linear / 4	25.0
40	Diurnal-Linear-ANN / 1	24.6
45	Diurnal-Linear-ANN / 1	25.1
50	Diurnal-Linear-ANN / 1	26.7
55	Diurnal-Linear-ANN / 1	31.9
60	Diurnal-Linear-ANN / 1	36.0
65	Diurnal-Linear-ANN / 1	38.0
70	Diurnal-Linear-ANN / 1	44.5
75	Diurnal-Linear-ANN / 1	46.1

4. Conclusion

The ability to forecast peak flow and ammonia events at WWTP could allow for a proactive aeration response. The current reactive control paradigm results in multiple short-lived peak aeration events per day, which is mechanically hard on individual blowers and requires the use of auxiliary blowers. Consequently, the control strategy is energy intensive and inefficient. In this work, DO and ABAC aeration control strategies are compared by their process stability using a TSV metric, to assist in operator and engineer decision-making. To further improve aeration control, the accuracy of a forecasted ABAC strategy was evaluated and shown to improve upon the current persistence forecast used in PID cascade control. Two data-driven modeling approaches were compared, statistical learning and machine learning, and we found that the diurnal-linear was able to more accurately forecast ammonia than a standard ANN. When the diurnal-linear model was combined with ANN, model forecast accuracy was improved for larger forecast horizons (>40 min), which suggests ANN are able to account for nonlinearities that are not consequential at shorter forecast horizons. The hybrid diurnal-linear-ANN forecast improvement over persistence is from 24% to 46% for the 40 to 75-minute forecast horizons.

For full-scale implementation of the forecast, tuning is required to determine the optimum number of diurnal pairs (e.g., sine/cosine pairs), training window length, observation frequency, and forecast horizon. Additionally, bulk metrics such as TSV and RMSE should be used to compare process variation and error and to identify the best diurnal-linear-ANN model configuration for a given forecasting horizon. By using a forecasted value of ammonia in lieu of measured ammonia, it may be possible to maintain a more stable effluent ammonia concentration; for example, when high ammonia loading is forecast, aeration will increase to keep pace with demand and will limit aeration when low ammonia loading is forecast. Proactive aeration control strategies that use high-quality forecasts could help small and medium-sized WWTP reduce energy requirements while providing high-quality effluent. The computationally feasible statistical and machine learning models explored here can be easily integrated into a WWTP's control system.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Science Foundation Partnership for Innovation: Building Innovation Capacity project 1632227; the National Science Foundation Engineering Research Center program under cooperative agreement EEC-1028968 (ReNUWit); a grant from the Colorado Higher Education Competitive Research Authority (CHECRA); the Water Research Foundation and the Water Environment Federation through the Leaders Innovation Forum for Technology (LIFT) initiative; and the City of Boulder's Water Resource Recovery Facility's engineering, operations, and process automation staff. The anonymous comments from three reviewers also greatly contributed to framing the content of this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jwpe.2020.101389>.

References

- [1] USEPA, Working in Partnership with States to Address Phosphorus and Nitrogen Pollution Through Use of a Framework for State Nutrient Reductions, (2011).
- [2] Eddy Metcalf, Wastewater Engineering: Treatment and Resource Recovery, 5th ed., McGraw-Hill, 2014.
- [3] M.W. Falk, J. Neethling, D.J. Reardon, Striking the Balance between Nutrient Removal in Wastewater Treatment and Sustainability, IWA Publishing., 2011, <https://doi.org/10.2166/9781780403298>.
- [4] D. Rosso, L.E. Larson, M.K. Stenstrom, Aeration of large-scale municipal wastewater treatment plants: state of the art, Water Sci. Technol. 57 (2008) 973–978, <https://doi.org/10.2166/wst.2008.218>.
- [5] M. Bellucci, I.D. Ofiteru, D.W. Graham, I.M. Head, T.P. Curtis, Low-dissolved-Oxygen nitrifying systems exploit ammonia-oxidizing Bacteria with unusually high yields, Appl. Environ. Microbiol. 77 (2011) 7787, <https://doi.org/10.1128/AEM.00330-11>.
- [6] L. Rieger, R. Jones, P. Dold, C. Bott, Myths about ammonia feedforward aeration control, Proceedings of the Water Environment Federation (2012), <https://doi.org/10.2175/193864712811726392>.
- [7] G. Duffy, S. Kestel, M. Gray, G. Lee, T. Stahl, J. Zhang, Dissolved oxygen control based on Real time oxygen uptake rate estimation, Proceedings of the Water Environment Federation 2010 (2010) 7400–7408, <https://doi.org/10.2175/193864710798207594>.
- [8] A. Visioli, Practical PID Control, Springer Science & Business Media, 2006.
- [9] N. Banadda, I. Nhapi, R. Kimwaga, A review of modeling approaches in activated sludge systems, Afr. J. Environ. Sci. Tech. 5 (2011) 397–408.
- [10] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis - Forecasting and Control, 4th ed., Wiley, 1994.
- [11] G.W. Ellis, X. Ge, D. Grasso, Time series analysis of wastewater quality, in: R. Briggs (Ed.), Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems. Pergamon, 1990, pp. 441–448, <https://doi.org/10.1016/B978-0-08-040776-0.50059-7>.
- [12] J. Huo, W.L. Seaver, R.B. Robinson, C.D. Cox, Application of Time Series Models to Analyze and Forecast the Influent Components of Wastewater Treatment Plants (WWTPs), in: Impacts of Global Climate Change, American Society of Civil Engineers, Anchorage, Alaska, United States, 2005, pp. 1–11, [https://doi.org/10.1061/40792\(173\)97](https://doi.org/10.1061/40792(173)97).
- [13] X. Qin, F. Gao, G. Chen, Wastewater quality monitoring system using sensor fusion and machine learning techniques, Water Res. 46 (2012) 1133–1144, <https://doi.org/10.1016/j.watres.2011.12.005>.
- [14] J.-J. Zhu, L. Kang, P.R. Anderson, Predicting influent biochemical oxygen demand: balancing energy demand and risk management, Water Res. 128 (2018) 304–313, <https://doi.org/10.1016/j.watres.2017.10.053>.
- [15] D.-J. Choi, H. Park, A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process, Water Res. 35 (2001) 3959–3967, [https://doi.org/10.1016/S0043-1354\(01\)00134-8](https://doi.org/10.1016/S0043-1354(01)00134-8).
- [16] L.J. Zhang, N. Li, J.J. Zhang, X.Y. Tian, Application of neural network in modeling of activated sludge wastewater treatment process, 2017 36th Chinese Control Conference (CCC), Presented at the 2017 36th Chinese Control Conference (CCC), 2017, pp. 4556–4561, <https://doi.org/10.23919/ChiCC.2017.8028074>.
- [17] H.-C. Pu, Y.-T. Hung, Artificial neural networks for predicting municipal activated sludge wastewater treatment plant performance, Int. J. Environ. Stud. 48 (1995) 97–116, <https://doi.org/10.1080/00207239508710981>.
- [18] P. Holubar, L. Zani, M. Hager, W. Fröschl, Z. Radak, R. Braun, Advanced controlling of anaerobic digestion by means of hierarchical neural networks, Water Res. 36 (2002) 2582–2588, [https://doi.org/10.1016/S0043-1354\(01\)00487-0](https://doi.org/10.1016/S0043-1354(01)00487-0).
- [19] A. Ekster, J. Carr, T. Elliott, J. Hagstrom, V. Alchakov, I. Meleshin, Model predictive control saves 47% of energy, improves Nutrient Removal and reduces effluent TSS, Proceedings of the Water Environment Federation, Chicago, IL, 2019.
- [20] L. Vezzaro, J.W. Pedersen, L.H. Larsen, C. Thirsing, L.B. Duus, P.S. Mikkelsen, Evaluating the performance of a simple phenomenological model for online forecasting of ammonium concentrations at WWTP inlets, Water Sci. Technol. (2020), <https://doi.org/10.2166/wst.2020.085>.
- [21] D.Ömer Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell. 23 (2009) 586–594, <https://doi.org/10.1016/j.engappai.2009.09.015>.
- [22] D.S. Lee, C.O. Jeon, J.M. Park, K.S. Chang, Hybrid neural network modeling of a full-scale industrial wastewater treatment process, Biotechnol. Bioeng. 78 (2002) 670–682, <https://doi.org/10.1002/bit.10247>.
- [23] K.B. Newhart, R.W. Holloway, A.S. Hering, T.Y. Cath, Data-driven performance analyses of wastewater treatment plants: a review, Water Res. 157 (2019) 498–513, <https://doi.org/10.1016/j.watres.2019.03.030>.
- [24] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175, [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
- [25] H. Moeni, H. Bonakdari, Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model, Stoch. Environ. Res. Risk Assess. 31 (2017) 1997–2010, <https://doi.org/10.1007/s00477-016-1273-z>.
- [26] K. Lotfi, H. Bonakdari, I. Ebtehaj, F.S. Mjalli, M. Zeynoddin, R. Delatolla, B. Gharabaghi, Predicting wastewater treatment plant quality parameters using a novel hybrid linear-nonlinear methodology, J. Environ. Manage. 240 (2019) 463–474, <https://doi.org/10.1016/j.jenvman.2019.03.137>.
- [27] G. Burger, C. Sigmon, C. Marks, C. Tyler, C. Bye, D. Conidi, K. Brian, Modeling aeration performance for energy reduction, Proceedings of the Water Environment Federation (2019).

- [28] H. Zou, The adaptive Lasso and its oracle properties, *J. Am. Stat. Assoc.* 101 (2006) 1418–1429, <https://doi.org/10.1198/016214506000000735>.
- [29] G. James, D. Witten, T. Hastie, R. Tibshirani, *An introduction to statistical learning*, Springer Texts in Statistics, Springer, New York, NY, 2013, <https://doi.org/10.1007/978-1-4614-7138-7>.
- [30] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B* 58 (1996) 267–288.
- [31] G. Dreyfus, Neural networks: an overview, *Neural Networks SE - 1* (2005) 1–83, https://doi.org/10.1007/3-540-28847-3_1.
- [32] M.A. Nielsen, *Neural Networks and Deep Learning*, Determination Press., 2015.
- [33] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.* 61 (2015) 85–117, <https://doi.org/10.1016/j.neunet.2014.09.003>.
- [34] S. Ruder, *An Overview of Gradient Descent Optimization Algorithms*, arXiv:1609.04747 [cs] (2017).