Introduction

The goal of this work is to reduce the cost of disinfection using peracetic acid (PAA) at the Robert W. Hite Treatment Facility operated by the Metro Wastewater Reclamation District (MWRD) in Denver, CO. Due to differences between the initial PAA pilot and full-scale disinfection installation (e.g., geometry and residence time of disinfection basin, variable influent *E. coli* concentrations, variable PAA initial demand), MWRD experienced an instance of exceeding its *E. coli* discharge limit for a single day (252 MPN/100 mL based on a 7-day geometric mean, 126 MPN/100 mL based on a 30-day geometric mean) while operating in constant CT dosing mode. To ensure an exceedance does not occur and that proper dosing is achieved, MWRD is currently operating at a constant initial dose of PAA (1.2 mg/L PAA at the time of this report. This approach has increased PAA chemical costs substantially and has resulted in a re-evaluation of the PAA dosing strategy.

Manoli et al. (2019) proposed a novel CT-based PAA dosing strategy derived from first principals. A double-exponential model of microbial inactivation was solved given a first order model of PAA decay and an n-CSTR hydraulic model. The formulation predicted effluent fecal coliform concentrations given influent fecal coliform concentration and the integrated CT (ICT). ICT was solved given the initial concentration of PAA, the PAA decay constant, and the initial demand of PAA. The PAA decay constant was estimated by solving for various ICTs using Excel Solver and in for Manoli et al. ranged from 0.01-0.02 min^{-1}. Given the average ICT for a given hour, fecal coliform samples were taken at the inlet and outlet to fit the microbial inactivation model. The fitted parameters (β, k_d, m, k_p) varied with each batch, which demonstrates that the first order model may not fully describe PAA demand and decay kinetics in a real water matrix, requiring four degrees of freedom to fit the model to the observed data.

Manoli et al.'s approach assumes that there is in initial immediate consumption of PAA that occurs prior to exponential decay. From PAA concentration profiles of the disinfection basin at MWRD (Figure 1), this does not appear to be the case. If the initial demand were instantaneous, there would not be such a change in slope between 0-5 minutes and 5-60 minutes. A single exponential model fit for the two sampling campaigns in Figure 1 (excluding the initial dose at time 0) gives a coefficient of determination (R^2) of 0.94, which is slightly less than the double exponential approach of 0.99. When integrated, the difference in CT is 0.4 mg/L*min (24.8 mg/L*min for single exponential and 25.2 mg/L*min for double exponential when HRT is 60 minutes). However, the double exponential approach could not be fit in real time given the position of the current online analyzer. The difference between the time of sample collection and the exponential model prediction is approximately 10-13 minutes. However, this is just for one snapshot in time.

Alternatives to predicting PAA concentration using first order models are non-deterministic approaches, such as statistical models and neutral network models. Both approaches have the advantage of being able to consider the impact of multiple variables without a known relationship. However, water quality and operational parameters of a wastewater treatment system are too complex for many statistical models (e.g., linear regression, generalized linear models, random forest model, support vector machines). Therefore, neural networks (NN) were used to predict concentrations of PAA throughout the disinfection basin and exponential model fits of the predicted concentrations were used to calculate instantaneous CT.

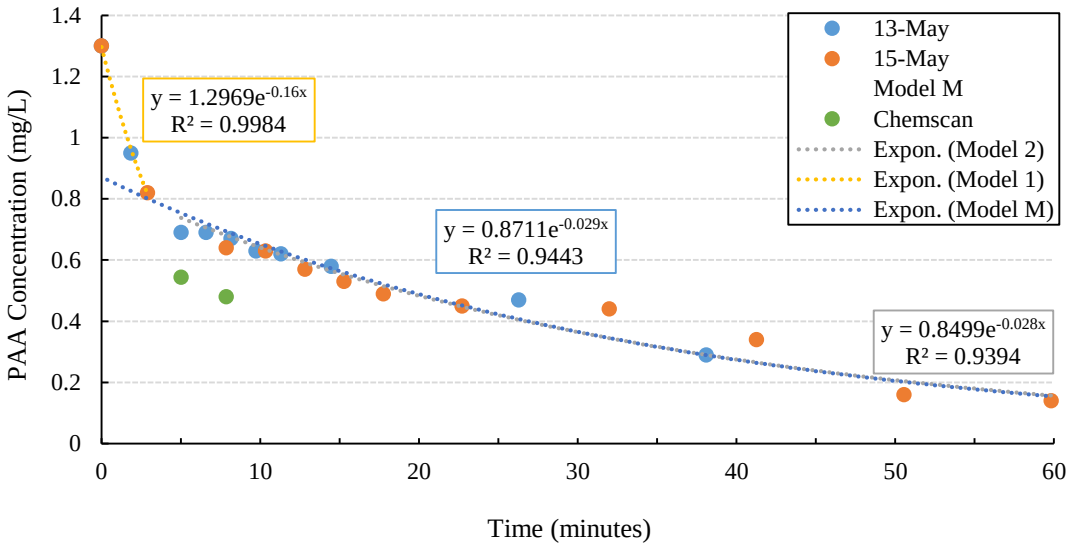

Figure 1. Exponential model fits of two PAA sampling events throughout the MWRD Hite North disinfection basin (observations in blue and orange, model fits in gray and yellow). Online PAA data collected by a Chemscan is plotted in green. The online PAA data is plotted at the sample collection time, not accounting for the accelerated degradation rate with exposure to the autosampler tubing itself, which is a better approximation of conditions at 16-20 minutes of hydraulic retention time as opposed to 5-8 minutes.

Methods

CT is the sum of the area of the curve of PAA concentration as a function of time. Assuming a single exponential model describes the consumption of PAA throughout the disinfection basin, CT is calculated using Equation (1):

$$\int_0^t C dt = \int_0^t C_0 e^{-kt} dt \rightarrow CT = \frac{C_0}{k} - \frac{C_0}{k} e^{-kt} \quad (1)$$

where C is the concentration at time t given the rate constant k and initial concentration C_0. To calculate the single exponential model parameters, known and predicted values of C at two hydraulic retention times (HRT) were used to calculate C_0 and k. The CT model parameters (and the NN model fitting) were calculated using a rolling window approach in which a set number of observations were used to calculate model parameters and the next observation was predicted. To predict the following observation, the oldest datapoint is removed from the training dataset and the most recent is included to keep a constant number of observations in the training dataset. NN model fit was performed in *R* using half of the number of inputs as the number of hidden nodes. The results were compared using root mean squared error (RMSE) and the squared Pearson correlation coefficient (R^2).

NN models use a set of algorithms, modeled loosely after actual brain neural connections, to classify input variables and predict output variables without any prior knowledge of the relationship between inputs and outputs (e.g., predictor and response). NN models can be fit by iteratively fitting NN model parameters (between each variable and at each transformation function or node), which can be computationally intensive. However, once a model is fit, it is quick to calculate predictions but the model itself has no physical significance.

Results

A variety of NN structures were used in each phase of development. Initially, a large number (18) of online process variables were used to predict PAA concentration using a NN with two nodes in the hidden layer (Figures 2). 80% of data observations were randomly selected used to train the models and the remaining 20% of observations were used to test the accuracy of the model by calculating RMSE. Using all variables, PAA concentration at the inlet and halfway through the disinfection basin was predicted within +/- 0.15 mg/L (Figure 3).

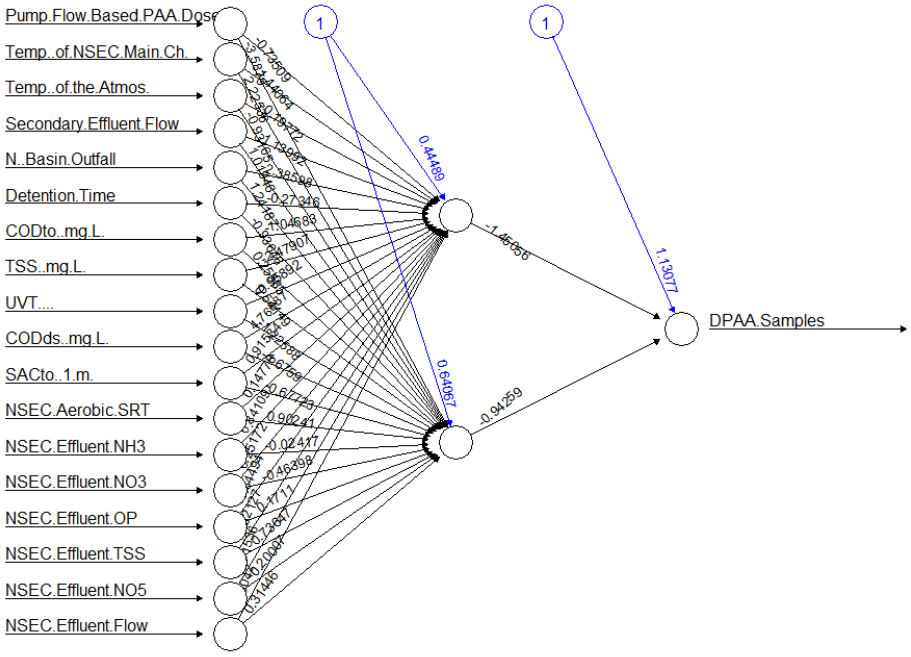

Figure 2. Example neural network model structure to predict PAA concentration from 18 online process variables with two hidden nodes.

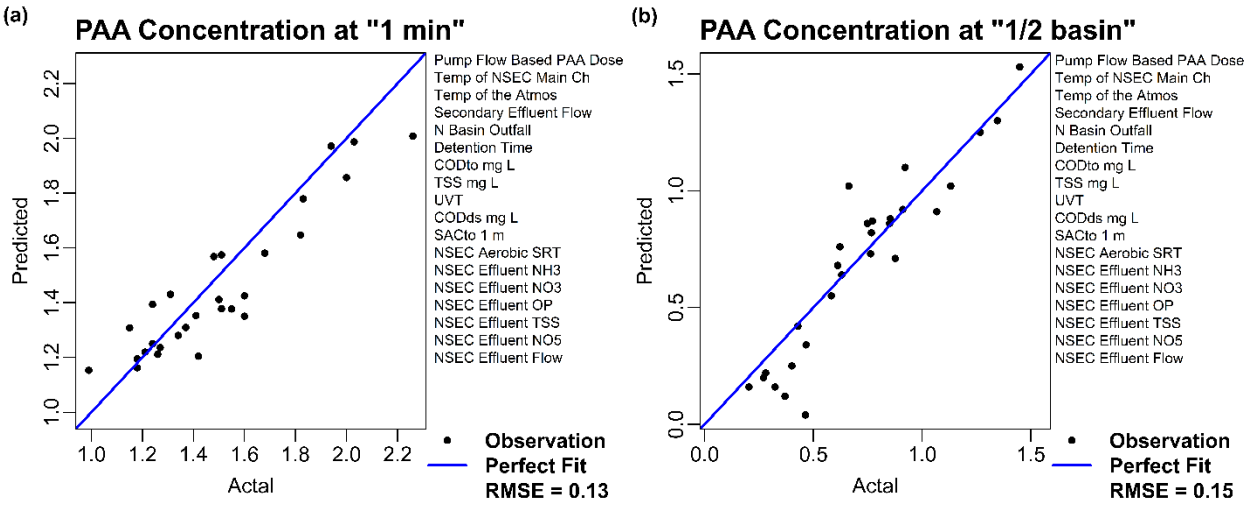

Figure 3. Neural network predictions of PAA concentration at the (a) "1 minute" and (b) "1/2 basin" grab sampling locations given the variables listed to the right of each parity plot.

To identify which variables are most important to the prediction of PAA concentration throughout the basin, combinations of different number of process variables (from Figure 2) and hidden nodes were used to train and test neural network predictions. The most frequently included variables in the models that were successfully able to predict the PAA concentration at the sampling location labeled "1-minute" include flow and water quality parameters of the North secondary effluent (i.e., disinfection basin influent) measured by nutrient sensors and a visual spectrum analyzer at the point of dosing (Table 1).

Table 1. Process variables used as inputs for neural network models to predict PAA concentration at the entrance to the disinfection basin immediately after dosing (labeled "1-minute" sample point, models 1-5) and halfway through the disinfection basin (labeled "1/2 basin" sample point, models 6-10). Each model had one hidden layer with half of the number of nodes as inputs (rounded up when appropriate) for combinations of 3 – 9 variables. Each model was trained 10 times on 80% of randomly selected data from a sampling campaign from October 2, 2018 to October 15, 2018. The remaining 20% of randomly selected observations were used to compare model predictions to actual observed values and calculate R^2.

Model Response	PAA Concentration					PAA Concentration					Instances of Predictor Variable
	Entrance of Basin					Half-Basin					
Model Number	1	2	3	4	5	6	7	8	9	10	
Average R^2	0.82	0.82	0.82	0.82	0.82	0.86	0.86	0.86	0.86	0.86	
PAA Dose	■	■	■	■	■	■	■	■	■	■	10
Temp of NSEC	■	■	■	■	■		■	■			7
Air Temp	■	■	■		■		■				5
NSEC Eff Flow	■		■	■	■				■		5
NSEC ASRT			■			■					2
NSEC Eff TSS				■		■					2
NSEC Eff NO3						■			■	■	3
NSEC Eff NO5	■	■	■		■	■	■	■	■	■	9
NSEC Eff OP		■	■	■	■			■			5
CODto	■			■				■			3
TSS	■					■		■		■	4
CODds	■		■			■		■			4
UVT						■					1
SAC									■		1
Outfall Flow		■					■	■	■		4
Total Predictors	8	6	8	6	6	8	5	8	6	4	

CT is calculated using Equation 1, using PAA concentration values from actual measurements and NN predictions for the models included in Table 1. CT calculations for Model 2 are compared to actual CT calculations in Figure 4. The decline in accuracy (R2) between the individual PAA concentration predictions and the CT predictions indicates that the single exponential CT model either (i) needs more than 2 points to fit or (ii) is not a true representation of PAA consumption throughout the disinfection basin. More PAA measurements during individual sampling events are needed to improve the model fit or to explore more complex PAA consumption models.

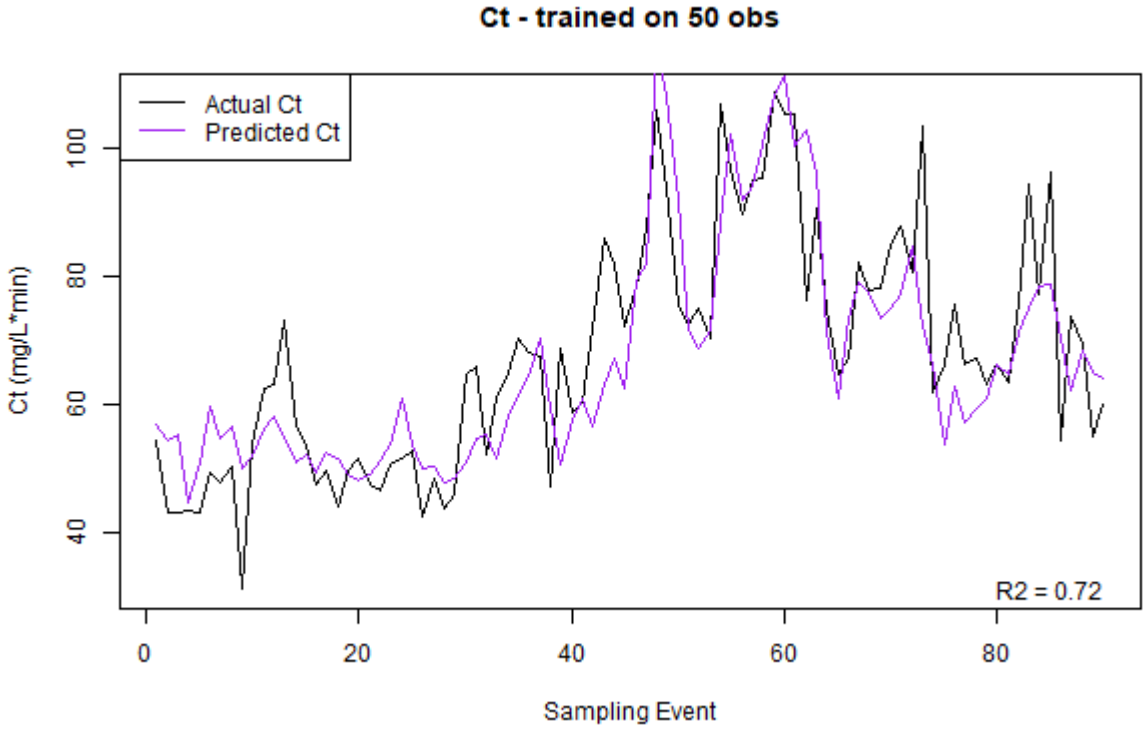

Figure 4. Comparison of calculated CT from actual PAA concentration measurements (black line) and from NN predictions of PAA concentration measurements (purple line). Model parameters were calculated using the previous 50 observations to both fit the single exponential concentration model (Equation 1) and fit the NN. Each time the model updated the oldest observation was excluded and the newest observation was included for the next model fitting. R^2 is calculated from the difference in "actual" and "predicted" CT. For this case, in which a rolling window of 50 observations were used to train Model 2 (Table 1) for both initial and half basin PAA concentration, $R^2 = 0.72$.

Future Work

Predicting CT, and subsequently log inactivation, in real time could act as setpoints for a future control logic. To achieve this, a more detailed sampling campaign that includes more PAA and *E. coli* measurements along the process will improve the CT model fit and prediction. However, we will need at least 50 observations to train and more to validate the model. Additionally, the residence time of the online PAA analyzer needs to be approximated in order to include the data into the NN predictions. Incorporation of the *E. coli* measurements will allow the CT predictions to be fit to a microbial inactivation model. However, it has not yet been feasible to predict influent *E. coli* to the disinfection basin, and prediction of a final effluent *E. coli* is not possible within the scope of this work. Using the double exponential microbial inactivation model will connect PAA predictions and CT/inactivation model fits to predict valuable control parameters for the PAA disinfection system.