Real-time Integrated CT Control for PAA Disinfection in Municipal Wastewater Treatment Using Artificial Neural Networks

Kathryn Newhart, Joshua Goldman-Torres, Daniel Freedman, Blair Wisdom

**Keywords** peracetic acid, machine learning, predictive control, ICT

**Key takeaway** Artificial neural networks can integrate online and lab data to fit process models for use in control, such as integrated CT utilizing peracetic acid.

# 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 while operating at the dose initially recommended by the pilot study (1.2 mg/L). To ensure another exceedance does not occur, MWRD has continually increased PAA dose to respond to poor E. coli inactivation. 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 fitted model parameters (β, kd, m, kp) varied with each batch, demonstrating that a first order model with constant model parameters may not fully describe PAA demand and decay kinetics in a real, constantly changing water matrix, requiring four degrees of freedom to fit the model to the observed data. Alternatives to predicting PAA concentration using first order models are non-deterministic approaches, such as machine learning. Here, artificial neural networks (ANN) were used to predict concentrations of PAA throughout the disinfection basin using water quality and process variables as inputs. The purpose of integrating a neural network model with the CT-dosing strategy is to accurately predict CT by adapting to real-time changes in *k* and *D*, providing a more robust and dynamic disinfection system.

# Materials and methods

A total of 143 observations were collected for this study. From October 2, 2018 through Oct 15, 2018, three daily grab samples were taken (1) immediately downstream of the PAA dosing location and (2) halfway through the disinfection basin. Online data recorded at the time of collection were also included; specifically from nutrient and total suspended solids (TSS) sensor measurements at the end of the secondary treatment process and ultraviolet-visual spectrum measurements (YSI CarboVis®, Yellow Springs, OH, USA) at influent of the disinfection basin.

| Process Variables | Location | Sampling |
| --- | --- | --- |
| Pump Flow Based PAA Dose | Disinfection | Online |
| PAA 1 min Sample | Disinfection | Grab |
| PAA 1 2 Basin Sampling | Disinfection | Grab |
| N Eff TSS Conc | Disinfection | Online |
| Temp of NSEC Main Ch | Secondary | Online |
| Temp of the Atmos | Secondary | Online |
| Secondary Effluent Flow | Secondary | Online |
| N Basin Outfall | Disinfection | Online |
| CODto mg L | Disinfection | Online |
| TSS mg L | Disinfection | Online |
| UVT | Disinfection | Online |
| CODds mg L | Disinfection | Online |
| SACto 1 m | Disinfection | Online |
| NSEC Aerobic SRT | Secondary | Online |
| NSEC Effluent NH3 | Secondary | Online |
| NSEC Effluent NO3 | Secondary | Online |
| NSEC Effluent OP | Secondary | Online |
| NSEC Effluent TSS | Secondary | Online |
| NSEC Effluent NO5 | Secondary | Online |
| NSEC Effluent Flow | Secondary | Online |

To predict and simulate real-time disinfection, neural networks were trained on 80% of the avalible data (114 observations) and tested on the remaining 20% (29 observations) using a rolling window approach. To illustrate: (1) the neural network model is trained to predict the PAA concentration at one of the two sampling locations using observations 1-114, (2) the trained model is used to predict the next observation in time (115); (3) the rolling window moves forward one timestemp and steps 1 and 2 are repeated using observations 2-115 for training and 116 for testing. The rolling window continues to move forward and the model retrained-tested until all 29 observations have been tested. To evaluate the performance of the neural network approach, the training model fit is calculated using R2 and the actual PAA concentration is compared to the model prediction using root mean squared error (RMSE).

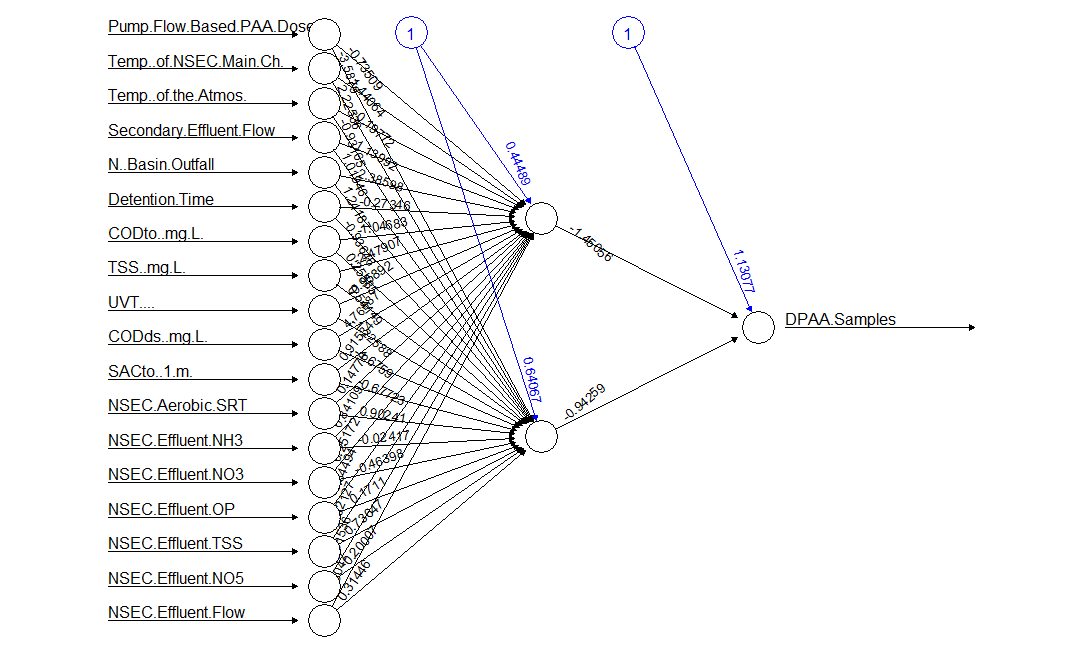
Various neural network configurations are tested in order to determine (1) the optimal model inputs and nodes and (2) which water quality variables impact PAA disinfection potential. For each test, models containing 3-9 process inputs are constructed using 2 hidden layers. The first hidden layer contains half of the number of process inputs and the second hidden layer contains half the number of the first (both rounded up in the case of a fraction). For each configuration, model R2 and RMSE are averaged across the 29 predictions and ranked. The best performing models are analyzed for how frequently each process variable is included to identify correlations between PAA concentration in the disinfection basin, water quality, and upstream treatment performance.

From the PAA concentration predictions at two points in the disinfection basin, total disinfection potential can be calculated by *CT*. 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 from:

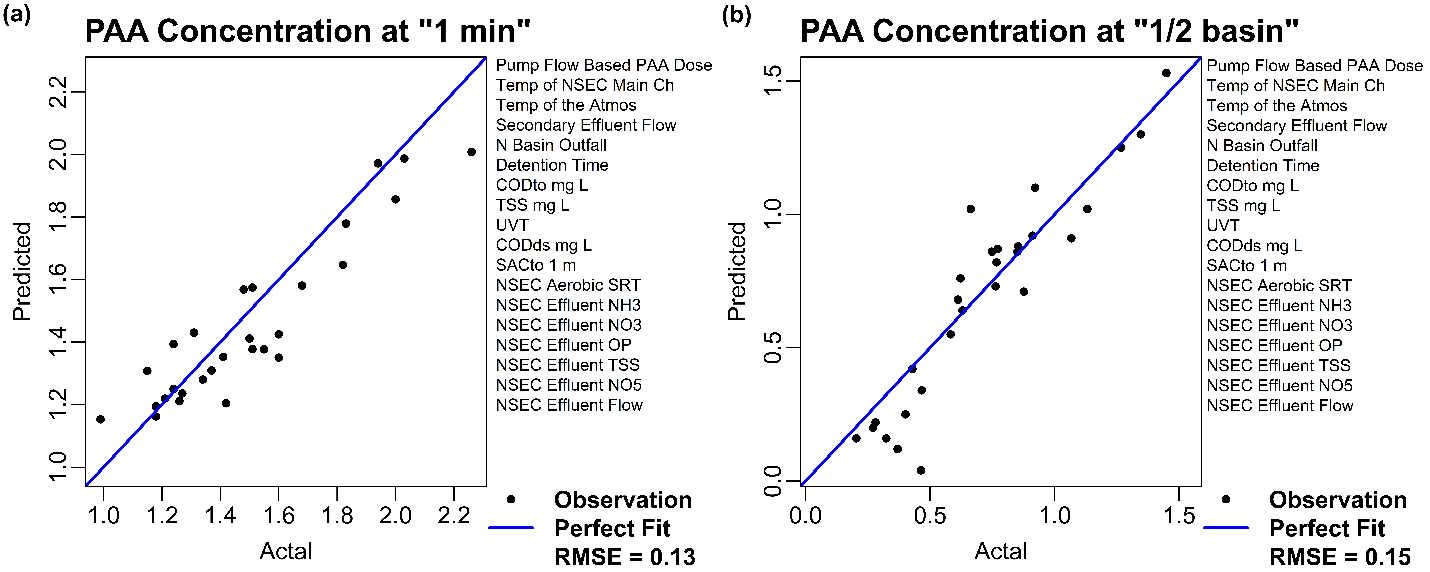
where is the concentration of PAA as a function of hydraulic retention time, is the total hydraulic retention time in the disinfection basin, is the 1st order exponential decay constant, and is the solution to the 1st order exponential decay at . is equivalent to the initial PAA dose minus instantaneous PAA demand (). The curve is fit to the two PAA samples collected at each of the 143 sampling events and compared to the curve fit by the neural network predcition to determine if neural networks could be used in lieu of an online PAA analyzer for CT-based dose control.

# Results

A variety of ANN 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 ANN with two nodes in the hidden layer (Figure 1). Using all variables, PAA concentration at the inlet and halfway through the disinfection basin was predicted within +/- 0.15 mg/L (Figure 2).



**Figure 1.** Example neural network model structure to predict PAA concentration from 18 online process variables with two hidden nodes.



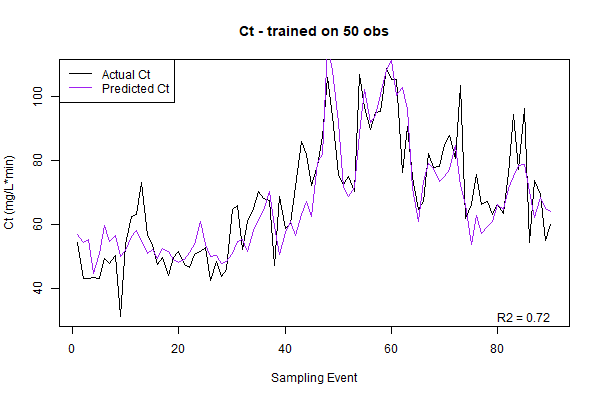
**Figure 2.** Neural network predictions of PAA concentration at the (a) “1 minute” and (b) “1/2 basin” grab sampling locations given the variables listed to the right of each parity plot.

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 1) 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 2).

**Table 2.** 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 R2.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 R2 | 0.82 | 0.82 | 0.82 | 0.82 | 0.82 | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 |
| PAA Dose | x | x | x | x | x | x | x | x | x | x | 10 |
| Temp of NSEC | x | x | x | x | x |  | x | x |  |  | 7 |
| Air Temp | x | x | x |  | x |  | x |  |  |  | 5 |
| NSEC Eff Flow | x |  | x | x | x |  |  |  | x |  | 5 |
| NSEC ASRT |  |  | x |  |  | x |  |  |  |  | 2 |
| NSEC Eff TSS |  |  |  | x |  | x |  |  |  |  | 2 |
| NSEC Eff NO3 |  |  |  |  |  | x |  |  | x | x | 3 |
| NSEC Eff NO5 | x | x | x |  | x | x | x | x | x | x | 9 |
| NSEC Eff OP |  | x | x | x | x |  |  | x |  |  | 5 |
| CODto | x |  |  | x |  |  |  | x |  |  | 3 |
| TSS | x |  |  |  |  | x |  | x |  | x | 4 |
| CODds | x |  | x |  |  | x |  | x |  |  | 4 |
| UVT |  |  |  |  |  | x |  |  |  |  | 1 |
| SAC |  |  |  |  |  |  |  |  | x |  | 1 |
| Outfall Flow |  | x |  |  |  |  | x | x | x |  | 4 |
| Total Predictors | 8 | 6 | 8 | 6 | 6 | 8 | 5 | 8 | 6 | 4 |  |

CT is calculated using PAA concentration values from actual measurements and ANN predictions for the models included in Table 2. 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 3.** Comparison of calculated CT from actual PAA concentration measurements (black line) and from ANN predictions of PAA concentration measurements (purple line). Model parameters were calculated using the previous 50 observations to both fit the single exponential concentration model and fit the ANN. Each time the model updated the oldest observation was excluded and the newest observation was included for the next model fitting. R2 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 2) for both initial and half basin PAA concentration, R2 = 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 ANN 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.