DEMONSTRATION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN POTABLE REUSE PROJECTS

Ozan Erdal, ARCADIS, 320 Commerce, Irvine, CA 92602
Ufuk G. Erdal, ARCADIS, 320 Commerce, Irvine, CA 92602
Email: Ufuk.Erdal@arcadis.com, Phone: 714 508-2642

Abstract

In many potable reuse projects, we largely rely on traditional control logics along with the conservative set points to take a unit process or entire treatment train offline to prevent the public consumption of unsafe water. While it represents a safe solution, it does not necessarily reflect the best management strategy due to potential losses of revenue as well as operational and maintenance challenges due to intermittent operations resulting from advanced treatment trains going frequently offline. Artificial intelligence (AI) and machine learning (ML) have the ability to constantly and quickly adapt and process large amounts of data in real-time with great accuracy, it may be an ideal platform for supporting our decision for managing potable reuse projects and assets. Unfortunately, AI/ML has not been widely used in the water industry despite its huge potential. The objective of this paper is to demonstrate if AI/ML can be successfully applied in potable reuse projects. To do that, AI/ML concepts have been applied to operational data obtained from two different potable reuse projects (one pilot, one full-scale).

Feedforward artificial neural networks (ANNs) using Pytorch were developed to predict:

- RO permeate TOC in a pilot scale IPR train
- Energy use of the RO facility and third stage RO fouling in a full-scale IPR facility

Despite substantial changes in feed water quality and other operational parameters, the ANN model developed well predicted RO permeate TOC with R^2 value of 0.945. Similarly, despite changes in RO feed water quality, operational parameters, and variations in membrane ages in the RO trains, the ANN models well predicted energy use (R^2 of 0.971 to 0.989) and third stage fouling as reflected in the specific flux (R^2 of 0.991) for plant optimization. This study clearly showed that AI/ML is a powerful tool to support our decision making in potable reuse projects while reducing errors and boosting regulatory and public confidence.

Key Words

Artificial intelligence, artificial neural networks (ANNs), direct and indirect potable reuse, machine learning, data analytics

Background and Objectives

Potable reuse has become an alternative water supply choice that is drought proof, locally controllable, sustainable, and in some cases more affordable than other water supply sources (e.g.

sea water, brackish groundwater). Indirect potable reuse (IPR), the augmentation of groundwater or a reservoir with highly purified water and direct potable reuse (DPR), the blending of highly purified recycled water with raw water before treatment or the direct introduction of it into a public water distribution system, are the two main potable reuse strategies. Figure 1 shows process flow schematic of the IPR train used at the Ground Water Replenishment System (GWRS) of Orange County Water District.

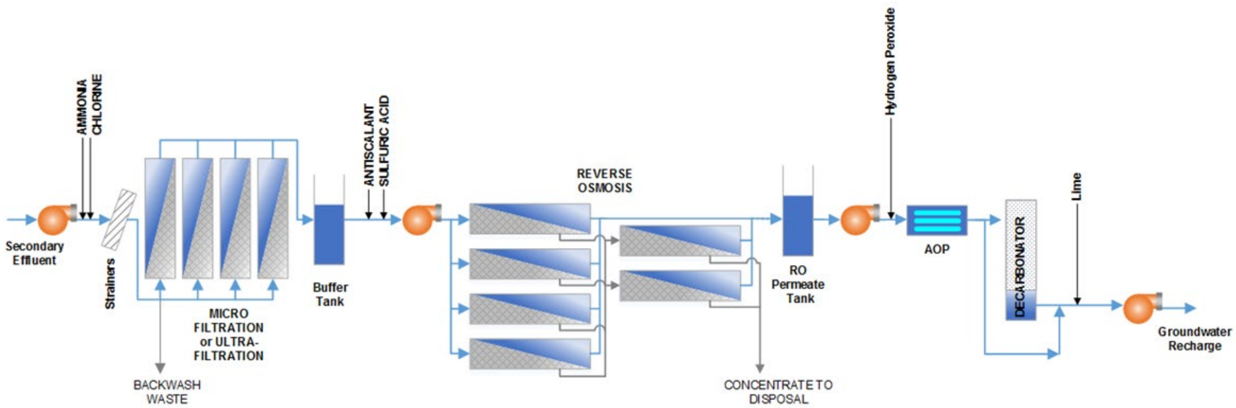

Figure 1. Process Flow Schematic of the IPR train used at GWRS

Significant reduction or elimination of environmental buffers in some potable reuse projects requires additional treatment barriers and/or the implementation of more sophisticated on-line monitoring and control in critical control points to verify the performance and integrity of unit treatment processes. Today, we largely rely on traditional control logics along with the conservative set points to take a unit process or entire treatment train offline to prevent public consumption of unsafe water. While it represents a safe solution, it does not necessarily reflect the best management strategy due to potential losses of revenue as well as operational and maintenance challenges due to intermittent operations resulting from advanced treatment trains going frequently offline.

Reverse osmosis (RO) is one of the main requirements in most IPR and future DPR trains in California, Texas, Arizona and Florida. RO is an energy intensive process that requires continuous chemical pretreatment, periodic clean in place, and periodic membrane replacement, all of which result in relatively high O&M costs. To minimize the O&M and net present value of IPR and DPR treatment, practitioners are looking for ways to optimize O&M costs since predicting energy requirements under various operating conditions (e.g. CIP intervals, chemical doses, etc.) and as a function of membrane replacement may provide a highly valuable information.

AI and ML have the ability to constantly and quickly adapt and process large amounts of data in real-time with great accuracy, it may be an ideal platform for supporting our decision for managing potable reuse projects and the assets we are using. AI is the capability of a machine to imitate intelligent human behavior. AI refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. A subset of artificial intelligence is ML which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Unfortunately, neither AI nor ML have been widely used in the water industry despite the huge potential. The objective of this paper is to demonstrate that AI&ML can be successfully leveraged in potable reuse projects by using data obtained from pilot and full-scale potable reuse treatment trains in California.

Methods

In this study, various ML techniques were explored in search of an effective means for predictions regarding different aspects of the RO process. Although there is an abundance of available learning methodologies to select from, the results of each approach will largely vary, and not every technique will be a perfect fit for every problem area. Based on prior experience with similar prediction tasks, ANNs will generally perform quite well. Due to this effectiveness as well as a series of encouraging preliminary results, an ANN was selected as the structure for the project.

An ANN is a network structure consisting of artificial neurons which are loosely modeled after a human neuron. Though far simpler than a real neuron, the effectiveness of artificial neurons is undeniable. Just as human neurons have dendrites, bodies, and axons, these components are emulated in artificial neurons through inputs, functions, and outputs. What varies from neuron to neuron is the number of inputs and outputs as well as the function representation. Since this definition is general, it is applicable to many problem areas. A simple example for demonstrating the function of a neuron is to represent the equation of a line in two dimensions. Figure 2 shows such a representation.

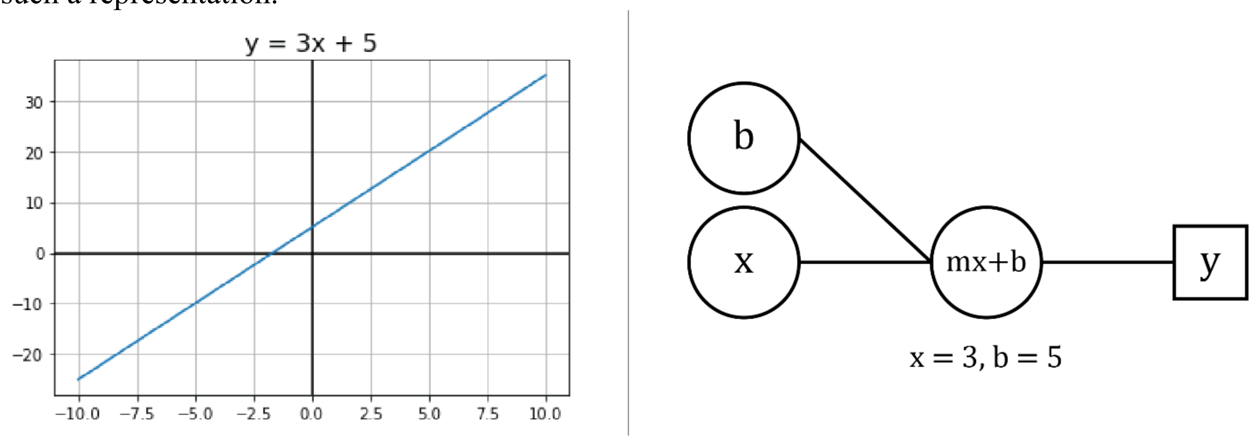

Figure 2. Graphical and Neuron Representation of the 2D Line Equation

The formula of $y = mx + b$ is ubiquitous, serving as a foundation for statistics, linear algebra, engineering, and much more. In fact, this simple function serves as the basis for ANNs as well. In ML, the concept of *slope*, or 'm', is referred to as the *weight*, and the letter 'b' refers to the *bias*, which is analogous to the *y-intercept*; although in both cases, the concepts are synonymous and serve as industry-specific terminology; as such, the terms "weight" and "bias" will be used throughout. Additionally, the 'x' and 'y' terms are the same and are called the input and output. An equation of a surface in three dimensions could be represented with 2 inputs and 1 output, and the neuron diagram would simply have another node appended.

The strength of an ANN lies in its ability to systematically learn the correct value(s) of the weight terms across a network containing a massive number of nodes. After inputs make their way through the network, the weights are adjusted slightly as the model aims to learn the problem space. Once a neural network is understood from its building blocks, it stops appearing as a "garbage in, garbage out" process for analysis. As a matter of fact, these building blocks can be strung together in countless arrangements to form the "network" of neurons – the neural network.

The above figure demonstrates a single neuron computing the final output. If instead, this output was used as the input for other neurons, with their own weights and biases, the network begins to form. These intermediary neurons which do not directly receive raw input are designated as *hidden*. The repetition of this process of feeding computed outputs to other neurons is, when done many times, known as *deep learning*. The inputs for the ANN along with its structure were decided upon based primarily on the combined subject matter expertise in both water treatment and data analysis. Through subsequent testing and experimentation an effective ANN was created. In all cases for this project, an ANN similar to the one shown in Figure 3 was used. Concretely, networks with multiple inputs (green dots in Figure 3), one hidden layer (blue dots in Figure 3), and a single output (yellow dot in Figure 3) were used.

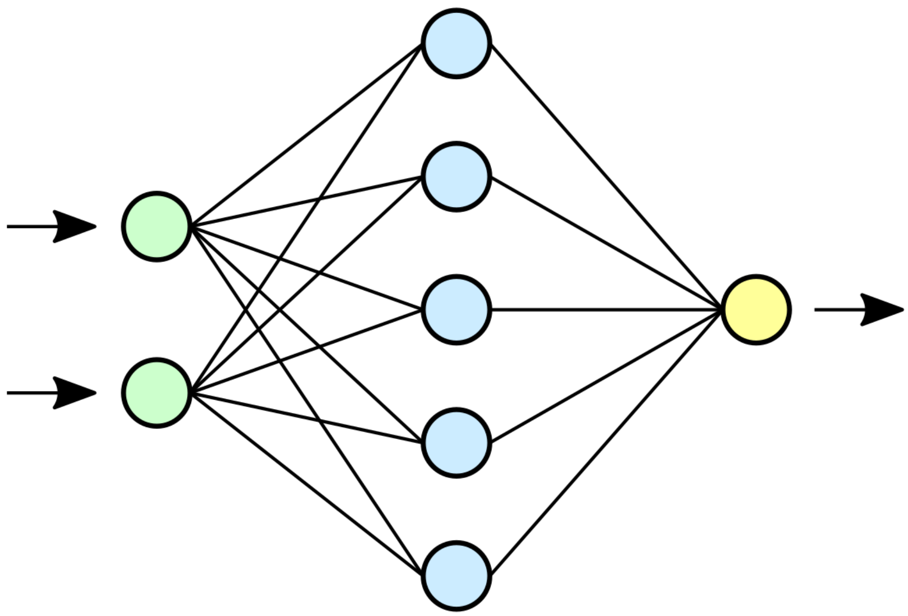

Figure 3. Example of ANN with 1 Hidden Layer and a Single Output (Adapted from Wikimedia.org)

Adding additional hidden layers did not yield significant increases in accuracy; the rule of thumb is to keep the network as simple as possible for the given problem, so the use of a single hidden layer was upheld. Additionally, a 4:3 ratio of hidden neurons to input neurons was determined to work well for this project. Overcomplicating the structure can lead to a significantly slower training process for the model, and it can result in overfitting. Another tactic used to prevent overfitting was the use of a train/test split in the data – by withholding a portion of the data from the training process, a more accurate understanding of the true accuracy of the predictor was reached. In general, 20-30% of the data was withheld for testing purpose.

In this study, the ANN models were developed using on-line data provided from a pilot and full-scale IPR trains. The pilot study data was limited to 120 days of continuous operation whereas two years of continuous data was inputted into the models for predicting full-scale facility performance. The quality of the model was expressed as using the mean squared error of the predicted value compared to the known value for each observation in the test dataset (average of the residuals). The R^2 value of the actual vs. expected for these predictions was also used throughout the project as an indicator of model accuracy.

Results and Discussion

RO Permeate TOC Predictions:

Figure 4 shows the ANN model RO permeate TOC prediction results below.

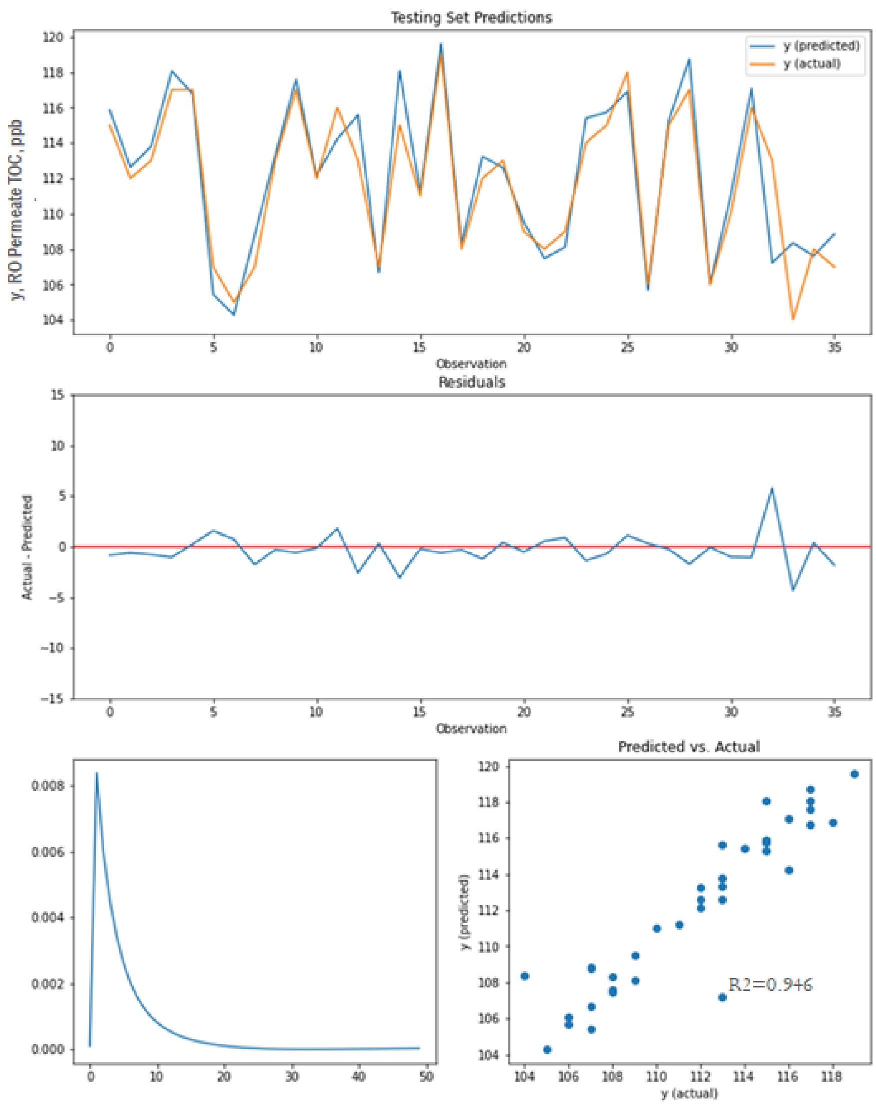

Figure 4. RO Permeate TOC Prediction, Residual and Model Accuracy Analysis

70% of the 120 random data (a total of 84 data) was used for training purposes whereas 30% of the remaining random data (36) was used for prediction purposes. Despite substantial changes in advanced treatment facility feed quality including temperature, the model predicted accurately, and actual RO TOC concentrations were in very good agreement with an R^2 value of 0.946. Modifying training and testing percentages as well as learning rate did not further improve model accuracy.

RO Train and Sub-Group Power Consumption Prediction:

Figure 5 shows the ANN model predictions for full-scale RO facility power consumption for a randomly selected single train (Train A01) on the left, and a randomly selected RO subgroup (Group A, trains A01, A02 and A03) on the right.

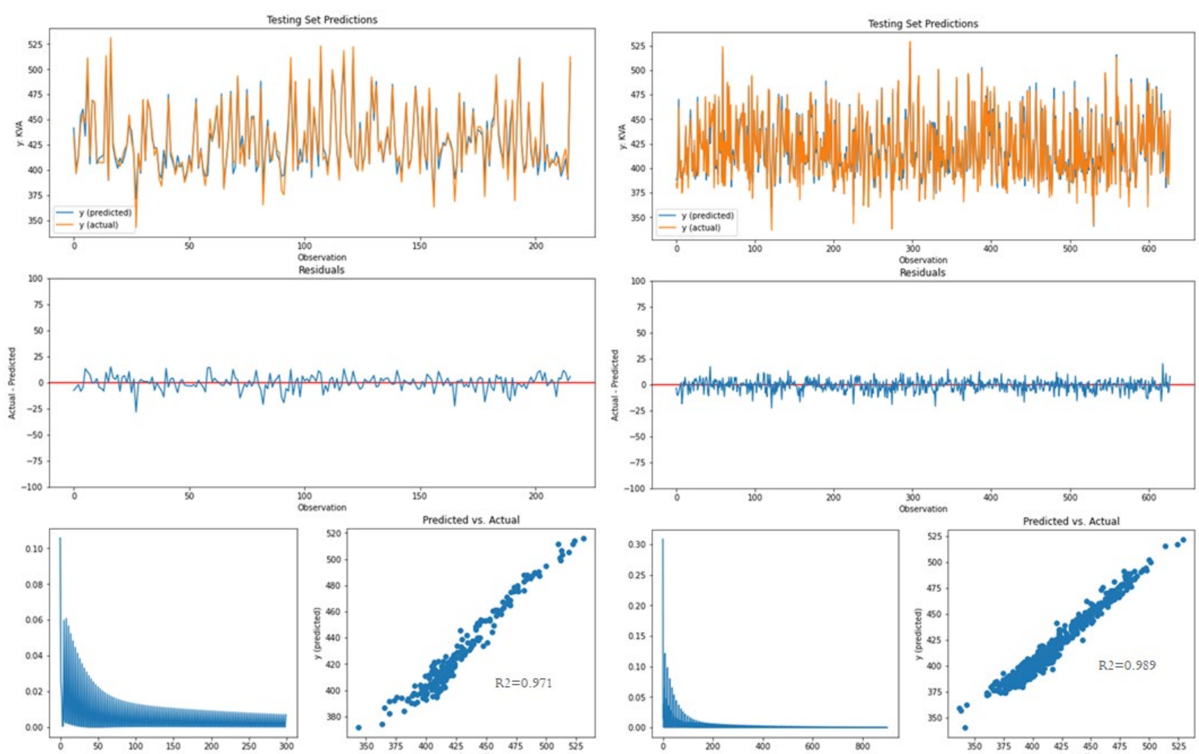

Figure 5. Power Consumption Prediction for Full-Scale RO Facility in an IPR Train

Roughly 233 random data (70%) was used training purposes and the remaining 99 random data (30%) was used for predicting train A01's power consumption out of 332 data total data. Similarly, roughly 80% (train) and 20% (test) of the data out of 718 data was used for predicting RO subgroup A's power consumption. Despite substantial changes in RO operating conditions (e.g. water temperature, third stage RO feed and concentrate RO conductivity, third stage feed pressure, etc.) the accuracy of the models was great, each having an R^2 value of greater than 0.970. Since model accuracies were great, no attempt was made to further improve them.

RO Train Third Stage Fouling (Specific Flux) Prediction:

Specific flux is the amount of pressure required to push a target amount of flow through an RO membrane surface area (gfd/psi). Declining specific flux indicates that the target RO permeate production goals can be met at increased pressures as a result of increased scaling/fouling of RO membranes. The third stage RO specific fluxes were predicted by training and testing approximately 70% and 30% respectively of the total data obtained for Train A01. Figure 6 shows the ANN model predictions for the A01 train excluding (left in the picture) and including RO membrane ages (right in the picture) to see if membrane age has an impact on model prediction.

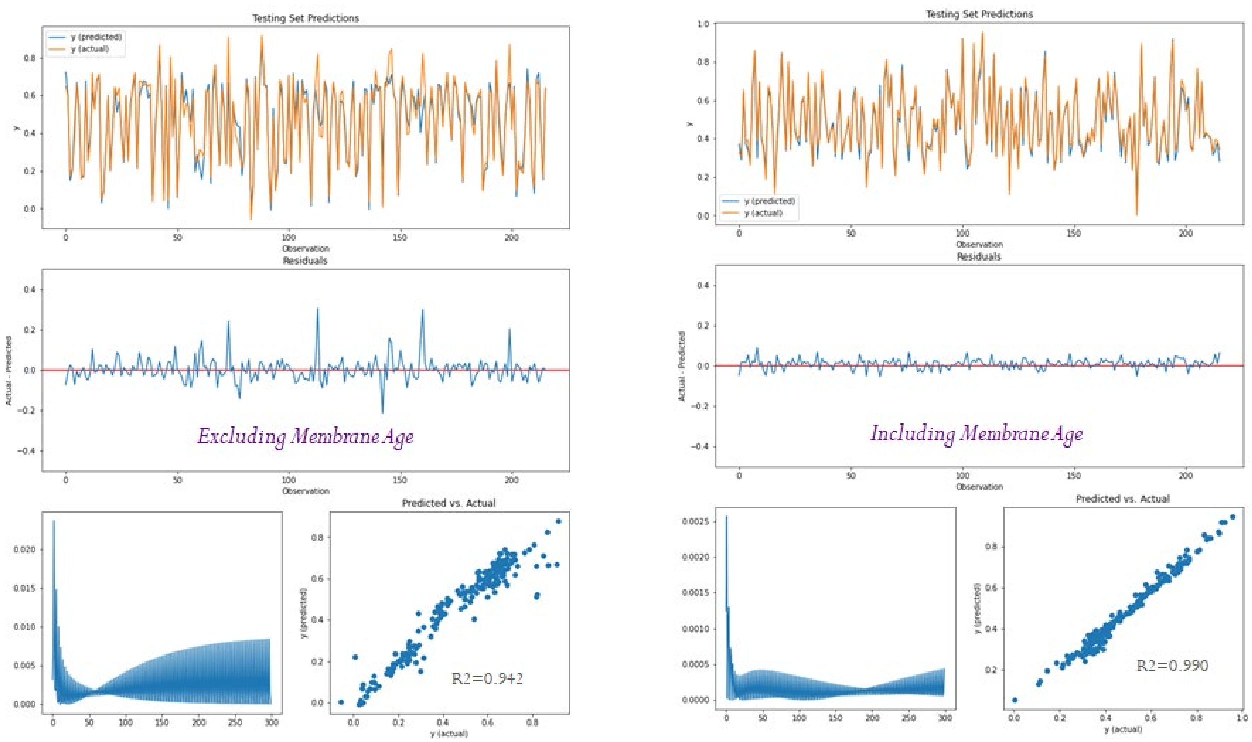

Figure 6. Third Stage Specific Flux Prediction for RO Train A01

Despite substantial changes in RO operating conditions (e.g. water temperature, third stage RO feed and concentrate RO conductivity, etc.) presented in Table 1, the accuracy of the models was great, each having an R^2 value of greater than 0.940. As membranes age, permeability of RO membranes progressively declines over the time which is a normal behavior. As expected, including membrane ages as a model input parameter has considerably improved the model prediction accuracy and resulted in better R^2 value (0.942 vs. 0.990). Since the model accuracy with the inclusion of membrane age was excellent, no attempt has been made to further refine the ANN models at that point.

Table 1. Train A01 Key Model Input Parameters and Data Range

	Average	**Minimum**	**Maximum**
Temperature, ºC	26.3	22.3	29.7
Stage 3 Feed Flow, gpm	957	751	1,050
Stage 3 Permeate Flow, gpm	346	139	438
Stage 3 Feed Electro Conductivity, µS/cm	7,232	5,744	8,750
Stage 3 Permeate Electro Conductivity, µS/cm	65.8	14.2	291

Conclusions

AI and ML have not been widely used in water business despite its huge potential. Despite changes in water quality and operational parameters, the ANN models developed in this study successfully predicted RO permeate TOC content in a pilot IPR facility as well as power consumption and specific fluxes of RO trains in a full-scale IPR facility. This study clearly showed that AI/ML is a powerful tool that can be used to support our decisions while potentially reducing errors and boosting regulatory and public confidence in potable reuse projects.

Acknowledgements

The authors sincerely thank to following OCWD staff for providing GWRS plant data and guidance:

- Mehul Patel, PE - Executive Director of Operations
- Megan Plumlee, PhD, PE – Research Director
- Han Gu, PhD - Scientist

References

https://commons.wikimedia.org/wiki/Artificial_neural_network