The application of artificial neural networks for the optimization of coagulant dosage
K. A. Griffiths and R. C. Andrews

ABSTRACT

Filtration is the final physical barrier preventing the passage of microbial pathogens into public drinking water. Proper pre-treatment via coagulation is essential for maintaining good particle removal during filtration. To improve filter performance at the Elgin Area WTP, artificial neural network (ANN) models were applied to optimize pre-filtration processes in terms of settled water turbidity and alum dosage. ANNs were successfully developed to predict future settled water turbidity based on seasonal raw water variables and chemical dosages, with correlation ($R^2$) values ranging from 0.63 to 0.79. Additionally, inverse-process ANNs were developed to predict the optimal alum dosage required to achieve desired settled water turbidity, with correlation ($R^2$) values ranging from 0.78 to 0.89.

Key words | alum dosage, artificial neural networks, optimization, prediction, water treatment

INTRODUCTION

Conventional drinking water treatment, involving coagulation, flocculation, sedimentation, and filtration, is commonly used to remove turbidity, color, and pathogens from raw water. A chemical coagulant is typically added to the water during coagulation/flocculation to destabilize particles by neutralizing their surface charge. The water is slowly mixed to promote the aggregation of smaller particles into large floculates that settle out during sedimentation. Prior to disinfection, the water is typically passed through a granular media filter, which acts as a final physical barrier preventing the passage of particle-bound pathogens (i.e. *Giardia* and *Cryptosporidium*) into drinking water.

Proper pre-treatment via coagulation is essential for maintaining good particle removal during filtration (Cleasby *et al.* 1989; Huck *et al.* 2001). Pre-treatment optimization does not necessarily imply that settled water turbidity should be minimized; however, maintaining a lower target for settled water turbidity can extend filter run-time, increase media life, and improve filter water quality (Satterfield 2006). The Partnership for Safe Waters, a voluntary program developed by the U.S. Environmental Protection Agency (EPA), American Water Works Association (AWWA) and other drinking water organizations, recommends that settled water turbidity should be kept below 2.0 NTU 95% of the time, when raw water turbidity exceeds 10 NTU (Logsdon *et al.* 2002). The objective of this study was to improve filter performance at the Elgin Area WTP (Union, Ontario, Canada) through the optimization of pre-filtration processes in terms of settled water turbidity and alum dosage. To meet this objective, artificial neural networks (ANNs) were applied to model pre-filtration processes.

Overview of the Elgin area water treatment plant

The Elgin Area WTP employs conventional treatment processes including coagulation, flocculation, sedimentation, and filtration. Raw water from Lake Erie enters the intake where sodium hypochlorite (NaOCl) is added during the Summer/Fall for zebra mussel control (Figure 1). Water is then pumped at the low lift station, where online analyzers measure raw water turbidity, conductivity, temperature, pH, and flow rate. Once the water reaches the flash mixing chambers, alum, polymer, and powdered activated carbon

(PAC) are added. PAC is added during the Summer/Fall for the removal of taste and odor compounds. The water then travels through flocculation and sedimentation; online analyzers measure settled water turbidity. Settled water passes through dual-media filtration (sand/anthracite) and is sent to a clear well for disinfection via chlorination.

Artificial neural networks

ANNs are trained using large historical data sets, typically encompassing one year. During training, the internal weights of an ANN adjust to establish the best non-linear relationship possible for a particular data set and architecture (Rodriguez & Sérodes 2004). ANNs can be developed to predict an input or output variable of a process. In a process model, the values of one or more outputs are predicted in terms of input and control variables (Baxter et al. 2001; Maier et al. 2004). This type of model can be used to optimize a process through trial and error. An inverse-process model is used to predict the value of a process input, when all other input variables are known and target values have been selected for the output (Baxter et al. 2001; Maier et al. 2004). An inverse-process model is useful for optimizing a unit process in terms of a process control variable (i.e. alum dosage).

Previous applications

ANNs have been previously applied in water treatment for modeling coagulant dosage (Gagnon et al. 1997; Zhang & Stanley 1999; Maier et al. 2004). Gagnon et al. (1997) developed an ANN capable of predicting alum dosage at the City of Sainte-Foy’s WTP (Quebec, Canada) based on four raw water variables (pH, turbidity, temperature, and conductivity); predictions were not necessarily optimal as they were based solely on previous operator behavior. To overcome this limitation, desired water quality parameters were included as inputs in subsequent studies. Zhang & Stanley (1999) developed an ANN for the Rossdale WTP (Edmonton, Alberta, Canada) capable of predicting the optimal alum dosage required to achieve a desired clarifier effluent turbidity. Maier et al. (2004) developed an ANN that could predict the optimal alum dosage in southern Australian surface waters required to achieve a desired finished water quality in terms of several parameters including filtered water turbidity, pH, color, and UV adsorbance at 254 nm (UV254). In the latter study, the ANN was trained using data collected through jar testing since operational WTP data was not available.

In the current study, two different ANNs were designed for the Elgin Area WTP to optimize pre-filtration processes. First, a Settled Water Turbidity ANN was developed to predict settled water turbidity in terms of raw water variables and chemical dosages, and second, an optimal alum dosage ANN was developed to predict the optimal dosage required to achieve a desired settled water turbidity. The optimal alum dosage ANN was comparable to the model developed by Zhang & Stanley (1999) in the sense that (i) operational WTP data was used for training the ANN and (ii) settled water turbidity was included as an input parameter. However, this study involved more rigorous data pre-processing including the use of data lagging, hourly averaging, and data transformation.

METHODOLOGY

Data collection and input selection

Historical data was collected at the Elgin Area WTP for the period of August 2006 to June 2007. Several raw water variables were selected as inputs to the settled water turbidity and optimal alum dosage ANNs including: raw water turbidity, conductivity, color, temperature, and pH. Additionally, chemical dosages (NaOCl, polymer, and PAC) and the incoming flow rate were selected as inputs. In the Settled Water Turbidity ANN, alum dosage was selected as the final input parameter and settled water turbidity was selected as the output.
parameter. Conversely, in the optimal alum dosage ANN, settled water turbidity was selected as the final input parameter and alum dosage was selected as the output parameter.

For the purpose of ANN training, data was split into three seasonal sets: (i) Fall (1st August–16th November), (ii) Winter (17th November–29th January), and (iii) Spring (1st March–2nd June). Seasonal ANNs were developed, rather than an annual model, since the full data set was not consistent throughout the entire year. Specifically, NaOCl and PAC were only applied during the Summer/Fall and the acidified alum type was changed from Clarion® A5 to A3.

**Pre-processing**

Prior to ANN development, raw data was processed using three steps: (i) data lagging, (ii) hourly averaging, and (iii) transformation. Data lagging was necessary to ensure that the true input/output relationships were learned by the ANN, as the incoming water at a current time step influences the future settled water turbidity, rather than the current. To lag the data, input readings were moved forward several time steps with respect to their corresponding settled water turbidity reading. Lag calculations for each input parameter were dependent on the hydraulic retention times (Equation (1)) between processes. Since flow rate was not constant during the period of study, three different retention times were calculated based on the approximate minimum (420 L/s), average (603 L/s), and maximum (785 L/s) flow rates through the plant.

\[
tr = \frac{V}{Q}
\]

where \(tr\) is the hydraulic retention time (s), \(V\) the internal tank or pipe volume (L), \(Q\) the plant flow rate (L/s).

Next, data lags for each parameter were calculated using the three different retention times (Equation (2)). The set of lags (minimum, average, or maximum) that yielded the best predictive performance in terms of the mean square error (MSE) and correlation coefficient (\(r\)) were determined during training. The MSE and \(r\) were calculated using Neurosolutions® version 5.07 (NeuroDimension Inc., Gainesville, FL). MSE is a measure used to quantify the amount by which an estimator (i.e. ANN output) differs from the actual value (NeuroDimension Inc. 2008). Optimal lag values which correspond to the lowest MSE and the strongest correlation are listed in Table 1. Three lags were required to properly represent the raw water turbidity in both the settled water turbidity and optimal alum dosage ANN, due to the high variability between readings with respect to flow rate fluctuations.

\[
\text{Lag} = \left( \frac{tr_{\text{output}} - tr_{\text{input}}}{f} \right)
\]

where \(\text{Lag}\) is the data lag for input parameter, \(tr_{\text{output}}\) the hydraulic retention time from intake to settled water conduit (s), \(tr_{\text{input}}\) the hydraulic retention time from intake to the point of input data collection (s), \(f\) the frequency of data collection = 10 min = 600 s.

Hourly averaged values were based on six instrument readings. Outliers were defined as points that lie outside of two standard deviations from the mean in a normally distributed data set. Since chemical dosages were varied on a daily basis, hourly values were approximated using a step function.

<table>
<thead>
<tr>
<th>ANN Input parameters</th>
<th>Data lags (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1: predicts settled water turbidity</strong></td>
<td>Turbidity(^b,c) 22, 31, 41</td>
</tr>
<tr>
<td>pH</td>
<td>22</td>
</tr>
<tr>
<td>Conductivity</td>
<td>22</td>
</tr>
<tr>
<td>Flow rate</td>
<td>19</td>
</tr>
<tr>
<td>Temperature</td>
<td>22</td>
</tr>
<tr>
<td>Color</td>
<td>22</td>
</tr>
<tr>
<td>Alum dose</td>
<td>19</td>
</tr>
<tr>
<td>Polymer dose</td>
<td>19</td>
</tr>
<tr>
<td>PAC dose(^d)</td>
<td>19</td>
</tr>
<tr>
<td>NaOCl dose(^d)</td>
<td>26</td>
</tr>
<tr>
<td><strong>Model 2: predicts optimal alum dosage</strong></td>
<td>Turbidity(^b,c) 22, 31, 41</td>
</tr>
<tr>
<td>pH</td>
<td>22</td>
</tr>
<tr>
<td>Conductivity</td>
<td>22</td>
</tr>
<tr>
<td>Flow rate</td>
<td>19</td>
</tr>
<tr>
<td>Temperature</td>
<td>22</td>
</tr>
<tr>
<td>Color</td>
<td>22</td>
</tr>
<tr>
<td>Polymer dose</td>
<td>19</td>
</tr>
<tr>
<td>PAC dose(^d)</td>
<td>19</td>
</tr>
<tr>
<td>NaOCl dose(^d)</td>
<td>26</td>
</tr>
<tr>
<td>Settled water turbidity(^c)</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) Lag = 10 min.
\(^b\) Raw water turbidity was lagged at the minimum, average, and maximum hydraulic retention times. All three lags were included as inputs due to the variability in raw water turbidity with respect to flow rate fluctuations.
\(^c\) A natural logarithm transformation was applied to the parameter to normalize its distribution prior to training.
\(^d\) Only included as an input to the Fall ANNs.
approach. Hourly readings throughout each day were assumed to remain constant until the dosage was changed at the start of the next day (12:00 am). Hourly raw water colour values were approximated using linear interpolation between daily readings.

In the final data pre-processing step, transformations were applied to the hourly averaged data sets. The raw water turbidity data exhibited a logarithmic distribution, with approximately 70% of the readings falling below 100 NTU and the remaining 30% lying between 100 and 900 NTU. Neural networks trained using normally distributed inputs have been shown to perform better when compared to those trained using non-normally distributed data (Stein 1993). As such, a natural logarithm transformation was applied to the raw water turbidity data to shift its distribution closer to Gaussian. Likewise, a natural logarithm transformation was applied to the settled water turbidity data to normalize its distribution prior to training the optimal alum dosage ANN.

**Development of ANN models**

ANNs were developed using Neurosolutions® version 5.07 (NeuroDimension Inc., Gainesville, FL). A Neurosolutions® for Excel add-on was used to randomize the processed data and divide it into three data sets for use in training, validation and testing. The majority of the data (60%) was reserved for training the ANN, while the remainder was divided equally into validation (20%) and test sets (20%). Validation data was used to assess the model during training to ensure that it was learning, rather than purely memorizing the training data. The test data was used subsequently to evaluate the model’s performance.

In terms of architecture, a multi-layer perceptron was selected for all ANNs since it was previously applied successfully for modelling coagulant dosage (Maier et al. 2004). The number of hidden layers and processing elements were determined during the training and validation process using a trial and error approach. A single hidden layer was found to yield the lowest cross-validation error upon training. The TanhAxon, a non-linear hyperbolic tangent function, was selected for the hidden layer in each ANN as it was recommended for solving regression problems (NeuroDimension Inc. 2008).

The models were calibrated using the momentum learning rule, a type of back-propagation (BPN) learning algorithm. In BPN learning, the difference between the actual and desired output is used to modify the network weights in a manner that decreases the prediction error (Nelson & Illingworth 1991). Momentum learning involves the selection of a momentum coefficient ($\mu$) and learning rate ($\gamma$). The $\mu$ increases the effective learning rate when weight change occurs consistently in the same direction. This ability allows the network to break out of a local minima that it would otherwise get caught in during the training process (NeuroDimension Inc. 2008). A large $\gamma$ will accelerate the training process by modifying the weights significantly between cycles; however, if it is too large the search will oscillate rather than converge to a minimum error (Basheer & Hajmeer 2000). The optimal values for the $\mu$ and $\gamma$ were found to be 0.7 and 0.5, respectively. These values produced the lowest cross-validation error during training.

**Training and validation**

Training and validation were run simultaneously in NeuroSolutions® to determine the optimal number of hidden neurons and training epochs. The validation set was used to test the ANN at the end of each epoch such that training was ceased when the cross-validation error began to rise. The optimal number of hidden neurons for each ANN was determined by comparing the cross-validation error for several different architectures. The Bailey & Thompson (1990) estimate (75% of the input neurons) was used to approximate the initial number of hidden neurons and several values surrounding the estimate were also investigated. For example, since there were 12 inputs in the Fall ANNs, the use of architectures containing 7–11 hidden neurons were examined.

**Performance evaluation**

Two criteria were selected for comparing the performance of seasonal ANNs: mean absolute error (MAE) and the coefficient of determination ($R^2$). MAE is determined by taking an average of the absolute errors for each testing exemplar; hence, it provides an absolute measure of how close the predicted values are to an actual outcome. The $R^2$ provides an indication of the accuracy of the model in terms of the percentage of variation that can be explained by the regression equation.
RESULTS AND DISCUSSION

Model 1: settled water turbidity ANN (Process model)

The performance of each seasonal Settled Water Turbidity ANN, in terms of MAE and $R^2$, is presented in Table 2. The Fall Settled Water Turbidity ANN had the greatest $R^2$ value (0.79) and lowest MAE (0.21 NTU), followed by the Spring ($R^2 = 0.71; \text{MAE} = 0.30 \text{ NTU}$) and Winter ($R^2 = 0.63; \text{MAE} = 0.35 \text{ NTU}$) ANNs. The performance of each model was most likely influenced by the size of the training data set, as the Fall, Winter, and Spring training sets contained 1366, 1033, and 1278 exemplars, respectively.

A correlation plot of the actual versus predicted settled water turbidity for the Fall ANN is presented in Figure 2(a) with the corresponding trend plot shown in Figure 2(b). An examination of the regression line ($y = 0.80x + 0.19$) on the correlation plot suggests that the model had a tendency to under-predict settled water turbidity for values over 1 NTU. This was likely attributable to momentary spikes in settled water turbidity (Figure 2(b)). Removal of these data points from the data set would shift the regression line closer to an ideal fit ($y = 0.85x + 0.17$); however, there was no evidence to suggest that these data points were in error, and as such they were not removed from the training/test data sets.

The results obtained suggest that the ANNs were capable of predicting settled water turbidity on a seasonal basis for the Elgin Area WTP. As such, the Settled Water Turbidity ANNs could be used as a simulation tool to train new operators on coagulant dosing, such that prior to dosage selection, the ensuing settled water turbidity for several different dosages could be forecasted and compared.

Model 2: optimal alum dosage ANN (Inverse-process model)

The performance of individual optimal alum dosage ANNs for each season, in terms of MAE and $R^2$, is presented in Table 2. The optimal alum dosage ANNs were trained

Table 2 | Results for settled water turbidity and optimal alum dosage ANNs

<table>
<thead>
<tr>
<th>Performance parameter</th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.21 NTU</td>
<td>0.35 NTU</td>
<td>0.30 NTU</td>
<td>2.7 mg/L</td>
<td>3.6 mg/L</td>
<td>3.0 mg/L</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.79</td>
<td>0.63</td>
<td>0.71</td>
<td>0.89</td>
<td>0.78</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Figure 2 | Correlation/trend plots for the Fall (2006) Settled Water Turbidity (a,b) and optimal alum dosage (c,d) ANNs.
with the same data that was used to train the Settled Water Turbidity ANNs. Once again, the Fall ANN had the best performance in terms of MAE (2.7 mg/L) and $R^2$ (0.89), followed by the Spring (MAE = 3.0 mg/L; $R^2$ = 0.82) and Winter (MAE = 3.6 mg/L; $R^2$ = 0.78) ANNs. In terms of $R^2$-values, the optimal alum dosage ANNs performed better than the Settled Water Turbidity ANNs for all three seasons. This was anticipated, as the alum dosage data sets consisted of constant daily readings, which were easier to predict than the settled water turbidity.

A correlation plot of the actual versus predicted alum dosage for the Fall ANN is presented in Figure 2(c), with the corresponding trend plot shown in Figure 2(d). An examination of the regression line ($y = 0.86x + 4.5$) on the correlation plot suggests that the Fall model had a tendency to under-predict alum dosages above 50 mg/L. However, dosages exceeding 50 mg/L occurred for only a brief period in mid-October (Figure 2(d)). If more training exemplars were available in the 50–75 mg/L range, the slope of the regression line would likely increase since the prediction accuracy above 50 mg/L would improve.

The optimal alum dosage ANNs developed during this study were comparable to the ANNs developed for coagulant optimization in previous studies in terms of MAE. Maier et al. (2004) trained an ANN using 126 exemplars containing alum dosages ranging from 2.5–100 mg/L and achieved an MAE of 3.2 mg/L upon testing. Zhang & Stanley (1999) trained an ANN using 1745 exemplars and achieved an MAE of 1.8 mg/L upon testing. The slightly lower MAE achieved by Zhang & Stanley (1999) may have been due to the larger training data set (i.e. 1745 vs. 1366 exemplars). The results obtained in this study suggest that the ANNs are capable of predicting alum dosage on a seasonal basis in terms of a desired settled water turbidity level. As such, the optimal alum dosage ANNs could be implemented as an optimization tool at the Elgin Area WTP to assist operators in selecting an appropriate alum dosage to attain a target settled water turbidity.

CONCLUSIONS

This study investigated the feasibility of applying ANNs for the optimization of the pre-filtration process in terms of settled water turbidity and alum dosage at the Elgin Area WTP. Process models were successfully developed for three separate seasons (Fall, Winter, and Spring) to predict the settled water turbidity resulting from a combination of raw water parameters and process control variables. Inverse-process models were successfully developed to predict the optimal alum dosage required to attain desired settled water turbidity levels. The size of training data sets was found to influence the performance of each ANN such that data sets containing a greater amount of training exemplars generated ANNs that performed better in terms of $R^2$-values and MAEs during testing. The ANNs are now undergoing online validation at the Elgin Area WTP.

ACKNOWLEDGEMENTS

This work was funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Chair in Drinking Water Research. The authors would like to thank Carolyn de Groot and personnel from the Elgin Area WTP for their contributions.

REFERENCES

Bailey, D. & Thompson, D. 1990 How to develop neural network applications. AI Expert 5 (6), 38–47.


Satterfield, Z. 2006 Turbidity Control. *Tech Brief,* 6 (2). The National Environmental Services Center, West Virginia University, Morgantown, West Virginia.


First received 18 November 2010; accepted in revised form 3 January 2011