APPLICATION OF NEURAL NETWORKS TO THE MODELING OF WATER TREATMENT PARTICULATE REMOVAL PROCESSES

by

Ayala Chai

A thesis in conformity with the requirements for the degree of Master of Applied Science
Graduate Department of Civil Engineering
University of Toronto

© Copyright by Ayala Chai, 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
APPLICATION OF NEURAL NETWORKS TO THE MODELING OF WATER TREATMENT PARTICULATE REMOVAL PROCESSES
Ayala Chai, M.A.Sc., 1998
Department of Civil Engineering, University of Toronto

ABSTRACT

This study examined the application of artificial neural networks (ANNs) to particulate removal processes. Using data collected from the Manheim Water Treatment Plant (WTP) and the Britannia WTP, ANNs were developed to model settled water turbidity and post-filter particle counts. Particle counts in certain size channels of the instrument may be used as a surrogate for pathogens such as Giardia and Cryptosporidium. Once modelled, the underlying relationships learned by the ANN were established by determining the sensitivity of the particulate quantity to the network input parameters, such as pH, alkalinity, temperature, coagulant dosage, coagulant aid dosage, and flow rate. These underlying relationships were then used to obtain optimal alum and polymer dosages, or filter flow rates in order to achieve a pre-selected level of post-filter particle counts. Furthermore, given the cost of the coagulant and coagulant aid, the underlying relationship may be utilized for cost minimization.
ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Dr. R.C. Andrews, professor of Civil Engineering at the University of Toronto, for his co-operation and assistance. The constructive comments provided by Dr. B. McCabe in her review has further enhanced this thesis. I would like to thank Enviromega Ltd. and NSERC Industrial Scholarship for providing financial support. In addition, Brian Pett, Doug Stendahl and Joe Pirera of the Manheim Water Treatment Plant (Waterloo, Ontario) and Ian Douglas and John van Den Oever of the Britannia Water Treatment Plant (Ottawa, Ontario) supplied and facilitated in the collection of data for this project.

Dr. G. Hinton and Dr. R. Neal are to be thanked for providing guidance with respect to neural network architecture and pre-processing.

I would also like to thank my family and friends who have provided me with endless support, patience and access to their computers. I would especially like to thank my parents for always being there and providing me with the solid foundation that has brought me to this point.
# TABLE OF CONTENTS

| LIST OF TABLES |  |
| LIST OF FIGURES |  |
| LIST OF ABBREVIATIONS |  |
| ABSTRACT |  |

## 1.0 INTRODUCTION

### 2.0 LITERATURE REVIEW

- **2.1 Neural Network Background**
  - **2.1.1 General Supervised Learning Algorithm**
  - **2.1.2 Net Architecture**
  - **2.1.3 Existing Applications of Neural Networks to Environmental Engineering**

- **2.2 Parameter Identification**
  - **2.2.1 Raw Water Quality**
  - **2.2.2 Plant Hydraulics and Flow Patterns**
  - **2.2.3 Recycle**
  - **2.2.4 Chemical Addition**
    - **2.2.4.1 Pre-oxidation**
    - **2.2.4.2 Coagulation**
  - **2.2.5 Flocculation and Rapid Mix**
  - **2.2.6 Sedimentation**
  - **2.2.7 Filtration**
    - **2.2.7.1 Influent Filter Concentration**
2.2.7.2 Influent Filter Particle Size Distribution........................33
2.2.7.3 Run Length............................................................35
2.2.7.4 Filtration Rate..........................................................38
2.2.7.5 Backwashing Efficiency.............................................41
2.2.8 Research Needs..........................................................43

3.0 NET CONSTRUCTION.........................................................44
3.1 Data Pre-Processing.........................................................44
  3.1.1 Data Inspection..........................................................47
  3.1.2 Trend Detection..........................................................48
  3.1.3 Underlying Distribution.................................................49
  3.1.4 Outlier, Error, and Correlation Removal.........................51
  3.1.5 Data Scaling..............................................................54
  3.1.6 Dimensionality Reduction.............................................56
3.2 Preliminary Net Architecture Design.................................56
  3.2.1 Alkalinity and pH.......................................................59
  3.2.2 Flow Rate.................................................................60
  3.2.3 Temperature...............................................................62
  3.2.4 Coagulant and Polymer Dosage......................................62
  3.2.5 Storage.....................................................................63
  3.2.6 Filter Influent............................................................65
  3.2.7 Filter Effluent: Particle Count......................................69
3.3 Training and Validation.....................................................71
  3.3.1 Minimization Method...................................................71
3.3.2 Hidden Layer .............................................................. 73
3.3.3 Transfer Function ....................................................... 73
3.3.4 Stopping Conditions ................................................... 73
3.3.5 Weights .................................................................. 76
3.3.6 Replicates ................................................................. 76

4.0 NETWORK ANALYSIS AND RESULTS ......................................................... 78

4.1 Settled Water Turbidity Net: Manheim WTP ....................... 79

4.1.1 Network Pre-Processing and Design: Settled Water
Turbidity Net for the Manheim WTP ....................................... 80

4.1.2 Network Results and Discussion: Settled Water Turbidity
Net for Manheim WTP ....................................................... 83

4.1.3 Application of Settled Water Turbidity Net: Manheim
WTP .................................................................................. 91

4.2 Particle Count Net: Manheim WTP ....................................... 97

4.2.1 Data Pre-Processing and Network Design: Particle Count
Net for Manheim WTP ....................................................... 99

4.2.2 Results and Discussion: Particle Count Net for Manheim
WTP .................................................................................. 105

4.2.3 Application of Results: Particle Count Net for Manheim
WTP .................................................................................. 112

4.3 Particle Count Net: Britannia WTP .................................... 117

4.3.1 Initial Filter Ripening Network: Britannia WTP ............. 118

4.3.1.1 Data Pre-Processing and Network Design ............ 118
4.3.1.2 Network Discussion and Results........................................................................122

4.3.2 Filter Ripening Curve and Breakthrough Network:
Britannia WTP..........................................................................................128

4.4 Summary.................................................................................................129

5.0 CONCLUSIONS.......................................................................................133

6.0 RECOMMENDATIONS.............................................................................134

7.0 REFERENCES..........................................................................................135

APPENDIX A Data Specification.................................................................142
APPENDIX B Data Pre-Processing...............................................................155
APPENDIX C Network Design.................................................................172
APPENDIX D Training and Validation Procedure.................................183
APPENDIX E Final Network Architectures.............................................202
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Pretreatment Parameters Affecting Particulate Removal Process</td>
<td>31</td>
</tr>
<tr>
<td>2.2 Filtration Parameters Affecting Particulate Removal Processes</td>
<td>42</td>
</tr>
<tr>
<td>3.1 Statistical Properties of Input Parameters obtained from the Manheim Water Treatment Plant to be used in the Settled Water Net</td>
<td>52</td>
</tr>
<tr>
<td>3.2 Range of Sub-Sample Periods for Manheim WTP in 1997</td>
<td>52</td>
</tr>
<tr>
<td>3.3 Summary of Possible Expressions of Storage</td>
<td>66</td>
</tr>
<tr>
<td>3.4 Advantages and Disadvantages for Various Expressions of Influent Turbidity</td>
<td>68</td>
</tr>
<tr>
<td>3.5 Advantages and Disadvantages for the Expression of the Particle Count Trend</td>
<td>70</td>
</tr>
<tr>
<td>4.1 Summary of Overall Sensitivities (%) of each Input Parameter for the Settled Water Turbidity Net – Manheim WTP</td>
<td>85</td>
</tr>
<tr>
<td>4.2 Summary of Error Statistics of the Settled Water Turbidity Net for the Manheim Water Treatment Plant</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Summary of Error Statistics for the Initial Filter Ripening Net-Britannia WTP</td>
<td>123</td>
</tr>
<tr>
<td>4.4 Summary of Overall Sensitivities for the Britannia Initial Filter Ripening Net for Particles in the 2 to 3 µm Range</td>
<td>125</td>
</tr>
<tr>
<td>4.5 Summary of Error Statistics for the Filter Ripening and Breakthrough Expert Network – Britannia WTP</td>
<td>137</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A Three Layered Feed-Forward Neural Net (adapted from Bishop, 1995)</td>
</tr>
<tr>
<td>2.2</td>
<td>Suitable Activation Functions for Regression Models (adapted from Swingler, 1996)</td>
</tr>
<tr>
<td>2.3</td>
<td>General Conceptualization of Weight Adjustment Algorithm for Network Learning (adapted from Caudill and Butler, 1996a)</td>
</tr>
<tr>
<td>2.4</td>
<td>A Time Delay Neural Network</td>
</tr>
<tr>
<td>2.5</td>
<td>A Recurrent Neural Network</td>
</tr>
<tr>
<td>2.6</td>
<td>A Network with a Time Node</td>
</tr>
<tr>
<td>2.7</td>
<td>Filter Ripening Trend (adapted from Amirtharajah, 1988)</td>
</tr>
<tr>
<td>2.8</td>
<td>Proposed Filter Attachment and Detachment Model (adapted from Ginn et al., 1992)</td>
</tr>
<tr>
<td>2.9</td>
<td>Ripening for a Given Particle Size Range Independent of Total Particle Count (adapted from Hargesheimer et al., 1992)</td>
</tr>
<tr>
<td>3.1</td>
<td>Net Development Process</td>
</tr>
<tr>
<td>3.2</td>
<td>Pre-Processing Flow Chart for Pre-Selected Input Layer Design</td>
</tr>
<tr>
<td>3.3</td>
<td>Simplified Schematic of the Plant Layout</td>
</tr>
<tr>
<td>3.4</td>
<td>Optimal Net Architecture in Terms of Number of Nodes in Hidden Layer</td>
</tr>
<tr>
<td>3.5</td>
<td>Training and Validation Error vs. Number of Epochs for “Early Stopping” Method</td>
</tr>
<tr>
<td>3.6</td>
<td>Example of Hinton Diagram</td>
</tr>
<tr>
<td>4.1</td>
<td>Settled Water Turbidity Network Architecture for the Fall and Spring Season – Manheim Water Treatment Plant</td>
</tr>
<tr>
<td>4.2</td>
<td>Actual and Predicted Settled Water Turbidity for Manheim WTP: A-Spring Season; B- Fall Season</td>
</tr>
</tbody>
</table>
4.16 Particle Count Architecture for the Initial Filter Ripening Net-Britannia WTP.. 120

4.17 Test Set Results for the Initial filter Ripening Net- Britannia WTP: A- Correlation Plot, B- Error Histogram........................................................................................................... 123

4.17 Test Set Results for the Initial filter Ripening Net- Britannia WTP: C- Cumulative Distribution Plot of Error, D- Confidence for Interval ................................................. 124

4.18 Time Trend Plot of Actual and Predicted Values of the Initial Filter Ripening for the Britannia WTP........................................................................................................... 125

4.19 Contour Plot of Post-Filter Particle Counts at a Run Time of 20 minutes With Respect to Flow Rate and Rate of Change in Flow Rate – Britannia WTP........ 127

4.20 Contour Plot of Post-Filter Particle Counts at a Run Time of 20 minutes and a Rate of Change of Flow of 0.8 ML/D*min With Respect to Flow Rate and Silica Dosage – Britannia WTP ................................................................. 127

4.21 Actual (Historical) and Predicted Filter Ripening and Breakthrough Trend for the 2 to 3 μm Channel- Britannia WTP........................................................................................................... 130
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Momentum Constant</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Learning Rate Constant</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Allowable Fraction of Test Set Error ($&lt; 1/8$)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Slope of Linear Transfer Function</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Filter Bed Porosity</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Iwasaki’s Coefficient for Modelling Particles in the Filter</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean of the Sub-sample of Data Provided</td>
</tr>
<tr>
<td>$\mu m$</td>
<td>Microns</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Scaled Number of Days in the Year</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>Standard Deviation of Sample</td>
</tr>
<tr>
<td>$\circ$</td>
<td>Degrees</td>
</tr>
<tr>
<td>$A$</td>
<td>Power Law Density Coefficient</td>
</tr>
<tr>
<td>$A_f$</td>
<td>Filter Cross-Sectional Area</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>Amp[I]</td>
<td>Amplitude</td>
</tr>
<tr>
<td>$B$</td>
<td>Power Law Slope</td>
</tr>
<tr>
<td>BP</td>
<td>Back-propagation</td>
</tr>
<tr>
<td>C</td>
<td>Turbidity in Iwasaki’s Equation</td>
</tr>
<tr>
<td>$C_H$</td>
<td>Hazen’s Coefficient</td>
</tr>
<tr>
<td>D</td>
<td>Arithmetic Mean Particle Diameter</td>
</tr>
<tr>
<td>$D_{10}$</td>
<td>Effective Size of Filter Media</td>
</tr>
</tbody>
</table>
\begin{itemize}
\item \textbf{d} = Number of Training Sets Required
\item \textbf{d}_{50} = Arithmetic Mean Particle Diameter
\item \textbf{E} = Error of the Node
\item \textbf{e} = Filter Bed Void Ratio
\item \textbf{FFT} = Fast-Fourier Transform
\item \textbf{G} = Mixing Energy
\item \textbf{GAC} = Granular Activated Carbon
\item \textbf{H} = Head
\item \textbf{\Delta H}_0 = Headloss through a Clean Bed
\item \textbf{h}_L = Headloss
\item \textbf{I} = Input Parameter Number
\item \textbf{i} = Index of Node in Layer Above the Current Layer
\item \textbf{I(t)} = Particulates Entering the Filter
\item \textbf{j} = Index of Node in Current Layer in Back-propagation Algorithm
\item \textbf{k} = Coefficient of Permeability
\item \textbf{L} = Liters
\item \textbf{Lowerbound} = Minimum Value Specified by User for Normalization of Data
\item \textbf{m} = Number of Hidden Nodes
\item \textbf{min} = Minutes
\item \textbf{mg} = Milligrams
\item \textbf{MAE} = Mean Absolute Error
\item \textbf{Max[I]} = Maximum Value for Input Parameter I
\item \textbf{Min[I]} = Minimum Value for Input I
\end{itemize}
MSE = Mean Square Error
N = Total Number of Data Sets
ΔN = Channel Particle Concentration
n = Total Number of Input Nodes
NMSE = Nominal Mean Square Error
NOM = Natural Organic Matter
NTU = Nephelometric Turbidity Units
Off[I] = Offset
O(t) = Output of Filter in Particle Counts
PAC = Powder Activated Carbon
Q = Filtration Rate
r = Linear Correlation Coefficient
RMSE = Root-Mean-Square Error
RN = Recurrent network
s = seconds
S(t) = Filter Storage
SWTR = Surface Water Treatment Rule
t = Time
tj = Target Value
TDNN = Time Delay Neural Network
TDS = Total Dissolved Solids
TOC = Total Organic Carbon
Upperbound = Maximum Value Entered by User to Normalize Data
\( w \) = number of weights

\( w_{ij} \) = Weight to Node \( i \) from Node \( j \)

WTP = Water Treatment Plant

\( x \) = Input to Node

\( y' \) = Derivative of the Activation Function

\( y_j \) = Output Value of Node

\( z \) = Standardized Data Entry
INTRODUCTION

The water industry has traditionally relied on mathematical and physical models, expert systems, and/or knowledge based systems for simulating and dealing with complex problems. An artificial neural network (ANN) may also be used to resolve the same problems without the need for an understanding of physical laws or detailed accounts of operator experience (Rodrigues and Serodes, 1994), thus making it a powerful tool when applied to the field of water treatment. ANNs are adaptive, highly parallel, non-linear, and capable of generalization (Hammerstrom, 1993) as well as being stochastic and robust against process noise or instrument bias (Boger, 1992). When given a set of historical data consisting of parameters which are known to influence the process, an ANN can be used for modeling, forecasting, plant and instrument monitoring, fault detection and diagnosis, expert rule extraction (Boger, 1992), and process control (Hammerstrom, 1993).

There are several disadvantages of applying ANN. First, ANNs have difficulty extrapolating beyond their training experience (Boger, 1992). In other words, if the ANN is not exposed to a given set of conditions during the training procedure, the ANN’s output will contain error without warning when this set of conditions occurs. Maintenance of the network is therefore required in order to update the training set to include a broader range of conditions. Secondly, the computation time increases with additional layers and connection weights, especially for complex problems (Hammerstrom, 1993; Hansen and Messier, 1991; Bishop, 1995). The increased data requirement with increasing network complexity (Baum and Haussler, 1989) contributes
to the increased computation time. Thirdly, the data requirement for an ANN approach is larger than for other methods of modeling (Boger, 1992) and may dictate project feasibility. Finally, there is no guarantee of convergence to an acceptable solution (Hansen and Messier, 1991) and the network construction methods are not completely understood (Hammerstrom, 1993). The network construction method is an art form rather than a procedure and the optimal network is achieved through a trial and error procedure. Although the objective of the ANN training is to obtain the smallest validation error possible, the magnitude of the validation error, and consequently the test set error, may be large resulting in poor model predictions. The network construction approach employed in this research may be applied to modeling any process in any field of study.

There are some cases in the literature where ANNs have been considered for application in the drinking water industry. In the water treatment industry, ANNs have been applied for the prediction of alum dosage (Baba et al., 1990; Collins and Ellis, 1992), chlorine residual prediction in storage tanks and distribution systems (Serodes and Rodrigues, 1994; Serodes and Rodrigues, 1996a), and chlorine dosage requirement modelling (Serodes and Rodrigues, 1996b). Networks have also been applied for plant operation purposes such as water demand forecasting (Guillon and Crommelynck, 1991). In the wastewater treatment industry, an ANN has been used for treatment train optimization in attempt to meet target effluent concentrations of various contaminants (Krovvidy and Wee, 1990). The application of ANNs to the drinking water industry to model particulate removal in the treatment train is investigated in this research.
A water treatment plant contains a “train” of unit processes that physically and chemically removes undesirable matter in order to produce potable water. A conventional treatment train consists of coagulation, followed by flocculation, then sedimentation, and finally filtration. The coagulation process consists of the addition of (a) chemical(s) referred to as coagulants and coagulant aids to enhance the particulates ability to settle by aiding agglomeration of the particulates in the water. The flocculation process agitates the water by mixing and therefore increases the probability of collision between particle resulting in increased agglomeration. Sedimentation allows the floc leaving the flocculation tanks to separate from the fluid through gravitational forces. The filtration process consists of passing the sedimentation tank effluent through a granular media bed for removal of the smaller particles that were not removed in the sedimentation process.

Particle removal, in terms of nephelometric turbidity units (NTU), has traditionally been used to assess filter effluent quality and treatment train efficiency. The Surface Water Treatment Rule (SWTR) (EPA, 1989), a U.S. rule concerned with filtration of surface water supplies to protect against microbial contaminants, requires that filtered water turbidity for a well designed and operated conventional treatment plant have 95% of monthly measurements less than 0.5 NTU, never have turbidity measurements exceed 5 NTU, and never exceed 1 NTU more than 5% of monthly measurements. Under the SWTR, 99.5% of *Giardia* cysts and 99% of viruses must be removed in conventional water treatment systems.

Although traditionally turbidity is used to measure particulates in the water, particle counts measure particle distribution in addition to particle concentration. It has
been debated in the literature whether the use of particle counters as a means of describing effluent quality in terms of particle concentration as well as size distribution could potentially allow for surrogate measures of *Giardia* and *Cryptosporidium* (Hargesheimer *et al.*, 1992; Li *et al.*, 1997; LeChevallier and Norton, 1992; McTigue *et al.*, 1995; Andrew, 1994; McTigue, 1995). Low turbidity waters (<1.0 NTU) cannot estimate pathogen removal sufficiently by the use of a turbidimeter in which case a particle counter may provide a better measure of effluent quality (Crozes *et al.*, 1994). There are cases that have been reported where high concentration of organics or colloidal particles have resulted in a detected change in turbidity earlier than particle count (Myers *et al.*, 1994).

Particle counting has been shown to be useful for the following purposes (Hargesheimer *et al.*, 1992; Myers *et al.*, 1994; McTigue, 1995) of which an ANN may aid in its application:

- analysis of raw water and pre-filter particulates for optimization of their removal
- optimization of chemical dosages and combinations of coagulants and coagulant aids
- measurement of the efficiency of flocculation, rapid mix, and coagulation
- assessment of plant hydraulics, filtration rate, run length, filter to waste, and recycle
- comparison and measurement of filter efficiency within the plant
- control of treatment processes to allow for a given finished water quality (as using filter ripening data to reduce spikes in effluent quality immediately prior to backwash or observing in advance when breakthrough will occur)
- serve as a possible surrogate for pathogens such as *Giardia lamblia* and *Cryptosporidium parvum*
• pilot testing of alternate technologies, filter media and treatment chemicals

The primary objective of this project is to model post-filter particle count using an ANN approach. The ANN approach to modelling post-filter particle counts was applied to two water treatment plants (WTP): the Manheim WTP located in Waterloo, Ontario and the Britannia WTP, located in Ottawa, Ontario. Once modelled, the ANN can be used as a substitute particle counter. If the ANN is put on-line and the original particle counter is not removed, any divergence of the network’s predictions from the actual measurements indicates a fault in the particle counter or a disturbance upstream in the treatment train has been detected, assuming the network is adequately maintained. The ANN may be used to generate values of post-filter particle count given the characteristics of the water and the operational properties. It will be demonstrated that the ANN predictions will enable plant personnel to optimize water treatment and minimize treatment cost. For example, the Manheim WTP particle count data sets indicate that the filter run was terminated after breakthrough occurred since turbidity measurements are used as the criteria for taking the filter off-line. The Manheim WTP data was therefore used to demonstrate minimization of particulates and pathogens in the filter effluent and the minimization of chemical costs. The Britannia WTP particle count data do not contain filter breakthrough data because run time is the limiting backwash criteria. The Britannia WTP does not, however, practice filter to waste at the beginning of the filter run. Therefore, by modelling post-filter particle counts, the initial filter ripening peak can be reduced by selecting an appropriate filtration rate at an appropriate rate of increase.
CHAPTER 2

LITERATURE REVIEW

2.1 Artificial Neural Network (ANN) Background

An ANN is a computing system that is composed of interconnected processing elements, which through a training processes using an external set of data attains a representation of the data at a state of equilibrium of minimum error (Caudill, 1987). An illustration of an ANN is provided in Figure 2.1. The circular objects represent the processing elements (otherwise known as nodes, neurons, processing units) and the interconnecting lines between them represent weights \( w_{0,1}, w_{0,2}, \ldots, w_{i,j}, \ldots, w_{n,m} \) for their respective layer. A weight is an adjustable value associated with the link between two nodes in the ANN (Caudill, 1987; Caudill and Butler, 1996a; Bishop, 1995; Swingler, 1996; Hammerstrom, 1993). The arrows pointing into the input neuron layer are the components of the input vector \( (x_1, x_2, \ldots, x_i, \ldots x_n) \) and the arrow pointing out of the ANN is the output vector. The bottom layer is a set of passive elements that read in the input parameters and is denoted as the input layer (Caudill, 1987; Caudill and Butler, 1996a; Bishop, 1995; Swingler, 1996; Hammerstrom, 1993). An extra input is referred to as the bias or threshold that functions as a reference level (Bishop, 1995). The number of identified input parameters required to model a given phenomenon is denoted by subscript \( n \). The hidden layer consists of nodes that indirectly connect the input to the output. The number of nodes in the hidden layer is denoted by the subscript \( m \). The output layer is the layer that contains the target (or desired outcome) values from the external set of data (Caudill, 1987; Caudill and Butler, 1996a; Bishop, 1995; Swingler, 1996; Hammerstrom, 1993). A training process consists of getting the output vector as
Figure 2.1  A Three Layered Feed-Forward ANN (adapted from Bishop, 1995)
close as possible to the target value by finding the global minimum error (Swingler, 1996). This process of minimizing the error until convergence results in a "black box" that has learned in a supervised manner the underlying relationships between the input parameters and the output. Considering that back-propagation algorithm is the most common method of supervised learning in industry (Hammerstrom, 1993), a description of other algorithms will not be considered.

2.1.1 General Supervised Learning Algorithm

The algorithm for basic supervised learning given by Rumelhart and McClelland (1989) is briefly summarized as follows. To initialize the ANN, the weights are randomized, and the training process begins in the forward direction by reading the input parameters into each input node (Hansen and Messier, 1991).

\[ X_j = \sum w_{ij} x_i \]  \hspace{1cm} (2-1)

The weighted sum of the inputs then enters the processing element in the layer above, referred to as the hidden layer. This summation is fed through a transfer function, usually a logistic sigmoid function for the back-propagation algorithm as illustrated in Figure 2.2. The sigmoid function has the properties of attenuated upper and lower limbs thereby constraining the weighted sum within fixed bounds and squashing the extreme input values. The output from the processing element inevitably becomes the input of processing element in the next layer above where this process is repeated. The output from the ANN is compared the target value to determine the error. The error at each output node and hidden node is calculated respectively to be

\[ E^{\text{output}} = t_j - y_j \]  \hspace{1cm} (2-2)

\[ E^{\text{hidden}} = y^* \sum w_{ij} E^{\text{output}} \]  \hspace{1cm} (2-3)
Figure 2.2  Suitable Activation Functions for Regression Models (adapted from Swingler, 1996)

\[
\text{LINEAR: } y = \gamma X \\
\text{LOGISTIC: } y = \frac{1}{1 + e^{-X}} \\
\text{TANH: } y = \tanh(\gamma X) = \frac{e^{\gamma X} - e^{-\gamma X}}{e^{\gamma X} + e^{-\gamma X}}
\]

where X is the cumulative sum of the (weights * scaled input), \( \gamma \) is the slope of the activation function (set to equal one), and y is the output of the node.

Figure 2.3  General Conceptualization of Weight Adjustment Algorithm for Network Learning (adapted from Caudill and Butler, 1996a)
The weight adjustment is accomplished with the aid of a predetermined mathematical criterion that minimizes the ANN's sum of squared errors (Hammerstrom, 1993). The accuracy of the ANN is improved as the weights reach a steady state at the minima. The process of weight adjustment, illustrated in Figure 2.3, is best visualized as a surface search for the global minimum on the error surface, the direction of which is defined by the value of the weights. Weight adjustment as described by Bishop (1995) can be accomplished by more than one algorithm. Gradient or steepest descent algorithm adjusts the weights when the user specifies the small finite learning rate (β). The small learning rate reduces the likelihood that the learning process does not skip over the minimum but is inefficient due to the large number of steps that must be taken. High dimensional spaces, or many input parameters, have relatively few local minima (McClelland and Rumelhart, 1989).

One improvement to the process is the addition of a momentum term (Caudill and Butler, 1996; Bishop, 1995; Swingler, 1996) to the gradient descent formulation so that oscillations are smoothed out. Note however that this requires the addition of a second parameter that must be selected by the programmer. Conjugate gradient method and line searches (Bishop, 1995) are other training algorithms used to find the minimum error that removes oscillations in the descent by using the previous weight vectors when determining the weight update. By smoothing out the descent in this manner, the number of epochs required to reach the minimum is greatly reduced.

There are three sets of data that is used in the trial and error process (described in Chapter 3) to arrive at the best ANN are: i) the training set, ii) the validation set, and iii) the test set. During the training process, the training data set is used for weight
adjustment. The validation data set is used during the training process after a weight update to evaluate the network's performance to an unseen data set. The network training algorithm is stopped once the minimum validation error is achieved. An evaluation of the ANN's learning abilities is conducted after the training procedure is complete. This evaluation utilizes the test set and is completely independent of the training procedure.

2.1.2 ANN Architecture

The ANN architecture and weight training processes, both of which are defined in the design process, affect the response of the ANN and its accuracy in representing the phenomenon being modeled. ANN architecture refers to the arrangement of nodes into layers and the weighted connection pattern within and between layers. An ANN can have layers that omit connections or connections that even skip layers (Hammerstrom, 1993). For time trend modeling, a time-delay neural network (TDNN) or a recurrent network (RN) is considered suitable architecture (Bishop, 1995; Swingler, 1996). Another potential architecture for accounting for trend is the use of a time node (Daniell and Wundke, 1993).

A TDNN, as illustrated in Figure 2.4, has the input parameters of previous and current time steps explicitly read into the network in a feed forward direction and are considered to be independent of each other. The previous input vectors do not have to be at a consistent time interval apart or adjacent time vectors. A feed-forward ANN allows a transfer of information only in the direction of input to output (Daniell and Wundke, 1993). The resolution of the window size, defined as the number of previous time steps required to model time trend, can be adjusted at the programmer's discretion allowing for
Figure 2.4  A Time Delay Neural Network
a smaller number of required nodes for a given window size. A trial and error process is used to define the window size dimension (Rodrigues and Serodes, 1996b).

Unlike a TDNN, a recurrent neural network allows for dependency of the current time step’s behavior on the previously entered vector. A RN contains a closed loop back to the unit and is generally used for sequence generation of time varying patterns (Caudill, 1993). Recurrent networks allow for short-term memory with the addition of a recurrent node layer. Although RN’s require less input parameters than a TDNN, the number of hidden units increases, as illustrated in Figure 2.5. The use of a time node drastically reduces the number of input nodes required, as illustrated in Figure 2.6. The disadvantage of using a time node is that the ANN is only applicable in the linear range of the time range upon which it has been trained and it does not learn attenuation of the time trend due to its lack of previous values. Hybrid networks, or networks that are a combination of two of the three previously mentioned architectures, results in an ANN that has greater benefits than one architecture alone due to the combined advantages of each of the aforementioned architecture type.

2.1.3 Existing Applications of Neural Networks to Environmental Engineering

Collins and Ellis (1992) compared predictive ANNs to regression relationships for alum dosage control. The three inputs to the ANN were the previous alum dosage, the current turbidity and current season. Considering this approach is dependent on prior values of alum dosage and does not contain other relevant parameters such as pH, a network analysis of the weights of this strict time series approach will not relate such parameters to the predicted dosage itself. Three hidden nodes were required to predict the current alum dosage. It was found that the ANN predicted slightly better than the
Figure 2.5  A Recurrent Neural Network

Figure 2.6  A Network with a Time Node
regression model. Due to the meager amount of data, both the ANN and the regression model have difficulty performing at higher dose levels, but the ANN's performance was not as effective as the regression model at higher dose levels.

Rodrigues and Serodes (1996a) modeled chlorine residual for the purpose of chlorine dosage adjustment in a storage tank. Two models were constructed, one each for the winter and summer season. The input parameters included the current flow rate, the temperature, the re-chlorination dosage, and the intake and discharge residual concentration. Considering the dynamics of chlorine evolution, a moving window is used such that trend is included in the ANN's prediction. The number of time steps in advance the programmer decides upon sets the size of the window in the TDNN. If the size of the window is three time steps, with five input parameters there will be a total of fifteen inputs to the ANN. Finding the appropriate window size is a trial and error procedure during training and was found by Rodrigues and Serodes (1996a) to be equal to the average residence time in the tank. The model was evaluated on the basis of the statistical root mean square error (RMSE). During preliminary training, it was observed that temperature was not an important criterion for the winter ANN. This can be verified on the basis that temperature data does not have a great variability in the winter. It was observed that a less favorable prediction resulted when trying to predict more than one day in advance. Rodrigues and Serodes (1996a) conclude that this ANN is easily generalized and applied to other residual chlorine simulations since routine parameters are used.

Rodrigues and Serodes (1994) uses ANNs to model post chlorine dosage and predict residual chlorine in a clear water reservoir. The post chlorine dosage ANN
mimics the operator's knowledge and provides the advantage of a less time consuming and less complicated rule than knowledge based system approaches. The eight inputs include a water temperature parameter, flow rate parameter, clear water turbidity parameter, current and previous post-chlorination dosage parameter, clear water residual chlorine parameter, residual chlorine in the distribution system parameter, and a constant term equal to one. The RMSE of the model was less than 0.1, which the authors considered adequate for the ANN to have learned the underlying relationship adequately. In order to predict residual chlorine dosages in the distribution system (which will serve the purpose of defining the chlorine dosage adjustment) a moving window was utilized. The moving window was used since the exact travel times within the pipe distribution and the reaction kinetics is unknown. Using flow rate, chlorine dosage, and residual chlorine as the input parameters, the window size increased when predicting more than one step in advance.

ANNs have also been applied to determining re-chlorination dosages in storage tanks by Rodrigues and Serodes (1996b). Since this ANN mimics the operator's experience, the data was carefully edited such that cases where the dosage was not appropriate were filtered out. The data exclusion thresholds specified by the authors are then evaluated by determining the RMSE. Editing of this type will allow for more accurate chlorine dosages and improved operation practices.

2.2 Parameter Identification

To model any phenomenon, the appropriate parameters must be identified. Parameters that affect the final effluent particle counts include water quality characteristics of the raw water (Andrew, 1994; Lawler et al., 1980; LeChevallier and
Norton, 1992; Myers, 1994) and filter influent in terms of particle concentration and particle size distribution (Clark et al., 1992; Darby and Lawler, 1990; Kavanaugh et al., 1980). Plant hydraulics and flow patterns (Logsdon, 1987) affect treatment efficiency. Pre-treatment processes such as coagulation (Cleasby, 1972; Hargesheimer et al., 1992; Hilmoe and Cleasby, 1986; Kawamura; 1976; O'Melia, 1972; McCormick and King, 1982; McTigue, 1995), pre-oxidation (Edwards et al., 1994; Hargesheimer et al., 1992; Wilczak et al.; 1992), flocculation (Lawler et al., 1980; Letterman, 1987; Logsdon et al., 1985), and sedimentation (Lawler et al., 1980; Logsdon et al., 1985; Treweek, 1979) aid in the removal of raw water particulates. The particle removal efficiency of the filtration process is dependent on parameters such as filter run length, filtration rate, and pre-treated effluent. The ANN cannot quantitatively or adequately model the raw water source, plant hydraulics, plant layout, equipment type, media type, filter configuration, backwashing efficiency, bed depth, chemical feedpoints, chemical storage time, coagulant and/or chemical type, short circuiting, particle counting instrumentation (Letterman, 1987; Logsdon, 1987) or any other qualitative, invariable parameters. The site-specific, qualitative parameters limit the ANN’s applicability solely to the plant modelled. Any changes in these intrinsic factors will render the ANN useless. The relevant parameters that can be explicitly quantified are summarized in this section.

2.2.1 Raw Water Quality

Seasonal events such as spring runoff, summer and fall algae blooms, and soil erosion control affect final effluent quality (Logsdon, 1987). For example, winter months dictate that influent turbidities will be high (Sweazy et al., 1995). Wet weather conditions may result in poor filter capabilities but dry weather will result in poor particulate
removal (Andrew, 1994). Spring or lake turnover causes an increase in finished water particle counts (Myers, 1994).

The raw water source affects final effluent water quality (Logsdon, 1987). Due to various particles' affinity towards adsorption or ion exchange, two source waters with identical turbidity, pH, and alkalinity may required different coagulant dosages (Kawamura, 1976). Both water color and particle color were found to affect the sizing of the particles and Giardia cysts as measured by the particle counter (Hargesheimer et al., 1992). River and creeks are more susceptible to pathogen contamination than lakes and reservoirs (McTigue et al., 1995) and have higher particulate concentrations (Kavanaugh et al., 1980). The level of pathogens, such as Giardia and Cryptosporidium, and in the filtered water is related to their respective levels in the raw water (LeChevalier and Norton, 1992).

Raw water temperatures (and wind) are primary factors controlling lake and reservoir source water (Tchobangolous et al., 1987). Large differences in day and night temperatures in shallow lakes or reservoir sources can lead to changes in water temperature in and out of the plant (Logsdon, 1987) except for direct filtration plants with impoundments (McCormick and King, 1982). Coagulation, pre-oxidation, or any other process that consists of a chemical reaction should be provided with 30 minutes detention time for sufficient chemical reaction completion when temperatures are low (Logsdon, 1987). Temperature changes also alter the viscosity and impede floc formation for the given reaction time and therefore a modification to the applied mixing energy may be required in order to remove a certain level of particulate matter from the water (Logsdon, 1987). Brownian motion, a removal mechanism for small particles, and sedimentation (in
the sedimentation tank or in the pores of the filter) are temperature dependent mechanisms for particle removal (Montgomery, 1985). Temperature differences result in density currents in sedimentation tanks if there are large diurnal temperature differences (Montgomery, 1985) thereby affecting tank performance.

An increased measure of raw water particle count results in an increase in the 50% tile of particulates in the filtered water and an increase in particulate removal throughout the plant (LeChevallier and Norton, 1992). Lawler et al. (1980) developed a numerical treatment plant model that predicted the effects of raw water concentration on the effluent quality. High raw water solids concentration has been observed to result in rapid aggregation in flocculators, extensive additional floc formation in sedimentation tanks, great sedimentation, long filter runs, and good effluent quality through the run under the given set of conditions specified for modeling purposes. At high concentrations, the model predicts a reduction in all sizes except for the sub-micron particles, which are removed with greater efficiency during flocculation and sedimentation than at lower raw water concentrations. When inputting a low solids concentration to the model, long filter runs but slow ripening process (thereby resulting in an increased number of particles in the filter effluent) is predicted. The smallest filter run is predicted to occur at some concentration between the pre-designated high and low solids concentrations investigated. At low concentrations, there was little change in the particle size distribution, except in the sub-micron range where particles are reduced significantly by flocculation and sedimentation.

Furthermore, Lawler et al. (1980) also observed the model's response to a varying raw water particle size distribution. The raw water particle size distribution effects the
coagulation, sedimentation and filtration unit processes as well as the performance of the entire plant. It has been predicted that the best treatment plant performance is predicted at a power law slope coefficient (defined in equation 3-6) value of 3 (from the power law slope coefficient values of 3, 4 and homogeneous that were tested in the model).

2.2.2 Plant hydraulics and flow patterns

The quantity and quality of the effluent are affected by the flow rate and flow patterns (Logsdon, 1987). As the flow rate increases, the detention time decreases causing shorter flocculation and sedimentation time and increased filtration rates leading inevitably to poorer water quality (especially for weak floc in filtration processes) (Logsdon, 1987). Unequal division of flow between parallel treatment lines may lead to overloading of a unit process. The distribution of flow within a unit process is also of importance since short-circuiting can decrease the processes’ efficiency (Logsdon, 1987).

2.2.3 Recycle

In a few WTPs, the filter backwash water used to clean the filter may be sent back, or recycled, to the plant intake. The influent to the plant must account for the addition of the recycle water flow and concentration (Cornwell and Lee, 1994). Recycling increases the cyst concentration to the plant (McTigue, 1995; LeChevallier et al., 1991). Depending on its feedpoint, pre-chlorination can eliminate the accumulation of pathogens (and algae) (Kawamura, 1976) contributed by the recycle line. Polymer addition to the backwash water may settle the cysts before they are recycled to the head of the plant (McTigue, 1995). Recycling pre-formed floc such as fresh and turbid filter wash waste, settled wash waste, or settled fresh sludge may improve flocculation and
sedimentation, alter coagulant dosage, and reduce detention times, especially for low turbidity waters (Kawamura, 1976).

2.2.4 Chemical Addition

Chemical feed capabilities, the kind of chemical feed, techniques of measuring chemical feed, and the means of dispersing the chemicals affect treatment efficiency (Logsdon, 1987). Chemicals that may be added to the treatment train include inorganic coagulants, acid, activated silica, polymers, caustic soda, lime, corrosion inhibitors, soda ash, powder activated carbon (PAC), ammonia, disinfectant, and fluoride (O'Melia, 1972; Logsdon, 1987). Not all these chemicals impact the final effluent particle count. For example, PAC is generally removed during filtration and alters the organic content in the water. Furthermore, the order of this addition is important. Caution should be taken that it reacts with the water but does not react with other chemicals added. For example, fluoride added before filtration will increase the alum required (Logsdon, 1987), residual chlorine can react with polymers thereby decreasing their effectiveness (Kawamura, 1976), and PAC may decrease the chlorine residual (Logsdon, 1987).

2.2.4.1 Pre-oxidation

Wilczak et al. (1992) observed the effects of various pre-oxidants (such as ozone, potassium permanganate, chlorine and chlorine dioxide) on the particle concentration and distribution in filter effluent as measured by a laser particle counter. Ozonation of raw water reduces the turbidity and particle concentration in filtered water and the filter head loss to a greater extent than other pre-oxidants. For turbidity removal, ozonation increases the particle removal by 1.5 log (or 95%) when compared to potassium permagnate,
chlorine, chlorine dioxide, or no pre-oxidant (Wilczak et al., 1992). The particle
distribution, however, remains the same regardless of ozonation (Wilczak et al., 1992).

Higher dosages of ozone do not have a negative effect on flocculation, and
actually increases particle removal in the filter (Wilczak et al., 1992). Pre-oxidation does
not affect filter-ripening time or head loss build-up rate (Wilczak et al., 1992). Pre-
ozonation may enhance coagulation and extend filter runs (Wilczak et al., 1992). The
ratio of metal coagulant to polymer determines whether ozone reduces coagulant demand
and head loss, and increases the concentration of natural organic matter (Edwards et al.,
1994). Ozonation decreases the head loss build up rate and the filter particle loading
when cationic polymer or the combination of cationic polymer and alum is used, but,
decreases filter particle removal when ferric chloride is used (Edwards et al., 1994). A
decrease in polymer and natural organic matter (NOM) reactions results in a reduced
chemical coagulant demand and head loss buildup due to ozonation (Edwards et al.,
1994).

In addition to ozone dosage, the raw water quality also determines the effects of
ozone on turbidity removal and enhanced coagulation. Singer and Chan (1989) found that
total organic carbon (TOC) and hardness affect particle stability. In enhanced
coagulation, the particle aggregation rate is dependent on the hardness to TOC ratio of
the raw water. Optimal coagulation with ozone occurs for waters with an ozone dosage
of 0.4 to 0.8 mg/mg and hardness (CaCO₃ as mg/L) to TOC (mg/L) ratios of >25. The
lower the ratio, the less the application of ozone can enhance the rate of aggregation.
Furthermore, a change in pH affecting particle stability is one of the mechanisms that
explain destabilization via ozonation (Wilczak et al., 1992). Ozonation alters the influent
pH and the best particle removal is found for certain initial raw water conditions at lower pH (Collins et al., 1987). However, when using ozone as a coagulant aid, an increase in pH and particle count results in an increase removal of raw water turbidity (Saunier et al., 1983).

Wilczak et al. (1992) used the measurement in the 5-12 μm range as a surrogate for pathogen removal via pre-oxidation as specified by the SWTR. Caution is required when applying these results since the calibration procedure required to find the equivalent pathogen size measured by the particle count instrumentation (Hargesheimer et al., 1992) was neglected. An additional 1 log removal of Giardia and Cryptosporidium occurred when ozone was used instead of potassium permanganate, chlorine, chlorine dioxide, or no pre-oxidant (Wilczak et al., 1992). The filtration process achieved 2.5 log (99.5%) removal of Giardia and Cryptosporidium sized cysts with pre-ozonation (Wilczak et al., 1992).

2.2.4.2 Coagulation

Coagulation is required to remove colloids (McTigue, 1995). Reducing repulsive forces of the solution thereby destabilizing the particles and aiding their agglomeration achieves coagulation. In order to destabilize the particle, the energy barrier is reduced by the addition of a chemical that will adsorb to the particle’s surface thereby neutralizing it. The ionic strength and particle charge determines the potential energy surrounding the particle. The coagulation reaction is a function of pH, temperature, ionic strength, reaction time, coagulant dosage and the characteristics of the coagulant. In direct filtration, alum dosage is more important than raw water conditions or the operational and
pretreatment parameters (Collins et al., 1987). Effective coagulation and clarification can remove 1 log (90%) of pathogens (LeChevallier et al., 1991).

One of the most important variables for aggregation is pH (Kawamura, 1976). The speciation and solubility of a coagulant is dependent on the pH of the water. Aquometallic ions are more significant at lower pH due to the increased positive charge on the metal species, enhancing destabilization by charge neutralization. There is a chance, however, of re-stabilization by overdosing at a low pH (O’Melia, 1972) and an exceptionally low pH results in ineffective coagulation. Alum coagulant has the most effective charge neutralization at a pH of 5.5 while ferric chloride is optimal at a pH of approximately 8. Removal of natural organic matter by coagulation is pH dependent and is optimal at a pH near to 5.0 for alum and 4.0 for ferric sulphate (Kawamura, 1976) and increases with increasing metal coagulant dosages (Edwards et al., 1994). Furthermore, the type of coagulant or coagulant aid may affect or be affected by the pH of the water. Hydrolyzing metal salt coagulants are acids (O’Melia, 1972). Pre-hydrolyzed forms minimize the depletion of pH and can be used in conditions of low pH (Letterman, 1987). Polymer charge is also affected by pH but not to the same extent as alum (Kawamura, 1976). Acids or alkalis such as sulfuric acid, carbon dioxide, or excess alum can correct the pH before or after coagulant addition (Kawamura, 1976).

In addition to pH, other raw water parameters such as alkalinity and particle concentration affects coagulation (O’Melia, 1972). Coagulant addition lowers the pH and if there is insufficient alkalinity the pH may drop and floc formation is inhibited. When there is insufficient alkalinity, lime or soda ash is added to the water. Water turbidities dictate which mechanism will remove particles from the water. At high particle
concentration, charge neutralization and adsorption dominate. This is especially true at low pH where hydroxometal polymers are highly charged but the colloid concentration is high enough to prevent overdosing (O'Melia, 1972). At lower concentrations, a greater coagulant concentration is required to achieve sweep flocculation and as a result, alkalinity may be added to maintain the pH (O'Melia, 1972). Bentonite, a soil substance that supplements the existing turbidity, may be added to reduce the required coagulant dosage at the expense of shorter filter runs (McCormick and King, 1982).

Furthermore, large amounts of sulfate ion or humic substances can suppress the re-stabilization zone. For natural particles in surface water, colloidal stability is dependent on the hardness (caused principally by calcium) and humic substances (or the dissolved organic carbon). The concentration of divalent ions such as Ca$^{2+}$ and Mg$^{2+}$ combined with anionic polymers increases the aggregation of negative colloids and combined with cationic polymers results in a broader range of dosage (O'Melia, 1972). Humic and fulvic acids stabilize natural particles but calcium destabilizes them (O'Melia, 1985).

According to Langelier et al. (1952), the raw water particle size has impact on coagulation (Kawamura, 1976). Particles in the 1 to 5 μm range are the foundation for dense, rapidly settleable floc (Kawamura, 1976). When alum is applied, particles of approximately 1 μm in size give a binding action for flocculent growth resulting in bulky, porous floc with a very slow settling velocity (Kawamura, 1976). If both colloidal and coarse fractions exist in the water, good quality floc will form (Kawamura, 1976).

Coagulant type affects the properties of the floc in the filter. Hilmoe and Cleasby (1986) observed that alum produces better effluent quality but the cationic polymer
increases the filter run length for both constant and declining flow controls. Furthermore, the controlling mechanism of cationic polymers in filtration processes is inter-particle bridging between the media and the particle thereby allowing for a wider variation in the filtration rate without the risk of breakthrough. Only in one case did higher filtration rate decrease the filter cycle output. Even though inter-particle bridging makes the floc more resistant to shearing during the filter run, it also makes the removal of particles during backwash more difficult (Cleasby, 1972). Low floc strength occurs for ferric and aluminum substances resulting in a limited capacity for their removal due to shearing in the filter bed. As well, higher particle compaction coefficients imply that more void space is available resulting in reduced head loss development. Alum or iron based coagulants have particle densities near unity affecting sedimentation (Montgomery, 1985). Other polymers are anionic and nonionic which are used more readily in wastewater treatment (Kawamura, 1976) or coagulant aids or filter aids (McCormick and King, 1982). Jar tests were conducted with ferric chloride (9 mg/L), ferric chloride and polymer (1 mg/L), alum (25 mg/L), and alum and polymer for one water type (LeChevallier et al., 1991). Ferric chloride removed no or few cysts or oocysts in the water even when combined with polymer. Alum was more effective, especially when combined with polymer for the removal of Giardia, Cryptosporidium, and turbidity for the water tested. Therefore, some coagulants remove turbidity but not cysts or oocysts. Furthermore, coagulants react differently for different source waters.

Possible points of addition of a coagulant or coagulant aid include at the beginning of the plant (mixed with the raw water), just before rapid mix, the rapid mix tank, just before flocculation, before filtration, after filtration as it flows into the
clearwell, or in the backwash water (Logsdon, 1987). Each possible point of addition has advantages and disadvantages in terms of treatment plant efficiency given the coagulant or coagulant aid type. For example, polymers can be added immediately prior to filtration when the attachment to the filter media is desirable (Letterman, 1987). When the filter surface is covered with polymers, however, the filter efficiency may decrease.

Jar tests determined that the optimal coagulation dosages for total particle and turbidity removal were similar (Sweazy et al., 1995). By conducting a series of jar tests, it was determined that the total particle count decreased as coagulant dose increased (Sweazy et al., 1995). If the optimal polymer or coagulant dosage is exceeded, the particle can re-stabilize. Pilot scale tests demonstrated that increased coagulant dose to enhanced feed rates decreases loading to the filter thereby increasing filter run time and increasing post-filter particle count. This finding supports Crozes et al. (1995) statement that enhanced coagulation may not achieve optimal turbidity removal.

Polymer addition to a GAC filter decreases the cumulative total particle count during the time of addition (Hargesheimer et al., 1992). After polymer addition ceases, the particle count jumps back up to a value similar to the original value. The addition of polymers results in an increased percentage of smaller particles and decreased percentage of larger particles to the original distribution at the beginning of the filter run. When polymer addition ceased, there was an immediate decrease in percentage of small particles. When using a dual media filter instead of a GAC filter, the particle distribution did not change significantly and the cumulative particle count dropped greatly with polymer addition.
2.2.5 Flocculation and Rapid Mix

Flocculation occurs after coagulation in conventional and direct filtration systems. Direct in-line filtration has no flocculation. In direct filtration, the plant is optimized such that small dense flocs are produced so that the floc is filtered out effectively and headloss is minimized (Edzwald et al., 1987; Treweek, 1979). Conventional filtration is optimized for settling and so there is a loss of ability to control head loss development in the filter (Letterman, 1987; McCormick and King, 1982). If flocculation is provided prior to filtration, shorter initial ripening periods after backwashing, lower turbidity, and lower rate of headloss development occur. However, earlier filter breakthrough is also caused (Logsdon, 1987).

The mixing energy, G, will effect the final effluent quality but is unlikely to be modified by treatment plant personnel. Temperature, mixing time (t), and mixing energy are all factors that affect rapid mix (Logsdon, 1987). Optimal aggregation for the flocculation process occurs when the optimal coagulant concentration, mixing duration, and mixing intensity is practiced. Floc shearing due to excessive energy (which is dependent on the surface shear and floc size) can break the aggregates back into the original primary particles. Jar tests determine the optimal Gt, or product of mixing energy and time (Logsdon, 1987). The optimal Gt is dependent on the coagulant type and dosage, water temperature, and raw water qualities. Argaman and Kaufman (1970) reported the existence of a minimum residence time (in which no flocculation occurs at any value of G) and that compartmentalization reduces the detention time and energy. The baffles, or walls found in the flocculation tank, allow plug flow to develop and uniformity in flocculation time. Tapered energy input, or reduced mixing energy for each
progressive flocculation chamber in the series, allows for larger particles to form and less transport energy is required.

Bearing in mind that an increase in particle concentration will increase the rate of aggregation in the flocculators, the settling velocity, and density of the floc, the efficiency and stability of particle removal are increase as the influent concentration increases (Letterman, 1987). The rate of disappearance of primary particles is a function of particle diameter according to the von Smouchowski equation. A wider particle distribution results in a greater number of collisions due to differential settling.

Lawler et al. (1980) developed a plant performance model that simulates flocculation via Brownian diffusion and fluid shear using expressions by von Smouchowski and some simplifying assumptions. The model predicted that flocculation reduces the number of sub-micron particles indicating that fluid shear can be effective for smaller particles if larger particles are present. The aggregation continues in the sedimentation tank where differential settling occurs and particle concentrations are reduced through the whole size range (as well as an increase in beta). When the model's G value is increased up to the optimal value (after which breakthrough occurs), the filtration performance is improved and the rate of aggregation is increased. More efficient sedimentation and less head loss development then ensues since the concentration applied to the filters decreases and the size applied to the filters increases. Lawler et al. (1983) compares the model's results to experimental results and found that the model overestimated the collision frequency by fluid shear.
2.2.6 Sedimentation

Sedimentation removes larger particles by gravitation. Sedimentation efficiency is dependent on particle size, density and temperature. Overflow rate, detention time, and tank depth define the site-specific design parameters of the sedimentation tank in order to achieve a certain level of particle removal. Logsdon et al. (1985) reports that in sedimentation processes, the Giardia removal is similar to turbidity removal. Lawler et al. (1980) reports through a numerical model’s response that by reducing the overflow rate to half of its original value by halving the depth but maintaining the detention time, higher suspended solids removal on a volume average diameter basis and slightly longer filter runs result. Aggregation, however, is reduced such that smaller particles are able to pass through the filter, resulting in long ripening periods and solids continually passing through the filter. A higher concentration of particles is observed for all particle sizes at the smaller overflow rate. If the overflow rate is halved by doubling the detention time, additional aggregation occurs and both settling and filter performance improves. The sedimentation removal efficiency is dependent on the influent particle size distribution (Treweek, 1979).

The parameters that affect particulate removal prior to the filters are summarized in Table 2.1.

2.2.7 Filtration

Removal of non-settleable solids in the water depends on the physical and chemical characteristics of the water such as pH, alkalinity, Ca\(^{2+}\), ionic strength, diameter of the particle (affected by coagulation), particle shape, rate of filtration, and viscosity of the fluid. Media size, bed depth, particle size and density, temperature, and filtration rate
Table 2.1  Pretreatment Parameters Affecting Particulate Removal Efficiency

<table>
<thead>
<tr>
<th>Process</th>
<th>Parameter</th>
<th>Impact on final particle count</th>
<th>Applicability to ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw water quality</td>
<td>Source</td>
<td>Different sources respond differently to pretreatment.</td>
<td>Not quantifiable</td>
</tr>
<tr>
<td></td>
<td>Color</td>
<td>Affects calibration for pathogens but is not an input parameter unless color changes with time.</td>
<td>Assumed not to vary drastically with time.</td>
</tr>
<tr>
<td></td>
<td>Algae</td>
<td>Filter clogging for direct filtration</td>
<td>Magnitude of parameter depends on filter configuration.</td>
</tr>
<tr>
<td></td>
<td>Ionic strength</td>
<td>Affects destabilization. A function of total dissolved solids or specific conductance.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>Affects chemical processes: pre-oxidation and coagulation.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td></td>
<td>Alkalinity</td>
<td>Affects chemical processes: coagulation and attachment.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td></td>
<td>Hardness</td>
<td>Affects chemical process: pre-ozoneation and enhanced coagulation. Divalent cations such as $\text{Ca}^{2+}$ and $\text{Mg}^{2+}$ also of significance since can destabilize particles.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td></td>
<td>Temperature / Season</td>
<td>Affects particle concentration and reaction rates.</td>
<td>For a large lake these two parameters are equivalent.</td>
</tr>
<tr>
<td></td>
<td>Particle concentration</td>
<td>Affects final effluent quality.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td></td>
<td>Influent flow rate</td>
<td>Flow pattern through the plant.</td>
<td>May be difficult to model intrinsically.</td>
</tr>
<tr>
<td></td>
<td>Particle distribution</td>
<td>Affects final effluent quality.</td>
<td>Depends on its variability with time.</td>
</tr>
<tr>
<td>Recycle</td>
<td>Particle concentration</td>
<td>Affects final effluent quality.</td>
<td>Depends on recycle protocol and variability with time.</td>
</tr>
<tr>
<td></td>
<td>Particle distribution</td>
<td>Affects final effluent quality.</td>
<td>Depends on recycle protocol and variability with time.</td>
</tr>
<tr>
<td></td>
<td>Recycle flow</td>
<td>Flow pattern through plant.</td>
<td>Depends on recycle protocol and variability with time.</td>
</tr>
<tr>
<td>Process</td>
<td>Parameter</td>
<td>Impact on final particle count</td>
<td>Applicability to ANN</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Pre-oxidation/ Pre-disinfection</td>
<td>Raw water quality parameters</td>
<td>TOC, hardness. pH, temperature, particle concentration</td>
<td>Previously described.</td>
</tr>
<tr>
<td></td>
<td>Reaction time</td>
<td>Function of flow rate</td>
<td>May not be required.</td>
</tr>
<tr>
<td></td>
<td>Pre-oxidant type</td>
<td>Potassium permanganate, chlorine, and chlorine dioxide have no significant impact (Wilczak et al., 1992)</td>
<td>Not quantifiable.</td>
</tr>
<tr>
<td></td>
<td>Pre-oxidant concentration</td>
<td>Affects particle concentration if ozonation used.</td>
<td>Depends on variability with time.</td>
</tr>
<tr>
<td>Coagulation</td>
<td>Raw water quality parameters</td>
<td>Alkalinity, pH, temperature, ionic strength, particle concentration and distribution</td>
<td>Previously described.</td>
</tr>
<tr>
<td></td>
<td>Reaction time</td>
<td>Possible function of flow rate.</td>
<td>May not be required.</td>
</tr>
<tr>
<td></td>
<td>Lime/Soda ash addition</td>
<td>For alkalinity adjustment.</td>
<td>Depends on variability with time.</td>
</tr>
<tr>
<td></td>
<td>Coagulant/ Polymer dosage</td>
<td>The optimal dosage can be determined by a jar test for conventional treatment.</td>
<td>Depends on variability with time.</td>
</tr>
<tr>
<td></td>
<td>Coagulant/ Polymer type</td>
<td>Metal coagulants have different properties than polymers. Each type will have a different coagulant dosage and have a different backwash removal efficiency.</td>
<td>Not quantifiable.</td>
</tr>
<tr>
<td></td>
<td>Bentonite addition</td>
<td>The addition of turbidity can reduce the coagulant required for aggregation.</td>
<td>Depends on variability with time.</td>
</tr>
<tr>
<td>Flocculation and Rapid Mix</td>
<td>Mixing time</td>
<td>Possible function of flow rate.</td>
<td>May not be required.</td>
</tr>
<tr>
<td></td>
<td>Mixing energy</td>
<td>Affects aggregation.</td>
<td>Alterable if utilizes a variable speed drive.</td>
</tr>
<tr>
<td></td>
<td>Raw water quality</td>
<td>Temperature, influent concentration and distribution</td>
<td>Previously described.</td>
</tr>
<tr>
<td>Sedimentation</td>
<td>Raw water quality</td>
<td>Particle concentration and distribution, temperature</td>
<td>Previously described.</td>
</tr>
<tr>
<td></td>
<td>Overflow rate</td>
<td>Possible function of flow rate.</td>
<td>May not be required.</td>
</tr>
<tr>
<td></td>
<td>Detention time</td>
<td>Possible function of flow rate.</td>
<td>May not be required.</td>
</tr>
<tr>
<td></td>
<td>Depth of tank</td>
<td>Fixed.</td>
<td>Not required.</td>
</tr>
</tbody>
</table>
affect the transport mechanisms. Water quality, physical characteristics of a particle, and the filter media decide the mechanism a particle will deviate from its streamline. Chemical attachment to filter media is dependent on pH, chemical dosage, alkalinity, interactions with specific ions and molecules, and ionic strength (Tobiason and O'Melia, 1988) in the same manner as coagulation of two particles and is likely to be the important factor in filtration (Tobiason and O'Melia, 1988; McCormick and King, 1982 sites Adin et al. (1979)).

2.2.7.1 Influent Filter Concentration

Clark et al. (1992) reports that higher concentrations (in terms of turbidity) have greater removal efficiency since at lower influent concentrations there is less chance of collision with a captured particle in a specified amount of time. If the filter influent concentration is high, several smaller particles can enter a pore and simultaneously block the pore, removing them from the suspension (Montgomery, 1985). The time lapse may result in the captured particle moving along the surface to the bottom of the media grain, causing a reduced chance of collision and reduced ripening and removal. As well, at higher concentrations, the lower portion of the bed initially had a higher rate of removal (Clark et al., 1992) and the filter run length was shorter (Letterman et al., 1979 sites Shea et al. (1971)).

2.2.7.2 Influent Filter Particle Size Distribution

Brownian motion, interception, and gravity forces are all dependent on the particle size (Lawler et al., 1980). As the ratio of particle size to media size increases, the removal efficiency by interception increases. The fact that larger ratios of concentration of larger to smaller particles result in increased removal efficiency was
referenced (Darby and Lawler, 1990). Straining occurs for particles greater than 100 microns (Kavanaugh et al., 1980). Minimum removal efficiency occurs at particle size ranges of 1 to 2 microns (where interception and gravitation mechanisms are dominant) for conventional water treatment according to models of a single collector (Darby and Lawler, 1990). Heterogeneous suspensions have a removal efficiency that is not dependent on particle size as predicted by the models. Clark et al. (1992) observed that although the removal efficiency increased with size at the beginning of the filter run, the increase was less extreme than that predicted by the Rajagopalan and Tien (1979) filtration model.

Darby and Lawler (1990) investigated the effects of mono-disperse, bimodal, and tri-modal distributions to determine the effects of particles of differing sizes on each other by a Coulter counter. One of the observations noted was that immediately after backwash, the behavior of the filter media was defined by the particles first captured, which was influenced by particle size. Formation and break-off changed the particle size distribution with time and depth. As well, particles of a given size ripen the bed for other particles of the same size (Clark et al., 1992).

Kavanaugh et al. (1980) mention the fact that particle size may be less important than particulate destabilization for granular media. If the particle size distribution parameter can be neglected, then perhaps turbidity is sufficient to describe influent particle concentration instead of particle count. Andrew (1994) reports that the only consistency in the literature is that water with turbidity less than 0.1 NTU does not have a turbidity and particle count that correlate. Logsdon et al. (1985) stated that similar trends exist between turbidity and particle or Giardia cyst removal although relatively small
changes in turbidity has large changes in cyst concentrations. Hargeshiemer et al. (1992) reported that turbidity and total number of particles are linearly related and have a good correlation, however, the slope of the regression line changed with filter run time as the particle distribution changed. The correlation of turbidity and filter effluent particle count is dependent on seasonal changes in source water, filter performance with time, filter media, chemical addition and instrumentation used (Hargeshiemer et al., 1992). LeChevallier and Norton (1992) observed that log removal of Giardia and Cryptosporidium may have the same proportionality to log removal of particle counts.

2.2.7.3 Run length

Removal efficiency increases for smaller particles with increased filter run time (Clark et al., 1992). However, the larger particles are removed less efficiently after a certain run length and indicate deterioration of water quality before turbidity and suspended solids measurements detect such an event. This reduction of removal efficiency can theoretically be due either to surface chemistry differences for certain size particles or floc break-off.

Adin and Rebhun (1977) noted that a "working layer" is formed in the filter bed by the coagulated floc. The working layer moved deeper into the bed as a front (completely saturated above the working layer and relatively clean below) at a rate dependent on the polymer concentration, coagulant/polymer type, media grain size, pretreatment efficiency, and the filtration rate.

Particle removal is achieved through straining, attachment to the media, or attachment to a previously deposited particle (Darby and Lawler, 1990). With increasing run time, attachment to previously deposited particles is more likely (Tobiason et al.,
1993). Considering that particles can be removed due to (an) existing particle(s) in the pores, removal efficiency increases with time. This is known as ripening. More than 90 percent of particles travel through the filter bed during the initial stages of ripening following backwash (Amirtharajah, 1988). The filter ripening curve shows effluent quality variation with filter run time and is illustrated in Figure 2.7. The attachment and detachment forces responsible for particle removal due to filtration are demonstrated in Figure 2.8. The backwash efficiency and volume of backwash remnant water can distort the two individual peaks (Amirtharajah and Wetstein, 1980). The first peak is due to backwash water remnants within and above the media and the peak is dependent on the backwashing efficiency and time. The second peak is due to particles passing through the filter due to the filter’s initial lower efficiency and is related to the filtration rate, influent particle concentration and size (Amirtharajah and Wetstein, 1980). *Giardia* passes through the filter in higher concentrations right after backwash than after the filter has started to ripen (Logsdon *et al.*, 1985) especially when turbidity reduction is less than 90% (Ongerth, 1990). Cyst and turbidity removal was improved with the use of polymer and higher concentrations of alum. Li *et al.* (1997) notes that during filtration, the *Cryptosporidium* oocysts can pass through the filter membrane of smaller pore sizes than the diameter of the pathogen itself. Small changes in turbidity cause *Giardia* cysts to pass through the filter (Logsdon *et al.*, 1985).

There is a change in the beta value with filter ripening (Hargesheimer *et al.*, 1992). The maximum value of beta occurred at the same time of the maximum total particle count and then gradually decreased from that point. This decrease is due to the fact that the filter efficiency is increased with time due to a reduction in porosity and can
Figure 2.7  Filter Ripening Trend (adapted from Amirtharajah, 1988)

- Window size dependent on slope of each section. First window includes the period of backwash.
- Plant data may not capture the increase in this window depending on the frequency of particle count measurements.
- Increase in counts due to large initial pore space that cannot capture all of the influent particulates.
- Negative slope as the pore space decreases due to ripening.
- Window e shows a steady slope.
- Window f will be differentiated from others since it does not include the last trend node as a zero value or the first filter ripening bump. This window may not occur in the data if the plant backwashed the filter prior to breakthrough.

Figure 2.8  Proposed Filter Attachment and Detachment Model (adapted from Ginn et al., 1992)

Attachment mechanisms:
1. Straining: Initial and cumulative loading, particle size
2. Sedimentation: Stokes equation
3. Interception: Q/A_d, size, inertial forces
4. Flocculation: void volume, Q/A_d, chemical attachment
5. Chemical attachment: coagulant and coagulant aid dosage, temperature, alkalinity, pH, reaction time
6. Diffusion: temperature

Detachment mechanisms:
1. Influent concentration (avalanche effects, rate of change of interstitial flow rate): Settled water turbidity changes. ΔQ/(A_d*void ratio)/Δt
2. Hydrodynamic forces (fluid drag, shear): upstream velocity
3. Deposit morphology
therefore remove smaller particulate matter. The change in particle size distribution is demonstrated in terms of percentage of the total particle count on Figure 2.9.

Slow start, filter bed conditioning with polymer for direct filtration, and filter-to-waste are methods of starting a new filter run. Slow start is a different means of re-starting the filter used after backwashing is used instead of filter to waste. In such a scenario slow start may only delay the time the particles pass through and elongate the ripening curve. This allows for dilution of this ripening from the other mature filters.

2.2.7.4 Filtration Rates

High flow velocity will carry particles deeper into the bed before they can be captured, allowing for a reduced head loss and more effective bed depth. Therefore a greater quantity of water is treated before backwashing (Clark et al., 1992). Sedimentation is dependent on the superficial velocity (Montgomery, 1985). Cleasby et al. (1963) demonstrated that any change in filtration rate will cause notable deterioration in effluent quality (Amirtharajah, 1988). Particle removal efficiencies decrease for increasing filtration velocities (Clark et al., 1992; Sweazy et al., 1995; Hargesheimer et al., 1992). With time, the filter adjusts to higher flow rates and the cumulative particle count decreases (Sweazy et al., 1995; Hargesheimer et al., 1992). A decrease in filtration rate after the filter had stabilized from the previous filtration rate increase resulted in a decrease in particle count less than the original count. Furthermore, a further decrease in filtration rate resulted in a release of particles followed by a reduction in particle count (Hargesheimer et al., 1992). The magnitude of rate changes and the rapidity of these changes affect the effluent quality unless stronger floc of alum and non-ionic polymer was used such that the floc resists breakthrough at a filtration rate increase (Logsdon,
Figure 2.9  Ripening for a Given Particle Size Range Independent of Total Particle Count (adapted from Hargesheimer et al., 1992)
There are different means of flow control through a filter. Constant rate filtration is accomplished by maintaining a constant filtration flow rate through the filter by opening the outflow valve with run time. There are three types of constant rate filtration: constant rate, constant level, and influent flow splitting. Constant level control maintains a constant water level above the filter bed by adjusting the filter effluent flow valve. The valve is throttled if the water level drops due to very high filtration rates. For uniform operation, the water level in the clearwell and the influent flow rate must remain constant. If the filtration rate surges due to a jump in influent flow or due to one of the filters taken off-line, then there will be a jump in effluent particle concentration due to flushing of particles from the bed.

Declining rate filtration can either be unrestricted, influent restricted, or effluent restricted (Hilmoe and Clearsby, 1986). The filtration rate decreases as particles accumulate in the bed and the influent is redistributed among the beds to account for the decreasing capacity of the dirtier beds. The reduction in flow rate reduces the stress on the filter thereby reducing the effluent deterioration. As the flow is redistributed among the filters, the water level rises providing the hydraulic head required maintaining a constant quantity of filtered water. When the water elevation reaches the maximum, the dirtiest filter is backwashed. Spikes occur in filters and the flow rate to the filters change when their neighbouring filters are being backwashed (McTigue, 1995). When all filters are on-line, the flow is redistributed and the water level in all filters drops. A disadvantage of declining rate filters is that although high water levels are desirable, the rate in the filter cannot be adjusted and so the plant operator must take increased flows into consideration.
Filter start up, rate changes and backwashing must be done smoothly. Rate of flow controllers should be well maintained and function smoothly or they may cause head loss surges (to which piezometers are sensitive) (Logsdon, 1987; Hilmo and Clearsby, 1986).

2.2.7.5 Backwashing Efficiency

Backwashing is effective if there is no mudball formation and the filter is not dirty for long periods of time (Amirtharajah, 1988). Air scour (throughout the filter bed) and surface wash (which causes collisions at the top of the filter bed) can aid in cleaning the bed, especially if particles with high adhesive forces are present (such as polyelectrolytes) (Amirtharajah, 1988; Cleasby, 1972). Wash rate and temperature also affect backwash efficiency (Qureshi, 1982; Cleasby, 1982). The rate the backwash valve is shut has impact on the first peak in the filter ripening curve (Amirtharajah and Wetstein, 1980). Cranston and Amirtharajah (1987) observed that alum or the coagulant used as the primary coagulant is optimal for backwash water, the time of injection can be optimized, the volume of backwash water can be optimized, as well as the coagulant dose.

Table 2.2 summarizes all parameters affecting post-filter particle counts during the filtration unit process.
### Table 2.2  Filtration Parameters Affecting Particulate Removal Efficiency

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Impact on Post-Filter Particle Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influent filter concentration</td>
<td>Dependent on pretreatment and raw water quality.</td>
</tr>
<tr>
<td>Influent filter particle size distribution</td>
<td>Dependent on pretreatment and raw water quality.</td>
</tr>
<tr>
<td>pH, ionic strength, and major divalent cations</td>
<td>Affects chemical attachment. (Tobiason &amp; O'Melia, 1988; Tobiason et al., 1993)</td>
</tr>
<tr>
<td>Pretreatment</td>
<td>Certain attachment mechanisms are dependent on pretreatment (such as coagulation, oxidation, flocculation and sedimentation).</td>
</tr>
<tr>
<td>Filter media</td>
<td>Does not vary.</td>
</tr>
<tr>
<td>Bed depth</td>
<td>Does not vary.</td>
</tr>
<tr>
<td>Filter run time</td>
<td>Filter ripening curve trend to be accounted for.</td>
</tr>
<tr>
<td>Filtration rate</td>
<td>If constant rate filtration is used then this parameter can be neglected.</td>
</tr>
<tr>
<td>Temperature</td>
<td>Sedimentation and Brownian motion mechanisms are dependent on temperature.</td>
</tr>
<tr>
<td>Backwash Efficiency</td>
<td>Applicability depends on the variability of the backwash protocol used by the plant operator after each filter run.</td>
</tr>
</tbody>
</table>
2.2.8 Research Needs

Based on an extensive literature review, existing models that are used to describe particulate removal include a numerical treatment plant model by Lawler et al. (1980) and theoretical filtration equations. The theoretical basis of these models does not account for non-ideal behaviour or the site-specific nature of this problem. By using an ANN, all of the relevant identified parameters affecting particle removal are used to model the non-ideal characteristics of a specific plant.

Various applications may be derived from a neural network model. By constructing such a model, a sensitivity analysis of the settled water turbidity and/or post-filter particle count will enable plant personnel to maximize particle removal (and minimize cost). By modelling post-filter particle count rather than post-filter turbidity, the appropriate size channels may be used as a surrogate for Giardia and Cryptosporidium. This is particularly important for the channel size that represents a surrogate size for Cryptosporidium, which may not be inactivated by chlorination. Removal of Cryptosporidium is a recent problem and research has only begun to grow in this area in the past several years. Furthermore, there is no known model in the literature that allows for the minimization of pathogens in the filter effluent. The sensitivities of the ANN's output may aid plant personnel in making WTP efficiency improvements and to adjust their operation to minimize the pathogen content in the water.

An ANN approach needs to be developed to demonstrate its ability to accurately model non-ideal, site-specific behaviour. The application of this powerful tool needs to be demonstrated in terms of cost, post-filter particle count, and pathogen minimization.
CHAPTER 3

ARTIFICIAL NEURAL NETWORK CONSTRUCTION

Particle counts, pH, total dissolved solids (TDS), mixing energy, flow rate, temperature, chemical addition dosage (such as a coagulant or pre-oxidant), alkalinity, colour, and filter run time have been identified in the previous chapter as possible input parameters to model the target parameter (or the desired model response), settled water turbidity or particle counts, using an artificial neural network (ANN). Some of the input parameters identified can be neglected due to the site-specific nature of network construction. For example, a plant that does not practice ozonation, a pre-oxidation process, will not require an ozone dosage input node. Once the input and target parameters have been identified and collected, one can proceed with the next phases of ANN construction. The phases of neural network construction include data pre-processing, network architecture design, training, validation, and test error analysis, as illustrated in Figure 3.1. Data pre-processing, network architecture design, and the training and validation procedure are described in this chapter and demonstrated in Appendix A to Appendix D.

3.1 Data Pre-processing

Data pre-processing can include trend detection, outlier removal, dimensionality reduction, correlation elimination, error elimination, and scaling as illustrated in Figure 3.2. The data pre-processing in this section will deal specifically with the data obtained from the Manheim Water Treatment Plant (WTP) for example purposes. The process of identifying the applicable input parameters required by the ANN (thereby identifying the quantity of data sets required for modelling purposes) for the Manheim WTP is
Figure 3.1  ANN Development Process

DATA SPECIFICATION

DATA COLLECTION

DATA PREPROCESSING

NET ARCHITECTURE DESIGN

ALTER ARCHITECTURE, NETWORK PARAMETERS, OR PREPROCESSING METHODS

TRAINING

ALTER HIDDEN LAYER

VALIDATION

REPLICATE

NO

YES

OPTIMAL MINIMUM ERROR ACHIEVED?

NO

YES

OUTPUT, ERROR, AND NETWORK ANALYSIS

GOOD FIT

NO

YES

FINAL NET
Figure 3.2 Data Pre-processing Flow Chart for a Pre-Selected Input Layer Design

- **Trend Detection**
- **Construct Histograms and Calculate Statistics**
- **Outlier and Error Removal**
- **Data Scaling**
- **Train Net**
- **Analyse Results**

- **Process Prior to Training**
- **Process in Parallel with Training**

- **Net Error (MSE) and $r^2$ Are Acceptable**
  - **Yes**
  - **Acceptable Architecture**
  - **No**
    - Alter Net Architecture, Transform Data, Remove Additional Outliers, or Reduce Dimensionality
3.1.1 Data Inspection

The quantity and quality of the available data sets will ultimately determine both the performance and the complexity of the ANN. Baum and Haussler (1989) have provided guidelines to determine the required quantity of data sets for training, denoted by $d$, for a simple back-propagation ANN as:

$$d > \frac{w}{\varepsilon}$$

(3-1)

where $w$ is the number of weights and $\varepsilon (<1/8)$ is the allowable fraction of error on the test set (which is double the error on the training set) for a classification ANN. For example, if a classification ANN learns to classify the training set to no more than $(\varepsilon/2)=0.05$ then the test error can approach $(\varepsilon) = 0.1$. The number of data sets required is approximately $(1/0.05) = 20$ times the number of weights ($w$) in the ANN (as determined from the Baum and Haussler equation). Knowledge of the required quantity of data sets allows for a rough estimation of the maximum number of nodes in a hidden layer. Conversely, by specifying the number of weights given a defined number of input nodes, the number of data sets can be approximated using the Baum and Haussler equation ($w \leq \varepsilon d$). These calculations are demonstrated for the data provided by the Manheim WTP in Example B.1 and B.2 of Appendix B- Data Pre-Processing.

Ideally, more than one year of data would be required for the ANN to learn annual variations. Unlike turbidity measurements, however, particle counters represent relatively new technology and lack regulatory guidelines (Veal and Riebow, 1994). The Manheim WTP had only four periods of adequate data in 1997. Each of these periods represents
one sub-sample that is equivalent to one season. Data pre-processing, output analysis, and error analysis become more labour intensive when constructing seasonal ANNs since each season is best analysed independently. Due to the discontinuous nature of sub-sampled data, the most suitable network architecture is a separate ANN for each season. If there is an insufficient amount of data to develop seasonal ANNs, a season node may be applied (as described in Section 3.1.2).

3.1.2 Trend Detection

Trend detection, using historical data versus time plots, box and whisker plots, or by applying non-parametric tests, is a necessary step for architecture design and pre-processing (Stein, 1993b). A plot of historic data with respect to time for each season will indicate whether a trend exists. Furthermore, a time trend plot is useful for the observation of outliers that result from instrumental checks and calibrations, as shown in Figure B.1 and B.2 of Appendix B, representing the raw data and the pre-processed data respectively. Time trend plots for Manheim’s pre-processed input data set are shown in Figures B.4 and B.5 for the fall and spring seasons, respectively. Since seasonal variations are observed in the data, as illustrated in Figures B.4 and B.5, two nodes for season were added. A radial view eliminates the discontinuity in time between January 1\textsuperscript{st} (day one) and December 31\textsuperscript{st} (day 365) that would appear on a linear scale. By radial view it is implied that given a unity circle (with a radius, r, equal to one), \(\cos \Theta \) and \(\sin \Theta \), where \(\Theta \) is equal to the (day in the year) \(\times(360 \text{ degrees per 365 days})\), allows for a continuous perspective of the variability due to season.
The trend plot can also give clues to the pre-processed form of the input parameters that may be required before and during training. For example, due to the discrete step-wise nature of the alkalinity data, the ANN may approximate the target values as a step-like function if alkalinity is determined to be a dominant factor. A moving average approach (given a user-defined window size) will smooth out the trend and allow the ANN to train on a continuous range of data thereby resulting in a smoother network output with time. Pre-processing of time trend data can be accomplished using a fast Fourier transform (FFT), using filters for high frequency data, or using moving averages (Stein, 1993b).

Another option is to “lag” settled water turbidity to account for the travel time from the input parameters, such as plant intake, to ANN output parameter, that being the settled water turbidimeter measurements for the settled water turbidity ANN. “Lagging” the data implies that the value of settled water turbidity account for the travel time from the intake to the sedimentation tank effluent. This is accomplished by shifting the data values by the amount corresponding to the minimum, maximum, and mean detention times in an attempt to represent the travel time of a slug input.

3.1.3 Underlying Distribution

Stein (1993a) suggests that ANNs perform best if the data are normally distributed. Non-linear transforms such as inverses, exponentials, logarithms, roots, and exponents can convert the input parameter’s underlying distribution to a normal distribution (Stein, 1993b). Neal (1998) states that only the output needs to be normally distributed, to result in quicker pre-processing. The fact that ANNs are non-parametric, and therefore more robust for non-linear and non-Gaussian data (Hansen and Messier,
suggests that raw non-normal data are suitable as the input to the ANN. If the ANN's distribution of residual error fails to be normal distributed after training, a transform can be used as suggested by Neal (1998). Application of a transform, however, does impact the ANN's final results and the sensitivity associated with the input parameters. The implication, for example, of a logarithmic transform is that there is more emphasis placed on data of smaller magnitude and less on that which is closer to the highest value (Swingler, 1996). This type of transformation is not recommended when the high values associated with an input parameter range have a larger effect on the target value than smaller values.

Moreover, the use of intermediate functions such as ratios or products of two input variables will result in smaller, faster ANNs (Stein, 1993b; Bailey and Thompson, 1990b). Alum and polymer dosage can be expressed in terms of milligrams per litre (mass/volume basis) or milligrams (mass basis). One form of expression may yield a lower test set error. It is expected that the milligrams (mass basis) form will perform better since the expression of the dosages in units of milligrams per litre has an obvious step-like nature resulting in a finite amount of values over which the ANN can generalize.

The distribution of the input and target parameters can be visually illustrated by plotting the data in histogram format. The magnitude of skewness and kurtosis determine whether a normal distribution exists (for normal distribution skewness between −0.5 and +0.5 and kurtosis between −1 and +1) and guides the appropriate outlier removal process (Stein, 1993a) as shown in Figure B.3 of Appendix B. Other statistics, such as the maximum and minimum value of each input parameter, define the ANN's range of
applicability since the ANN cannot extrapolate far beyond its experience. The statistics for the Manheim WTP are summarized in Table 3.1 for the complete data set. The sub-sample ranges are summarized in Table 3.2. The statistics summary for the Britannia ANN is summarized in Table B.1 of Appendix B.

3.1.4 Outlier, Error, and Correlation Removal

It is desirable to have a large, good quality data set. The quality of the data is dependent on the quantity and magnitude of outliers and errors as well as correlation in the historical input and target (desired outcome) data.

Outliers are identified following a review of data presented in histogram format for each of the individual input parameters. In general, it is recommended that data points which lie outside two standard deviations (95% percentile) of a normally distributed data set are to be considered as outliers (Swingler, 1996). Caution must be exercised that the general trend is not omitted as a result of the outlier removal process.

Missing data of an individual parameter cannot be accurately replaced if more than one of its adjacent values is non-existent. Therefore, the data point must be discarded. It is important to be aware of the situations that lead to missing values. For example, exceeding instrumental detection limits may result in missing values. The omission of these values will result in inaccurate, under-approximated ANN output in areas where the detection limit is exceeded. Other methods for dealing with missing data, in order of least labour intensive to most, include; i) substitution by simple interpolation (Hinton, 1997), ii) substituting with the mean, mode, or a simulated value (if one is available) (Crooks, 1992), or iii) substitution with a value produced from an auto-encoder ANN (Fausett, 1994).
Table 3.1  Statistical Properties of Input Parameters obtained from the Manheim Water Treatment Plant to be used in the Settled Water ANN

<table>
<thead>
<tr>
<th>Season</th>
<th>Property</th>
<th>pH</th>
<th>Temperature (°C)</th>
<th>Alkalinity (mg/L)</th>
<th>Raw Water Turbidity (NTU)</th>
<th>Alum (mg/L)</th>
<th>Polymer (mg/L)</th>
<th>Flow rate (L/s)</th>
<th>Settled Turbiditya (NTU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Maximum</td>
<td>8.06</td>
<td>19.3</td>
<td>170</td>
<td>11.30</td>
<td>48.42</td>
<td>0.08</td>
<td>468.5</td>
<td>3.175</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>7.56</td>
<td>14.7</td>
<td>160</td>
<td>2.42</td>
<td>39.33</td>
<td>0.06</td>
<td>281.4</td>
<td>0.455</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>7.86</td>
<td>16.6</td>
<td>165.5</td>
<td>3.74</td>
<td>44.82</td>
<td>0.07</td>
<td>412.6</td>
<td>1.478</td>
</tr>
<tr>
<td>Spring</td>
<td>Maximum</td>
<td>8.33</td>
<td>11</td>
<td>180</td>
<td>13.86</td>
<td>57.80</td>
<td>0.2</td>
<td>327.9</td>
<td>2.68</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>7.92</td>
<td>6.7</td>
<td>176</td>
<td>3.59</td>
<td>35.16</td>
<td>0.1</td>
<td>159.4</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>8.18</td>
<td>9.7</td>
<td>178.9</td>
<td>4.56</td>
<td>41.94</td>
<td>0.11</td>
<td>288.7</td>
<td>1.52</td>
</tr>
</tbody>
</table>

* Total Turbidity of the two sedimentation tanks

Table 3.2  Range of Sub-sample Periods for Manheim WTP in 1997

<table>
<thead>
<tr>
<th>Season</th>
<th>Sub-Sample Range (by date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>1/5-1/23</td>
</tr>
<tr>
<td>Spring</td>
<td>4/23-5/19</td>
</tr>
<tr>
<td>Summer</td>
<td>8/6-8/27</td>
</tr>
<tr>
<td>Fall</td>
<td>9/13-10/10</td>
</tr>
</tbody>
</table>
Any data that are known to be outliers due to instrumental checks or calibration are removed immediately upon inspection. Only data that are clearly due to external factors, which are not desirable to model, are eliminated. There must be a clear explanation as to why a data point would not be valid before it is eliminated. The outlier removal process is demonstrated for the Manheim WTP in Figures B.1 and B.2 of Appendix B.

Aside from outlier removal due to instrumental checks or re-calibration, there are other instances where data are omitted. For example, at the end of the winter sub-sample, the settled water turbidity in some instances is higher than the raw water turbidity. Mr. Stendahl (1998) of the Manheim WTP mentioned that during that the winter season a different coagulant was tested until 1/20/97. A different coagulant was used for 70% of the data sets from the winter sub-sample. These data points were therefore omitted. During the Summer months there are two instances where the flow rate dips unusually low on 8/16/97. Mr. Stendahl has indicated that these points are valid and no instrumental checks or calibration was conducted during that time. An error in flow pacing the polymer dosage is assumed to have occurred in the Summer months. Due to the inconsistent nature of the sub-sample's data, the data sets from the Summer season were discarded during the trial and error process as it became clear that the ANN could not learn unsteady flow or inconsistent practice of polymer addition.

Entropy analysis, a method of measuring mutual information between input and output parameters or between two input parameters, can be applied prior to training to detect ill-posed problems where the same input data set predicts more than one target value (Swingler, 1996). If the input data set is capable of accurately predicting the target
value, then the conditional probability of the output, given the input, is close to one. Entropy analysis can be applied to: i) measure the quality of individual parameters, ii) can indicate a parameter’s independence from other input parameters, and iii) evaluate it’s ability to predict the target value by determining their respective conditional probabilities. Neal (1998) suggests that an ANN should be able to learn without removal of errors in the data.

If the correlation removal step is eliminated there is a slower learning speed. The correlation may be removed during the scaling each of the input parameters (Bishop, 1995) or by other methods such as matrix transformation (Haykin et al., 1995). When using WTP data, some parameters are dependent on another (i.e. alum dosage is dependent on influent raw water turbidity) or are naturally correlated (i.e. temperature affects particle size distribution). Since no experimental values outside of the operationally desirable range defined by the data are available, correlation cannot be removed.

3.1.5 Data Scaling

Data scaling is the most important pre-processing step since it will allow the ANN to learn the relative importance of each parameter independent of its magnitude. The target (desired ANN output) values are to be scaled with an activation function in mind. For example, an output sigmoid activation function must have target data scaled between 0 and 1, and an output tanh function between −1 and +1 (NeuroSolutions® Manual, 1996). For regression problems, the output unit uses a linear function resulting in added flexibility of the output’s scaling range since the linear function has a range that approaches infinity. Note that the range of the output unit will affect the training,
validation, and testing error as well as the magnitude of the weights. If the range is too large, the network weights will explode due to the learning algorithm and an inefficient net will result. If the range is small, the training error will be small, but the test set error is expected to be higher.

Data must be scaled for the input parameters as well as target parameters. As long as the maximum and minimum of the scaled input parameters is within the same order for each input parameter, the value of the maximum and minimum are not of significance. Using NeuroSolutions® for Excel developed by NeuroDimension Inc. Gainesville, Florida, the ANN software utilized for training, the data are scaled using the following equations:

\[
\text{Amp}[I] = \frac{\text{UpperBound}-\text{LowerBound}}{(\text{Max}[I]-\text{Min}[I])} \quad (3-2)
\]

\[
\text{Off}[I] = \text{UpperBound} - \text{Amp}[I]*\text{Max}[I] \quad (3-3)
\]

\[
\text{Data}[I] = \text{Amp}[I]*\text{Data}[I]+\text{Off}[I] \quad (3-4)
\]

Max[I] and Min[I] are the maximum and minimum values found for a given input parameter I. Upperbound and LowerBound are the values entered by the user as the pre-specified range of the normalized data. Once the amplitude (Amp[I]) and offset (Off[I]) are calculated using equation (3-2) and (3-3) respectively, the normalized value is determined using equation (3-4). This method of scaling is called linear scaling (Swingler, 1996). Another method utilizes the standard normalization formula:

\[
z = \frac{x - \mu}{\sigma_s} \quad (3-5)
\]

where x is the parameter value, \( \mu \) is the mean of the data, and \( \sigma_s \) is the standard deviation of the data (Swingler, 1996). Standard scaling accounts for both the variations in
magnitude as well as the variations in the standard deviation (Crooks, 1992). The built in scale of the NeuroSolutions® program, which is a variation of linear scaling, is lacking with this respect, but is considered suitable for scaling nevertheless.

3.1.6 Dimensionality Reduction

Dimensionality reduction is based on a thorough literature review and operator's experience, similar to the method used by Rodrigues and Serodes (1994). Any parameters that remain constant or are measured infrequently, such as ozone dosage, hardness, total organic carbon (TOC), mixing energy and bulk underflow velocity, have been eliminated. During training, if the weights approaches a value of zero then the corresponding parameter is assumed to not contribute to the prediction of the target parameter (Bailey and Thompson, 1993a). NeuroSolutions® for Excel can be used to conduct a sensitivity analysis in order to determine the relative importance of each of the individual parameters. Other methods of dimensionality reduction are entropy based analysis, statistical testing for statistically significant variations, and principal component analysis (Swingler, 1996; Bishop, 1995).

Based on operator experience, a further reduction of the original parameters may be warranted. The operator has mentioned that pH and alkalinity will only have influence during periods spring runoff (Stendahl, 1998). During the episode of spring runoff, however, the pH and alkalinity measurements were not taken. If it is decided to remove the run-off event from the data set, the pH and alkalinity nodes may be removed.

3.2 Preliminary Network Architecture Design

Considering that the network architecture is site-specific, a clear example of the construction of an ANN must be demonstrated. The example that is being used is the
particle counter ANN for the Manheim Water Treatment Plant (WTP). When the data sets are inspected, it is necessary to determine the detention time through the plant for proper network architecture design. The backwash protocol, the length of filter runs, the plant and filter detention time, the frequency of sampling, backwash criteria, and alterations or temporary interruptions to the treatment train must be defined and identified for proper ANN construction. Other useful knowledge includes the plant personnel's experience (Rodrigues and Serodes, 1996b).

A suitable network architecture is selected and the input units are designed accordingly. For time trend modelling, a time delay neural network (TDNN) or a recurrent network (RN) is considered suitable architecture (Swingler, 1996). The addition of a time node may be an alternate scheme to model trend (Daniell and Wundke, 1993). A brief summary of these architectures and their respective advantages and disadvantages are presented in Table C.1 of Appendix C-Artificial Neural Network Design. The design proposed here is a simplified time delay neural network (TDNN) containing nodes for current values of all the input parameters and nodes for previous values of those input parameter that varies significantly within the window size, as described in Section 2.1.2. The proposed architecture allows for a smaller required training set, better ANN analysis, and better generalization for parameters that are not continuously measured when compared to a TDNN. The proposed network architecture design for the particle count ANN is demonstrated in Figure C.1.

A control volume approach is used to model post-filter particle count. This means that a volume balance of the particulates that have entered and are being stored within the filter accounts for all particulates leaving the filter. Two potential designs that account
Figure 3.3  Simplified Schematic of a Typical Conventional Treatment Plant Layout

1 Represents network architecture design #1 used for Manheim Water Treatment Plant (WTP). Separate ANNs are used for the coagulation-flocculation-sedimentation processes and the filtration process. This approach accounts for the noise due to the filtration process alone when modelling post-filter particle counts. This design appears to be better than design #2, particularly when there is a limiting amount of data available for modelling.

2 Represents network architecture design #2 used for the Britannia WTP. The ANN models the post-filter particulates as a function of the raw water parameters. This approach is larger, more complex, and requires longer training time than design #1.
for the particle counts leaving the filter using a control volume approach are presented in Figure 3.3 and are further described in Section 3.2.5.

The size of the ANN is dependent on the selected architecture and the number of nodes per layer. Given the identified input parameters, some of which remain constant throughout treatment, the revised list of applicable parameters determines the number of input units in the ANN. Since there are restrictions on the size of the hidden layer, the total number of nodes required is calculated and the data requirement is specified, as shown in Figures B.1 and B.2 of Appendix B. The final size of the hidden layer is determined during the training process. Adjustments to the input layer size which account for trends of a specific parameter depend on the magnitude of change, time frame in which change takes place, and overall influence of the parameter in the water treatment plant. The various forms of expressing the potential input parameters are further described in Section C.2 to C.5 in Appendix C- Artificial Neural Network Design.

3.2.1 Alkalinity and pH

Alkalinity is measured periodically through the month and pH is continuously measured using a Rosemount Model 1054A with an accuracy of ±0.01 pH units. The variations in pH over the Fall and Spring season are presented in Figure B.4-A and B.5-A of Appendix B. Changes in alkalinity were assumed to occur at 8:00 AM, the beginning of the work day, and remain constant until the next alkalinity measurement was taken. Other methods of expressing the periodic alkalinity measurements include moving averages. Moving averages smooth the step-like trend of the periodic alkalinity measurements and therefore allow the ANN to be exposed to potential alkalinity values occurring between measurements. It is the operator’s opinion that alkalinity and pH are
generally insignificant when determining effluent particle count due to their small variation. Alkalinity and pH may be discarded as input parameters if the ANN finds their respective impacts on settled water turbidity to be small. These parameters, however, are included in the model since, theoretically, the pH and alkalinity change during periods of snowmelt run-off thereby improving the coagulation, even though these changes were not captured in the data. Furthermore, the raw water pH and alkalinity data have minute variations over the plant residence time and therefore may require fewer nodes than other parameters when representing the time trend.

3.2.2 Flow rate

Flow rate (L/s) is measured using a Fischer Porter Model flow meter, series 10D 1465 copa -X, which has an accuracy of 1. The plant flow was varied seasonally, as illustrated in Figures B.4-F and B.5-F in Appendix B. The influent flow rate was calculated to be the sum of the filtration rates since no plant influent rate measurements were available. It has been assumed that there is negligible accumulation or depletion of water stored in the plant. If this assumption holds true, the filtration rate and influent flow rate are considered to be directly proportional at any point in time and any change in filtration flow rate will result in an approximately instantaneous change throughout the plant. This assumption, however, does not hold true during the Summer months when the plant flow rate was dropped close to zero flow at certain times during this season in order for the plant personnel conduct tests. The drop in flow rate causes the water to be stored and accumulated in the treatment train until the flow recommences and reaches steady state again. The filter flow configuration changed during either the month of February or August, from a variable flow rate (such that when one of the filters are backwashed, the
other filter flow compensate for the filter out of service) to a fixed flow rate configuration.

The flow rate through the filters during their run is maintained at relatively constant rate. Changes in filtration rate are conducted in 20 L/s gradations and can be slowly altered over a time period of several hours such that there are no spikes of post-filter particulates. The particle count prediction ANN would only require filtration rate as an input since both the filtration rate and plant flow rate are proportional and the plant flow rate only affects the reaction time of the chemicals added, most of which react prior to the filter. Since flow rate and velocity are directly proportional (by a factor of the cross-sectional area of the filter), the ANN considers flow rate and velocity to be interchangeable.

The form of the flow rate parameter depends on the dominant attachment and/or detachment mechanisms during the filter run, as shown by Figure 2.8. For example, velocity describes the reaction time for the chemical attachment mechanisms, interstitial velocity influence detachment mechanisms, and the inverse of the approach velocity impacts the hydrodynamic forces. In addition, a sudden dramatic increase in velocity can cause a spike in the filter effluent particle count. The change in velocity parameter can either be expressed by one node representing the velocity difference over the specified time interval, or two nodes, one of which is the current velocity and the other is a previous value of velocity taken at a time proportional to the filter detention time. The calculation for determining the pre-specified time interval that is proportional to the filter detention time, otherwise referred to as a “lag”, is described in Section C.1 of Appendix C. The advantage of using the latter scheme is that the ANN may learn that the prior
flow rate is more important and interpolate between the two nodes accordingly. Another means of expressing filter flow rate, which may be useful for some filter configurations, is as a percentage of the plant flow rate.

3.2.3 Temperature

Temperature is a continuously measured seasonal parameter, as illustrated in Figures B.4-B and B.5-B of Appendix B. Temperature measurements occur at the same location that the raw water turbidity measurements are taken. In the operator's opinion, the diurnal nature of the water's temperature is negligible and the temperature throughout the day can be safely assumed to be constant. Temperature is an important parameter since it affects the chemical reaction time in the coagulation-flocculation-sedimentation processes (which carries over into the filter). Furthermore, water viscosity and density are dependent on temperature, which may affect sedimentation, Brownian motion, and hydraulic permeability. The explicit form of the viscosity and density may aid the ANN in modelling the particle count trend.

3.2.4 Coagulant and Polymer dosage

Alum and LT274 are used as the coagulant and polymer produced by General Chemical and Allied Colloid respectively at the Manheim WTP. It was assumed that any change in dosage occurred at 8:00 AM, which is the approximate start of the workday. Coagulant and polymer dosages (mg/L) are varied at most on a weekly basis and therefore can be assumed to remain constant over a given filter run time. The various alum and polymer dosages applied is presented in Figure B.4-D,E and B.5-D,E for the Fall and Spring seasons respectively. Dosages are dependent on the flow rate, except for an unidentifiable episode when the polymer was not flow paced that has been assumed to
occur during the Summer season. The coagulant and polymer dosage nodes may not be required to model post-filter particle count since they can be assumed to have completely reacted prior to filtration. It has been approximated by the operator that 10-15% of the coagulant and polymer carries over into the filtration process. If alum alone is used as a coagulant, the pore flocculation mechanism can be ignored due to the low resistance of an alum floc to shearing forces in the filter bed, resulting in further justification for the omission of this node.

3.2.5 Storage

In order to account for filter ripening, the ANN must be able to recognize when the filter has reached its ultimate capacity under a given set of conditions. The storage nodes account for the filter-ripening trend through the continuity equation

\[ O(t) = \Sigma I(t) - \Sigma O(t-1) - S(t) \]  

where \( O(t) \) represents the filter output (otherwise known as post-filter particle count), \( I(t) \) is the filter input (using either turbidity or particle counts), and \( S(t) \) represents the storage of the filter (otherwise known as the deposit in the filter. Another way to express the continuity equation is in the rate form of equation (3-6). This kind of scheme, however, results in poorer model predictions since this form is more sensitive to instrumental noise in the data representing the change in storage or the settled water turbidity than storage itself and the cumulative sum of settled water turbidity. In order to be conservative, the change in storage (a term in the rate form of the continuity equation) was included in the ANN by adding an extra node consisting of the storage calculated in the previous time step.
There are numerous ways in which to express storage. Headloss may be used to express storage since it is proportional to the surface area of the particles (Darby and Lawler, 1990) and therefore may adequately describe the quantity of solids in the pores. Another means of expressing storage is the specific deposit, or the mass of solids per unit volume of the filter, approximated as (Ives, 1982)

\[
\text{Specific deposit} \propto (1 - \Delta H / \Delta H_0)
\]  

(3-7)

where \( \Delta H_0 \) represents the head loss through a clean filter bed. This equation, however, is only valid when the volume of the solids take up less than 10% of the pore volume (Montgomery, 1985).

Other methods of describing the quantity of deposits stored within the filter originate from a form of describing the pore space. Porosity is defined as the ratio of the void volume to the solids volume; void ratio is defined as the ratio of the volume of voids to the total volume. If the initial porosity or void ratio is known by the plant or the manufacturer of the filter media is known, a clean bed estimate of the porosity or void ratio will allow for the measurement of their change. The changes in these respective measurements are related to the amount of deposits in the filter bed. Furthermore, the hydraulic conductivity has various forms of proportionality to porosity (Whitlow, 1990), one of which is

\[
k \propto e^3
\]  

(3-8)

where \( k \) denotes the hydraulic permeability and \( e \) is the void ratio. The coefficient of permeability is calculated using the formula

\[
k = Q / h_L.
\]  

(3-9)
The inverse of the permeability is proportional to the volume of the solids per volume of the voids. The use of the inverse is considered a highly powerful form of the storage parameter since it is a measure of the volume of solids present, and proportional to the inverse of the volume of the voids, which accounts for interstitial velocity. If the effective size and uniformity coefficient, and/or the permeability are known, the amount of solids per unit volume \( = \frac{1}{1+e} \) can be quantified using the relationship

\[
k = C_H D_{10}^2
\]  

(3-10)

where \( D_{10} \) is the effective size of the filter media, and \( C_H \) is a experimental constant dependent on the nature of the media. The \( C_H \) constant, otherwise known as the Hazen coefficient, is determined by the media's effective size and uniformity coefficient (Whitlow, 1990). These calculations are further described in Appendix C.5. A brief summary of the forms of expressing storage is summarized in Table 3.3.

3.2.6 Filter or Plant Influent

Turbidity is measured continuously using a Hach Model 1720C low range turbidimeter with an accuracy of 2% for the 0-30 NTU range. The trend of the raw water turbidity for the Manheim WTP is presented in Figures B.4-C and B.5-C for the Fall and Spring season, respectively, and is measured prior to rapid mix. There are limitations of using turbidity as opposed to particle count when expressing the influent particulate matter but an estimate is still required. These limitations are due to the fact that turbidity is a one-dimensional parameter and particle count is a two dimensional parameter. If the particle size distribution varies, turbidity and particle count cannot be correlated (Hargesheimer et al., 1992). It is therefore advisable to develop seasonal ANNs along with separate ANNs for events such as spring runoff to ensure that extreme conditions
<table>
<thead>
<tr>
<th>Expression</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headloss</td>
<td>Headloss is proportional to the surface area of the particles (Darby and Lawler, 1990).</td>
</tr>
<tr>
<td>Coefficient of Permeability</td>
<td>Hydraulic permeability is correlated to porosity (Whitlow, 1990) and can be easily calculated with the data collected.</td>
</tr>
<tr>
<td>Void Ratio (as the square root of permeability)</td>
<td>Related to the volume of solids present. Void Ratio = volume of voids per volume of solids (Whitlow, 1990)</td>
</tr>
<tr>
<td>Inverse of Permeability</td>
<td>Correlated to inverse of void ratio. No transform to void ratio required</td>
</tr>
<tr>
<td>Inverse of Void Ratio</td>
<td>Directly related to the volume of solids. The volume of voids is in the denominator, which may aid in modelling the interstitial velocity.</td>
</tr>
<tr>
<td>Porosity</td>
<td>Whitlow, 1990 (see Appendix C for calculations)</td>
</tr>
</tbody>
</table>
may be modelled to a higher degree of accuracy. The window dimension of filter influent must be large enough to include all process detention times as well as the window size required for the particle count trend. A way to reduce the size of the window (and therefore the number of nodes required) when using raw water turbidity, is to calculate the pre-filter plant detention time and include it implicitly in a lag design (when applying Design #2) as done in Section C.1 of Appendix C, or use settled water turbidity (Design #1). A summary of the advantages and disadvantages of using various forms to represent the loading onto the filter are summarized in Table 3.4. Settled water turbidity was used to describe filter influent for the Manheim WTP.

Turbidity units may be expressed as mg/L if the readings are less than 100 turbidity units (Nielsen et al., 1973). The form of the filter influent loading depends on the form of the continuity equation applied. It has been previously established that the form described in equation (3-6) is most suitable. Since turbidity units can be converted to mg/L, the cumulative input, $\Sigma I(t)$, is simply the sum of the product of the settled water turbidity and the filtration rate from the beginning of the filter run. This assumes constant particle density, which may not hold true during episodes of spring runoff. Other parameters of importance that may be required as inputs to the ANN include the initial loading and the change in loading. Sharp increases in particle concentration can potentially lead to spikes in the filter effluent particle count. Various forms of expressing the change of influent loading include a node of the difference between the current and immediately prior settled water turbidities, or using two nodes one of which is the current turbidity loading and the other the loading prior to that time step. The advantages and disadvantages of using these different forms of the change in influent loading are the
Table 3.4 Advantages and Disadvantages for Various Expressions of Influent Turbidity

<table>
<thead>
<tr>
<th>Influent Turbidity</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Water Turbidity or</td>
<td>• Will allow direct relationship between raw water parameter and the final</td>
<td>• There is an accumulation of noise as it travels through the pre-filter unit</td>
</tr>
<tr>
<td>Particle Count</td>
<td>effluent particle count</td>
<td>processes</td>
</tr>
<tr>
<td></td>
<td>• ANN potentially learns the reaction of the raw water turbidity or particle</td>
<td>• The raw water pre-filter detention time is dependent on the flow rate. The</td>
</tr>
<tr>
<td></td>
<td>count with the chemicals added prior to the filter</td>
<td>ANN is assumed to be able to calculate despite the varying raw water</td>
</tr>
<tr>
<td></td>
<td></td>
<td>particulate concentration at varying flow rates</td>
</tr>
<tr>
<td>Settled Water Turbidity or</td>
<td>• Measure of solids applied directly onto the filter</td>
<td>• Large number of nodes potentially required over a larger time interval</td>
</tr>
<tr>
<td>Particle Count</td>
<td>• Does not include the pre-filter process, therefore the accumulation of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>noise in these pre-processes are eliminated</td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>• Accounts for both raw and the settled water turbidity</td>
<td>• Two separate ANNs are required to illustrate the effects of raw water</td>
</tr>
<tr>
<td></td>
<td>• The reaction with alum and polymer can be added since there may be spill</td>
<td>turbidity or particle count on the final effluent particle count</td>
</tr>
<tr>
<td></td>
<td>over onto the filters</td>
<td>• May be required to assume that all coagulant and polymer have already</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reacted prior to entry into the filter</td>
</tr>
</tbody>
</table>


same as using different forms of the change in flow rate.

3.2.7 Filter effluent: particle count

In order to reduce the number of nodes in the ANN for reduced network complexity, run time, and data quantity requirements, it is recognized that the particle count data for all channels can be expressed by the power law relationship (Kavanaugh et al., 1980)

$$\log(\Delta N/\Delta d) = \log(A) - B \log(d_{50})$$

(3-11)

where $\Delta N$ is the channel particle concentration, $d_{50}$ is the arithmetic mean particle diameter, $A$ is the power law density coefficient, and $B$ is the power law slope coefficient. The calculation of power law slope requires that the counts in each channel investigated be greater than 10 counts/mL (Hargesheimer et al., 1992). $\log A$ (which is related to the concentration of particles in suspension) and $B$ (which is related to particle size distribution) can therefore be used to represent the particle count data. The Manheim water treatment plant has an 8-channel particle counter, but only 4 of the channel’s data representing the 1 - 2 $\mu$m, 2 - 5 $\mu$m, 5 - 10 $\mu$m, and 10 to 15 $\mu$m particle sizes were available from the SCADA. These channels represent the range where the majority of the post-filter particle counts are found and represent the channels that may be used as a surrogate for Giardia and Cryptosporidium. There are various methods of expressing filter effluent particle count, both as an input to account for trends, such as using the cumulative total particle count from the prior time step or prior values of particle count in trend ($O(t-1)$), $\Sigma O(t-1)$, and as the target values as summarized by Table 3.5.
### Table 3.5 Advantages and Disadvantages for the Expression of the Particle Count Trend

<table>
<thead>
<tr>
<th>Particle count trend expression</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>B and log(A)</td>
<td>• Can represent the complete distribution with two parameters</td>
<td>• Added pre-processing required</td>
</tr>
<tr>
<td></td>
<td>• There are two parameters that could potentially aid the ANN to predict backwash</td>
<td>• Labour intensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requires an experienced operator to interpret the results</td>
</tr>
<tr>
<td>Individual ANN per particle size channel</td>
<td>• May more accurately be able to represent trend, especially when predicting in micron units</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Can be read directly off of SCADA</td>
<td></td>
</tr>
<tr>
<td>All individual size channels</td>
<td>• May more accurately be able to represent trend, especially when predicting in % micron channel size per total particle count</td>
<td>• Accuracy of each individual particle channel size may be compromised</td>
</tr>
<tr>
<td>Effluent turbidimeter (not recommended unless combined with one of the above expressions)</td>
<td>• Can possibly correlate the particle count with the use of one node</td>
<td>• Particle count cannot be correlated to turbidity unless the particle size distribution remains constant</td>
</tr>
<tr>
<td>Natural logarithm transformation</td>
<td>• Limits the range of the data so that it may train better</td>
<td>• When transforming the data back, the error at the higher range of the set will be greater than the rest of the set.</td>
</tr>
<tr>
<td>Number of particles in channel / total particle count (%)</td>
<td>• May train to a lower error due to reduced range of data</td>
<td>• Requires an addition ANN to model total particle count</td>
</tr>
<tr>
<td>Volume basis</td>
<td>• May aid the ANN to comprehend the mass balance occurring</td>
<td>• Must account for this additional ANN’s noise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Added pre-processing required</td>
</tr>
</tbody>
</table>

*a* These expressions can be utilized in combination

*b* Expression utilized may depend on the form of the continuity equation
3.3 Training and Validation

Prior to training, the pre-processed data are split into three sets: i) a training set, ii) a validation set, iii) and a test set. Validation and test sets are required to ensure the ANN is capable of generalization beyond the data used for training. Validation occurs during the training process and allows the programmer to identify the minimum validation error and consequently stop ANN training in a method referred to as "early stopping" (Bishop, 1995). The recommended size of the validation set is at least 20% of the training set (Swingler, 1996). Testing occurs after the training process is complete and is used for ANN analysis purposes as described in Chapter 4. Using the software package Neurosolutions® for Excel (NeuroDimensions, 1997), all the data were randomised and then the training, validation, and test set sizes were defined. Randomisation of the inputs is required before training to ensure that the weight adjustments allow for generalization. ANN parameters defined within the ANN design step include the minimisation method, number of hidden layers, number of hidden units per layer, transfer function used, noise addition or regularisation, and the number of replicates selected during the training and validation steps in ANN construction.

3.3.1 Minimization method

In order to achieve a minimum validation error, a search of the error surface is conducted using a pre-selected minimisation method (as illustrated in Figure 2.3). The weights are adjusted in accordance with the chosen minimisation method. Steepest descent, momentum descent, conjugate gradient descent (combined with line searches), quickprop method, and delta-bar-delta method are some of the minimisation methods that may be applied (Bishop, 1995; Swingler, 1996; Caudill and Butler, 1996; Caudill, 1987).
Momentum descent is the minimization method that was selected. The advantages of momentum descent over steepest descent are that the added momentum term allows for faster convergence that aids in situations where the error surface forms a narrow ravine (Bishop, 1995; Caudill and Butler, 1996). Although the conjugate gradient method when combined with line searches allows for even faster convergence and greater accuracy, the ANN may get trapped in local minima (Bishop, 1995). The quickprop method tends to be a little coarse in finding the minima (Bishop, 1995). The delta-bar-delta method has four parameters that need to be specified, which is more labor intensive (Bishop, 1995).

The momentum method utilises two parameters, the learning rate ($\beta$) and the momentum constant ($\alpha$). These parameters are altered during the training process. The more complex the problem, the smaller the learning rate required (Caudill and Butler, 1996). As the weights approach a local minimum, the learning rate is typically reduced to a value of 0.1 and the momentum term is approximately equal to 0.9 (Swingler, 1996). Bishop (1995) defines the weight update for the momentum descent method as:

$$\Delta w_{ij} = \beta E x_i + \alpha \Delta w_{ij}^{prev}$$

(3-12)

where $\Delta w_{ij}$ is the weight change, $E$ is the error of the node and $x_i$ is the input to the node along the connection that is being updated. The momentum term allows the previous weight update to contribute to the current update. This is analogous to a ball gaining speed as it rolls down a hill such that a weight vector (or ball) has enough "speed" and does not get stuck in a local minimum.
3.3.2 Hidden layer

The ANN should not contain more than two hidden layers since the back-propagation of the gradients will lose their significance with each additional hidden layer (Caudill, 1991). When starting training, it is advisable to begin with one layer (Bailey and Thompson, 1990a; Swingler, 1996). Two hidden layers are preferred over one for regression problems, but a single layer has been shown to be adequate (Swingler, 1996; Bishop, 1995).

During the training process, the number of hidden units is modified and the architecture with the minimum validation error is selected. A good starting size for the hidden layer is approximately 75% of the number of nodes in the input layer (Bailey and Thompson, 1990a).

3.3.3 Transfer Function

The sigmoid function is the transfer function of choice for back-propagation (Kempa, 1993). The tanh function is an alternative to the logistic function. This function ranges the output between +1 and -1 while the sigmoid logistic function ranges between 0 and 1, as illustrated in Figure 2.2. The tanh function, therefore, elongates the error surface causing steeper gradients and faster convergence at the expense of computer storage space (Swingler, 1996). For regression problems, and a sigmoid transfer function is applied to the hidden layer(s), and a linear transfer function for the output node is most suitable since it allows for an unbound range of output values (NeuroDimensions, 1997).

3.3.4 Stopping Conditions

When a large external data set is available (as defined by equation 3-1 for a pre-specified error limit), the ANN can over-fit, or memorise, the data. In order to assure
generalisation, the ANN is stopped and the weights are saved when the validation error reaches a minimum. The termination of training before the validation error increases due to memorisation is referred to as "early stopping". The optimal network architecture in terms of hidden units is shown in Figure 3.4. The concept of early stopping is demonstrated in Figure 3.5.

The average cost function utilized is the mean square error expressed as

$$E = \frac{1}{N} \sum (t_i - y_i)^2$$

(3-13)

The average cost is the sum of the square of the difference between the target value ($t_i$) and the output value ($y_i$) divided by the total number of data sets ($N$). The cost function allows for minimisation of network complexity (simply expressed as the number of weights) as well as the error. Since early stopping is applied when large data sets are available, the cost function is simplified to the error function.

When small amounts of data are available, a regularization term, or noise addition, may be added to the cost function to reduce network complexity and limit weight growth (Swingler, 1996). The addition of a small amount of noise (approximately 10%) to the input values will aid generalization (Caudill and Butler, 1996). The amount of noise added depends on the particular process the ANN is modeling. Noise addition is another way of promoting regularization. If the ANN validation error attains a pre-specified error designated by the ANN designer (ideally zero error), a pre-specified difference in error between epochs (ideally the difference between the errors in the training process should be zero), or a pre-designated maximum number of epochs, the ANN training is terminated.
Figure 3.4  Optimal Net Architecture in Terms of Number of Nodes in the Hidden Layer

Figure 3.5  Training and Validation Error vs. Number of Epochs for “Early Stopping” Method
3.3.5 Weights

Before training, the weights are randomly selected to be in a range of +1 to -1. A fully trained ANN has weights of a magnitude no greater than approximately 10 (Swingler, 1996). As training proceeds, Hinton diagrams, as illustrated in Figure 3.6, can be used to visually observe the weights (Eberhart and Dobbins, 1990). The ANN is training properly if the weights stop fluctuating between positive and negative values and arrives at a steady state over time. Rapid weight change from positive to negative values from one pattern to the next requires a decrease in the learning rate. Alternatively, an increased learning rate is called for when the weights change continually in the same direction or if the error decrease is small (Swingler, 1996).

3.3.6 Replicates

In order to find the global minimum, different initial starting points on the error surface is required. Each randomly selected initial set of weights specifies a different starting point for the surface search for the global minimum error. Each starting point, or seed number used to generate the random set of weights, is referred to as a replicate. Three replicates (Carpenter and Hoffman, 1995), or starting point on the error surface, were used for the training of the ANNs in this project. The training procedure is briefly demonstrated in Appendix D.
Figure 3.6  Example of a Hinton Diagram (adapted from Eberhart and Dobbins, 1990; NeuroDimension, 1997)

All weight values of this node are small ⇒ possibly only 3 nodes in hidden layer required.

Black ⇒ positive value, Small size ⇒ small magnitude

White ⇒ negative value, Large size ⇒ Large magnitude

Number of inputs to the net = 5

Number of hidden nodes = 4
CHAPTER 4

NETWORK ANALYSIS AND RESULTS

Artificial neural networks (ANNs) are applied in this project to model the post-filter particle count of the Manheim Water Treatment Plant (WTP) and the Britannia WTP. A site description for the Manheim WTP is shown in Appendix A. The Manheim WTP is a conventional treatment plant that makes use of the coagulation, flocculation, and sedimentation process. Alum and polymer are applied to two equally efficient parallel treatment trains. After the raw river source water travels through four flocculation chambers and a sedimentation tank with lamella plate settlers, it is assumed that this effluent has reacted with all the coagulant and coagulant aid such that there is no carry-over onto one of the four filters, two GAC filters and two anthracite and sand filters. Although ozonation is practiced at the Manheim WTP, the dosage is set typically to 3.5 mg/L. The Britannia WTP is a conventional treatment plant that utilizes alum and silica as a coagulant and coagulant aid to remove the raw river source water color and particles. The raw water particulates are measured with a particle counter. It has been observed by treatment plant personnel that silica does carry over onto the filters. Due to increasing water demand, another treatment train using present day technology was added to the existing processes resulting in different treatment train efficiencies. Since the particle counter is on Filter #1 which primarily treats the water from the old treatment train with the old dual media filters consisting of anthracite and sand, it has been assumed only this portion of the treatment train is sufficient to model post-filter particle counts.

Using a trial and error approach of altering network parameters and architecture, the network architectures for the Manheim Water Treatment Plant (WTP) and the
Britannia WTP were developed. The design for the Manheim WTP consisted of two separate artificial neural networks (ANNs) in attempt to reduce noise attributed to individual processes in the treatment train: one ANN modelling settled water turbidity using raw water parameters as input, and the modelling post-filter particle counts using settled water characteristics as input. This design is preferred since it is possible to assume that no carry over of coagulant or coagulant aid onto the filters occurs, and therefore the filtration process efficiency is independent of upstream conditions such as coagulant of coagulant aid dosage. The Britannia Plant post-filter particle count is modelled using one ANN representing processes found from the plant intake to the filter particle counter. This design is preferred since raw water particle counter measurements are available thereby allowing the ANN to learn the affects of particle size distribution on treatment plant efficiency. Furthermore, the personnel at the Britannia WTP requested that the affects of carry over of silica onto the filters be investigated. In this section, each of these final network architectures (previously shown in Figure 3.3) is demonstrated, analyzed, evaluated, and applied to demonstrate the various benefits of incorporating an artificial intelligence based model for plant operation and optimization purposes.

4.1 Settled Water Turbidity ANN: Manheim WTP

The settled water turbidity network is designed to mimic the Manheim WTP coagulation- flocculation- sedimentation processes. Separate ANNs have been designed for the Fall and Spring season due to the lack of polymer flow pacing, as described in Section 3.2.2, during the Summer season. A drop to approximately zero flow for two short instances during the Summer season reflects the fact that the treatment train performance was temporarily interrupted (Stendahl, 1998). These drops in flow rate
cause the range of flow values to increase more than three-fold. This extreme value affects the network architecture and the scaling of the flow input parameter. Since it was assumed that during the Summer season that polymer dosage was accidentally not flow-paced, the data from the Summer season were discarded.

4.1.1 Network Pre-processing and Design: Settled Water Turbidity Network for Manheim WTP

The network input parameters include pH, temperature, alkalinity, raw water turbidity (at various travel time lags), alum dosage, polymer dosage, and flow rate (at a pre-determined lag) as shown in Figure 4.1. A lag is defined as a set number of time steps prior to the current time step. For example, if there is an average filter detention time of 20 minutes, a lag of 2 (20 minute / 10 minute frequency between data sets) is applied implicitly in the network architecture. By using a previous time step's data, the “lagged” design may model the travel time of a slug input from the plant intake to the settled water turbidimeter. The turbidity lags were determined by calculating the theoretical maximum and minimum detention time, as described in Section C.1 of Appendix C. Then, lags in this approximate time range were entered into the ANN. The overall sensitivity of settled water turbidity to the raw water turbidity “lag” nodes, as calculated by NeuroSolutions®, are calculated and listed in order of greatest to smallest magnitude. A large overall sensitivity implies that this specific lag has a large impact on the settled water turbidity. Two lags are then chosen from all the lags investigated due to their high sensitivities and their neighbouring values to the calculated theoretical maximum and minimum detention times. These selected lags are considered to be the actual rather than the theoretical detention times such that non-ideal behaviour, such as
Figure 4.1 Settled Water Turbidity Network Architecture for the Fall and Spring Seasons – Manheim Water Treatment Plant

For the Fall season lag 1, 2, and 3 correspond to raw water at a travel time from the raw water turbidimeter to the settled water turbidimeter of 50, 70, and 90 minutes respectively; for Spring it corresponds to 80, 150, 200 minutes respectively.

Alum and polymer dosage in mg/L.

Flow is the sum of all the filter flow rates (L/s). Flow has a lag of 20 minutes due to the high correlation of backwashing with the depressions in settled water turbidity.

Temperature in degrees Celsius

Alkalinity in mg/L as CaCO_3.

Turbidity in NTU.
short-circuiting, is accounted for. Since three points may be used to define a curve, the average turbidity lag between the "maximum" and "minimum" lag nodes is added to the network architecture in order to improve the approximation of the particulate matter concentration as it has been routed down the treatment processes. The smallest error for an ANN with a "lagged" design would therefore occur when the plant flow rate is constant since the raw water turbidity lags implicit in the network architecture would not change. Even though there are few variations in alum or polymer dosages over a given season, the ANN error (mean square error) was smaller when the units used are mg/L that when using units of mg/sec. The plant flow rate is at a lag of two since the plant flow rate is the sum of the filter flow rates, each with a detention time of 20 minutes. This lag was determined by correlating the historical record of settled water turbidity with the plant flow rate. The highest correlation between settled water turbidity and plant flow rate occurs at a lag of 20 minutes, where the drop in flow rate coincides with the depression in settled water turbidity when one of the filters is taken off-line.

A different number of hidden units were required for the Fall and Spring season ANN, as illustrated in Figure 4.1. The different number of nodes required reflects the difference in the ranges of the input parameters investigated, the variations of other parameters neglected in the model (such as TOC), and the quality of the data. For example, the Spring season has a snow melt period, in which the change in pH and alkalinity were not measured. This impacts the ability of the ANN to adequately learn the effects of alkalinity and pH on settled water turbidity. Furthermore, there are only four measurements of alum and polymer dosages over the entire Fall season. The ANN
may be unable to interpolate between these values. A larger data set containing more variations would allow the ANN to learn the underlying relationships more accurately.

4.1.2 Network Results and Discussion: Settled Water Turbidity Network for Manheim WTP

The network output for the Manheim WTP are shown in Figure 4.2-A and 4.2-B for the Spring and Fall seasons respectively. The dates correspond to the duration of the Spring and Fall data sets provided by the utility, as presented in Table 3.2. The amount of variation of a given parameter that the ANN is trained on impacts the overall sensitivities calculated. The overall sensitivities of the ANN (as calculated by Neurosolutions®), are summarized in Table 4.1. The larger the overall sensitivity, the more significant the parameter is contributing to turbidity removal in the coagulation-flocculation – sedimentation processes. Parameters that do not vary much within a data set will not have a large sensitivity even though the parameter is of great significance to the process being modelled. No other studies reported in the literature within the environmental engineering field calculates overall sensitivities to show which parameters are the most significant.

The data for the Fall season show more variation in the raw water turbidity than the Spring season, as shown by the large magnitude of the calculated overall sensitivity (Table 4.1). This impacts the ANN results once training is complete since the Fall season ANN is capable of generalizing over a wider range of raw water turbidities than the Spring season. If the data range is wide enough such that all input parameters are adequately varied, the input parameter with the greatest sensitivity is the most significant parameter in terms of impacting the settled water turbidity, followed by the second
Figure 4.2  Actual and Predicted Settled Water Turbidity for Manheim WTP: A-Spring Season; B-Fall Season
Table 4.1  Summary of Overall Sensitivities (%) of Individual Input Parameter for the Settled Water Turbidity Network – Manheim WTP\textsuperscript{a}

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fall Sensitivity</th>
<th>Spring Sensitivity</th>
<th>Manheim's Experience\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>8.2</td>
<td>3.1</td>
<td>Relevant during events such as Spring snow melt</td>
</tr>
<tr>
<td>Temperature</td>
<td>8.5</td>
<td>20.4</td>
<td>Significant</td>
</tr>
<tr>
<td>Alkalinity</td>
<td>7.1</td>
<td>7.4</td>
<td>Relevant during events such as Spring snow melt</td>
</tr>
<tr>
<td>Raw Water Turbidity (lag 1)</td>
<td>8.4</td>
<td>4.3</td>
<td>Significant</td>
</tr>
<tr>
<td>Raw Water Turbidity (lag 2)</td>
<td>7.1</td>
<td>17.3</td>
<td>Significant</td>
</tr>
<tr>
<td>Raw Water Turbidity (lag 3)</td>
<td>11.6</td>
<td>5.4</td>
<td>Significant</td>
</tr>
<tr>
<td>Alum (mg/L)</td>
<td>10.0</td>
<td>29.0</td>
<td>Significant</td>
</tr>
<tr>
<td>Polymer (mg/L)</td>
<td>11.6</td>
<td>8.6</td>
<td>Significant</td>
</tr>
<tr>
<td>Flow rate (L/s at lag 2)</td>
<td>27.4</td>
<td>4.4</td>
<td>Significant</td>
</tr>
<tr>
<td>SUM (\Sigma 100%)</td>
<td>(\Sigma 100%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Calculated overall sensitivities have been rounded to one decimal place.

\textsuperscript{b} Based on the opinion of Mr. Doug Stendahl – Water Quality Coordinator.
largest, and so forth. The Spring season has alum, temperature, and raw water turbidity at a lag of 2 as the most significant parameters affecting settled water turbidity, as demonstrated in Table 4.1. The temperature varies in the Spring season such that the ANN has learnt the implications of a range of temperature values. The Fall season sensitivities demonstrate that raw water turbidity, coagulant and polymer dosage and flow rate are the most significant parameters impacting settled water turbidity. Settled water turbidity is most sensitive to flow rate as indicated by the large depressions in settled water turbidity which occurs when one of the filters is taken off-line.

Error histograms and correlation plots are used to evaluate the accuracy of the ANN’s predicted values to a test set, which consists of unseen data set constituting 25% of the original data set. The error histogram of the test set of a healthy network has a mean at zero and the same distribution as the noise in the data (Swingler, 1996). Error is defined as the difference between the output and the target (or the desired output) values. A summary of the error as calculated by Neurosolutions® for the Spring and Fall seasons, shown in Table 4.2, allow for a quantitative understanding of the distribution of error and evaluation of the ANN (Zhang and Stanley, 1997a). The error histograms are shown in Figure 4.3 for the Spring and Fall seasons.

The histograms of the network error, expressed in turbidity units, are skewed towards the positive direction. This appears to be due to either the rapid changes in flow rate that are not captured in the data or the large quantity of positive noise in the data that could not be discarded as outliers during the pre-processing phase. This noise was not discarded due to its relatively small magnitude when compared to outliers such as those due to instrumental re-calibration, and due to the proposed purposes of the ANN for
Table 4.2 Summary of Error Statistics for the Settled Water Turbidity Fall and Spring Season Networks for the Manheim Water Treatment Plant

<table>
<thead>
<tr>
<th>Performance</th>
<th>Spring Error</th>
<th>Fall Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE(^a)</td>
<td>0.0054</td>
<td>0.0248</td>
</tr>
<tr>
<td>NMSE(^b)</td>
<td>0.0976</td>
<td>0.1402</td>
</tr>
<tr>
<td>MAE(^c)</td>
<td>0.0550</td>
<td>0.1144</td>
</tr>
<tr>
<td>Min Abs Error(^d)</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td>Max Abs Error(^e)</td>
<td>0.4780</td>
<td>0.8366</td>
</tr>
<tr>
<td>(r^2)</td>
<td>0.9501</td>
<td>0.9301</td>
</tr>
</tbody>
</table>

\(^a\) The mean square error or root mean square error (RMSE) is the criteria used to evaluate which network architecture is best (Rodrigues and Serodes, 1996b; Comrie, 1997). The ANN is trained to achieve the smallest mean square error (MSE) and the architecture and replicate with the smallest MSE of the validation set is selected and tested. Optimally, it is desirable to have the mean square error of the test set equal to zero.

\(^b\) The nominal mean square error indicates the variance of the test set results. It is desirable to have as small a variance as possible.

\(^c\) The mean absolute error (MAE) allows for a visualization of the error distribution (Zhang and Stanley, 1997a). The direction of skewness is established when comparing the MSE to the MAE. The minimum absolute error would optimally be a value of zero.

\(^d\) The maximum absolute error indicates whether outliers exist and if the ANN achieves a certain level of required confidence.

\(^e\) The linear correlation is also demonstrated in the correlation plots and demonstrates the scatter along the regression line. The \(R^2\) value is the other statistic (besides RMSE) used for network error assessment (Zhang and Stanley, 1997a).
Figure 4.3  Histogram of Settled Water Turbidity Network Error – Manheim WTP: A-Spring Season, B-Fall Season
on-line applications using non-ideal data to predict settled water turbidity. In order for
the ANN to learn the underlying relationship, the positive noise of the data set will result
in an over-estimation of settled water turbidity depressions in order to minimize the
average error during the training process.

Correlation plots provide another means of evaluating the ANN’s accuracy
(Comrie, 1997; Rodrigues and Serodes, 1994). The regression line through the
correlation plot of the network output to the target value ideally has a slope of one,
intercept of zero, and correlation coefficient of one. The final ANNs for the Fall and
Spring season have been deemed to be accurate by a visual comparison of the correlation
plots to the ideal regression line of the test set data as shown in Figure 4.4. The
correlation \( r^2 \) of the settled water turbidity ANNs is 0.9027 for the Spring and 0.8651
for the Fall. Other neural network applications within the environmental engineering
field have reported \( r^2 \) values ranging from 0.998 (Zhang and Stanley, 1997a) for a water
demand forecasting model, to 0.67 (Comrie, 1997) for an atmospheric ozone forecast
model. The slope of the network correlation plots (0.9004 for Spring; 0.9293 for Fall) are
slightly less than one and the intercepts (0.1551 for Spring; 0.0931 for Fall) are greater
than zero. The value of these statistics demonstrate that the ANN is learning to avoid the
noise in the data at the expense of over-approximating the settled water turbidity when
one of the filters is taken off-line. The ANN for the Fall season shows a reduced
accuracy in predicting at lower settled water turbidities, as demonstrated by the greater
variance at the base of the regression line. The increased variance for the lower settled
water turbidity results may be due to the smaller quantity of data available at lower flow
rates and the training algorithm itself. Outliers in the data set impact the slope, intercept,
Figure 4.4  Correlation Plot of Actual (Target) and Predicted (Output) Values of Settled Water Turbidity for Manheim WTP: A- Spring Season; B- Fall Season

A

Potential outlier

\[ y = x \]
\[ R^2 = 1 \]
\[ y = 0.9004x + 0.1551 \]
\[ R^2 = 0.9027 \]

B

\[ y = x \]
\[ R^2 = 1 \]
\[ y = 0.9293x + 0.0931 \]
\[ R^2 = 0.8651 \]

* if falls out of 95% confidence interval and it is a conflicting data point or there is reasonable justification, such as instrument re-calibration, this point may be discarded
and correlation coefficient of the regression line and can be easily identified with these plots during the training process.

4.1.3 Application of Settled Water Turbidity Network: Manheim WTP

In order to demonstrate the use of the final network architecture's underlying relationships for process modeling and optimization, a sensitivity analysis was conducted to illustrate the underlying relationships learned by the ANN. Due to the small data set, only a small range of values has been investigated, as demonstrated in Figures 4.5 and 4.6 for the Spring and Fall seasons, respectively. Zhang and Stanley (1997b) used this approach to determine the underlying relationships for an enhanced coagulation model. The alum and polymer dosages examined in these plots are the existent dosages applied to the water in the Manheim Plant during their respective seasons. All other parameters were kept constant at the values that occur most frequently in the data set in order to ensure that the ANN has learnt the underlying relationship well and that extrapolation beyond it's experience does not occur.

The sensitivity of turbidity removal to alum and polymer dosage respectively is shown in Figure 4.5-A and 4.5-B. The underlying relationships for the Spring season illustrate that an increase in raw water turbidity yields and increase in settled water turbidity. The sensitivity analysis was conducted for a raw water alkalinity of 180 mg/L and a pH of 8.2, the most recurrent values that occur for in the training set. The alum sensitivity plot is conducted for a polymer dosage that was typical over the data set provided by Manheim WTP of 0.112 mg/L. For the sensitivity to polymer, the alum dosage is set to 42 mg/L, the most frequent value (over the data set provided by Manheim WTP). Due to the limited amount of data points at high levels of raw water turbidity
Figure 4.5  Sensitivity of Settled Water Turbidity during Spring Season at a Temperature of 9°C and a Plant Flow Rate of 300 L/s – Manheim WTP: A- to Alum dosage\textsuperscript{a}; B-to Polymer Dosage\textsuperscript{b}

\textsuperscript{a} For the most frequent polymer dosage (= 0.112 mg/L)

\textsuperscript{b} For the most frequent alum dosage (= 42 mg/L)
 (> 3 NTU), the network relationships are not accurate in this range and may have required extrapolation beyond the ANN’s experience. During the Fall season, however, an increase in raw water turbidity results in a decrease in settled water turbidity at an alkalinity of 170 mg/L and a pH of 7.9, as demonstrated in Figure 4.6. The decreasing trend apparent in the Fall season may be due to the ANN extrapolating beyond it’s experience since the only raw water turbidity value that the ANN is exposed to in the training set is approximately 3 NTU when the pH = 7.9, alkalinity =170 mg/L, temperature = 15°C, and flow rate = 400 L/s. Another less probable explanation of the decreasing trend during the Fall season is that there is a small colloid concentration and with the alkalinity and coagulant and polymer dosages, turbidity must be added in order to achieve sweep flocculation. Turbidity addition has not traditionally been practiced at the Manheim Plant. The personnel at the Manheim WTP may want to consider such an option for settled water turbidity reduction, particularly if the addition of turbidity cost is less than increasing the alum and polymer dosage in order to attain sweep flocculation. Both seasons indicate that a slight increase in alum dosage is less effective at lower raw water turbidities than at high raw water turbidities. A slight increase in polymer dosage, however, is more effective in turbidity removal at low raw water turbidities.

If the data set contains at least nine combinations of applied alum and polymer dosages, a surface plot could be constructed such that for a desired level of settled water turbidity and a given raw water turbidity, an optimal combination of alum and polymer can be selected. Even though the ANN is extrapolating beyond its experience at polymer dosages near 0.2 mg/L where only a few training data sets exist, a contour plot of settled water turbidity as a function of alum and polymer dosage for the Spring data set
Figure 4.6  Sensitivity of Settled Water Turbidity during Fall Season at a Temperature of 15°C, and a flow rate of 400 L/s– Manheim WTP: A- to Alum Dosage\textsuperscript{a}; B- to Polymer Dosage\textsuperscript{b}

\textsuperscript{a} at a polymer dosage of 0.06 mg/L
\textsuperscript{b} at an alum dosage of 47 mg/L
was prepared and is illustrated in Figure 4.7. MathCad® was used to created the contour plot iso-turbidity lines (from the Excel® spreadsheet data that contained the sensitivity analysis conducted by NeuroSolutions®) that now can be used as a tool for determining the optimal coagulant and polymer concentration for a pre-selected settled water turbidity and a given raw water turbidity. For example, as demonstrated in Figure 4.5 by the dashed arrows, if the objective is to achieve a typical Spring settled water turbidity of 1.4 NTU given a raw water turbidity of 4 NTU, the addition of 42 mg/L of alum and 0.11 mg/L of polymer represents an optimal condition (option #1). This option will cost approximately $12.35/ML as shown by Figure 4.8. If a settled water turbidity of 1.4 NTU is required, the alum dosage can be reduced to 39 mg/L and the polymer dosage may be increased to 0.15 mg/L (option #2), as demonstrated in Figure 4.7, at a cost of $12.60/ML. Assuming that the floc produced at these two combinations of polymer and alum are equivalent in strength, size, and density, clearly option #1 is better.

At a raw water turbidity of 4 NTU, increasing the alum dosage to over 40 mg/L becomes less effective than a 0.05 mg/L increase in polymer dosage for Spring Season. The optimal cost for a desired level of settled water turbidity can be found if the plots shown in Figure 4.7 and 4.8 are combined. For example, for a desired settled water turbidity of 1.5 NTU, the optimal cost is approximately $11.50/ML over the range of dosages in the data set provided by the Manheim WTP at an alum dosage of 38 mg/L and a polymer dosage of 0.11 mg/L. For a settled water turbidity of 1.4 NTU, the minimum cost is $12.20/ML over the range of dosages investigated for an alum dosage of 43 mg/L and a polymer dosage of 0.1 mg/L. For a settled water turbidity of 1.3 NTU, the
Figure 4.7  Contour Plot of Settled Water Turbidity as a function of Polymer and Alum Dosages for the Spring Season - Manheim Water Treatment Plant

* Applicable for raw water turbidity = 4 NTU

Figure 4.8  Contour Plot of Cost ($/ML) for Various Combinations of Alum and Polymer Dosages

* constructed using alum costs $0.216/kg and polymer (Percol LT27A) costs $5.75/kg dry weight added with a dilution factor of 0.2 mg/L.
minimum cost is approximately $12.50/ML for an alum dosage of 35 mg/L and a polymer dosage of 0.16 mg/L. Therefore, in order to reduce the settled water turbidity by 0.1 NTU over the range of values investigated, the cost increases at a diminishing rate.

For the Fall Season, at higher raw water turbidities (such as 4 NTU), the alum appears to be the more effective in removing turbidity than polymer. At low raw water turbidities (such as 2.5 NTU), a 0.01 mg/L change in polymer dosage is the more effective chemical added in terms of particulate removal. The optimal polymer dosage appears to be 0.07 mg/L from the sensitivity plots, but this ANN result may be due to the fact that the only polymer dosage investigated at the selected input parameter values (such as pH, alkalinity, and so forth) is 0.07 mg/L. Therefore other polymer dosages investigated in the sensitivity analysis are subject to a greater amount of error.

A path of minimum turbidity and cost is clearly illustrated in Figure 4.9 for the Fall season. The trade-off between settled water turbidity and cost becomes further complicated when applying the particle counter network for the minimization of post-filter particle counts. An example of cost minimization and post-filter particle count minimization is described in Section 4.2.3. Using the sensitivity results illustrated here in combination with the post-filter particle count sensitivities, the alum and polymer dosages can be determined and continuously altered on-line in order to achieve optimal particulate removal at a pre-selected cost, or the optimal cost at an allowable level in the filter effluent.

4.2 Particle Count Network: Manheim WTP

Once an ANN that is capable of modelling post-filter particle counts has
Figure 4.9 Contour Plot of Settled Water Turbidity and Cost ($/ML) as a function of Polymer and Alum Dosages for the Fall Season - Manheim Water Treatment Plant

Path of optimal turbidity removal and cost within the range investigated (in direction of minimizing cost)

a Applicable for pH = 7.9, Temperature = 16.46 °C, alkalinity = 170 mg/L, plant flow rate = 400 L/s

b Constructed using alum costs $0.216/ kg and polymer (Percol LT27A) costs $5.75/ kg dry weight added with a dilution factor of 0.2 mg/L
been developed, a sensitivity analysis of the particle counts to the settled water turbidity (and filter flow rate) may be conducted using Figure 4.9 for the Fall season. Since settled water turbidity is a function of the alum and polymer dosage, the post-filter particle count can ultimately be expressed in terms of alum and polymer dosages for a given set of raw water characteristics such as turbidity, pH, alkalinity, plant flow rate, and filter flow rate. Furthermore, the cost of particulate removal in terms the amount of chemicals added for coagulation may be optimized on the basis of post-filter particle counts.

4.2.1 Data Pre-Processing and Network Design: Particle Count Network for Manheim WTP

Particle counter networks for the Manheim WTP have been developed individually for the Fall and Spring season, due to the variations in seasonal data quality. For example, during the Fall season, an increasing trend in particle counts (up to a maximum of 999.9 counts/mL) at the end of the filter run signifies that breakthrough occurs prior to the end of the filter run and hence produces a poorer quality filter effluent than during the Spring season. The distribution of the data for the post-filter particle count contains a large number of points during filter ripening and comparatively fewer points of larger particle counts due to breakthrough when compared to filter ripening. An ANN operates best if the distribution is even over the range of the target and input parameters (Swingler, 1996), since the ANN learns the underlying relationship by minimizing the average error. Therefore, it is hypothesized that the variance of network error during breakthrough is greater than the variance during filter ripening, resulting in less accurate predictions of particle count during breakthrough. In addition, during the
Spring season the particle counter does not detect breakthrough at an earlier run time than the turbidimeter, since run time is the defining backwash criteria. Therefore, the accuracy of the ANN for the Spring season is dependent on the ANN’s ability to distinguish noise in the data from valid fluctuations in particle counts. It is therefore hypothesized that a decreased ANN learning accuracy will result for the larger particle counter channels where fluctuations are more predominant.

Several assumptions were made when constructing the ANN. It is assumed that the water colour, which affects particle count measurements (Hargesheimer et al., 1992), remains constant (in the range of <1, 10, or 50 TCU) in the filter effluent. Since the ozone dosage remains fairly constant, the effects of slight changes in the ozone dosage cannot be modelled by the ANN and are deemed negligible. It is assumed that all the alum and polymer reacted prior to the filters such that there is no carry-over of these chemicals onto the filters and therefore no nodes for alum and polymer are required. Any optimization of the alum and polymer dosages with respect to post-filter particle count is done in parallel with the ANN developed for settled water turbidity. Since turbidity (NTU) is the measurement used to define the quantity of particulate mass (mg/L) entering the filter (Nielson et al., 1973), there is no means of describing the particle distribution into the filter. Therefore, it is assumed that the particle distribution remains constant and that any variations in the raw water particle distribution will be dealt with in the coagulation – flocculation – and sedimentation processes. Furthermore, since the quantity of particles entering the filter are measured in units of mass (Nielson et al., 1973) but a volume balance of the particles entering, stored, and discharging from the filter (as described by equation 3-6) is implicit in the neural network design, it is assumed
that the particle density entering the filter remains constant. Periods of zero flow are considered to be due to backwash and have been eliminated from the data set.

The network architecture, as demonstrated in Figure 4.10, has been designed such that subsequent sensitivity analyses can be conducted with ease. Therefore, a time node was used to account for the filter ripening trend rather than incorporating other network architecture designs described Table C.1. One target node (representing the channel size analyzed) per ANN will result in less data required to obtain a certain level of network error than an ANN with four output nodes. The ANN, however, is designed with four target nodes in order to reduce the training time four-fold (or one time for each ANN representing only one particle count channel).

The network target values are expressed in terms of percent of total particle counts for the Fall, and in actual particle counts for the Spring. The format of percent total particle count for the output units results in a more even distribution of the target values over a smaller range and therefore a potentially smaller network error. In other words, the scale of the particle count data is reduced to a maximum of 100% rather than a maximum of 999.9 counts/mL, and, the transformation into units of percent allow for a more even number of counts in the low to high range than units of counts. In order to convert from percent of total particle count (channel count/total particles in the channels *100%) to actual particle count, a separate total particle count ANN was constructed, as illustrated in Figure 4.11, such that the percent of total particle counts can be multiplied by the results from the total particle count ANN to get the channel counts. Any error from the total particle count network will result in a consistent error in all particle
Figure 4.10  Particle Count Architecture - Manheim WTP

Channel Particle Count expressed as percent of total particle count for the fall season, and in particle count for the spring
Cumulative sum of the settled water turbidity in mg (turbidity [NTU= mg/l] multiplied by the flow rate)
The previous time step’s flow rate (1/s) is required so that abrupt changes in flow rate is accounted for.
Initial head loss in percent. Aids in modeling smaller particle size ranges.
Head loss in units of %. One head loss unit is at the current time and the other at the previous time step to account for changes in filter storage.
Run time in minutes starting from the time the filter is brought back on-line.
The inverse of the coefficient of permeability= headloss/ filter flow rate. Provides aid to modelling storage and the change in porosity such that interstitial flow forces may be considered.
Figure 4.11. Total Particle Count Architecture for Fall Season - Manheim WTP

- **pH**
- **Temp.**
- **Alk.**
- **Σ I(t)**
- **Filter Flow**
- **Filter flow**
- **h₁(t)**
- **h₁(t - 1)**
- **run time(t)**
- **1/(K)**

* Total particle count in units of microns.

* Cumulative sum of the settled water turbidity in mg (turbidity [NTU]= mg/L multiplied by the flow rate)

* The previous time step's flow rate (L/s) is required so that abrupt changes in flow rate is accounted for.

* Initial head loss in percent. Aids in modeling smaller particle size ranges.

* Head loss in units of %. One head loss unit is at the current time and the other at the previous time step to account for changes in filter storage.

* Run time in minutes starting from the time the filter is brought back on-line.

* The inverse of the coefficient of permeability= headloss/ filter flow rate. Aid to modelling storage and the change in porosity such that interstitial flow forces may be considered.

* The filter particle count distribution is the 1 µm, 2µm, 5 µm, and 10 µm particle count in percent. The distribution assists this network's accuracy.
channels when converting from percent to actual counts. The Spring season utilizes the actual channel particle count due to the large quantity of noise, which distorts the calculations of the percent particle count for each particle counter channel. Furthermore, the particle count data for the Spring season demonstrate that the filter is backwashed on a run time criteria rather than a turbidity criteria used in the Fall. Therefore, the Spring season does not have a noticeable amount of increased particle counts at the end of the filter run since the maximum run time is exceeded before the post-filter turbidity reaches 0.1 NTU.

Trend in the post-filter particle count data set influences the design and analysis of the ANN model. For example, the required window size is dependent on the initial filter ripening, (as illustrated in Figure 2.7 as window a, b, and c,) that does not appear in Manheim’s data. Therefore, inclusion of a time node alone is sufficient to distinguish between different parts of the curve. Furthermore, the overall sensitivities of the particle counts to the input parameters are dependent on the distribution of the values of the input parameters. Therefore, the ANN learns that parameters such as the sum of the turbidity (expressed on a mass basis) into the filter is more predominant than it potentially is in reality on the post-filter particle counts due to its similar steadily increasing trend with time.

The input parameters have been selected on the basis of principles described in section 3.2. Input parameters include pH, temperature, alkalinity as CaCO₃, the cumulative turbidity (expressed on a mass basis) from the beginning of the filter run, storage (within the voids) and change in storage, the flow rate and change in flow rate, run time, and the ratio of headloss to flow rate (as defined in Appendix C.5 as the
coefficient of permeability). The temperature node is assumed to implicitly model changes in water viscosity and density, which may influence filter attachment mechanisms. The performance of the ANN was observed to improve when using a mass expression for the filter influent turbidity expressed as the cumulative sum of the turbidity multiplied by the flow rate (otherwise expressed as \(\Sigma(\text{turbidity}\times\text{flow rate})\)). The type of flow rate, such as the superficial versus the interstitial flow rate, may be dependent on the predominant attachment and detachment mechanisms, as illustrated in Figure 2.8. Since porosity measurements affect interstitial velocity, the coefficient of permeability was added to the ANN due to its proportionality to porosity.

4.2.2 Results and Discussion: Particle Count Network for Manheim WTP

The network predictions are compared to actual data in Figures 4.12 to 4.14. For the Spring season, the ANN under-predicts an incident of breakthrough occurring at 4/30/97 in the middle of a filter run despite the sudden drop in headloss, as shown in Figure 4.12-A. Furthermore, the Spring ANN (shown in Figure 4.12), is incapable of predicting the minute amount of breakthrough due to the large quantity of particle count fluctuations occurring in the 5 and 10 \(\mu\text{m}\) channels and a lack of data sets which represent breakthrough. For the 1 \(\mu\text{m}\) channel, as demonstrated in Figure 4.12-A, the ANN over-predicts in locations where the particle count detection limit of 999.9 particle counts was surpassed, potentially providing a more realistic result than the particle count measurements itself since the particle counter cannot measure counts greater than the detection limit.

For the Fall season, the “channel particle count” ANN has the most difficulty in accurately predicting at the end of each filter run, as demonstrated in Figure 4.13.
Figure 4.12  Actual and Predicted Values for the Spring Particle Count Network – Manheim WTP: A- 1 µm channel, B- 2µm channel
Figure 4.12  Actual and Predicted Values for the Spring Particle Count Network – Manheim WTP: C- 5 \( \mu \)m channel, D- 10 \( \mu \)m channel
Figure 4.13  Actual and Predicted Values in Percent Particle Count for the Fall Channel Particle Count Network – Manheim WTP:  A- 1 μm channel, B- 2μm channel
Figure 4.13 Actual and Predicted Values in Percent Particle Count for the Fall Channel Particle Count Network – Manheim WTP: C- 5 μm channel, D- 10 μm channel
Figure 4.14  Actual and Predicted Values for the Total Particle Count Network- Fall Season- Manheim WTP

Large under-prediction of peak
The "channel particle count" network predicts the particle counts for each channel of the particle counter in terms of percentage of total counts of all the channels. The ANN difficulty in predicting the peaks of the filter ripening curve may be due to the ANN training procedure of finding the minimum average error. Since there are less available data sets during breakthrough that the number of data sets available for filter ripening, in order to minimize the average error the ANN will underestimate the fewer points at the end of the run in order to achieve a better fit for the rest of the filter ripening curve. Furthermore, a data set larger than the Fall data set provided by Manheim WTP containing sixteen filter runs (and sixteen periods of backwash) would improve filter breakthrough predictions. The "channel" ANN's output (in % of total counts) is used as input to the "total particle count" ANN since the distribution of the particle sizes is a function of the run time (Hargesheimer et al., 1992) and therefore allows for improved predictions of the filter ripening trend. The maximum error for the Fall season occurs at the end of the third filter run, as shown in Figure 4.13-C and Figure 4.14, in terms of actual counts. At the end of this third filter run, there is an approximate error of 10% of the total particle count for the 5 μm channel.

All of Manheim's particle count ANNs are evaluated with the help of error summary tables and correlation plots. The error summary table can be used as another means (instead of error histograms) of evaluating the distribution of error and are summarized in Figure E.11 for the Fall channel particle count network, Figure E.9 the Fall total particle count network, and Figure E.6 the Spring particle count network. The correlation plots, illustrated in Appendix E, demonstrate the ANN's predictive capabilities. The variance of the correlation plot tends to increase in the range of values.
where the end of the filter run occurs. The skewed regression line of the correlation plots reflects the ANN’s relative inaccuracy at the end of the filter run during the Fall season due to the fewer data sets available during breakthrough than filter ripening. During the Spring season, the regression line of the correlation plots is skewed due to the random variations in count. These random variations were not removed during pre-processing because there was no indication that such fluctuations were due to an external source of a disturbance, such as human error made by the plant personnel, or equipment failure. The ideal regression line, y=x, appears to have a good fit (average error of 68.6 counts/mL, variance of error at low particle count range smaller than at higher range) with the majority of the data points.

4.2.3 Application of Results: Particle Count Network for Manheim WTP

Considering that particle counts in a given channel size may be used as surrogates for pathogens such as Cryptosporidium and Giardia, the behavior of the individual particle size channels are of importance. Hargesheimer et al. (1992) evaluated the HIAC Royco particle counter Model 8000A with a HR-LD150 sensor and determined that the mode for Cryptosporidium is measured in the <1-2 µm range and the mode of Giardia is found in the 4-5 µm range. It is therefore assumed that the HIAC Royco 8000A measurements for Manheim’s water will have the 1 µm channel as the most suitable as a surrogate for Cryptosporidium and the 5 µm for Giardia. The sensitivity of the 1 µm and 5 µm channel particle counts to filtration rate and settled water turbidity has been explicitly determined and are shown in Figure 4.15, such that the particulate and pathogen removal efficiency can be maximized by minimizing the channel counts. The sensitivity analysis is conducted for a typical filter run of 1600 minutes, which is a typical
Figure 4.15  Particle Count as a Function of Filtration Rate and Settled Water Turbidity at a Run Time of 1600 minutes- Fall Season- Manheim WTP: A- 1 μm channel (Cryptosporidium surrogate) ; B- 5 μm channel (Giardia surrogate)
run length under the given set of conditions defined by the other input parameters. Ultimately, sensitivity results of this type may aid plant personnel in improving plant removal efficiencies for both particulate matter and pathogens. Caution must be exercised when using the sensitivity analysis results since the ANN may have in some cases extrapolated beyond its experience as a result of the limited amount of data available for training.

Since filter-to-waste is practised, only the end of the filter-ripening curve where breakthrough occurs has been investigated to determine the underlying relationships such that minimization of particulates in the filter effluent may be pursued. Combining the error of the Fall channel (%) network and the total particle count network (found respectively in Table E.5 and E.4 of Appendix E), the root mean square error (RMSE) of the 1 μm channel is 1.4 counts/mL and the RMSE for the 5 μm channel is 0.4 counts/mL. The Fall data set contains filter runs that are terminated on the basis of turbidity or headloss rather than filter run time. Therefore there are varying filter run lengths over this data set (from a minimum of 1340 minutes to a maximum of 3690 minutes) but at a pH of 7.9, temperature of 16.5 °C, and alkalinity of 170 mg/L, the filter run length is approximately 1600 minutes. The input parameters were designated at these values since they occur most frequently in the data set, and therefore, the ANN may have learned them to a higher degree of accuracy. At a typical filter run of approximately 1600 minutes, a reduction in initial flow rate by 5 L/s results in an approximate decrease of 150 counts for the 1 μm channel and 50 counts for the 5 μm channel, as shown in Figure 4.15. Furthermore, as shown in Figure 4.15, a decrease of 0.1 NTU results in a decrease of roughly 450 counts and 35 counts for the 1 μm and 5 μm channels, respectively.
It is currently unknown how many *Giardia* or *Cryptosporidium* are in the filter effluent for the given particle count measurements and therefore it is unknown what a removal of 450 counts actually means in terms of pathogen content. If the concentration of *Giardia* or *Cryptosporidium* can be approximated, an allowable level of post-filter particle counts may be used as criteria for backwash and optimization of particulate removal processes is possible based on the number of potential pathogens in the post-filter effluent. For example, if a removal of 450 counts is equivalent the removal of 1 pathogen, the additional cost of the alum and polymer added to reduce the settled water turbidity by 0.1 NTU in order to attain a reduction of 450 counts in the filter effluent may not be an feasible option in terms of pathogen removal (rather than particulate removal).

Once breakthrough begins, as observed by an increase post-filter particle counts and headloss for a given filtration rate, it is best to initiate backwash since an increase in settled water turbidity or flow rate will lead to a sharp increase in particle counts. Other possible means of control of particle counts at run times greater than 1100 minutes include: i) a lower flow rate, or ii) settled water turbidity. If a lower flow rate is not possible, the minimization of particle counts can occur by minimizing settled water turbidity using the underlying relationships obtained from the corresponding settled water turbidity ANN.

A small decrease in settled water turbidity (0.1 NTU) will result in a large decrease in particle count (450 counts/mL), particularly for the 1 µm channel, which is a surrogate for *Cryptosporidium*. In addition, a reduction in initial filtration rate by 5 L/s will result in a decrease in particle counts (150 counts/mL for 1 µm channel). Since the impact of a change in flow rate was not investigated explicitly in the ANN (because the
flow rate is never changed abruptly at the Manheim WTP), an overall lower filtration rate or a very gradual decrease in filtration rate at the time breakthrough begins is required. A lower flow rate implies that less potable water is available. This cost must be compared to the cost of the reduction in average settled water turbidity (equivalent to the (cumulative sum of (settled water turbidity * flow rate))/run time * flow rate) by using various combinations of alum and polymer dosages in order to determine which alternative is best.

The settled water turbidity network and the particle counter network can be used to minimize cost for a desired level of post-filter particle counts. For example, at a run time of 1600 minutes, filtration rate of 105 L/s, and a maximum allowable particle count of 600 counts for the 1 μm channel and 60 counts for the 5 μm channel, the maximum allowable settled water turbidity is approximately 1.4 NTU (at a pH of 7.9, temperature of 16.5 °C, and alkalinity of 170 mg/L). Using Figure 8, a polymer dosage of 0.077 mg/L and 40.5 mg/L of alum is one of the possible combinations to achieve a settled water turbidity of 1.4 NTU if the raw water turbidity is 4 NTU. The cost of this alternative is $11.00/ML, as shown in Figure 4.9. Moreover, by travelling along the 1.4 NTU iso-turbidity line in Figure 4.9, at a polymer dosage of 0.075 and at an alum dosage of 39 mg/L the cost is $10.75/ML. But at a cost of $10.75/ML, a settled water turbidity of 1.25 can be achieved if the polymer dosage is 0.066 mg/L and the alum dosage is 40 mg/L. The path of minimum turbidity for a given cost is demonstrated in Figure 4.8. Assuming that the floc characteristics, such as floc strength, do not vary over the iso-turbidity lines, the optimal combination of alum and polymer may be selected to minimize chemical feed costs while obtaining a pre-specified post-filter particle count.
4.3 **Particle Count Network: Britannia Water Treatment Plant (WTP)**

The post-filter particle count of a second water treatment plant was modelled in order to demonstrate network architecture design variations. The pre-processed data obtained from the Britannia WTP in Ottawa, Ontario are summarized in Table B.1 of Appendix B. The Britannia Plant has a particle counter at the plant intake allowing the ANN modelling the process from the intake to the post-filter particle counter (as demonstrated in Figure 3.3) to potentially learn the effects of influent particle distribution on the post-filter particle counts. Filter-to-waste is not practised at the Britannia WTP. Furthermore, there are approximately only 5 filter runs, \(\approx 0.1\%\) of the complete data set, that contain data representing filter breakthrough. The initial filter ripening represents approximately 0.25\% (100 minutes for the initial filter ripening /40000 minutes a typical filter run) of the complete data set in which the magnitude of the particle counts is about ten times larger than the counts during breakthrough. As such, separate “expert” ANNs are required to model the initial filter ripening occurring in the first 100 minutes of the filter run, and the rest of the curve.

The criteria used for filter backwash at the Britannia WTP over the duration of the data set is run time and therefore backwash typically occurs prior to filter breakthrough. Since only 0.1\% of the data represents filter breakthrough, the ANN will have difficulty in modelling the curve from 100 minutes until the end of the filter run because the ANN learns by generalizing the underlying principles and minimizing the average error. In addition to the minimal amount of data available to represent filter breakthrough, the headloss record was not available. Although the headloss parameter may aid in
modelling the maximum storage capacity of the filter and therefore breakthrough, it does not aid the initial filter ripening to the same degree.

4.3.1 Initial Filter Ripening Network: Britannia WTP

Since only 0.25% of the complete data set is used to model initial filter ripening, less than 2,000 data sets are available for network training, validation, and testing. This limited amount of data affects the form of expression of the input parameters such that the number of input nodes is minimized. In addition to influencing the design, the size of the testing and validation sets are altered such that sufficient data are available for training. In order to provide enough data for training, only 10% of the data sets were designated for validation and 15% of the data set for testing. The size of the testing set affects the evaluation of the ANN model since it may not contain an even distribution of the complete data set. Three randomizations of the pre-processed data sets were investigated in order to obtain a representative sample from the limited amount of data available for modelling. The trial that contained the most representative data of the average of the network output for the complete data set was used for the sensitivity analysis.

4.3.1.1 Data Pre-processing and Network Design

In order for the ANN to learn properly, all outliers must be removed which further reduces the data set. Data that falls out of the 95% confidence interval (if the data of the input parameter are normally distributed) are considered to be outliers according to Swingler (1996). Any large abrupt unexplained increases in the parameter's value have been discarded. For the IBR particle counter used in the Britannia WTP, any particle count measurements greater than 20,000 counts/mL, or a conservative estimate of 16,000...
counts/mL, have been discarded since they have exceeded this detection limit (Michelow, 1998). The episode of run-off that occurs during April causes the particle counts to drastically increase such that the particle counter obtains poor readings. In order to remedy this, data sets that contained data which do not exceed the detection limit of the particle counter was assumed to remain constant over periods in which the counts jump into the range of millions of counts. In addition, a filter flow rate less than 1 ML/d occurs during the period of time in which the filter has been taken off-line to be backwashed (Van Den Oever, 1998), has been discarded.

The data provided by the Britannia WTP contain unequal frequencies between adjacent data sets. For example, some data sets are taken 5 minutes apart while others are taken 3 minutes apart. In order to lag the data, the VLOOKUP command in Excel® allows the programmer to roughly determine the value of a given input parameter at a pre-specified detention time. No lags were used for the raw water particle count since the settled water turbidity is only measured every 12 hours and therefore assumed constant, regardless of the raw water particle counts.

The final form of expressing the input parameter in the network design, as illustrated in Figure 4.16, also may reduce the number of data sets required for training. For example, there are over 5 raw water particle count channels in the raw data, each requiring one node, unless substituted with the power law slope and intercept coefficient (as described by equation 3-11). This reduces the number of input parameters for raw water particle count to two, the power law intercept (proportional the particle concentration) and the power law slope (proportional to the particle distribution), thereby reducing the data quantity required to model the initial filter ripening curve. The post-
Figure 4.16  Particle Count Architecture for the Initial Filter Ripening Net- Britannia WTP

Channel Particle Count expressed in incremental counts within range
Power law coefficient representing raw water particulate concentration and distribution
Inverse of plant flow rate. Units are in d/ML.
Filter flow rate measured 10 minutes prior to current time
Slope of all lags from current flow rate to flow rate 20 minutes ago.
filter particle counts cannot be expressed in terms of the power law coefficients since there is only a maximum of two channels that have all their counts above 10 counts/mL (Hargesheimer et al., 1992). The post-filter particle counts for the Britannia WTP are expressed in terms of incremental particle counts, for example, the the 2 μm channel represents the counts between 2 to 3 μm, as similarly expressed for the Manheim WTP.

Due to the small amount (<300 sets) of data available for testing, the ANNs were evaluated on the basis of network architecture complexity (the amount of nodes contained in the network), network input layer design for easy determination of sensitivity of particle counts to input parameters, and the mean square error (MSE) and linear correlation coefficient (Zhang and Stanley, 1997a) of the test set and the complete data set. The rate of change of the flow at the beginning of the run affects the peak of the initial filter ripening curve. One method of accounting for the rate of change of the filter flow is the addition of previous lags of flow rate as input to the ANN. However, in order to reduce the number of input nodes to the ANN, the filter flow rate and the change of flow rate are inputted to the ANN as the flow rate at a lag (or filter detention time) of 10 minutes and the slope of flow rates over the last 20 minutes, respectively. The cumulative filtrate volume from the end of the last filter run affects the amount of particles left over after backwash, but was removed from the ANN during the training process due to it's low overall sensitivity value. No node was designated for backwash efficiency since no quantifiable measure was available. The plant flow rate (ML/d) is expressed as it's inverse since this is proportional the theoretical reaction time (d/ML) of alum per unit volume of filtrate so that the ANN will not get confused between plant flow rate and filtration rate.
Although 3 different types of acidified alum have been applied to the water over the duration of the data set (including 5%, 3%, and 0% acidified alum), only the alum content on a dry weight basis requires a node since the pH measured after the alum has been added (at the low lift pump) may account implicitly for the additional acid content added to the raw water.

4.3.1.2 Network Results and Discussion

Once an optimal ANN has been trained and validated, the ANN was evaluated on the basis of the error summary table, correlation plot, and error histogram of the test set, as demonstrated in Table 4.3 and 4.17 for the 2 to 3 μm channel. The 2 to 3 μm channel may be used as a surrogate for Cryptosporidium for the IBR particle counter (Michelow, 1998). The results indicate that 79.7 percent of the error values fall in the +150 to -150 particle count range, as shown in Figure 4.17-C,D. The error histogram (Figure 4.17-B) is slightly shifted to the right due to the ANN’s under-estimation of the peaks. The correlation plot (Figure 4.17-A) contains a comparatively smaller number of points than the ANN for the Manheim WTP, demonstrating the pitfalls of evaluating the ANN on a small amount of data selected randomly from the complete data set provided by Britannia. The regression line of the correlation plot (slope = 0.8712, intercept = 19.406, and $r^2=0.8049$) is dependent on the contents of the randomly selected test set due to the limited amount of data available. The time trend plot demonstrating the ANN’s ability to learn is illustrated in Figure 4.18.

The overall sensitivities, (as calculated by NeuroSolutions®), and summarized in Table 4.4, may be used to identify which input parameters impart a significant sensitivity
Table 4.3  Summary of Error Statistics for the Initial Filter Ripening Network-Britannia WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>2 to 3 μm</th>
<th>3 to 5 μm</th>
<th>5 to 10 μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>81508.6</td>
<td>12144.2</td>
<td>600.1</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.2</td>
<td>0.2</td>
<td>0.25</td>
</tr>
<tr>
<td>MAE</td>
<td>141.5</td>
<td>54.2</td>
<td>11.7</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.49</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>1409.6</td>
<td>571.9</td>
<td>142.2</td>
</tr>
<tr>
<td>( r )</td>
<td>0.8972</td>
<td>0.8967</td>
<td>0.8675</td>
</tr>
</tbody>
</table>

Figure 4.17  Test Set Results for the Initial Filter Ripening Network for the 2 μm Channel - Britannia WTP: A-Correlation Plot, B- Error Histogram

A

\[ y = x, \quad R^2 = 1 \]

\[ y = 0.8712x + 19.406, \quad R^2 = 0.8049 \]

B

![Error Histogram](image)
Figure 4.17  Test Set Results for the Initial Filter Ripening Network for the 2 μm Channel- Britannia WTP: C: Cumulative Distribution Plot of Error, D- Confidence for Interval

\[ a \] represents the error limit, such that the probability that the network error (X) falls between the positive and negative error limit is defined as \( P(-a<X<a) \)
Figure 4.18  Time Trend Plot of Actual and Predicted Values of the Initial Filter Ripening Network for the 2 to 3 μm Channel -Britannia WTP

Table 4.4  Summary of Overall Sensitivities for the Britannia Initial Filter Ripening Network for Particles in 2 to 3 μm Range

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Overall Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Law Slope (Raw Water Particle Concentration)</td>
<td>34.5</td>
</tr>
<tr>
<td>Power Law Intercept</td>
<td>3.3</td>
</tr>
<tr>
<td>1/Plant flow Rate</td>
<td>5.8</td>
</tr>
<tr>
<td>Filter Flow Rate 10 minutes prior to current time</td>
<td>15.1</td>
</tr>
<tr>
<td>Rate of Filter Flow Change</td>
<td>2.6</td>
</tr>
<tr>
<td>Run Time</td>
<td>19.9</td>
</tr>
<tr>
<td>Temperature</td>
<td>10.7</td>
</tr>
<tr>
<td>Low Lift pH</td>
<td>14</td>
</tr>
<tr>
<td>Conductivity</td>
<td>1.4</td>
</tr>
<tr>
<td>Alum</td>
<td>1.4</td>
</tr>
<tr>
<td>Silica</td>
<td>1.1</td>
</tr>
</tbody>
</table>
on the post-filter particle counts (based on the limited amount of data available). The ANN has learned that the post-filter particle count is primarily dependent on the raw water particle concentration. Other parameters of significance are run time, filter flow rate, pH, temperature, reaction time, influent particle distribution, the calculated rate of change of flow, conductivity, alum, and silica content, in order of greatest sensitivity to least. The rate of change of flow was calculated by taking the slope of various lags of the filtration rate over a given time period. From these sensitivity values it is observed that physical parameters such as raw water concentration, run time, and flow rate, influence the initial filter ripening more than the chemical parameters. It is possible that the ANN has learned to assign alum and silica a low sensitivity since it is added in proportion to raw water turbidity. This may affect the optimization of alum and silicate dosage in order to minimize post-filter particle counts.

A sensitivity analysis, as shown in Figure 4.19 was conducted to illustrate the influences of flow rate and the rate of change in flow rate at a run time of 20 minutes, the approximate time in the filter run where the peak of the initial filter ripening occurs. An increase in flow rate results in an increased number of post-filter particle counts, but an increase in the rate of change of the flow rate (only for the first 20 minutes of the filter run) results in a decrease in particle counts. Since the flow rate parameter to the ANN is a measurement taken 10 minutes prior to the current time step, a “hold time” prior to the filter run which allows the backwash water to drain from the filter will reduce post-filter particle counts. In addition, it has been theorized (Van den Oever, 1998) (prior to obtaining the neural network results) that a large initial rate of change in flow is beneficial since it’s force will cause the filter bed to go into compression reducing the
Figure 4.19  Contour Plot of the 2 to 3 μm Channel Post-Filter Particle Counts at a Run Time of 20 minutes With Respect to Flow Rate and Rate of Change in Flow Rate- Britannia WTP

![Contour Plot of the 2 to 3 μm Channel Post-Filter Particle Counts at a Run Time of 20 minutes With Respect to Flow Rate and Rate of Change in Flow Rate- Britannia WTP](image)

Extrapolation beyond experience or this combination of silica, raw water concentration, and small filter detention time results in poor floc quality that does not adhere or settle within the filter.
pore space in the filter bed. Once the filter bed is compressed, a shorter period of initial filter ripening in addition to a smaller post-filter particle count measurement will result.

An investigation of the impact of various silica dosages on post-filter particle counts was conducted by request of the plant personnel to assess the impact of carry over of the silica onto the filter. It is not recommended that the results shown in Figure 4.20 be used for practical application since the ANN may not be able to distinguish between the raw water concentration and silica (and alum) dosage due to the application of a dosage proportional to the raw water particulate concentration. If continuous measurements of settled water turbidity are available, the design used for Manheim WTP is more suitable for optimizing the post-filter particle counts with respect to chemical dosages. The design for the Manheim WTP, however, assumes that there is no carry over of silica on to the filter. One approach to solve this problem is to measure the amount of silica entering the filter after sedimentation and using the residual silica as the input parameter to the ANN. Another method is to include another node for the settled water turbidity (if enough data are available) such that the ANN may learn the relationship between the chemical dosages added and the resulting settled water turbidity given a certain level of raw water particle counts.

4.3.2 Filter Ripening Curve and Breakthrough Network: Britannia WTP

As discussed earlier, it has been hypothesized that based on the principles of network design, that a neural network will be unable to learn the end of the filter-ripening curve for two reasons. Firstly, there are fewer data sets available to represent filter breakthrough than the filter ripening portion of the curve such that the algorithm of minimizing the mean square error will result in an ANN that is incapable of learning the
few incidents of filter breakthrough. As the ANN generalizes the data, it will not learn the underlying relationships well since it is essentially learning to model a horizontal line because there are comparatively fewer data sets available to represent the higher particle counts that occur during breakthrough for a 40000 minute filter run (compared to the 1600 minute run for the Manheim WTP). Secondly, the Britannia WTP backwashes the filters on the basis of maximum run time, and therefore the filter has not reached it’s maximum storage capacity such that the data do not include many data sets representing breakthrough.

The results for the filter ripening curve and breakthrough ANN for the Britannia WTP are demonstrated in Table 4.5 and Figure 4.21 for the 2 to 3 μm channel. As hypothesized, the ANN has not learned the underlying relationships adequately. The input layer design is the same as the one shown in Figure 4.16 except there is an additional node representing the cumulative filtrate volume that has passed into the filter from the beginning of the filter run.

4.4 Summary

This chapter demonstrates i) the feasibility of developing an ANN, ii) accuracy of the ANN, and iii) the applications of the ANN used to model settled water turbidity and post-filter particle count. Using the data obtained from the Manheim WTP, ANNs were developed that was capable of modelling settled water turbidity, channel particle counts in terms of percent of total counts, channel particle counts in terms of actual counts (count/mL), and the total particle count for the filter ripening curve. Using the data obtained from the Britannia WTP, an ANN was developed that was capable of modelling particle count for the initial filter ripening curve.
Table 4.5  Summary of Error Statistics for the Filter Ripening and Breakthrough Expert Network – Britannia WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>2-3 μm</th>
<th>3-5 μm</th>
<th>5-7 μm</th>
<th>7-10 μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>6.2</td>
<td>1.3</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.864</td>
<td>0.897</td>
<td>0.898</td>
<td>0.904</td>
</tr>
<tr>
<td>MAE</td>
<td>0.96</td>
<td>0.45</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>43.63</td>
<td>34.65</td>
<td>9.18</td>
<td>6.05</td>
</tr>
<tr>
<td>r</td>
<td>0.398</td>
<td>0.364</td>
<td>0.339</td>
<td>0.313</td>
</tr>
</tbody>
</table>

Figure 4.21  Actual (Historical) and Predicted Filter Ripening and Breakthrough Trend for the 2 to 3 μm Channel-Britannia WTP
Using a trial and error procedure, various forms of parameters influencing settled water turbidity and particle count were investigated as potential inputs to the ANN. In addition, various network parameters (previously identified in Section 3.3) were investigated.

It has been observed that unless the maximum turbidity criterion has been met as established by the utility, it is unlikely that filter breakthrough will occur as evidenced in the data set. Therefore, in order to model filter breakthrough, the plant must backwash primarily on the basis of turbidity so that filter breakthrough data are available in terms of particle count. Since post-filter particle counts generally increase prior to turbidity measurements (Crozes et al. 1994), the beginning of filter breakthrough may be modelled and optimization can therefore ensue.

The network design used for the Manheim WTP consists of two separate ANN; i) one that models settled water turbidity from raw water input parameters and ii) a second which models post-filter particle counts from the settled water turbidity parameters. This design assumes that there is no carry over of alum and polymer onto the filter. For the Britannia WTP, one ANN representing the plant from the plant intake to the post-filter particle count was used since silica has been observed to carry over onto the filters and since raw water particle count measurements are available. It is plausible that when using a design similar to Britannia that both the raw water and settled water particulate concentrations should be used as input to the ANN such that the network may learn how much alum and silica carry over onto the filter. In any case, both designs suffer from the fact that chemicals for coagulation or to aid coagulation are added in proportion to the raw water turbidity. This results in dependent ANN inputs. A neural network learns best
when correlation between input parameters is eliminated. In order to eliminate the correlation, experimental values out of the optimal alum dosage range may be added to the data set. It has been assumed that the continuously varying raw water turbidity measurements combined with the step-like trend of specified alum dosages will intrinsically contain enough data that are un-correlated for the ANN to distinguish between the parameters.

The results shown here illustrate the potential applications of a settled water turbidity and particle count network. The trend plots and test set results demonstrate the ANN’s ability to be used as a substitute settled water turbidimeter or particle counter, respectively. As long as the data are not extrapolating beyond its experience, in cases where the network predictions diverge from the actual measurements, the ANN indicates that there is a fault with the instrument it is mimicking or there is a disturbance upstream in the treatment train. The sensitivity analysis plots demonstrate the extent of which neural networks can be used as a powerful optimization tool. Not only may these plots be used to optimize particulate removal, they may be used to reduce cost. Once installed online, the ANN’s underlying relationships can be used to adjust the coagulant and coagulant aid dosage for optimal settled water turbidity removal, particle count (of which certain channels are surrogates to pathogens) removal, or minimum cost. It must be ensured that the ANN does not extrapolate beyond its range of experience in order for the network and its sensitivity plots to be valid. In order to ensure that extrapolation does not occur, the ANN must be adequately maintained. Maintenance consists of updating the training set with a broader range and combination of input parameters.
CHAPTER FIVE
CONCLUSIONS

Neural networks are particularly useful in the field of environmental engineering particulate removal processes due to the quantity of site-specific data available. Neural networks are the anticipated wave of the future that may eventually replace mathematical regression and physical models, expert systems, and knowledge based systems for modeling purposes. In this study, neural networks have successfully been applied to:

1. model the settled water turbidity of the Manheim Water Treatment Plant (WTP)
2. model the post-filter particle count in the form of total particle counts, channel particle counts in terms of percentage of total, and channel particle counts in terms of actual counts/mL for the Manheim WTP.
3. model the post-filter particle counts during initial ripening for the Britannia WTP.

The ability to model these particulate removal processes have been demonstrated to have applications for:

1. fault detection (provided the network is adequately maintained and is not extrapolating beyond it’s experience).
2. optimization of settled water turbidity and post-filter particle counts (which may be used as a surrogate for pathogens such as *Giardia* and *Cryptosporidium*) for various alum and polymer concentrations from the network’s underlying relationships.
3. optimization of the filtration rate and the rate of change of filter flow to reduce the post-filter particle counts during initial filter ripening.
4. optimization of cost for a given level of settled water turbidity or particle count.
5. a substitute particle counter or settled water turbidimeter.
CHAPTER SIX

RECOMMENDATIONS

The following are recommendations for further research in this area:

1. Extend the data set beyond the "optimal" chemical dosage at a given raw water turbidity by including lab data thereby un-correlating the two input parameters such that the network recognizes the difference between the chemical dosage and the raw water turbidity.

2. In order to ensure a high level of accuracy in the network's predictions, separate seasonal networks are recommended (if enough data is available). In addition, a separate network should be developed for events such as a snowmelt period.

3. In order for the network to predict under a broad range of upstream conditions and for changes in water quality over a period of years, the addition of current data is essential. Network maintenance must be conducted on a regular basis in addition to times where the network predictions diverge from the actual instrumental values. Maintenance will be required at a higher frequency at the beginning of the network implementation and the frequency of subsequent maintenance is dependent on plant operations and variations in the raw water quality.
CHAPTER SEVEN

REFERENCES


HARGESHEIMER, E.E.; LEWIS, C.M.; and YENTSCH, C.M. Evaluation of Particle Counting as a Measure of Treatment Plant Performance. AWWA Research Foundation and AWWA, Denver, CO. (1992).


The first step in data specification is to understand which parameters are required to model the particle quantity for that specific water treatment plant. A brief example of a sample questionnaire aiding data specification is demonstrated in here.

All data italicised is based on the correspondences with Mr. Doug Stendahl from the Manheim Treatment Plant located in Waterloo, Ontario, Canada. Some of these correspondences occurred via fax or phone at a later date.

I. Raw Water Quality

Please describe the water source (river or lake source, is there soil erosion control, lake turnover, etc.). *River Source: shallow, easily eroded clay and silt. Use 32 million gallon reservoir for presedimentation at river.*

Is there a significant quantity of algae in the water such that it can clog a filter or influence the removal of particulates? Please quantify the variation in algae over the data record. *Yes, significant quantity of algae. Report available. Does not hamper filtration.*

Is there a particle counter monitoring (parameter #1) the raw water? yes no

If so, does the raw water colour vary significantly (by 5 TCU)? yes no

*Colour varies on a monthly basis.*

Parameter #1: Raw Water Particulate Quantity

What kind of turbidimeter and/or particle counter is utilized? *This refers to the plant influent turbidimeter or particle counter. Manheim uses a Hach Model 1720C low range turbidimeter that has an accuracy of 2% for the 0 to 30 NTU range.*

*The most important parameters are raw water turbidity and temperature.*

Parameter #2,#3,#4,#5,#6: pH, Alkalinity, Hardness, Total Organic Carbon, Ionic Strength (as measured by Total Dissolved Solids)
For each of these parameters, please answer the following questions by filling in the table below.

1. Is the parameter measured?

2. In the operator's opinion, does the raw water parameter vary significantly as to impact the final particle count (can it not be assumed to remain constant or have a great predictive power for the final effluent particle count)? (yes, no)

3. What degree of confidence does the operator have in the parameter reading? (high, low, or medium)

4. What method or instrument is used to measure these parameters?

<table>
<thead>
<tr>
<th>Question #</th>
<th>pH</th>
<th>Alkalinity</th>
<th>Hardness</th>
<th>TOC</th>
<th>TDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>continuous</td>
<td>Weekly/monthly</td>
<td>monthly</td>
<td>Monthly</td>
<td>Monthly</td>
</tr>
<tr>
<td>2</td>
<td>constant</td>
<td>Weekly/monthly</td>
<td>monthly</td>
<td>Monthly</td>
<td>Monthly</td>
</tr>
<tr>
<td>3</td>
<td>medium</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>4</td>
<td>Rosemount Model 1054A accuracy ±0.01 pH units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The variations in alkalinity, hardness, TOC, and TDS occur seasonally or are episodic.

In general, these parameters are constant except during i.e. spring snow melt. Episodic events such as i.e. periods of runoff indicated by a decrease in alkalinity and other times in which an input parameter such as raw water turbidity jump or drop to an extreme value can be identified by changes in the filter ripening curve.

Later in the treatment train, is the pH adjusted? If so, where in the treatment train does this adjustment take place, with what chemical, and what is the adjusted pH?

The addition of alum reduces pH early in the process by 0.3 units.

Chlorination after filtration also reduces the pH by 0.2 units.

Although pH is an important parameter, it remains constant.
Parameter #7: Influent flow rate

Is the influent flow rate constant? If so, at what is the influent flow rate? No, it can be altered.

Is the flow evenly distributed throughout the plant? Yes

Are there any flow alterations, such as bypassing or partial flow treatment, in the treatment train? Fixed filter flows maintained.

Parameter #8: Season and Temperature

What are the dates of the start and finish of the data set collected? There are 4 data sets from January to September '97. Each of these data sets have a different quality:

1/5/97-1/27/97  22 days  poor quality
4/23/97-5/18/97  25 days  good quality
8/6/97-8/27/97  21 days  poor quality
9/13/97-10/10/97  27 days  medium quality

The majority of the winter subsample (the first subsample) applies a different type of coagulant than the other subsamples. This coagulant will not be applied in the future. Furthermore, during the summer subsample (the third subsample) there is a change in filter configuration. The flow rate is modified such that during backwash the blender flow is reduced to compensate for the filter out of service.

In the operator’s opinion, over the seasons the data was collected, can it be assumed that temperature is the only seasonal effect (i.e., surface runoff, etc.)? Yes  no

Alkalinity and pH during the spring are considered seasonal effects during the snow melt period. Alkalinity and pH are less dominant factors.
Has a record of day and night raw water temperatures been recorded? \(\checkmark\) yes \(\bigcirc\) no

*Temperature is measured continuously.*

II. Recycle

Does the plant have a recycle line? Yes but not used. \(\checkmark\) yes \(\bigcirc\) no

Is the recycle a fixed proportion of the flow through the plant? \(\checkmark\) yes \(\bigcirc\) no

From where and to where is there a recycle line? *Residue plant to raw water storage*

Are there any existing measurements of the particulates and cysts (parameter #9) in recycle water? \(\checkmark\) Not much. \(\bigcirc\) yes \(\bigcirc\) no

Does the flow (parameter #10) in the recycle line vary? \(\checkmark\) yes \(\bigcirc\) no

III. Preoxidation and Predisinfection

What type of preoxidant or predisinfectant (parameter #11) is used (i.e., ozone, chlorine, potassium permanganate, etc.)? *Ozonation after sedimentation*

What is the criterion for the determined required dosages of the preoxidant or predisinfectant? 3 log inactivation *Cryptosporidium*

IV. Coagulation

What type of coagulant (parameter #12) /polymer (parameter #13) is used? *alum, LT27A*

Is there lime addition (parameter #14)? \(\checkmark\) yes \(\bigcirc\) no

Is there soda ash addition (parameter #14)? \(\checkmark\) yes \(\bigcirc\) no
Is there bentonite addition (parameter #15)?

What are the criteria for the determined required dosages of the above parameters?

Turbidity of raw water. Jar testing experience

What other chemical substances (parameter #16) have been added (i.e. powder activated carbon, etc.)? Where and how much is added? *None are added.*

V. Flocculation and Rapid mix

Does the mixing energy (parameter #17) vary (i.e., is it variable speed drive)? *yes*

*Variable speed drive but use fixed low rate for flocculation and fixed high rate for rapid mix.*

If so, what criterion is used to alter the mixing energy?

Is there a flocculation unit process?

Is flocculation tapered?

How many compartments are there? *2 tanks in series, 8 tanks total*

VI. Sedimentation

What is the overflow rate, detention time, and depth of tank(s)? *Lamella plate settler.* 17 min. at rated design flow 440 L/s over flow rate = 428 m/hr at rated design flow. 2 tanks, 430 m³ each, 5.7 m to overflow.

What is the bulk underflow velocity (parameter #18)? Does this rate of sludge withdrawal remain constant? *Sludge withdrawal constant.* 54 m³/day.

VII. Filtration

What is the filter configuration (declining rate, constant rate, constant level)? If the filter configuration is not constant rate, please include a record of the filtration rate (parameter #19). *Filters individually controlled. Constant rate.*
What is the backwashing protocol? Is alum added to the backwash water? Is the backwash time fixed to a constant value? If they are not constant, please indicate this (parameter #20) and provide a record of its variation. As well, on what criterion is backwashing started? Please identify filter aids utilized (parameter #21). No alum is added to backwash. No filter aids used. Backwash criteria: Turbidity 0.1 NTU, headloss, or run time. Run time varies seasonally and is proportional to the plant flow rate, turbidity, or headloss.

What filter media is being used and what is the bed depth? 4 filters, two 56” GAC, two 48 “ anthracite and sand.

VIII. Particle count Data:

What is the frequency of particle count measurements? Every 10 minutes

How many replicates are taken at a given point in time? 1 minute count per 25 mL/min

In the operator’s opinion, what is the minimum interval of measurements (Δt) required for adequate resolution for the filter ripening curve? 10 minutes is a good resolution

Is the particle counter located at a point in which the filter to waste data is included in the data set? Yes.

What is the minimum annual filter run length? Rough estimate of 20 hours.

What is the maximum annual filter run length? Rough estimate of 60 hours.

What is the maximum filter run length in the data that has been collected? ~60 hrs.

(Please ensure that the time units provided are consistent.)

Give an estimate of the time required for filter ripening and filter to waste. _Filter ripening: 2hrs max, filter to waste: > 1hr estimate._

How many months of data are available for the data set? 4 subsamples. ~ 90 days
What is the reliability of the data? Has the data been analysed by the operator to ensure that data quality is adequate? During that process, has poor data been eliminated or replaced? *Data was quickly reviewed and unreliable data excluded.*

What type of particle counter is being used? *Hiac Royco*

How many of the particle counter’s channels are monitored? *8 channels, only 4 monitored*

How does the particle counter take its samples? *1 minute interval fixed 25 mL flow*

Does the particle counter take replicate samples? If so, how many? *No, continuous over minute.*

Specify the number of hours a day and days a week that data is collected and available.

*24 hr/day, 7 days/week.*

**Plant:**

What is the minimum, maximum, and average plant residence time (from the plant intake to the particle counter)? *1.1 hr, 4.2 hr, 2.1 hr. Currently both sides are operating.*

Is there a measure of influent filter particle count or turbidity, pH, temperature, ionic strength (in terms of Total Dissolved Solids), and major divalent cations. If so, these records may be required as a supplement. *Influent particle count unavailable, turbidity is available. Other parameters remain reasonably constant, either monthly or weekly measured.*

In addition, please provide a record of the run time and headloss of the filter being investigated.

The target values will be particle count and its distribution (using the power law slope coefficient). How consistent is the particle counter in its sampling and what degree of confidence does the operator have in the quality of the data? *Usually high,*
but in 1997 we have not given this a lot of attention since we are specifying a new particle count system for 1998. The data provided is rough and primarily useful for software development.

Please indicate as well any disturbances in the treatment train during the duration of data collection. Include the alarm record, if available. As well, include a diagrammatic of the plant layout with all unit processes and chemical additions throughout the treatment train. If the flow is unevenly distributed, please indicate it in the diagrammatic. Please specify the detention times of the individual unit processes. Include tracer tests, if available. In addition, a record of settled water turbidity and filter effluent turbidity may aid the accuracy of the net and should be included.
Volume of Water

1. Raw Water

   Hidden Valley High Lift Reservoir  $4 \times 36.32 \text{ ML} = 145.28 \text{ ML}$

   Bleams Road Main  $10 \text{ km} \times 1.19 \text{ m}^2 = 11.90 \text{ ML}$

   Raw Water Storage Tanks  $2 \times 2400 \text{ m}^3 = 4.80 \text{ ML}$

   $\text{TOTAL} = 161.98 \text{ ML}$

2. Plant Water

   Flocculation Tanks  $8 \times 212 \text{ m}^3 = 1.70 \text{ ML}$

   Sedimentation  $2 \times 430 \text{ m}^3 = 0.86 \text{ ML}$

   Ozonation  $2 \times 100 \text{ m}^3 = 0.20 \text{ ML}$

   Filtration  $4 \times 67.44 \text{ m}^2 \times 1.6 \text{ m} = 0.43 \text{ ML}$

   Treated Water Clear  $2 \times 7.64 \text{ ML} = 15.28 \text{ ML}$

   $\text{TOTAL} = 18.47 \text{ ML}$

3. Manheim Reservoir  $5 \text{ cells} = 194.50 \text{ ML}$

   $\text{TOTAL} = 284.84 \text{ ML}$
Process Schematic

From reservoir

From turbidimeter

Rapid mix

Flocculation tanks

Rapid mix

Sedimentation Tanks

Filter #1

Filter #2

Filter #3

Filter #4

To clearwells
In order to determine the number of parameters to be used in the neural net, the parameters that remain constant, are observed to have little variation throughout the duration of the record, or have not been recorded or measured by the WTP's personnel can be eliminated. The above questions are there to help identify the parameters of significance. Below is a table of all possible parameters. Consider particle count data used to measure particle concentration and distribution as 2 separate parameters. If turbidimeter data is only available for the raw water concentration, count that as one parameter. If the quantity of algae is recorded with time and it affects filtration, count it as a parameter.

<table>
<thead>
<tr>
<th>#</th>
<th>Parameter</th>
<th>Significant?</th>
<th>Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Influent particle concentration and distribution (turbidity or particle count) (2x)</td>
<td><em>Raw turbidity.</em>&lt;br&gt;<em>Significant</em></td>
<td><em>Y</em></td>
</tr>
<tr>
<td>2</td>
<td>pH</td>
<td><em>Little variation.</em>&lt;br&gt;<em>Seasonal</em></td>
<td><em>Y (data set available)</em></td>
</tr>
<tr>
<td>3</td>
<td>Alkalinity</td>
<td><em>Little variation.</em>&lt;br&gt;<em>Seasonal</em></td>
<td><em>Y (data set available)</em></td>
</tr>
<tr>
<td>4</td>
<td>Hardness</td>
<td><em>Little variation.</em>&lt;br&gt;<em>seasonal</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>5</td>
<td>Total Organic Carbon (TOC)</td>
<td><em>Little variation.</em>&lt;br&gt;<em>seasonal,</em>&lt;br&gt;<em>independent and constant</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>6</td>
<td>Ionic Strength (Total Dissolved Solids)</td>
<td><em>Little variation.</em>&lt;br&gt;<em>seasonal</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>7</td>
<td>Influent Flow rate</td>
<td><em>Significant,</em>&lt;br&gt;<em>varies seasonally</em></td>
<td><em>Y</em></td>
</tr>
<tr>
<td>8</td>
<td>Temperature</td>
<td><em>Significant</em></td>
<td><em>Y</em></td>
</tr>
<tr>
<td>9</td>
<td>Recycle flow</td>
<td><em>Not applicable</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>10</td>
<td>Recycle concentration and distribution</td>
<td><em>Not applicable</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>11</td>
<td>Preoxidant/ Predisinfectant</td>
<td><em>Ozone,</em>&lt;br&gt;<em>Little variation</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>12</td>
<td>Coagulant</td>
<td><em>Significant</em></td>
<td><em>Y (for settled water turbidity ANN)</em></td>
</tr>
<tr>
<td>13</td>
<td>Polymer</td>
<td><em>Significant</em></td>
<td><em>Y (for settled water turbidity ANN)</em></td>
</tr>
<tr>
<td>14</td>
<td>Lime/ Soda Ash</td>
<td><em>Not applicable</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15</td>
<td>Bentonite</td>
<td>Not applicable</td>
<td>N</td>
</tr>
<tr>
<td>16</td>
<td>Other chemical additions (also state the number of different chemicals that are added)</td>
<td>Not applicable</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>Mixing energy</td>
<td>No variation</td>
<td>N</td>
</tr>
<tr>
<td>18</td>
<td>Bulk underflow velocity</td>
<td>No variation</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>Filtration rate</td>
<td>Proportional to #7</td>
<td>Y (for particle count ANN)</td>
</tr>
<tr>
<td>20</td>
<td>Backwash variations</td>
<td>No. Automated, operator initiated.</td>
<td>N</td>
</tr>
<tr>
<td>21</td>
<td>Filter aids</td>
<td>Not applicable</td>
<td>N</td>
</tr>
<tr>
<td>22</td>
<td>Settled Water Turbidity</td>
<td>Significant.</td>
<td>Y (as target value for settled water turbidity net and input to particle count network)</td>
</tr>
<tr>
<td>23</td>
<td>Filter Particle count (Net Output) (2x)</td>
<td>Yes.</td>
<td>Y (as target value of particle count ANN)</td>
</tr>
</tbody>
</table>

Total number of records to retrieve from SCADA = 9 parameters

Minimum number of input parameters for the settled water turbidity net = 7 parameters

Minimum number of input parameters for the particle count net = 6 parameters recorded on the SCADA - 1 parameter for run time = 7 parameters.

Network may require more nodes than the minimum if a "lagged" design described in Section C.1 is applied. The various forms of expressing the input parameters are summarized in Appendix C. Each form of expression requires a different number of input nodes per input parameter.
APPENDIX B

DATA PRE-PROCESSING
Once the data have been specified (as demonstrated in Appendix A), the data are formatted such that each column in the spreadsheet represents a continuous input parameter. Data formatting consists of compounding the day by day data files obtained from the plant SCADA system. The format allows the input parameters to be directly accessed by the network from the SCADA with minimal transformations (discussed in Appendix C). Once the file is constructed and the statistics of the parameters are calculated (such as the mean, standard deviation, maximum, and minimum), the removal of conflicting, error-filled data is completed.

The criteria for removal of a potential outlier is as follows:

1. Modifications done to the plant that are not represented by the input parameters to the network are considered to be outside of the network’s experience and therefore neglected. For example, the settled water turbidity network does not include the winter season since a different coagulant is applied during that time.

2. Any abrupt changes are investigated by asking the personnel at the water treatment plant whether these changes are relevant. For example, the settled water turbidity network requires input for the raw water turbidity. The raw water turbidity values can change abruptly. After consulting with Doug Stendahl, it was determined that these points are likely due to instrument re-calibration and are therefore discarded from the data set, as demonstrated in Figure B.1 and B.2.

3. If data are missing, they may be interpolated or replaced by an experience-based or theoretical-based assumption. If this is not available, the point is discarded (Stein, 1993b). For example, for the settled
Figure B.1 Raw Data from the Manheim WTP: A- Spring Season Settled Water Turbidity, B- Fall Season Settled Water Turbidity
Figure B.2
Turbidity, B: Fall Season Settled Water Turbidity
A: Spring Season Settled Water
water turbidity network, an episode of snow-melt runoff results in a change in raw water pH, alkalinity, and raw water turbidity. According to Doug Stendahl, Water Quality Coordinator of the Manheim WTP, pH and alkalinity are not measured frequently enough to capture these changes (although they are known to occur). In order to ensure that there is sufficient data for modeling (as defined in Equation 3-1), the run off data were not discarded from the set eventhough in all other circumstances the run-off data are discarded.

4. One approach for discarding outliers is that any data falling outside of the 95% confidence interval of a normally distributed input parameter, may be neglected (Swingler, 1996). The distribution of the target parameter for the settled water turbidity network is illustrated in Figure B.3. Another method of describing the distribution is through statistics such as the maximum, minimum, mean, and standard deviation, such as Table 3.1 for Manheim WTP and Table B.1 for Britannia WTP. The approach, however, used here is that no data are discarded unless there is a logical explanation for doing so.

5. If two data sets have nearly the exact same input parameter values but completely different target values, the data set less likely to be accurate (i.e., it does not follow the underlying trend), is discarded. This is done to prevent confusion of the ANN during the training process.

6. If the underlying distribution of an input parameter is unclear due to abrupt fluctuations, these data points that interrupt the smooth trend of the input parameter are to be investigated for potential logical explanations to discard the data sets.
Figure B.3 Frequency Histograms of Settled Water Turbidity from the Raw Data of the Manheim WTP: A-Spring Season, B-Fall Season

The distribution of the settled water turbidity appears to be logarithmically distributed. Later in the training and validation process (described in Appendix D), a logarithmic transform of the settled water turbidity may result in a smaller network error (Swingler, 1996).
Table B.1  Statistical Properties of Input Parameters obtained from the Britannia Water Treatment Plant over the Initial Ripening Period

<table>
<thead>
<tr>
<th>Property</th>
<th>pH</th>
<th>Temperature (°C)</th>
<th>Conductivity (μm/cm)</th>
<th>Raw Water: logA</th>
<th>Raw Water: beta</th>
<th>Detention Time (D/ML)</th>
<th>Alum (mg/L)</th>
<th>Silica (mg/L)</th>
<th>Flow rate (L/s)</th>
<th>Run Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>6.30</td>
<td>19.0</td>
<td>85.5</td>
<td>5.36</td>
<td>3.73</td>
<td>0.01054</td>
<td>34.0</td>
<td>2.50</td>
<td>19.1</td>
<td>107.1</td>
</tr>
<tr>
<td>Minimum</td>
<td>5.73</td>
<td>0.3</td>
<td>52.0</td>
<td>4.65</td>
<td>2.70</td>
<td>0.00322</td>
<td>28.0</td>
<td>0.75</td>
<td>0.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Mean</td>
<td>5.99</td>
<td>4.5</td>
<td>67.3</td>
<td>4.89</td>
<td>3.37</td>
<td>0.00549</td>
<td>30.5</td>
<td>1.35</td>
<td>8.9</td>
<td>48.0</td>
</tr>
</tbody>
</table>
The data quality must be checked for an even distribution of training examples. For example, if 1000 of the points in your data set has input parameter #1 between 1 and 10 and 25 points between 100 and 1000, it may be best to model these 25 points with a different network used for extreme circumstance to obtain an even distribution of the input parameters. Trend plots, as illustrated in Figure B.4 and B.5 for various input parameters for the Manheim plant, aid the visualisation of the frequency of values of an input parameter. For example, the raw water turbidity trend for the spring season indicates that there is a run off period occurring at approximately 5/13/97. In cases of spring run off periods, it may be best to design a separate network were there are not many data sets available to describe these events. If one of the parameters remain constant throughout the training set, the network does not require a node for this constant parameter. For example, it has been assumed that the Total Organic Carbon (TOC) content is constant throughout the data set and so therefore no input node represents the TOC of the raw water. During the training process, the overall sensitivity is calculated for the best final architecture using NeuroSolutions®. The input parameter that has the smallest sensitivity, which is approaching a value of zero, may be discarded as an input to the network. The network is then re-trained to observe whether that parameter was necessary.

The size of a network determines the amount of data required. The following formulae and guidelines (Swingler, 1996; Bishop, 1995) can be used to determine the quantity of data sets required.

- The number of hidden units must be less than or equal to (twice the number of input units +1) (Swingler, 1996).
Figure B.4  Trend of Input Parameters for the Manheim WTP Fall Season Settled Water Turbidity Network: A- pH, B- Temperature
Figure B.4  Trend of Input Parameters for the Manheim WTP Fall Season Settled Water Turbidity Network: C- Raw Water Turbidity, D- Alum Dosage
Figure B.4  Trend of Input Parameters for the Manheim WTP Fall Season Settled Water Turbidity Network: E- Polymer Dosage, F- Plant Flow Rate

**E**

![Graph showing the trend of Polymer dosage from 9/13 to 10/10 in 1997](image)

**F**

![Graph showing the trend of Plant Flow Rate from 9/13 to 10/9 in 1997](image)
Figure B.5  Trend of Input Parameters to the Spring Settled Water Turbidity Network for Manheim WTP: A-pH, B-Temperature

A

pH

8.4
8.3
8.2
8.1
8.0
7.9
7.8
7.7


Date (1997)

B

Temperature (°C)

12
10
8
6
4
2
0


Date (1997)
Figure B.5  Trend of Input Parameters to the Spring Settled Water Turbidity Network for Manheim WTP: C-Raw Water Turbidity, D-Alum Dosage
Figure B.5  Trend of Input Parameters to the Spring Settled Water Turbidity Network for Manheim WTP: E- Polymer Dosage, F- Flow Rate
• The Baum and Haussler (1989) formula allows for an approximation of network size based on the target error. Assuming the error limit, \( \varepsilon \), is restricted such that \( 0 < \varepsilon \leq 1/8 \), the network trains to achieve an error of \( \varepsilon/2 \) if \( d \geq w/\varepsilon \), where \( d \) is the size of the training set and \( w \) is the number of weights. This formula is commonly used in industry.

• A network that is a feature extractor, otherwise stated as a non-linear regression type model that learns the underlying relationships of the data that contains intrinsic properties, the network has fewer hidden units than input units.

• The fewer the number of nodes in the network, the greater are its generalization capabilities. Furthermore, run time is shortened.

• A good starting point for the number of hidden units is 75% of the input layer (Bailey and Thompson, 1990b) when training the network.

• For monitoring network learning capabilities, a test set and validation set are required. The validation set is a minimum of 20% of the training set. The test set is approximately 10-30% of the data. As well, one must consider that approximately 20% of the data will be discarded due to outliers or error in the data.

**Example B.1: Manheim Water Treatment Plant Settled Water Network**

• Test and validation sets are set-aside after outlier removal, each set is 25% of the original data. This leaves 50% of the pre-processed data for training purposes.

• To calculate the amount of data, an error limit must be set. The error limit, \( \varepsilon \), was chosen to be 0.1. Therefore, the number of training sets required \( (d) \) is 10 times the number of network weights \( (w) \).
The minimum number of parameters has been specified in Appendix A. Suppose the number of input units for the settled water network is 12 once accounting for network design, such as the additional lagged raw water turbidity nodes (as described in Section C.1). The number of units exceeds the number of parameters since prior values have been inputted into the network to account for trend and the travel time of a slug input.

A hidden layer of 12 units was assumed. This is more conservative than the initial "75% of the input layer" approximation as an estimation of the number of data sets required since it is the architecture with the maximum number of hidden nodes that is capable generalization.

The output contains one unit, namely the settled water turbidity.

The number of weights required is \(12^2 + 12 = 156\) weights.

Given an error limit of 0.1, 1560 data sets are required for training.

Given that the training set is 50% of the original data set, the set must contain \(1560 \times 2\) or 3120 data sets.

Allowing 20% of the data to be removed due to outliers or errors, \(3120 / 0.8 = 3900\) data sets are required.

Therefore, if one wants to restrict the number of input nodes knowing that a given amount of data, \(N\), is available,

\[
\frac{(N \times 0.8 / 2)}{10} = I^2 + I
\]

The quadratic formula is used to solve for \(I\), the maximum number of input units allowed.
Example B.2: Calculating the number of nodes for the Manheim WTP particle count network

Assume that no\# of hiddens = no\# of inputs

Use $\epsilon = 0.1$.

Manheim has 7924 sets total that are suitable for simulation. Accounting for possible errors in the data and the validation and test set, the number of input nodes allowed are:

For one output node:

$I^2 + I = 7924 \times 0.8/20 \quad : \quad I = 17$

For two output nodes:

$I^2 + 2I = 7924 \times 0.8/20 \quad : \quad I = 16$

<table>
<thead>
<tr>
<th>No# of output units</th>
<th>Seasonal network (1/2 data set)</th>
<th>Whole data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>16</td>
</tr>
</tbody>
</table>

The input parameters to the network include filtration rate, headloss, run time, alum, polymer, pH, alkalinity, temperature, turbidity, and filter ripening trend. The network design is restricted such that the maximum number of input nodes should not be exceeded.
To demonstrate design decisions, the particle counter network for the Manheim WTP is used as an example. The proposed general design is illustrated in Figure C.1. The general design encompasses all potential designs that will be investigated prior to the training validation procedure. The forms of the input parameters, which may vary for a different plant, are defined through a trial and error procedure found in Appendix D. Various network architectures that may be applied to model time trend data are summarized in Table C.1. The network design selected for the Manheim Water Treatment Plant (WTP) is a combination of a simplified time delay neural network (TDNN) and an architecture that contains a run time node. The simplified time delay neural network contains "lags" only for parameters that are not constant over the window size being investigated.

C.1 Calculating Lag to Account for Detention Time

The data from Manheim for the design of the particle count network are used as an example of calculating the detention time or node "lag".

Given:

Filter volume = 0.1075 ML

Minimum flow rate during filter run over data set provided = 60 L/s

Maximum flow rate during filter run over data set provided = 120 L/s

@ flow rate of 60 L/s:

- detention time = 0.1075 * 10^6 L / ((60 L/s)* 60 s/min) = 30 minutes

- Since the data are taken at a frequency of every 10 minutes, the lag required is 30/10=3. (In other words, the input parameter used is 3 rows of data prior to the current row in an Excel® spreadsheet.)
Proposed General Design for Particle Count Neural Network - Design #2

- Unknown number of hidden units
- Number of hidden units ≤ 2*(number of input units + 1)
- For generalization: Number of hidden units ≤ number of input units

Temperature, pH, Alkalinity, Influent Particulate Quantity, Filtration Rate, Filter Storage, Run Time

Potential formats of parameter described in Table 3.5 and Section C.2
Potential formats of parameter described in Table 3.4 and Section C.3
Potential formats of parameter described in Section 3.2.2 and Section C.4
Potential formats of parameter described in Table 3.3 and Section C.5
As described in Figure 3.3. This design was applied to the data from the Manheim WTP. The input parameters listed here are those that have been identified for the Manheim WTP. Other water treatment facilities may require other input parameters, for example change in flow rate, as described in Sections C.2-C.5.

The minimum number of input nodes to the particle count network is 7 (as specified in Appendix A). The form of expressing the various input parameters dictates the number of nodes in the design. For example, if the change of filtration rate is a significant factor, it may be expressed as either i) one node representing (Q(t) - Q(t-1))/sampling frequency, ii) two nodes representing the current flow rate and the previous flow rate, or iii) three nodes representing Q(t), Q(t-1), and Q(t-2), and so forth.
<table>
<thead>
<tr>
<th>Architecture</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| Time delay neural network (TDNN)  | • Can easily input vectors at an unequal resolution (that are not equidistant in time)  
• Good for very large networks  
• Easy to add extra nodes when required  
• Better than recurrent network when there is a complex network interaction  
• A strict TDNN is not required. If a parameter does not vary, the number of nodes required to model the parameter may be less than the window size | • Requires a large amount of input nodes due to the explicit input required for the input vectors  
• Long run time required due to higher complexity of the network  
• Quantity of data required affected by outlier removal in the lags  
• Best if have small window with small number of input parameters  
• Window size and resolution must be found using a trial and error process |
| Recurrent network (RN)            | • Non-independent input channels  
• Internal representation allows for potentially easier learning | • Cannot randomize the sequence of data sets due to short term memory of the network  
• Only the temporal parameters have context/recurrent nodes  
• Requires adjacent input vectors  
• The number of hidden and input units are difficult to approximate |
| Time node                         | • Requires less input units than the other architectures thereby requiring less data for design and allowing for a shorter run time  
• Easy to analyze the effects of time as a parameter in itself | • Cannot deal well with variations in the filter ripening curve such as attenuation since it does not account for previous values  
• The variation of filter run times will not allow for an equal number of examples, especially near the programmer's specified maximum filter run time  
• Inaccurate since detention time requires prior values |
@ flow rate of 120 L/s:

- detention time = 15 minutes; lag = 1.5

Raw water turbidity lags, however, were not applied in the final network.

Due to dispersion in the filter (since ideal conditions are not realistic), the range of lags to be investigated (or rather inputted intrinsically into the network architecture), is not strictly the theoretical lags mentioned above. If tracer tests are available and the flow rate is maintained constant through the filter, the peak of the E curve may be the only lag required. Otherwise, a trial and error process to find the peak may be required. It would be best if a module was built to identify the peak of the E curve under various conditions of flow such that the input to the network is always under conditions of maximum throughput. For the Manheim WTP settled water turbidity network, three nodes were used to identify the lag of the input to the filter since 3 points are the minimum required to define a curve. The lags used are labeled lags 1, 2, 3 referring to a detention time of 10, 20 and 30 minutes respectively.

Data that do not have a constant frequency (i.e., sometimes the data sets are taken 5 minutes apart, 4 minutes, or even 10 minutes apart) cannot be lagged by inserting cells at the top of column of the parameter in the spreadsheet in order to account for the detention time implicit in the lag. The columns of the spreadsheet represent the input parameter and the rows represent the time at which the measurement was taken. If lags are required, the VLOOKUP command in Excel enables the user to lookup to value of a given input parameter at the detention time you are investigating and therefore provides an approximation of the lag value. For example, if the detention time is five minutes, the
VLOOKUP command looks up the raw water turbidity value in the spreadsheet row representing the detention time equivalent to five minutes prior to the current time.

C.2 Forms of Expressing Post-Filter Particle Count

The forms of describing post-filter particle counts are summarized in Table 3.5. The post-filter particle count of the previous time step may be used as a last resort to gain a smaller network error as done by time series simulation models.

The power law slope and intercept may be calculated using equation (3-11). Considering that it is unlikely for more than 2 particle counter channels to have all their counts above 10 counts/mL (as demonstrated by the Manheim and Britannia WTP data), it is unlikely that this means of expressing particle count may be utilized.

The natural logarithm transformation, percent of total counts transformation (in either number or volume format), and the change in particle counts over one time step reduces the range of the post-filter particle counts and alters the distribution such that the network may be able to learn the underlying relationships better. The natural logarithmic transformation is calculated by

\[ y = \ln(x) \]  

where \( x \) the post-filter particle counts in counts/mL from the raw data set, and \( y \) is the transformed equivalent. The natural logarithm transform suffers from pitfalls such as under-approximation of higher values of the target range. The percentage format is generally the best format since it distributes the data more evenly and compresses the range as well. The change in particle count may be the best format depending on the rapidity of particle count changes and the network design. For example, if there are abrupt changes in flow rate resulting in a rapid increase in post-filter particle counts, this
format may enable the network to learn this relationship more effectively. Furthermore, if previous values of post-filter particle counts are used as input to the network to learn the filter ripening curve more effectively, the change in particle count may be the best format due to its comparatively lower correlation to this type of input parameter and the easy calculations (previous count + change in count) required to obtain the post-filter particle counts in terms of counts/mL.

In addition to the various forms of expressing post-filter particle count, the number of output nodes in the architecture is defined in the design step. The Baum and Haussler criteria (equation 3-1) for determining the amount of data required for a specified number of weights was derived for an architecture with only one input unit. It is assumed here that the relationship holds true for more than one output unit. If sufficient data is available, a network that contains an output node for each channel being modeled will take less time to train. It is possible that a network that contains an output node for each particle counter channel may be more accurate than individual networks per channel (even though the data requirements are more stringent). Due to the mutual hidden layer shared by all particle count channels and the network algorithm that minimizes the average validation error of all channel sizes, the network may be able to learn the particle size distribution changes with increasing run time, especially if particle counts are expressed in percent of total count format.

C.3 Forms of Expressing the Influent Particulate Quantity

The influent particle quantities can be expressed as raw water turbidity or particle count, settled water turbidity or particle count, or both, as summarized in Table 3.4. The influent particle quantity is described here in terms of settled water turbidity (since not
enough data were available to describe it in terms of raw water turbidity). The forms of expressing settled water turbidity is the cumulative settled water turbidity in mg/L or milligrams and the actual settled water turbidity in either mg/L or milligrams (at the appropriate filter lag). The form selected depends on the dominant mechanism in the filter. For example, if avalanche effects are occurring at the end of the filter run during breakthrough or if the initial filter ripening is being investigated, the post-filter particle count network may be best described if a settled water turbidity node is available. During the filter run, the cumulative settled water turbidity is important in order to keep track of the accumulation within the filter such that the network can identify when filter breakthrough begins.

The particle count network for the Britannia plant models the post-filter particle count from the plant intake. The raw water particle counts are provided instead of raw water turbidity. Due to the restrictions on network design with respect to the amount of data available, the power law was utilized since it reduces all particle channels into two input parameters. The power law relationship is suitable for the raw water particle counts since all channels have counts greater than 10 counts/mL.

C.4 Forms of Expressing Filter Flow Rate

The various expressions of filter flow rate, such as flow rate, interstitial velocity, the inverse of the flow rate, the square of the flow rate, and so forth, is dependent on the dominant attachment and detachment mechanisms, as described in section 3.2.2 and Figure 2.8. Notice that the filter velocity can be expressed without accounting for filter area in terms of filter flow rate since the network scales the input parameters and therefore the constant filter area is not of importance. For example, the inverse of the
flow rate is proportional to the reaction/detention time and can be entered into the network without multiplying by the constant volume of the unit process.

In addition to filter flow rate, rapid changes in flow rate affect post-filter particle count. Filtration rates always change at the beginning of the filter run as the filter is put back on-line. If filter-to-waste is practiced and there are no significant changes in filter flow rate, this parameter does not need to be included into the network. If, however, as for the Britannia network, filter to waste is not practiced, the change in flow rate is a significant parameter.

The change in flow rate can be expressed in the network by the lags corresponding to the range of filter detention times, the change in filtration rate from the previous time step (provided that the previous flow rate is provided), the slope of the filtration rate lags, the average velocity over the filter detention time, or the cumulative filtrate volume to the network. If there is a limited amount of data available, the latter three measures require fewer input nodes. If the underlying relationships are to be investigated, the best format for the change in filter flow rate is the slope of the filtration rate such that one node represents the change in filter flow rate. The cumulative filtrate volume may be used to calculate the time averaged filtration rate by dividing by the run time. Furthermore, this form of expressing the quantity of filtrate that has passed through the filter since the beginning of the filter run may be used instead of the cumulative settled water turbidity node (in terms of milligrams) when the post-filter particle counts are modeled for the plant intake.

Another input parameter that may affect the initial filter ripening is the cumulative filtrate volume or cumulative settled water turbidity form the previous filter run in
combination with a measure of backwash efficiency. Since backwash efficiency is difficult to describe if the protocol changes each time backwashing occurs, these parameters are neglected for the Manheim WTP.

C.5 Forms of Expressing Storage

It is useful to have storage as an input parameter such that the network can understand that breakthrough occurs once the maximum storage capacity is reached. A storage node is not required if modeling initial filter ripening due to the negligible changes in storage. Although the storage node aids modeling, the amount of particulates stored within the filter is a function of the settled water turbidity, run time, and flow rate as well as characteristics of the floc due to coagulation-flocculation-sedimentation processes. The various forms of expressing storage are summarized in Table 3.3 in terms of headloss, hydraulic permeability, porosity, inverse of permeability, inverse of porosity, and the void ratio. Darcy's Law is defined as

\[ q = k^* h_L \]  \hspace{1cm} (C-2)

where \( q \) is the velocity, \( k \) is the coefficient of permeability, and \( h_L \) is the hydraulic gradient. The hydraulic gradient is defined as

\[ h_L = \Delta H / \Delta L \]  \hspace{1cm} (C-3)

or the difference in pressure head over the flow path length. In order to calculate the coefficient of permeability, simply isolate for \( k \).

\[ k = q / h_L \]  \hspace{1cm} (C-4)

The coefficient of permeability is affected by temperature and requires a correction factor for temperatures other than 20°C (Whitlow, 1983).
The coefficient of permeability is empirically related to the porosity by

\[ k \propto \varepsilon^2 \quad k \propto D_{10} \]

where \( \varepsilon \) is the void ratio, or volume of voids/volume of solids and \( D_{10} \) is the effective size. In order to quantify the porosity (\( \eta \)), or the volume of voids/total volume (\( = \varepsilon/(1+\varepsilon) \)), which expresses the amount of storage in filter, one must first take the square root of \( k \) to find a value proportional to the void ratio. The total volume is taken to be equal to one. Once this value is calculated, however, one cannot determine the porosity since this value is only proportional to, not equal to, the void ratio. In attempt to find the porosity of the filter, the Hazen formula may provide some aid.

\[ k = C_H D_{10}^2 \quad (C-5) \]

where \( C_H \) is an experimental coefficient dependent on the nature of the soil. Given that the effective size of the filter is 1.0 to 1.1 mm with a uniformity coefficient of 1.5 mm maximum, the value of the experimental coefficient is within the range of 8 to 12 (Whitlow, 1983). The initial \( k \) value then ranges from 4.8 m/min to 8.712 m/min. From the data provided, the maximum \( k \) value is approximately 5 m/min. Therefore, if

\[ k = J^* \varepsilon^2 \quad (C-6) \]

The value of \( J \) ranges between 0.96 and 1.64. The volume of the solids is then calculated by \( 1/(1+\sqrt{(k/J)}) \).
APPENDIX D

TRAINING AND VALIDATION PROCEDURE
The following section describes the procedure that was employed in obtaining the final network. The settled water turbidity net for the Manheim Water Treatment Plant (WTP) is used as an example of the main concepts applied in the trial and error approach to training and validation. A cross-section of three trials have been selected from the process, which best summarizes the concepts employed.

The first trial has the following input parameters to the network: season (described in Appendix C), pH, temperature, alkalinity, plant flow rate, plant flow rate in the previous time step (also expressed as plant flow rate at a lag of one), alum dosage in units of mg/s (=mg/L*L/s), polymer dosage (mg/s), and the natural logarithm of raw water turbidity at a lag of 50 minutes, 100 minutes, and 200 minutes. The calculations required to obtain these parameters, given the raw data, have been described in Appendix C. There are a total of 12 input nodes to the network. The raw water turbidity nodes enter the network as its natural logarithmic form since the distribution of raw water turbidity is logarithmic and it is hypothesized at this point in the training process that this form of expression is best. Furthermore, the alum and polymer dosages are described in units of mg/s since it is hypothesized that the discrete expression of mg/L that remains constant for days at a time will not allow the network to truly understand the importance of this parameter. The network contains the data from both the spring and fall seasons.

Network parameters including the momentum and learning rate parameter, scaling, and transfer functions are defined in Table D.1. The learning algorithm that the neural network utilizes for the training and validating procedure is momentum descent. The momentum and learning rate parameters are set to 0.7 and 0.01 respectively (Bishop,
Table D.1  Summary of Network Parameters for the First Trial

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum constant</td>
<td>0.7</td>
</tr>
<tr>
<td>Learning rate constant</td>
<td>0.01</td>
</tr>
<tr>
<td>Network inputs</td>
<td></td>
</tr>
<tr>
<td>• Season</td>
<td></td>
</tr>
<tr>
<td>• pH</td>
<td></td>
</tr>
<tr>
<td>• Temperature</td>
<td></td>
</tr>
<tr>
<td>• Alkalinity</td>
<td></td>
</tr>
<tr>
<td>• ln(raw water turbidity - 50)</td>
<td></td>
</tr>
<tr>
<td>• ln(raw water turbidity - 10)</td>
<td></td>
</tr>
<tr>
<td>• ln raw water turbidity - 20)</td>
<td></td>
</tr>
<tr>
<td>• Plant flow rate</td>
<td></td>
</tr>
<tr>
<td>• Plant flow rate - 1</td>
<td></td>
</tr>
<tr>
<td>• Alum * flow rate</td>
<td></td>
</tr>
<tr>
<td>• Polymer *flow rate</td>
<td></td>
</tr>
<tr>
<td>Number of epochs/run</td>
<td>300</td>
</tr>
<tr>
<td>Input and output scale</td>
<td>-1,+1</td>
</tr>
<tr>
<td></td>
<td>-2,+2</td>
</tr>
<tr>
<td>Noise addition</td>
<td>None</td>
</tr>
<tr>
<td>Convergence criteria</td>
<td>300 epochs/run</td>
</tr>
<tr>
<td>Transfer function used</td>
<td>tanh for hidden; Linear for output</td>
</tr>
</tbody>
</table>
1995; Swingler, 1996). The convergence criteria, which specifies when the network is to stop training, is either the point of minimum validation error (known as "early stopping" (Bishop, 1995)) or surpassing the maximum number of epochs per run, whichever happens first. The maximum number of epochs per run has been set to 300 epochs. A tanh unit was applied to the hidden layer and a linear unit for the output node (as described in section 3.3.3) since this network's function is that of regression. A tanh unit requires that the input units be scaled between −1 and +1 and the linear output unit allows for a user defined scale for the target values (NeuroSolutions® manual, 1996). The scale used for the target settled water turbidity values is −2 to +2. The target scale was selected to be double the range of the input values such that the network will place heavier weights on those parameters of importance and therefore potentially improve the accuracy of the network. Caution is to be exercised when selecting the target range since too large a range (greater than five to ten-fold of the input range) will result in the network weights exploding to exceptionally large values rendering the network useless.

The network training and validation steps occur in parallel. At the end of each epoch, the network runs the validation set through the network and records the network error. During the training procedure, the training and validation error decreases. Several architectures are investigated for each trial. In this trial, architectures with 7 through 11 hidden nodes have been investigated and their minimum validation error recorded, as demonstrated in Table D.2. The network weights are initiated using a randomly selected seed value. More than one seed value, from now on referred to as a replicate, is investigated such that different start points on the error surface will increase the chances of the network to discover the global minimum rather than a local minimum. Three
Table D.2  Minimum Average Validation Cost of Architectures Investigated for the First Trial

<table>
<thead>
<tr>
<th># hiddens</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed #1</td>
<td>0.0349378</td>
<td>0.0365944</td>
<td>0.0351884</td>
<td>0.0338472</td>
<td>0.0398006</td>
</tr>
<tr>
<td>Seed #2</td>
<td>0.034519</td>
<td>0.0365497</td>
<td>0.0332878</td>
<td>0.033908</td>
<td>0.0332608</td>
</tr>
<tr>
<td>Seed #3</td>
<td>0.035388</td>
<td>0.0344522</td>
<td>0.0356239</td>
<td>0.0339741</td>
<td>0.0332348</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.034519</td>
<td>0.0344522</td>
<td>0.0332878</td>
<td>0.0338472</td>
<td>0.0332348</td>
</tr>
</tbody>
</table>

Optimal Architecture
replicates were used for the settled water turbidity network (Carpenter and Hoffman, 1995). Each replicate for each architecture investigated is trained for 300 epochs or until the validation error begins to increase. The minimum validation error of the run is tabulated, as demonstrated in Table D.2. The minimum of all replicates investigated for each architecture is then identified. If the network has learned well and sufficient numbers of replicates have been used, the overall minimum validation error for a given network architecture is the global minimum. The architecture with the minimum overall validation error is the architecture that has best learned the underlying relationships. The architectures with fewer nodes than the optimal network does not learn the underlying relationships adequately, and the architectures with more nodes memorize the underlying relationships. This network is then evaluated with a test set to determine the degree of accuracy of the network in predicting settled water turbidities for an unseen data set that is independent of the training-validation procedure.

The optimal network is evaluated based on the error summary tables produced by NeuroSolutions®, correlation plot of the actual and predicted settled water turbidity, and error histograms. The summary table of error statistics, as seen in Table D.3, contains the mean square error (MSE), the nominal mean square error (NMSE), the mean absolute error (MAE), the minimum absolute error, the maximum absolute error, and the linear correlation coefficient. If the network is consistent in its representation of the data of the independent test set, the correlation coefficient of the correlation plot, as demonstrated in Figure D.1, approaches the value of 1. Furthermore, the network correlation plot ideally has a slope of one and intercept of zero. The correlation plot enables the programmer to
Table D.3  Test Results for Trial #1: Summary of Error Statistics for Settled Water Turbidity Network- Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0292</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.1911</td>
</tr>
<tr>
<td>MAE</td>
<td>0.1204</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0001</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>1.1078</td>
</tr>
<tr>
<td>r</td>
<td>0.9004</td>
</tr>
</tbody>
</table>
Figure D.1: Test Results for Manipulated WTP Sediment Water Turbidity Net Trial #1.
identify ranges where the network has difficulty learning in addition to outliers. The error histogram of the test set of a healthy network has a mean at zero and takes on the same distribution as the noise in the data (Swingler, 1996). Error is defined as the difference between the output (predicted) and target (actual) values. A positive error on the histogram denotes that the predicted value is greater than the actual value. The error histogram shown here is skewed in the positive direction and has a mean greater than zero.

In order to better understand what the network has learned, and therefore be able to make the appropriate amends to the input layer, the time trend of the complete data set and the network's predicted values over the complete set is constructed, as shown in Figure D.2. The parameters that have the dominant effects on the results are identified by comparing the predicted trend with the trends of the input parameters shown in Appendix B, or correlating the error from each data set with each input parameter. From the trend plots, it has been deduced that one definite source of error is due to the expression of alum and polymer in units of milligrams per second since the drop in settled water turbidity occurs at the same time a drop in the chemicals added in mg/s. Another potential error is that the network gives too much importance to the raw water turbidity parameter. It is possible that the wrong lags have been used. In order to correct this potential error, all lags corresponding to the investigated flow rates have been added in. After training in trial two, one will be able to easily identify by calculating the overall sensitivity (using NeuroSolutions®) and eliminating the lags with low sensitivity. Although two changes have been suggested here at the same time for demonstration purposes, in practice amends are made one at a time.
Figure D.2: Comparison of Actual and Predicted Values for the First Trial: Settled Water Turbidity

- C: Summer Season
- B: Spring Season
- A: Winter Season

Legend:
- Actual
- Predicted

Notes:
- Presence of two flow nodes
- Excessed backwash
The next trial changes the units of alum and polymer to mg/L, omits the current flow rate node, and adds raw water turbidity lags. Only one flow rate parameter at a lag of 10 minutes is used such that the network can better understand that for a given flow rate there is an appropriate raw water travel time to the settled water turbidimeter. The raw water turbidity nodes are still expressed in terms of the natural logarithm. The lags investigated include 50, 60, 70, 80, 90, 100, 140, and 210 minutes. After training, the validation errors are recorded for the architectures investigated, as demonstrated by Table D.4. As already demonstrated, the best architecture is then tested using an independent test set and the results are analyzed using a summary of error statistics, correlation plot, and error histogram as demonstrated in Table D.5 and Figure D.3. The correlation plot coefficients indicate a lower slope, higher intercept, and smaller correlation than the first trial. The network is potentially not learning as well due to its dependence on the amount of data available and the number of nodes required for modelling. The error histogram is again skewed to the right. By plotting the time trend plots for the spring, summer, and fall season, as illustrated in Figure D.4, the reason for the poor learning ability of the network is identified. Consistently throughout the previous trials (not illustrated in this cross-section) the summer season is poorly predicted. It appears that the network is predicting the settled water turbidity earlier than it actually occurs. When investigating the data set, it was discovered that the flow rate drops to zero during the summer. If there is no flow in the plant, a slug input entering the treatment plant is detained for a longer period of time than the calculated detention time and therefore the network is incapable of learning the underlying relationships. Furthermore, it has been mentioned by Doug Stendahl of the Manheim
More replicates or a longer run may be needed to get a more accurate view of which architecture is correct.

Table D.4\textsuperscript{b}  Minimum Average Validation Cost of Architectures Investigated for the Second Trial

<table>
<thead>
<tr>
<th># hidden</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed #1</td>
<td>0.0480957</td>
<td>0.0608591</td>
<td>0.0435511</td>
<td>0.0426157</td>
<td>0.0452453</td>
<td>0.0437055</td>
<td>0.0417843</td>
</tr>
<tr>
<td>Seed #2</td>
<td>0.0423061</td>
<td>0.0458622</td>
<td>0.0437849</td>
<td>0.041843</td>
<td>0.0412213</td>
<td>0.0428316</td>
<td>0.0424351</td>
</tr>
<tr>
<td>Seed #3</td>
<td>0.044931</td>
<td>0.0461388</td>
<td>0.0427642</td>
<td>0.0445228</td>
<td>0.0437252</td>
<td>0.0437024</td>
<td>0.043674</td>
</tr>
<tr>
<td>Min.</td>
<td>0.0423061 \textsuperscript{a}</td>
<td>0.0461388</td>
<td>0.0427642</td>
<td>0.041843</td>
<td>0.0412213</td>
<td>0.0428316</td>
<td>0.0417843</td>
</tr>
</tbody>
</table>

\textsuperscript{a} It is possible that there may be a better architecture with fewer hidden units than 8.

\textsuperscript{b} More replicates or a longer run may be needed to get a more accurate view of which architecture is correct.

Table D.5  Test Set Results: Summary of Error Statistics for the Second Trial

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0339</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.2296</td>
</tr>
<tr>
<td>MAE</td>
<td>0.1262</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0000</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>1.3990</td>
</tr>
<tr>
<td>$r$</td>
<td>0.8779</td>
</tr>
</tbody>
</table>
Figure D.3 Test Set Results for the Second Trial: A-Correlation Plot, B- Error Histogram

A

B

Error [NTU]
Figure D.4 Comparison of Actual and Predicted Values for the Second Trial: A-Spring Season, B-Summer Season, and C-Fall Season

A

Still having difficulty due to backwash. From the data, turbidity is more closely correlated to a flow with a lag of 20 minutes.

B

This shift is suspicious since predicted values occur in advance of the actual. There appears to be an interference with the treatment train.

C

Difficulty when flow rate is low.
WTP that there were tests being taken in which the plant flow rate was greatly decreased, as well as a period of time where the polymer dosage was not flow paced. Due to the potential lack of flow paced polymer addition during the summer season and the observed difficulty of the network to predict under unsteady conditions, this season has been discarded. Another area of difficulty is the depressions in settled water turbidity due to taking one of the filters off-line. The settled water turbidity depressions correlate best when a flow rate at a lag of 20 minutes is used. Once these changes have been made, the third trial is investigated.

The architectures for the third trial have been investigated and the summary of the minimum overall validation errors are summarized in Table D.6. The test set results are illustrated in Table D.7 and Figure D.5. The correlation plot demonstrates an improvement in network learning due to the higher slope (m=0.8059) and correlation coefficient ($r^2=0.8397$) and lower intercept (b=0.2819). The network, however, still over-estimates the lower settled water turbidities (which are due to one of the filters being taken off line). The mean of the histogram reflects the over-estimation during periods of backwash by its shift to the right. The natural logarithmic transform of the raw water turbidity nodes may also contribute to the error histogram’s skew. These deficiencies are reflected in the time trend plots demonstrated in Figure D.6. The time trend plots demonstrate the impact of the neural network algorithm on the predicted settled water turbidity. The network learns by minimizing the average network error, resulting in over-approximating the effects of backwash in spring and underestimating during the fall. Once the fall and spring seasons
The minimum values follow a difficult pattern. It is likely from the results that the minimum is located at either an architecture of 9 or 10. Clearly, more replicates are needed. One may want to try training without noise to see if this problem still arises. Another approach is to alter the learning parameters, such as lowering the learning rate constant. It is assumed here that the minimum occurs at an architecture of 9.

---

### Table D.6
Minimum Average Validation Cost of Architectures Investigated\(^a\) for the Third Trial

<table>
<thead>
<tr>
<th># hidden</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed #1</td>
<td>0.0247497</td>
<td>0.0266733</td>
<td>0.38116</td>
<td>0.0264287</td>
<td>0.0246439</td>
<td>0.0248872</td>
<td>0.0236172</td>
</tr>
<tr>
<td>Seed #2</td>
<td>0.023865</td>
<td>0.0244124</td>
<td>0.0235552</td>
<td>0.0243679</td>
<td>0.0235369</td>
<td>0.0251402</td>
<td>0.0247085</td>
</tr>
<tr>
<td>Seed #3</td>
<td>0.0252481</td>
<td>0.0252817</td>
<td>0.0234269</td>
<td>0.0265886</td>
<td>0.0239425</td>
<td>0.0244036</td>
<td>0.0252367</td>
</tr>
<tr>
<td>Min.</td>
<td>0.023865</td>
<td>0.0244124</td>
<td>0.0234269</td>
<td>0.0243679</td>
<td>0.0239425</td>
<td>0.0244036</td>
<td>0.0236172</td>
</tr>
</tbody>
</table>

\(^a\) The minimum values follow a difficult pattern. It is likely from the results that the minimum is located at either an architecture of 9 or 10. Clearly, more replicates are needed. One may want to try training without noise to see if this problem still arises. Another approach is to alter the learning parameters, such as lowering the learning rate constant. It is assumed here that the minimum occurs at an architecture of 9.

### Table D.7
Test Results: Summary of Error Statistics for the Third Trial

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0237</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.1624</td>
</tr>
<tr>
<td>MAE</td>
<td>0.1064</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0001</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>0.7195</td>
</tr>
<tr>
<td>(r)</td>
<td>0.9164</td>
</tr>
</tbody>
</table>
Figure D.5 Test Set Results for the Third Trial: A- Correlation Plot, B-Error Histogram

A

Larger variance at lower settled water turbidities (during a time when one filter is taken off-line)

\[ y = 0.8059x + 0.2819 \]

\[ R^2 = 0.8397 \]

B

The error histogram is skewed to the right since the network overestimates the majority of values, such as when one of the filters is taken off-line for backwashing.
Figure D.6 Comparison of Actual and Predicted Values for the Third Trial: Settled Water Turbidity Network at Manheim WTP: A-Spring Season, B-Fall Season

A Under-estimates due to prior raw water run-off period. May want to reduce the number of raw water nodes to obtain a more accurate representation of raw water turbidity.

B Over-estimate settled water turbidity due to method in which net learns, i.e. minimizing the average error over the entire data set.
are modelled by two separate networks, the season nodes are no longer required. Furthermore, the predicted settled water turbidity is more sensitive to the raw water turbidity parameter than the actual settled water turbidity. This can be remedied by eliminating those raw water turbidity lag that have a low overall sensitivity, as calculated by NeuroSolutions®.

Once these changes have been made, the network was trained using a learning rate constant of 0.001 and a momentum constant of 0.7 for a total of 1000 epoch. The results from this trial are illustrated in Appendix E- Final Network Results.
APPENDIX E

FINAL NETWORK RESULTS
Table E.1  Summary of Error Statistics of the Test Set for the Spring Settled Water Turbidity Network for the Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0054</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.0976</td>
</tr>
<tr>
<td>MAE</td>
<td>0.0550</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0003</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>0.4780</td>
</tr>
<tr>
<td>r</td>
<td>0.9501</td>
</tr>
</tbody>
</table>

Figure E.1  Test Set Results for Spring Season of Settled Water Turbidity Network for the Manheim WTP: A- Correlation Plot, B- Error Histogram

A  
\[ y = 0.9004x + 0.1551 \]  
\[ R^2 = 0.9027 \]

B  
Frequency

Error (NTU)
Figure E.2  Actual (Historical) and Predicted Trend of the Complete Spring Season Data Set for the Manheim WTP Settled Water Turbidity Net
Figure E.3 Underlying Relationships of Spring Settled Water Turbidity Net at a Temperature of 9°C, pH of 8.2, Alkalinity of 180, and a Plant Flow of 300 L/s- Manheim WTP: A- Sensitivity of Settled Water Turbidity to Alum and Raw Water Turbidity, B- Sensitivity to Polymer and Raw Water Turbidity

A

- alum = 35 mg/L
- alum = 39 mg/L
- alum = 42 mg/L
- alum = 49 mg/L
- alum = 57 mg/L

B

- polymer = 0.1 mg/L
- polymer = 0.15 mg/L
- polymer = 0.2 mg/L

Extrapolation beyond network experience occurs here
Table E.2  Summary of Error Statistics of the Test Set for the Fall Settled Water Turbidity Network for the Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0248</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.1402</td>
</tr>
<tr>
<td>MAE</td>
<td>0.1144</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0001</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>0.8366</td>
</tr>
<tr>
<td>r</td>
<td>0.9301</td>
</tr>
</tbody>
</table>

Figure E.4  Test Set Results for Fall Season of Settled Water Turbidity Network for the Manheim WTP: A- Correlation Plot, B- Error Histogram

![A](image1.png)  ![B](image2.png)
Figure E.5  Actual (Historical) and Predicted Trend of the Complete Fall Season Data Set for the Manheim WTP Settled Water Turbidity Net
Table E.3  Summary of Error Statistics of the Test Set for the Spring Post-Filter Particle Count Network for the Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>1 μm</th>
<th>2 μm</th>
<th>5 μm</th>
<th>10 μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>4711.0225</td>
<td>87.4935</td>
<td>11.4875</td>
<td>0.8813</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.2117</td>
<td>0.3671</td>
<td>0.3376</td>
<td>0.4582</td>
</tr>
<tr>
<td>MAE</td>
<td>27.4267</td>
<td>2.3780</td>
<td>1.3101</td>
<td>0.3789</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0009</td>
<td>0.0027</td>
<td>0.0006</td>
<td>0.0034</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>889.9346</td>
<td>234.9355</td>
<td>52.0909</td>
<td>14.5057</td>
</tr>
<tr>
<td>r</td>
<td>0.8887</td>
<td>0.7987</td>
<td>0.8168</td>
<td>0.7369</td>
</tr>
</tbody>
</table>

Figure E.6  Test Set Results for the Spring Post-Filter Particle Count Network for the Manheim WTP: A- Correlation Plot\(^a\) of 1 to 2 micron channel, B- Correlation Plot of 2 to 5 micron channel

\(^a\) The y=x line on all correlation plots from this point on represents ideal network predictions
Figure E.6  Test Set Results for the Spring Post-Filter Particle Count Network for the Manheim WTP:  C- Correlation Plot of 5 to 10 micron channel, B- Correlation Plot of 10 to 15 micron channel.
Figure E.7  Actual (Historical) and Predicted Trend of the Complete Spring Season Data Set for the Manheim Particle Count Net: A- 1 to 2 micron channel, B- 2 to 5 micron channel
Figure E.7  Actual (Historical) and Predicted Trend of the Complete Spring season Data Set for the Manheim Particle Count Net: C- 5 to 10 micron channel, D- 10 to 15 micron channel
Figure E.8 Underlying Relationships of Spring Particle Count Net at a Temperature of 10°C, pH of 8.2, and Alkalinity of 180 mg/L at Manheim WTP. Sensitivity of Particle Count to Average Settled Water Turbidity: A- at a run time of 20 minutes ($\Sigma I(t) = 200$ mg, headloss = 2.42 (%), filter flow rate = 45-75 L/s), B- at a run time of 1000 minutes ($\Sigma I(t) = 15000$ mg, headloss = 6 (%), filter flow rate = 45-75 L/s)

A

B
Figure E.8  Underlying Relationships of Spring Particle Count Net at a Temperature of 10°C, pH of 8.2, and Alkalinity of 180 mg/L at Manheim WTP- Sensitivity of Particle Count to Average Settled Water Turbidity: C- at a run time of 2000 minutes (ΣI(t) = 25000 mg, headloss = 10 (%), filter flow rate = 45-75 L/s)
Table E.4  Summary of Error Statistics of the Test Set for the Fall Total Particle Count Network for the Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>10572.5986</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.1216</td>
</tr>
<tr>
<td>MAE</td>
<td>61.8725</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0574</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>675.8777</td>
</tr>
<tr>
<td>r</td>
<td>0.9569</td>
</tr>
</tbody>
</table>

Figure E.9  Test Set Correlation Plot for the Fall Total Particle Count Network for the Manheim WTP
Figure E.10  Actual (Historical) and Predicted Trend of the Complete Fall Season Data Set for the Manheim WTP Fall Total Particle Count Net
Table E.5  Summary of Error Statistics of the Test Set for the Fall Particle Count Network (%) for the Manheim WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>1-2 μm (%)</th>
<th>2-5 μm (%)</th>
<th>5-10 μm (%)</th>
<th>10-15 μm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>3.1060</td>
<td>1.8681</td>
<td>0.2145</td>
<td>0.0093</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.0576</td>
<td>0.0771</td>
<td>0.0421</td>
<td>0.0539</td>
</tr>
<tr>
<td>MAE</td>
<td>1.1188</td>
<td>0.9104</td>
<td>0.2772</td>
<td>0.0542</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>12.3894</td>
<td>8.9685</td>
<td>2.9645</td>
<td>0.7501</td>
</tr>
<tr>
<td>r</td>
<td>0.9712</td>
<td>0.9610</td>
<td>0.9793</td>
<td>0.9727</td>
</tr>
</tbody>
</table>

Figure E.11  Test Set Results for Fall Particle Count Network (%) for the Manheim WTP: A- Correlation Plot of the 1 to 2 micron channel (% of total count), B- Correlation Plot of the 2 to 5 micron channel (%)

![A](image)

![B](image)
Figure E.11  Test Set Results for the Fall Particle Count Network (%) for the Manheim WTP: C- Correlation Plot of 5 to 10 micron channel (%), D- Correlation Plot of 10 to 15 micron channel (%)
Figure E.12  Actual (Historical) and Predicted Trend of the Complete Fall Season Data Set for the Manheim WTP Particle Count Net: A- 1 to 2 micron channel (%), B- 2 to 5 micron channel (%)
Figure E.12  Actual (Historical) and Predicted Trend of the Complete Fall Season Data Set for the Manheim WTP Particle Count (%) Net: C- 5 to 10 micron channel (%), D- 10 to 15 micron channel (%)
Figure E.13  Underlying Relationship of the Fall 1 to 2 micron Channel Particle Counts When Using Both the Fall Particle Count Net (%) and Total Particle Count Net at a pH of 7.2, Alkalinity of 170 mg/L, Temperature of 16.5°C for Manheim WTP: A- at a run time of 10 minutes (headloss = 6 %), B- at a run time of 20 minutes (headloss = 6 %)
Figure E.13  Underlying Relationship of the Fall 1 to 2 micron Channel Particle Counts When Using Both the Fall Particle Count Net (%) and Total Particle Count Net at a pH of 7.2, Alkalinity of 170 mg/L, Temperature of 16.5°C for Manheim WTP: C- at a run time of 600 minutes (headloss = 6 %), B- at a run time of 1100 minutes (headloss = 20 %)

C

D
Figure E.14  Underlying Relationship of the Fall 5 to 10 micron Channel Particle Counts When Using Both the Fall Particle Count Net (%) and Total Particle Count Net at a pH of 7.2, Alkalinity of 170 mg/L, Temperature of 16.5°C for Manheim WTP: A- at a run time of 10 minutes (headloss = 6 %), B- at a run time of 20 minutes (headloss = 6 %)
Figure E.14 Underlying Relationship of the Fall 5 to 10 micron Channel Particle Counts When Using Both the Fall Particle Count Net (%) and Total Particle Count Net at a pH of 7.2, Alkalinity of 170 mg/L, Temperature of 16.5°C for Manheim WTP: C- at a run time of 600 minutes (headloss = 6 %), D- at a run time of 1100 minutes.
Table E.6 Summary of Error Statistics of the Test Set for the Fall Particle Count Network (%) for the Britannia WTP

<table>
<thead>
<tr>
<th>Performance</th>
<th>2-3um</th>
<th>3-5um</th>
<th>5-10um</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>81508.6172</td>
<td>12144.2451</td>
<td>600.1399</td>
</tr>
<tr>
<td>NMSE</td>
<td>0.2030</td>
<td>0.2017</td>
<td>0.2529</td>
</tr>
<tr>
<td>MAE</td>
<td>141.4586</td>
<td>54.2397</td>
<td>11.7039</td>
</tr>
<tr>
<td>Min Abs Error</td>
<td>0.4924</td>
<td>0.0180</td>
<td>0.0189</td>
</tr>
<tr>
<td>Max Abs Error</td>
<td>1409.5914</td>
<td>571.9019</td>
<td>142.2351</td>
</tr>
<tr>
<td>r</td>
<td>0.8972</td>
<td>0.8987</td>
<td>0.8675</td>
</tr>
</tbody>
</table>

Figure E.15 Test Set Results for the Initial Filter Ripening Period Network for the Britannia WTP: A- 2 to 3 micron range, B- 3 to 5 micron range, C- 5 to 10 micron range
Figure E.15  Test Set Results for the Initial Filter Ripening Period Network for the Britannia WTP: A- 2 to 3 micron range, B- 3 to 5 micron range, C- 5 to 10 micron range
Figure E.16  Actual (Historical) and Predicted Trend of the Initial Ripening for the Data Set provided by the Britannia