University of Alberta

FULL-SCALE ARTIFICIAL NEURAL NETWORK MODELLING OF ENHANCED COAGULATION

By

Christopher Wayne Baxter 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**

in

Environmental Science

Department of Civil and Environmental Engineering

Edmonton, Alberta

Fall 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-34335-9

Canada

ABSTRACT

As water treatment regulations for the removal of disinfection-by-products (DBPs) become more stringent, water utilities must actively search out new technologies that improve the removal of DBP precursors, namely natural organic matter (NOM). Enhanced coagulation has been identified as the best practical technology for the removal NOM. Few attempts have been made to develop a full-scale model of the enhanced coagulation process. Models derived from bench-scale and pilot scale experiments often fail when applied to full-scale systems. This thesis describes the development of full-scale artificial neural network (ANN) models for the removal of NOM by enhanced coagulation at the Rossdale and E.L. Smith Water Treatment Plants (WTPs) in Edmonton, Alberta. The models derived from this research can be used by the WTP operators at these facilities to optimize the enhanced coagulation process. Alternatively, the models can be used as virtual full-scale laboratories to provide insight into the enhanced coagulation process.

This work is dedicated to my family, without whose love and encouragement I would never have reached this level of education, and to Sheri Patterson for her continued love, friendship, and support.

ACKNOWLEDGEMENTS

This work would not have been possible without the financial and organizational support of AQUALTA and the American Water Works Association Research Foundation. In addition, several individuals deserve special recognition for their contributions:

Stephen J. Stanley, my supervisor, for providing valued assistance in developing and implementing the research, Riyaz Shariff and Simon Thomas of AQUALTA for digging up obscure data and for answering countless questions, and Qing Zhang for his assistance in developing and evaluating the models.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

°C	degrees Celsius
%	percent
AI	artificial intelligence
ANN	artificial neural network
A-units	association cells
C1	clarifier 1
C2	clarifier 2
C3	clarifier 3
$CaCO_3$	calcium carbonate
d	day
D/DBP	disinfectant / disinfection by-product ratio
DBP	disinfection by-product
DLL	dynamic link library
DOC	dissolved organic carbon
FA	fulvic acid
G	mean velocity gradient
g/mol	grams per mole
GAC	granular activated carbon
GDR	generalized delta rule
GRNN	general regression neural network
HA	humic acid
HAA	haloacetic acid

IMC	internal model control
km	kilometer
L	liter
m	meter
m/d	meters per day
m^2	square meters
m^3	cubic meters
m^3/d	cubic meters per day
Max.	maximum
MCL	maximum contaminant level
mg/L	milligrams per liter
Min.	minimum
ML	megaliter
ML/d	megaliters per day
MSE	mean squared error
MWC	molecular weight cut-off
MWD	molecular weight distribution
NaOH	sodium hydroxide
NF	nanofiltration
nm	nanometer
NOM	natural organic matter
NTU	nephelometric turbidity unit
PAC	powdered activated carbon
PACL	polyaluminum chloride
PNN	probabilistic neural network

r^2	coefficient of regression
R^2	coefficient of multiple regression
RO	reverse osmosis
R-units	responses
s	seconds
SCADA	supervisory control and data acquisition
SCD	streaming current detector
SOM	self-organizing map
S-points	sensory units
Std. Dev.	standard deviation
SUVA	specific ultraviolet absorbance
T	flocculation time
t	time
T : T : P	training: testing: production ratio
TCU	true color unit
TDS	total dissolved solids
THM	trihalomethane
THMFP	trihalomethane formation potential
TOC	total organic carbon
TTHMR	total trihalomethane regulation
UF	ultrafiltration
um	micrometer
us/cm	microSeimens per centimeter
USEPA	United States Environmental Protection Agency
UVA	ultraviolet absorption
VOC	volatile organic compound

WTP water treatment plant

1.0 INTRODUCTION

1.1 Purpose of the Study

The purpose of the current study is to develop full-scale artificial neural network (ANN) models for the enhanced coagulation process at the E.L. Smith and Rossdale Water Treatment Plants in Edmonton, Alberta. The completed models can be used to optimize the process by providing plant operators with a robust tool to determine the effects of their proposed control actions on the removal of disinfection by-product precursors. Alternatively, the models can be used as virtual full-scale laboratories to provide insight into the enhanced coagulation process. The effects of simultaneously varying the values of multiple input parameters on the process output can be assessed without the prohibitive costs and scale-up concerns of bench-scale and pilot-scale experimentation.

1.2 General Problem Description

1.2.1 Natural Organic Matter and Enhanced Coagulation

In conventional water treatment where free chlorine is used as the disinfectant, disinfection by-products, such as trihalomethanes (THM) and haloacetic acids (HAAs) can form by the reaction of residual chlorine with some forms of natural organic matter (NOM) in the treated water. In addition, in the presence of bromide, the residual free chlorine oxidizes bromide to hypobromous acid which, in turn, reacts with NOM to produce brominated-DBPs (Singer 1994). As many DBPs are suspected to be

carcinogenic, it is generally desirable to remove them from the drinking water stream. Strategies for controlling DBPs in treated water include precursor removal, the use of alternative disinfectants, and removal of DBPs following formation (Singer 1994). The precursor removal strategy involves the removal of NOM in the water treatment process by enhanced coagulation or softening, granular activated carbon (GAC) adsorption, or membrane filtration. Enhanced coagulation, which involves the use of coagulant doses that exceed those required for turbidity removal, is considered to be the best available technology for the reduction of DBPs in treated water (Krasner and Amy 1995). Membrane filtration and GAC adsorption can be effective at removing DBP precursors, although both of these technologies are prohibitively expensive (Singer 1994). With respect to the use of alternative disinfectants, common substitutes for chlorine in disinfection include chloramines, chlorine dioxide, and ozone. Both chloramines and chlorine dioxide are able to form DBPs, albeit at lower concentrations than free chlorine, and ozone can result in the production of brominated DBPs when bromide-containing waters are ozonated (Crozes, White and Marshall 1995). In addition, the use of substitute disinfectants may not provide adequate disinfection; the balance between adequate disinfection and the formation of DBPs must be assessed on a case-by-case basis. The final DBP control strategy involves the use of air stripping apparatus to remove DBPs from the treated water. The main disadvantages of this strategy include the inability of the air stripping apparatus to remove non-volatile DBPs and the enormous costs associated with treating the entire product stream in large WTPs.

1.2.2 Current and Proposed Disinfection By-Product Regulations

Over the past quarter-century, the United States Environmental Protection Agency (USEPA) has attempted to find a balance between promoting microbial reduction in drinking water by disinfection and discouraging the formation of disinfection by-products. This balance can be quantified through the use of the disinfectant / disinfection by-product (D/DBP) ratio. The first attempt to control D/DBP was made by the USEPA in 1974, when THM formation in drinking water was identified as a by-product of chlorination (Clark, Adams and Lykins 1994). The Total Trihalomethane Regulation (TTHMR) set initial limits of 0.10 mg/L in treated water for the compounds in November, 1979. More recently, the USEPA proposed the two-stage D/DBP rule which will reduce the maximum contaminant level (MCL) of TTHM to 0.080 mg/L and HAA_5 to 0.060 mg/L in stage 1 (Crozes, White and Marshall 1995). The TTHM group includes chloroform, bromodichloromethane, dibromochloromethane, and bromoform while the HAA_5 group includes monochloroacetic acid, dichloroacteic acid, trichloroacetic acid, monobromoacetic acid, and dibromoacetic acid. In addition to these two groups of DBPs, both bromate and chlorite will be regulated in stage 1 with MCLs of 0.010 and 1.00 mg/L respectively. Long-term standards for THMs and HAAs will be reduced to 0.040 mg/L and 0.030 mg/L respectively as part of stage 2 of the D/DBP rule. The D/DBP has met much opposition from water treatment utilities since the treatment costs associated with meeting the proposed new legislation are significant. In response to pressures from public utilities, the U.S. Congress has postponed the promulgation of stage 1 of the D/DBP rule until November 1998, with stage 2 coming into effect in November 2002. These dates

have been established for facilities which provide water for populations exceeding 10 000; smaller facilities will have a more relaxed schedule of compliance.

With respect to Canadian regulations, under the Guidelines for Canadian Drinking Water Quality, the current interim maximum acceptable concentration of THMs in drinking water is 0.1 mg/L (Federal-Provincial Subcommittee on Drinking Water 1995). The guideline is designated as interim until such time as the risks from DBPs other than chloroform are ascertained. In Alberta, the Guidelines for Canadian Drinking Water Quality are enacted into law under the Environmental Protection and Enhancement Act. Water utilities are required demonstrate compliance with the 0.1 mg/L guideline through periodic sampling for TTHMs at the extreme end of their distribution (Standards and Guidelines Branch 1997). The sampling frequency is dependent on the population served by the utility and varies from every three months to once per year.

1.2.3 Raw Water Quality at the Rossdale and E.L. Smith Water Treatment Plants

Both the Rossdale and E.L. Smith Water Treatment Plants obtain their source water from the North Saskatchewan River. The river has its headwaters in the Canadian Rocky Mountains and flows in an easterly direction for approximately 500 km before reaching the city of Edmonton. Due to substantial contributions from spring-melt, the river water quality varies considerably on a seasonal basis. Turbidities range from less than 2 NTU in winter to over 2000 NTU during spring thaw. Similarly, color ranges from 5 TCU to 80 TCU and TOC ranges from 1.5 mg/L to 8 mg/L. The water is moderately hard (100-135

mg/L as $CaCO_3$) and slightly alkaline (pH 7.8-8.8). The large variations in water quality will allow for an assessment of the robustness of the model under varying water quality conditions.

1.2.4 Artificial Intelligence Modelling Techniques

As an alternative to conventional modelling approaches, several artificial intelligence (AI) modelling techniques have been developed over the past 50 years. These techniques can be classified into one of three categories: knowledge-based expert systems, fuzzy logic-systems, and artificial neural networks. The use of AI techniques is currently on the rise due to the advancement in computer technology in recent years.

Knowledge-based expert systems combine human expertise with the computational power of specialized systems to solve problems in a specific area (Hushon 1990). The systems require a database of knowledge about a particular area which, when combined with a skill pool, allows for a greater understanding of the problem at hand. The combination of the knowledge base and skill pool also allows for the generation of rules or heuristics that can be used in problem solving and modelling.

Fuzzy logic systems classify input data according to its degree of membership in the fuzzy set. The fuzzy set is a loosely defined category to which an input may belong to a certain degree of probability. The output is reported as a value between 0 and 1, with a value of 1 indicating 100 % membership in the fuzzy set. Fuzzy logic systems, when

coupled with statistical tools, are able to formulate rules from historical data without the use of rules generated by experts.

Artificial neural networks mimic the human brain's problem solving capabilities. ANN systems are comprised of interconnected arithmetic computing units that are analogous to the brain's network of neurons. Artificial neural networks are capable of self-organization and learning; concepts and patterns can be extracted directly from historical data without human subjective inference. ANNs are data intensive; problems for which large set of data are available are more likely to be solved by this technique. The technique can also be successfully applied when the problem can not be described by well-defined algorithms, heuristics, or rules.

The ANN modelling technique has several advantages over conventional modelling approaches that make it especially applicable to the current study. The ANN approach does not require mathematical algorithms, only a knowledge of the important factors governing the process. In the water treatment industry, many uncertainties exist because of the micro-scale physical and chemical reactions involved. Conventional modelling techniques require mathematical algorithms to describe these uncertainties, where a neural network simply learns the process based on historical data (Stanley, Zhang and Baxter 1998). In addition, the ANN technique is fast and flexible. Once the historical data has been compiled, ANN models can be generated rather quickly. Changes in the underlying process, which would render a conventional model invalid, can be incorporated into ANN models following a brief period of retraining. The ANN

technique can also handle the non-linearity of changes in the input parameters. In water treatment, many of the raw water quality parameters vary on a daily and seasonal basis; models that assume a linear structure are not able to cope with such changes. Finally, ANNs tend to be fault-tolerant as their data structure is loosely organized and there is no boundary limit on the input parameters (Stanley, Zhang and Baxter 1998). This feature allows the ANN models to be incorporated into real-time process control.

1.3 Thesis and Approach

1.3.1 Summary of Expected Outcomes

Due to the success of previous applications of the ANN technique to the modelling of water quality parameters (Zhang and Stanley 1997) and water treatment processes (Stanley and Zhang 1997), it is expected that the current study will yield favorable results. Each of the models is expected to capture the interactions between the input parameters and the output parameter in historical data in order to allow for accurate output parameter predictions using new input data. When tested online for use in process control, the technique will also prove to be less time consuming and more robust than current methods of enhanced coagulation process control. Finally, it is expected that the models, when used as a virtual full-scale laboratory, will provide valuable insights into the enhanced coagulation process, allowing plant operators to make more effective and economical decisions with respect to plant operations.

1.3.2 Artificial Neural Network Model Development

The artificial neural network modelling process involves three distinct stages: source data analysis, ANN model development, and model evaluation using simulated real-time data. The primary objectives of the source data analysis are to gain a familiarity with the study domain and to examine the applicability of available data for model development. In the model development stage, the goal is to design and evaluate a series of network architectures that can be used as an effective process model. In the final stage, candidate models, developed using historical data, are evaluated retroactively on data selected to simulate real-time operations in order to assess their potential for use in process control.

1.4 Criteria for Success

The success of the study is dependent on three key measures: the performance of the models with respect to the historical data, the ability of the models to cope with continuous real-time data, and the ability of the models to generate meaningful results from virtual lab experiments. With respect to the first measure, it is desirable to obtain a high degree of correlation between actual outputs and model predicted outputs on the training, testing, and production data sets. With respect to the second measure, the models will only be useful for process control if they can be applied to continuous real-time data. The models will be developed using historical data averaged on a daily basis. In order to ensure that the models are actually capturing the unique features, relationships, and interactions of the enhanced coagulation system, they will be subjected

to a period of testing in simulated real-time. Based on conversations with plant operators, a tolerance of 0.5 TCU in the 2.0 - 3.0 TCU range during the real-time simulations is desirable if the models are to be used in process control applications (Thomas and Shariff 1998). Finally, once the performance of the models on historical and online data has been proven, the models can be used as virtual full-scale laboratories. Successful models will be able to reproduce results obtained from bench-scale and pilot-scale experimentation, and will provide insight that does not conflict with operator knowledge or literature.

1.5 Outline of the Document

The remainder of this document is organized into five sections: background information, methodology, results and discussion, applications, and conclusions and recommendations. The background information section consists of an overview of the treatment scheme at each of the water treatment plants, as well as more detailed discussions of NOM and enhanced coagulation, coagulation modelling and control, and artificial neural networks. In the methodology section, much of the discussion focuses on the protocol used to develop the ANN models. Both the software and the source of the data used in modelling are also discussed. In the results and discussion section, results from each of the modelling stages are presented, and the results obtained from each of the models are discussed. The applications section discusses potential applications of the model in process control. Finally, a summary of the current study, as well as recommendations for future study are presented in the conclusions and recommendations section of this document.

2.0 BACKGROUND INFORMATION

2.1 Overview of the Rossdale and E.L. Smith Water Treatment Plants

2.1.1 Rossdale WTP

The Rossdale Water Treatment Plant (WTP), owned and operated by AQUALTA, is located on the North Saskatchewan River, a major tributary in the Saskatchewan-Nelson river system, within the boundaries of the City of Edmonton. The river has its headwaters in the Canadian Rocky Mountains and flows in an easterly direction for approximately 500 km before reaching the city. Much of the upstream watershed is uninhabited forest with little industrial or residential development, although there is a significant amount of agricultural land-use closer to the city. There has been a water treatment facility at the site since 1903, although the current facility was constructed in 1947 and was expanded in 1955 to meet the needs of the expanding Edmonton population.

The Rossdale WTP is composed of two independent treatment trains, identified as Plant # 1 and Plant # 2, with a combined operating capacity of 260 ML/day. With respect to the clarification process equipment, each plant has one square cross-flow clarifier that is comprised of a rapid-mix chamber, three stages of tapered flocculators, and one sedimentation basin. The rapid mix uses inline mechanical mixers with a 2.4 m^3 mixing volume and a design Gt of 2000. Each of the three stages of flocculators consists of 5 cells with total volumes of 522, 693, and 690 m^3 for Plant # 1 and 571, 701, 652 m^3 for Plant # 2. The design detention times though the clarifiers is 20 minutes during normal

operating conditions, and 30 minutes during cold weather operations. The sedimentation basin for Plant # 1 measures 35.0 m x 35.2 m x 4 m, while that for Plant # 2 measures 46.4 m x 49.8 m x 4 m. Sedimentation is assisted by banks of up-flow tube settlers that cover approximately 30 % of the sedimentation basin area, and are mounted at the effluent end of the basin. With respect to the clarification process, the Rossdale facility practices enhanced alum coagulation with an anionic polymer coagulant aid. Powdered activated carbon (PAC) can also be added on demand in order to control severe taste and odour problems, which are especially prevalent during spring runoff.

Following clarification, the effluent is softened using lime and is recarbonated in order to adjust the pH. Disinfection occurs through the use chlorinated disinfectants. Ammonia is added within minutes of free-chlorine addition resulting in the formation of chloramines. The water is fluoridated and passes through a stilling basin for increased disinfection contact time. The effluent is then filtered via mono-media (crushed-quartz) rapid sand filtration before being pumped into 100 ML on-site reservoirs. A schematic diagram of the treatment process at Rossdale WTP is presented in Figure 2.1.

2.1.2 E.L. Smith WTP

The E.L. Smith WTP, owned and operated by AQUALTA, is also located on the North Saskatchewan River, approximately 15 km upstream of Rossdale WTP on the western fringes of the City of Edmonton. The facility was originally built in 1976 and was expanded to its current configuration in 1984.

The E.L. Smith facility is currently a single-train facility with a design capacity of 190 ML/day. The coagulation equipment consists of three identical upflow solids-contacting clarifiers. Under normal operating conditions two of the clarifiers are used for alum coagulation, and one is used for lime softening. Each clarifier is a 42.7 m x 42.7 m x 7.7 m deep concrete basin, with a draught tube and impeller mixer in the center and high rate shallow depth plastic tube settlers installed approximately 1 m below the launder system along the outer edges of the clarifier. A cross-section of half of one of the clarifiers is presented in Figure 2.2. With respect to the clarification process, the E.L. Smith facility also practices enhanced alum coagulation with an anionic polymer coagulant aid. Again, PAC can be added on demand in order to control taste and odour problems. The chemicals are introduced and rapid mixing occurs by means of an in-line mixer in the raw water supply line at a point just upstream of the clarifiers. The influent raw water is introduced directly into the circular draught tube via a tangential entry. The raw water influent is mixed with recirculated sludge and forced upward and out of the draught tube due to the presence of an impeller-generated recirculated flow. After exiting the draught tube, the flow may either be recycled back into the draught tube through the bottom or may exit the clarifier through the tube settlers and launders. There are two primary purposes for the draught tube and impellers: 1) to provide mixing energy for the flocculation of particles; and 2) to recycle sludge to increase particle concentration and improve the efficiency of flocculation. The existing impellers and clarifiers are a proprietary design of Ecodyne Ltd. They are typically operated at impeller speeds of 4 to

10 RPM depending upon the conditions necessary to produce an acceptable degree of flocculation.

Following clarification, the effluent is softened using lime and is recarbonated in order to adjust the pH. Disinfection occurs through the use of free-chlorine, followed by the addition of ammonia in order to ensure a chloramine residual in the distribution system. The water is also fluoridated. The effluent is then filtered via dual-media (anthracite and sand) rapid sand filtration before being pumped into 125 ML on-site reservoirs. A schematic diagram of the treatment process at E.L. Smith WTP is presented in Figure 2.3.

2.2 NOM and Enhanced Coagulation

2.2.1 NOM and DBPs

Natural organic matter consists of humic substances, amino acids, sugars, aliphatic acids, and a large number of organic molecules (Malcolm Pirnie Inc. 1993). Organic matter in natural waters is mostly humic in nature, imparting a yellow or brown colour to the water (Edwards and Amirtharajah 1985). Humic substances are formed by the biodegradation of plant and animal matter in both the aquatic and terrestrial environments. As such, the humic material reflects the aquatic, soil, and vegetative conditions of their originating drainage basin (Joyce, DiGiano and Unden 1984). Humic substances are amorphous, acidic, predominantly aromatic, hydrophilic, chemically complex polyelectrolytes (Edwards and Amirtharajah 1985). In natural water pH conditions, humic substances exist as negatively charged macromolecules with the negative charge resulting from the

presence of functional groups (Edwards and Amirtharajah 1985). The major functional groups are carboxyl, phenolic hydroxyl, and alcoholic hydroxyl; depending on which functional groups are present, the chemical composition of humic matter varies between approximately 40-60 % carbon, 30-50% oxygen, 3-6% hydrogen, and 1-4 % nitrogen (by weight) (Malcolm Pirnie Inc. 1993).

Humic matter is further sub-divided into a number of classes depending on its solubility in acid, alcohol, and base (Table 2.1) (AWWA Committee 1979).

Table 2.1 Classification of humic substances

Group Name	Definition
Humus Coal	Insoluble in water and NaOH
Fulvic Acid	Soluble in mineral acid and NaOH
Hymatomelanic Acid	Soluble in NaOH and alcohol, insoluble in mineral acid
Humic Acid	Soluble in NaOH; insoluble in alcohol and mineral acid

In natural waters, the distribution of the humic substances is > 80 % fulvic acid, ~10 % hymatomelanic acid, and ~ 2 % humic acid (AWWA Committee 1979). Humus coal, being insoluble in water, is not found in natural waters in any appreciable quantities. Many researchers do not distinguish between the hymatomelanic acid and the humic acid fractions. As such, much of the literature focuses on differentiating the humic acid and fulvic acid fractions based on their physical and chemical characteristics. Fulvic acids have molecular weights in the range of 200 - 1000 g/mol and therefore exhibit a much higher charge density than their humic counterparts, which have molecular weights in the 200000 g/mol range (Kavanaugh 1978). Fulvic acids also contain a higher concentration of oxygen-containing functional groups and have fewer carbon and nitrogen atoms per unit weight than the humic acid fractions (AWWA Committee 1979). These differences

result in the preferential coagulation of certain NOM fractions over others, as will be discussed further on.

The reasons for removing organic matter from the drinking water stream are many (AWWA Committee 1979). Natural organic matter imparts colour, taste, and odour to water, all of which can be aesthetically unpleasing. Organic matter can also interfere with the oxidation and removal of heavy metals such as iron and manganese and can foul anion exchange resins, thereby inhibiting demineralization. Natural organic matter is also a precursor to disinfection by-products and some organics can be toxic and/or carcinogenic alone or in association with heavy metals and pesticides. Finally, in distribution systems, the presence of organics may contribute to increased biological growth and thereby increases the rate of corrosion.

Currently, the most cited reason for removing NOM from the drinking water stream is to impede the formation of disinfection by-products (DBPs). In conventional water treatment where chlorinated disinfectants are used, disinfection by-products such as trihalomethanes (THMs) and haloacetic acids (HAAs) can form by the reaction of residual chlorine with natural organic matter (NOM) in the treated water (Table 2.2). As many DBPs, such as chloroform, are suspected to be carcinogenic, it is generally desirable to remove them from the drinking water stream.

2.2.2 Measurement of NOM

As there is no bulk parameter that measures the concentration of disinfection by-product precursors in water, surrogate parameters are used. The ideal surrogate is one that can be

rapidly measured, does not require special equipment or training, and permits the accurate estimation of the target parameter (Malcolm Pirnie Inc. 1993). A list of the most common surrogates used for the determination of NOM is presented in Table 2.3.

Table 2.2 Chlorinated disinfection by-products (after Clark, Adams and Lykins 1994)

Class	Representative DBPs
Trihalomethanes	Chloroform
	Bromodichloromethane
	Chlorodibromomethane
Haloacetic Acids	Dichloroacetic acid
	Trichloroacetic acid
Haloacetonitriles	Bromochloroacetonitrile
	Dichloroacetonitrile
	Trichloroacetonitrile
Haloaldehydes	Dichloroacetaldehyde
	Trichloroacetaldehyde
Haloketones	1,1-dichloropropanone
	1,1,1-trichloropropanone
	1,1-dichloro-2-butanone
	3,3-dichloro-2-butanone
	1,1,1-trichloro-2-butanone
Chlorophenols	2-chlorophenol
	2,4-dichlorophenol
	2,4,6-trichlorophenol
	Chloropicrin
	Cyanogen chloride

Table 2.3 Methods for the determination of NOM in water

Method	Abbreviation	Reference
Total Organic Carbon	TOC	5310*
Dissolved Organic Carbon	DOC	5310*
Ultra-Violet Absorption (253.7 nm)	UVA	5910*
Specific UV Absorbance	SUVA	(Edzwald 1993)
Trihalomethane Formation Potential	THMFP	5710 B and C*
Colour		2120*

*in *Standard Methods*, 19th Ed. (Eaton, Clesceri and Greenberg 1995)

2.2.2.1 TOC and DOC

Total organic carbon (TOC) measures the amount of organically bound carbon in water samples. Dissolved organic carbon (DOC) is the portion of TOC that remains after the water is passed through a 0.45 μm filter (Malcolm Pirnie Inc. 1993). Both are independent of the oxidation state of the organic matter and do not measure other organically bound elements, such as nitrogen and hydrogen (Eaton, Clesceri and Greenberg 1995). TOC and DOC may not consistently provide an accurate measure of DBP precursors as they do not indicate aromaticity, aliphatic nature, functional group chemistry, or chemical bonding associated with natural organic molecules (Malcolm Pirnie Inc. 1993). The removal of TOC is a conservative indicator of the removal of the precursors of trihalomethanes and haloacetic acids, the most common DBPs, and of UV-absorbing compounds (Miltner, Nolan and Summers 1994). As previously mentioned, TOC will be used by the USEPA for the D/DBP rule.

2.2.2.2 UVA and SUVA

Ultra-violet absorption at a wavelength of 253.7 nm is used to provide an indication of the aggregate concentration of UV-absorbing organic constituents, such as humic substances and various aromatic compounds (Eaton, Clesceri and Greenberg 1995). As such, it is often considered to be a surrogate measure of TOC (Edzwald, Decker and Wattier 1985). Measuring the organic content of water by UV is a fast, cheap, automatable process (Reid, Cresser and Macleod 1980). As with any spectrophotometric

technique, interferences are common. Certain oxidants and reducing agents such as ozone, chlorate, chlorite, chloramines, and thiosulphate also absorb light at 253.7 nm (Eaton 1995). UV analyses are also affected by the pH and turbidity of the solution being tested (Malcolm Pirnie Inc. 1993). The specific ultra-violet absorbance (SUVA) is operationally defined as the UVA / DOC of the sample (Edzwald 1993). SUVA is an indicator of the humic content of the organic matter and should be used in addition to TOC when determining the amenability of a water to enhanced coagulation (Krasner et al. 1997). Waters with low SUVA values contain primarily non-humic matter that is not amenable to enhanced coagulation.

2.2.2.3 THMFP

The trihalomethane formation potential (THMFP) of a sample is defined as the concentration of total trihalomethanes (TTHM) that are formed when the sample is incubated in the presence of excess free chlorine (Eaton, Clesceri and Greenberg 1995). THMFP determinations provide a worst-case scenario of the concentration of THMs that may be formed and do not necessarily provide an accurate description of what will happen in a full-scale system (Malcolm Pirnie Inc. 1993). In addition, THMFP determinations are extremely sensitive to solution pH, chlorine dosages, and contact times.

2.2.2.4 Colour

Colour determinations are reported as "true colour", that is, the colour of the water from which turbidity has been removed. Colour in water may result from the presence of

natural metallic ions, humus and peat materials, plankton, weeds, and industrial wastes (Eaton, Clesceri and Greenberg 1995). In water treatment applications, colour is a surrogate measure for the humic content of a sample (Edzwald, Decker and Wattier 1985). As the humic fraction of natural organic mater is largely responsible for the formation of DBPs, the removal of colour can be used as surrogate for the removal of DBP precursors, provided that site-specific correlations are made. Reductions in colour and THM precursors were found to occur almost concurrently (Amirtharajah, Dennett and Studstill 1993).

2.2.2.5 Correlations between the NOM Surrogates

As each of the surrogates measures different fractions of the NOM present in a sample, certain surrogates are preferentially removed over others. In general, the order of removability is UVA > THMFP > TOC (Reckhow and Singer 1990 ; Cheng et al. 1995). In spite of being preferentially removed, UVA was found to be linearly correlated with both TOC and THMFP, although such correlations are case-specific (Edzwald, Decker and Wattier 1985). Correlations between UVA and TOC are affected by the presence of chloramines and turbidity; efforts should be made to remove these interferences before correlations are made (McCarty and Aieta 1984 ; Dobbs, Wise and Dean 1972). The UVA/TOC correlation may also need to be sub-divided on a seasonal basis as the composition of the organic matter is likely to vary from season to season (Reid, Cresser and Macleod 1980). With respect to the correlation between colour and TOC, TOC

removal curves at varying alum doses parallel colour removal curves, although colour is preferentially removed (Babcock and Singer 1979).

2.2.3 DBP Removal Techniques

Disinfection by-product control strategies commonly fall into one of three categories: precursor removal, disinfectant modification, and removal after formation. An assessment of the relative cost of the various control strategies is presented by Clark, Adams and Lykins (1994).

2.2.3.1 DBP Precursor Removal

The most common and most widely supported means of DBP control is through the removal of DBP precursors, namely dissolved and suspended natural organic matter. Precursor removal can be accomplished through precipitative processes, adsorptive processes, and membrane processes (Clark, Adams and Lykins 1994).

With respect to the precipitative processes such as enhanced coagulation and enhanced softening, the physical and chemical properties of the organic matter are altered such that agglomeration is enhanced and the precipitate is removed by settling out or by filtration (Malcolm Pirnie Inc. 1993). The enhanced coagulation process involves the use of additional coagulant in the clarification stage of water treatment in order to improve the removal of disinfection by-product (DBP) precursors, namely natural organic matter

(Crozes, White and Marshall 1995). NOM removal by enhanced coagulation is accomplished by chemical precipitation, formation of insoluble complexes through charge-neutralization, or adsorption onto floc (Bell et al. 1997). The mechanism of removal is highly dependent on the pH of coagulation; at low pH (4.0 - 4.5) the charge-neutralization mechanism predominates while the adsorption mechanism predominates at higher pH levels (6.0-8.0) (Edwards and Amirtharajah 1985).

Adsorption techniques are often employed to remove low molecular weight organics following or during coagulation. The two main techniques involve using activated carbon in either the granulated (GAC) or powdered (PAC) form. GAC is either placed in a gravity bed or in pressure contactors following coagulation, in order to maximize contact with the pre-treated water. If GAC adsorption is not preceded by coagulation, high molecular weight organics, which adsorb to a lesser extent on GAC, will block the pores on the GAC. Adsorption increases with decreasing pH. The overall optimum pH for coagulation / GAC adsorption can be expected to be somewhat acidic and highly dependent on the alkalinity of the system (Randtke 1988). Historically, PAC has been used primarily for the control of taste and odour problems on a seasonal basis. When high pore-volume PAC is used in elevated doses (>50 mg/L) however, it can assist in the removal of DBP precursors. PAC is often more economical than GAC since the need for organics removal often varies considerably throughout the year. PAC is easily turned on and off as needed, creates no headloss, and involves only a very minor capital investment (Randtke 1988).

Membrane processes are generally used in relatively low-turbidity waters in order to remove total dissolved solids (TDS). The membranes employed in these processes are defined by their effective pore size which is often expressed as the molecular weight cut-off (MWC), the molecular weight above which most species are retained by the membrane. Ultrafiltration (UF), nanofiltration (NF), and reverse osmosis (RO) membranes, with MWCs of >500, 200-500, and <200 respectively, are commonly employed in the water treatment industry (Malcolm Pirnie Inc. 1993). Membranes with lower effective pore sizes and MWCs have higher operating pressure requirements and are more susceptible to fouling than those with larger effective pore sizes and MWCs.

2.2.3.2 Disinfectant Modification

In removing DBPs through disinfectant modification, two methods are commonly employed. The most cost-effective method involves replacing free chlorine with an alternate disinfectant such as ozone, chlorine dioxide, and chloramine, although this method has several drawbacks. Both chloramine and chlorine dioxide are able to form DBPs, albeit in lower concentrations, and ozonation of bromine-containing waters can result in the formation of brominated DBPs (Crozes, White and Marshall 1995). In addition, because ozone decomposes rapidly in water, it cannot be used to maintain a residual in distribution systems (Tate and Arnold 1990). As such, a chlorinated disinfectant must be used as the residual disinfectant, potentially forming DBPs in the distribution system if precursors are present. The second method involves moving the point of disinfection further downstream. Since DBP formation is time dependent,

moving the disinfection point further downstream will help to reduce DBP concentrations in the finished water (Clark, Adams and Lykins 1994). In addition, if disinfection is delayed until after filtration, fewer precursors remain to react with the disinfectant. Unfortunately however, moving the point of disinfection further downstream reduces the disinfection contact time thereby reducing the effectiveness of disinfection.

2.2.3.3 Removal of DBPs

The physical removal of DBPs after formation is a costly alternative to precursor removal and can generally be accomplished by either GAC adsorption or by air stripping (Clark, Adams and Lykins 1994). GAC adsorption is commonly used in the manner described for precursor removal above. Since DBPs are, for the most part, volatile organic compounds (VOCs), they can also be removed by air stripping. The process involves bringing air in contact with the water in order to transfer the contaminants from the aqueous to the gaseous phase (Cornwell 1990). Air-stripping for the removal of THMs commonly involves the use of a packed tower with a counter-current flow pattern with water falling down through the tower packing, and air passing upward. This arrangement maximizes the turbulence required for high transfer efficiencies. One of the major disadvantages of the technique is the production of off-gas that is either dispersed to the air or treated by adsorption onto GAC, a costly alternative. In addition, the formation of DBPs occurs in two stages, a rapid stage and a long-term (7 day) stage (Knocke, West and Hoehn 1986). If the DBP precursors are not removed by one of the methods highlighted above, the

potential for DBP formation in the distribution system will exist regardless of whether or not air stripping is employed.

2.2.4 Enhanced Coagulation

Enhanced coagulation is considered to be the best available technology for the reduction of DBPs in treated water (Krasner and Amy 1995). The process involves the use of additional coagulant in order to improve the removal of disinfection by-product (DBP) precursors, namely natural organic matter (Crozes, White and Marshall 1995). The enhanced coagulation process consists of two steps: the conversion of dissolved NOM to particulate matter during coagulation and flocculation, and the removal of the particulates during sedimentation and filtration (Malcolm Pirnie Inc. 1993). A general discussion of the enhanced coagulation process, as well as detailed descriptions of the mechanisms of removal is presented by Malcolm Pirnie Inc. (1993).

2.2.5 Factors Affecting NOM Removal by Enhanced Coagulation

Many factors, both chemical and physical, must be controlled in order to maximize DBP precursor removal by enhanced coagulation.

2.2.5.1 Nature of the Organics

Organic matter occurring in natural waters can be characterized by its molecular weight distribution (MWD), humic substance content, and hydrophobicity, all of which affect its removal by enhanced coagulation. Typically, enhanced coagulation removes humic and

high molecular weight fractions better than it removes non-humic and low molecular weight fractions (Cheng et al. 1995). In addition, hydrophobic organic molecules are preferentially removed over their hydrophilic counterparts (Collins, Amy and Steelink 1986).

Separating the humic substances into humic acid and fulvic acid fractions yields additional information on the removability of natural organic matter by enhanced coagulation. Humic acids, having higher molecular weights than fulvic acids, are preferentially removed (Krasner and Amy 1995). Fulvic acids typically require approximately twice the coagulant dose as humic acids for effective removal (Kavanaugh 1978). With respect to the relative reactivities of the two humic components, the fulvic acid fraction is less reactive than the humic acid fraction (Joyce, DiGiano and Unden 1984). As such, even though fulvic acids comprise over 80 % of the organic matter found in natural waters, the more reactive high molecular weight humic acid fraction is preferentially removed.

2.2.5.2 Alkalinity

For the USEPA D/DBP rule, TOC removal requirements are based on raw water alkalinity, suggesting that the alkalinity is an important parameter for DBP precursor removal. Alkalinity is an extremely important economic parameter since, along with the coagulant type and dose, alkalinity governs the pH which can be achieved without resorting to the use of supplemental acid or base (Randtke 1988). Waters with higher alkalinity (80 -100 mg/L as $CaCO_3$) will have good buffering capabilities and will resist pH depression by the coagulant (Malcolm Pirnie Inc. 1993). If a low coagulant dose is

applied to high alkalinity water however, the pH of the system may be well above the optimum for organics removal. Conversely, if the alkalinity of the raw water is low, the pH obtained may be satisfactory for organics removal, but the quantity of solids precipitated may be too low to coprecipitate some of the soluble organics that could otherwise be removed (Randtke 1988). In summary, NOM removal will be favored by lower raw water alkalinities, given the same dosing conditions, provided that the pH of the water does not drop below the optimal range upon coagulant addition (Malcolm Pirnie Inc. 1993).

2.2.5.3 Inorganics, Suspended Solids, and Turbidity

The effect of divalent cations and anions, suspended solids, and turbidity-causing particles on the removal of NOM has been thoroughly studied. Divalent cations can broaden the pH range for effective removal and lower the required coagulant dosage (Randtke 1988). In particular, the Ca^{2+} cation was found to enhance the removal of fulvic acid, which is generally difficult to remove, when alum or PACL is used as the coagulant (Hundt and O'Melia 1988). Conversely, the SO_4^{2-} and OH^- anions may inhibit the coagulation of organics by competing with organics in solution for hydrolyzed coagulant species and adsorption sites (Jacangelo et al. 1995). Other monovalent anions, such as Cl^-, were found to have little effect on the removal of organics (Harrington, Chowdhury and Owen 1992). Most authors agree that suspended solids, as well as turbidity-causing particles, have little effect on the removal of humic substances and colour (Dempsey, Ganho and O'Melia 1984 ; Ollier, Miltner and Summers 1997). In raw waters with low colour concentrations however, a small amount of turbidity may decrease the required

coagulant dose by providing adsorption sites for the organic molecules (Edwards and Amirtharajah 1985).

2.2.5.4 Temperature and Biological Activity

Seasonal variations in both raw water temperature and biological activity affect the influent water quality and its treatability. During the warmest months, high levels of NOM are likely to exist, coinciding with the season of peak water demand. As such, it is necessary to ensure that treatment processes are able to cope with increased organic loading. Water temperature does not directly affect the removal of TOC by coagulation, although colour removal is impaired at low temperatures (< 2°C) (Knocke, West and Hoehn 1986). In addition, the solubility of alum is affected by low temperatures (< 4°C), as the pH of minimum alum solubility is lowered by approximately 0.5 units (Edzwald 1993). This reduced solubility can result in an increase in the required alum dose. With respect to biological activity, the effectiveness of reducing NOM does not appear to be affected by variations in algal or bacterial populations (Hoehn et al. 1984). Biota can however, contribute to the organic loading of the raw water and thereby increase the concentration of NOM.

2.2.5.5 Coagulation pH

The optimal pH of coagulation for the removal of organic matter is dependent on the coagulant used. For alum-based treatment, the optimal pH range has been variously reported as 5 - 6 (Malcolm Pirnie Inc. 1993) (Kavanaugh 1978), 5.5 - 6.0 (Singer 1994),

and 5.0 - 6.5 (Edwards and Amirtharajah 1985). For Fe(III)-based coagulants, the optimal range is shifted to lower values and has been reported as 3 - 5 (Kavanaugh 1978), 4 - 6 (Amirtharajah, Dennett and Studstill 1993), and 4 - 5 (Malcolm Pirnie Inc. 1993). More detailed studies suggest that the optimal coagulation pH is also dependent on the nature of the organic matter. For alum coagulation, colour caused by humic acids is best removed at pH < 5.5, while colour caused by fulvic acids is best removed at 4.5 < pH < 5 (Babcock and Singer 1979). More recent studies suggest that the coagulation pH is not statistically significant by itself when determining optimal coagulation conditions (Harrington, Chowdhury and Owen 1992). Rather, optimal regions of organic removals are defined by specific combinations of alum dose and coagulation pH. The first region, in which the charge-neutralization mechanism predominates, typically occurs when a coagulation pH is 4.00 - 4.75 is combined with alum doses of 15 - 80 mg/L (Edwards and Amirtharajah 1985). At pH 5.75 - 7.50 and alum doses > 30 mg/L, the organic matter is adsorbed to aluminum-hydroxide precipitates.

2.2.5.6 Type of Coagulant

Many compounds including alum, ferric chloride, polyaluminum chloride (PACL), cationic polymers, aluminum chlorohydrate, ferric sulphate, sodium aluminate, ferric chlorosulphate, and polyiron chloride have been used as primary coagulants in water treatment facilities (Edzwald 1993). In spite of the variety of coagulants available, most utilities use either alum or ferric chloride as the primary coagulant. In general, ferric coagulants outperform alum coagulants at higher doses, on an equal dose basis, while alum has the edge at lower doses (Edwards 1997 ; Randtke and al. 1994). Ferric chloride

produces approximately 2.8 times more metal hydroxide, produces less sludge, and is more effective at lowering the coagulation pH than alum, making it more suitable to enhanced coagulation (Crozes, White and Marshall 1995). In comparing these metal-salt coagulants to cationic polymers, the metal-salt coagulants are typically more effective at removing low molecular weight NOM (Randtke 1988). Synthetic organic cationic polymers can achieve colloidal charge neutralization and possibly participate in the precipitation of humic and fulvic acids, but they do not provide a substrate for adsorption of the organic matter (Crozes, White and Marshall 1995). Polyaluminum chloride is sometimes used as the primary coagulant in low turbidity waters, although the doses used are generally insufficient to remove NOM (Malcolm Pirnie Inc. 1993). Other coagulants such as aluminum sulphate (Edwards and Amirtharajah 1985), aluminum chlorohydrate (Williams et al. 1996), and magnesium salts (AWWA Committee 1979) have been proposed for colour and organics removal, although very few full-scale applications of these coagulants have been identified.

2.2.5.7 Coagulant Dose

The coagulant dose required for optimal DBP precursor removal is highly dependent upon the source water characteristics, as well as the type of coagulant used. As such, there is no single optimal coagulant dose for NOM removal, only general trends and dose/removal profiles. In general, except at very low colour concentrations, the coagulant dose required for colour removal is greater than that required for turbidity removal alone (Edwards and Amirtharajah 1985). The removal of NOM generally follows one of two dose/removal patterns. Removal either increases very sharply in the

vicinity of a particular coagulant dosage (type 1 pattern) or increases very gradually with increasing dosage and eventually becomes asymptotic (type 2 pattern) (Randtke 1988). Type 1 behavior is indicative of a precipitation mechanism and is associated with low pH (5-6 for alum) and high concentrations of NOM that is predominantly humic. Type 2 is associated with relatively high pH, low NOM, and non-humic materials. A stoichiometric relationship between the NOM and the coagulant exists for the type 1 pattern, but not necessarily for type 2. Type 2 is most common for surface waters, especially where turbidity and/or alkalinity are high. With the type 2 pattern, it is possible to reach a point of diminishing returns where dosing beyond the minimum required for effective removal of organics does not substantially improve removals (Vrijenhoek et al. 1998).

2.2.5.8 Mixing Conditions

Colour removal by enhanced coagulation is a function of the rapid mix time, the mean velocity gradient (G), and the flocculation time (T). For most waters, the optimal rapid mix time for colour removal was found to be 30 - 60 seconds (Bowie and Bond 1977). With respect to the flocculation characteristics, the mean velocity gradient should not exceed 100 s^{-1} in order to ensure proper floc development (Kawamura 1973). Organics removal was found to be unaffected by changes in the flocculation time in the range of 10 - 30 minutes, however shorter flocculation times may not produce adequate floc for organics removal (Semmens and Field 1980).

2.2.5.9 Addition of PAC

Powdered activated carbon (PAC) is commonly used in concentrations of 5 - 10 mg/L on a seasonal basis to control taste and odour problems. At these low concentrations, PAC is typically ineffective at removing NOM due to insufficient contact with the organics (Randtke 1988). At elevated concentrations (> 50 mg/L) however, the PAC acts as a flocculation nucleus and can significantly reduce the amount of NOM present in the water (Chadik and Amy 1983). The effectiveness of PAC addition can be improved by using smaller diameter PAC in order to maximize surface area, and by adding the PAC before the coagulant in order to maximize contact time (Randtke 1988).

2.2.5.10 Addition of Polymer

While generally ineffective as a primary coagulant for the removal of NOM, polymeric coagulant aids can reduce the dose of coagulant required to achieve acceptable NOM removals by as much as 44 % (Hubel and Edzwald 1987). Most studies conclude that the addition of polymeric coagulant aids does not increase organics removal (Randtke 1988) (Hubel and Edzwald 1987), although one study suggests that elevated concentrations (2 mg/L) of cationic polymer can reduce the THMFP of finished water by up to 22 % (Chadik and Amy 1983). At such elevated polymer doses however, the floc may restabilize resulting in poor turbidity removals (Edzwald, Haff and Boak 1977).

2.2.5.11 Preoxidation

The use of strong oxidants, such as ozone, can alter the chemical properties of NOM, making it less amenable to coagulation. Preozonation can convert some humic material into non-humic material and reduce the molecular weight of the organic matter by severing the carbon backbone. As such, preozonation can lead to a reduction in UV absorbance and colour without reducing TOC (Edzwald, Decker and Wattier 1985). In addition, ozonation increases the polarity and the average acidity of the NOM, resulting in a diminished tendency of the NOM to bind to the coagulant (Owen et al. 1995).

2.3 Conventional Control and Modelling of Coagulation

2.3.1 General Strategies

2.3.1.1 Batch Dose Control

The most common method of determining optimal coagulation conditions is the jar test, a batch test consisting of a series of 0.5 - 2.0 L jars that serve as bench-scale clarifiers. The jar test may be used for coagulant selection, dose selection, coagulant aid and dosage selection, determination of optimal pH, determination of the point of chemical addition, optimization of mixing times and energies, and determination of coagulant feed concentration (Amirtharajah and O'Melia 1990). While many utilities develop their own jar test methodologies, a sample protocol for the determination of coagulant dose is presented by Amirtharajah and O'Melia (Amirtharajah and O'Melia 1990). In order to

approximate the full-scale system, it is essential that the relative dimensions of the jar, as well as the method of chemical preparation, temperature, mixing intensity, and mixing time are all similar to that found in the full-scale system. When used for coagulant dose selection, any one of a number of methods including turbidity, colour, streaming current detection, and particle counts, can be applied to the jar effluent (Dentel 1991). The optimal dose is often selected by plotting the selected parameter against the doses used in the jars and selecting the dose that best meets the utility's process criteria.

There are several key disadvantages to using jar tests for coagulation optimization. First and foremost is the frequency of testing. Being batch tests, jar tests are only able to provide a snapshot of influent water quality and are unable to represent the continuously changing dynamics of the full-scale system. Jar tests are often performed infrequently throughout the plant operator's shift and often after clarified water quality begins to degrade. Adjusting coagulant doses based on the results of jar tests is reactive rather than proactive; dosing levels generally can not be adjusted until an upset occurs. As such, the magnitude of the upset is often magnified due the time lag between the change in influent water quality and the chemical dosing adjustments. In addition the requirement to now determine the optimal dose for both particulate and organic removal adds significant complexity to jar testing methodologies. The optimal dosing levels determined by the bench-scale jar tests may also differ from those in full-scale operations due to the differences in the hydrodynamics of the two systems. In addition, clarifiers with sludge recycling are inherently difficult to approximate at the bench scale. In spite of these

limitations, the jar-test is widely used for determining dosing levels as few alternatives exist.

2.3.1.2 Continuous Dose Control

Continuous control of coagulant dose can be accomplished via flow-paced chemical feed, electrokinetic methods, and turbidimetric methods. In flow-paced chemical feed systems, the coagulant feed rate is adjusted according to the raw water flow rate, thereby maintaining a constant coagulant dose (Dentel 1991). This method does not address the need for varying the coagulant dose according to influent water characteristics and is therefore unsuitable for all but the best quality raw waters.

The main electrokinetic method involves the estimation of the zeta potential of a sample using a streaming current detector (SCD). The zeta potential is a measure of the electric potential surrounding particles in solution. Suspensions that are well destabilized by the charge-neutralization mechanism of coagulation have a zeta potential value near zero. A streaming current detector is a cylindrical device containing a piston with a reciprocating motion and an annular space between the piston and the cylinder. A sample of water is drawn into and pushed out of the annular space as the piston moves up and down, generating an alternating or streaming current at electrodes affixed to the ends of the cylinder (Amirtharajah and O'Melia 1990). The streaming current is directly proportional to the charge on the particles in the sample, and a pH specific correlation can be made between the SCD reading and the zeta-potential of the sample. The SCD can be

implemented in feedback loop control systems to automatically adjust the coagulant dose for effective destabilization of the particles (Amirtharajah and O'Melia 1990). Electrokinetic methods are only useful if the charge-neutralization mechanism of coagulation predominates.

Turbidimetric methods involve the use of continuous online turbidity measurements following either flocculation or sedimentation for feedback loop coagulant dose adjustment. Such methods are rarely used due to the difficulties in the provision of a trouble-free means of continuous turbidity measurements and the implementation of a successful control algorithm (Dentel 1991). Conventional turbidity meters are subject to clogging, and damage through continuous use. With respect to the control algorithm, conventional set-point control strategies are not always applicable. If excessive coagulant dosing results in elevated turbidity measurements, for example, the controller would attempt to control the turbidity by increasing the coagulant dose, clearly the wrong control action.

2.3.2 Proposed Models

A model for the removal of NOM by coagulation was developed by Harrington using a forward stepwise variable selection procedure with multiple linear regression analysis (Harrington, Chowdhury and Owen 1992). Initially, the model considered the effects of raw water TOC, alum dosage, coagulation pH, temperature, and chloride concentration on finished water TOC, although the later two parameters were dropped due to lack of

statistical significance. The resulting model had an adjusted r^2 of 0.965 based on 44 observations. When applied to data from associated databases, the model tended to under-predict finished water TOC by approximately 7 %. Harrington's model verification is based on a narrow range of water quality and operating conditions (Greiner, Obolensky and Singer 1992). As such, it may not be able to accommodate site-specific deviations from the reference databases, resulting in even greater prediction errors.

General and site-specific models for the removal of DOC during enhanced coagulation, proposed by Edwards, predicted the final DOC with a standard error of approximately 12 % (Edwards 1997). The model inputs were coagulant dosage, coagulation pH, raw water UVA, and raw water DOC. The author cautions that the excellent predictive ability of the model should not be construed to imply conceptual validity until further tests are completed.

2.3.3 Process Optimization and Control at Rossdale and E.L. Smith WTPs

The coagulation process at both of the WTPs has traditionally been controlled by manual operator control. For typical on-line operation control, the operators assess the quality of the water, using jar-tests or their experience, and then adjust the operational parameters as they see fit. Both facilities have a supervisory control and data acquisition (SCADA) system in place to assist in process operation, however these systems serve only to relay control information and do not generate control actions themselves. In many large water treatment plants, there may be some dosing control strategy in place however, such

dosing strategies normally only serve as a guideline and are not rigorously followed. At Rossdale WTP for example, a strategy exists for selecting alum dose levels based on clarifier effluent colour. This strategy is rarely invoked however, as each operator maintains operational control using strategies derived from personal experience. At E.L. Smith WTP, the overriding goal for the coagulation process is to reduce clarifier effluent turbidity. Again, each operator achieves this goal using a different strategy based on personal experience.

2.4 Overview of ANN Modelling

2.4.1 General Characteristics of ANNs

The ANN modelling technique is an artificial intelligence technique that simulates the human brain's problem solving processes. Unlike the human brain, which consists of an estimated 10^{11} neurons with 10^{14} connections, the most complex ANNs consist of less than 1000 processing elements or artificial neurons and 10^{6} connections (Flood and Kartam 1997). Artificial neural networks are capable of self-organization and learning; concepts and patterns can be extracted directly from historical data without the need for complex mathematical formulae or algorithms. In general, artificial neural networks can be applied to the following types of problems: pattern classification, clustering and categorization, function approximation, prediction and forecasting, optimization, associative memory, and process control (Jain, Mao and Mohiuddin 1996). ANNs are particularly useful if the characteristics of the specific problem are known, but the

connections between the data that are necessary to solve the specific problem are not (Buscema 1997). In the water treatment industry, for example, ANNs can be applied to processes where the important factors are known but the reactions and interactions between the factors are difficult to elucidate.

The ANN technique holds several advantages over other statistical methods of data analysis. With respect to data processing, the type of relationship (linear, non-linear, exponential, etc.) between input and output data is determined purely from the information presented, with no in-built assumptions from the ANN (Harvey and Harvey 1998). In addition, discontinuities in the data, different levels of data measurement precision and noise, and data scatter are easily accommodated (Foody and Arora 1997). The feature that really sets ANNs apart from other statistical methods however, is the ability to self-organize or learn. This feature allows ANNs to produce correct or nearly correct responses when presented with partially incorrect or incomplete stimuli, and to generalize rules from the training cases and apply these to new cases (Garret, Ghaboussi and Wu 1992).

With respect to the disadvantages of the technique, ANNs are data intensive and are only suited to problems where large quantities of data exist (Stanley, Zhang and Baxter 1998). In addition, the ANN technique is considered to be a "black-box" technique; ANN models do not yield explicit mathematical formulae (Harvey and Harvey 1998). Current research efforts in the field are aimed at unraveling the black box in order to extract heuristics and rules from model architectures. In addition, little research has been

conducted on the ability of ANNs to extrapolate beyond their knowledge base. As such, if new patterns contain data which is more that 2 standard deviations away from the mean, ANN predictions should be used with caution (Boger 1997). Alternatively, the ANN can be retrained to enlarge its knowledge base. Finally, due to the ease of obtaining accurate fits to data sets, ANNs are often misused and misapplied (Chen et al. 1996). The technique is constantly being applied to problems with insufficient or incorrect data, resulting in unfounded conclusions concerning ANN performance. Models are also often constructed without due consideration to optimization, resulting in sub-standard performance.

2.4.2 Historical Development of ANNs

The history of artificial neural networks can be divided into three eras according to the technique's acceptance by the scientific community: initial enthusiasm, lack of support, and re-emergence. An excellent account of the history of neural networks can be derived from the collection of key papers assembled by Anderson and Rosenfeld (1988). What follows is a brief synopsis of some of the key developments in the field of artificial neural networks.

In 1943, McCulloch and Pitts published what is considered by many to be the founding paper of the field. In the paper, the authors list five main assumptions governing the operation of biological neurons: 1) the activity of a neuron is an "all-or-none" process, 2) a certain fixed number of synapses must be excited in order to excite a neuron 3) the only

significant delay within the nervous system is synaptic delay 4) the activity of any inhibitory synapse absolutely prevents excitation of the neuron at that time 5) the structure of the neural net does not change with time (McCulloch and Pitts 1943). The resulting model of a neuron, referred to as the McCulloch-Pitts Neuron, operates in a similar fashion to the nodes found in modern ANNs. The neuron responds to the activity of its synapses, which reflect the state of the presynaptic cells (Anderson and Rosenfeld 1988). If no inhibitory synapses are active, the neuron sums its synaptic inputs and becomes activated if the sum exceeds a certain threshold.

The next major development came in 1958 when Rosenblatt designed and introduced the perceptron, a conceptual model of neural activity in the human brain. The perceptron consists of three layers: sensory units (S-points), association cells (A-units), and responses (R-units) (Rosenblatt 1958). Stimuli impinge on the S-points, which in turn transmit impulses to the A-units. If the sum of the impulses arriving at a particular A-unit exceed a threshold value, the unit is activated and transmits an impulse to a particular R-unit. Rosenblatt recognized the inherent power in the theories he developed, suggesting that the perceptron could be useful in pattern recognition, associative learning, temporal and spatial pattern recognition, and trial-and-error learning (Rosenblatt 1958). Rosenblatt's work was well received; seemingly overnight, a hundred learning algorithms bloomed and a hundred schools of learning machines contended (Anderson and Rosenfeld 1988). Perceptrons, with a single layer of A-units, were developed and were able to discriminate between linearly separable patterns. When such systems were tested

with complex patterns that were not linearly separable however, little progress was seen and many researchers became disenchanted with the field (Harston and Maren 1990).

In 1969, Minsky and Papert published a book that effectively eliminated funding for neural network development. *Perceptrons* , a book on the mathematics and the theory of computations, discusses the limits of Rosenblatt's perceptrons. Minsky and Papert suggest that much of the early work on perceptron development was without scientific value but proceeded due to the romantic enchantment with the new ideas of machine learning (Minsky and Papert 1969). The authors suggest that simple perceptrons are limited with respect to their connectedness. Two such classes of limitations include diameter-limited perceptrons, where each associative unit can only be associated with an area of fixed diameter on the sensory surface, and order-restricted perceptrons where each associative unit can only be connected to a set number of points on the sensory surface. In either case, simple perceptrons are not able to determine whether all the parts of a geometric figure are connected to each other (Minsky and Papert 1969). In order to overcome this difficulty, the network would have to be excessively large, making it impractical for all but the simplest of problems. In the concluding chapter of the book, the authors suggest that multi-layer variants of the simple three-layer perceptron would not prove any more fruitful than their less complex counterparts. This assertion became the mainstream opinion until the resurgence of neural networks in the mid-1980's.

In 1986, Rumelhart, Hinton, and Williams published a two-volume book *Parallel Distributed Processing* which contains, among other things, the first well-known

description of the back-propagation learning algorithm. The algorithm had been described earlier by Werbos in his Ph.D. thesis in 1974 and was simultaneously re-discovered by Parker in 1985 and Le Cun in 1986, yet Rumelhart is generally considered to be the father of back-propagation (Anderson and Rosenfeld 1988). Since 1987, the back-propagation learning algorithm had been the most popular algorithm for multilayer networks. Error propagation in the back-propagation algorithm involves the use of the *generalized delta rule*, which minimizes the squares of the differences between the actual and the desired output values summed over the output units and all pairs of input/output vectors (Rumelhart, Hinton and Williams 1986). The ability of the back-propagation algorithm to solve a wide variety of problems using multi-layer architectures has since been demonstrated many times, thereby effectively negating the misplaced comments of Minsky and Papert with regards to multi-layer perceptrons.

Since Rummelhart's developments, research in the area of ANNs has once again intensified. As will be discussed further on, applications using the back-propagation and other algorithms have provided new insight into a wide variety of processes and systems.

2.4.3 Components of ANN Architectures

There are seven major components to an ANN model, which are collectively known as the ANN architecture: 1) processing units 2) a state of activation 3) an output function for each unit 4) a pattern of connectivity among units 5) a propagation rule for propagating patterns of activities through the network connectivities 6) an activation

function for combining the inputs impinging on a unit with the current state of that unit to produce a new level of activation for that unit 7) a learning rule whereby patterns of connectivity are modified by experience (Rumelhart, Hinton and McClelland 1986).

2.4.3.1 Processing Units

]
ANN models are comprised of interconnected arithmetic processing units, also called artificial neurons, neurodes, or units, that are analogous to the biological neurons in the human brain. Each neuron is an elementary processor that performs primitive operations, like summing the weighted inputs coming to it and then amplifying or thresholding the sum (Pal and Srimani Pradip 1996). A schematic diagram of a single neuron is presented in Figure 2.4. In multi-layer architectures, there are three basic types of processing units: input units, output units, and hidden units (Garrett, Gunaratman and Ivezic 1997). The units are arranged in layers, the input layer, the output layer, and the hidden layer(s). Input units receive input from external sources, compute their activation level, compute their output as a function of activation level, and transmit this output to the next layer of units in the system. Output units receive input from the rest of the network and compute or transmit their output to external receivers or, in the case of recurrent networks, back to the input layer for further processing. Hidden units only receive their input from, and transmit their computed output to, other layers in the network.

2.4.3.2 State of Activation

The state of activation is an overall picture of the activation level of each of the neurons in the network, which typically have a value in the continuous range between 0 and 1 (Garrett, Gunaratman and Ivezic 1997). Input and output unit activation levels represent the current external input to the input processing units, and the current output being transmitted to the external receivers, respectively. The activation level of the hidden units represents the features within the input pattern that are present, and influence the output produced by the network (Rumelhart, Hinton and McClelland 1986).

2.4.3.3 Output Function

Associated with each processing unit is an output function which defines how the output value for the processing unit is determined from its activation (Garrett, Gunaratman and Ivezic 1997). In some models, this function is unity. In other models the output function is a threshold function; a processing unit produces no output unless the activation exceeds some predefined level of activation (Rumelhart, Hinton and McClelland 1986).

2.3.3.4 Pattern of Connectivity

Processing units are connected to each other and communicate with each other via connection weights. The pattern of connectivity influences how a network will respond to a given set of inputs (Garrett, Gunaratman and Ivezic 1997). A positive weight represents

an excitatory input, a negative weight represents an inhibitory input, and a zero weight represents an inactive input; the absolute value of the weight is the connection strength (Rumelhart, Hinton and McClelland 1986).

2.3.3.5 Rule of Activation Propagation

The rule of propagation describes how the inputs connected to a processing unit and the strengths of the connections are combined to compute the net input (Garrett, Gunaratman and Ivezic 1997). The input is generally expressed as the weighted sum of the outputs from connected neurons (Figure 2.4).

2.3.3.6 Activation Function Employed

The activation function defines how the net input received by a processing unit is combined with its current level of activation to compute a new level of activation (Garrett, Gunaratman and Ivezic 1997). When ANN learning is accomplished using the backpropagation algorithm, a sigmoidal activation function, which is continuous and differentiable, is often used to maintain the value of activation for a processing unit between 0 and 1. The mathematical expression of the logistic function, the most popular sigmoidal function, is (Jain, Mao and Mohiuddin 1996) :

$$f(x) = 1/(1+\exp\{-\beta x\})$$

where β is the slope parameter.

2.3.3.7 Rule of Learning

The learning rule defines how the network is modified in response to training patterns presented to the system (Garrett, Gunaratman and Ivezic 1997). Network modification can take place via the development of new connections, the loss of existing connections, and the modification of the strengths of connections that already exist. As will be discussed in the next section, the last of the three methods is the most common.

2.4.4 The ANN Learning Process

While many learning algorithms are available, the backpropagation algorithm is the most common and is used in the current study. As such, much of the ensuing discussion concerning learning focuses primarily on the backpropagation algorithm. The ability to learn is a fundamental trait of intelligence ; the learning process for ANNs can be viewed as a problem of updating network architecture and connection weights so that a net can perform a specific task (Jain, Mao and Mohiuddin 1996). In the learning process, each presentation of a pattern and subsequent modification of connection weights is called a learning cycle. A set of cycles, one for each pattern is called an epoch. The learning cycle has 3 steps (Garret, Ghaboussi and Wu 1992). 1) For the training pattern to be learned, the network is presented with the input pattern and then propagates the activation through to the output processing units. 2) The output units then backpropagate their errors back to the hidden processing units according to the generalized delta rule (GDR), or another learning rule. 3) The units then modify their incoming connection strengths using this

backpropagated error and the learning rule. The entire cycle is repeated until the ANN produces a sufficiently small error on a previously unseen data set (Boger 1992).

There are three broad paradigms of learning: supervised, unsupervised, and reinforcement (Pal and Srimani Pradip 1996). In supervised learning, adaptation occurs when the system directly compares the network output with a known output. In unsupervised learning, the network is tuned to the statistical regularities of the input data so that it can form categories by optimizing, with respect to the network's free parameters, a task-independent measure of the quality of the net's category representation. Reinforcement learning attempts to learn the input-output mapping through trial and error with a view to maximizing a performance index called the reinforcement signal. The system knows whether the output is correct or not, but does not know the value of the correct output. The backpropagation algorithm uses supervised learning implemented in two stages. In the forward stage, the network node outputs are computed, in the backward one, weights are adjusted to minimize the error between the observed and desired model output (Pal and Srimani Pradip 1996).

There are four basic types of learning rules: error-correction, Bolltzmann , Hebbian, and Competitive (Jain, Mao and Mohiuddin 1996). The generalized delta rule (GDR) used in backpropagation networks is a gradient descent error-correction learning rule. The GDR is applicable only for semi-linear activation functions and layered feed-forward nets where the input units form the bottom layer, the outputs form the top and all others are

hidden (Garret, Ghaboussi and Wu 1992). The rule can be summarized by the following three equations (Rumelhart, Hinton and Williams 1986):

$$\Delta_p w_{ji} = \eta \delta_{pj} o_{pi}$$

$$\delta_{pj} = (t_{pj} - o_{pj}) f_j'(net_{pj})$$

$$\delta_{pj} = f_j'(net_{pj}) \sum_k \delta_{pk} w_{kj}$$

In the first equation, the change in the weight (Δw) from the j^{th} to the i^{th} unit following the presentation of pattern p is proportional to the product of an error signal, δ, available to the unit receiving input along that line and the output (o) of the unit sending activation along that line. The symbol η represents the learning rate of the system and has a value between 0 and 1. The other two equations represent the error signal, the determination of which is a recursive process starting with the output unit. If a unit is an output unit, its error signal is given by the second equation where t_{pj} is the target input for the j^{th} component of the output pattern for pattern p, o_{pj} is j^{th} element of the actual output pattern produced by the presentation of input pattern p, and $f_j'(net_{pj})$ is the derivative of the semilinear activation function which maps the total input to the unit to an output value. Finally, the error signal for hidden units for which there is no specified target is determined recursively in terms of the error signals (o_{pk}) of the units to which it directly connects and the weights (w_{kj}) of those connections. For a more detailed description of the generalized delta rule, as well as for the complete derivation of the above equations, please refer to the text by Rumelhart (1986).

2.4.5 ANN Architectures

ANN architectures can be separated into two broad categories, feed-forward and feedback, according to the direction of communication of input information (Garret, Ghaboussi and Wu 1992). In feed-forward nets, communication takes place in the forward direction only, although errors can be backpropagated. In feedback or recurrent nets, there is iterative computation between the nodes as communication is cyclic. Within each of the two main classes of ANNs, there are a number of families that arise from the variances in the activation functions and learning rules applied. A list of the most common families is presented in Table 2.4; a more detailed description of these network architectures can be found in the text by Jain (1996). The most common family of feed-forward nets, multi-layer perceptrons, have layers of neurons that are completely connected, although there is no intra-layer connectivity (Pal and Srimani Pradip 1996). A sample diagram showing the interconnectivity of multi-layer perceptrons is presented in Figure 2.5. Feed-forward nets are memoryless in the sense that their response to an input is independent of the previous network state. (Jain, Mao and Mohiuddin 1996). Recurrent nets are dynamic systems; when a new input pattern is presented, the neuron outputs are computed. Because of feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state. (Jain, Mao and Mohiuddin 1996).

Table 2.4 Classification of network architectures (after Jain, Mao and Mohiuddin 1996)

Feed-Forward Networks	Feedback Networks
Single-Layer Perceptron	Competitive
Multilayer Perceptron	Kohonen's Self-Organizing Mapping
Radial Basis Function	Hopfield
	Adaptive Resonance Theory

2.4.6 Model Development and Evaluation

In conventional modelling, model development consists of three stages: data selection, mathematical model specification, and evaluation (Schwerk 1996). With respect to ANN model development, the model specification stage includes the selection of training cases, the selection of architecture and form details, and the actual training of the network (Garret, Ghaboussi and Wu 1992).

2.4.6.1 Data Selection

The selection of the type and quantity of data is a key step in model development. A network will struggle if presented with too little or too much data or if presented with data that doesn't emphasize differentiating components (Villarreal and Baffes 1992). If the network is not trained using relevant data, it will be unable to make accurate predictions on a generalized data set (Schwerk 1996). Most researchers advocate splitting the data randomly into training and testing data sets (Boger 1997 ; Garret, Ghaboussi and Wu 1992); the training set is used to train the model, while the testing set is used to test its performance on previously "unseen" data. The ratio of data in each set can be arbitrarily assigned (Boger 1997) or determined experimentally (Bernieri et al. 1996).

2.4.6.2 Model Specification

With respect to the selection of training cases, there should be at least 10-30 times as many training patterns as there are input variables (Foody and Arora 1997). Increasing

the number of training patterns provides more information about the shape of solution surfaces and thus increases the potential level of accuracy that can be achieved by the network. Large training sets can however, overwhelm some training algorithms, leading to the net getting stuck in a local minimum (Flood and Kartam 1997). Since ANNs are not usually able to extrapolate, the input patterns should be selected to encompass the entire problem domain. It is also advisable to have the training patterns evenly distributed throughout the domain, else the net will tend to focus on areas where patterns are tightly clustered. In this respect, it may be beneficial to have more patterns in areas where the response surface is more complicated. In practice though, it is not always possible to control the distribution of data, especially when historical data are used (Flood and Kartam 1997).

There are many aspects of network architecture design that will affect the ability of and ANN model to generalize, that is the ability to correctly predict outputs for new or previously unseen data cases. Some of the key considerations include: architecture type, the number and size of the hidden layers, and the learning rate and momentum (Foody and Arora 1997 ; Schwerk 1996). There are currently no good heuristics or rules of thumb for selection of network architecture, however experience has shown that there are definitely good and bad architectures for a given problem (Garret, Ghaboussi and Wu 1992). Until further advances in ANN model development techniques are made, empirical methods and experimentation must be used to determine optimal architecture configurations (Villarreal and Baffes 1992 ; Yabunaka, Hosomi and Murakami 1997). With respect to the size of the network, two issues must be addressed: the number of

hidden layers, and the number of neurons per layer. When the output is a continuous function, some researchers advocate the use of a single hidden layer (Hecht-Nielsen 1987), while others suggest the use of two hidden layers (Flood and Kartam 1997). In general, two hidden layers provide more flexibility at modelling more complex surfaces however, only one layer is required for problems where the network has only one input or where the input values are binary (1 or 0) (Flood and Kartam 1997). With respect to the number of neurons in each layer, input and output layers contain the same number of neurons as there are inputs and outputs (Villarreal and Baffes 1992). Generally, there is no direct and precise way of determining the most appropriate number of neurons to include in each hidden layer. Increasing the number of hidden nodes will increase the likelihood that the network will converge, but an excess of hidden nodes will result in poor generalization capabilities (Villarreal and Baffes 1992 ; Shang and Wah 1996). Some researchers have made broad suggestions, with regards to the number of nodes, that may not be applicable to every problem. Hecht-Nielson, for example, suggests that when a single hidden layer architecture is used, an upper limit of $2n+1$ hidden layer neurons, where n is the number of inputs, should be used (Hecht-Nielsen 1987). For networks with two hidden layers, a 3:1 ratio of neurons in the two hidden layers is suggested (Maier and Dandy 1996). Weigend suggests that the number of hidden neurons should be less than 1/10 the number of training patterns in order to avoid overfitting (Weigend, Huberman and Rumelhart 1990).

The prediction accuracy and generalization capabilities of a network can be affected by two other factors, the learning rate and the momentum factor, which alter the speed of

convergence (Schwerk 1996). When a net is trained, the weights are initially randomized. The amount by which the weights are modified in any given training step is the learning rate. If the learning rate is small (<0.2), the weights will be changed in smaller increments resulting in slower convergence with less oscillation (Garret, Ghaboussi and Wu 1992). Larger learning rates (>0.5) will cause the weights to be adjusted more drastically, but this may cause the optimum combination of weights to be overshot, resulting in oscillations about the optimum. (Garret, Ghaboussi and Wu 1992). In general, lower learning rates should be used for more noisy data (Schmuller 1990). The momentum factor, which has a value between 0 and 1, is a smoothing factor that allows for faster learning without oscillation by making the weight change a function of the previous weight change (Ward Systems Group Inc. 1996).

With respect to the actual training of the network, it is possible to over-learn if training proceeds too long. The situation is analogous to learning to drive a car by driving around the block over and over again. All of the necessary skills are learned, but the driver would have difficulty applying those skills correctly to highway driving (Villarreal and Baffes 1992). DeSilets suggests a maximum of 2000 iterations (epochs) (DeSilets et al. 1992), although in practice, this number must be experimentally determined. Another issue involved with lengthy training periods is the possibility that the network will try to fit noise in the data. Assuming that the noise is small in comparison to the features, it is only in the later stages that the net will try to fit noise. Stopping the training before the noise is learned will improve generalization. (Weigend, Huberman and Rumelhart 1990).

2.4.6.3 Model Evaluation and Validation

In order to evaluate the performance of a model, a variety of qualitative and quantitative means can be employed. Most researchers advocate the use of only two data sets, the training set and the testing set, as previously discussed. Other studies advocate the use of a third data set, the production set (Weigend, Huberman and Rumelhart 1990). The utility of the production set will be discussed in the methodology section of this report. In the case of the former, the testing set is most commonly used to assess a model's performance, while for the later, the production set is used. With respect to qualitative means of evaluation, a visual inspection of a plot of actual outputs versus predicted outputs can serve as measure of the fitness of a model (Flood and Kartam 1997). Quantitatively, the fitness of a model can be assessed using a statistical indicator such as the mean squared error, the mean absolute error, or the R^2 parameter (Schwerk 1996).

In order to ensure that the ANN is not simply memorizing the training patterns, the model must be validated. The most common means of validating a model involves swapping the training and testing sets, retraining the model, and comparing the new model's performance with that of the old one (Schwerk 1996). This cross-validation technique can also be applied when 3 data sets are used by swapping the production and training or production and testing sets, as will be discussed in the methodology section.

2.4.7 Applications of ANN Modelling

The ANN technique has been applied to virtually every field where complex non-linear error minimization problems exist. An extensive list of known and potential applications in fields as diverse as finance and health-care is presented in the text by Buscema (1997). The remainder of this section focuses on known and developing applications in the field of environmental engineering.

2.4.7.1 Environmental Engineering Applications

Civil and environmental engineering applications of ANN modelling date to the late 1980's when models for optimizing construction tasks were developed (Flood and Kartam 1997). Since that time, interest in ANN applications has increased at a steady rate, as indicated by an increase in the number of publications. Specifically, applications have been developed in the areas of hazardous waste management and groundwater exploration (Schmuller 1990), hydrology and water resources engineering (Daniell 1991), groundwater remediation (Garrett, Ranjithan and Eheart 1992), and biological waste treatment (Cote et al. 1994). More recently, ANNs have been used in transient drainage design (Shukla et al. 1996), air quality monitoring (Hasham, Stanley and Kindzierski 1998), and wastewater treatment plant operations (Boger 1997). As the focus of the current study is the ANN modelling of water treatment plant processes, known applications in the water treatment industry are discussed separately.

2.4.7.2 Water Quality and Treatment Applications

To date, while several ANN applications in the areas of water quality monitoring and water treatment have been proposed, the potential for application in these areas remains largely unexploited. With regards to water quality applications, the California Department of Water Resources has used ANN models to predict THM formation and speciation in the Sacramento Delta (Hutton *et. al.*, 1996). In addition, ANN models have been developed to forecast salinity in the Chesapeake Bay in order to determine appropriate pumping schedules for water supply (DeSilets et al. 1992). Finally, research in the Environmental Engineering Program at the University of Alberta has led to the development of successful models for colour in the North Saskatchewan River (Zhang and Stanley 1997).

With respect to water treatment operations, much of the research has focussed on modelling chemical dosing levels in the coagulation process. To this end, models have been developed for alum and polymer dose forecasting at a water treatment plant in New South Wales (Mirsepassi, Cathers and Dharmappa 1995). While good correlations between predicted and actual doses were obtained, the models have little practical use. The model input parameters are exclusively time-series data, resulting in the development of models which essentially follow the slope of historical dosing levels without capturing the unique features of the treatment processes. Operational full-scale models for the removal of turbidity at the two City of Edmonton water treatment facilities have been

developed (Stanley and Zhang 1997) and will be used, in conjunction with the results of the current study to optimize the enhanced coagulation process at each facility.

In addition to the coagulation models, a model has been developed to improve control and to reduce backwash water volume for the rapid gravity filters in London, U.K. (Koutsakos 1995). As well, successful models for the concentration of residual chlorine in distribution systems in Quebec and Ste. Foy have been developed (Rodriguez et al. 1997). Finally, models are currently being developed for predicting breaks in city water mains by members of the Environmental Engineering Program at the University of Alberta.

2.4.8 Process Control Using ANNs

In process control applications, neural networks can be incorporated in internal model control (IMC) in either direct or indirect methods (Psichogios and Ungar 1991). In the direct method, a neural network is trained with observed input-output data from the system in order to represent its inverse dynamics. As such, given the current state of the system, the network can be trained to produce the control action required to reach a desired target state. In the indirect method, the network is trained with input-output data from the dynamic system to represent the forward dynamics. As such, given the current state and the current control action, the network can be trained to predict the next state of the system. According to IMC theory, if a good plant process model exists, the closed-loop IMC system will achieve exact set-point following unmeasured disturbances acting

on the plant (Yan et al. 1996). A direct dual-model IMC system is presented schematically in Figure 2.6. In this system, the control action, u , is the output of an ANN controller trained to represent the plant inverse dynamics (Psichogios and Ungar 1991). The process model, also an ANN model, is used in parallel to the plant in order to provide the feedback error between the actual and predicted value of the plant's output.

Similar ANN control strategies have been proposed for chemical process systems (Bhat and McAvoy 1990), anaerobic digestion (Wilcox et al. 1995), and, as previously discussed, water treatment dosing levels (Koutsakos 1995 ; Zhang and Stanley 1998). While most of these proposed control schemes have yet to be tested on full-scale systems, preliminary results suggest that full-scale integration of ANN-IMC systems will occur in the near future.

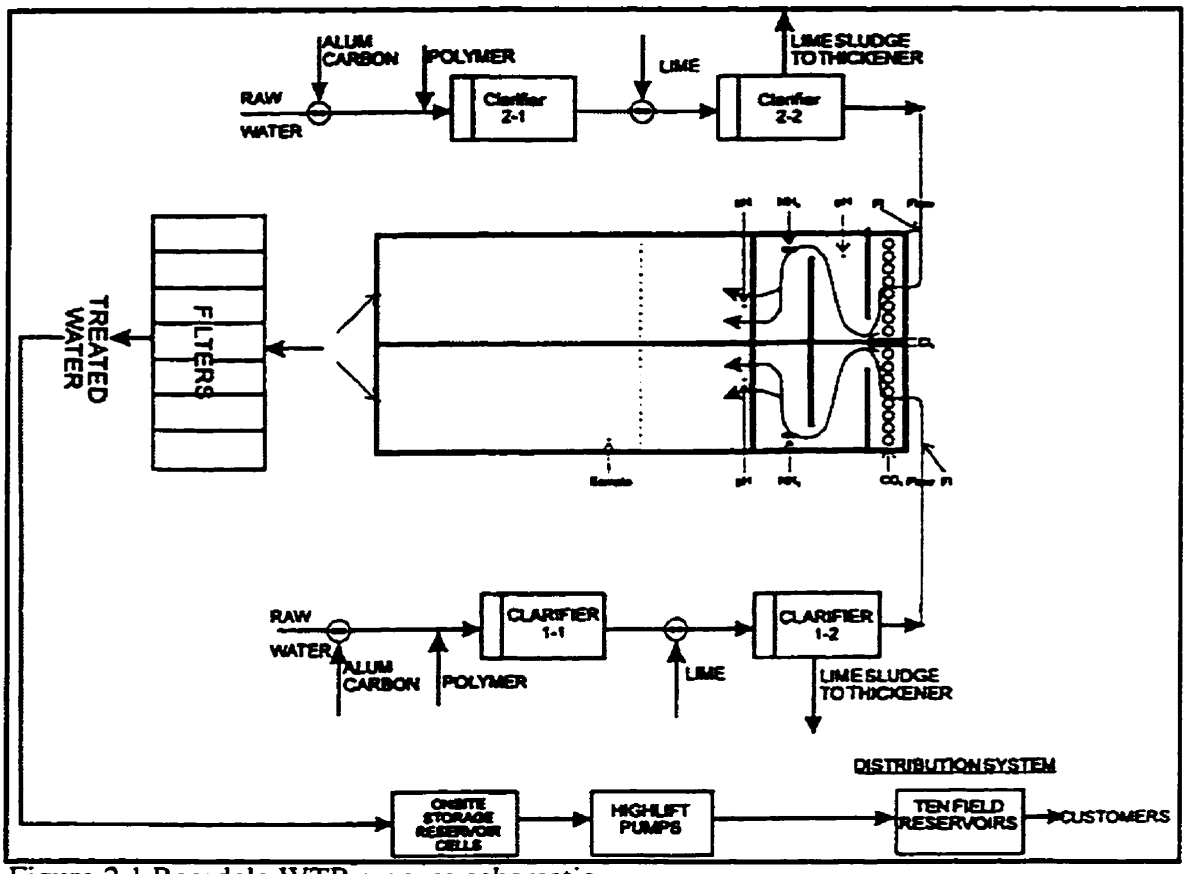


Figure 2.1 Rossdale WTP process schematic

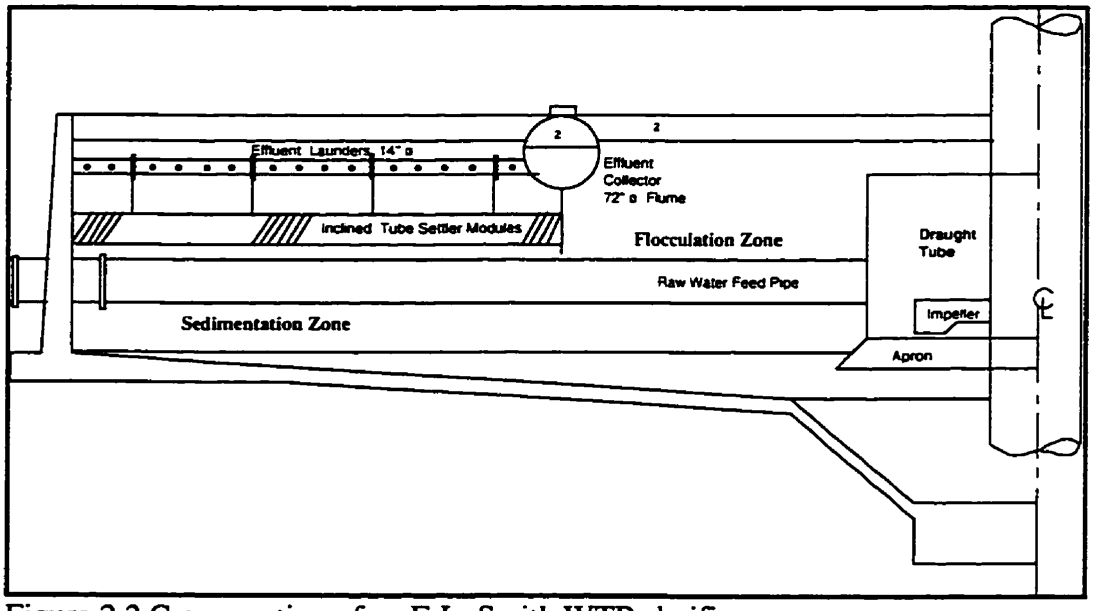


Figure 2.2 Cross-section of an E.L. Smith WTP clarifier

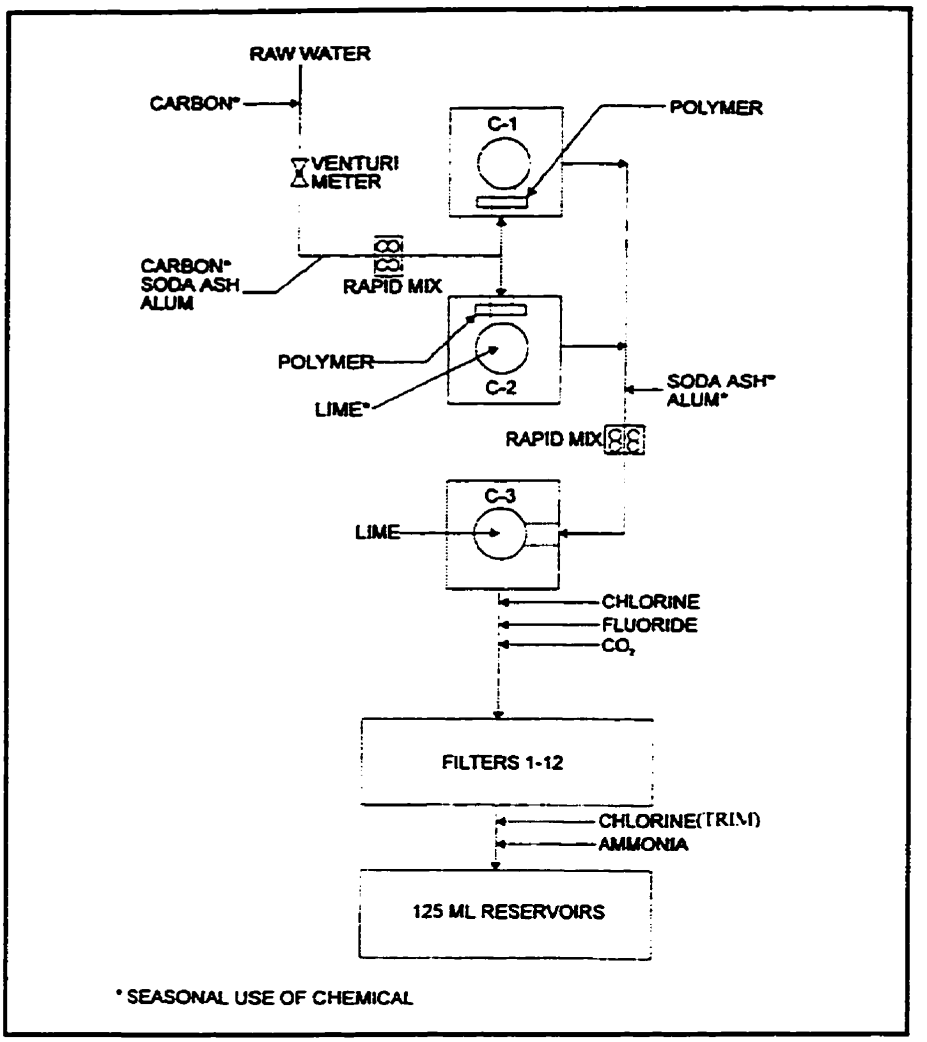


Figure 2.3 E.L. Smith WTP process schematic

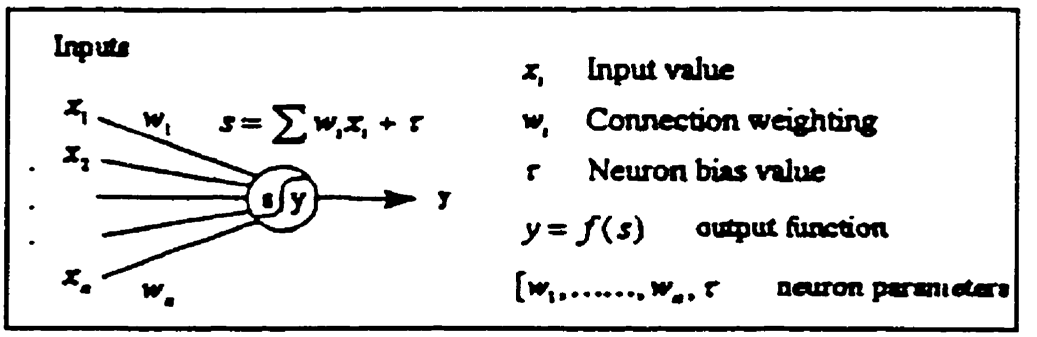


Figure 2.4 Schematic representation of a single neuron (after Tang et al. 1995)

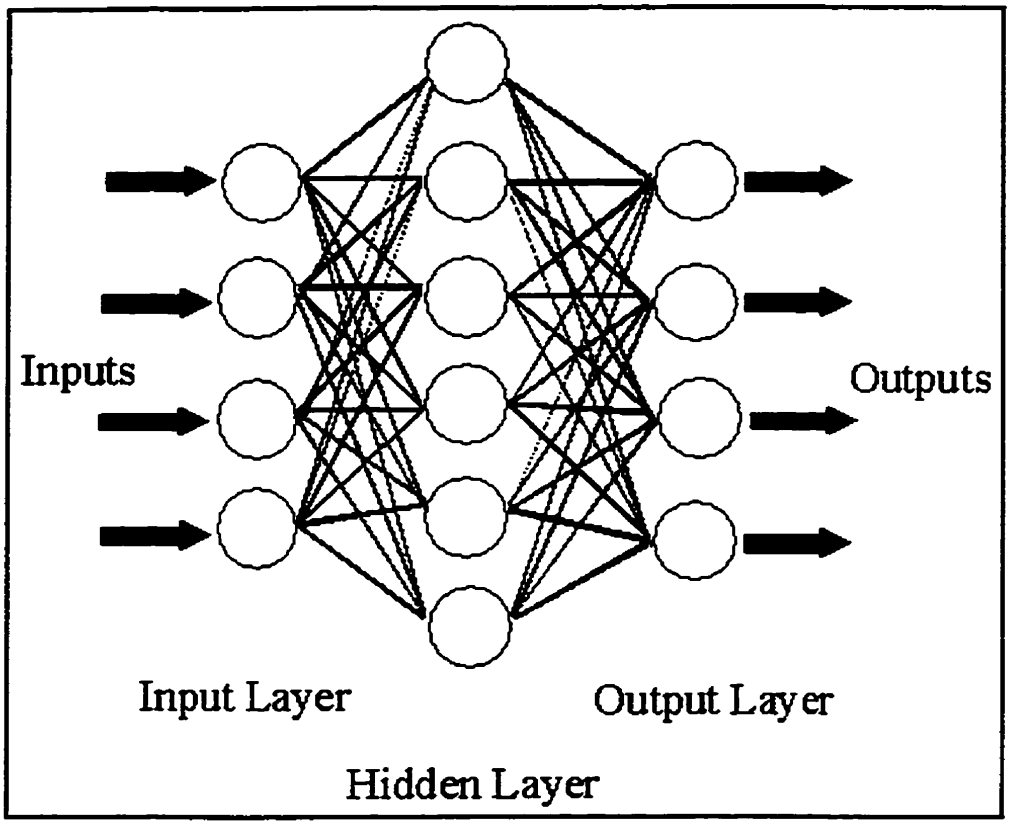


Figure 2.5 Neuron connectivity in ANNs

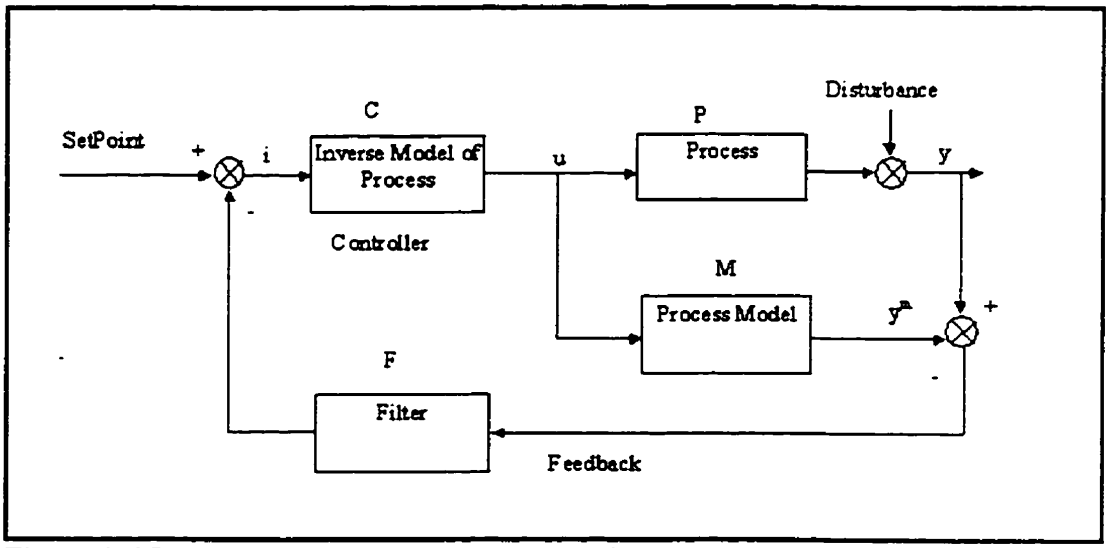


Figure 2.6 Internal model control structure (after Psichogios and Ungar 1991)

3.0 METHODOLOGY

3.1 Data Collection and Management

All of the data used in the development of the ANN models was obtained from AQUALTA, the parent company that owns and operates both Rossdale WTP and E.L. Smith WTP. The data from both facilities is stored on a central supervisory control and data acquisition (SCADA) system and can be accessed at will by plant operators. As will be discussed, three primary types of data were used in model development: raw water quality, process data, and performance data.

3.1.1 Raw Water Quality Data

The raw water quality data used in model development were measured by lab technicians on a 4-hour cycle, and were acquired from the SCADA system as daily average values. The daily high values of some parameters, such as turbidity and colour, were also recorded although they were not used in model development, as will be discussed. A list of the raw water quality parameters, as well as the error associated with their measurement, is presented in Table 3.1.

3.1.2 Process Data

The process parameters, which include chemical dosing as well as plant flow information, are measured in real-time online and stored in the SCADA system. For the

Table 3.1 Error associated with raw water quality data applications (Thomas and Shariff 1998)

Parameter	Error
PH	± 0.2 pH units
Colour (TCU)	± 1 TCU
Turbidity (NTU)	± 10 %
Alkalinity (mg/L)	± 5 mg/L
Temperature (°C)	± 2 %
Hardness (mg/L as $CaCO_3$)	± 5 mg/L

purposes of the current study, daily average values were used. A list of instrumental errors associated with each of the process parameters is presented in Table 3.2.

Table 3.2 Error associated with process data applications (Thomas and Shariff 1998)

Parameter	Error
Plant flow (ML/d)	± 3 %
Alum dose (mg/L)	± 7 %
PAC dose (mg/L)	± 10 %
Polymer dose (mg/L)	± 6 %

3.1.3 Performance Data

The model output parameter, clarifier effluent colour, is measured by laboratory technicians on a 4-hour cycle, and is reported as a daily average value. The error associated with colour determinations is the same as that reported for raw water colour.

3.2 Software

ANN model development was accomplished using NeuroShell 2 Release 3.0 from Ward Systems Group, Inc. of Frederick, Maryland, USA. The software was chosen due to the high degree of control over model development given to the user. In addition, NeuroShell 2 is completely compatible with Microsoft Excel spreadsheets, allowing for seamless importing and exporting of data. Finally, the software has a user-friendly Windows-based interface that allows for the simultaneous construction of multiple models. Much of the information that follows concerning NeuroShell 2 features can be found in detail in the NeuroShell 2 User's Manual (Ward Systems Group Inc. 1996).

3.2.1 Architecture Types

3.2.1.1 Feed-Forward Architectures

The NeuroShell 2 software supports both supervised and unsupervised feed-forward networks. With respect to the former, backpropagation neural networks, probabilistic neural networks (PNNs) and general regression neural networks (GRNNs) are supported. Three types of backpropagation networks, which differ in the connectivity of the neurons, are supported, and all use Rummelhart's backpropagation learning algorithm (Section 2.4.4). In backpropagation networks with standard connections, commonly termed backpropagation networks, every layer is connected to the immediately previous layer. In backpropagation networks with jump connections, or jump networks, every layer is connected to every previous layer. Ward Networks, proprietary backpropagation

architectures developed by the Ward Systems Group Inc., allow hidden neurons to be grouped into slabs so that more than one activation function can be used simultaneously.

PNNs are a type of supervised network known for their ability to train quickly on sparse data sets. These networks separate data into a specified number of output categories, and are therefore not useful for continuous valued outputs.

GRNNs are a type of supervised network that, like backpropagation networks, have the ability to produce continuous valued outputs. GRNNs are three-layer networks and have one hidden neuron for each pattern in the training set. While these networks are particularly good at generalizing cases similar to those on which they were trained, they tend to have poor interpolation capabilities.

With respect to unsupervised feed-forward architectures, the software supports the Kohonen Self Organizing Map (SOM). Kohonen SOMs have the ability to learn without being shown correct outputs in sample patterns and are able to separate data into a number of predetermined categories.

3.2.1.2 Feedback Networks

The NeuroShell 2 software supports three types of feedback or recurrent networks. All three types use the backpropagation algorithm in training. Recurrent networks can be easily distinguished from feed-forward networks due to the presence of an extra slab in

the input layer that is connected to one of the layers via feedback connections. This extra slab holds the contents of one of the layers as it existed when the previous pattern was trained. As such, "memory" is added to the network, making it particularly useful for time-series data.

3.2.2 Scaling Functions

In order to scale data from its numeric range to a range that the networks are able to use more effectively, a scaling function is employed. The NeuroShell 2 software supports both linear and non-linear scaling functions. In either case, data can be scaled into one of two ranges: -1 to 1 or 0 to 1. Both ranges can be either bounded, where values outside the parameter boundaries are scaled to the boundary value, or unbounded, where values outside the boundaries are scaled proportionately. The linear scaling function scales data using a common divisor, while non-linear scaling functions, such as logistic and tan h, tend to compress values near the extreme ends of the data ranges.

3.2.3 Activation Functions

NeuroShell 2 supports 8 separate activation functions that can be used in the hidden layers of network architectures. While functions such as the logistic function have been widely reported in literature others, such as the sine function, are rarely used. Some of the functions map values to the -1 to 1 range, while others map values to the 0 to 1 range. A

list of the activation functions, as well as their respective equations and mapping ranges, supported by the software is presented in Table 3.3.

Table 3.3 Activation functions supported by NeuroShell 2

Activation Function	Equation	Mapping Range
Logistic	$f(x) = 1/(1+e^{-x})$	0 to 1
Linear	$f(x) = x$	-1 to 1 or 0 to 1
Tanh	$f(x) = \tanh(x)$	-1 to 1
Tanh1.5	$f(x) = \tanh(1.5x)$	-1 to 1
Sine	$f(x) = \sin(x)$	-1 to 1
Symmetric Logistic	$f(x) = (2/(1+e^{-x})) - 1$	-1 to 1
Gaussian	$f(x) = e^{-x^2}$	0 to 1
Gaussian Complement	$f(x) = 1-(e^{-x^2})$	0 to1

3.2.4 Weight Updates

The software supports three methods of updating weights: vanilla, momentum, and turboprop. The vanilla method uses a proprietary algorithm to apply only a learning rate term to the weight updates. Weight updates by the momentum method include not only a learning rate term, but also incorporate a portion of the last weight change as the momentum term. The turboprop method is not sensitive to either learning rate or momentum. Training proceeds through an entire epoch before weights are updated. The turboprop method uses an independent weight update size for each different weight as opposed to having a single learning rate or momentum term applied to all weights.

3.2.5 Method of Pattern Selection

During training, the software can select patterns in either a random fashion or in rotational order. When the turboprop method of weight updates or recurrent network architectures are used, only the rotational method of pattern selection is available.

3.3 ANN Model Development and Evaluation

As previously discussed, there is no set protocol for the development of ANN models. What follows is a description of the three-step protocol employed in the current study. As ANN model development is very much a trial and error process, the protocol was continuously updated to reflect the knowledge gained from working with ANN systems.

3.3.1 Source Data Analysis

The primary objectives of the source data analysis were to gain a familiarity with the study domain and to examine the applicability of available data for model development. Initially, the problem domain was thoroughly examined through a review of pertinent literature on enhanced coagulation and DBPs, as blind application of the ANN technique to problems that have not been thoroughly studied will lead to the development of models with poor generalization capabilities. Following the domain study, all the available pertinent data were examined and subjected to comprehensive statistical analyses in order to determine the range, seasonal and daily trends, and other important data characteristics.

3.3.2 ANN Model Development

3.3.2.1 Rossdale WTP, Plant #1

The artificial neural network modelling process used involved two distinct stages: preliminary model development and model optimization. The objective of the preliminary model design stage was to design and evaluate a series of network architectures that, when optimized, could be used as an effective process model. This objective was best met through the use of a four-step scheme which included the selection of input and output parameters, the organization of the data patterns, the selection of preliminary factors and levels of analysis, and the evaluation of potential architectures. With respect to the selection of input and output parameters, the output parameter which best represents the process was selected. Each input parameter was selected based on data availability and the likelihood of there being a cause-effect relationship between it and the output parameter. Once the model parameters were selected, the data patterns were selected to reflect the availability and reliability of the data. Incomplete patterns, as well as those that appeared to be inconsistent with the remaining data were removed.

The data was initially organized into two categories based on the value of the output parameter. The boundary was selected according to process performance criteria and separated regular operating conditions from process upset or special case conditions. In

order to develop a successful model, the data was further divided into three fractions: the training set, the test set, and the production set. The training set consisted of data patterns that the network processes repeatedly in order to learn trends and patterns in the data. During the learning process, the network was periodically evaluated using the test set patterns in order to ensure that the network was not simply memorizing the training data. The trained network was applied to the production set which consists of data that the network had never "seen" before in order to assess the performance of the model. Each of these data sets contained an equal percentage of special case data in order to ensure that the model was trained, tested, and evaluated over a similar range of effluent quality.

In designing the preliminary architectures, many factors needed to be considered including the type of architecture, the number of layers, the number of neurons in each layer, the type of scaling and activation functions, and the learning approach. In order to determine the optimal levels of each of these factors, the factorial experimental design approach was applied. This statistical method is generally used for studying the effects of varying the levels of multiple parameters in a limited number of runs. For an in-depth discussion on the mechanisms of factorial design experimentation, please refer to the text by Box and Hunter (1978). Even with the aid of factorial design experimentation, determining the effects of the multitude of factors that can be altered in network design was beyond the scope of the current study. As such, a number of factors were held constant throughout the preliminary design stage. All input data were scaled using an unbounded linear scaling function in the -1 to 1 range. The linear function was selected in order to ensure that data in the extreme ends of each parameter's range were not

compressed. The unbounded -1 to 1 range allowed for the inclusion of negative values, which were present for some of the parameters, as well as values beyond the preliminary data range that may have surfaced when the model was applied to real-time data. With respect to the activation function used in the hidden layers, the logistic function was employed exclusively in the preliminary model design stage, as it is the default activation function used by the software. In addition, the turboprop method of weight updates was used, as it is insensitive to learning rate and momentum. Finally the range of the output parameter was exclusively positive, necessitating the use of an activation function that maps values into the 0 to 1 range. The logistic activation function, being the function used most often in ANN applications, met this criterion and was always used on the output layer.

In order to assess the model's performance, two separate statistical indicators were applied to the production data set. The coefficient of multiple regression, R^2, compares the accuracy of the model to the accuracy of a trivial benchmark model wherein the prediction is just the mean of all the samples. A perfect fit would result in an R^2 value of 1, a very good fit near 1, and a very poor fit near 0. The equation the NeuroShell 2 uses to calculate R^2 is:

$$R^2 = 1 - \frac{SS_r}{SS_T}$$

where SS_r is the residual sum of squares and SS_T is the total sum of squares of the model output. The R^2 indicator was applied to the entire production data set and therefore serves as a measure of the model's performance in periods of routine operation as well as during

special-case scenarios. The second statistical indicator, the mean absolute error, was used to compare the actual process outputs with the network predictions. This indicator can serve to highlight inconsistencies in model predictions and can also be used to determine whether the model predictions are adequate for process control. In the initial model development stage, the mean absolute error was used to assess the performance of the model during the special-case scenarios only. The ideal network will have a low mean absolute error on the special-case scenarios without sacrificing the goodness of fit, as measured by R^2, for the entire production data set.

In the model optimization stage, the most promising candidate models were optimized through the fine-tuning of the network architectures in order to minimize the error on the production set data. The activation functions, method of weight updates, learning rate, and momentum, all factors that were held constant throughout the preliminary model design stage, were varied. The optimal model will be able to follow daily trends in plant operations in addition to predicting the special case patterns. In addition the model should produce consistent results for all three data sets. The model should also be insensitive to retraining following a swapping of the testing and production sets. Finally, a plot of the model residuals should be free of obvious trends.

3.3.2.2 Rossdale WTP, Plant # 2

Building on experience gained from developing an ANN model for Plant #1, a new protocol was developed for ANN model development and was applied to Plant # 2. In

developing the model for Plant #1, all of the backpropagation networks including those with standard, jump, Ward, and recurrent configurations were evaluated for use in model development. Of these architectures, only the 3-layer and 4-layer backpropagation networks with standard connections yielded acceptable results. In the interest of increasing the efficiency of the ANN protocol, only these two architectures were evaluated for use in the model for Plant #2.

One of the glaring deficiencies in the protocol used for Plant #1 is its heavy reliance on factorial design to determine the optimal level of each factor. The factorial design approach is particularly useful if all factors are continuous and there is some overriding trend with regards to the relationship between the factor and the model output. Unfortunately, this is not the case for many of the factors associated with neural network design. Many of the factors being evaluated in ANN model design are categorical, with more than two possible levels, and can not be ordered to allow for the application of response surface methodology. There are, for example, 8 separate activation functions that can be used in ANN model development. The factorial design approach assumes that there is some definable relationship between each of the possible activation functions. If the logistic and linear activation functions were selected as the initial levels of analysis, for example, the factorial design approach would not be able to tell the experimenter which function to try next, only which of the two functions works best under the given conditions.

Keeping these limitations in mind, the new protocol involved the use of classical one-variable-at-a-time experimental design. The optimal values of each of the factors were determined sequentially, keeping the values of the other factors constant. A factorial design was employed in the final stages in order to determine the appropriate learning rate and momentum, as these factors are not subject to the limitations highlighted above. Overall, the protocol involved the following 9 steps: selection of input and output parameters, selection and organization of data patterns, determination of the appropriate training : testing : production ratio, determination of the appropriate number of neurons for each hidden layer, determination of the appropriate activation function, reassessment of the number of hidden layer neurons, evaluation of model stability, model fine-tuning, and comparison of candidate models. The protocol was applied separately to each of the two architectures being investigated. Unless otherwise discussed the scaling functions, activation functions, and weight updates were kept constant as described in the Plant #1 protocol.

In order to evaluate the model's performance as it was being developed, a composite parameter, termed the performance code or the run effect, was developed. The value of the code is based on the R^2 and mean squared error on the production set, as well as the percentage of production set predictions that have a relative error greater than 30 %. As previously discussed, the best models will have a high R^2 and low mean squared error. The "percent > 30 % error" also serves as a measure of prediction accuracy; lower values are best. As shown in Table 3.4, each of the three parameters is coded over a pre-

determined range. The run effect or performance code is simply the sum of the three coded parameters; higher values indicate better performance.

Table 3.4 Generation of performance codes

R^2		Mean Squared Error		% >30%	
Value	Coded	Value	Coded	Value	Coded
< 0.3	1	>0.5	1	> 30 %	1
0.3 - 0.34	2	0.45 - 0.5	2	25 - 30 %	2
0.35 - 0.39	3	0.4 - 0.44	3	20 - 24 %	3
0.4 - 0.44	4	0.35 - 0.39	4	15 - 19 %	4
0.45 - 0.49	5	0.3 - 0.34	5	10 - 14 %	5
0.5 - 0.54	6	0.25 - 0.29	6	< 10 %	6
0.55 - 0.59	7	0.2 - 0.24	7		
0.6 - 0.64	8	0.15 - 0.19	8		
0.65 - 0.69	9	0.10 - 0.14	9		
0.7 - 0.74	10	<0.10	10		
0.74 - 0.79	11				
0.80 - 0.84	12				
0.85 - 0.89	13				
0.90 -	14				

As will be discussed, the model parameters used for the Plant #2 were identical to those used for Plant #1. For the Plant #2 model, no differentiation was made between special case scenarios and normal operating conditions. Instead, the data was initially sorted according to the value of the output parameter. The separation of the data into the training, testing, and production sets, was done in order, thereby insuring an equal distribution of high and low clarifier effluent colour data in each set. Following these steps, the appropriate training : testing : production ratio was selected by preparing a number of data files with varying ratios and testing the model performance on each. Ratios in the range of 1:2:2 to 4:1:1 were prepared and a series of preliminary runs was

performed on each. The ratio that gave the best performance, based on both R^2 and mean absolute error, was selected.

Keeping both the T : T : P ratio and the activation function constant, the number of hidden layer neurons for each layer was varied. For single layer networks, the number of neurons was increased in five-neuron increments over a pre-selected range. For two-layer networks, a grid was used to determine the appropriate number of neurons in each layer. Again, the number of neurons was increased in five-neuron increments over a pre-selected range. Following this initial assessment, the range of values for the number of hidden layer neurons that yielded the best results was expanded. Within this range, the number of neurons was varied in single-neuron increments.

Holding the T : T : P ratio and the number of hidden layer neurons constant, the best activation function was selected by comparing the results for models trained with each of the eight separate activation functions. For multiple-layer models, a grid system was again employed.

If the activation function selected in previous step was not the logistic activation function, the appropriate number of hidden layer neurons was again determined, as previously described, to ensure that the best possible combination of number of neurons and activation function was obtained.

In order to ensure that the results obtained were not dependent on the manner in which the data were extracted, an alternate data set in the appropriate T : T : P ratio was prepared. The results of the models were compared to ensure that there was no significant difference when different data sets were used.

The model was retrained using the momentum and vanilla weight updates in order to determine if any improvements could be made. Using weight updating methods other than turboprop necessitated the evaluation of the effects of learning rate and momentum, as well as the method of pattern selection. All four factors were evaluated simultaneously using factorial design experimentation.

Finally, the candidate models derived from each of the two base architectures were compared with respect to their performance on the training, testing, production, and cross-validation sets. The candidate with the best performance, as defined by R^2, mean squared error, and mean absolute error, was selected.

3.3.2.3 E.L. Smith WTP

Once again, experience gained in the development of the models for Rossdale WTP allowed for a further pairing-down of the steps involved in developing successive models. First, the ratio of data in the training, testing, and production sets fixed at 3:1:1 Based on the first two models, this ratio was found to be most appropriate. In addition, the determination of the appropriate architecture step was eliminated. Based on the first

two models, the logistic activation function always provided the best results and, as such, was the sole activation function evaluated in the E.L. Smith model. As only one activation function was evaluated, the reassessment of the number of hidden layers was also eliminated from the protocol. The protocol used in the development of the model for E.L. Smith WTP therefore, had 6 distinct steps: selection of input and output parameters, selection and organization of data patterns, determination of the appropriate number of hidden layer neurons, evaluation of model stability, model fine-tuning, and comparison of candidate models. The methodology employed for each of these steps is identical to that used for Rossdale WTP, Plant #2, with the exception of the selection and organization of data patterns. As previously mentioned, two clarifiers are used for coagulation at E.L. Smith WTP. While both have identical dimensions and receive the same raw water, different dosing levels are sometimes employed in each. In order to develop a single model for the WTP, the values of each of the process and performance parameters were averaged between the two clarifiers.

3.3.3 Model Evaluation Using Real-time Data

In order to assess the models' performance on real-time data, the three models were applied retroactively to data supplied by AQUALTA. The time frame of the study was selected to correspond with spring-thaw, typically a difficult time for water treatment plant operations. The frequency of data was selected to reflect the frequency of updates of model parameters in the SCADA system. The performance of the models was assessed using R^2, mean absolute error, and other statistical indicators.

4.0 RESULTS

4.1. Source Data Analysis

As suggested in the methodology section, the first step in developing an effective ANN model is a careful study of the problem domain. With respect to the enhanced coagulation process, the data can be divided into major categories: raw water quality parameters, process parameters, and performance parameters. The raw water quality parameters provide an assessment of the quality of the raw water that must be treated by the WTPs and include, for example, colour, turbidity, alkalinity, and pH. The process parameters are those that can be controlled by the plant operators in order to optimize the enhanced coagulation process and include the doses of alum, PAC, and polymer, as well as the flow through each of the plants. The performance parameters allow the operators to assess the performance of the enhanced coagulation process and include clarifier effluent turbidity and colour.

4.1.1 Rossdale Water Treatment Plant

4.1.1.1 Raw Water Quality Parameters

In order to effectively operate the facility, a number of raw water quality parameters are monitored at Rossdale WTP on either a real-time, or scheduled basis. The availability of data for many of these parameters is presented in Table 4.1. Due to substantial seasonal variations in the North Saskatchewan River flow and ambient air temperature, the river water quality varies considerably. Raw water daily average turbidities range from approximately 2 NTU in winter, when the river is under ice cover, to over 1400 NTU

during spring thaw (Table 4.2). Similarly, raw water colour ranges from approximately 2 TCU to 80 TCU throughout the year. The seasonal nature of these parameters is presented graphically in Figures 4.1 and 4.2. Other parameters, such as total alkalinity and total hardness, show far less variation with mean values of 133.7 ± 9.3 mg/L and 166.9 ± 12.6 mg/L (as $CaCO_3$) respectively.

Table 4.1 Rossdale WTP, daily data availability for raw water quality parameters

Parameters	1990	1991	1992	1993	1994	1995	1996
PH	X	X	X	X	X	X	X
Temperature (°C)	X	X	X	X	X	X	X
River flow (m^3/s)	X	X	X	X	X	X	X
Turbidity, daily low (NTU)	X						
Turbidity, daily high (NTU)	X	X	X	X	X	X	X
Turbidity, daily average (NTU)	X	X	X	X	X	X	X
Colour, daily low (TCU)	X						
Colour, daily high (TCU)	X	X	X	X	X	X	X
Colour, daily average (TCU)	X	X	X	X	X	X	X
Total hardness (mg/L as $CaCO_3$)	X	X	X	X	X	X	X
Total alkalinity (mg/L)	X	X	X	X	X	X	X
VOC (mg/L)	X	X	X	X	X	X	X
Maximum conductivity (µS/cm)					X	X	X
UV-absorbance (200 nm)	X	X					
UV-absorbance (254 nm)			X	X	X		
Raw TOC (mg/L)			X				

Table 4.2 Rossdale WTP, data analysis for raw water quality parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95th	50th	5th
pH	92-96	8.2	0.2	7.8	8.8	1.0	8.5	8.3	7.9
Temperature (°C)	92-96	10.3	6.2	0.5	25.0	24.5	20.7	8.0	1.0
River flow (m^3/s)	92-96	190.4	104.8	18.0	1050.0	1032.0	368.6	159.0	95.0
Turbidity, daily high (NTU)	92-96	49.6	153.4	2.0	2400.0	2398.0	170.0	12.0	3.0
Turbidity, daily ave. (NTU)	92-96	31.7	88.9	1.6	1481.0	1479.4	116.2	8.0	2.4
Colour, daily high (TCU)	92-96	10.2	10.3	2.0	82.0	80.0	32.0	7.0	3.0
Colour, daily ave. (TCU)	92-96	9.0	9.1	2.0	77.0	75.0	6.0	2.0	2.0
Total hardness (mg/L as $CaCO_3$)	92-96	166.9	12.6	104.0	204.0	100.0	188.0	167.0	144.0
Total alkalinity (mg/L)	92-96	133.7	9.3	94.0	174.0	80.0	149.0	134.0	119.0
VOC (mg/L)	94-96	0.01	0.02	0.00	0.44	0.44	0.03	0.00	0.00
Raw TOC (mg/L)	92	2.7	1.3	0.2	7.90	7.70	5.1	2.3	1.2

4.1.1.2 Process Parameters

With respect to the operating conditions at the WTP, a list of the parameters that can be controlled by the operator in order to optimize the enhanced coagulation process is presented in Table 4.3. As discussed, the Rossdale facility is split into two separate treatment trains, Plant #1 and Plant #2. The mean flow through Plant # 1 is approximately 61 ML/d, while that for Plant #2 is 86.9 ML/d (Table 4.4). With respect to the alum dose, the range for both plants is from 9 mg/L under the most favorable water quality conditions, to 164 mg/L for poor quality source water. PAC is used extensively during spring runoff in order to remove taste and odour causing compounds. Doses of up to 146 mg/L have been used, although the dose exceeds 45 mg/L less than 5 % of the time

(Table 4.4). The anionic polymer dose is typically 0.30 mg/L, although higher doses may be added during periods of high alum use.

Table 4.3 Rossdale WTP, daily data availability for process parameters

Parameter	1990	1991	1992	1993	1994	1995	1996
Raw flow, Plant # 1 (ML/d)	X	X	X	X	X	X	X
Raw flow, Plant # 2 (ML/d)	X	X	X	X	X	X	X
Alum dose, Plant # 1 (mg/L)	X	X	X	X	X	X	X
Alum dose, Plant # 2 (mg/L)	X	X	X	X	X	X	X
PAC dose, Plant # 1 (mg/L)	X	X	X	X	X	X	X
PAC dose, Plant # 2 (mg/L)	X	X	X	X	X	X	X
Polymer dose, Plant # 1 (mg/L)			X	X	X	X	X
Polymer dose, Plant # 2 (mg/L)			X	X	X	X	X

Table 4.4 Rossdale WTP, data analysis for process parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95th	50th	5th
Raw flow, P #1 (ML/d)	92-96	61.2	23.0	0.0	125.0	125.0	97.6	64.0	0.0
Raw flow, P #2 (ML/d)	92-96	86.9	27.8	0.0	165.0	165.0	136.6	85.0	40.4
Alum dose, P #1 (mg/L)	92-96	31.8	19.7	9.0	164.0	155.0	71.1	26.0	15.0
Alum dose, P #2 (mg/L)	92-96	32.8	20.6	9.0	164.0	155.0	73.2	27.0	14.0
PAC dose, P #1 (mg/L)	92-96	10.7	18.1	0.0	145.6	145.6	46.0	4.4	0.0
PAC dose, P #2 (mg/L)	92-96	10.2	16.7	0.0	142.1	142.1	44.2	4.9	0.0
Polymer dose, P #1 (mg/L)	92-96	0.29	0.15	0.0	0.87	0.87	0.51	0.30	0.00
Polymer dose, P #2 (mg/L)	92-96	0.28	0.14	0.00	0.86	0.86	0.49	0.30	0.00

4.1.1.3 Performance Parameters

The most common measures of enhanced coagulation treatment performance employed at Rossdale WTP include clarifier effluent turbidity and clarifier effluent colour. The availability of data for these performance parameters is presented in Table 4.5. With respect to clarifier effluent turbidity, mean values of 2.5 NTU and 2.8 NTU are obtained for Plant #1 and Plant #2, respectively (Table 4.6). With respect to clarifier effluent colour, a mean value of 2.1 TCU is shared between the two plants. Neither the effluent turbidity nor the effluent colour show any seasonal variations, as isolated cases of high effluent turbidity and colour occur throughout the year. With respect for the values obtained for the 5^{th} percentile of both clarifier effluent colour and clarifier effluent turbidity, values of 1.0 TCU and 1.0 NTU were obtained respectively, for both plants. Rather than being purely coincidental, this occurrence can be explained by the manner in which the lab technicians recorded values below 1.0 TCU or 1.0 NTU in their log books.

Table 4.5 Rossdale WTP, daily data availability for performance parameters

Parameters	1990	1991	1992	1993	1994	1995	1996
Effluent turbidity, Plant # 1 (NTU)	X	X	X	X	X	X	X
Effluent turbidity, Plant # 2 (NTU)	X	X	X	X	X	X	X
Effluent colour, Plant # 1 (TCU)	X	X	X	X	X	X	X
Effluent colour, Plant # 2 (TCU)	X	X	X	X	X	X	X
Residual TOC (mg/L)			X				

In some cases, the technicians simply entered a value of "<1" for values below 1.0. In order to include this data in the preliminary statistical analysis, a nominal value of 1.0 was assumed. Data for which a value of "<1" was entered were not, however, used in model development.

Table 4.6 Rossdale WTP, data analysis for performance parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95th	50th	5th
Effluent turbidity, P # 1 (NTU)	92-96	2.5	1.3	0.4	11.6	11.2	4.8	2.2	1.0
Effluent turbidity, P # 2 (NTU)	92-96	2.8	1.8	0.3	18.7	18.4	6.1	2.4	1.0
Effluent colour, P # 1 (TCU)	92-96	2.1	0.9	0.5	7.4	6.9	3.8	2.0	1.0
Effluent colour, P # 2 (TCU)	92-96	2.1	0.8	0.4	7.5	7.1	3.5	2.0	1.0

4.1.2 E.L. Smith Water Treatment Plant

The modelling process for E.L. Smith WTP was started almost a full year later than that for Rossdale WTP. As such, the availability and analysis of the data, as listed below, is more current than that for Rossdale.

4.1.2.1 Raw Water Quality Parameters

The availability of data used to measure influent water quality from 1990 - present is presented in Table 4.7. Both the pH and the temperature of the influent show seasonal variations, with low values occurring during the winter months and higher values during the summer months (Figure 4.3). The pH of the influent typically dips to a minimum of 7.7 under ice cover and reaches a high of 8.6 during July and August (Table 4.8).

Table 4.7 E.L. Smith WTP, daily data availability for raw water quality parameters

Parameter	1990	1991	1992	1993	1994	1995	1996	1997	1998
pH	X	X	X	X	X	X	X	X	X
Temperature (°C)	X	X	X	X	X	X	X	X	X
River flow (m^3/s)	X	X	X	X	X	X	X	X	X
Turbidity, daily low (NTU)	X								
Turbidity, daily high (NTU)	X	X	X	X	X	X	X	X	X
Turbidity, daily average (NTU)	X	X	X	X	X	X	X	X	X
Colour, daily low (TCU)	X								
Colour, daily high (TCU)	X	X	X	X	X	X	X	X	X
Colour, daily average (TCU)	X	X	X	X	X	X	X	X	X
Total hardness (mg/L as $CaCO_3$)	X	X	X	X	X	X	X	X	X
Total alkalinity (mg/L)	X	X	X	X	X	X	X	X	X
VOC (mg/L)	X	X	X	X	X	X	X	X	X
Maximum conductivity (µS/cm)					X	X	X	X	X
UV-absorbance (200 nm)	X	X							
UV-absorbance (254 nm)			X	X	X				
Total organic carbon (mg/L)			X						

Similarly, water temperature varies between 0 and 25 °C. As previously mentioned, variations in raw water turbidity and colour are due primarily to seasonal fluctuations in river flow and the contributions of run-off during spring-thaw and summer storm events. Raw water daily average turbidity varies between approximately 1 NTU, under ice cover,

to well over 1300 NTU during spring thaw and summer storm events (Table 4.8). Similarly, daily average colour varies between 2 and 98 TCU. The erratic nature of these two parameters is presented in Figures 4.4 and 4.5 respectively.

Table 4.8 E.L. Smith WTP, data analysis for raw water quality parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95th	50th	5th
pH	94-97	8.2	0.2	7.7	8.6	0.9	8.5	8.2	7.9
River flow (m^3/s)	92-96	190.4	104.8	18.0	1050.0	1032.0	368.6	159.0	95.0
Turbidity, daily high (NTU)	94-97	53.2	178.5	2.0	2000.0	1998.0	191.0	8.0	2.0
Turbidity, daily average (NTU)	94-97	32.3	94.4	1.0	1364.0	1363.0	133.0	6.0	2.0
Temperature, °C	94-97	7.8	7.8	0.0	25.0	25.0	20.2	4.0	0.0
Colour, daily high (TCU)	94-97	11.8	13.2	2.0	109.0	107.0	39.0	6.0	3.0
Colour, daily average (TCU)	94-97	10.3	11.6	1.0	98.0	97.0	34.0	6.0	2.0
Total hardness (mg/L as $CaCO_3$)	94-97	164.9	13.4	102.0	202.0	100.0	189.0	164.0	144.0
Total alkalinity (mg/L)	94-97	132.0	9.0	96.0	162.0	66.0	147.0	132.0	116.0
VOC (mg/L)	94	0.0004	0.0005	0.0000	0.0060	0.0060	0.0010	0.0001	0.0001

4.1.2.2 Process Parameters

At E.L. Smith WTP, plant operators can modify the doses of alum, PAC, and anionic polymer and, to a lesser extent, the flow through the plant. Comprehensive records of these control actions are available, as listed in Table 4.9. The raw flow through the plant is subject to water supply and demand and typically ranges from a maximum of 240 ML/d to a minimum of 60 ML/d, although flows less than 160 ML/d occur less than 5 % of the time (Table 4.10). Alum doses vary concurrently with influent water quality and range from a low of 10 mg/L to a high of 182 mg/L. PAC doses show similar variability, with doses exceeding 150 mg/L during periods of poor quality influent. PAC is typically

used to control taste and odour problems that occur in January as well as during spring thaw. The anionic polymer coagulant-aid, used to assist in turbidity removal, is applied in doses up to 2.1 mg/L, although doses over 0.4 mg/L are rarely used (Table 4.10).

Table 4.9 E.L. Smith WTP, daily data availability for process parameters

Parameter	1990	1991	1992	1993	1994	1995	1996	1997	1998
Raw flow (ML/d)	X	X	X	X	X	X	X	X	X
Alum dose (mg/L)	X	X	X	X	X	X	X	X	X
PAC dose (mg/L)	X	X	X	X	X	X	X	X	X
Polymer dose (mg/L)			X	X	X	X	X	X	X

Table 4.10 E.L. Smith WTP, data analysis for process parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95th	50th	5th
Raw Flow (ML/d)	95-97	184.4	21.1	53.7	244.1	190.4	220.9	180.8	159.8
Alum dose (mg/L)	94-97	40.6	26.6	10.0	182.0	172.0	95.0	30.0	18.0
PAC dose (mg/L)	94-97	8.0	21.1	0.0	151.0	151.0	57.1	0.0	0.0
Polymer dose (mg/L)	94-97	0.2	0.1	0.0	2.1	2.1	0.4	0.3	0.1

4.1.2.3 Performance Parameters

Historically, clarifier effluent turbidity and colour have been used as a measure of the performance of the coagulation process (Table 4.11). Other methods, including streaming current detection and online TOC, have been used sparingly and very little data exists. Clarifier effluent turbidity rarely exceeds 3 NTU, although spikes as high as 8.9 NTU

have been observed (Table 4.12). Similarly, clarifier effluent colour exceeds 4.7 TCU less than 5 % of the time.

Table 4.11 Rossdale WTP, daily data availability for performance parameters

Parameter	1990	1991	1992	1993	1994	1995	1996	1997	1998
Effluent turbidity (NTU)	X	X	X	X	X	X	X	X	X
Effluent colour (TCU)	X	X	X	X	X	X	X	X	X
Residual TOC (mg/L)			X						

Table 4.12 E.L. Smith WTP, data analysis for performance parameters

Parameter	Date	Mean	Std. Dev.	Min.	Max.	Range	Percentile		
							95^{th}	50^{th}	5^{th}
Clarifier effluent turbidity (NTU)	95-96	1.2	0.8	0.0	8.9	8.9	2.7	1.1	0.3
Clarifier effluent colour (TCU)	95-97	2.3	1.2	0.1	6.9	6.8	4.7	2.1	0.5

4.2 ANN Model Development

4.2.1 Rossdale WTP, Plant #1

As previously mentioned, the model for Plant #1 at Rossdale WTP was developed first, using a five-step scheme which included the selection of input and output parameters, the organization of the data patterns, the selection of initial factors and levels of analysis, the evaluation of potential architectures, and model optimization.

4.2.1.1 Selection of Input and Output Parameters

In order to select appropriate input and output parameters for the model, the study domain was thoroughly examined. From recent literature in the areas of coagulation and

enhanced coagulation, as well as from plant operating records, a number of potential input parameters and output parameters were identified (Tables 4.13 and 4.14).

As previously mentioned, the ANN technique is data intensive. Only parameters for which large quantities of historical data exist are suitable for ANN model development. Of the potential output parameters suggested in the literature and listed in Table 4.14, only clarifier effluent colour has been continuously monitored at the Rossdale WTP. Data for the other potential outputs is either sketchy or non-existent. With respect to the model input parameters, historical data exists for colour, hardness, alkalinity, pH, turbidity, temperature, river flow, plant flow, alum dose, PAC dose, polymer dose, maximum

Table 4.13 Potential model input parameters

Parameter	Classification
TOC	Raw water quality parameter
DOC	Raw water quality parameter
UVA (254 nm)	Raw water quality parameter
Colour	Raw water quality parameter
THMFP	Raw water quality parameter
SUVA	Raw water quality parameter
Particle counts	Raw water quality parameter
Hardness	Raw water quality parameter
Alkalinity	Raw water quality parameter
pH	Raw water quality parameter
Turbidity	Raw water quality parameter
Water temperature	Raw water quality parameter
Conductivity	Raw water quality parameter
VOC	Raw water quality parameter
Electrophoretic mobility	Raw water quality parameter
River flow	Raw water quality parameter
Plant flow	Process parameter
Coagulant dose	Process parameter
Polymer dose	Process parameter
PAC dose	Process parameter
Mixing Gt	Process parameter

Table 4.14 Potential model output parameters

Parameter	Classification
Clarifier effluent TOC	Performance parameter
Clarifier effluent DOC	Performance parameter
Clarifier effluent Colour	Performance parameter
Clarifier effluent UVA (254 nm)	Performance parameter
Clarifier effluent THMFP	Performance parameter
Clarifier effluent SUVA	Performance parameter

conductivity, and VOC concentration. Of these, VOC concentration and maximum conductivity were eliminated due to the questionable accuracy of the recorded data. In addition, while the two parameters can be used to quantify a portion of the organics and inorganics in the raw water, respectively, there is no suggested direct link in the literature between either parameter and the effectiveness of the enhanced coagulation process. River flow was also eliminated as changes in river flow due to summer storms, ice cover, and spring-thaw are reflected in changes to the values of other parameters such as colour, turbidity, and water temperature.

Following the elimination process, 10 input parameters remained: colour, hardness, pH, alkalinity, turbidity, water temperature, plant flow, alum dose, PAC dose, and polymer dose. While most of the historical data for these parameters has been recorded as daily average values, the daily high values of both raw water turbidity and colour have also been recorded. Daily average values of these parameters were selected over the daily high values in order to ensure continuity with other input parameters. With respect to the plant flow parameter, most WTPs prefer to use the overflow rate as a measure of the flow through the clarifiers. The overflow rate, defined as the plant flow (m^3/d) / surface area of the clarifiers (m^2), was used in developing the model instead of the plant flow in order to allow for easier comparison between WTPs.

With regards to the 6 water quality input parameters, the value reported in any given day might be strongly correlated to the previous day's value. This time-series information can be used by the model to establish an operational baseline; an increase in turbidity of 500 NTU in 24 hours, for example, is an indication of recent summer storm activity. Autocorrelation coefficients are useful for measuring such correlations over a series of previous days. For the purposes of this study, a lag is defined as the difference between today's value and a previous day's value, with lag x being the difference between today's value and the value x days ago. The autocorrelation coefficients over 6 lags for each of the 6 parameters are presented in Table 4.15.

Table 4.15 Autocorrelation coefficients for water quality parameters

Parameter	Autocorrelation Coefficient					
	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
Influent turbidity	0.73	0.44	0.31	0.28	0.28	0.29
Influent colour	0.95	0.85	0.73	0.63	0.54	0.48
Influent alkalinity	0.93	0.90	0.88	0.86	0.85	0.83
Influent hardness	0.93	0.91	0.89	0.87	0.85	0.83
Influent temperature	0.98	0.96	0.95	0.94	0.93	0.92
Influent pH	0.93	0.91	0.90	0.88	0.86	0.84

The results of the autocorrelation study suggest that alkalinity, hardness, temperature, and pH are strongly autocorrelated until at least the 6^{th} lag. When examining the autocorrelation data however, it is important to determine whether the autocorrelations are meaningful or not. In the case of these parameters, the high degree of autocorrelation can be explained by the fact that while these parameters show seasonal variation, they do not vary to a great extent on either a daily or even a weekly basis. Including lag values of these parameters in the models would not be beneficial, as the complexity of the model would be increased without a corresponding increase in the available information provided to the network. Both turbidity and colour are strongly autocorrelated to the 1^{st}

lag. The large drop in the autocorrelation function for turbidity between lag 1 and lag 2 suggest that autocorrelations beyond the 1st lag are weak and should not be incorporated into the model. The more gradual difference between successive autocorrelation coefficients for colour suggests that the inclusion of time-series data for the first three lags may be beneficial. In the interest of keeping the model as simple as possible, however, only the 1st lag was used. As a result, two time-series parameters were added to the list of model inputs. The complete list of model input and output parameters is presented in Table 4.16.

Table 4.16 Rossdale WTP, Plant # 1 model input and output parameters

Parameter	Category	Classification
Influent pH	Input	Raw water quality parameter
Influent turbidity (NTU)	Input	Raw water quality parameter
Influent water temperature (°C)	Input	Raw water quality parameter
Influent colour (TCU)	Input	Raw water quality parameter
Influent hardness (mg/L as $CaCO_3$)	Input	Raw water quality parameter
Influent alkalinity (mg/L)	Input	Raw water quality parameter
Alum dose (mg/L)	Input	Process parameter
PAC dose (mg/L)	Input	Process parameter
Polymer dose (mg/L)	Input	Process parameter
Overflow rate (m^3/d)	Input	Process parameter
Lag 1 influent turbidity (NTU)	Input	Time-series parameter
Lag 1 influent colour (TCU)	Input	Time-series parameter
Clarifier effluent colour (TCU)	Output	Performance parameter

4.2.1.2 Selection and Organization of Data Patterns

In order to develop a successful model, an appropriate time frame for data acquisition must be selected. Three years of data, 1994-1996, were selected as the data records for this period are complete, and no major operational changes have taken place at Rossdale WTP since 1994. In order to avoid introducing uncertainties into the modelling process, incomplete or questionable data patterns were removed. The resulting data file consisted

Table 4.17 Statistical analysis of model input and output parameters for Rossdale WTP, Plant #1

Parameter	Mean	Std. Dev.	Max.	Min.	Percentile		
					5^{th}	50^{th}	95^{th}
Influent pH	8.2	0.2	8.6	7.9	8.0	8.2	8.5
Influent turbidity (NTU)	41.8	117.8	1481.0	1.6	2.5	9.0	170.6
Influent water temperature (°C)	9.7	6.1	25.0	0.5	1.0	7.0	20.0
Influent colour (TCU)	10.5	10.3	77.0	2.0	3.0	7.0	32.6
Influent hardness (mg/L as $CaCO_3$)	167.2	14.7	204.0	104.0	141.4	168.0	190.0
Influent alkalinity (mg/L)	135.1	10.1	169.0	94.0	116.0	136.0	150.6
Alum dose (mg/L)	34.3	21.9	164.0	9.0	15.0	29.0	78.0
PAC dose (mg/L)	9.9	19.2	145.6	0.0	0.0	0.0	47.5
Polymer dose (mg/L)	0.3	0.1	0.9	0.0	0.0	0.3	0.6
Overflow rate (m^3/d)	57.3	11.6	101.5	10.6	44.6	52.8	81.2
Lag 1 influent turbidity (NTU)	0.3	85.1	953.0	-969.0	-39.0	0.0	31.0
Lag 1 influent colour (TCU)	0.0	3.3	27.0	-18.0	-4.0	0.0	4.0
Clarifier effluent colour	2.2	0.8	5.1	0.4	1.0	2.1	3.5

of 889 separate days or patterns, spanning three years of water treatment at the Rossdale Water Treatment Plant. A statistical analysis of the data used in model development is presented in Table 4.17. The data was initially organized into two categories based on the.value of the output parameter. The boundary separating the data corresponds to the 90^{th} percentile of the clarifier effluent color and has a numerical value of 3.20 TCU. Data that exceeds this boundary falls into the special-case scenario category, while the remaining data corresponds to normal operating conditions at the WTP. The data file was then sorted into training, testing, and production sets according to the protocol previously discussed. The initial separation of data according to the value of the output parameter ensured that each of the three data sets contained an equal quantity of special-case scenario data

4.2.1.3 Selection of Factors and Initial Levels of Analysis

As previously mentioned, model development and optimization for Rossdale WTP Plant # 1 was accomplished using factorial design experimentation. This statistical technique allows for the evaluation of the effects of multiple factors on a single or multiple outputs while minimizing the number of experimental trials. For the preliminary model development stage, the effects of some of the most significant factors were evaluated. A sample list of factors, as well as their corresponding initial levels of analysis is presented in Table 4.18.

Table 4.18 Factors and initial levels of analysis used in preliminary model development

Factor	+	-
Ratio of training to testing data	1:1	2:1
Total number of hidden layer neurons	30	120
Activation function	Logistic	Gaussian Complement

For multiple hidden layer architectures, the ratio of neurons in each of the hidden layers was also evaluated. In addition, when Ward networks containing multiple hidden layer slabs were used, different activation functions were used for each slab. The initial levels of analysis were selected based on previous experience in ANN modelling.

4.2.1.4 Evaluation of Potential Architectures

Two separate statistical measures were used to assess the performance of the preliminary models. The R^2 indicator was applied to the entire production data set and therefore serves as a measure of the model's performance in periods of routine operation as well as during special-case scenarios. The second statistical indicator, the mean absolute error, was used to assess the performance of the model during the special-case scenarios only.

Initially, ten separate architectures were evaluated using the factors and initial levels of analysis previously outlined. For each of these architectures, the best values obtained for each of the statistical indicators is listed in Table 4.19.

Table 4.19 Results of preliminary runs for Rossdale WTP, Plant #1

Architecture	R^2	Mean Absolute Error (TCU)
3 Layer Backpropagation	0.5819	0.7860
4 Layer Backpropagation	0.5703	0.8127
5 Layer Backpropagation	0.5396	0.8003
2 Hidden Slab Ward	0.6129	0.7398
3 Hidden Slab Ward	0.6096	0.7581
2 Hidden Slab Ward + Jump Connection	0.5906	0.7902
Recurrent	0	NA
3 Layer Jump Connection	0.6107	0.7822
4 Layer Jump Connection	0.5593	0.8059
5 Layer Jump Connection	0.5798	0.8042

In these preliminary trials, the recurrent network performed extremely poorly since the separation of the data into multiple sets removes the time-series correlations between adjoining data patterns. The R^2 value of 0 obtained for the recurrent network suggests that the neural model predictions were worse than those that could have been obtained by using the mean of the model outputs as the network predictions (Ward Systems Group Inc. 1996). The remaining networks produced satisfactory results and were subjected to further optimization.

In the second set of trials, a second factorial design incorporating the most favorable levels of analysis derived from the results of the first factorial design was applied to each of the architectures. In general, both the R squared and average absolute error values improved by approximately 0.01 to 0.10. Based on the average absolute error results from the second set of trials, the best network is the 3 layer backpropagation network with 96 hidden layer neurons, a logistic activation function, and a 3:1 training data to testing data ratio. This network had an absolute average error for the special-case scenarios in the production set of 0.7074 and a production set R squared of 0.6621. A plot of the actual

data against the network predictions for this network is presented in Figure 4.6. This preliminary model follows daily operational trend well, although there are many inconsistencies in the model predictions for the special-case scenarios.

4.2.1.5 Model Optimization

From the preliminary model design stage, a number of potential candidate model architectures were selected for further optimization. Of these, the three-layer backpropagation architecture produced the most favorable results. When the trained network was applied to each of the data sets, the results were consistent, ranging from an R^2 of 0.71 for the testing set to 0.76 for the training set (Table 4.20). Similarly, the mean absolute error for all data ranged from 0.30 TCU for the training set to 0.32 TCU for the testing set. In addition, when the testing and production data sets were swapped and the model was retrained and applied to the new production set, the results are identical to those for the original test set. This suggests that the internal network structure is identical for both the original and swapped data, since the original test set contains the same data patterns as the new production set. As such, the model architecture is decidedly stable, a requirement for use in process control.

Table 4.20 Model results for Rossdale WTP, Plant # 1

Data Set	R^2	Mean Absolute Error (TCU)
Training	0.76	0.30
Testing	0.71	0.32
Production	0.75	0.31
Cross-validation	0.71	0.32

The model results for previously unseen data (production data) are presented graphically in Figure 4.7. The model follows the trends in the actual clarifier effluent data quite well, although two areas of apparent inaccuracy require a further examination. In the first 15 patterns, when the actual clarifier effluent colour ranges from approximately 1 to 2 TCU, the network tends to over-predict the actual values. From a process control standpoint however, this error is negligible since these patterns correspond to late-winter days where the raw water quality conditions are ideal and process control modifications are rarely required. With respect to the second area of concern, the model has some difficulty in predicting the value of the highest clarifier effluent colour peaks present during process upset conditions. For these cases the model clearly recognizes the existence of the peaks, however, it tends to under-predict the actual effluent colour by approximately 1 TCU. The model could be retrained to reduce the prediction error on these peaks, at the expense of reducing the predictive capacity during normal operating conditions. The ultimate goal of this study is to develop models that can eventually be used in process control applications to reduce the frequency and severity of plant upsets. As such, while it is important that the model recognizes plan upsets, it is far more important that the model better predictive capacity in the normal operating range (< 3 TCU) of clarifier effluent colour than during periods of upset.

In order to ensure that there are no obvious trends in the model residuals, a plot of the residuals against the model predictions for the production set is presented in Figure 4.8. The majority of the model residuals fall within a narrow band in the range of –0.5 TCU to

0.5 TCU. The clarifier effluent colour measurements are performed on an instrument that is only accurate to within 1 TCU. As such, all but a few of the residuals are smaller than the instrumental error. With respect to the distribution of the model residuals, a histogram delineating the frequency of error on the absolute scale is presented in Figure 4.9. From this figure, it is possible to see that approximately 95 % of the model residuals are within 0.8 TCU of the actual recorded values.

4.2.2 Rossdale WTP, Plant # 2

Building on knowledge gained through the development of the model for Plant # 1, the model for Plant # 2 was developed using a protocol that involved the following 9 steps: selection of input and output parameters, selection and organization of data patterns, determination of the appropriate training : testing : production ratio, determination of the appropriate number of neurons for each hidden layer, determination of the appropriate activation function, reassessment of the number of hidden layer neurons, evaluation of model stability, model fine-tuning, and comparison of candidate models.

4.2.2.1 Selection of Input and Output Parameters

The model input and output parameters used in the development of the model for Plant # 2 are identical to those used in the development of the model for Plant #1 (Table 4.16). Since the two plants share the same source water and have the same quality and quantity of historical data, the selection process used for Plant #1 is assumed to be valid for Plant #2.

4.2.2.2 Selection and Organization of Data Patterns

Using the same reasoning applied to the Plant #1 model, data from 1994 - 1996 was used in model development. Incomplete or ambiguous data patterns were removed, resulting in a final data file of 961 complete data patterns or days. The statistical analysis for each of the parameters used in the development of the model for Plant #2 is presented in Table 4.21. The data was sorted according to the value of the output parameter and separated into the training, testing, and production sets as discussed in the methodology section.

4.2.2.3 Determination of the Appropriate Training : Testing : Production Ratio

In developing the model for Plant # 1, a 3:1:1 ratio was found to be the most favorable. In order to confirm this result, a series of runs was performed for a variety of data ratios. For each series, a 3-layer backpropagation network was used, and the number of hidden layers was varied across a predetermined range. Runs were evaluated on the basis of production set R^2 and mean absolute error; the results are presented graphically in Figures 4.10 and 4.11, respectively. From the two figures, the best results are concentrated around the 3:1:1 ratio. As such, this ratio was selected for use in the development of the model for Plant #2.

Table 4.21 Statistical analysis of model input and output parameters for Rossdale WTP, Plant #2

Parameter	Mean	Std. Dev.	Min.	Max.	Percentile		
					5th	50th	95th
Influent pH	8.2	0.2	7.9	8.8	8.0	8.2	8.5
Influent turbidity (NTU)	42.1	116.6	1.7	1481.0	2.6	9.0	174.4
Influent water temperature (°C)	10.7	6.3	0.5	25.0	2.0	8.0	21.0
Influent colour (TCU)	10.8	9.7	2.0	77.0	3.0	7.0	33.0
Influent hardness (mg/L as $CaCO_3$)	167.1	13.6	114.0	204.0	144.0	167.0	190.0
Influent alkalinity (mg/L)	135.6	9.2	94.0	164.0	120.0	136.0	150.0
Alum dose (mg/L)	35.6	22.4	9.0	164.0	14.0	29.0	80.0
PAC dose (mg/L)	8.6	15.6	0.0	120.7	0.0	0.0	39.0
Polymer dose (mg/L)	0.3	0.1	0.0	0.8	0.0	0.3	0.5
Overflow rate (m^3/d)	41.1	9.8	13.0	71.4	28.2	38.9	60.6
Lag 1 influent turbidity (NTU)	-0.6	85.8	-969.0	953.0	-41.8	0.0	28.0
Lag 1 influent colour (TCU)	-0.1	3.2	-18.0	27.0	-4.0	0.0	3.0
Clarifier effluent colour	2.3	0.8	0.4	7.5	1.0	2.3	3.6

4.2.2.4 Determination of the Appropriate Number of Hidden Layer Neurons

For the 3-layer backpropagation network with a single hidden layer, the number of neurons was initially varied in 5 neuron increments over the range of 5 to 100 neurons. For the 4-layer backpropagation network with two hidden layers, a grid system was employed with the number of neurons in each hidden layer being varied in the range of 5 to 50 in 5 neuron increments. The initial network response, with respect to performance

codes, to changes in the number of hidden neurons in the two backpropagation networks is presented in Figures 4.12 and 4.13.

Areas of highest network response were further investigated by performing additional runs. For the 4-layer network, for example, additional runs were performed at 2 neuron intervals in the range of 25 to 35 neurons for the first hidden layer and 20 to 30 neurons in the second hidden layer. The process was continued until an optimal number of neurons was determined. For the 3-layer network, the optimal levels of performance were obtained for both 4 and 62 hidden layer neurons. For the 4-layer network, the optimal combination was 30 neurons in the first hidden layer and 27 neurons in the second hidden layer.

4.2.2.5 Determination of the Appropriate Activation Function

Keeping the number of hidden layer neurons constant at the optimal values, a series of 8 runs, one for each of the available activation functions, was performed for each of the 3-layer backpropagation networks. For the 4-layer backpropagation network, a grid system was again used. In the interest of saving time, the linear activation function was not evaluated for this network as it typically produces poor results. The results of the trials are presented in Tables 4.22 and 4.23.

Table 4.22 Determination of the appropriate activation function for the 3-layer backpropagation networks

Activation Function	Run Effect	
	4 Hidden Layer Neurons	62 Hidden Layer Neurons
Logistic	18	18
Tan h	13	17
Tan 1.5	12	17
Sine	4	16
Symmetric Logistic	15	16
Gaussian	14	14
Gaussian Complement	14	14
Linear	7	15

Table 4.23 Run effect for variations in activation function for the 4-layer backpropagation network

		Activation Function for Hidden Layer 1						
		Logistic	Tan h	Tan 1.5	Sine	Sym. Log.	Gaus.	G.C.
Activation Function for Hidden Layer 2	Logistic	20	20	16	16	18	18	17
	Tan h	17	18	12	16	14	8	14
	Tan 1.5	14	17	15	16	14	13	14
	Sine	14	13	14	10	17	16	17
	Sym. Log.	12	16	14	17	12	11	16
	Gaus.	14	12	14	15	11	12	14
	G.C.	18	16	12	12	18	16	14

In all cases, the best results were obtained using the logistic function. While the logistic / tan h combination also yielded favorable results for the 4-layer network, the effort required to reassess the number of hidden layer neurons using this combination of activation functions outweighed the potential benefits. As such, the logistic activation function was selected for all the hidden layers in each of the architectures.

4.2.2.6 Reassessment of the Number of Hidden Layer Neurons

Since the activation function selected did not differ from that used in the determination of the number of hidden layer neurons, this step was not necessary.

4.2.2.7 Evaluation of Model Stability

In order to ensure that the results obtained were not dependent on the manner in which the data were extracted, an alternate data set was prepared. The best models for each of the architectures were retrained on the alternate data set, and the results in terms of R^2 and mean squared error (MSE) on the production set data are presented in Table 4.24.

Table 4.24 Evaluation of model stability

	Original Data Set		Alternate Data Set	
Architecture	R^2	MSE	R^2	MSE
4-layer network	0.66	0.23	0.59	0.29
3-layer network (4 hidden neurons)	0.65	0.23	0.46	0.38
3-layer network (62 hidden neurons)	0.61	0.26	0.57	0.30

From these results, both the 4-layer network and the 3-layer network with 62 neurons appear to be resistant to the effects of changing the data sets, as only minimal changes in performance were observed. For the 3-layer network with 4 neurons however, there was a marked decrease in model performance when the alternate data set was used. As such, this architecture was eliminated from further consideration.

4.2.2.8 Model Fine-tuning

The final stage of ANN model development employed for Rossdale WTP Plant #2 involved the evaluation of the effects of the method of weight updates, values of learning rate and momentum, and method of pattern selection on model performance using factorial design experimentation. With respect to the 4-layer architecture, the best model was obtained using momentum weight updates, a learning rate of 0.2, a momentum of 0.8, and random pattern selection. For the 3-layer architecture, momentum weight updates with a learning rate of 0.1, a momentum of 0.9, and random pattern selection yielded the best results.

4.2.2.9 Comparison of Candidate Models

In order to determine which of the two models has the best performance, each was applied to the training, testing, production, and cross-validation sets as previously described. The results for the 3-layer model can be found in Table 4.25, while those for the 4-layer model are listed in Table 4.26.

Table 4.25 Results for the 3-layer backpropagation architecture

Data Set	R^2	Mean Squared Error	Mean Absolute Error (TCU)
Training	0.63	0.26	0.38
Testing	0.59	0.26	0.39
Production	0.68	0.19	0.35
Cross-validation	0.51	0.32	0.42

Table 4.26 Results for the 4-layer backpropagation architecture

Data Set	R^2	Mean Squared Error	Mean Absolute Error (TCU)
Training	0.63	0.26	0.38
Testing	0.59	0.26	0.38
Production	0.67	0.21	0.36
Cross-validation	0.57	0.28	0.38

Both of the two candidate models offer similar levels of performance on the production set data. When the testing and production sets were swapped and the models retrained however, the 3-layer model showed a marked decrease in performance. As such, the 4-layer backpropagation model was selected as the final model for Rossdale WTP, Plant #2.

The model results for the production data set are presented graphically in Figure 4.14. While the overall performance of the model, as measured by the statistical indicators in Table 4.24, is not as good as that for Plant #1, a visual inspection of the figure suggests that the model still follows operational trends quite well. The operators at Rossdale WTP acknowledge that Plant #2 is harder to optimize than Plant #1, as reflected by the differences in network performance. Recognizing the optimization difficulties with Plant #2, the operators tend to push Plant #1 to the limits. While Plant #2 has the larger capacity and the higher mean flow (Table 4.4), the overflow rate through Plant #1 (Table 4.17) is substantially higher than that for Plant #2 (Table 4.21).

In order to ensure that there are no obvious trends in the model residuals, a plot of the residuals across all of the patterns in the production set is presented in Figure 4.15. As

with the model for Plant #1, the majority of the model residuals fall within a narrow band in the range of –0.5 TCU to 0.5, although there is slightly more scatter for the Plant # 2 model. With respect to the distribution of the model residuals, a histogram delineating the frequency of error on the absolute scale is presented in Figure 4.16. From this figure, it is possible to see that approximately 95 % of the model residuals are within 0.8 TCU of the actual recorded values.

4.2.3 E.L. Smith WTP

Following the development of a model for Rossdale WTP, Plant # 2, a number of steps were removed from the model development protocol in order to increase its efficiency. First, the determination of the appropriate training : testing : production ratio step was eliminated. Based on the first two models, a 3:1:1 ratio was found to be most appropriate. In addition, the determination of the appropriate activation function step was eliminated. Based on the first two models, the logistic activation function always provided the best results and, as such, was the sole activation function evaluated in the E.L. Smith model. As only one activation function was evaluated, the reassessment of the number of hidden layers was also eliminated from the protocol. The protocol used in the development of the model for E.L. Smith WTP therefore, had 6 distinct steps: selection of input and output parameters, selection and organization of data patterns, determination of the appropriate number of hidden layer neurons, evaluation of model stability, model fine-tuning, and comparison of candidate models.

4.2.3.1 Selection of Input and Output Parameters

The model input and output parameters used in the development of the model for E.L. Smith WTP are identical to those used in the development of the model for Rossdale WTP Plant #1 and Plant #2 (Table 4.16). Both the Rossdale and E.L. Smith facilities share the same source water, although Rossdale WTP is approximately 15 km downstream of E.L. Smith. As such, the raw water quality characteristics at the two facilities are similar and can be described by the same raw water quality parameters. Operationally, while the two facilities use different types and configurations of clarifiers, the process parameters used to define the coagulation process at Rossdale WTP are equally applicable to E.L. Smith.

4.2.3.2 Selection and Organization of Data Patterns

As the E.L. Smith WTP model was the last to be developed, more current data were used. Almost three years of water treatment data, from May 1995 to April 1998, were used in model development. Data prior to May 1995 could not be employed as new tube settlers were installed, and a new spectrophotometer for colour measurement was purchased, at that time. Questionable or incomplete data patterns were removed, resulting in a final data file of 916 separate patterns or days. The statistical analysis of the parameters used in the development of the model is presented in Table 4.27. The data was extracted into the training, testing, and production sets as described in the methodology section.

Table 4.27 Statistical analysis of model input and output parameters for E.L. Smith WTP

Parameter	Mean	Std. Dev.	Min.	Max.	Percentile		
					5th	50th	95th
Influent pH	8.2	0.2	7.7	8.6	7.8	8.2	8.5
Influent turbidity (NTU)	37.6	110.7	1.0	1364.0	2.0	5.0	168.3
Influent water temperature (°C)	7.5	7.8	0.0	24.4	0.0	3.5	20.0
Influent colour (TCU)	11.3	11.5	2.0	74.0	3.0	7.0	35.3
Influent hardness (mg/L as $CaCO_3$)	163.8	14.0	102.0	199.0	142.0	163.0	189.0
Influent alkalinity (mg/L)	131.6	9.4	96.0	162.0	117.0	131.0	149.0
Alum dose (mg/L)	43.5	29.2	16.0	182.0	20.8	30.5	103.3
PAC dose (mg/L)	5.6	17.6	0.0	151.0	0.0	0.0	34.2
Polymer dose (mg/L)	0.2	0.1	0.1	2.1	0.1	0.2	0.4
Overflow rate (m^3/d)	56.8	14.8	14.7	109.1	44.2	52.1	93.5
Lag 1 influent turbidity (NTU)	-0.1	80.2	-909.0	902.0	-29.5	0.0	21.5
Lag 1 influent colour (TCU)	0.0	3.9	-15.0	35.0	-4.0	0.0	4.0
Clarifier effluent colour	2.2	1.3	0.1	6.9	0.5	2.0	4.8

4.2.3.3 Determination of the Appropriate Number of Hidden Layer Neurons

The optimal number of hidden layer neurons for each of the backpropagation architectures was determined in the same manner as described for Rossdale WTP, Plant #2. For the 3-layer network, the optimal number of neurons was found to be 85. For the 4-layer network, a combination of 40 neurons in the first hidden layer and 40 neurons in the second hidden layer offered the best network performance.

4.2.3.4 Evaluation of Model Stability

Once again, alternate training, testing, and production sets were prepared from the original data file. The two candidate models were retrained using the alternate data sets in order to ensure that the results obtained were not dependent on the manner in which the data was extracted. A comparison of each of the results obtained on the production data set from both the original and alternate data files for each model is presented in Table 4.28.

Table 4.28 Evaluation of model stability

	Original Data Set		Alternate Data Set	
Architecture	R^2	MSE	R^2	MSE
3-layer network	0.85	0.24	0.75	0.40
4-layer network	0.88	0.19	0.87	0.21

While the 3-layer network offered extremely good results on the original data set, there was a substantial decrease in performance when the architecture was retrained on the alternate data set. This suggests that the 3-layer model may not be as robust to process changes as the 4-layer model, whose performance on the alternate data set was only marginally inferior to that on the original data set. As such, the 3-layer architecture was eliminated from further consideration.

4.2.3.5 Model Fine-Tuning

The 4-layer backpropagation architecture was optimized in the manner previously described. When the model was retrained using momentum weight updates with a

learning rate of 0.1, a momentum of 0.9, and random pattern selection, the model performance was increased, as will be discussed in the following section.

4.2.3.6 Comparison of Candidate Models

As only one architecture, the 4-layer backpropagation architecture with 40 neurons in each hidden layer, was fully optimized, it is not possible to make a direct comparison with other architectures initially evaluated for use in the E.L. Smith WTP model. As demonstrated in Table 4.29, the model for E.L. Smith WTP offers better R^2 values, when applied to historical data, than those developed for Rossdale WTP (Tables 4.20 and 4.26).

Table 4.29 Model results for E.L. Smith WTP

Data Set	R^2	Mean Squared Error	Mean Absolute Error (TCU)
Training	0.92	0.13	0.27
Testing	0.89	0.17	0.31
Production	0.89	0.17	0.30
Cross-validation	0.89	0.17	0.31

A plot of the actual and predicted values of clarifier effluent colour across all of the values in the production set is presented in Figure 4.17. It appears as though there are no major deficiencies in the model; daily trends as well as plant upsets are predicted with similar levels of accuracy. In order to ensure that there are no obvious trends in the model residuals, a plot of the residuals across all of the patterns in the production set is presented in Figure 4.18. While the model residuals show more scatter than either of the two Rossdale WTP models, the majority of the residuals still fall within the -0.5 to 0.5

TCU range. With respect to the distribution of the model residuals, a histogram delineating the frequency of error on the absolute scale is presented in Figure 4.19. Once again, approximately 95 % of the model residuals are within 0.8 TCU of the actual recorded values.

4.3 Model Evaluation with Real-time Data

While all of the models developed for the two AQUALTA facilities performed admirably on historical data, the true measure of the models' performance is their ability to provide accurate predictions for real-time data. For the purposes of this study, a simulated real-time environment was employed; each of the models was applied retroactively to data obtained from AQUALTA's SCADA system on a four-hour cycle for the months of March, April, and May 1998. The four-hour cycle was chosen to coincide with the frequency of laboratory analysis updates, and the three-month time frame was selected to coincide with spring-thaw, one of the most operationally difficult time periods for water treatment facilities. Overall, the number of data patterns evaluated for each of the plants was 530 for each of the Rossdale plants and 550 for E.L. Smith. The differences in the number of patterns used for each of the facilities is due to the fact that 5 days of data from Rossdale were unusable. The results of the real-time data evaluation for each of the models is presented in Table 4.30. In addition, as mentioned in the introduction, a mean absolute error of 0.5 TCU in the 2-3 TCU range is desirable if the models are to be employed in process control applications. The results of the real-time data evaluation when the effluent colour was in the 2-3 TCU range are presented in Table 4.31.

Table 4.30 Results of model evaluation using real-time data

Model	R^2	Mean Absolute Error (TCU)
Rossdale WTP, Plant #1	0.71	0.25
Rossdale WTP, Plant #2	0.76	0.31
E.L. Smith WTP, Clarifier 1	0.67	0.57
E.L. Smith WTP, Clarifier 2	0.72	0.49

Table 4.31 Results of model evaluation using real-time data, 2.0 - 3.0 TCU effluent colour range

Model	Mean Absolute Error (TCU)
Rossdale WTP, Plant #1	0.25
Rossdale WTP, Plant #2	0.33
E.L. Smith WTP, Clarifier 1	0.75
E.L. Smith WTP, Clarifier 2	0.54

With respect to the results obtained for Rossdale WTP, Plant #1, the model showed slightly less correlation, in terms of the R^2 values, than when applied to historical data (Table 4.20). On the absolute scale, however, the mean absolute error on the production set improved by approximately 0.05 TCU on the real-time data. A plot of the actual clarifier effluent colour along with the model predicted clarifier effluent colour is presented in Figure 4.20. The model appears to follow daily plant operations quite well, with most of the variation being attributed to daily fluctuations in the measured clarifier effluent colour.

When applied to the real-time data, the model results for Rossdale, Plant # 2 improve considerably over those obtained on historical data (Table 4.26). A plot of the actual clarifier effluent colour and the predicted clarifier effluent colour is presented in Figure 4.21. While the model follows actual operational trends quite well, as with the model for Plant #1, most of the deviations stem from the daily fluctuations of the output parameter.

In some cases, the model tends to overpredict the value of the output parameter, although the overall mean of the model residuals was found to be essentially zero ($(10)^{-15}$).

In spite of demonstrating the best performance on historical data, the model for E.L. Smith WTP had the poorest performance of the three models when applied to real-time data. As was previously discussed, the model for E.L. Smith was developed using data averaged from the two clarifiers. While most of the input parameters had the same values for both clarifiers, the values of polymer dose, overflow rate, and clarifier effluent color were averaged between the two plants. With respect to the real-time evaluation it was necessary to apply the same model to each of the clarifiers, in spite of the presence of small operational differences between them. Plots of the actual and predicted clarifier effluent colour for clarifier 1 and 2 are presented in Figures 4.22 and 4.23 respectively. Of the two clarifiers, clarifier 1 demonstrated more variable performance, as the fluctuation in successive values of clarifier effluent colour is greater than that observed for clarifier 2. In both cases, the model tends to underpredict the value of clarifier effluent colour at higher values, although the mean of all the residuals is again essentially zero for both models.

With respect to the criteria established in the introduction section for successful model development, both of the Rossdale WTP models demonstrated excellent performance in the 2.0 - 3.0 TCU range of clarifier effluent colour, with mean absolute errors of 0.25 and 0.33 TCU for Plant #1 and Plant #2, respectively. When applied to clarifier 2, the model for E.L. Smith exceeded the 0.5 TCU mean absolute error target by only 0.04 TCU.

When applied to clarifier 1 however, this target was exceeded by 0.25 TCU. From a process control standpoint, these errors may still prove to be acceptable. Few studies have been conducted on the accuracy of the decisions made by the plant operators at either facility, and the proposed models may very well offer superior performance when used in process control.

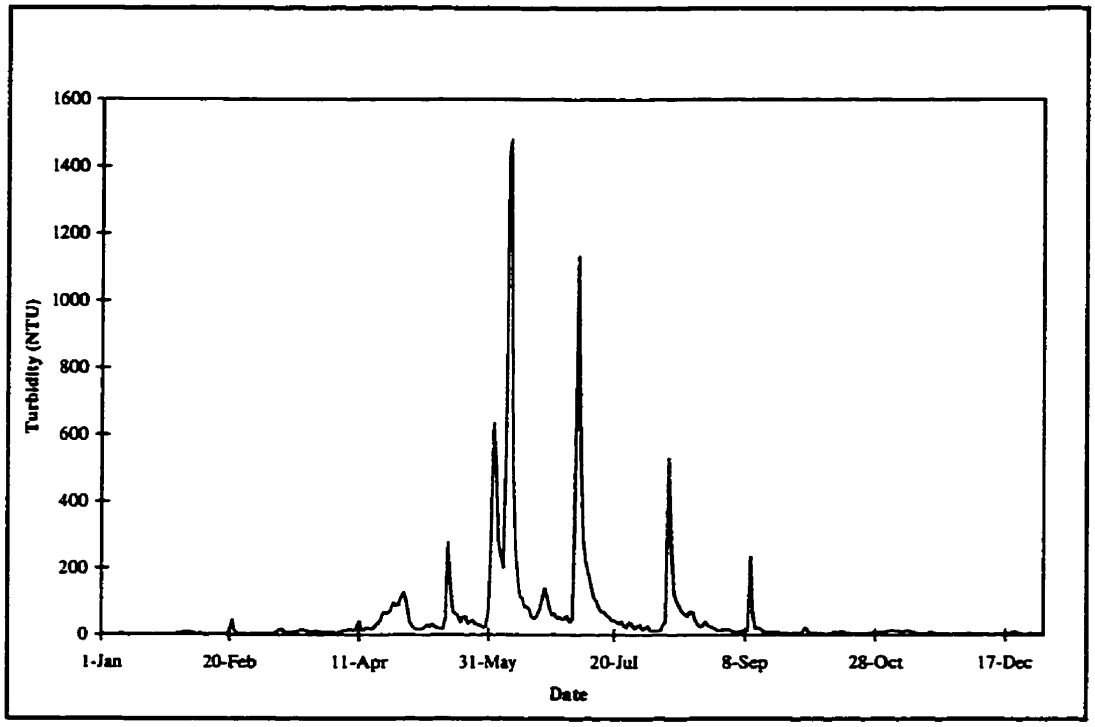


Figure 4.1 Rossdale WTP, raw water daily average turbidity, 1995

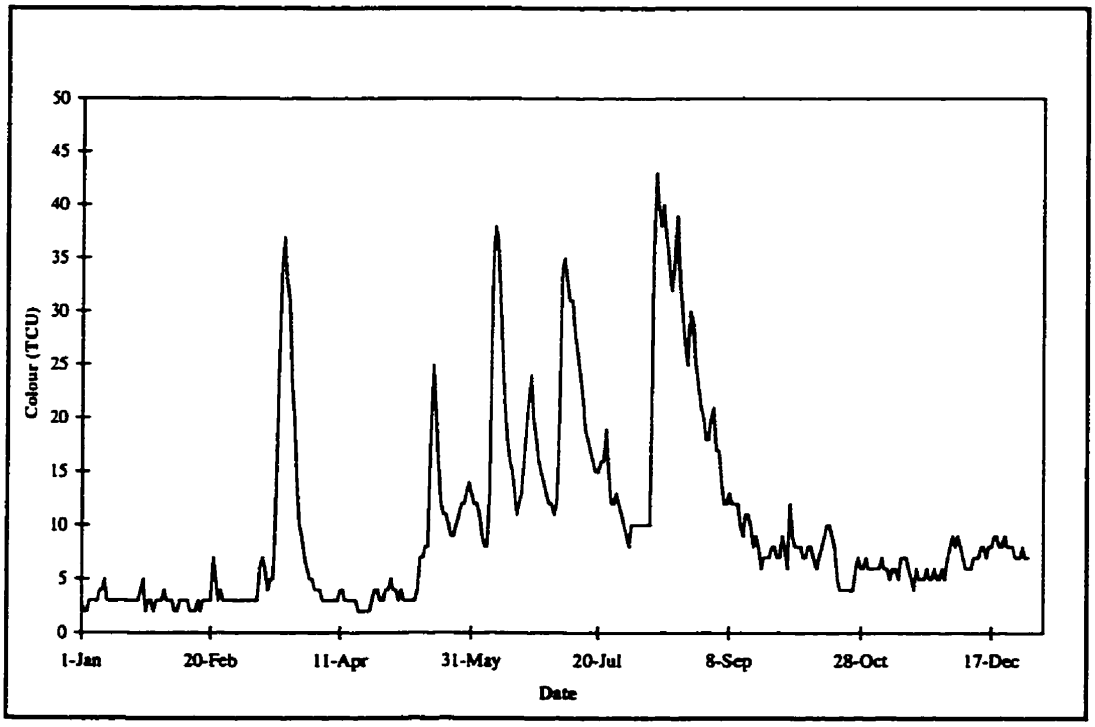


Figure 4.2 Rossdale WTP, raw water daily average colour, 1995

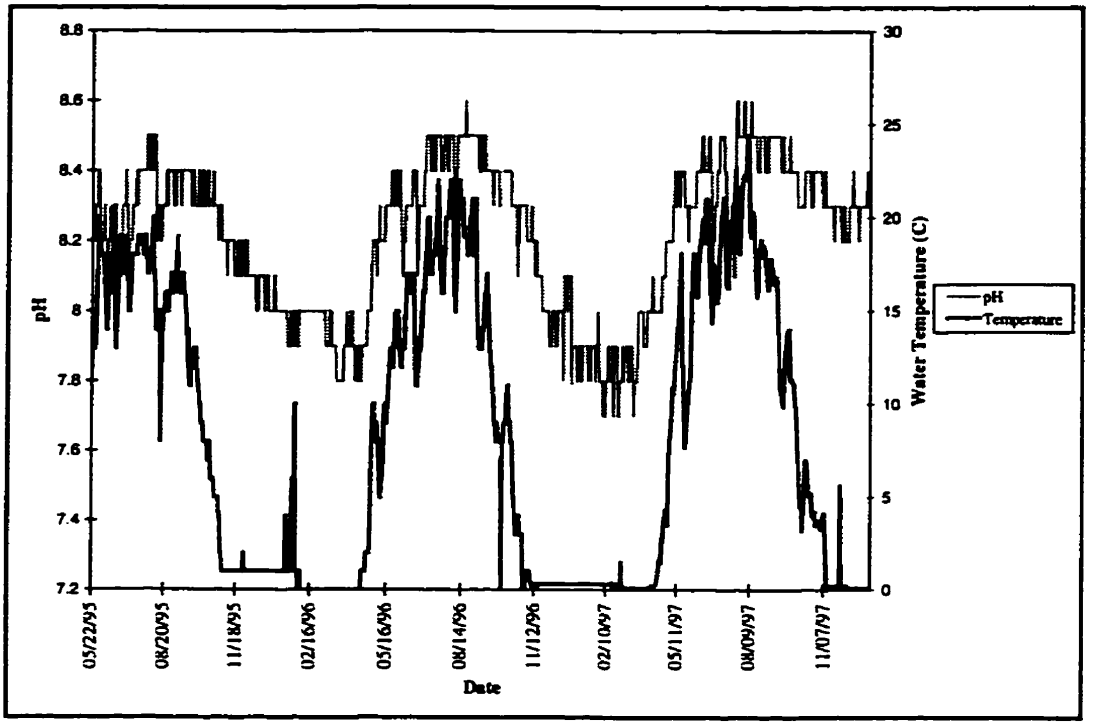


Figure 4.3 E.L. Smith WTP, raw water pH and temperature

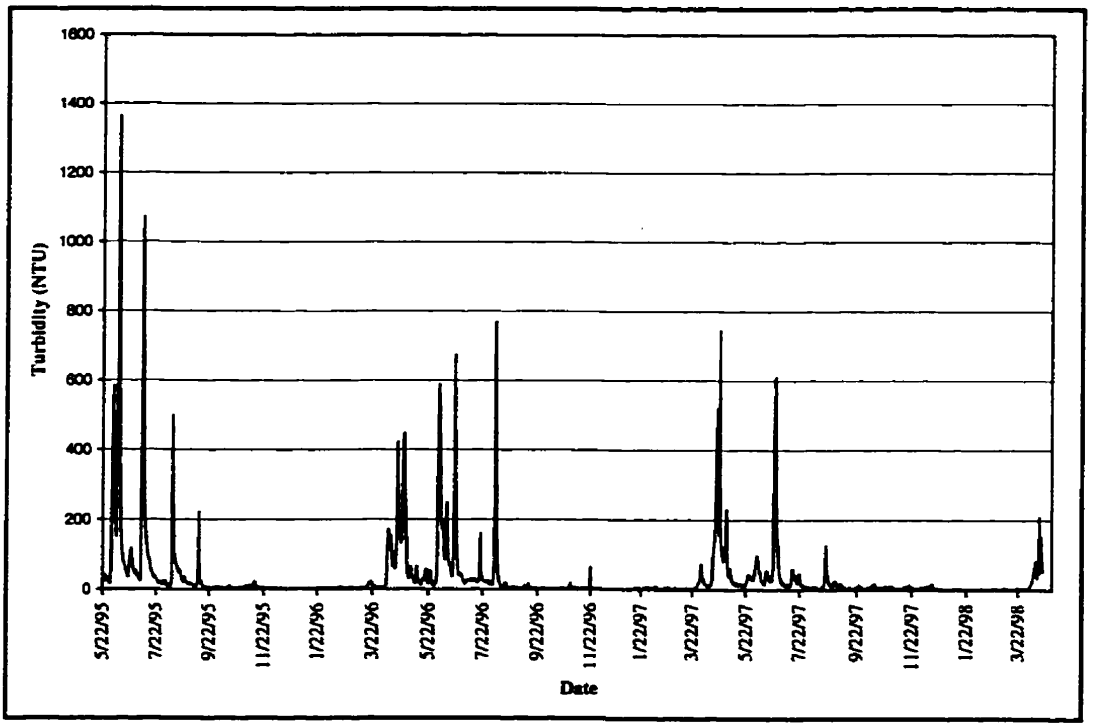


Figure 4.4 E.L. Smith WTP, raw water turbidity, 1995-1998

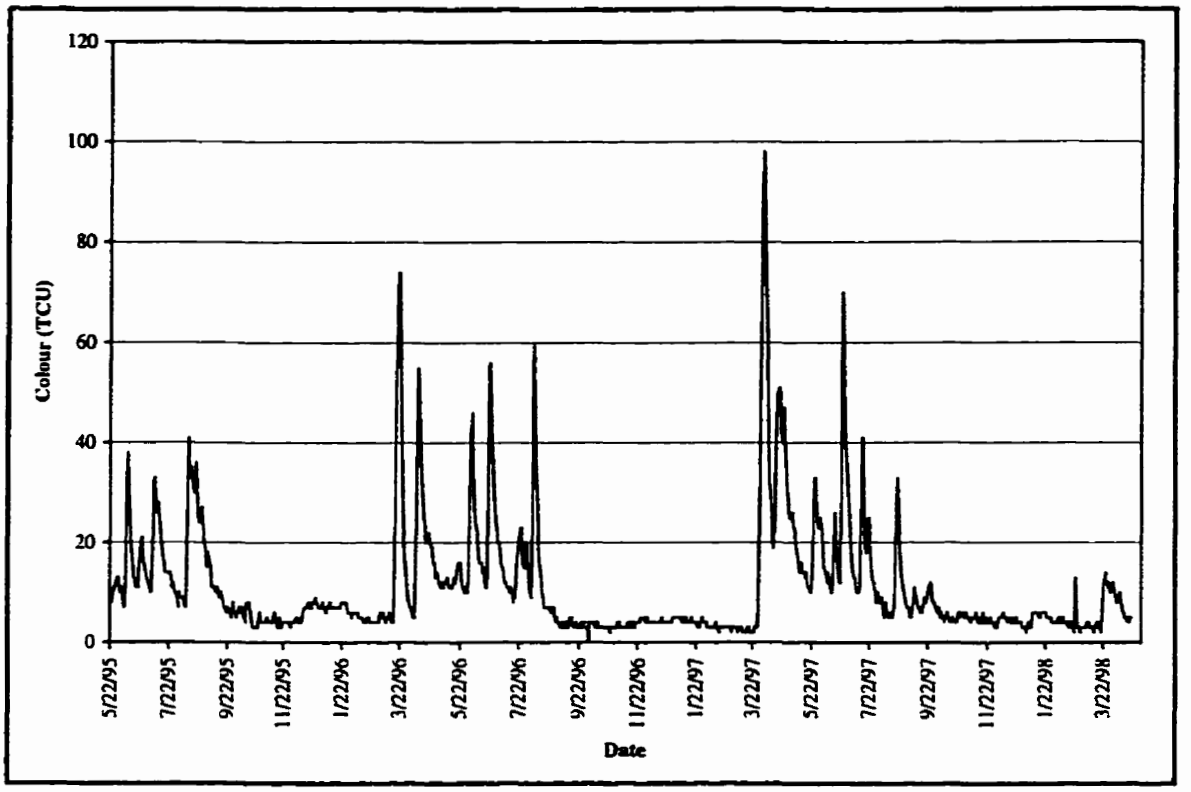


Figure 4.5 E.L. Smith WTP, raw water colour, 1995-1998

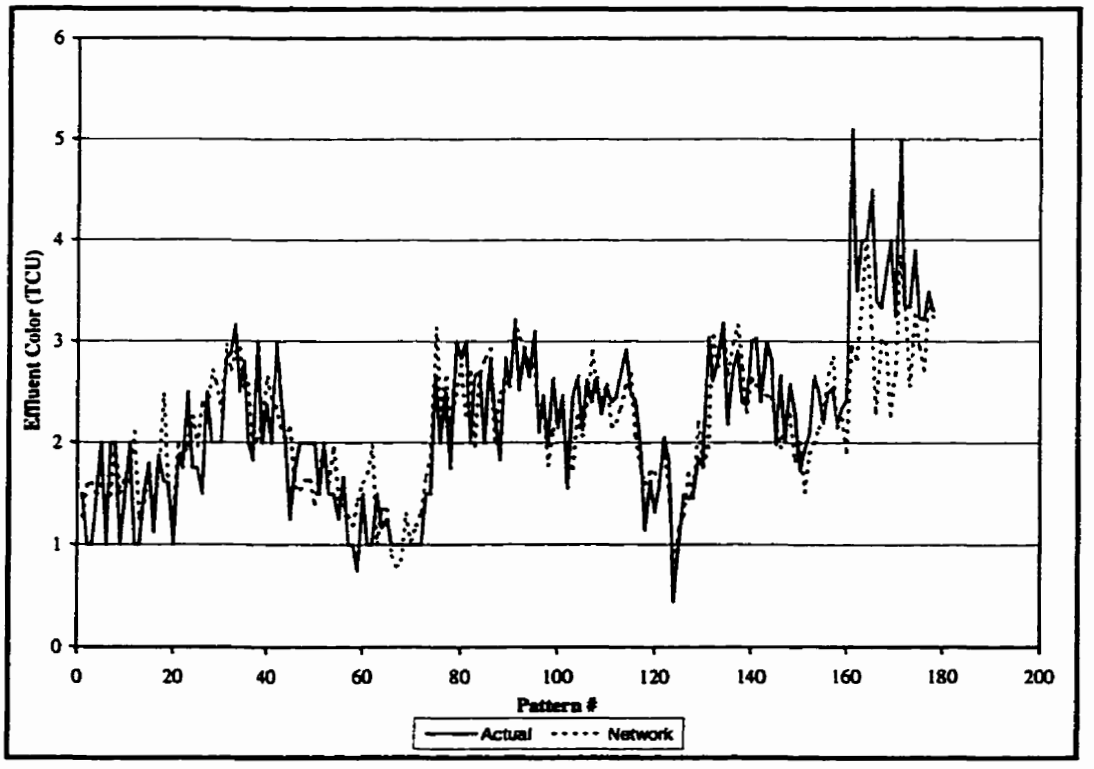


Figure 4.6 Preliminary model results for Rossdale WTP, Plant #1

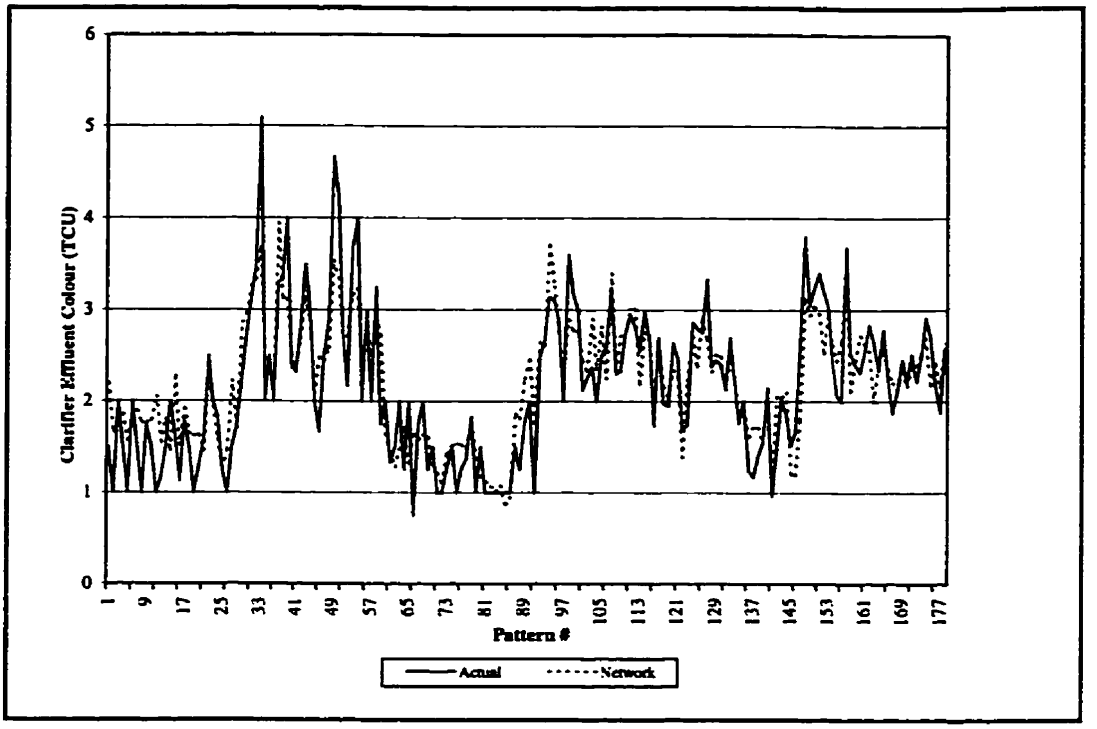


Figure 4.7 Model results for Rossdale WTP, Plant #1

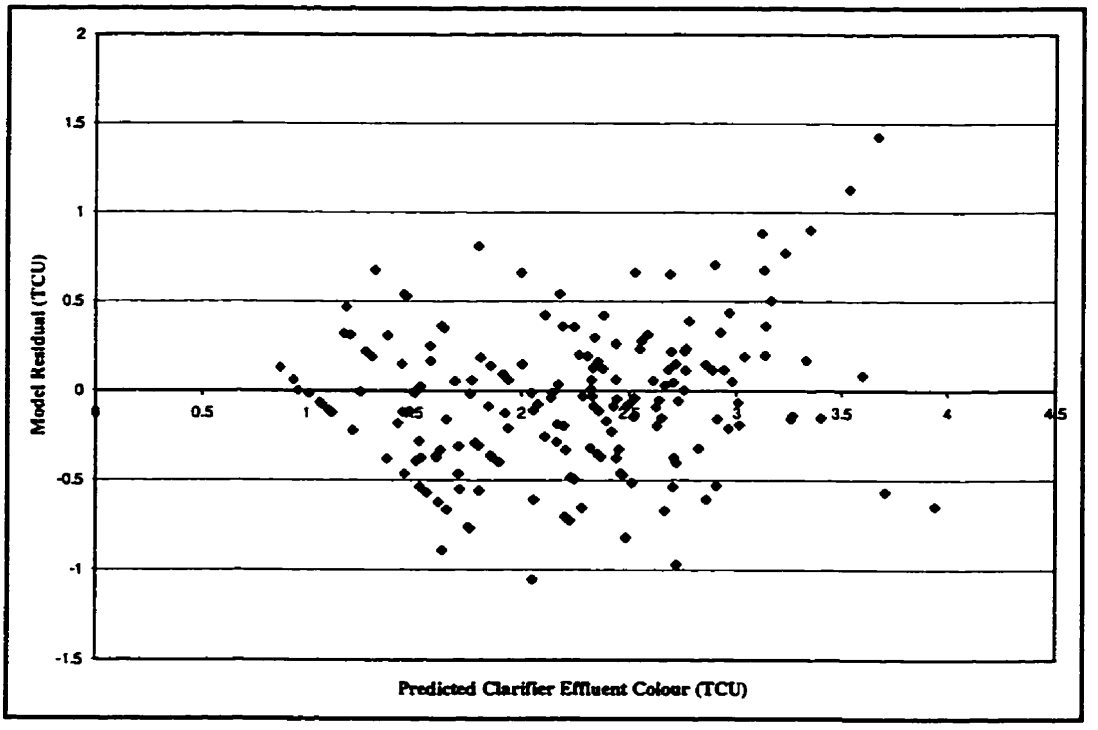


Figure 4.8 Rossdale WTP, Plant #1 model residuals

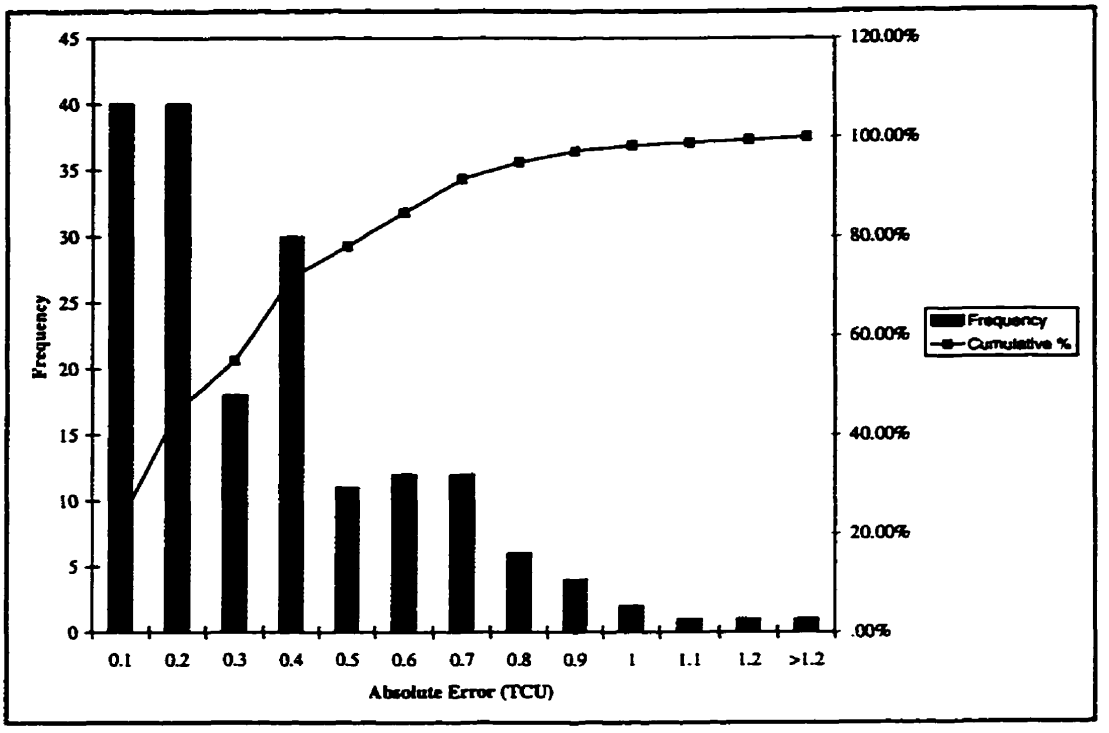


Figure 4.9 Distribution of absolute error for Rossdale WTP, Plant #1 model

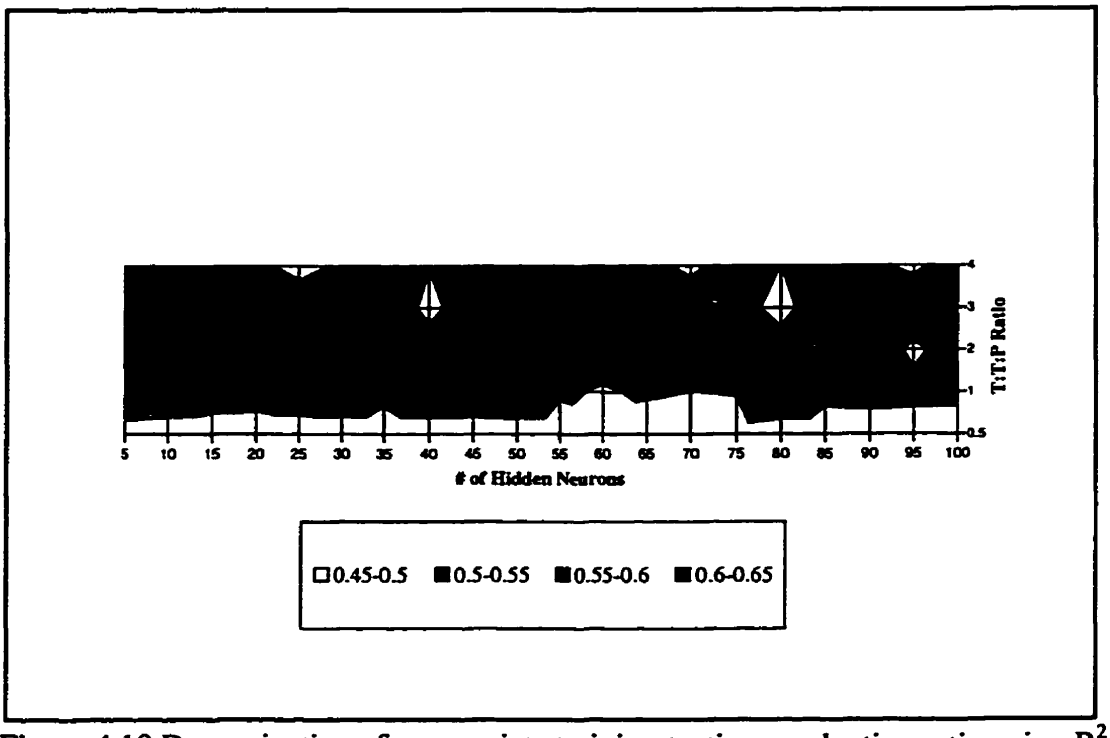


Figure 4.10 Determination of appropriate training:testing:production ratio using R^2

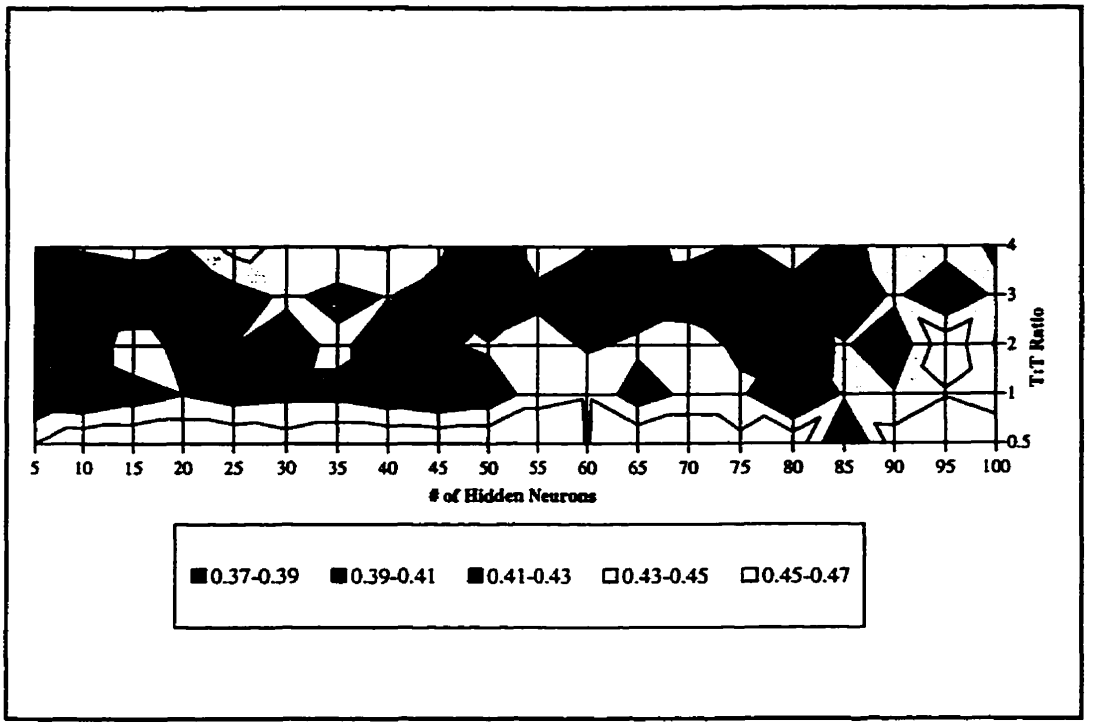


Figure 4.11 Determination of appropriate training:testing:production ratio using mean absolute error

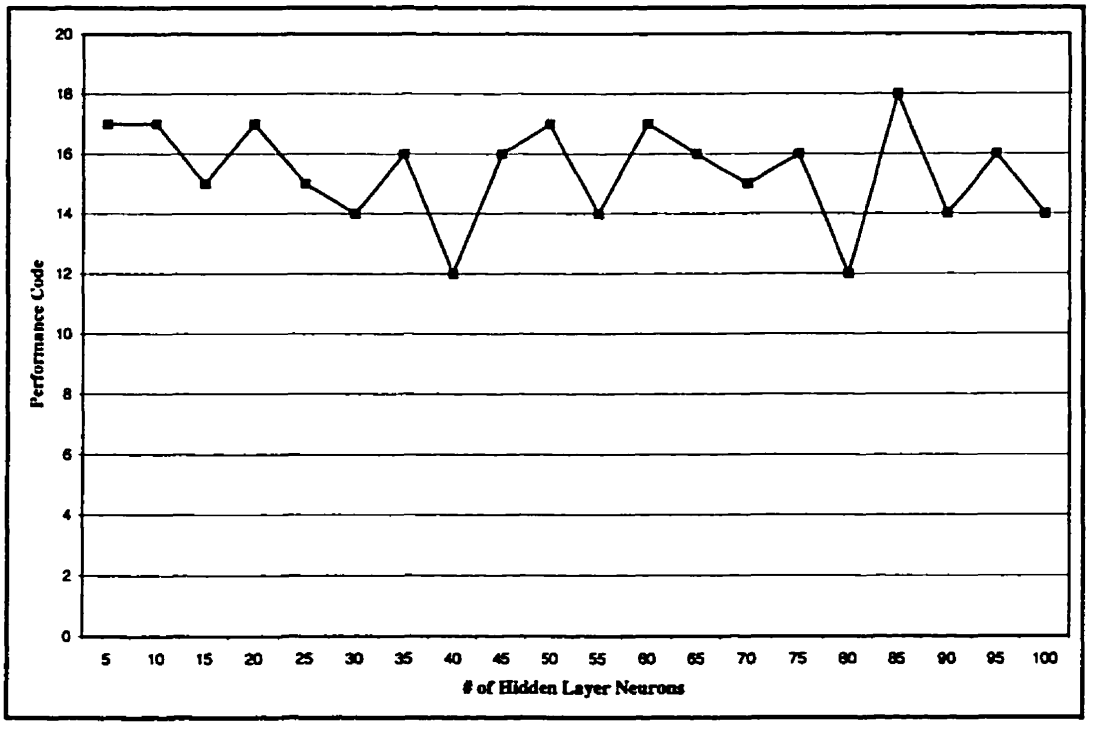


Figure 4.12 Determination of the appropriate number of hidden layer neurons for the 3-layer backpropagation network using performance codes

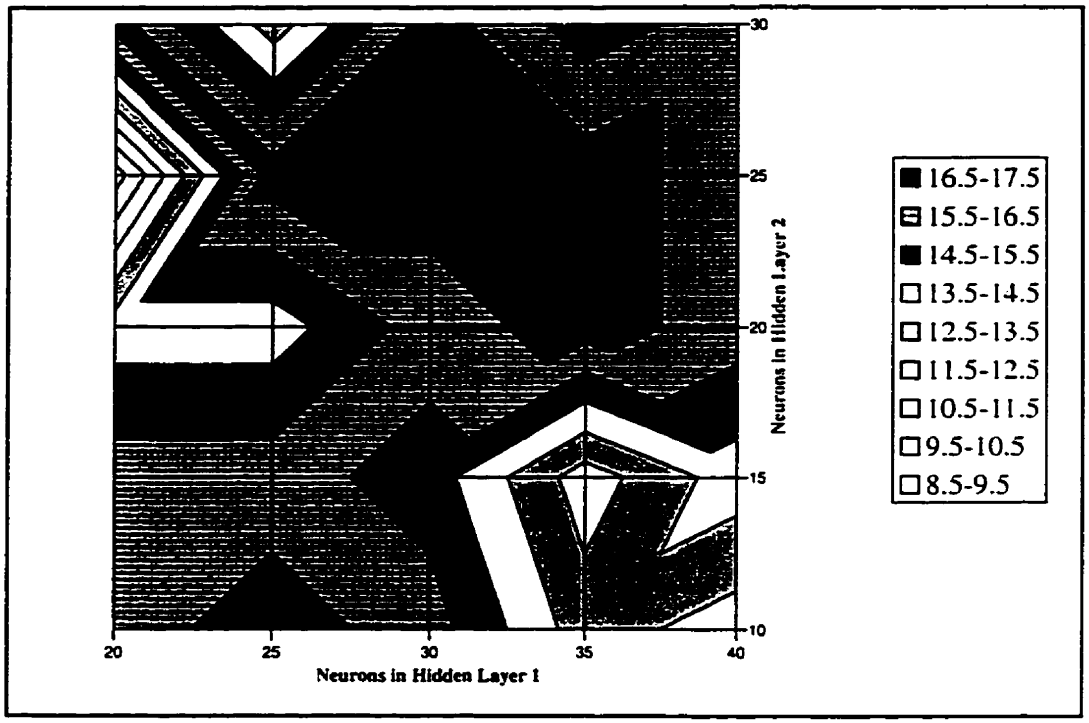


Figure 4.13 Determination of the appropriate number of hidden layer neurons for the 4-layer backpropagation network using performance codes

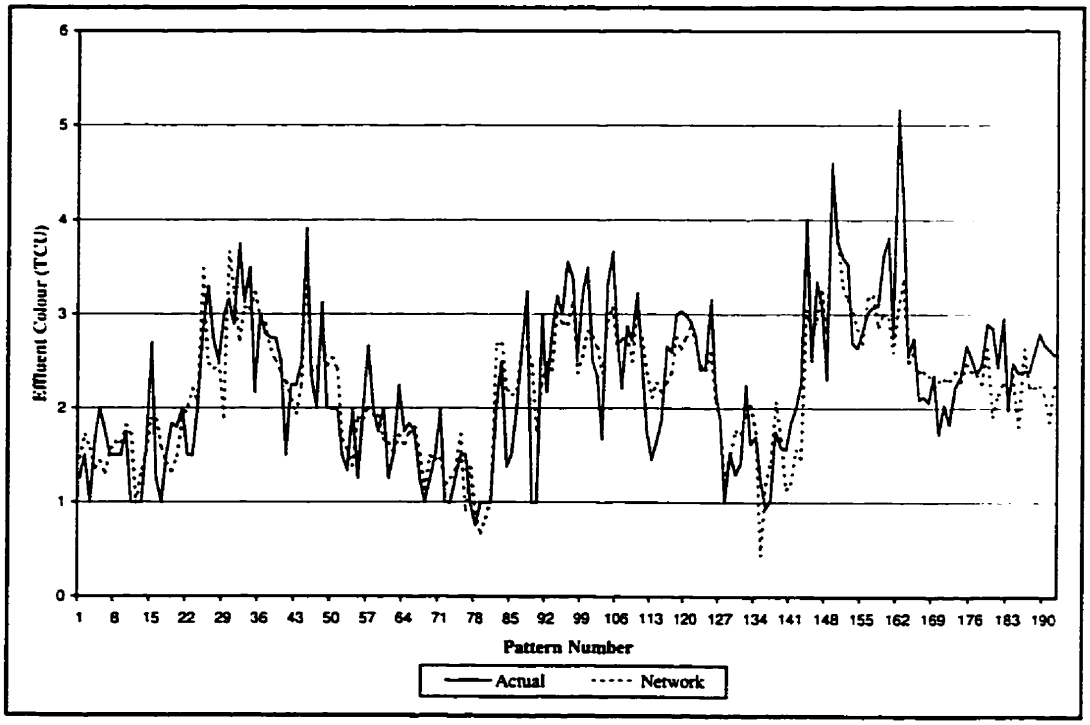


Figure 4.14 Model results for Rossdale WTP, Plant #2

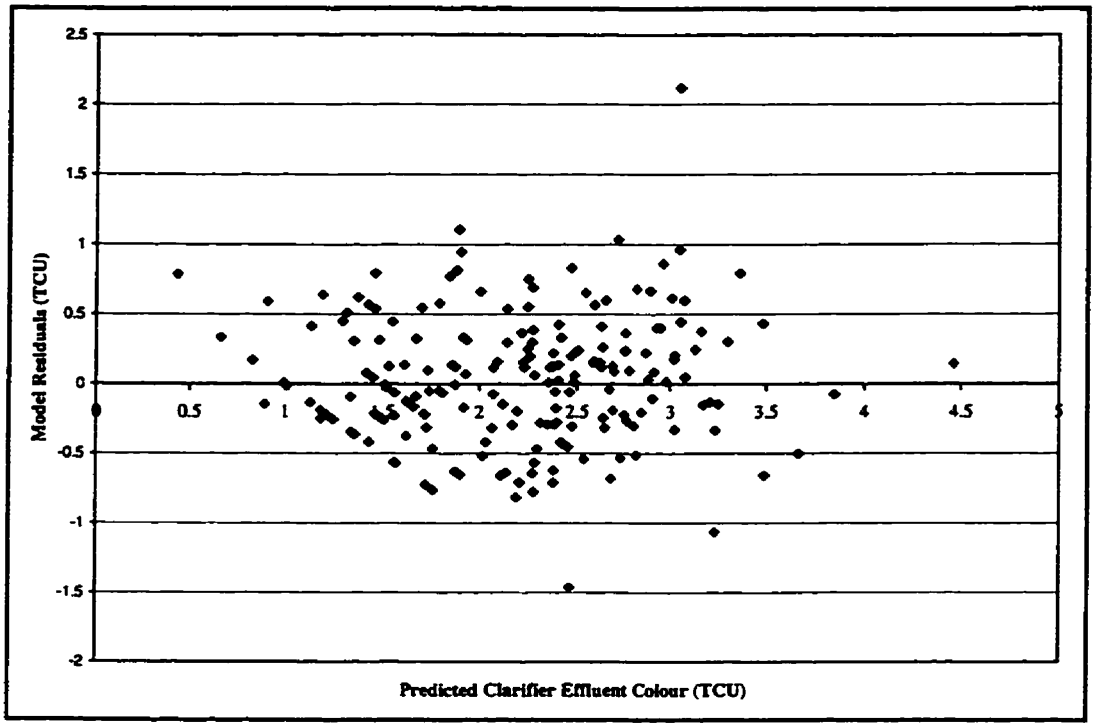


Figure 4.15 Rossdale WTP, Plant #2 model residuals

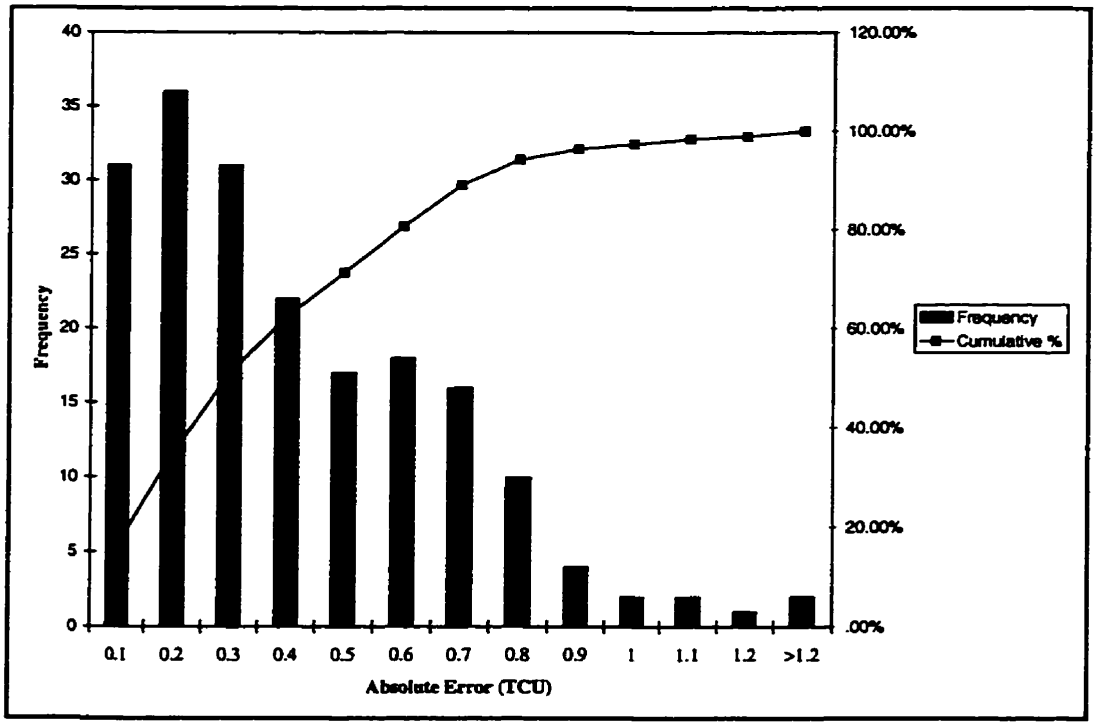


Figure 4.16 Distribution of absolute error for Rossdale WTP, Plant #2 model

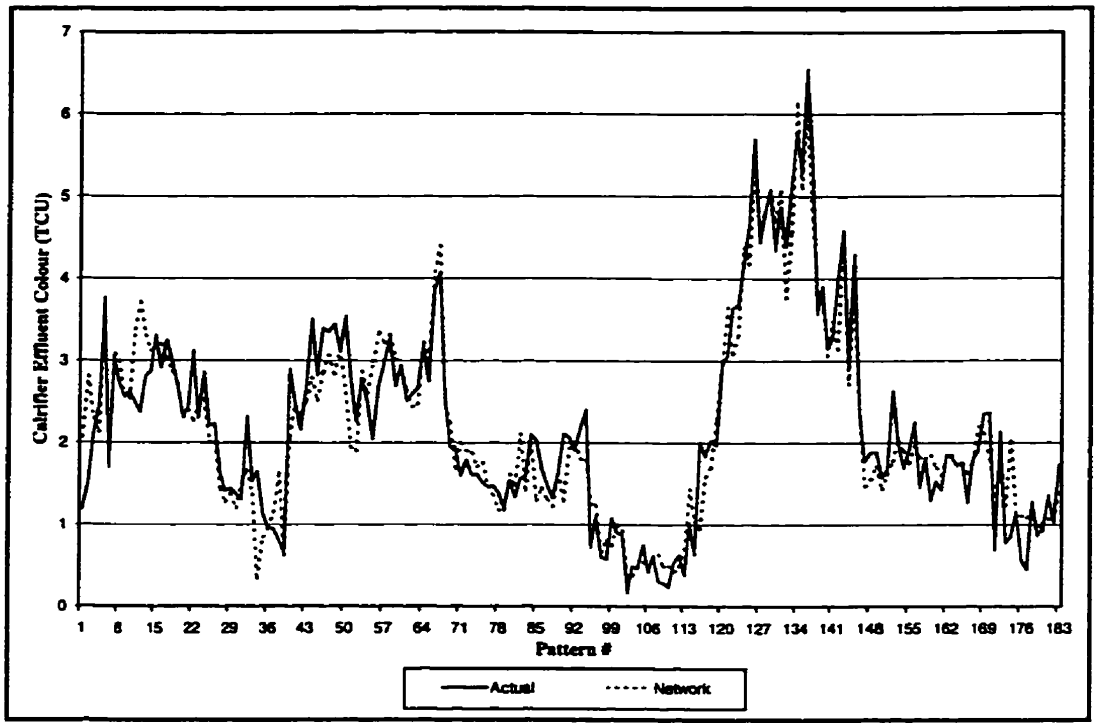


Figure 4.17 Model results for E.L. Smith WTP

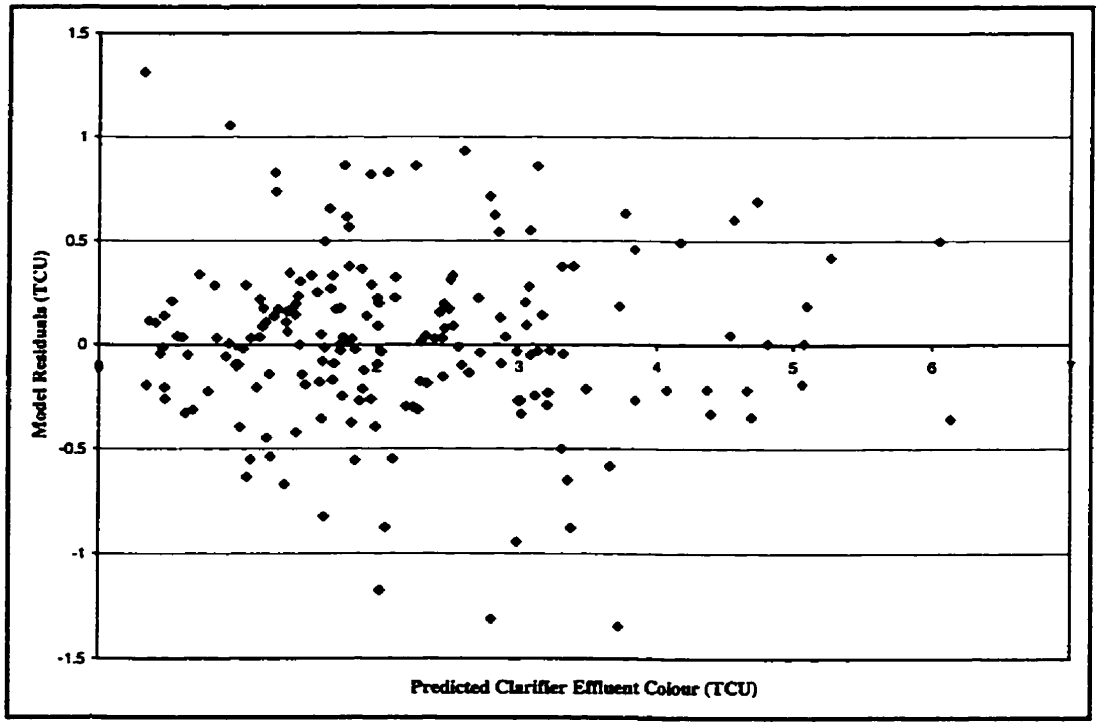


Figure 4.18 E.L. Smith WTP model residuals

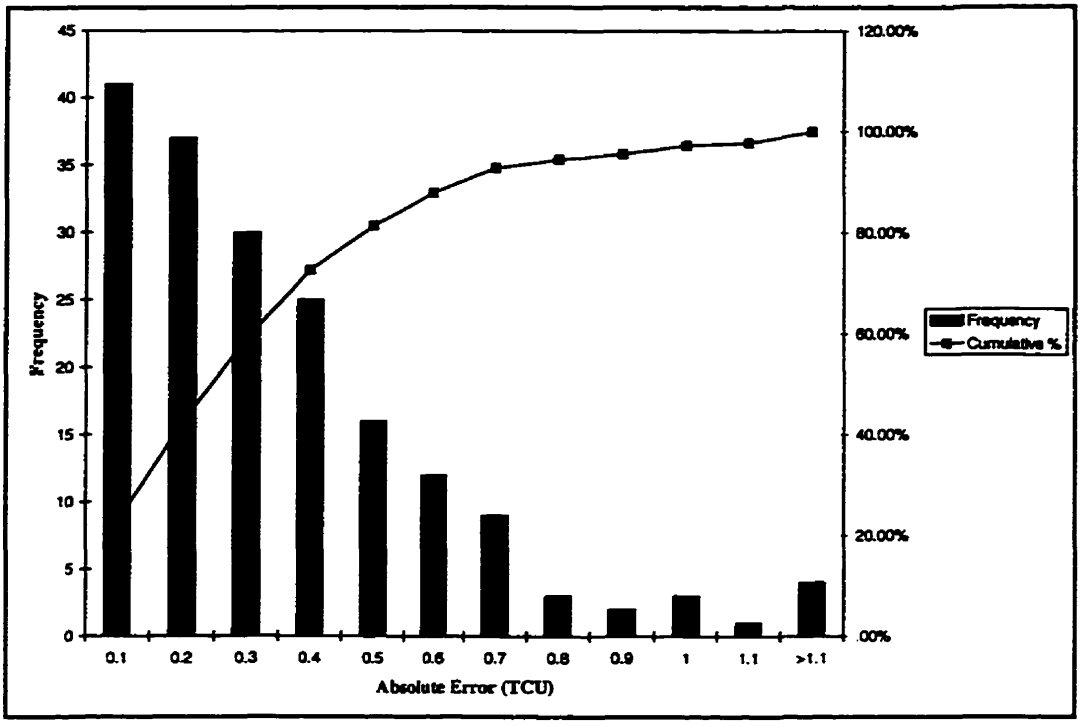


Figure 4.19 Distribution of absolute error for E.L. Smith WTP model

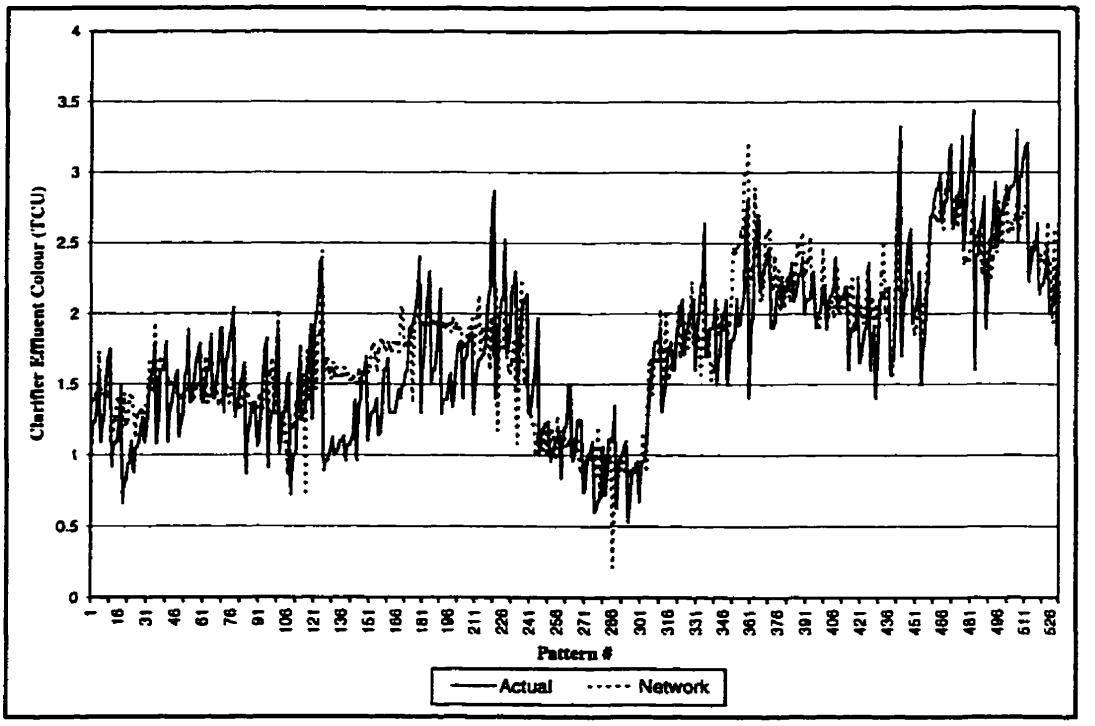


Figure 4.20 Model results for Rossdale WTP Plant #1 using real-time data

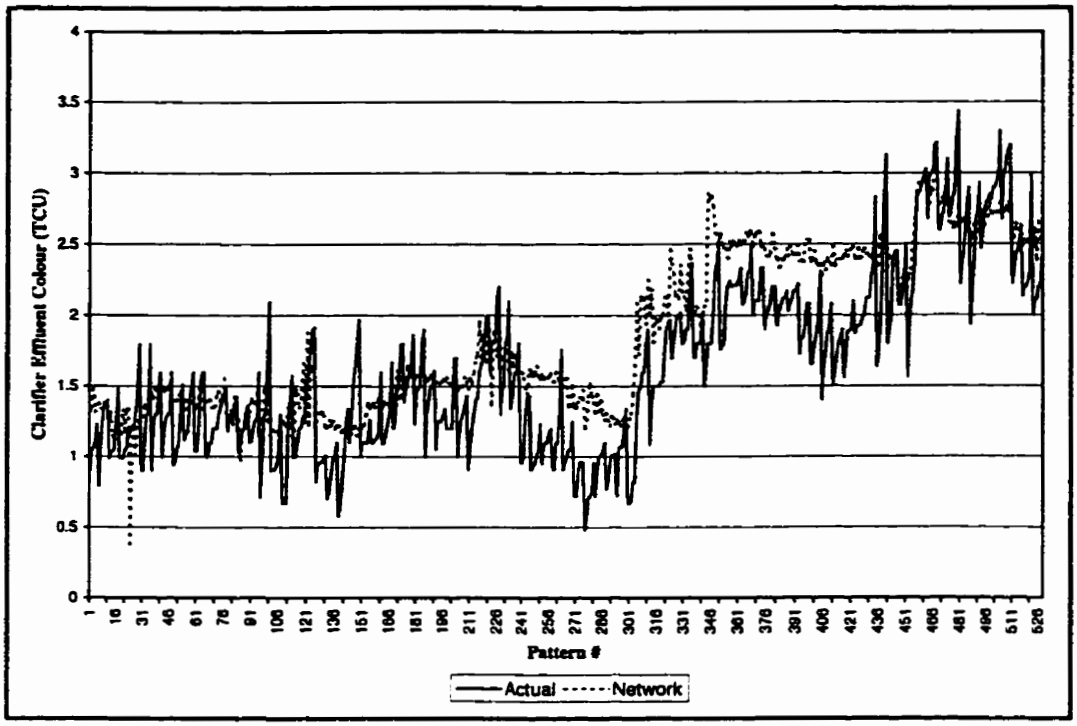


Figure 4.21 Model results for Rossdale WTP Plant #2 using real-time data

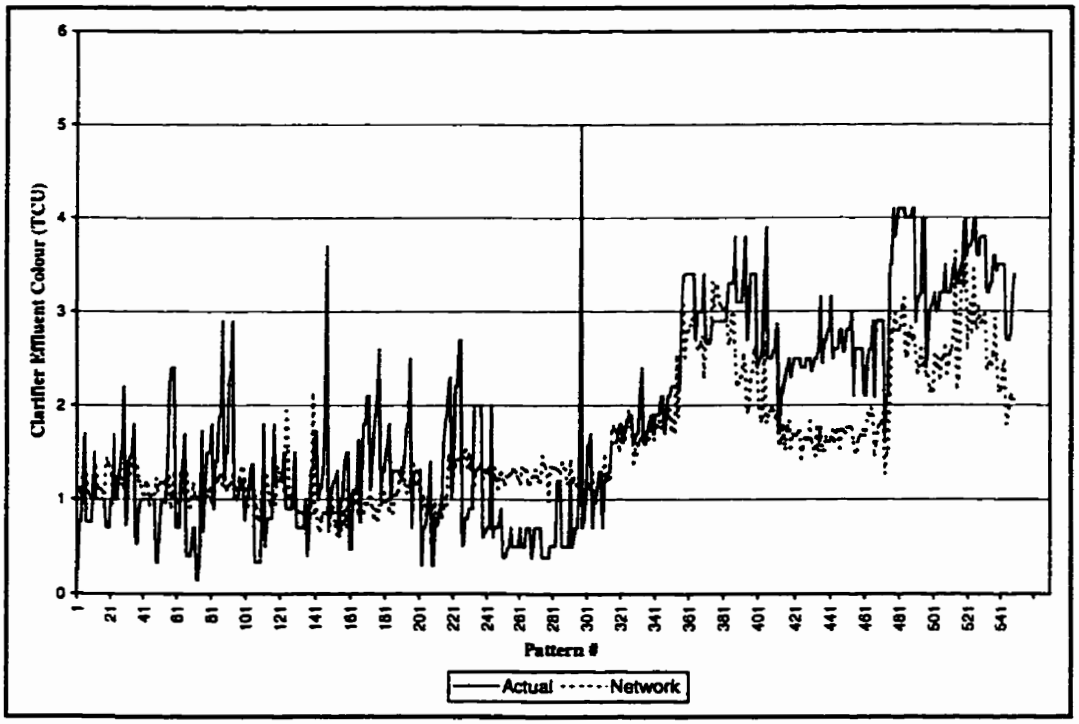


Figure 4.22 Model results for E.L. Smith WTP clarifier 1 using real-time data

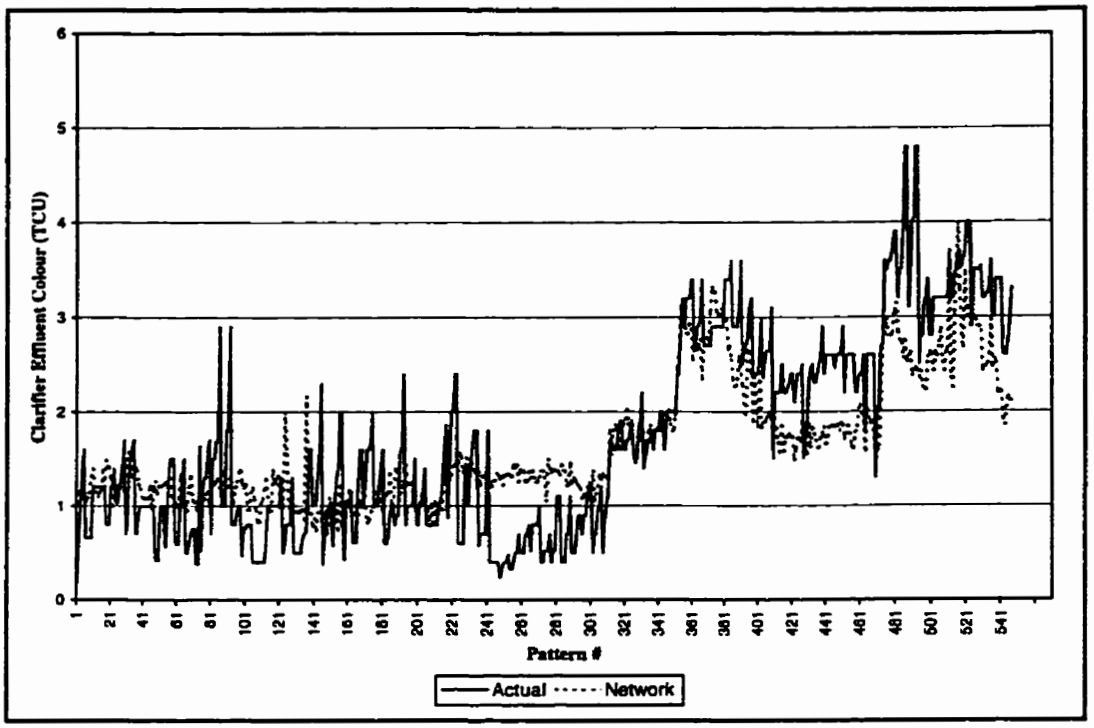


Figure 4.23 Model results for E.L. Smith WTP clarifier 2 using real-time data

5.0 APPLICATIONS

The completed models can be used in a wide variety of process control and related applications. In reference to the internal model control (IMC) scheme discussed in the background information section, the models have all been trained to represent the forward dynamics of the enhanced coagulation process; given a set of process inputs, the models are able to predict the process output. As such, the models can be used in indirect model control as a tool to help plant operators select the appropriate process operating conditions. In order to be used in direct process control, where the models select appropriate operating conditions independently, given the current raw water quality and desired clarifier effluent colour, the models would need to be retrained to represent the process inverse dynamics.

The remainder of this section will focus on the discussion of two potential offline methods of indirect process control. The operator decision verification method allows the operators to verify the effect of their control actions on the process prior to making changes in process operations. The virtual laboratory method can be used to generate heuristics or guidelines for a given influent water quality.

5.1 Operator Decision Verification

The NeuroShell 2 software allows completed ANN models to be accessed by a variety of other software applications using a dynamic link library (DLL). Languages such as

Visual Basic and C, as well as applications such as Microsoft Excel, can incorporate trained networks as part of their source code or as a spreadsheet function. In Excel, for example, the following function will return the predicted output of a trained network, given a series of inputs:

=CALL("NSHELL2.DLL","Predict","pppp",def_path,input_array, output_number)

where "def_path" is the path of the network file, "input_array" is the location of the values of the input parameters within the Excel spreadsheet, and "output_number" is the number of the desired output. The "output_number" has a value of 1 for networks with only one output, as is the case for the current study.

A spreadsheet can then be set up where the plant operators can enter the values of the 6 water quality and the 2 time-series model input parameters. The operators can then enter their proposed control actions for the given water quality information: alum dose, polymer dose, PAC dose, and overflow rate. The Excel spreadsheet will calculate the model prediction for clarifier effluent colour based on the information entered by the operators. If the predicted clarifier effluent colour is at an acceptable level, the operators can proceed with their proposed control actions. Otherwise, the levels of each of the process parameters can be varied in the spreadsheet until an acceptable level of clarifier effluent colour is achieved.

This process can also be used in the training of new plant operators. A spreadsheet containing numerous sets of water quality and time-series information can be made available so that operators-in-training can learn the effects of proposed control actions on the quality of the clarifier effluent.

5.2 Virtual Laboratory

The completed models can alternatively be used as virtual full-scale laboratories to provide insight into the enhanced coagulation process. The information generated from the virtual experiments can be used to assist plant operators to select the appropriate operating conditions for a given influent water quality.

Initially, a specific type of water is selected. For the purposes of the current discussion highly coloured, moderately turbid water, typical of the beginning of spring-thaw, is selected. By examining historical data records, appropriate values for each of the raw water quality and time series parameters can be selected (Table 5.1).

Table 5.1 Typical raw water characteristics of in early spring-thaw

Parameter	Value
PH	8.0
Turbidity (NTU)	19
Temperature (°C)	0.1
Colour (TCU)	50
Total hardness (mg/L as $CaCO_3$)	153
Total alkalinity (mg/L)	129
Lag-1 Turbidity (NTU)	10
Lag-1 Colour (TCU)	30

Typical ranges of process parameters for a given water treatment plant can then be identified. At E.L. Smith WTP, the ranges of each of the process parameters that would typically be used to treat the water described above are presented in Table 5.2

Table 5.2 Typical values of process parameters at E.L. Smith WTP during early spring-thaw conditions

Parameter	Range
Alum dose (mg/L)	30 - 70
PAC dose (mg/L)	40 - 80
Polymer dose (mg/L)	0.25 - 0.35
Overflow rate (m/d)	40 - 60

It should be noted that overflow rate is typically not varied in response to raw water quality, but rather in response to customer water demand. As such, colour control is maintained primarily by varying each of the chemical dosing levels in the described ranges. Plots of each of alum dose, PAC dose, and polymer dose over a range of overflow rates can be plotted against clarifier effluent colour. The clarifier effluent colour values are generated using an Excel spreadsheet, along with the "Call" function described above. Alternatively, one or more of the process parameters can be held constant, while other are varied alone or in combination. Figure 5.1 shows the variation in color with alum dose and PAC dose in the ranges of 30 - 90 mg/L and 40 - 80 mg/L, respectively. The polymer dose and overflow rate were held constant at 0.30 mg/L and 50 m/d, respectively. As shown in the figure, the model responds gradually and logically to changes in chemical doses, indicating that the model appears to have determined trends between input and output parameters. A similar plot of the effects of polymer dose on clarifier effluent colour is presented in Figure 5.2, with the PAC dose held constant at 60 mg/L. As shown, the model predicts that varying polymer dose in the range of 0.2 to 0.4 mg/L does

not significantly impact the removal of colour over a wide range of alum doses. This result is in agreement with the plant operators' knowledge base; polymer has historically been used to increase turbidity removals and has had little effect on the removal of colour.

The plots can be used by the plant operators to select appropriate dosing combinations for a given influent water quality. If a clarifier effluent colour of 2.0 TCU was desired at E.L. Smith WTP during spring thaw, for example, an alum dose of 60 mg/L in combination with a PAC dose of 60 mg/L and a polymer dose of 0.30 mg/L would be effective at typical overflow rates.

Once the Excel spreadsheet has been created, any one of the input variables can be changed and a new plot generated with ease. As such, a series of dose / response plots can be generated for any given water quality in order to give the plant operators a graphical representation of possible effective dose combinations.

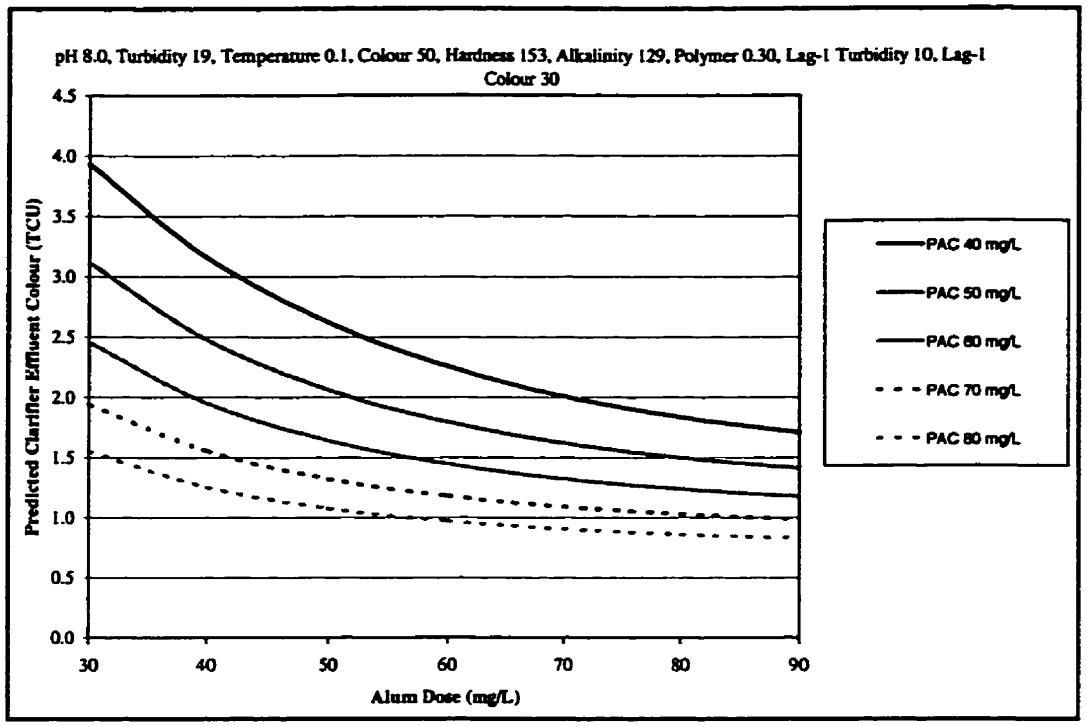


Figure 5.1 Effect of varying PAC dose and alum dose on clarifier effluent colour at E.L. Smith WTP during early spring-thaw

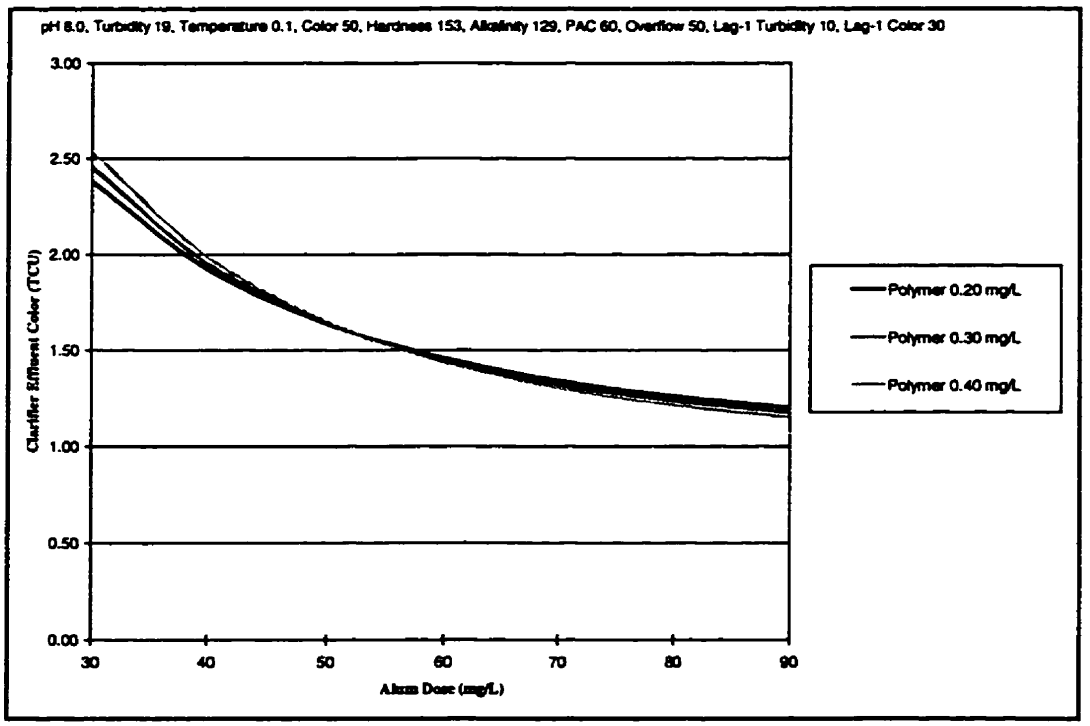


Figure 5.2 Effect of varying polymer dose and alum dose on clarifier effluent colour at E.L. Smith WTP during early spring-thaw

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The purpose of this study was to develop full-scale artificial neural network (ANN) models for the enhanced coagulation process at the E.L. Smith and Rossdale Water Treatment Plants in Edmonton, Alberta. For each facility, models were developed using historical data and were evaluated in a simulated real-time environment. With respect to performance on historical data, each of the models was able to predict clarifier effluent colour with a high degree of accuracy. When applied to simulated real-time data, the models for Rossdale WTP demonstrated better performance than on historical data. The E.L. Smith model exceeded the target maximum error by a small margin, although the validity of the target value has yet to be determined. Two methods of offline indirect model-based control were discussed, and the utility of the models in assisting operators to make appropriate control actions was demonstrated. Rather than being all encompassing however, this research demonstrates only a few of the many potential applications of the ANN modelling technique to the water treatment industry and related fields. Processes such as filtration and disinfection, for example, should be equally amenable to the technique.

6.2 Recommendations

Based on the knowledge gained in the development of the models, a number of areas requiring further study have been identified.

6.2.1 Applicability to Other Processes

As previously mentioned the ANN modelling technique should be equally applicable to other water treatment and related processes. Studies should be completed to confirm the applicability of the technique to filtration and disinfection. In addition, the technique may be applicable to the areas of maintenance scheduling, residual management, and distribution systems management.

6.2.2 Performance Criteria

If the developed models are to be used in process control applications, there is a need to establish criteria to determine whether the model performance is acceptable. The target of 0.5 TCU in the 2 - 3 TCU range was suggested by staff at AQUALTA, but does not have any real significance with regards to actual process operations. Studies should be undertaken to determine the accuracy and reproducibility of operator control actions, as well as the level of accuracy required for proper process control. Based on the results of these studies, more meaningful model performance criteria can be established.

6.2.3 Process Control

With regards to process control applications, the developed models should be implemented offline at the water treatment plants for control action verification and the training of new operators. Each of the models should also be retrained to represent the plant inverse dynamics so that they can eventually be used in a direct IMC system.

6.2.4 ANN Development Protocols

As mentioned several times, few ANN model development methodologies exist. While the protocols employed in the current study yielded acceptable models, further work in the area of protocol development is warranted.

6.2.5 Model Boundaries

While the ANN models provide accurate predictions in the domain on which they were trained, little is known about the ability of ANN models to extrapolate beyond the data ranges on which they were trained. A study should be conducted to determine the ANN model boundaries so that the models can be used with confidence in process control applications.

REFERENCES

Amirtharajah, A., K. E. Dennett and A. Studstill. 1993. Ferric Chloride Coagulation for Removal of Dissolved Organic Matter and Trihalomethane Precursors. *Wat. Sci. Tech* 27(11): 113-121.

Amirtharajah, A. and C. R. O'Melia. 1990. Coagulation Processes: Destabilization, Mixing, and Flocculation. *Water Quality and Treatment: a Handbook of Community Water Suplies*. F. W. A. Pontius. New York: McGraw Hill.

Anderson, J. A. and E. Rosenfeld, Eds. 1988. *Neurocomputing: Foundations of Research*. Cambridge, MA: The MIT Press.

AWWA Committee. 1979. Organics removal by coagulation: a review of research needs. *J. AWWA* 71(10): 588-603.

Babcock, D. B. and P. C. Singer. 1979. Chlorinaiton adn Coagulation of Humic and Fulvic Acids. *J. AWWA* 71(3): 149-152.

Bell, K. A., M. LeChevalier, M. Abbaszadegan, R. Mantiega, C. Volk and E. Ibrahim. 1997. *The Evaluation of Optimized Coagulation for Meeting Removal Requirements of the D/DBP and ESWT rules*. 1997 WQTC, Denver, CO.

Bernieri, A., G. Betta, C. Liguori and A. Losi. 1996. Neural Networks and Pseudo Measurements For Real Time Monitoring of Distribution Systems. *Ieee-Transactions-On-Instrumentation-and-Measurement* 45(2): 645-650.

Bhat, N. and T. J. McAvoy. 1990. Use of Neural Nets for Dynaimic Modelling and Control of Chemical Process Systems. *Computers chem. Engng* 14(4/5): 573-583.

Boger, Z. 1997. Experience in Industrial Plant Model Development Using Large Scale Artificial Neural Networks. *Information-Sciences* 101(3-4): 203-216.

Bowie, J. and M. T. Bond. 1977. Chemical Precipitaion-Coagulaiton for Organic Color Removal From Groundwaters. *Water Resources Bulletin* 13(6): 1269-1280.

Buscema, M. 1997. A General Presentation of Artificial Neural Networks. *Substance-Use-and-Misuse* 32(1): 97-112.

Chadik, P. A. and G. L. Amy. 1983. Removing Trihalomethane Precursors From Various Natural Waters by Metal Coagulants. *J. AWWA* 75(10).

Chen, V. C. P., J. Heldt, K. McGlynn and D. K. Rollins. 1996. *Critical issues in data collection and conditions when fitting predictive neural network models for dynamic processes*. ISA Conference Proceedings: Advances in Instrumentation and Control: ISA.

Cheng, R. C., S. W. Krasner, J. F. Green and K. L. Wattier. 1995. Enhanced Coagulation: A Preliminary Evaluation. *J. AWWA* 87(2): 91-103.

Clark, R. M., J. Q. Adams and J. Benjamin W. Lykins. 1994. DBP Control in Drinking Water: Cost and performance. *Journal of Environmental Engineering* 120(4): 759-782.

Collins, M. R., G. L. Amy and C. Steelink. 1986. Molecular Weight Distribution, Carboxylic Acidity, and Humic Substances Content of Aquatic Organic Matter: Implications for Removal during Water Treatment. *Environ. Sci. Technol* 20(10): 1028-1032.

Cornwell, D. A. 1990. Air Stripping and Aeration. *Water Quality and Treatment: a Handbook of Community Water Suplies*. F. W. A. Pontius. New York: McGraw Hill.

Cote, M., B. P. A. Grandjean, P. Lessard and J. Thibault. 1994. Dynamic Modelling of the Activated Sludge process: Improving Prediction Using Neural Networks. *Wat. Res.* 29(4): 995-1004.

Crozes, G., P. White and M. Marshall. 1995. Enhanced Coagulation: Its Effect on NOM Removal and Chemical Costs. *J. AWWA* 87(1): 78-89.

Daniell, T. M. 1991. *Neural Networks - Applications in Hydrology and Water Resources Engineering*. International Hydrology and Water Resources Symposium, Perth.

Dempsey, B. A., R. M. Ganho and C. R. O'Melia. 1984. The Coagulation of Humic Substances by Means of Aluminum Salts. *J. AWWA* 76(4): 141-150.

Dentel, S. K. 1991. Coagulant Control in Water Treatment. *Critical Reviews in Environmental Control* 21(1): 41-135.

DeSilets, L., B. Golden, Q. Wang and R. Kumar. 1992. Predicting Salinity in the Chesapeake Bay Using Backpropagation. *Computers and Operations Research* 19(3-4): 277-285.

Dobbs, R. A., R. H. Wise and R. B. Dean. 1972. The Use of Ultra-Violet Absorbance for Monitoring the Total Organic Carbon Content of Water and Wastewater. *Water Research* 6: 1173-1180.

Eaton, A. 1995. Measuring UV-Adsorbing Organics: A Standard Method. *J. AWWA* 87(2): 86-90.

Eaton, A. D., L. S. Clesceri and A. E. Greenberg, Eds. 1995. *Standard Methods for the Examination of Water and Wastewater*. Washington, DC: American Public Health Association.

Edwards, G. A. and A. Amirtharajah. 1985. Removing Color Caused by Humic Acids. *J. AWWA* 77(3): 50-57.

Edwards, M. 1997. Predicting DOC Removal During Enhanced Coagulation. *J. AWWA* 89(5): 78-89.

Edzwald, J., J. Haff and J. Boak. 1977. Polymer coagulation of Humic acid waters. *Journal of the Environmental Engineering Division of the ASCE* 103(EE6): 989-999.

Edzwald, J. K. 1993. Coagulation in Drinking Water Treatment: Particles, Organics and Coagulants. *Wat. Sci. Tech.* 27(11): 21-35.

Edzwald, J. K., W. C. Decker and K. L. Wattier. 1985. Surrogate Parameters for Monitoring Organic Matter and THM Precursors. *J. AWWA* 77(4): 122-132.

Federal-Provincial Subcommittee on Drinking Water Quality. 1995. *Summary of Guidelines for Canadian Drinking Water Quality.* Ottawa, ON: Environmental Health Directorate, Health Canada.

Flood, I. and N. Kartam. 1997. Neural Networks in Civil Engineering I : Principles and Understanding. *Journal of Computing in Civil Engineering* 8(2): 131-163.

Foody, G. M. and M. K. Arora. 1997. An Evaluation of Some Factors Affecting the Accuracy of Classification. *International-Journal-of-Remote-Sensing* 18(4): 799-810.

Garret, J. H., J. Ghaboussi and X. Wu. 1992. Neural Networks. *Expert Systems for Civil Engineers: Knowledge Representation.* R. H. Allen. New York: American Society of Civil Engineering: 104-143.

Garrett, J. H., D. J. Gunaratman and N. Ivezic. 1997. Introduction. *Artificial Neural Networks for Civil Engineers: Fundamentals and Applications.* N. Kartam, I. Flood and J. H. Garret. New York, NY: ASCE: 1-18.

Garrett, J. H., S. Ranjithan and J. W. Eheart. 1992. Application of Neural Network to Groundwater Remediation. *Expert Systems for Civil Engineers: Knowledge Representation.* R. H. Allen. New York: American Society of Civil Engineering: 259-269.

Greiner, A. D., A. Obolensky and P. C. Singer. 1992. Technical Note: Comparing Predicted and Observed Concentrations of DBPs. *J. AWWA* 84(11): 99-102.

Harrington, G. W., Z. K. Chowdhury and D. M. Owen. 1992. Developing a Computer Model to Simulate DBP Formation During Water Treatment. *J. AWWA* 84(11): 78-87.

Harston, C. T. and A. J. Maren. 1990. History and Development of Neural Networks. *Handbook of Neural Computing Applications.* A. J. Maren, C. T. Harston and R. M. Pap. San Diego, CA: Academic Press, Inc.: 13-27.

Harvey, S. and R. Harvey. 1998. An introduction to artificial intelligence. *Appita Journal. Jan* 51(1): 20-24.

Hasham, F. A., S. J. Stanley and W. B. Kindzierski. 1998. *Modelling of Urban Air Pollution in the Edmonton Strathcona Industrial Area Using Artificial Neural Networks.* Transportation, Land-Use, and Air Quality: Making the Connection, Portland, OR: ASCE.

Hecht-Nielsen, R. 1987. *Kolmogorov's Mapping Neural Network Existence Theorem.* First IEEE International Joint Conference on Neural Networks, San Diego, CA: IEEE.

Hoehn, R. C., K. L. Dixon, J. K. Malone, J. T. Novak and C. W. Randall. 1984. Biologically Induced Variations in the Nature and Removability of THM Precursors by Alum Treatment. *J. AWWA* 76(4): 134-141.

Hubel, R. E. and J. K. Edzwald. 1987. Removing Trihalomethane Precursors by Coagulation. *J. AWWA* 79(7): 98-106.

Hundt, T. R. and C. R. O'Melia. 1988. Aluminum-Fulvic Acid Interactions: Mechanisms and Applications. *J. AWWA* 80(4): 176-186.

Hushon, J.M. 1990. Overview of Environmental Expert Systems. *Expert Systems for Environmental Applications* . J.M. Hushon. ACS Symposium Series 431. New York: 1.

Hutton, P. H., N. Sandhu and F. I. Chung. 1996. *Predicting THMV Formation with Artificial Neural Networks.* Proceedings of the North American Water and Environment Conference: ASCE.

Jacangelo, J. G., J. DeMarco, D. M. Owen and S. J. Randtke. 1995. Selected Processes for Removing NOM: An Overview. *J. AWWA* 87(1): 64-77.

Jain, A. K., J. C. Mao and K. M. Mohiuddin. 1996. Artificial Neural Networks: a Tutorial. *Computer* 29(3): 31.

Joyce, W. S., F. A. DiGiano and P. C. Unden. 1984. THM Precursors in the Environment. *J. AWWA* 76(6): 102-106.

Kavanaugh, M. C. 1978. Modified Coagulation for Improved Removal of Trihalomethane precursors. *J. AWWA* 70(11): 613-620.

Kawamura, S. 1973. Coagulation Considerations. *J. AWWA* 65(6): 417-423.

Knocke, W. R., S. West and R. C. Hoehn. 1986. Effects of Low Temperature on the Removal of Trihalomethane Precursors by Coagulation. *J. AWWA* 78(4): 189-195.

Koutsakos, E. 1995. Modernisation and uprating of London's water treatment facilities. *Water Resources Management Under Drought or Water Shortage Conditions Proc EWRA*: 161-167.

Krasner, S., Z. Chowdhurdy, M. Edwards and K. Bell. 1997. *Use of SUVA in developing revised TOC removal requirements*. WQTC, Denver, Co.

Krasner, S. W. and G. Amy. 1995. Jar-Test Evaluations of Enhanced Coagulation. *J. AWWA* 87(10): 93-107.

Maier, H. R. and G. C. Dandy. 1996. The Use of Artificial Neural Networks For the Prediction of Water Quality Parameters. *Water-Resources-Research* 32(4): 1013-1022.

Malcolm Pirnie Inc. 1993. *Guidance Manual for Enhanced Coagulation and Enhanced Softening*. Mnwah, NJ: USEPA.

McCarty, P. L. and E. M. Aieta. 1984. Chemical Indicators and Surrogate Parameters in Water Treatment. *J. AWWA* 76(10): 98-106.

McCulloch, W.S. and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5: 115-132.

Miltner, M. J., S. A. Nolan and R. S. Summers. 1994. *Evaluation of Enhanced Coagulation for DBP Coagulation.* Critical Issues in Water and Wastewater Treatment: Proceedings of the 1994 National Conference on Environmental Engineering, Boulder, Co.: ASCE.

Minsky, M. L. and S. A. Papert. 1969. *Percptrons: An Introduction to Computational Geometry.* Cambridge, MA: The MIT Press.

Mirsepassi, A., B. Cathers and H. B. Dharmappa. 1995. Application of artificial neural networks to the real time operation of water treatment plants. *IEEE International Conference on Neural Networks Conference Proceedings* 1: 516-521.

Ollier, L. L., R. J. Miltner and R. S. Summers. 1997. *Microbial and Particulate Control Under Conventional and Enhanced Coagulation.* 1997 Water Quality Technology Conference, Denver, CO.: AWWA.

Owen, D. M., G. L. Amy, Z. K. Chowdhury, R. Paode, G. McCoy and K. Viscosil. 1995. NOM Characterization and Treatability. *J. AWWA* 87(1): 46-63.

Pal, S. and K. Srimani Pradip. 1996. Neurocomputing: motivation, models, and hybridization. *Computer* 29(3): 24-27.

Psichogios, D. C. and L. H. Ungar. 1991. Direct and Indirect Model Based Control Using Artificial Neural Networks. *Ind. Eng. Chem. Res.* 30: 2564-2573.

Randtke, S. J. 1988. Organic Contaminant Removal by Coagulation and Related Process Combinations. *J. AWWA* 80(5): 40-56.

Randtke, S. J. and e. al. 1994. A Comprehensive Assessment of DBP Precursor Removal by Enhanced Coagulation and Softening. *Proceedigns 1994 American Water Works Association Annual Conference.*

Reckhow, D. A. and P. C. Singer. 1990. Chlorination By-products in Drinking Waters: From Formation Potentials to Finished Water Concentrations. *J. AWWA* 82(4): 173-179.

Reid, J. M., M. S. Cresser and D. A. Macleod. 1980. Observations on the Estimation of Total Organic Carbon from U.V. Absorbance for an Unpolluted Stream. *Water Research* 14: 525-529.

Rodriguez, M. J., J. R. West, J. Powell and J. B. Serodes. 1997. Application of Two Approaches to Model Chlorine Residuals in Severn Trent Water LTD (STW) Distribution Systems. *Water Science ans Technology* 36(5): 317-324.

Rosenblatt, F. 1958. The perceptron: a probabalistic model for information storage and organization in the brain. *Psychological Review* 65: 386-408.

Rumelhart, D. E., G. E. Hinton and J. L. McClelland. 1986. A General Framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. D. E. Rumelhart, G. E. Hinton and R. J. Williams. Cambridge, MA: The MIT Press. 1: 45-76.

Rumelhart, D. E., G. E. Hinton and R. J. Williams. 1986. Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. D. E. Rumelhart, G. E. Hinton and R. J. Williams. Cambridge, MA: The MIT Press. 1: 318-362.

Schmuller, J. 1990. Neural Networks and Environmental Applications. .

Schwerk, T. 1996. Forecasting of Time Series With Neural Networks. *Z. Angew. Math. Mech* 76: 219-222.

Semmens, M. J. and T. K. Field. 1980. Coagulation: Experiences in Organics Removal. *J. AWWA* 72(8): 476-483.

Shang, Y. and B. Wah. 1996. Global Optimization for Neural Network Training. *Computer* 29(3): 45-54.

Shukla, M. B., R. Kok, S. O. Prasher, G. Clark and R. Lacroix. 1996. Use of Artificial Neural Networks in Transient Drainage Design. *Transactions-of-the-Asae* 39(1): 119-124.

Singer, P. C. 1994. Control of Disinfection By-Products in Drinking Water. *Journal of Environmental Engineering* 120(4): 727-744.

Standards and Guidelines Branch. 1997. *Standards and Guidelines for Municipal Waterworks, Wastewater and Storm Drainage Systems*. Edmonton, AB: Alberta Environmental Protection, Environmental Assessment Division.

Stanley, S. J. and Q. Zhang. 1997. *The Use of Artificial Neural Networks for Process Modelling of Enhanced Coagulation*. Water Quality Technology Conference, Denver, CO: AWWA.

Stanley, S. J., Q. Zhang and C. W. Baxter. 1998. *Use of Artificial Neural Network Models in Water Treatment*. Research and Development Conference on Rural Sanitation, Anchorage, AK: Alaska Water and Wastewater Management Association.

Tang, K. S., C. Y. Chan, K. F. Man and S. Kwong. 1995. Genetic structure for NN topology and weights optimization. .

Tate, C. H. and K. F. Arnold. 1990. Health and Asthetic Aspects of Water Quality. *Water Quality and Treatment: a Handbook of Community Water Suplies*. F. W. A. Pontius. New York: McGraw Hill.

Thomas, S. and R. Shariff. 1998. *Personal Communication.*

Villarreal, J. and P. Baffes. 1992. Time Series Prediction Using Neural Networks. *Expert Systems for Civil Engineers: Knowledge Representation.* R. H. Allen. New York: American Society of Civil Engineering: 270-284.

Vrijenhoek, E. M., A. E. Childress, T. S. Tanaka and M. D. Beuhler. 1998. Removing Particles and THM Precursors by Enhanced Coagulation. *J. AWWA* 90(4): 139-150.

Ward Systems Group Inc. 1996. *NeuroShell 2 User's Manual.* Frederick, MD: Ward Systems Group, Inc.

Weigend, A. S., B. A. Huberman and D. E. Rumelhart. 1990. Predicting the Future: A Connectionist Approach. *International Journal of Neural Systems* 1(3): 193-209.

Wilcox, S. J., D. L. hawkes, F. R. Hawkes and A. J. Guwy. 1995. A Neural Network. Based on Bicarbonate Monitoring, to Control Anaerobic Digestion. *Wat. Res.* 29(6): 1465-1470.

Williams, M., B. Badriyha, S.-C. Tu, J. Awad and M. Pirbazari. 1996. *Modified Jar Test Studies for Removal of Disinfection By-products (DBPs) adn Color Compounds from Groundwater.* North American Water and Environment Conference: ASCE.

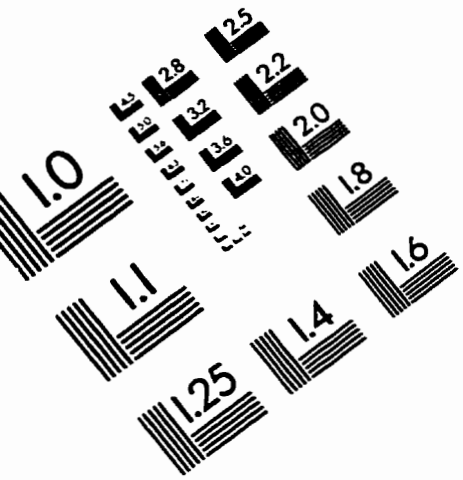
Yabunaka, K., M. Hosomi and A. Murakami. 1997. Novel Application of a Back-Propagation Artificial Neural Network Model Formulated to Predict Algal Bloom. *Water Science and Techology* 36(5): 89-97.

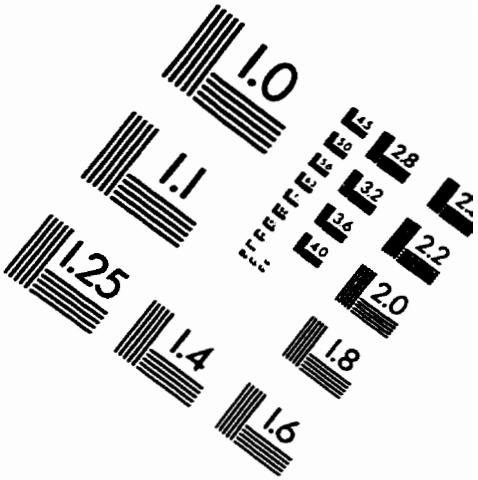
Yan, L., A. B. Rad, Y. K. Wong and H. S. Chan. 1996. *Model based control using artificial neural networks*. Proceedings of the 1996 IEEE International Symposium on Intelligent Control, Dearborn, MI: IEEE.

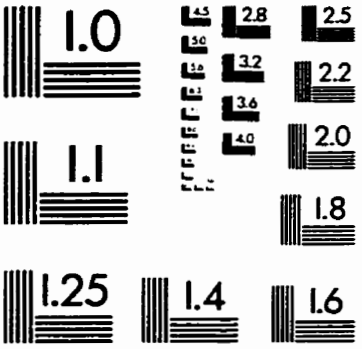
Zhang, Q. and S. J. Stanley. 1997. Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modelling. *Water Research* 31(9): 2340-2350.

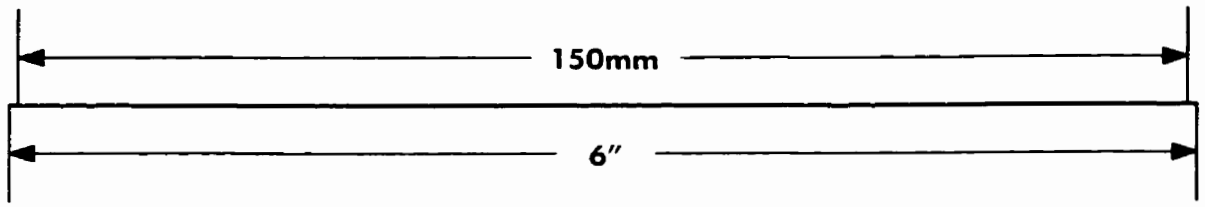
Zhang, Q. and Stanley, S.J. 1998. Use of neural networks for process modelling and control of coagulation. *Journal of Environmental Engineering, American Society of Civil Engineering*. (Accepted July).

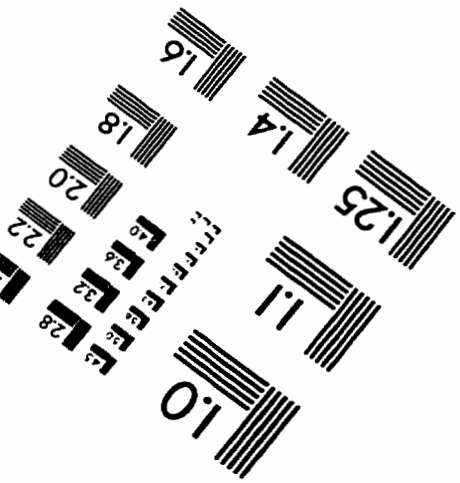
IMAGE EVALUATION
TEST TARGET (QA-3)

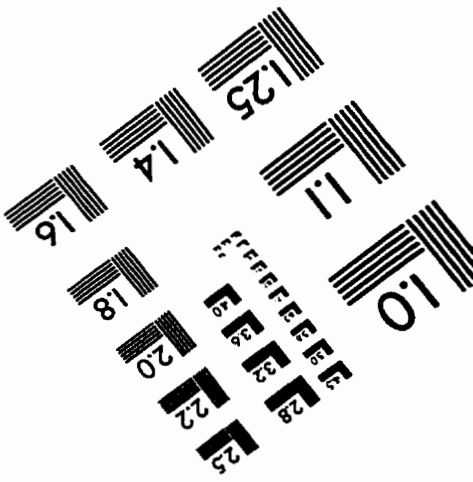












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989