

Empirical mathematical models and artificial neural networks for the determination of alum doses for treatment of southern Australian surface waters

J. van Leeuwen, C. W. K. Chow, D. Bursill and M. Drikas, *Australian Water Quality Centre, Private Mail Bag 3, Salisbury, South Australia, 5108, Australia*

ABSTRACT: The potential for predicting alum doses for surface waters from southern Australia based on physico-chemical parameters of the raw waters was studied. These parameters included dissolved organic carbon (DOC), absorbance at 254 nm, turbidity and alkalinity. Procedures used for assessing the predictability of alum dosing were empirical mathematical models and artificial neural networks.

Alum doses determined by jar tests were selected on the basis of target values for settled and filtered turbidities, colour and residual aluminium.

Regression equations which incorporated the parameters of DOC, UV absorbance (254 nm/cm), turbidity, alkalinity and pH gave correlation coefficients of greater than 0.9. These equations gave a high frequency of prediction within ± 10 mg/L alum of actual doses. Similarly, 86% of alum doses predicted by artificial neural networks were within 10 mg/L of the actual doses. Although a good prediction of coagulant dosing was achieved, it is likely that the models generated are specific for the types of waters studied and the criteria for alum dose selection.

INTRODUCTION

Coagulant dosing of surface and ground waters for drinking water purposes can be determined by jar tests, zeta potential measurement, streaming current detector (SCD), charge titration and through modelling empirical relationships that exist between coagulant dosing requirements and physico-chemical parameters of treated and/or raw water quality. An application of the above methods may be based on the specific requirements of individual water treatment facilities and on the costs of the various methods. For example, where changes in the quality of the source water are minor or occur slowly throughout the various seasons, periodic jar testing may be sufficient for the determination of coagulant doses. However, where raw water quality can change rapidly, such as in the case where the source water is from a river, other methods such as the streaming current detector or the charge titration unit [1,2] may be more suitable.

Several studies have been reported where water quality data have been used in mathematical relationships for the determination of required coagulant doses [3–6]. In this study, alum doses for 29 raw surface waters were determined by jar tests and were assessed for their relationships to physical and chemical parameters of the raw waters, including dissolved organic carbon (DOC), turbidity, colour and alkalinity.

MODELLING APPROACHES

Empirical mathematical models

These models are based on several treated and/or raw water

parameters and include polynomial [3], matrix factorial fractional and Doehlert lattice [5] and multivariate calibrations [4,6]. These models can involve a relatively large number of parameters (e.g. eight parameters) and their apparent interactions [5]. Bazer-Bachi *et al.* [3] developed two models based on polynomial equations that were used to determine coagulant feed rates for the Clairfont water treatment plant in France. One of the models, termed CALME, comprised of four variables (turbidity, resistivity, temperature and organic material) for waters of low turbidity (<20 NTU), while the CRUES model incorporated a fifth parameter (the ratio between colloidal titration and turbidity) for higher turbidity waters. The model developed by Girou *et al.* [5] was based on the concentration of calcium ions, bicarbonate ions, sulphate ions, humic acids, initial turbidity, temperature and pH. The parameters used in the model developed by Ratnaweera & Blom [6] are flow, sedimentation time, temperature, turbidity (influent and effluent water), orthophosphate concentration (influent), pH (influent) and conductivity (influent).

Ratnaweera & Blom [6] did not include a measure of the natural organic matter (NOM), and their model requires a measure of the coagulation/flocculation performance, i.e. treated water turbidity. If it is assumed that turbidity results from the coagulation/flocculation process involving NOM, then an adequate removal of treated water turbidity indirectly accounts for the removal of NOM. However, when rapid changes occur in raw water quality and dosing is dependent on the quality of previously treated water, then optimum dosing may not be consistently achieved during the period of coagulant adjustment.

The models vary in their complexity, with most incorporating NOM and turbidity. Some models include the use of alkalinity [5] or resistivity/conductivity [3,6].

Variation in the parameters used in these models may be due to the following:

- 1 some raw water parameters in some waters do not vary significantly;
- 2 some parameters exert much higher demands on coagulant use than others and the minor influencing parameters may or may not be included in the models;
- 3 some parameters that influence coagulant demand are recognised while others are not; and
- 4 minor parameters are not included due to their difficulty or cost for routine analysis.

In some circumstances, some parameters may have a good correlation with other more directly relevant but unrecognised parameters, e.g. nutrients correlating with organic levels in a particular catchment. This could lead to an apparent relationship between the measured parameter and coagulant demand that would not necessarily occur in another location.

Critchley *et al.* [7] described the assessment of coagulant dose based on the measurements of raw water quality parameters where relationships had been established, as a feed forward control. They described a feedback control as one which adjusts coagulant dosing based on the treated water quality.

A computer-based system for coagulant dosing based on feed forward and feed back control and using measurements of colour, turbidity, coagulant residual and other parameters was reported by Critchley *et al.* [7]. An equation used in this control process was as follows:

$$\text{Coagulant dose} = k_1 \text{ colloidal colour} + k_2 \text{ raw water turbidity} + k_3,$$

where k_1 , k_2 and k_3 are constants, and k_3 represents the pH, conductivity and influence of temperature.

According to Critchley *et al.* [7], the pH variable in the above algorithm is of secondary importance, which is contrary to the concept of correcting the pH to optimise flocculation. Consequently, these authors advocated the validation of the above chemical relationship for some waters.

Artificial neural networks

Artificial neural networks (ANN) are information processing systems that have been used for making predictions based upon their perceived influencing factors. Various applications of this technique have been reported [8–16] and a number of review articles regarding the use of artificial neural networks in chemistry have been published [17–19]. Mirsepassi *et al.* [20,21] have already evaluated the capacity of ANN for the prediction of alum doses at a water treatment facility in New South Wales, Australia. From a practical point of view, artificial neural networks represent a conceptually different approach to mathematical modelling, as the user is not required

to develop an application-specific algorithm. Instead, examples of previous outcomes and their perceived influencing factors are used to ‘teach’ the network. From this trained network, events may be predicted based on those input factors. This approach avoids the need to determine theoretical and empirical relationships between the influencing factors and the outcomes of interest. However, it would still be a requirement to search for and identify the important input factors, i.e. those that drive the coagulation/flocculation processes.

When approaching a task requiring the use of conventional computing techniques, an algorithm-based program that operates serially is needed. Information is stored at addressed locations in memory and is processed sequentially in a rather complex central processing unit. In contrast, artificial neural networks process information in parallel structures and the memory is of a distributed and inherently associative form. A large number of simple processing elements are employed with high interconnectivity between these elements. The processing unit computes a function of the weighted summation of all input signals. A given processing element’s output connects via specific weights to the inputs of many other processing elements. A number of such highly interconnected processing elements form the structure of a neural network. As the processor experiences its environment, it adapts through adjustment of the weights connecting processing elements. The altered state of the network through weight adaptation constitutes a form of ‘learning’. Explanations of this process have been described by Zupan & Gasteiger [17], Sperring *et al.* [11] and Spining *et al.* [19].

In most applications of ANN, the backward error propagation algorithm is used to train the network. The basic structure of this algorithm is formed by (i) entering the specific inputs and outputs, (ii) comparing the determined output with the actual output and calculating a quantitative error, and (iii) iteratively minimising the error by adjusting the weights of the connections in the network. To minimise the error, it is usual to begin at the output nodes and adjust their weights. Backward propagation to the hidden layer(s) adjacent to the output layer and recalculation of the errors are then performed, with weights adjusted accordingly. This step is repeated until the input layer is reached. The number of processing elements (nodes) in the hidden layer is selected based on achieving minimum differences between output and target values and a satisfactory processing time for the network.

In this study, raw water parameters of 40 samples were used in an attempt to predict alum doses that had been determined by jar tests. Predictions were made using empirical linear and multi-regression models and artificial neural networks.

MATERIALS AND METHODS

Collection of water samples

Water samples were collected from various locations in South

Table 1 Raw water parameters and alum doses determined for water samples by jar tests (ADJT) and used to predict doses by cross validation

No.	Water source	Turbidity	pH (NTU)	Colour (HU)	UV abs. @254 nm	Alkalinity as bicarbonate (mg/L)	DOC (mg/L)	ADJT (mg/L)
1	Anakie Basin	3.5	7.51	13	0.226	38	8.7	25
2	Anakie Basin	3.6	7.10	34	0.265	27	8.2	25
3	Hope Valley Reservoir	1.55	8.07	8	0.081	95	4.2	7.5
4	Little Para Creek	73	7.92	42	0.174	155	4.6	50
5	Little Para Reservoir	3.1	8.45	7	0.078	130	4.0	20
6	Mt. Bold Reservoir	7.0	8.09	8	0.068	72	4.1	15
7	Murray River	51	7.55	29	0.215	55	6.6	50
8	Murray River	47	7.48	26	0.143	53	4.3	40
9	Murray River	46	7.71	16	0.097	55	3.0	25
10	Murray River	30	7.97	24	0.144	65	4.7	30
11	Murray River	5.0	7.98	27	0.154	67	5.2	20
12	Murray River	0.33	7.80	18	0.157	65	5.1	15
13	Murray River	7.8	8.27	9	0.083	74	4.3	20
14	Myponga Reservoir	2.8	8.04	32	0.290	77	8.1	50
15	Myponga Reservoir	3.0	7.86	82	0.491	62	11.4	65
16	Myponga Reservoir	4.1	7.75	77	0.318	61	7.7	35
17	Myponga Reservoir	4.1	7.60	34	0.206	59	5.0	20
18	Onkaparinga River	7.3	7.48	34	0.205	96	6.3	40
19	Salisbury Wetlands inlet	104	7.98	43	0.164	82	3.8	45
20	Salisbury Wetlands outlet	71	8.13	18	0.142	123	3.9	40
21	South Para Reservoir	1.2	8.00	16	0.215	33	7.7	20
22	South Para Reservoir	2.9	8.11	53	0.373	57	9.5	45
23	South Para Reservoir	1.59	8.02	56	0.383	54	10.3	45
24	South Para Reservoir	2.1	7.84	37	0.254	55	7.8	25
25	South Para Reservoir	2.0	7.96	26	0.170	53	5.8	20
26	Tod Reservoir	5.8	8.63	19	0.314	206	12.3	80
27	Torrens River	10.6	7.79	19	0.181	150	4.2	35
28	Warren Reservoir	6.0	7.61	7	0.083	59	4.7	15
29	Wurdee Boluc Reservoir	2.1	7.48	18	0.090	84	4.4	10

Australia and Victoria in 10 L or 20 L plastic containers. The collected samples were transported to the laboratory within two days and stored at 4 °C.

Water samples

Forty water samples were obtained in 1994 and 1996 from 19 different sources, comprising a creek, rivers, wetlands and reservoirs in southern Australia from the states of Victoria and South Australia. Lists of water samples and their raw water parameters are shown in Tables 1 and 2.

Several of the above waters were altered prior to jar testing in order to vary parameters that had an apparent influence on alum dosing.

A Murray River water sample with a turbidity of 30 NTU (no. 10) was filtered (0.2 µm) resulting in a turbidity of 0.33 NTU (no. 12). A sub-sample of the filtered water was then mixed with raw water giving a turbidity of 5.0 NTU (no. 11).

The DOC concentrations of samples from the Murray River,

Myponga Reservoir and South Para Reservoir were altered by use of a magnetic ion exchange resin (MIEX™ ORICA, Australia). Natural organic matter was partially removed by slowly stirring the water samples to the resin for 30 min with slow mixing. Sub-samples (20 L) of South Para Reservoir water were treated with 0 (no. 23), 14 cm³ (no. 24) and 20 cm³ (no. 25) resin; Myponga Reservoir, 0 (no. 15), 25 cm³ (no. 16) and 50 cm³ (no. 17) resin; and Murray River, 0 (no. 7), 30 cm³ (no. 8) and 60 cm³ (no. 9). The alkalinities of water samples from Anakie Basin (no. 30), Millbrook Reservoir (no. 37) and South Para Reservoir (no. 39) were increased by about 50 mg/L (as bicarbonate) and samples from Anakie Basin (no. 31) and Millbrook (no. 38) were increased by about 100 mg/L (Table 2).

Coagulant dose determination by jar tests

Jar tests were performed using an instrument with six stirrers (Scientific Equipment Manufacturers Pty Ltd, South Australia). Throughout this study, flash mixing was performed at

Table 2 Raw water parameters and alum doses determined for water samples by jar tests (ADJT). For these waters, predictions of alum doses were based on the data of water samples listed in Table 1

No.	Water source	Turbidity (NTU)	pH	Colour (HU)	UV abs. @254 nm	Alkalinity as bicarbonate (mg/L)	DOC (mg/L)	ADJT (mg/L)
30	Anakie Basin	3.2	7.91	36	0.252	75	8.4	35
31	Anakie Basin	4.8	8.37	31	0.248	124	8.1	40
32	Barossa Reservoir	1.8	7.28	76	0.486	54	12.3	60
33	Blue Lake	1.2	8.38	3	0.005	169	0.56	5
34	Happy Valley Reservoir	1.9	8.19	13	0.094	76	4.17	12.5
35	Millbrook Reservoir	3.5	7.34	97	0.554	57	13.9	80
36	Millbrook Reservoir	3.7	7.76	65	0.421	64	10.5	60
37	Millbrook Reservoir	6.8	7.96	59	0.399	116	10.4	80
38	Millbrook Reservoir	5.8	7.86	54	0.385	165	10.2	100
39	South Para Reservoir	2.9	7.98	52	0.363	105	9.6	60
40	Upper Paskerville Reservoir	24.0	7.54	1	0.059	58	4.3	20

≈ 160 r.p.m. for 1 min, slow mixing at 28 r.p.m. for 14 min and settling for 15 min

The coagulant used throughout this study was $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ and pH levels were not controlled prior to or after the addition of alum. Parameters that were determined for the treated water included pH, unfiltered and filtered (filter No. 1, Cat. no. 1001240, Whatman) turbidities, colour (Hazen Units), UV absorbance (254 nm/cm) and residual aluminium. Additional parameters that were determined for raw water included alkalinity and dissolved organic carbon. The water temperatures were recorded at the time of the jar testing and ranged from about 10 °C to 25 °C. For this study, the required alum dosing was based on target values for the following treated water parameters: unfiltered turbidity of 3 NTU, filtered turbidity of 0.2 NTU, colour of 10 Hazen units and residual aluminium less than 0.2 mg/L. The attainment of a filtered turbidity of 0.2 NTU was mostly at or close to the optimum for this parameter. For some waters, the attainment of all of the target parameters was not possible, as increased dosing for the achievement of a target value for one parameter led to the deterioration of other parameters. In this study, a greater value was placed on the achievement of the target values for turbidities and the residual aluminium.

Measurement of UV absorbance (254 nm/cm)

Absorbances of filtered water samples (0.45 µm, cellulose nitrate filters, Sartorius, Germany) were measured using a Varian DMS 70, UV/visible Spectrophotometer.

Turbidity

Turbidities were determined using a Hach Ratio Turbidimeter (range 2–200 NTU), Model 18900-70.

Dissolved organic carbon (DOC)

Analysis for dissolved organic carbon was performed using a Skalar SK12 analyser (Skalar Analytical B.V., Breda, The Netherlands) after the samples were filtered through 0.45 µm cellulose nitrate filters (Sartorius, Germany). Samples were automatically acidified and sparged with nitrogen to remove inorganic and volatile organic carbon and organic carbon was oxidised by persulphate/UV treatment to carbon dioxide. The carbon dioxide was reduced to methane over a nickel catalyst at 400 °C and measured by flame ionisation detector.

Colour

The colour of the natural waters was determined by measuring the absorbance at 456 nm [22] using a Varian DMS 70, UV/visible spectrophotometer after filtration (0.45 µm cellulose nitrate, Sartorius, Germany).

Alkalinity

Alkalinities were determined by titration [23].

Aluminium, iron and manganese

Aluminium, iron and manganese were determined using inductively coupled plasma atomic emission spectroscopy method [23].

Modelling

Two approaches, regression analyses and artificial neural networks were used to predict alum doses for surface waters based on the raw water parameters.

Empirical mathematical modelling (linear and multi-regression analyses)

Raw water parameters were tested individually and in combination for their linear relationships to alum doses. The parameters tested were either transformed to their logarithmic values or unaltered. Logarithm transformation of some raw water parameters (turbidity, alkalinity and colour) was carried out in order to reduce their relative weighting compared with the other parameters (absorbance at 254 nm/cm, pH and DOC). Linear and multi-regression analyses were performed using ORIGIN™ version 4.1 (Microcal™ Software, Inc.). Predictions of alum doses for waters listed in Table 1 were made by cross-validation where the maximum data was used to make each prediction (i.e. 28 water samples). Predictions of alum doses for water samples listed in Table 2 were based on models developed from the data of waters listed in Table 1.

Artificial neural network

The artificial neural network (ANN) software used for this study is a multilayered neural network structure with backward error propagation (Back Error Propagation Simulator, BPS [24]). The processing elements are structured into three layers, and signal propagation is unidirectional. An IBM compatible Pentium™ 100 MHz personal computer (PC) was used to run this software.

Training was performed using the full data set with the parameters of turbidity, pH, colour, UV absorbance at 254 nm, alkalinity and DOC as input variables and alum doses determined by the jar test as the desired output variables. An illustration of the topology of the neural network used is shown in Fig. 1. The network first uses the input vector (turbidity, pH, colour, UV absorbance at 254 nm, alkalinity and DOC) to produce its own output vector via a forward pass. The activity of each layer is computed in turn using the already computed activity levels in the earlier layers. The final layer output is then compared with the desired output. Here we are concerned with only the single output of alum concentration.

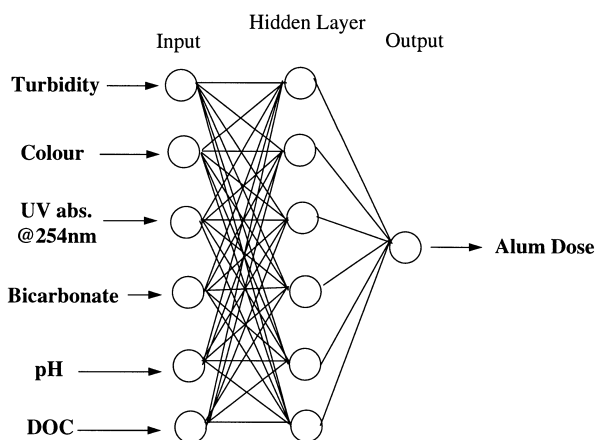


Fig. 1 The topology of the neural network used in this study. Each circle denotes a processing unit.

Prediction of alum doses for waters by cross validation

After the optimisation of the neural network, the 29 raw water parameters (Table 1) were used to predict alum doses by cross validation, as follows:

- (i) Training was performed using the parameters of 28 raw waters and predicting the alum dose for the excluded water. This procedure was repeated for all of the 29 water samples.
- (ii) Training and prediction with five samples randomly excluded.

Five water samples were randomly selected and the remaining 24 water samples were used as a training set. The procedures were repeated three times.

- (iii) Training and prediction of alum doses of waters based on their source.

In this case, all water samples collected from a particular location were removed from the training set and parameters of the remaining waters were used to predict the alum doses for the removed water samples. This procedure was used for waters from Myponga Reservoir, South Para Reservoir and Murray River.

Prediction of alum doses for waters not included in training of the ANN

Training was performed using the parameters of 29 water samples (Table 1) and predicting the doses for waters (Table 2) not included in any training of the ANN.

RESULTS

Selection criteria for alum dose

Values of the various parameters of the 29 water samples (Table 1) following alum treatment are shown in Table 3. For several waters, the selected dose resulted in some but not all of the target values being achieved. For example, the minimum unfiltered turbidity recorded for the sample from Little Para Creek over an alum dose range of 20–100 mg/L was 9.2 (at 50 mg/L). Although the colour was high (20 HU) at 50 mg/L, it remained mostly unchanged from 20 to 55 mg/L alum. At 60 mg/L the colour dropped to 15 HU but the unfiltered turbidity had increased to 9.6 NTU. At an alum dose rate of 100 mg/L, the colour dropped further to 10 HU but the unfiltered turbidity had increased to 15.3 NTU. Therefore, attainment of all target values for the various treated water parameters was not possible for some waters.

Empirical mathematical modelling

Correlation coefficients of linear relationships between raw water parameters (absorbance at 254 nm, turbidity, alkalinity, pH, colour and DOC) with alum doses are shown in Table 4.

Using the parameters of absorbance at 254 nm/cm; log (turbidity × 10); log (bicarbonate) and pH, a high correla-

Table 3 Treated water parameters at alum doses determined by jar tests (ADJT) and optimal for the selection criteria

No.	Water source	Unfiltered turbidity (NTU)	Filtered turbidity (NTU)	Colour (HU)	UV abs. @254 nm	pH	Residual aluminium (mg/L)	ADJT (mg/L)
1	Anakie Basin	2.8	0.13	4	0.134	6.89	0.047	25
2	Anakie Basin	1.9	0.06	7	0.116	6.82	0.030	25
3	Hope Valley Reservoir	1.69	0.15	5	0.067	7.72	0.185	7.5
4	Little Para Creek	9.2	0.31	21	0.090	7.84	0.065	50
5	Little Para Reservoir	3.7	0.06	3	0.057	8.08	0.147	20
6	Mt. Bold Reservoir	2.9	0.08	3	0.052	7.82	0.082	15
7	Murray River	4.2	0.09	6	0.081	6.72	0.016	50
8	Murray River	2.0	0.12	4	0.052	6.89	0.020	40
9	Murray River	2.9	0.09	2	0.036	7.55	0.037	25
10	Murray River	2.4	0.20	8	0.085	7.40	0.047	30
11	Murray River	2.2	0.16	10	0.100	7.29	0.079	20
12	Murray River	0.59	0.10	10	0.107	7.43	0.076	15
13	Murray River	1.27	0.13	6	0.060	7.74	0.179	20
14	Myponga Reservoir	4.0	0.07	9	0.130	7.35	0.027	50
15	Myponga Reservoir	3.2	0.09	10	0.137	7.10	0.026	65
16	Myponga Reservoir	2.6	0.05	10	0.115	7.23	0.042	35
17	Myponga Reservoir	2.4	0.08	4	0.066	7.40	0.069	20
18	Onkaparinga River	3.4	0.11	8	0.106	7.04	0.032	40
19	Salisbury Wetlands inlet	4.5	0.15	5	0.055	7.88	0.037	45
20	Salisbury Wetlands outlet	5.5	0.18	5	0.077	7.58	0.034	40
21	South Para Reservoir	1.9	0.23	7	0.169	7.62	0.195	20
22	South Para Reservoir	3.2	0.12	11	0.157	7.00	0.047	45
23	South Para Reservoir	2.0	0.09	12	0.148	7.22	0.047	45
24	South Para Reservoir	3.0	0.08	9	0.116	7.40	0.068	25
25	South Para Reservoir	1.37	0.10	5	0.078	7.61	0.059	20
26	Tod Reservoir	2.8	0.07	10	0.211	7.44	0.075	80
27	Torrens River	3.2	0.20	8	0.133	7.54	0.065	35
28	Warren Reservoir	3.5	0.08	1	0.059	7.23	0.034	15
29	Wurdee Boluc Reservoir	2.2	0.13	4	0.054	7.42	0.082	10

Table 4 Correlation coefficients of linear regression models of alum doses (mg/L) determined by a jar test procedure and selected raw water parameters. Models were based on empirical relationships between raw water parameters and alum doses determined by the jar test for 29 water samples

Model no.	Calculated doses =	r
1	$(\log(\text{turbidity} \times 10) \times \text{pH} \times \log(\text{alkalinity}) \times \text{UV}(254 \text{ nm/cm}) \times \log(\text{colour} \times 10) \times \text{DOC} + 48.3)/4.21$	0.8610
2	$(\log(\text{turbidity} \times 10) \times \text{pH} \times \log(\text{alkalinity}) \times \text{UV}(254 \text{ nm/cm}) \times \log(\text{colour} \times 10) + 0.53)/0.4$	0.9216
3	$(\log(\text{turbidity} \times 10) \times \text{pH} \times \log(\text{alkalinity}) \times \text{UV}(254 \text{ nm/cm}) - 0.21)/0.149$	0.9528
4	$(\log(\text{turbidity} \times 10) \times \log(\text{alkalinity}) \times \text{UV}(254 \text{ nm/cm}) - 0.051)/0.0181$	0.9397
5	$(\text{turbidity} \times \text{pH} \times \text{alkalinity} \times \text{UV}(254 \text{ nm/cm}) \times \text{colour} \times \text{DOC} + 298568)/20366$	0.5413
6	$(\text{turbidity} \times \text{pH} \times \text{alkalinity} \times \text{UV}(254 \text{ nm/cm}) \times \text{colour} + 40051)/3382$	0.3960
7	$(\text{turbidity} \times \text{pH} \times \text{alkalinity} \times \text{UV}(254 \text{ nm/cm}) + 932)/93.8$	0.4225

tion coefficient (0.953) was obtained. The standard deviation of the differences between the calculated and actual doses was 5.4 mg/L. Inclusion of the DOC and log (colour × 10) parameters did not increase correlation coefficients nor reduce standard deviations of differences between calculated and actual doses, i.e. 10 mg/L for the model with DOC and log (colour × 10) and 7.2 mg/L with log (colour × 10). Logarithm transformation of the alkalinity, colour and turbidity values resulted in comparatively high correlation coefficients while untransformed data gave low correlation coefficients, indicating the necessity for desensitising the absolute values of these parameters. Raw water colour and turbidity values varied from less than 10 NTU to greater than 100 NTU, and were multiplied by 10 to ensure that the subsequent logarithm transformed values were 1 or greater.

Linear regression equations incorporating pH, absorbance at 254 nm/cm, log (alkalinity) and log (turbidity (NTU) × 10) were used to assess the predictive capacity of this modelling approach by cross validation. In this case, the raw water

parameters of 28 samples were used to predict the alum dose for the excluded sample. The predicted alum doses using a linear regression equation (with the predictions of alum doses found by other methods) and the doses determined by the jar test are listed in Table 5. In most cases the differences between the predicted and actual doses were less than 10 mg/L, though for one water sample (no. 12) this level of variance represented a high percentage difference.

As the minimum alum dose increment in the jar test procedure used was generally 5 mg/L, variation or error in the selection of an actual dose could be expected at the magnitude of this increment size, e.g. where slightly higher dosing results in fluctuations at or near the target levels for the turbidity of treated water.

Standard deviations of the differences between alum doses determined by the jar test and by multiple regression models (MRM) using the complete data set are given in Table 6.

As found with linear regression models, improved multi-regression equations were obtained when some of the para-

Table 5 Comparison of alum doses determined by the jar test (ADJT) with predicted doses using cross validation and raw water parameters of 28 samples

No.	Water source	ADJT (mg/L)	Predicted doses (mg/L)		
			Linear	Multi-regression*	ANN
1	Anakie Basin	25	26	33	24
2	Anakie Basin	25	27	28	19
3	Hope Valley Reservoir	7.5	9	11	15
4	Little Para Creek	50	57	54	42
5	Little Para Reservoir	20	12	18	15
6	Mt. Bold Reservoir	15	11	17	14
7	Murray River	50	50	46	65
8	Murray River	40	31	30	30
9	Murray River	25	22	24	29
10	Murray River	30	33	33	40
11	Murray River	20	24	23	19
12	Murray River	15	5	9	18
13	Murray River	20	15	19	14
14	Myponga Reservoir	50	41	39	43
15	Myponga Reservoir	65	68	63	60
16	Myponga Reservoir	35	47	41	34
17	Myponga Reservoir	20	29	22	20
18	Onkaparinga River	40	37	38	34
19	Salisbury Wetlands Inlet	45	50	42	53
20	Salisbury Wetlands Outlet	40	45	45	50
21	South Para Reservoir	20	17	16	20
22	South Para Reservoir	45	51	48	46
23	South Para Reservoir	45	41	46	46
24	South Para Reservoir	25	30	31	29
25	South Para Reservoir	20	19	16	16
26	Tod Reservoir	80	70	71	76
27	Torrens River	35	41	37	36
28	Warren Reservoir	15	12	16	10
29	Wurdee Boluc Reservoir	10	10	12	12

*Multiple regression four parameters used.

Table 6 Standard deviations of differences in alum doses (mg/L) determined by the jar test and those calculated from multiple regression models. Models were based on empirical relationships between selected raw water parameters and alum doses determined by the jar test for 29 water samples

Model no.	Calculated doses =	SD (mg/L)
1	$15.53 \times \log(\text{turbidity} \times 10) + 0.5576 \times \text{pH} + 32.14 \times \log(\text{alkalinity}) + 71.01 \times \text{UV}(254 \text{ nm/cm}) - 0.3925 \times \log(\text{colour} \times 10) + 3.5132 \times \text{DOC} - 94.41$	3.95
2	$14.70 \times \log(\text{turbidity} \times 10) - 1.141 \times \text{pH} + 35.62 \times \log(\text{alkalinity}) + 178.71 \times \text{UV}(254 \text{ nm/cm}) - 15.35 \times \log(\text{colour} \times 10) - 50$	4.77
3	$12.89 \times \log(\text{turbidity} \times 10) + 3.02 \times \text{pH} + 32.73 \times \log(\text{alkalinity}) + 140.85 \times \text{UV}(254 \text{ nm/cm}) - 102.9$	5.40
4	$17.38 \times \log(\text{turbidity} \times 10) + 30.88 \times \log(\text{alkalinity}) + 59.26 \times \text{UV}(254 \text{ nm/cm}) + 4.19 \times \text{DOC} - 94.25$	3.82
5	$13.54 \times \log(\text{turbidity} \times 10) + 32.85 \times \log(\text{alkalinity}) + 143.24 \times \text{UV}(254 \text{ nm/cm}) - 81.61$	5.61
4	$0.3193 \times \text{turbidity} - 0.4677 \times \text{pH} + 0.1675 \times \text{alkalinity} + 110.3 \times \text{UV}(254 \text{ nm/cm}) - 0.116 \times \text{colour} + 2.133 \times \text{DOC} - 14.44$	5.54
5	$0.2907 \times \text{turbidity} - 1.308 \times \text{pH} + 0.1864 \times \text{alkalinity} + 174.9 \times \text{UV}(254 \text{ nm/cm}) - 0.2540 \times \text{colour}$	5.65
6	$0.2461 \times \text{turbidity} + 0.1998 \times \text{pH} + 0.1947 \times \text{alkalinity} + 134.5 \times \text{UV}(254 \text{ nm/cm})$	6.15

eters were transformed to their logarithm values. Models were obtained using ORIGIN™ version 4.1 software and regression coefficients were determined for each parameter, either transformed or untransformed. This software provided calculations of errors for the various models and *t*-values for regression coefficients and the individual regression coefficients were tested for their significance. A model incorporating the raw water parameters without pH and log (colour × 10) gave the lowest standard deviation of the differences between the calculated and actual alum doses (Table 6).

The predictive capacity of a MRM utilising the four selected parameters was determined by cross validation where equations were generated for each water sample based on the parameters of the other 28 raw waters. Alum doses predicted for each water sample are also shown in Table 5.

Artificial neural network

For prediction of alum doses using neural networks, the number of nodes used in the hidden layer was based on a minimal mean squared error (MSE) generated when the full data set of Table 1 was used, i.e.

$$\text{MSE} = \frac{1}{2} \sum (\text{target} - \text{output})^2 \tag{1}$$

The MSE generated for various numbers of nodes in the hidden layer after 32000 iterations is shown in Fig. 2. For this data set, a network with six hidden nodes (*n* = 6) resulted in a minimal MSE. A comparison of the actual alum dose with the neural network output dose using the full data set is shown in Fig. 3.

The ability of the neural network to predict alum doses was also determined by cross validation where one of the 29 water samples was removed from the training set (Table 5); randomly removing five samples from the training set (Table 7) and removing all samples collected from a particular location (Table 8). Where the data of 28 water samples were used to predict the alum dose for the remaining water sample, only one predicted dose differed from the actual dose by more than 10 mg/L (no. 7, Murray River). In the three trials where five randomly selected samples were removed, all predicted doses except one were within 10 mg/L of the actual doses. Similarly, where doses were predicted for samples based on location only one prediction exceeded the actual dose by 10 mg/L.

In this study, sensitivity analysis was not performed and all available parameters were therefore used in the ANN.

A comparison of the three types of procedures used for prediction (by cross validation) of alum doses is shown in Table 9. Based on the standard deviations of the differences between the predicted (by cross validation) and actual alum doses the best model of those tested is a multi-regression model using the parameters of turbidity, alkalinity, UV (254 nm/cm) and DOC. However, the standard deviations were similar for the three procedures used indicating that all could be utilised in

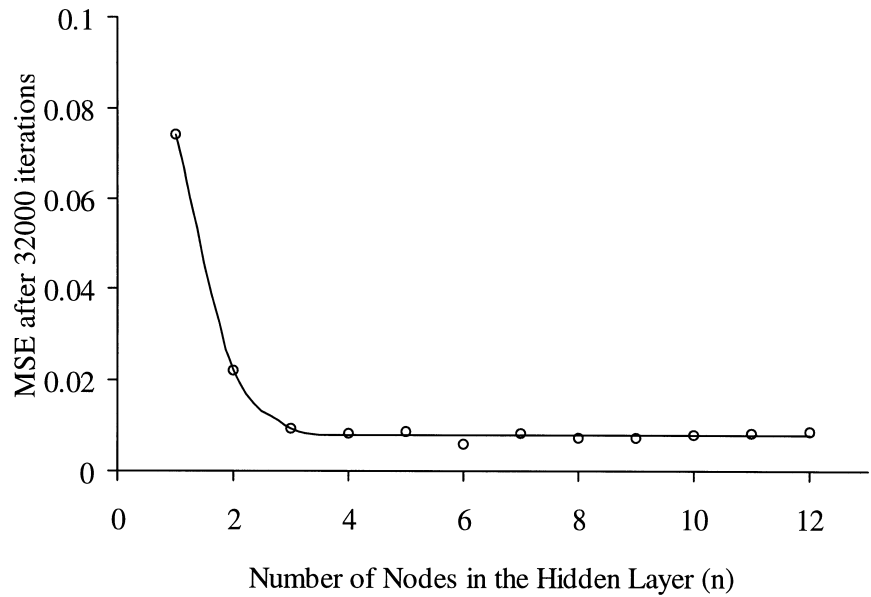


Fig. 2 Mean square errors (after 32 000 iterations) for hidden layer nodes using the complete data set ($n = 29$) and all available raw water parameters.

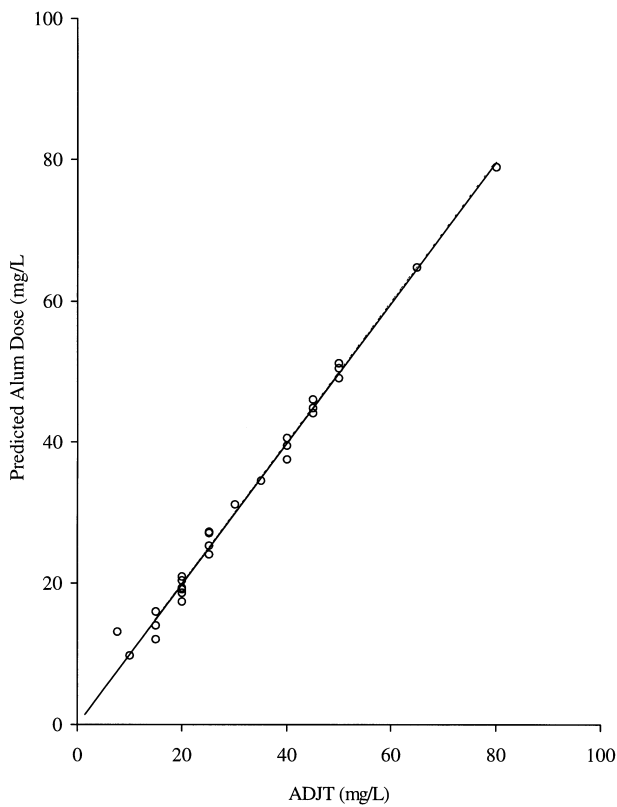


Fig. 3 Alum doses determined by ANN using all of the data for training.

Table 7 Comparison of alum doses determined by jar tests (ADJT) with those predicted using cross validation and raw water parameters of five randomly selected samples excluded

Water source	ADJT (mg/L)	Predicted alum dose (mg/L)
Trial 1		
5 Little Para Reservoir	20	19
10 Murray River	30	42
15 Myponga Reservoir	65	56
18 Onkaparinga River	40	41
24 Sth Para Reservoir	25	30
Trial 2		
4 Little Para Creek	50	49
6 Mt. Bold Reservoir	15	12
18 Onkaparinga River	40	39
27 Torrens River	35	45
28 Warren Reservoir	15	10
Trial 3		
1 Anakie Basin	25	25
2 Anakie Basin	25	22
4 Little Para Creek	50	55
26 Tod Reservoir	80	75
29 Wurdee Boluc Res.	10	10

Table 8 Comparison of alum doses determined by the jar test with those predicted using cross validation and raw water parameters excluded based on the location of the water source

Water source	ADJT (mg/L)	Predicted alum dose (mg/L)
Trial 1		
14 Myponga Reservoir	50	41
15 Myponga Reservoir	65	65
16 Myponga Reservoir	35	39
17 Myponga Reservoir	20	18
Trial 2		
21 South Para Res.	20	20
22 South Para Res.	45	50
23 South Para Res.	45	51
24 South Para Res.	25	30
25 South Para Res.	20	17
Trial 3		
7 Murray River	50	48
8 Murray River	40	26
9 Murray River	25	18
10 Murray River	30	28
11 Murray River	20	21
12 Murray River	15	19
13 Murray River	20	18

determination of alum dosing for the waters tested if the prediction tolerances are accepted.

When the data of the 29 raw water samples (Table 1) were used to predict the doses for the water samples listed in Table 2 by ANN, one prediction (Anakie Basin, no. 31) differed markedly (10 mg/L) from the actual dose (Table 10). Similarly, when the optima regression models determined from the data of Table 1 were used to predict the alum doses for the remaining 11 water samples (Table 2) one predicted dose was clearly different from the actual dose, i.e. Millbrook Reservoir (no. 38). However, the predictions for the other water samples from this source (nos 35–37) were generally comparable with the actual doses.

DISCUSSION

Factors affecting predictability of coagulant dosing

Although the predictions of alum doses were generally similar to the actual doses determined, it is likely that the models adopted are relevant only for the waters on which the models were based. According to Kawamura [25] proper dosage of coagulants such as alum, ferric salts and cationic polymers are site specific and there is no reliable formula to determine the effective dosage for waters. According to that author the most

Table 9 Standard deviations of differences between actual alum doses (ADJT) and the predicted doses obtained using the determined models

Model	SD (mg/L)
1 Linear: log (turbidity \times 10); pH; log (alkalinity); UV (254 nm/cm)	5.9
2 Multi-regression: log (turbidity \times 10); pH; log (alkalinity); UV (254 nm/cm); log (colour \times 10); DOC	5.5
3 Multi-regression: log (turbidity \times 10); log (alkalinity); UV (254 nm/cm); DOC	4.7
4 ANN, using all parameters described and untransformed	6.2

No.	Water source	ADJT (mg/L)	Predicted doses (mg/L)		
			Linear	multi-regression*	ANN
30	Anakie Basin	35	36	40	39
31	Anakie Basin	40	46	48	57
32	Baroosa Reservoir	60	49	61	67
33	Blue Lake	5	2	0	5
34	Happy Valley Reservoir	12.5	13	9	15
35	Millbrook Reservoir	80	69	78	71
36	Millbrook Reservoir	60	58	58	61
37	Millbrook Reservoir	80	75	69	81
38	Millbrook Reservoir	100	74	70	89
39	South Para Reservoir	60	53	55	71
40	Upper Paskerville Reservoir	20	13	23	16

Table 10 Comparison of alum doses determined by the jar test (ADJT) with predicted doses for samples not included for model building or training of ANN

*Multiple regression four parameters used.

reliable method to determine both the effective type of coagulant and their proper dosage is the bench scale test (jar test).

All waters used in this study were taken from locations in south Australia where the sources of natural organic matter may be from similar vegetation types or other biological sources. Natural organic matter is an important component of raw water when treating waters with coagulants for drinking purposes, with some studies showing a high correlation between organic matter content and coagulant requirements [26,27]. Fractions of natural organic matter (NOM) which are more readily removed by coagulation are the large molecular weight and humic (hydrophobic) material. Waters with a dominance of nonhumic or lower molecular weight material are recalcitrant to NOM removal by coagulation [28]. Where natural organic matter varies markedly, as might be expected in different geographical locations, it would appear that extrapolation of these models in a generalised context is not possible. However, if components or the character of NOM that influences coagulant demand could be readily analysed, it would then seem that a generalised model/ANN is possible.

The control of pH was not needed for waters used in this study as they were adequately buffered for treatment at the alum doses required. For these waters, pH was not found to be a significant parameter in multi-regression modelling and marginally improved the linear regression approach. Where waters are of low alkalinity and of high organic content as in the case of waters from a peat area, pH control would be necessary and the models determined in this study may not be applicable without further adjustment. Further, the criteria for selection of alum doses was specific to South Australia at the time of the study where the emphasis was not on total organic carbon (TOC) removal. Nonetheless, the results do indicate that relationships can exist between chemical and physical features of raw waters and the coagulant doses required for treated water.

The application of any model for determination of coagulant dosing is likely to be dependant on the costs for monitoring of the raw water parameters and the confidence in the models based on an acceptable frequency of correct predictions and the absence or infrequency of unacceptable predictions. One approach that may be taken for the establishment of models is to acquire data of raw water parameters that have been found to influence coagulant dose requirements over a period when other methods are the primary procedure for determination of coagulant dosing, and then attempting to correlate the raw water parameters with the determined doses by empirical relationships. As a database of raw water parameters and required alum doses for a particular water source becomes more extensive the predictive capacity by modelling can be assessed. The impact of seasonal variation on these parameters could also be determined, e.g. the character of NOM present in a source water changes together with its alum demand. Utilisation of historical data relevant to particular seasons or replace-

ment of old data with recently acquired data for model building may provide more accurate predictions.

Predictability of coagulants by empirical models and artificial neural network

Where application of models and artificial neural networks predicts the required alum dose at an acceptable frequency and monitoring for the relevant parameters is in place at a water treatment facility then reliance could be made on these models and/or ANNs. With the application of a modelling procedure for the prediction of coagulant doses it can be anticipated that the predicted doses are greater or less than the actual doses required. To account for under-dosing due to errors in prediction, increased dosing above that predicted could be performed, as suggested by Ellis *et al.* [4], based on the estimation of the magnitude of the error.

Artificial neural networks have been used in a wide range of applications, such as pattern recognition [10–13] and coagulant dose determination [20,21]. The capacity of ANN to make accurate predictions for coagulant dosing at a specific water treatment facility has been demonstrated by Mirsepassi *et al.* [20,21] where operational data of 1821 days was available and alum doses used for training varied in increments less than 5 mg/L. In the present study a high proportion of the predicted doses differed from the actual doses by 5 mg/L or more. This may have been due to the size of the alum dose increment used in the jar test procedure where the smallest was generally 5 mg/L.

Multiple regression analysis was found to be the optimum approach of the different methods attempted for modelling (based on predictions by cross validation). This procedure was used by Ellis *et al.* [4] who also transformed model variables, e.g. the response variable, to its logarithm value.

According to Ratnaweera & Blom [6], error in prediction can occur due to the possibility of instrument failure where a range of parameters is used in a model. To account for this, a series of models may be developed that incorporates various combinations of parameters which provide acceptable levels of accuracy of prediction, though are ranked from the most to the least preferable. For example, using the multivariate relationships found in this study the optimum model found included the parameter of DOC, but not for the linear regression models. Hence selection of a model could be based on the range of parameters available.

Some parameters were readily measured, e.g. turbidities, absorbance at 254 nm/cm and colour, while other parameters such as alkalinity and DOC required more time consuming analysis. Besides the actual costs for these analyses an essential feature would also be that measurements of the parameters can be made in sufficient time to enable use of the model/ANN for coagulant dose adjustment. DOC analysis can be rapidly measured by the use of an on-line instrument such as one manufactured by Sievers Instruments Inc. (i.e. TOC Analyser

Model 820) though the cost of such an instrument would be an important consideration for many water treatment authorities, particularly the smaller ones. In this study the alkalinity was found to be an important parameter for prediction of alum doses by models though the measurement procedure used was time consuming. Hence a quicker analysis procedure would need to be used if the modelling/ANN procedures described are to be used for coagulant dosing at a water treatment facility. Potentially, a surrogate determination of the buffering capacity of raw water could be determined by measuring the pH shift with a standard addition of acid to a sample of the water.

Besides the possibility of using models/ANN to determine coagulant doses in operating plants, other benefits may include a means by which cost estimates can be made for coagulant use at a new or proposed water treatment facility.

Although the relationships between the raw water parameters and coagulant dose requirements are empirical in the models detailed, they may enable a greater understanding of the relative impact of these various parameters on coagulant demand in water treatment. Where the correlation between the predicted and actual doses declines over time, further study of other raw water parameters that may have been confounded, would be warranted.

Variation in parameters needed for incorporation into the models was found, e.g. DOC was determined to be an important parameter in multi-regression but not in the linear model. Similarly, variation in the types of parameters used for model building is described in the literature, which may also be a reflection of the relative importance of these parameters in coagulant dose requirements at different water treatment facilities; instrument and historical data availability and the costs for parameter determinations. Selection of parameters for model building or ANN will depend also on an understanding of the potential rates of change of raw water. Where rapid changes occur in the quality of raw water, e.g. in turbidity and TOC of river water, the dose rates of previous days are unlikely to be able to be used as parameters for dosing prediction for subsequent days.

In this study empirical regression models and ANN provided similar predictions for alum dosing that were comparable with the actual doses. Higher confidence in predictions may be obtained if more than one type of modelling approach is applied and where the predictions made are similar.

CONCLUSIONS

The results of this study provide further evidence that regression models and artificial neural networks have the potential for predicting coagulant dosing. As in previously reported studies, a variety of water parameters were utilised including turbidity, pH, colour, UV absorbance at 254 nm, alkalinity and DOC. However, these parameters were of raw waters only and did not include those of the treated water quality. The predictive capacity of the empirical models and ANN were

evaluated using cross validation. Using the maximum data set for training or model building these methods were found to provide similar levels of accuracy and variance, with a predominance of the predicted doses being within 10 mg/L of the actual doses. As the character of DOC present in natural waters is likely to vary depending on its source it would seem that the regression models and the ANN developed in this study are relevant only for the types of waters used. Further, the selection of alum doses used for correlation to raw water parameters are site specific in that they relate to the attainment of treated water quality for South Australian consumers at the time of this study. With a general trend towards the adoption of enhanced coagulation for the minimisation of residual NOM for the reduction in disinfection by-products, the models would need to be readjusted or new models developed. Nonetheless, readjustment may be a relatively easy task once the general relationships have been established, particularly as modern personal computers and software enable complex and otherwise time consuming mathematical operations to be performed quickly.

Unlike empirical models, ANN does not require a mathematical equation to fit the experimental data, or to assume an ideal theoretical relationship between variables. Instead, the neural network is adapted through training to represent the interactive and nonlinear nature of the system.

The growth of artificial neural networks has undoubtedly been triggered by the availability and increasing power of the modern personal computer. In the past few years, there has been a major improvement in the running speed of computers, e.g. a training step which took several hours [11] using an IBM XT can be shortened to less than 2 min using a Pentium™ PC (as found in this work).

Thus, techniques which may have been unpopular because of the time required to obtain an answer are now being utilised.

ACKNOWLEDGEMENTS

This work, part of the Australia–Germany Collaboration on Water and Environment R&D, was financially supported by the Department of Industry, Science and Tourism.

BIBLIOGRAPHY

- 1 Bernhardt H, Schell H. Experience with an on-line determination of the flocculant dosage using the LADUNGTITRMAT. *Water Supply* 1995; **13**(3/4): 279–284.
- 2 Bernhardt H, Schell H. Experience in coagulant control by use of a charge titration unit. *J Water SRT—Aqua* 1996; **45**: 19–27.
- 3 Bazer-Bachi A, Puech-Coste E, Ben Aim R, Probst JL. Mathematical modelling of optimum coagulant dose in water treatment plant. *Revue Des Sci L'eau* 1990; **3**: 377–397.
- 4 Ellis GW, Collins AG, Ge X, Ford CR. Chemical dosing of small water utilities using regression analysis. *J Environ Engng* 1991; **117**(3): 308–319.
- 5 Girou A, Franceschi M, Puech-Costes E, Humbert L. Model-

- isation des phenomenes de coagulation et etude de la morphologie des floccs: optimisation du taux de coagulant. *Recents Prog Genie Procedes* 1992; **6**(20): 373–385.
- 6 Ratnaweera H, Blom H. Optimisation of coagulant dosing control using real-time models selective to instrument errors. *Water Supply* 1995; **13**(3/4): 285–289.
 - 7 Critchley RF, Smith EO, Pettit P. Automatic coagulation control at water treatment plants in the North-West region of England. *J Inst Water Environ Manage* 1990; **4**(6): 535–543.
 - 8 Bos M, Bos A, van der Linden WE. Processing of signals from an ion-selective electrode array by a neural network. *Anal Chim Acta* 1990; **233**: 31–39.
 - 9 Hoskins JC, Kaliyur KM, Himmelblau DM. Fault diagnosis in complex chemical plants using artificial neural networks. *Aiche J* 1991; **37**(1): 137–141.
 - 10 Smits JRM, Breedveld LW, Derksen MWJ, Kateman G, Balffoort HW, Snoek J, Holstraat JW. Pattern classification with artificial neural networks: Classification of algae, based upon flow cytometer data. *Anal Chim Acta* 1992; **258**: 11–25.
 - 11 Sperring DA, Chow CW, Mulcahy DE, Davey DE, Haskard MR. A neural network applied to sensory signal processing determination of copper in water. *J Intell Mater Syst Struct* 1992; **3**(3): 418–431.
 - 12 Hartnett M, Diamond D, Baker PG. Neural network based recognition of flow inject patterns. *Analyst* 1993; **118**: 347–354.
 - 13 Hitzmann B, Kullick T. Evaluation of pH field effect transistor measurement signals by neural networks. *Anal Chim Acta* 1994; **294**: 243–249.
 - 14 Wilcox SJ, Hawkes DL, Hawkes FR, Guwy AJ. A neural network, based on bicarbonate monitoring, to control anaerobic digestion. *Wat Res* 1995; **29**(6): 1465–1470.
 - 15 Maier HR, Dandy GC. The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Res* 1996; **32**(4): 1013–1022.
 - 16 Chow CWK, Davey DE, Mulcahy DE. A neural network approach to zinc and copper interferences in potentiometric stripping analysis. *J Intell Mater Syst Struct* 1997; **8**(2): 177–183.
 - 17 Zupan J, Gasteiger J. Neural networks: a new method for solving chemical problems or just a passing phase? *Anal Chim Acta* 1991; **248**: 1–30.
 - 18 Bos M, Bos A, van der Linden WE. Data processing by neural networks in quantitative chemical analysis. *Analyst* 1993; **118**: 323–328.
 - 19 Spining MT, Darsey JA, Sumpter BG, Noid DW. Opening up the black box of artificial neural networks. *J Chem Educ* 1994; **71**(5): 406–411.
 - 20 Mirsepassi A, Cathers B, Dharmappa HB. Application of artificial neural networks to the real time operation of water treatment plants. *IEEE International Conference on Neural Networks Proceedings*, Vol. 1, pp. 516–521. Perth, Western Australia, 27 November–1 December, 1995.
 - 21 Mirsepassi A, Dharmappa HB, Cathers B. Economic benefits of applying an artificial neural network to the operation of a water treatment plant. *Proceedings, Watertech*. The Centre, Darling Harbour, Sydney, 27–28 May, 1996.
 - 22 Bennet LE, Drikas M. The evaluation of colour in natural waters. *Wat Res* 1993; **27**(7): 1209–1218.
 - 23 Eaton AD, Clesceri LS, Greenberg AE, eds. *Standard Methods for the Examination of Water and Wastewater*, 18th edn. American Public Health Association, Washington, DC, USA, 1992.
 - 24 Apey EA. *Back Error Propagation Simulator (BPS)*. George Mason University, 1989.
 - 25 Kawamura S. Optimisation of basic water-treatment processes—design and operation: coagulation and flocculation. *J Water SRT—Aqua* 1996; **45**: 35–47.
 - 26 Bursill DB, Hine PT, Morran JY. The effect of natural organics on water treatment processes. *Proceedings of the Australian Water and Wastewater Association*. 11th Federal Convention, Melbourne, 1985.
 - 27 Hernebring C. Rapid methods for the determination of required alum dosage in surface water treatment plants. *Vatten* 1987; **42**: 209–219.
 - 28 Krasner SW, Amy G. Jar-test evaluation of enhanced coagulation. *JAWWA* 1995; **87**(10): 93–107.