ARTIFICIAL NEURAL NETWORKS AS ALTERNATIVE APPROACH FOR PREDICTING TRIHALOMETHANE FORMATION IN CHLORINATED WATERS

M. Hashem[1] **and H. Karkory**[*, 2]

[1] Assistant Professor, Civil Engrg Dept., Higher Institute of Engineering, Hon, Libya
[2] Assistant Professor, Head of Civil Engrg Dept., Higher Institute of Engineering, Hon, Libya

[*]Corresponding author: mhashem99@yahoo.com

ABSTRACT

Chlorination by-products (CBPs) are generated during water disinfection due to the reaction of chlorine with natural organic matter contained in raw water. There is more interest in modelling CBPs in order to better understand and manage the presence of these compounds in drinking water. This study is a trial to use artificial neural networks (ANN) to predict trihalomethane (THM) formation during chlorination bench-scale experiments. This trial is investigated and compared with the use of classical multivariate linear regression (MLR). The analysis of modelling results shows that, ANN in general is better than MLR for predicting THM formation for most water quality and chlorination conditions.

Keywords: Trihalomethanes, Modeling, Neural Networks.

INTRODUCTION

All natural waters and even treated drinking water exerts disinfectant demand due to the reactions with natural organic matter (NOM) and other constituents in water. Therefore, the applied disinfectant dose must be sufficient to meet the inherent demand in the treated water, to provide sufficient protection against microbial infection, and at the same time minimize exposure to Chlorination by-products (CBPs) (Rook 1974; Symons et al. 1975). Among these, trihalomethanes (known as THMs: chloroform, bromodichloromethane, dibromochloromethane and bromoform) have been the focus of particular attention because they are considered potentially carcinogenic (Cantor et al. 1987). Concerns about health risks associated particularly with chloroform and bromodichloromethane have resulted in the establishment of maximum acceptable levels of THM concentrations in drinking water by the World Health Organization (WHO) (Rodriguez & Serodes 2001). The US Environmental Protection Agency (USEPA) proposed a two-stage disinfection rule in which

maximum contaminant levels of total THMs are 80 µg/l based on an annual running average (Sharfenaker 2001). Canada set out drinking water guidelines stating a total THM maximum acceptable level of 100 µg/l which is the same standard established by the European Community (Rodriguez et al. 2003).

Some research effort has been made to develop predictive models for the formation of THMs in water. Models for THMs are aimed at better understanding the factors affecting their formation and thus are useful as decision-making tools. The modelling of THMs consists of establishing relationships between THM levels in water and the parameters of water quality and operational control which can influence their formation. The most important factors for THM formation are the levels of organic matter in water (generally designated by total or dissolved organic carbon, TOC or DOC, and by UV 254 nm absorbance), the applied chlorine dose, water pH, water temperature and the reaction time of residual chlorine in water (Rodriguez et al. 2003). Concentrations of bromide are also generally considered because they influence the distribution of the four THM components. In different studies, THM formation models have been developed both from data generated in full-scale studies at real water utilities or at laboratory-scale using controlled chlorination conditions. Some researchers have developed models to describe the formation of THMs based on kinetics involved during chlorine reactions (Clark & Sivaganesan 1998). However, most of the models presented in literature are empirical and are based on statistical regression equations which predict the levels of THMs from a number of operational and water quality parameters (Milot et al. 2000). Although regression-based models have shown acceptable predictive capacity of THM formation, they do not allow for consideration of the complex and non-linear interactions between operational and water quality parameters used as estimators for THMs.

The objective of this paper is to evaluate the ability of a recent non-linear modelling technology, artificial neural networks (ANNs), to predict the concentrations of THMs formed under controlled chlorination conditions. ANN modelling will be assessed by comparing with the classical multivariate linear regression (MLR) approach. The development of the model is based on database published by Amy et al. (1987)

THE SIMULATED DATA

To assess the utility of ANN for modeling trihalomethane formation, MLR model and ANN model are required. Model development is based on independent experimental database developed by Amy et al. (1987). The verification test set (200 observations) was used to evaluate the accuracy of all the model topologies that were tested and to select the best ANN structure using as performance criteria the mean square error (MSE) and the coefficient of multiple determinations (R2) between observed and estimated values of trihalomethane concentrations. Statistical distribution of water quality parameters for chlorination experiment data is shown in Table (1). Also the general performance of MLR and ANN models are shown in Table (2).

Table (1). Statistical distribution of water quality parameters for chlorination experiment data developed by Amy et al. (1987).

Parameter	**Minimum**	**Mean**	**Median**	**Maximum**
Dose (mg/l)	2	22	19	69
TOC(mg/l)	3	7	6	14
PH	4.5	7.3	7.2	9.7
UV (cm^{-1})	0.064	0.242	0.252	0.49
t (h)	0.1	32	4	168
T (°C)	10	20	20	30
Br (mg/l)	0.01	0.303	0.151	1.245
THM (µg/l)	4	280	176	2843

Table (2). Comparison of model performance according to different ranges of chlorination levels and of resulting THM concentrations

Model performance for whole results (n = 200)				***Case (I)*** **Chlorine dose range (mg/l)**	**Model performance for case (I)**				***Case (II)*** **THM level range (µg/l)**	**Model performance for case (II)**			
MLR		**ANN**			**MLR**		**ANN**			**MLR**		**ANN**	
MSE	**R^2**	**MSE**	**R^2**		**MSE**	**R^2**	**MSE**	**R^2**		**MSE**	**R^2**	**MSE**	**R^2**
7693	0.931	2143	0.982	0 to 16 (n = 65)	1547	0.960	723	0.98	0 to 100 (n = 60)	571	0.987	508	0.991
				17 to 25 (n = 72)	6011	0.794	1890	0.934	101 to 250 (n = 66)	2338	0.842	1287	0.912
				26 + (n = 63)	16070	0.943	3921	0.985	251 + (n = 61)	18430	0.924	4269	0.983

OVERVIEW OF MLRS AND ANNS

MLR analysis is a well-known modeling methodology used in many research fields to establish the strength of a linear relationship between a dependant variable and a set of independent variables (Menard 1995). The relationship between variables can be described using an equation in the following form (Rodriguez et al. 2003).

$$Y = \sum \beta_0 + \sum_{i=1}^{m} \beta_i X_i \quad (1)$$

Where Y is the dependent variable, X_i represents the independent variables with m denoting the number of independent variables considered, β_0 the intercept and β_i the partial slope coefficients providing a partial exploration or prediction for the value of Y. The parameters of MLR model are generally estimated using the ordinary least square (OLS) method which results in a line that minimizes the sum of squared vertical distances from the observed data points to the line (Neter et al. 1990).

ANNs are a modeling technique inspired by studies of brain's nervous system. They are capable of learning by example from representative data which describe a physical phenomenon or a decision process (Rumelhart et al. 1994). An ANN model provides certain theoretical advantages over conventional approaches such as MLR, including its high capacity for generalization and its increased tolerance to noisy data (Hammerstrom 1993).

An ANN consists of a set of nodes (neurons) organized into (1) an input layer, which receives the input data; (2) one or more hidden layer(s), which process the data; and (3) an output layer, which produces the network output. Many ANN structures have been proposed and explored since the 1950s. Among the most researched and widely used structures in hydrology and water-resource problems are multi-layer feed-forward networks (MFNs) with back-propagation (BP) training algorithms (Govindaraju & Rao, 2000). The typical topology is shown in Fig. 1. The nodes of one layer are connected to the nodes of another layer with connection weight, but they are not connected to nodes of the same layer. Thus, each node in a layer receives signals from nodes of the previous layer with connection weights, adds the weighted inputs of all nodes, converts the weighted sum into an output signal, and transmits the output signal to the nodes of the following layer.

The connection weights between nodes are optimized using the known input and target values through an iterative process and error-minimization technique, so that the network produces outputs close or equal to the known target values. The process is called training of the network. The trained network with an optimized set of connection weights is then applied to the validation data set to estimate the output. The network where data flow is in one direction is known as the feed-forward network; on the other hand, the network where the error estimated between the target and ANN-predicted values is propagated backward for connection weight optimization is called

the feed-forward with back-propagation network. The *newff* subroutine available in the Neural Network Tool box of MATLAB was used to create ANN model (MathWorks, Inc. 2001). The ANN model which performed best contains seven inputs (chlorine dose level, total organic carbon (TOC), pH, UV, T, t, and Br) one output (THM concentration) with two hidden layers of 15 hidden elements and required a learning rate of 0.1 and momentum of 0.5.

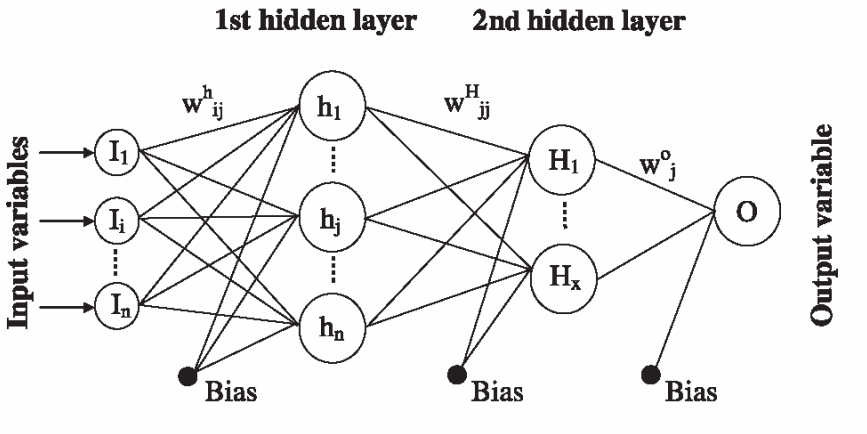

Fig. 1: Schematic diagram of ANN model.

The ANN modeling approach has been applied in several fields for solving classification problems and in the field of drinking water management they have been used for assessing the quality of raw water (Maier & Dandy 2000), for predicting urban water demand (Heller & Singh Thind 1994), for establishing coagulation dosage (Joo et al. 2000) and for predicting residual chlorine decay in distribution systems (Hashem & Hassan 2005). Considering how much interest in modeling THMs in drinking water has grown in recent years and the ability of the ANN to model complex and non-linear phenomena, it becomes important to assess the abilities of this approach in predicting THM formation in chlorinated waters.

COMPARISON OF MLR AND ANN

Table 2 presents the results for models applied to the verification dataset. The results is presented for the verification dataset and also for specific subsets according to the chlorine dose applied during the experiments and the range of resulting THM concentrations. Performance criteria values for the best models obtained using both modelling approaches show that, estimation of THMs based on ANN modelling is significantly more accurate than estimation with the MLR model.

For better explanation of these results, a more detailed analysis is necessary, comparing model performance for the specific conditions of chlorination experiments THMs generated. Information from Table 2 confirms that, in general, ANN models have a greater capacity than MLR models to predict formation. The ANN model gave

better results than the MLR model for any range of chlorination conditions and resulting THMs. Figure 2 illustrate these findings by showing some examples of model prediction fitting with the actual data, for bench-scale experiments carried out with both low, moderate and high chlorine doses (typically 6.9, 15.7 and 69 mg/l) and with low and high (typically 10 and 20°C) water temperature conditions.

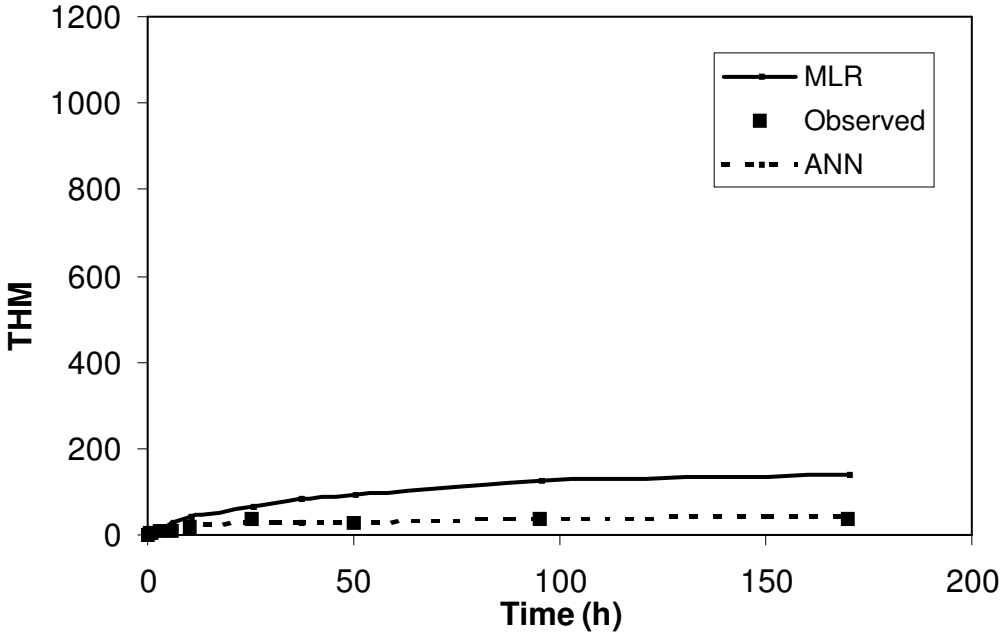

Fig. 2-a: Comparison of THM predictions using MLR and ANN models for specific experiments: low chlorine dose and water temperature.

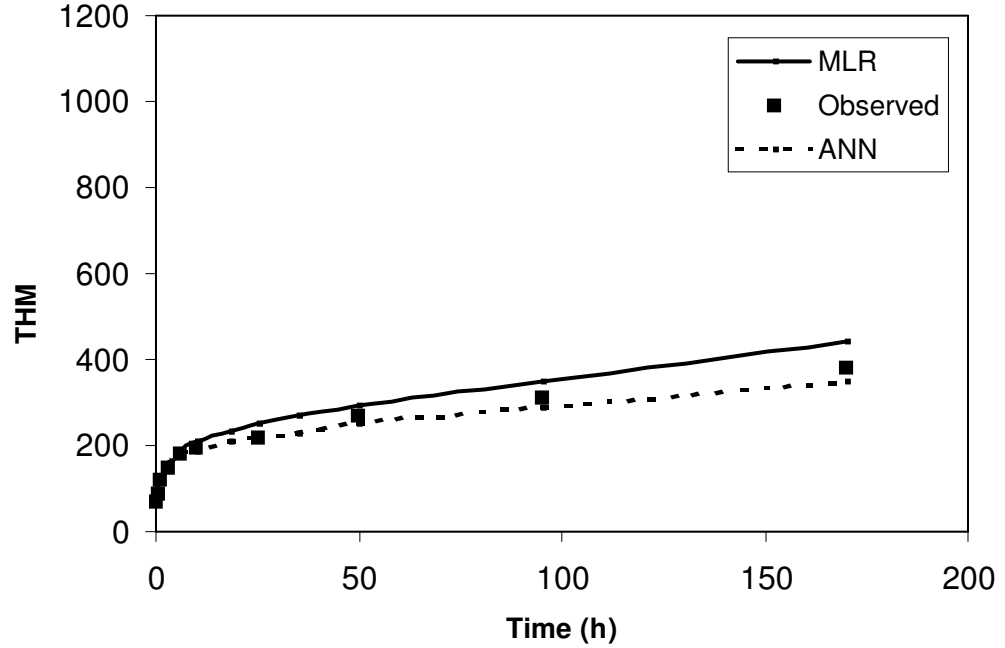

Fig. 2-b: Comparison of THM predictions using MLR and ANN models for specific experiments: moderate chlorine dose and high water temperature.

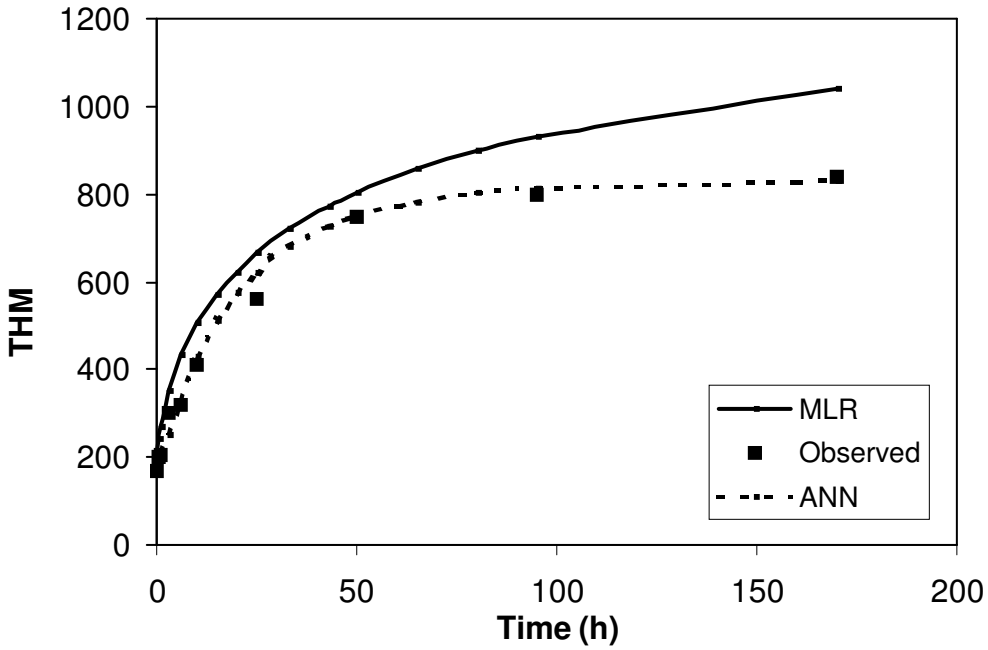

Fig. 2-c: Comparison of THM predictions using MLR and ANN models for specific experiments: high chlorine dose and water temperature.

CONCLUSION

In this paper, MLR and ANN models for the prediction of THM formation during bench-scale chlorination experiments was developed and compared. Even if, most of models published in the literature on THM formation in bench-scale conditions are based on the MLR modelling approach, the results of this paper demonstrate that performance of such models could be significantly improved by using ANN as an alternative modelling approach. The present investigation shows that, ANN can significantly reduce the calculated MSE between measured and predicted results.

The obtained results demonstrate that he developed ANN model provide a good fitted function. Moreover, the results demonstrate the flexibility of the ANN modelling approach for THM formation. The results of this paper show that ANN constitute a very promising alternative to MLR for modelling THM formation in experiments conducted at bench-scale conditions.

REFERENCES

1. Amy, G.L., Chadik, P.A. and Chowdhury, Z.K. 1987. "Devoloping models for predicting trihalomethane formation potential and kinetics". J. Am. Wat. Wks Assoc. 79(7), 89-97.
2. Cantor, K.P., Hoover, R.Hartge, P., Mason, T.J., Silverman, D.T., Altman, R., Austin, D.F., Child, M.A., Key, C.R., Marret, L.D., Myers, M.H., Narayana, A.S., Levin, L.I., Sullivan, J.W., Swanson, G.M., Thomas, D.B. and West, D.W.

1987. "Bladder cancer, drnking water source, and tap water consumption: a case-control study". J. National Cancer Inst. 79(6), 1269-1279.
3. Clark, R.M. and Sivaganesan, M. 1998. "Predicting chlorine residuals and formation of TTHMs in drinking water". J. Environ. Engrg. 124(12), 1203-1210.
4. Govindaraju, R.S., Rao, A.R. 2000. "Artificial Neural Networks in Hydrology". Kluwer Academic Publishers, Boston, USA
5. Hammerstrom, D. 1993. "Neural networks at work". IEEE Spectrum, 26-32.
6. Hashem, M. and Hassan, I.M. 2005. "Using artificial neural network model as a new technique for simulating residual chlorine". Journal of Engineering Sciences, Assiut University, Vol. 33, No. 3 , pp. 735-743.
7. Heller, M. and Singh Thind, H. 1994. "Forecasting with cascade correlation: an application to potable water demand". ANNIE'94. St. Louis, MO, PP. 115-1160.
8. Joo, D.S., Chio, D.J. and Park, H. 2000. "Determination of optimal coagulant dosing rate using an artificial neural network". J. Wat. Suppl.: Res. & Technol.-AQUA 49, 49-55.
9. Maier, H.R. and Dandy, G.C. 2000. "Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and application". Environ. Modelling & Software 15, 101-124.
10. MathWorks, Inc. 2001. "MATLAB". 3 Apple Hill Drive, Natick, MA, USA.
11. Menard, S. 1995. "Applied logistic regression analysis". Quantitative Appl. Social Sci., Sag Univ., CA, 7-106.
12. Milot, J., Rodriguez, M. J. and Serodes, J.B. 2000. "Modelling the susceptibility of drinking water utilities to form high concentrations of trihalomethanes". J. Environ. Mgmt 60(2), 155-171.
13. Neter, J., Wasserman, W. and Kutner, M. 1990. Applied linear statistical models, 3rd edition. Irwin, Homewood, IL.
14. Rodriguez, M.J., Milot, J., Serodes, J. 2003. "Predicting trihalomethane formation in chlorinated waters using multivariate regression and neural networks". J. Wat. Suppl.: Res. & Technol.-AQUA 52, 199-215.
15. Rok, J.J. 1974. "Formation of haloforms during chlorination of natural waters". Wat. Treat. & Examin. 23, 234-243.
16. Rumelhart, D.E., Widrow, B. and Lehr, M.A. 1994. "The basic idea of neural networks". Communications of the ACM 37(3), 87-92.
17. Sharfenaker, M.A., 2001. "USEPA offers first glimpse of stage 2 D/DBPR". J. Am. Wat. Wks Assoc. 93(12), 20-34.
18. Symons, J.M., Beller, T.A., Carswell, J.K., DeMrco, J., Krapp, K.L., Robeck, G.G., Seeger, D.R., Sloccum, C.J., Smith, B.L. and Steevens, A.A. 1975. "National organics reconnaissance survey for halogenated organics". J. Am. Wat. Wks Assoc. 67(11), 634-648.