

ARTIFICIAL NEURAL NETWORKS PREDICTIVE MODELS. A CASE STUDY: CARBON AND BROMINE CONCENTRATIONS PREDICTION BASED ON CHLORINATION TIME

E.G. FARMAKI^{1, 2}

S.A. SAMIOS^{2, 3}

N.S. THOMAIDIS^{1,*}

S. GOLFINOPOULOS⁴

C.E. EFSTATHIOU¹

T.D. LEKKAS^{2,3}

¹Laboratory of Analytical Chemistry, Department of Chemistry
University of Athens, Panepistimiopolis Zografou
15771 Athens, Greece

²Athens Water Supply and Sewerage Company (EYDAP SA)
11146 Athens, Greece

³Water and Air Quality Laboratory, Department of Environmental Studies
Faculty of the Environment, University of the Aegean
University Hill, 81100 Mytilene, Greece

⁴Department of Financial and Management Engineering
University of the Aegean, 41, Kountourioti str.
82100 Chios, Greece

Received: 02/09/11

Accepted: 13/10/11

*to whom all correspondence should be addressed:
e-mail: ntho@chem.uoa.gr

ABSTRACT

Artificial neural networks (ANNs) are being used increasingly to predict water variables. This study offers an alternative approach to quantify the relationship between time of chlorination in potable water (due to convectional treatment procedure) and chlorination by-products concentration (expressed as carbon and bromine) with an ANN model, i.e., capturing non-linear relationships among the water quality variables.

Thus, carbon and bromine concentrations in potable water (the second chosen due to the toxicity of brominated trihalomethanes, THMs) were predicted using artificial neural networks (ANNs) based mainly on multi-layer perceptrons (MLPs) architecture. The chlorination (detention) time as much as 58 hours in Athens distributed network, comprised the input variables to the ANNs models. Moreover, to develop an ANN model for estimating carbon and bromine, the available data set was partitioned into training, validation and test set. In order to reach an optimum amount of hidden layers or nodes, different architectures were tested. The quality of the ANN simulations was evaluated in terms of the error in the validation sample set for the proper interpretation of the results. The calculated sum-squared errors for training, validation and test set were 0.056, 0.039 and 0.060 respectively for the best model selected.

Comparison of the results showed that a two-layer feed-forward back propagation ANN model could be used as an acceptable model for predicting carbon and bromine contained in potable water THMs.

KEYWORDS: ANNs; chlorination by-products (CBPs), trihalomethanes (THMs).

1. INTRODUCTION

Chlorine has been used as a disinfectant agent since the beginning of the nineteenth century. It is the most commonly used disinfectant due to its efficiency, low cost and easy application. The use of chlorine as an acceptable disinfectant of drinking water has dramatically reduced the incidence of waterborne diseases. However, chlorination is incriminated for the formation of series of disinfection by-products (DBPs) or chlorination by-products (CBPs), like trihalomethanes (THMs: CHCl_3 , CHBrCl_2 , CHBr_2Cl , and CHBr_3), which are characterized as probable or possible human carcinogen

(Chowdhury *et al.*, 2009). Thus, the European Union has set a maximum contaminant level (MCL) of $100 \mu\text{g L}^{-1}$ for the sum of these substances (EC Directive, 1998).

The amount of chlorine added to the water seems to be very important. The dose cannot be very low as the residual left is crucial for the water protection along the whole distribution system. On the other hand, if the chlorine dose is high, its reactivity with DBP precursors or natural organic matter (NOM) can lead to elevated values of by-products.

The levels of THMs in chlorinated water are also associated with chlorine demand, pH, temperature and seasonal variability, chlorine contact time and organic materials or chemical elements (Chowdhury *et al.*, 2009; Nikolaou *et al.* 2002; Bo *et al.*, 2008; Sohn *et al.*, 2006; Ye *et al.*, 2009). Their formation is generally known to be increasing because of water high temperature, chlorine dosage and levels of dissolved organic carbon (DOC) in raw water (Ye *et al.*, 2009; Nikolaou *et al.*, 1999). In parallel, the presence of bromide ion seems very significant in THMs formation. Thus, THMs' level during chlorination of raw waters containing bromides was also studied by many investigators (Bo *et al.*, 2008; Sohn *et al.*, 2006; Ye *et al.*, 2009; Symons *et al.*, 1993; Heller-Grossman *et al.*, 1999; Mok *et al.*, 2005), as even at lower bromide concentration, brominated THMs can be produced (Bo *et al.*, 2008; Mok *et al.*, 2005). The molar yields of THMs and haloacetic acids (HAAs) increases as the initial bromide concentration increases. Bromide ion in natural waters is readily oxidized and incorporated into THMs during chlorination. The formation of THMs shifts to more brominated species with increasing bromide concentration (Symons *et al.*, 1993; Hua *et al.*, 2006).

However, the determination of THMs concentration in treated water seems to be rather a laborious work. Thus, statistical models have been developed in the past for the prediction of THMs or chlorine residual concentration. Gibbs *et al.* (Gibbs *et al.*, 2003) uses linear regression models (LR) and multi-layer perceptron (MLPs), to predict chlorine concentrations in Hope Valley water distribution system in South Australia. A total of six (6) inputs are included in the models: the water temperature at the sites of interest and at the water treatment plant (WTP), the WTP flow 5 days before the chlorine measurement being predicted, and the chlorine concentrations at the WTP, at the tank inlet and at the site under prediction the previous week. Rodriguez *et al.* (1997) implemented a single smoothing factor general regression neural network to predict the chlorine residual in a simple trunk main using hourly chlorine measurements. Milot *et al.* (2002) presents the application of ANNs to model THMs occurrence in drinking water. ANNs are compared with other modeling approaches, logistic regression and multivariate regression, to predict concentrations of formed THMs with variable chlorination conditions. May *et al.* (2008) investigates the application of the modified Partial Mutual Information-based Selection (PMIS) algorithm to the development of ANN models for forecasting disinfectant residual within water distribution system. Additionally in a recent work, Ye *et al.* (2011) uses seven inputs for developing a predictive model of THMs and HAAs concentrations. More applications are reviewed elsewhere (Chowdhury *et al.*, 2009).

In this work, assuming pH and temperature fluctuations as negligible, THMs concentrations (specifically carbon and bromide content) are predicted, based on the estimated chlorinated time (depended on the length of distribution network and the season) in treated water samples of Athenian WTP (Acharnes and Aspropyrgos facilities). Multi-layer perceptrons (MLPs) and Radial Bases Function (RBF) models are used, evaluated and compared in order to determine the parameters that define the best model predicting THMs concentration in chlorinated samples.

2. EXPERIMENTAL/METHODOLOGY

2.1 Area description

About three hundreds (295) samples were collected from two WTPs used for the water supply of Athens, named Aspropyrgos and Acharnes (Figure 1). Distribution system terminals were also sampled during a period from 2004 to 2007. The Athenian Distribution System is operated in such a way that very few points have detention time greater than 48 hours.

The distribution network that starts from Aspropyrgos WTP supplies drinking water up to Salamina (an island located 27 km away from the WTP reservoir tank). According to hydraulic data, detention time between pre-chlorination and the exit from flocculation tank is 2 hours, sand filtering is about 1 hour and its storage in reservoir tank is estimated to 12 hours. Chlorine detention time till Salamina sampling point is estimated to 27 hours depending on the water outflow.

The Acharnes WTP primarily supplies the portion of the distribution network that covers the areas of Attica with higher altitudes; however, it also provides supplemental quantities of water to the Municipalities of Athens and Piraeus. Its current total water treatment capacity is $800.000 \text{ m}^3 \text{ d}^{-1}$.

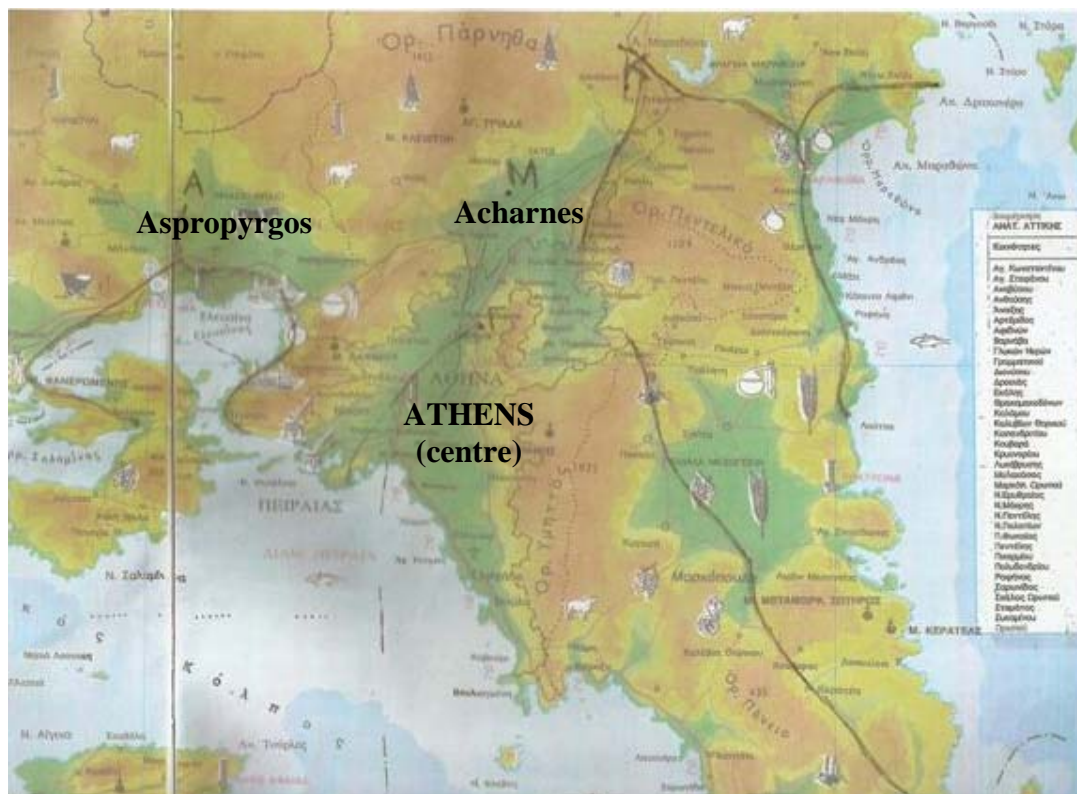


Figure 1. Acharnes and Aspropyrgos WTPs in Athens

2.2 Samples collection and analysis

Sample collection was performed in 40 mL sample vials capped with PTFE-faced silica septum. The THMs were preserved by the addition of a common dechlorination reagent (sodium sulfite), before shipment to the field. All samples were collected in duplicate. The sample bottles should be overflowed in order to avoid the formation of air bubbles through the sample. The samples were chilled to 4°C on the day of collection and maintained at that temperature until analysis. The samples were analyzed within 14 days of collection (there was an attempt to analyze them within 3 days of collection so as to prevent the hydrolysis of haloacetonitriles and halo ketones and further CHCl_3 formation).

The method employed for the determination of THMs, was a modification of 551.1 EPA method developed by Nikolaou *et al.* (2001) as it allows to extract them from drinking water quantitatively (Cancho and Ventura, 2005). Methyl-t-butyl ether (MTBE) was the primary extraction solvent. A 35 mL sample aliquot was extracted with 2 mL of MTBE after addition of 2.5 g sodium sulfate. One μL of the extract was then injected into a GC (Trace 2000 Thermo-Finnigan gas chromatograph) equipped with a fused silica capillary column and electron capture detector (^{63}Ni ECD) for separation and analysis. A capillary column AT-1 60 m x 0.25 mm i.d. x 0.25 μm was used. Procedural standard calibration was used to quantify analytes. The instrumental parameters used are given in Table 1. The method was validated and recoveries and precision figures were obtained for all the studied THMs (Table 1). Limits of detection (LOD) and quantification (LOQ) were calculated (Table 1). The interlaboratory scheme recorded, is a LEAP test purchased from FAPAS in June 2006.

Table 1. Quality control / validation data and GC instrumental parameters and for the determination of THMs in water

	LODs /LOQs ($\mu\text{g L}^{-1}$)	QCs recoveriers (%) in 10fold concentration of LOD (% RSD, n=7)	z-score (interlaboratory scheme)
CHCl_3	1.0 / 3.0	98 (4.5)	-0.018
CHBrCl_2	0.50 / 1.5	98 (3.5)	-1.3
CHBr_2Cl	0.50 / 1.5	102 (4.6)	-0.57
CHBr_3	0.50 / 1.5	98 (5.2)	-1.6

The procedure for the analysis of THMs was splitless injection at temperature 175°C; Carrier gas: helium; Flow rate 2.0 mL min⁻¹.
Oven conditions: Initial temperature: 37°C for 20 min; Ramp to 0.2°C min⁻¹ up to 39°C; Ramp to 2.3°C min⁻¹ up to 100°C.
Retention time (min): 8.5 for CHCl_3 , 13.2 for CHBrCl_2 , 17.6 for CHBr_2Cl and 20.6 for CHBr_3

2.3 Data pretreatment and calculations

For the calculation of the detention time, distance and water supply data of the sampling points were used. Except this, the season seemed to play an important role, as during summer, water consumption (and supply) is increased and thus the detention time is minimum.

The calculations of carbon and bromine content in THMs for every sample were performed as follows:

$$[\text{C}] = \text{Ar}_{\text{C}}([\text{CHCl}_3]/\text{Mr}_{\text{CHCl}_3} + [\text{CHCl}_2\text{Br}]/\text{Mr}_{\text{CHCl}_2\text{Br}} + [\text{CHClBr}_2]/\text{Mr}_{\text{CHClBr}_2} + [\text{CHBr}_3]/\text{Mr}_{\text{CHBr}_3})$$

$$[\text{Br}] = \text{Ar}_{\text{Br}}([\text{CHCl}_2\text{Br}]/\text{Mr}_{\text{CHCl}_2\text{Br}} + 2[\text{CHClBr}_2]/\text{Mr}_{\text{CHClBr}_2} + 3[\text{CHBr}_3]/\text{Mr}_{\text{CHBr}_3}),$$

where [C] and [Br] and $[\text{CHCl}_3]$, $[\text{CHCl}_2\text{Br}]$, $[\text{CHClBr}_2]$, $[\text{CHBr}_3]$ are the weight concentrations of carbon and bromine and the four individual THMs and Ar, Mr the atomic and molar relative masses, respectively.

Mean values of carbon and bromine content were also calculated for the samples corresponding to the same detention time. Thus, twenty-five samples were finally used for the construction of the ANNs models. Detention time was ranged from 1.5 to 58.3 hours, while carbon and bromine concentrations were fluctuating from 1.5 to 2.4 and 2.7 to 6.4 $\mu\text{g L}^{-1}$, respectively. Samples were randomly divided in three sets: training (12), validation or selection (6) and test (7). Training sample set was used for the models' learning process; validation set was used for an independent check of this process and test set evaluated the final predictive ability of the model (Farmaki *et al.*, 2010).

2.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) have seen an explosion of interest over the last two decades and have been successfully applied in all fields of analytical chemistry. Inspired from biological systems and originated from the perceptron, i.e. a program unit that learns concepts, ANNs are capable of gradual learning over time and modeling extremely complex functions. In addition to the traditional multivariate chemometric techniques, ANNs are often applied for prediction, clustering, classification, modeling of a property, process control, procedural optimization and/or regression of the obtained data (Farmaki *et al.*, 2010). However, unlike the traditional model-based methods, ANNs are self-adaptive data-driven methods: they learn from examples and are capable of capturing relationships, even if these are hard to be described or explained (Zhang *et al.*, 1998). Thus, ANNs are appropriate for problems that knowledge is hard to be extracted.

Moreover, ANNs are non-linear models that can generalize efficiently, are noise and fault tolerant and their operation is based on less assumptions and restrictions for the data compared to the traditional techniques. Theoretically, they can also approximate any function, mainly the MLPs (Farmaki *et al.*, 2010).

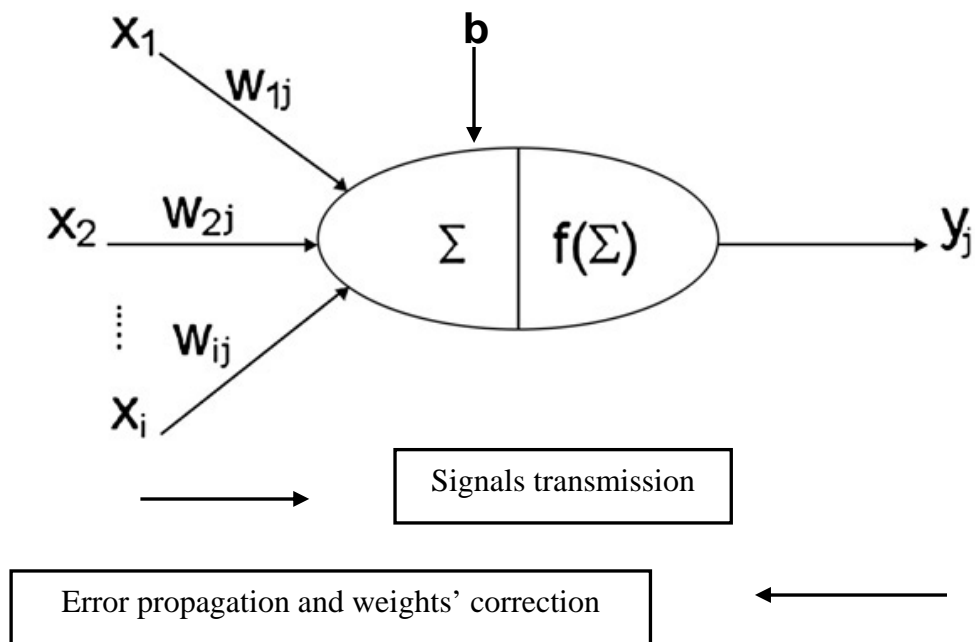


Figure 2. The neuron: the basic processing unit of an ANN model

ANNs receive a number of inputs in the processing units that are called neurons and are capable of:

1. processing data received from the previous units with the use of weights W_{ij} and an activation function f ,
2. sending signals y_j to the next ones and
3. propagating the error received from the next neurons to the previous ones (back-propagation, BP algorithm).

Figure 2 summarizes the aforementioned features.

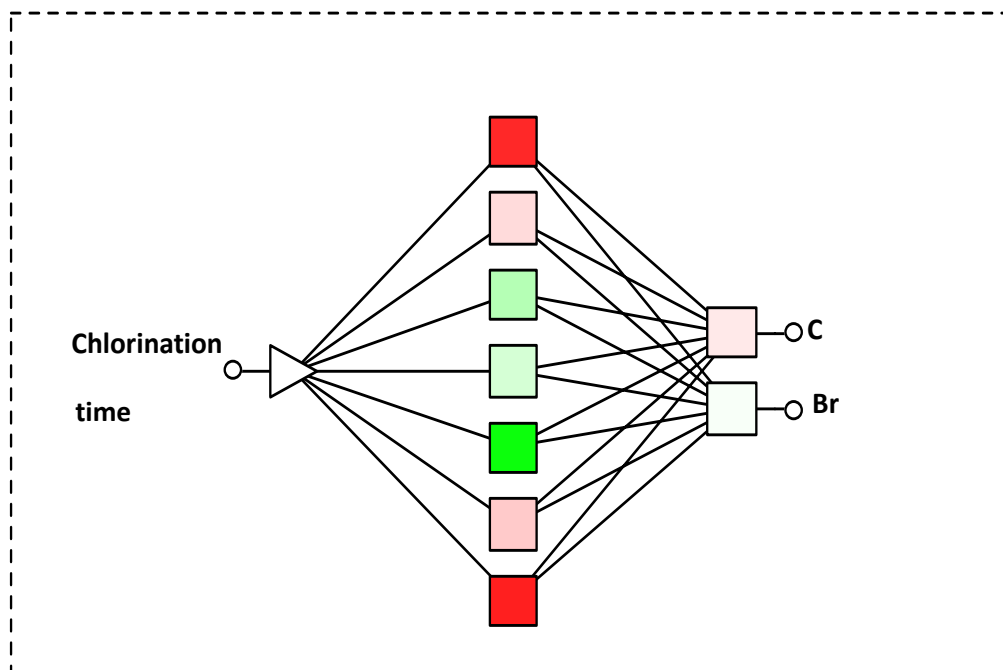


Figure 3. Architecture of the final MLP model (1:1-7-2→2)

Two famous ANNs techniques were applied in this work as a non-parametric chemometric tool to find a regression model for quantifying the relationship between time chlorination in potable water and THMs concentration (expressed as carbon and bromine).

Multi-layer perceptrons (MLPs) trained by the back-propagation algorithm (BP), and Radial Basis Function (RBF) models were selected for this work.

All the calculations and plots were made using Excel 2003 by MicroSoft, and Statistica for Windows, Version 7.0 by StatSoft Inc., 2004.

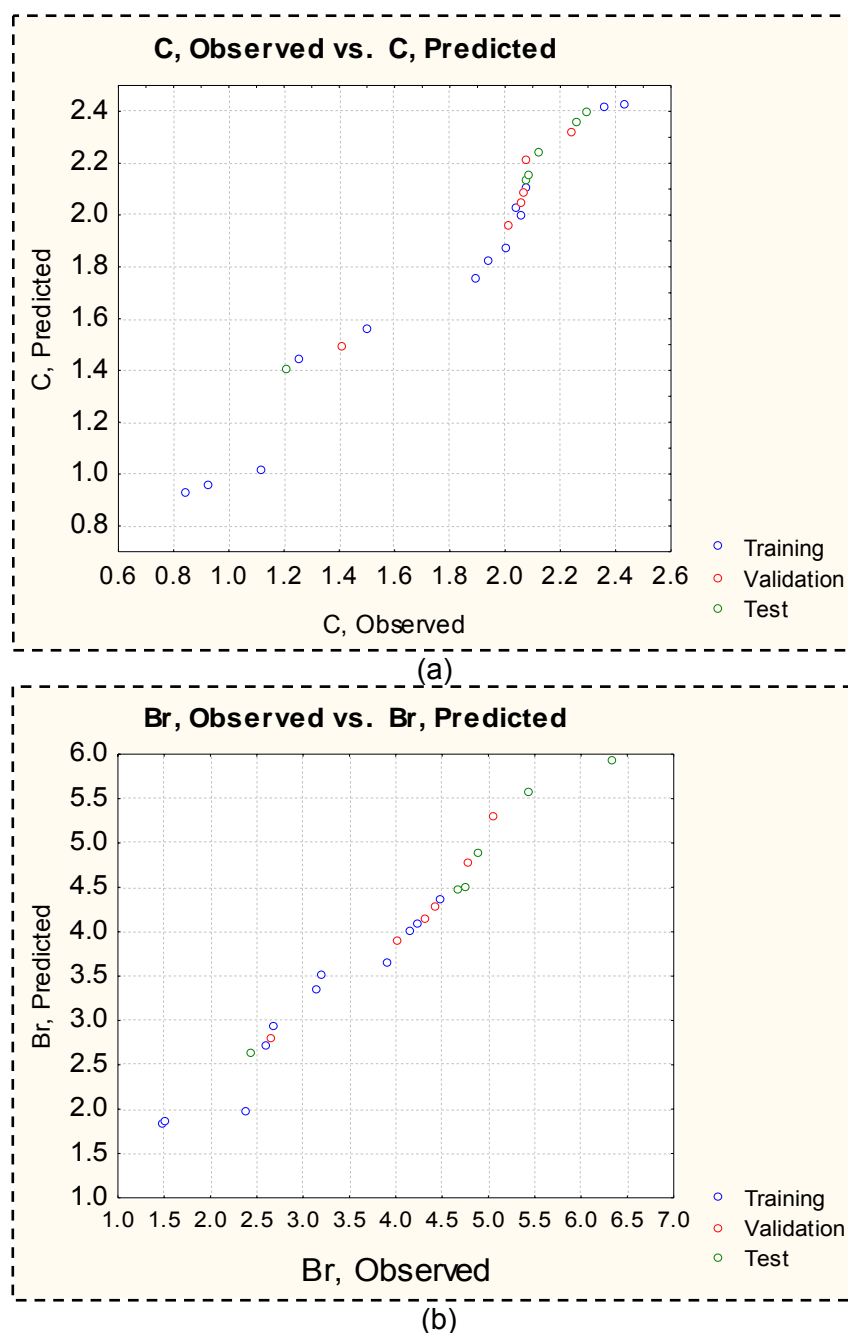


Figure 4. Best MLP model: Observed vs predicted values plots (a) for C and (b) for Br

3. RESULTS AND DISCUSSION

3.1 MLPs models

Multi-layer perceptrons networks used in this work were optimized through the parameters of the number of the hidden units and the hidden layers. The criterion used was the RMS (Root Mean

Square) error for the validation sample set. For each trial, different networks were tested. Thus, the initial parameters for the model construction are different and independent.

The number of hidden units was fluctuated from 2 to 9, for one or two layers. Finally, the best model chosen succeeded RMS error = 0.056, 0.039 and 0.060 for training, validation and test set, respectively. The architecture of the model included seven hidden units in its unique layer (Figure 3). The final model MLP 1:1-7-2:2 reflected the number of inputs (1), the number of the neurons in the hidden layer (7) and the final outcome prediction (2). Predicted and observed values were also compared by the correlation plot (Figure 4). Training, validation and test samples are marked with different colors. The performance of the artificial neural networks model was excellent ($r = 0.96$ and 0.98 respectively for carbon and bromine concentration).

3.2 RBF models

RBF models were optimized in terms of the number of units in the hidden layer and the spread of the Gaussian function. The criterion used was also the RMS error for the validation sample set. The recognition (concerning training samples) and predictive (concerning test samples) ability of these models seemed to be lower than the MLPs ones. The best model chosen included six hidden neurons in the intermediate layer and succeeded RMS error = 0.074, 0.068 and 0.082 for training, validation and test set, respectively (Table 2).

Table 2. Error for the optimized MLP and RBF models

Model	Error		
	Training	Validation	Test
MLP 1:1-7-2:2	0.056	0.039	0.060
RBF 1:1-6-2:2	0.074	0.068	0.082

4. CONCLUSIONS

The application of ANNs for the prediction of carbon and bromine concentrations in THMs of potable water in WTPs and distribution terminals in Athens was demonstrated. The final optimized MLPs and RBF models used only one variable as input: the chlorination time assuming minimum variations in the other affecting parameters. The RMS error was used as the evaluation criterion of the models, while three samples sets were used (training, validation and test). In particular, the best ANNs model (MLPs with one hidden layer) gave 0.056, 0.039 and 0.060 respectively. It has been shown that under special circumstances, ANNs modeling approach is suitable for estimating the real carbon and bromine concentrations in chlorinated waters.

REFERENCES

- Bo L., JiuHui Q., HuiJuan L. and Xu Z., (2008), Formation and distribution of disinfection by-products during chlorine disinfection in the presence of bromide ion, *Chinese Sci. Bull.*, **53**(17), 2717-2723.
- Cancho B. and Ventura F., (2005), Optimisation of methods for the determination of DBPs, *Global Nest J.*, **7**(1), 72-94.
- Chowdhury S., Champagne P. and McLellan P. J., (2009), Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review, *Sci. Total Environ.*, **407**, 4189-4206.
- EC Directive of the European Parliament and of the council 98/83/EC of 11 November 1998 on the Quality of water for human consumption, *Off. J. Eur. Commun.*, L330, 32-54.
- Farmaki E.G., Thomaidis N.S. and Efstathiou C.E., (2010) Artificial Neural Networks in water analysis: Theory and applications, *Int. J. Environ. An. Ch.*, **90**, 85-105.
- Gibbs M.S., Morgan N., Maier H.R., Dandy G.C., Holmes M. and Nixon J.B. (14-17 July 2003), Use of Artificial Neural Networks for Modelling Chlorine Residuals in Water Distribution Systems, *Proceedings of International Congress on Modelling and Simulation*, 789-794.
- Heller-Grossman L., Manka J., Limoni-Relis B. and Rebhun M., (1999), Formation and distribution of haloacetic acids, THM and TOX in chlorination of bromide-rich lake water, *Water Res.*, **27**(8), 1323-1331.
- Hua G., Reckhow D.A. and Kim J., (2006), Effect of Bromide and Iodide ions on the formation and speciation of disinfection byproducts during chlorination, *Environ. Sci. Technol.*, **40**, 3050-3056.

- May R.J., Dandy G.C., Maier H.R. and Nixon J.B., (2008), Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems, *Environ. Modell. Softw.*, **23**, 1289-1299.
- Milot J., Rodriguez M.J. and Sérodes J., (2002), Contribution of neural networks for modeling THM occurrence in drinking water, *J. Water Resour. Plan. Manage.*, **128**(5), 370-376.
- Mok K.M., Wong H. and Fan X.J., (2005), Modeling bromide effects on the speciation of trihalomethanes formation in chlorinated drinking water, *Global NEST J.*, **7**(1), 1-16.
- Nikolaou A.D., Golfinopoulos S.K. and Lekkas T.D., (2002), Formation of organic by-products during chlorination of natural waters, *J. Environ. Monitor.*, **4**(4), 910-916.
- Nikolaou A.D., Kostopoulou M. and Lekkas T.D., (1999), Organic By-Products of drinking water chlorination (Review Article), *Global Nest J.*, **1**(3), 143-156.
- Nikolaou A.D., Golfinopoulos S.K., Kostopoulou M.N. and Lekkas T.D., (2001), Investigation of the behaviour of haloketones in water samples, *Chemosphere*, **44**, 907-912.
- Rodriguez M. J., West J. R., Powell J. and Sérodes J. B., (1997), Application of two approaches to model chlorine residuals in Severn Trent Water Ltd (STW) distribution systems, *Water Sci. Technol.*, **36**(5), 317-324.
- Sohn J., Amy G. and Yoon Y., (2006), Bromide ion incorporation into brominated disinfection by-products, *Water Air Soil Pollut.*, **174**, 265-277.
- Symons J.M., Krasner S.W., Simms L.A. and Scilimenti M., (1993), Measurement of THM and precursor concentrations revisited: the effect of Bromide Ion, *J. Am. Water Works Assoc.*, **85**(1), 51-62.
- Ye B., Wang W., Yang L. Wei J. and E X., (2009), Factors influencing disinfection by-products formation in drinking water of six cities in China, *J. Hazard. Mater.*, **171**, 147-152.
- Ye B., Wang, W., Yang L. Wei J. and E X., (2011), Formation and modelling of disinfection by-products in drinking water of six cities in China, *J. Environ. Monitor.*, **13**, 1271-1275.
- Zhang G., Patuwo B. E. and Hu M. Y., (1998), Forecasting with artificial neural networks: The state of the art, *Int. J. Forecast.*, **14**, 35-62.