RESEARCH

Using water quality parameters to prediction of the ion-based trihalomethane by an artificial neural network model

**Ali Akbar Babaei · Yaser Tahmasebi Birgani ·
Zeynab Baboli · Heydar Maleki ·
Kambiz Ahmadi Angali**

Abstract Trihalomethanes (THMs) are the first disinfectant by-products in the drinking water distribution network and are classified as potential carcinogens. The presence of THMs in chlorinated water depends on the pH, water temperature, contact time between water and chlorine, type and dose of disinfection, bromide ion concentration, and type and concentration of natural organic materials (NOMs). In the present study, the formation of THMs was evaluated by six simple and easy water quality parameters and modeled by an artificial neural network (ANN) approach through five water distribution networks (WDNs) and the Karoun River in Khuzestan province. The results of this study that was conducted from October 2014 to September 2015 showed that THM concentration ranged in five WDNs, including Shoushtar, Ahvaz (2), Ahvaz (3), Mahshahr, Khorramshahr, and total WDNs through N.D.–9.39 µg/L, 7.12–28.60, 38.16–67.00, 17.15–90.46, 15.14–29.99, and N.D.–156, respectively. The concentration of THMs exceeded Iran and EPA standards in many cases in Mahshahr and Khorramshahr WDNs. Evaluation of R^2, MSE, and RMSE showed the appropriate correlation between measured and modeled THMs, indicating a reasonable ANN potential for estimating THM formation in water sources.

A. A. Babaei · Y. Tahmasebi Birgani
Department of Environmental Health Engineering, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

A. A. Babaei · Y. Tahmasebi Birgani · K. Ahmadi Angali
Environmental Technologies Research Center (ETRC), Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

Z. Baboli (✉)
Department of Environmental Health Engineering, Behbahan Faculty of Medical Sciences, Behbahan, Iran
e-mail: baboliz87@gmail.com

H. Maleki
Department of Environmental Health Engineering, School of Public Health, Student Research Committee, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

K. Ahmadi Angali
Department of Statistic and Epidemiology, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

Introduction

Water-related epidemics have considerably decreased by using chlorine as disinfection for drinking water resources in 1904. Researchers have discovered a new compound called trihalomethanes (THMs) in chlorinated disinfected water that forms within the reaction of chlorine and NOM, both of which are known to be precursors of THMs. Due to the low molecular weight of these precursors, they do not eliminate by

conventional water treatment units (Garcia-Villanova et al., 1997; Kim & Yu, 2005; Liang & Singer, 2003). The type and concentration of THMs depend on several factors, i.e., the type of chlorinated disinfectants, contact time, pH water, and concentration of added disinfectants, organic materials, remaining chlorine, and bromide ion (Khan et al., 2019; Xu et al., 2022).

Chloroform $CHCl_3$ is the most frequent compound among other THMs over 70% of THM compounds belong to chloroform in many cases that are classified as B_2 according to the international agency for research on cancer IARC. Effects of THMs on health are divided into acute and chronic effects. Skin lesions, allergic symptoms, and poisoning are acute complications, and cancer infection is a chronic complication (Bull, 1991). Due to the health effects of THMs, USEPA in 1979 imposed regulatory controls on the amount of THMs in drinking water. Accordingly, the maximum THM concentration in drinking water of 100 µg/l was considered the annual average, which was reduced to 80 µg/l in 1998 (EPA, 1998).

Researchers implement different models for the evaluation of variations of environmental pollutants for the better management of environmental resources. The use of these models is very complicated and requires a significant amount of field data for analysis. In addition, many of the statistical models consider the relationship between response and predicted variables as linear with normal distribution. However, evaluation of environmental issues is under influence of the numerous factors. Thus, the traditional models may not be practical and robust enough to solve the environmental issues. In other words, they have weak accuracy for nonlinear modeling relationships with many different variables. Artificial neural network (ANN) is capable of evaluating the complex nonlinear relationships with high accuracy (Hong et al., 2020). At the same time, the ANN technique is flexible enough and can reveal the hidden relationships among data. Therefore, it facilitates the modeling of nonlinear behavior. ANN is modeled on the biology of the human brain, in which millions of neurons are linked together to process different complex information (Aleboyeh et al., 2008). To model THM compounds, a suitable method is needed. The compounds that are affected by different factors and standard mathematical models are not capable of analyzing them. Due to the simplicity and strangeness of ANN for simulation, prediction, and modeling, many researchers used it (Aleboyeh et al., 2008; Elmolla & Chaudhuri, 2011). Recently, many scientific branches, including water engineering and biological and environmental sciences, used the neural network approach. For example, it is used to simulate and predict the concentration of different pollutants in the air, water, and earth (Alimissis et al., 2018; Ghaedrahmat et al., 2019; Maleki et al., 2019; Tahmasbi, 2019; Takdastan et al., 2019; Tinelli & Juran, 2019; Vasanthi & Kumar, 2019; Wu et al., 2011; Xu et al., 2022).

Recent studies showed that ANN is a vital tool for enhancing the performance of water resources management systems (Daosud et al., 2005; Khan et al., 2019; Rizal et al., 2022). This approach also describes the behavior of water quality parameters with higher accuracy than other methods, e.g., linear regression (Tinelli & Juran, 2019). Since surface water is one of the main drinking water sources in Iran, and the application of chlorination systems is the most frequent method of disinfection for drinking water, the formation of THM compounds increased in treated water. In recent years, drinking water resources in Khuzestan province, including Ahvaz (2), Ahvaz (3), Mahshahr, Khorramshahr, and Shoushtar water treatment plants and total WDNs confronted with a high level of pollution and water shortage crisis. Therefore, the potential of THMs formation is high during the water treatment, and the necessity of momentary management of THMs with robust tools like ANN is vital. In this study, ANN is used to predict the concentration of THMs in WDNs by the influencing parameters.

Material and methods

Site selection

In this descriptive, analytical, and cross-sectional study, quantitative and qualitative levels of THMs and their precursors were evaluated and modeled at water withdrawal points of the Karoun River, including Shoushtar, Ahvaz, Mahshahr, and Khorramshahr during 12 months of sampling from October 2014 to September 2015. The map of Khuzestan province and the sampling points along the Karoun River are presented in Fig. 1.
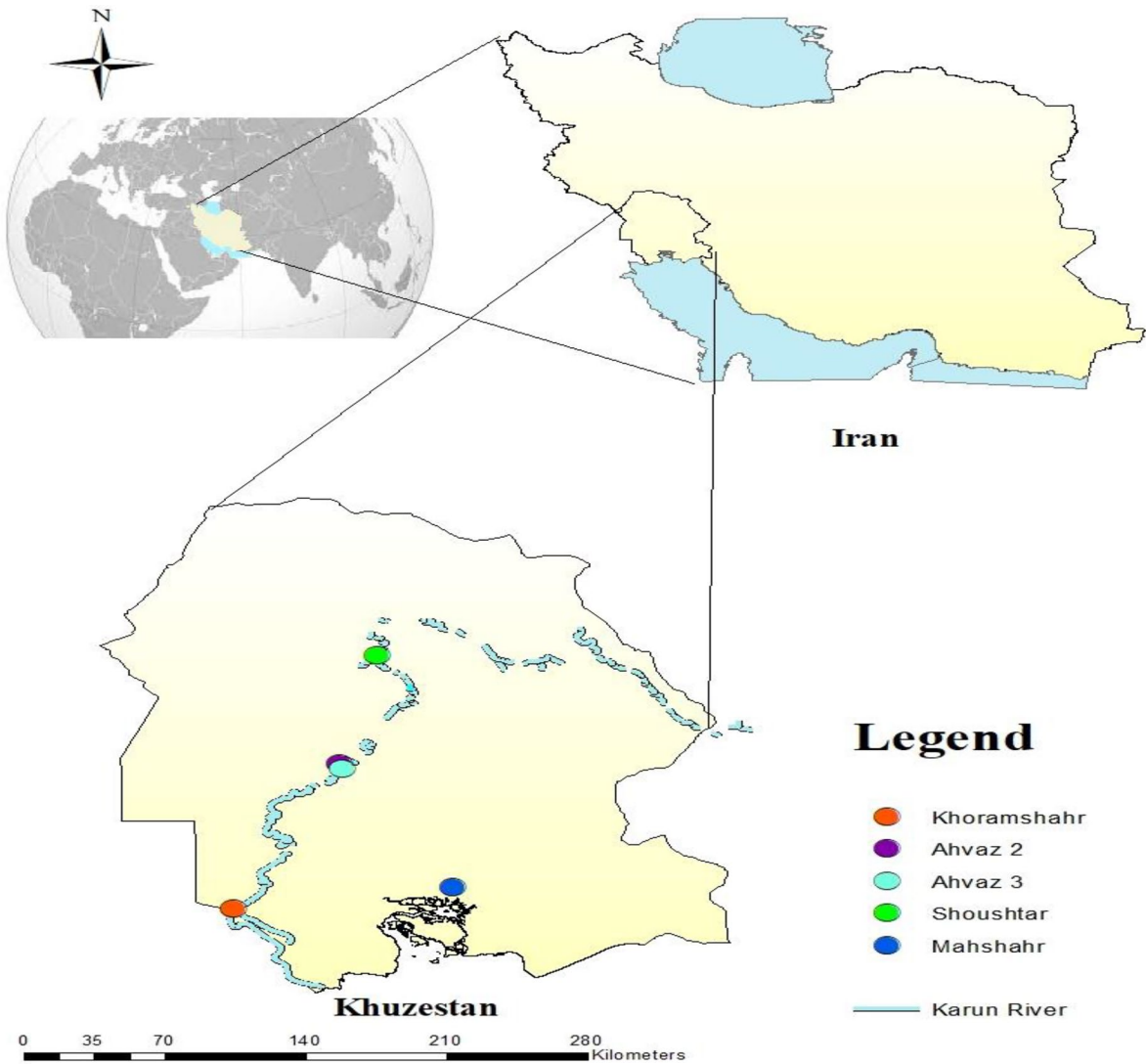
Fig. 1 The map of the studied locations

Water sampling procedure

The sampling points of the distribution networks were in three points of WDNs of Shoushtar, Ahvaz 2, Ahvaz 3, Mahshahr, and Khorramshahr. The parameters of DOC, pH, water temp, UV254, chlorine demand, and bromide were measured and entered into the software. In every sampling, one sample was recorded from raw water before the water treatment plant, and three samples were taken through the first, middle, and end of WDN. Sampling was performed twice each month. Water temperature, pH, free residual chlorine free residual chlorine (FRC), and UV_{254} were measured during sampling. Water temperature (°C) was detected by a digital thermometer, made in Germany, with an accuracy of around ± 0.05. FRC and pH were observed by digital chlorine/pH meter, made in Palintest company from England model Multi 1000. In this colorimetric method, red phenol and diethyl-p-phenylenediamine DPD tablets are used as pH and FRC markers, respectively. Observing ranges of the devices are 0–5 mg/l and 6.8–8.4 for FRC and pH meters, respectively.

The concentration of THMs was measured by an Agilent 6890 Gas Chromatograph USA with a micro-electron capture detector µECD. All stages of the study, including sampling, sample preparation and stabilization, and measurements, were conducted according to USEPA standards EPA-METHOD 551.1 in the hydrology laboratory of Ahvaz Water Treatment Plant No. 2, Health Faculty of Ahvaz Jondishapour University of Medical Sciences, and Iran Mineral Processing Research Center.

The samples taken by grab sampling were used to analyze dissolved organic carbon DOC and ultraviolet absorption at a wavelength of 254 nm. DOC samples were analyzed by a Shimadzu TOC Analyzer-VCSH Japan. The water samples collected for measuring bromine ions were analyzed by a Waters Alliance 2695 ion chromatography USA equipped with a Waters 2465 electrochemical detector USA.

Modelling by ANN

The gathered water data were processed in Excel. Then, the artificial neural network for predicting THM concentration was built in MATLAB. The inputs of this network were six parameters including DOC, pH, water temperature, UV_{254}, bromide, and chlorine demand that

affect THM formation, and its output was the concentration of THMs in water.

An artificial neural network is a system of simple processing parameters called neurons that take a number of inputs, give them weights, add them together, and add errors to them. Finally, they use the output values as the only numerically valued function, while they are obtained at the output of the neurons. In general, the neural network creates a non-linear relationship by creating a non-linear mapping between a set of input and output data, which after creating this relationship can be used to simulate the output based on the values of other inputs. Figure 2 shows the architecture of a neural network (Elmolla & Chaudhuri, 2011; Hong et al., 2020; Kadali et al., 2014; Rizal et al., 2022; Ubah et al., 2021). As shown in Fig. 2, the neural network model includes the input values that are multiplied by a set of weights. The results are aggregated in the neurons of the middle layer, and finally, the outputs are calculated using Eq. 1 (Maleki et al., 2019; Rizal et al., 2022; Ubah et al., 2021).

$$y = \sum W_{ij}^T . P + b \qquad (1)$$

In this equation, y is the simulated output, W_{ij}^T is the transpose of the weight of input i for neuron j, P is the input vector, and b is the bias.
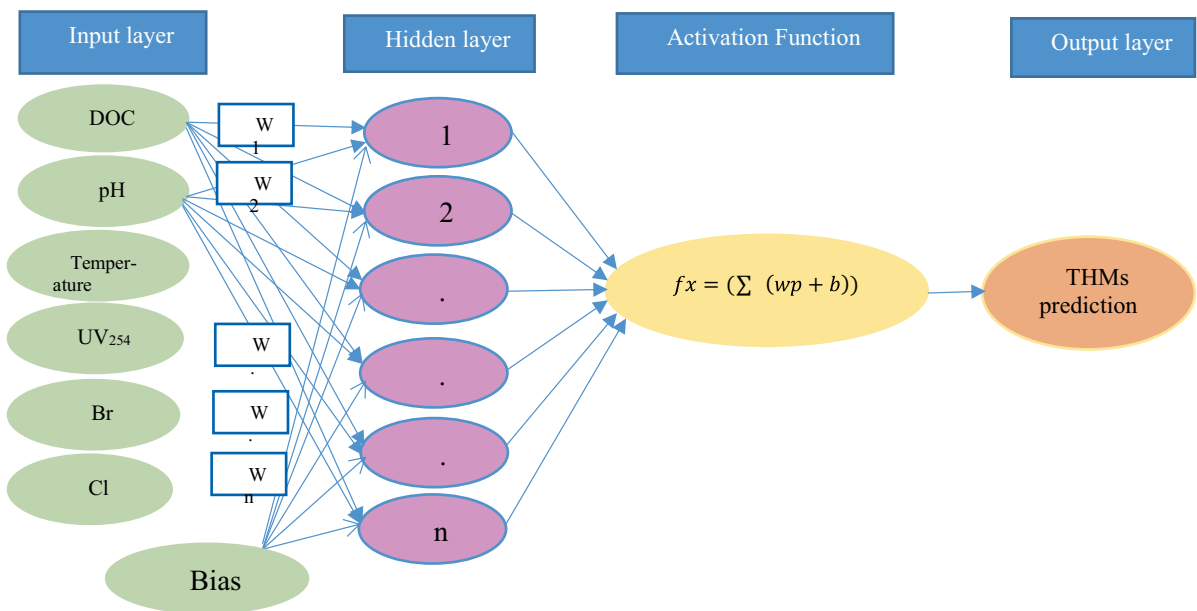

Fig. 2 Structural diagram of ANN

Based on this model, the parameters involved in air pollution in this research at any time and previous times related to it are selected as input and the value of that parameter is considered as output in the following times. First, the neural network was trained for a portion of the data to determine the best values for its weights and biases, and then tested for the desired time period. In this research, a MATLAB neural network toolbox is used to implement the network. Data preprocessing is necessary to improve neural network training. Input and output data should be scaled between 0.1 and 0.9.

The architecture of the artificial neural network is determined by the number of layers and the number of neurons in each layer, which is an important step for the development of the model to determine this architecture. To determine the number of different neurons in the hidden layer, the neural network is tested several times for different numbers of neurons (5–15 neurons in the hidden layer) and trained several times for each specific number of neurons, and the results were compared in terms of root-mean-square error (RMSE), mean squared error (MSE), and coefficient of determination R^2 as formulated in Eqs. 2–4. In these equations, N is the number of data items, P_i is the value predicted by the network, O_i is the value obtained from the experiments, and i is the subscript of data items (Rizal et al., 2022).

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(P_i - O_i\right)^2 \qquad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(P_i - O_i\right)^2} \qquad (3)$$

$$R^2 = \left(\frac{\sum_{i=1}^{N} \left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{N} \left(O_i - \overline{O}\right)^2} \sqrt{\sum_{i=1}^{N} \left(O_i - \overline{O}\right)^2}} \right)^2 \qquad (4)$$

Sensitivity analysis for data

To determine the relative importance of each input variable for the solution, a sensitivity analysis was performed using Eq. 5, Garson's equation (Elmolla

& Chaudhuri, 2011; Kadali et al., 2014; Rizal et al., 2022).

$$I_j = \frac{\sum_{m=1}^{m=N_h} \left(\left(\left|W_{jm}^{ih}\right| / \sum_{k=1}^{N_i} \left|W_{km}^{ih}\right| \right) \times \left|W_{mn}^{ho}\right| \right)}{\sum_{k=1}^{k=N_i} \left\{ \sum_{m=1}^{m=N_h} \left(\left(\left|W_{km}^{ih}\right| / \sum_{k=1}^{N_i} \left|W_{km}^{ih}\right| \right) \times \left|W_{mn}^{ho}\right| \right) \right\}} \qquad (5)$$

In this equation, I_j is the relative importance of the jth input for the output; N_i and N_h are the numbers of input neurons and hidden neurons, respectively; W_s denotes the connection weights; i, h, and o are the subscripts for input, hidden, and output layers, and k, m, and n are the subscripts for input, hidden, and output neurons, respectively (Elmolla & Chaudhuri, 2011; Maleki et al., 2019).

Results and discussion

### THM concentration at the points of water withdrawal from Karoun River

Changes in THM concentration along Karoun River from Shoushtar to Khorramshahr in different seasons and the standard concentration of THMs in water are illustrated in Fig. 3. As Fig. 3 shows, the concentration of THMs in the Karoun River gradually increases as it flows from Shoushtar to Khorramshahr. The THM value of Karoun River in Ahvaz is almost twice that of Shoushtar. While passing through Ahvaz, the river receives significant amounts of municipal and industrial wastewater, which increase its DOC and, therefore, THM level. As a result, in Mahshahr, the river has on average 1.5 times more THM than when it arrives at Ahvaz. In Khorramshahr, the DOC level of Karoun River is even higher than in Mahshahr, further increasing the potential for THM formation.

Both Mahshahr and Khorramshahr sections of the river are at risk of THM concentrations above the EPA standard. In all four points, the highest THM concentrations emerge during summer.

Variations and the relative importance of parameters affecting THM formation

The average values of six parameters affecting THM formation in the studied water samples DOC, pH,

Fig. 3 Changes in THM concentration in the studied sites of Karoun River
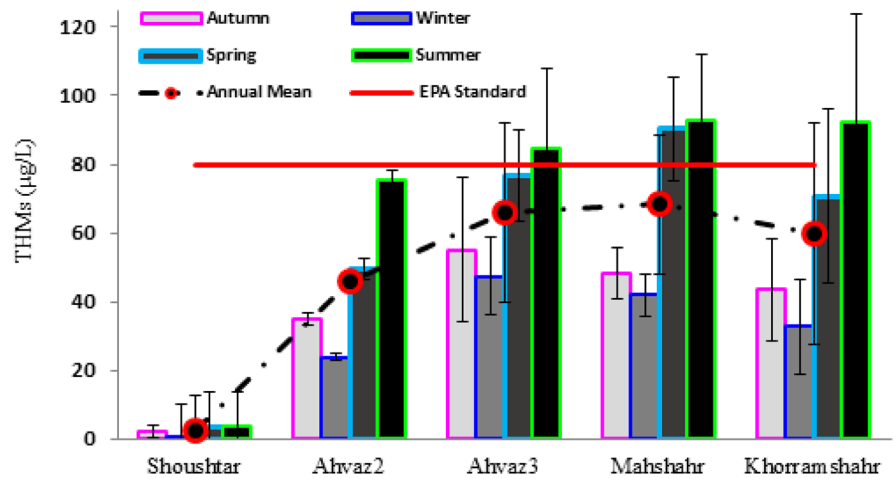
water temperature, UV_{254}, bromide, and chlorine demand are provided in Table 1.

Table 2 shows the relative importance of the input parameters including DOC, pH, water temperature, UV_{254}, bromide, and chlorine demand for predicting the concentration of THMs based on Eq. 5. As is evident, among the considered inputs, the factors most greatly affecting the concentration of THMs in Ahvaz Water Treatment Plant No. 2, Ahvaz Water Treatment Plant No. 3, the Shoushtar section, the Mahshahr section, the Khorramshahr section, and Karoun River as a whole are DOC, DOC, water temperature, chlorine demand, bromide, and DOC, respectively. The results of the present study are consistent with the results of other studies in terms of the most important parameters affecting the concentration of trihalomethanes using artificial neural network. Xu et al. showed that 4 tap water parameters, i.e., temperature, Cl_2, UVA_{254}, and pH are the most important parameters. Also, Hong et al. showed that the most important

Table 1 The concentration of parameters affecting the concentration of THMs in the studied sites

WDNs	Parameters	DOC (mg/L)	pH	Temp (°C)	UV_{254} (1/cm)	Br (µg/L)	Cl demand (mg/L)
Ahvaz2	Mean ($\pm$ SD)	1.14 ± 0.43	195 ± 18	0.016 ± 0.004	24.98 ± 5.75	7.92 ± 0.07	0.84 ± 0.15
	Max	1.27	8.07	33.9	0.028	232	3.6
	Min	0.57	7.81	15.5	0.01	156	2.4
Ahvaz3	Mean ($\pm$ SD)	1.14 ± 0.26	252 ± 37	0.015 ± 0.003	23.58 ± 5.75	7.64 ± 0.13	0.97 ± 0.14
	Max	1.66	7.95	35.6	0.026	321	4
	Min	0.75	7.35	15	0.011	200	2.6
Khoramshahr	Mean ($\pm$ SD)	0.64 ± 0.27	292 ± 100	0.026 ± 0.009	25.04 ± 5.57	7.87 ± 0.14	1.44 ± 0.32
	Max	2.21	8.18	34.7	0.048	552	3.5
	Min	1.02	7.65	16	0.013	196	2.5
Mahshar	Mean ($\pm$ SD)	0.55 ± 0.23	283 ± 30	0.033 ± 0.003	24.29 ± 5.24	7.98 ± 0.12	2.11 ± 0.19
	Max	2.58	8.19	36.1	0.043	350	4.5
	Min	1.76	7.8	16.6	0.026	220	3.5
Shoushtar	Mean ($\pm$ SD)	1.39 ± 0.38	162 ± 17	0.003 ± 0.001	17.84 ± 2.28	7.43 ± 0.12	0.090
	Max	0.135	7.72	24.6	0.004	210	1
	Min	0.027	7.18	14	0.002	138	1
Total WDNs	Mean ($\pm$ SD)	1.87 ± 0.88	236.8 ± 71.4	0.02 ± 0.01	23.15 ± 5.57	7.77 ± 0.24	1.09 ± 0.69
	Max	2.58	8.19	36.1	0.048	552	4.5
	Min	0.027	7.18	14	0.002	138	1

Table 2 Relative importance 100% of the inputs of the artificial neural network model for the concentration of THMs in the studied sites

WDNs Relative importance, %	DOC	pH	Water Temp	UV$_{254}$	Bromide	Chlorine Demand
Ahvaz2	18.81	18.75	14.34	17.87	11.52	18.69
Ahvaz3	30.09	16.03	9.05	8.62	13.46	22.72
Shoushtar	4.54	14.86	38.12	21.15	6.2	15.1
Mahshahr	23.63	5.85	23.08	11.55	5.5	30.36
Khorramshahr	19.33	14.46	10.25	10.63	23.84	21.46
Total WDNs	32.76	12.87	12.64	13.67	10.08	17.98

influencing parameters are pH, temperature, and UVA$_{254}$, followed by DOC and Br-, while NH^{4+}-N and Cl$_2$ residues played a very small role (Hong et al., 2020; Xu et al., 2022).

The DOC level of Karoun River gradually increases from Shoushtar to Khorramshahr. The DOC level of Ahvaz is almost twice that of Shoushtar. Because of the discharge of municipal and industrial wastewater of Ahvaz into the Karoun River, the DOC level of the river almost doubles before it reaches the Mahshahr section. In Khorramshahr, the discharge of industrial wastewaters, particularly from soap-making plants and fish and shrimp farms, further increases the DOC water level. According to the results of Table 2 and the artificial neural network model, DOC is the factor with the most significant impact on THM concentration in Ahvaz water treatment plant nos. 2 and 3 and Karoun River as a whole and is among the most important determinants of this parameter in other sections, particularly Mahshahr and Khorramshahr. Although the DOC level of Karoun is not high enough to justify control and removal strategies, the seasonal changes of the DOC level suggest that the river tends to exhibit higher average DOC levels during summer, when river discharge is markedly lower. Examining the changes in the UV$_{254}$ absorbance of Karoun, it was found that this parameter rises and falls with the river's DOC level. The results indicate that the only place where the amount of UV$_{254}$ has a significant effect on THM concentration is the Shuoshtar water distribution network, where the THM level is so low that no control measures are required.

The only place where pH significantly affects the concentration of THMs is Ahvaz water treatment plant no. 2. In general, THM concentration tends to increase with increasing pH. In a study by Kim et al., it was reported that the potential for THM formation increased with increasing pH, resulting in THM concentrations of 9.7, 20.7, and 41.6 µg/l at pH levels of 5.5, 7, and 7.9, respectively (Kim et al., 2003). A study by Liang and Singer has also shown that more THM tends to form at pH = 8 than at pH = 6 (Liang & Singer, 2003). Some studies have reported a linear relationship between pH and the formation of THMs (Garcia-Villanova et al., 1997). However, in our study, the effect of pH on the THMs concentration in the studied areas and Karoun River was not significant in general, which could be due to reasonably low variations in water pH over the length of Karoun River and during each year.

As shown in Fig. 3, the concentration of THMs in the drinking water of the studied networks usually exceeds the recommended level in summer and at the same time as the water temperature rises. Water temperature is an uncontrollable factor dictated by environmental conditions. Since rising temperature greatly accelerates the decrease of residual chlorine in water, it is challenging to maintain a specific chlorine concentration in water distribution networks during the hot months of the year. High doses of chlorine should be used to ensure sufficient residual chlorine in the water (Rodriguez & Serodes, 2001). According to Villanova et al. and Rodriguez et al., water temperature is one of the factors that significantly affect the formation of THMs in water (Rodriguez & Serodes, 2001; Uyak et al., 2005). One study reported that the total THM concentration in three water distribution systems was 34.2, 35.5, and 35.7 µg/l when water was more relaxed than 15 °C and increased to 64.2, 40.6, and 60.8 µg/l when the water had a temperature above 15 °C (Rodriguez & Serodes, 2001). Our results showed that water temperature had a notable impact on the formation of THMs in the Shoushtar and Mahshahr water distribution networks.

Examining the bromide ion concentration along Karoun, it was observed that this parameter also

gradually increases from Shoushtar to Khorramshahr. The results showed an increase in bromide ion levels due to a gradual increase in water EC throughout the river. This increase is much more pronounced in the Khorramshahr section, where the bromide ion concentration increases an average of three times as much as in the Ahvaz region. According to the results, the only place where the bromide ion concentration significantly affects THM concentration is the Khorramshahr water distribution network. One of the reasons for the high bromide concentration ion in the Khorramshahr segment of Karoun is its proximity to the Persian Gulf and the effects of the tides. Bromide ion is an inorganic precursor for the formation of disinfectant by-products. This ion is naturally present in the groundwater of coastal areas because of the seawater seepage. In chlorinated water, bromide ions are oxidized by hypochlorous acid HOCl, forming hypobromous acid HOBr, which reacts with natural organic matter to form disinfectant by-products. Many studies have shown that the simultaneous presence of bromide and chlorine in a drinking water source during the chlorination process can lead to bromine and bromochlorine by-products (Duong et al., 2003; Kawamoto & Makihata, 2004; Westerhoff et al., 2004).

In a study by Kampioti et al. on the Greek coastal city of Heraclion, they observed high concentrations of bromide ions in raw water 4.0–4.2 mg/L. They reported that the bromine components of THMs were dominant over the chlorine components of disinfectant by-products in drinking water (Kampioti & Stephanou, 2002). In the present study, the amount of residual chlorine was found to be the factor with the most significant effect on THM concentration in the Mahshahr water distribution network and also an essential determinant of this parameter in other places, including Ahvaz water treatment plants (2) and (3), Khorramshahr, and Karoun River as a whole. The significance of the effect of free residual chlorine concentration on THM concentration in the studied water distribution networks is directly associated with the dose of chlorine used.

Modeling and prediction of THM concentration

In the present study, the optimal numbers of hidden neurons in the artificial neural network model for sampling sites were tested by examining 5 to 15 neurons, and in each case, they were trained several

times, and the results were compared in terms of MSE, RMSE, and R^2.

Figure 4 shows the error of the models with different numbers of neurons for each sampling point. The best neuron has the lowest MSE and RMSE while having an R^2 of greater than 0.9. As can be seen, the network tries to find the best weights for the connections coming from every input and going into every neuron. At some point, the model has obtained the best possible weights, while producing worse results with more significant errors with any further change in the weight matrix (Maleki et al., 2019).

In this study, for all sampling points, the model inputs were six parameters affecting the concentration of THMs including DOC, water temperature, pH, bromide ion concentration, UV_{254} absorbance, and chlorine demand content of water, and the model output was the concentration of THMs. Accordingly, the model was built with six neurons including six water parameters in the input layer and one neuron in the output layer simulate THM concentrations. The hidden neurons for the Shoushtar, Ahvaz (2), Ahvaz (3), Mahshahr, and Khorramshahr water distribution networks were 7, 13, 8, 7, and 7 neurons, respectively (see Fig. 4). Neural network training was performed with 70% of the database to determine the best weights and biases; then, 15% of the database was used to validate the model and the last 15% of the database was used to test the ability of the model to predict and to simulate THM concentrations. Rizal et al. used three data sets to predict water quality parameters using ANN model; the percentages of data used for training, validation, and testing stages were 70, 15, and 15%, respectively (Rizal et al., 2022).

The results of the testing of the developed artificial neural network for all sites are presented in Fig. 5. Figure 5a shows the relationship between predicted and measured THM concentrations at all sites. The THM concentration values predicted by the network for each site are incredibly close to and highly consistent with the measured THM concentrations at those points. Accordingly, no difference was observed between the predicted and measured THM concentrations. The matching of the measured and predicted concentrations by the model indicates the proper performance of the ANN model in predicting the THM concentration based on the input parameters (Ghaedrahmat et al., 2019; Xu et al., 2022).
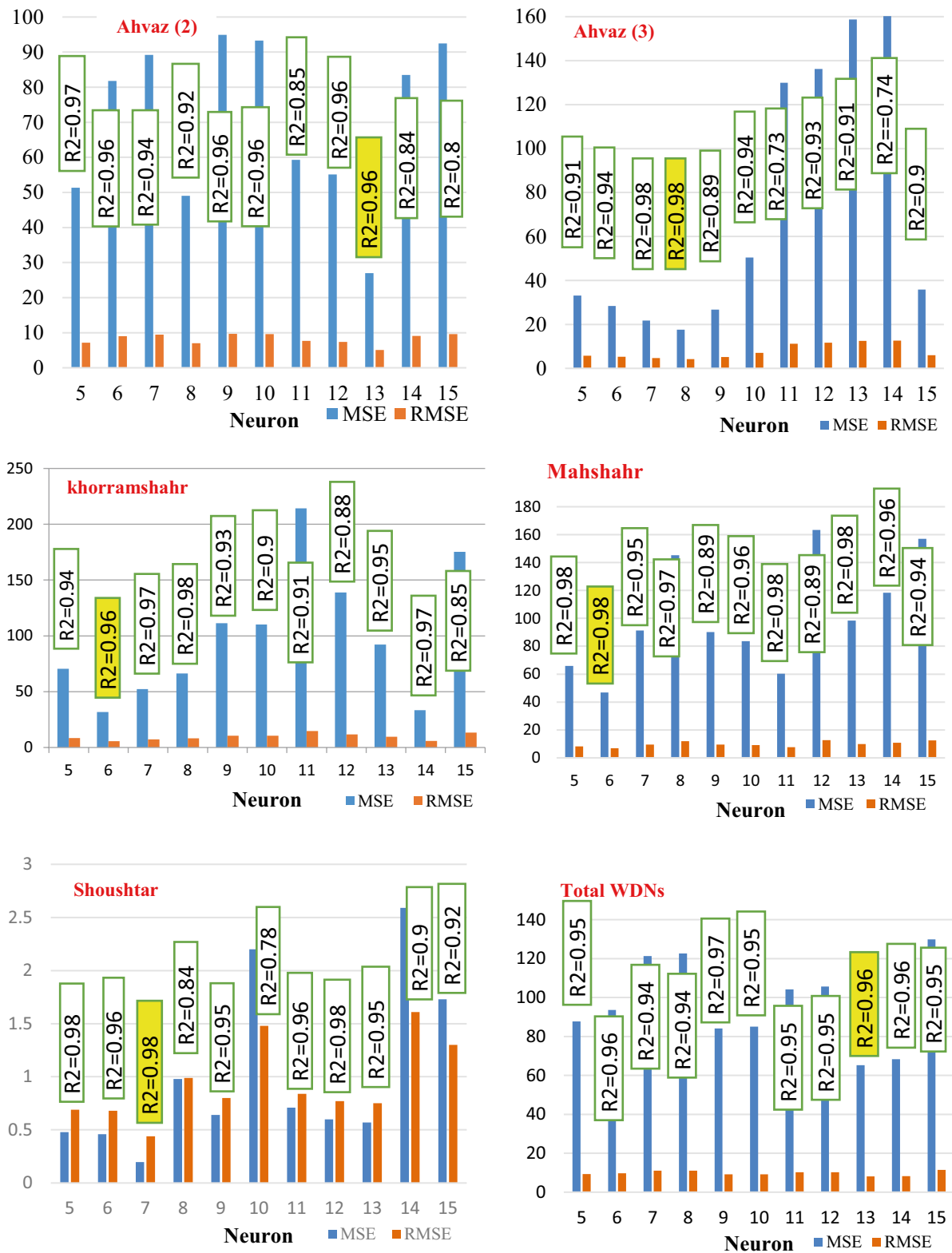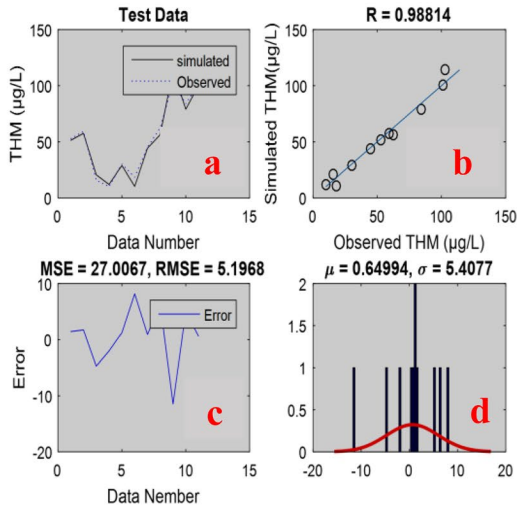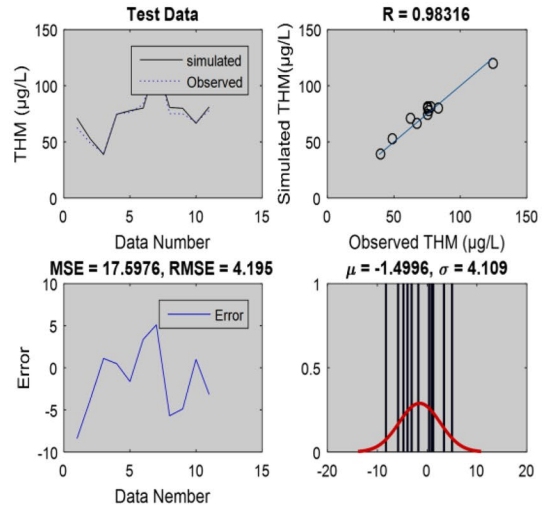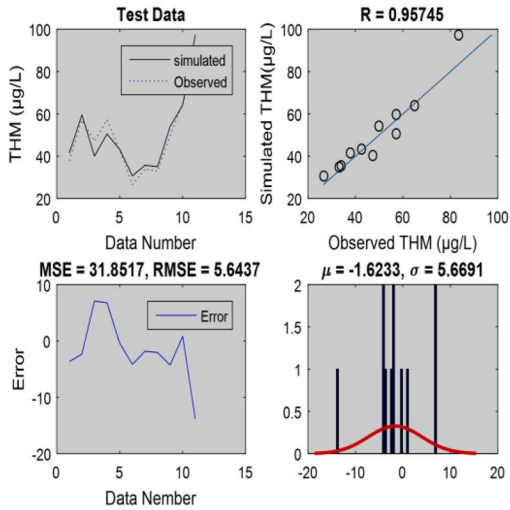
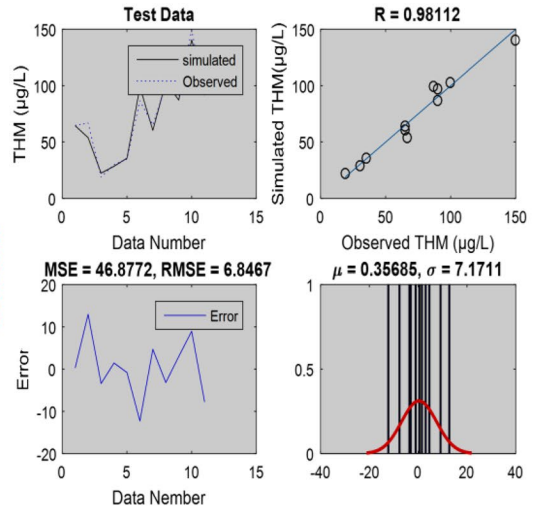**Fig. 4** Selection of the best number of hidden neurons in the ANN model for all sampling sites
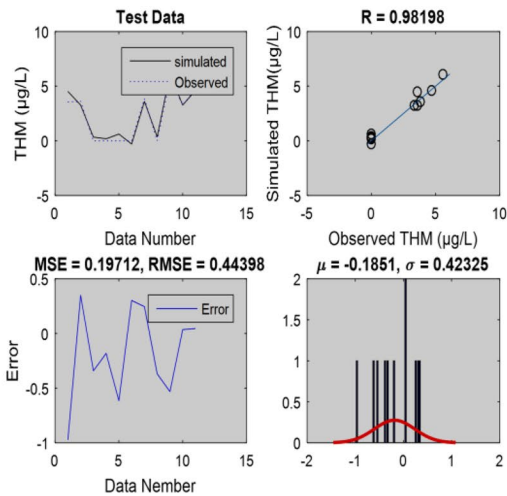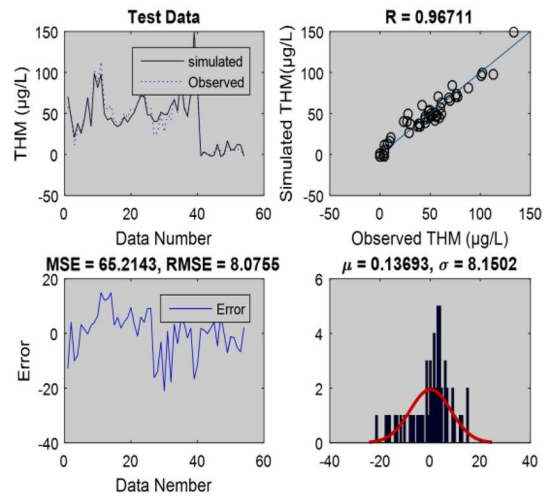
◄**Fig. 5** Testing stage performance of the developed artificial neural network for all sites. **a** The relationship between measured and predicted concentrations of THMs. **b** Scatterplots measured versus predicted concentrations. **c** The error values. **d** The histogram of error values

This consistency is shown more clearly in Fig. 5b, where the calculated error between the values of the simulated efficiencies and the measured efficiencies is displayed. The values simulated by the network are plotted against the measured values. In this diagram, most points are close to the bisector line representing $R^2 = 1$, indicating that R^2 is more significant than 0.95 for all sites. This is a pretty desirable level of consistency for environmental data.

Figure 5c shows the error of the simulated values relative to the measured concentrations. As can be seen, over 90% of the data have almost zero error, indicating a high level of accuracy. The high accuracy of the model is also reflected in of Fig. 5d, which shows the histogram of error values. As this diagram demonstrates, the error histogram has a normal-like distribution, with the data points being more frequently located around the zero error (Ghaedrahmat et al., 2019; Maleki et al., 2019; Takdastan et al., 2019; Vasanthi & Kumar, 2019). This indicates the excellent performance of the ANN in modeling and predicting the concentration of THMs.

The results of Xu et al. showed that RBF ANN (radial basis function ANN) models with a higher correlation coefficient (0.886–0.813) than linear regression models (LRM) with low R^2 (0.153 to 0.445) and low correlation coefficient (0.272–0.459) had better performance in predicting THMs. In Xu et al. study, the performance of the RBF ANN with 4 input parameters was better than the RBF ANN with 3 input parameters (Xu et al., 2022). The lower percentage of error and the higher coefficient of the ANN method in the present study compared to the study of Xu et al. indicate the better performance of the designed network and the more suitable selection of input parameters. Therefore, the correct selection of the effective input parameters and the precise design of the network will have a significant impact on its model performance (Rizal et al., 2022; Ubah et al., 2021; Xu et al., 2022).

Conclusion

The analysis of water samples taken from all points of water withdrawal from Karoun River between Shoushtar and Khorramshahr showed a gradual increase in the amount and concentration of water quality parameters that affect THM formation, including water temperature, DOC, bromide ion, and consequently a gradual increase in THM concentration in water distribution networks from Shoushtar to Khorramshahr. All studied water distribution networks showed much higher THM concentrations in hot seasons spring and summer than in cold seasons autumn and winter, with the difference being more pronounced in Ahvaz water treatment plant no. 3, the Mahshahr, and the Khorramshahr water distribution network. In the Shoushtar water distribution network, the concentration of THMs and their components is much lower than the standard levels of Iran and WHO guidelines. However, in Ahvaz water treatment plants (2) and (3), the Mahshahr water distribution network, and the Khorramshahr water distribution network, these concentrations occasionally exceed Iranian and WHO standards in spring and summer. DOC and free residual chlorine were the most significant impact on THM formation in all studied sections and Karoun River as a whole. The presence of free chlorine remaining in the studied sections can be partly attributed to the excessive use of chlorine in high doses, especially in the warm seasons. While THM formation is typically influenced by water pH, the higher the pH, the higher the THM formation; in this study, water pH had little effect on THM levels in the studied water networks due to small pH changes along the river Karoun and during each year.

The results showed that the developed artificial neural network could produce acceptably accurate predictions of THM concentration in the studied water distribution networks. This model can be used with reasonable accuracy to estimate THM concentrations, so it can help organizations and authorities to avoid costly THM measurements. Given the parameters included in the model, it can also facilitate the adoption of appropriate strategies to control THMs. The modeling results of this study suggest that the majority of the studied water treatment plants will benefit from more aptly chosen DOC and chlorine dose control measures for controlling THMs, although further technical and economic assessments are needed to decide which strategy would be more appropriate for and responsive to the situation.

Abbreviations   THMs: Trihalomethanes; NOMs: Natural organic materials; ANN: Artificial neural network; WDNs: Water distribution networks; IARC : International Agency for Research on Cancer; FRC: Free residual chlorine; DPD: Diethyl-p-phenylenediamine; DOC: Dissolved organic carbon

Declarations

Consent for publication   This is an experimental study on environmental samples (water samples).

Competing interests   The authors declare no competing interests.

References

Aleboyeh, A., Kasiri, M., Olya, M., & Aleboyeh, H. (2008). Prediction of azo dye decolorization by UV/H2O2 using artificial neural networks. *Dyes and Pigments, 77*(2), 288–294.

Alimissis, A., Philippopoulos, K., Tzanis, C., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment, 191*, 205–213.

Bull, R. J. (1991). *Health effects of disinfectants and disinfection by-products*: Amer Water Works Assn.

Daosud, W., Thitiyasook, P., Arpornwichanop, A., Kittisupakorn, P., & Hussain, M. A. (2005). Neural network inverse model-based controller for the control of a steel pickling process. *Computers & Chemical Engineering, 29*(10), 2110–2119.

Duong, H. A., Berg, M., Hoang, M. H., Pham, H. V., Gallard, H., Giger, W., & von Gunten, U. (2003). Trihalomethane formation by chlorination of ammonium-and bromide-containing groundwater in water supplies of Hanoi. *Vietnam. Water Research, 37*(13), 3242–3252.

Elmolla, E. S., & Chaudhuri, M. (2011). The use of artificial neural network (ANN) for modelling, simulation and prediction of advanced oxidation process performance in recalcitrant wastewater treatment. *Publisher in Tech*, 105–124.

EPA, U. (1998). National primary drinking water regulations: Disinfectants and disinfection byproducts. *Final Rule, 63*, 69390–69476.

Garcia-Villanova, R. J., Garcia, C., Gomez, J. A., Garcia, M. P., & Ardanuy, R. (1997). Formation, evolution and modeling of trihalomethanes in the drinking water of a town: II. *In the Distribution System. Water Research, 31*(6), 1405–1413.

Ghaedrahmat, Z., Vosoughi, M., Birgani, Y. T., Neisi, A., Goudarzi, G., & Takdastan, A. (2019). Prediction of O 3 in the respiratory system of children using the artificial neural network model and with selection of input based on gamma test, Ahvaz. *Iran. Environmental Science and Pollution Research, 26*(11), 10941–10950.

Hong, H., Zhang, Z., Guo, A., Shen, L., Sun, H., Liang, Y., & Lin, H. (2020). Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and grey relational analysis able to well predict trihalomethanes levels in tap water. *Journal of Hydrology, 591*, 125574.

Kadali, B. R., Rathi, N., & Perumal, V. (2014). Evaluation of pedestrian mid-block road crossing behavior using an artificial neural network (ANN). In *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems* (pp. 1911–1922).

Kampioti, A. A., & Stephanou, E. G. (2002). The impact of bromide on the formation of neutral and acidic disinfection by-products (DBPs) in Mediterranean chlorinated drinking water. *Water Research, 36*(10), 2596–2606.

Kawamoto, T., & Makihata, N. (2004). Distribution of bromine/chlorine-containing disinfection by-products in tap water from different water sources in the Hyogo Prefecture. *Journal of Health Science, 50*(3), 235–247.

Khan, M. Y. A., Hasan, F., & Tian, F. (2019). Estimation of suspended sediment load using three neural network algorithms in Ramganga River catchment of Ganga Basin. *India. Sustainable Water Resources Management, 5*(3), 1115–1131.

Kim, H.-C., & Yu, M.-J. (2005). Characterization of natural organic matter in conventional water treatment processes for selection of treatment processes focused on DBPs control. *Water Research, 39*(19), 4779–4789.

Kim, J., Chung, Y., Shin, D., Kim, M., Lee, Y., Lim, Y., & Lee, D. (2003). Chlorination by-products in surface water treatment process. *Desalination, 151*(1), 1–9.

Liang, L., & Singer, P. C. (2003). Factors influencing the formation and relative distribution of haloacetic acids and trihalomethanes in drinking water. *Environmental Science & Technology, 37*(13), 2920–2928.

Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Birgani, Y. T., & Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy, 21*(6), 1341–1352.

Rizal, N. N. M., Hayder, G., & Yusof, K. A. (2022). Water quality predictive analytics using an artificial neural network with a graphical user interface. *Water, 14*(8), 1221.

Rodriguez, M. J., & Serodes, J. -B. (2001). Spatial and temporal evolution of trihalomethanes in three water distribution systems. *Water Research, 35*(6), 1572–1586.

Tahmasbi, H. (2019). *Interatomic potentials based on artificial neural network: structural and thermal properties of matters.* Institute for Advanced Studies in Basic Sciences.

Takdastan, A., Samarbaf, S., Tahmasebi, Y., Alavi, N., & Babaei, A. A. (2019). Alkali modified oak waste residues as a cost-effective adsorbent for enhanced removal of cadmium from water: Isotherm, kinetic, thermodynamic and artificial neural network modeling. *Journal of Industrial and Engineering Chemistry, 78*, 352–363.

Tinelli, S., & Juran, I. (2019). Artificial intelligence-based monitoring system of water quality parameters for early detection of non-specific bio-contamination in water distribution systems. *Water Supply, 19*(6), 1785–1792.

Ubah, J., Orakwe, L., Ogbu, K., Awu, J., Ahaneku, I., & Chukwuma, E. (2021). Forecasting water quality parameters using artificial neural network for irrigation purposes. *Scientific Reports, 11*(1), 24438.

Uyak, V., Toroz, I., & Meric, S. (2005). Monitoring and modeling of trihalomethanes (THMs) for a water treatment plant in Istanbul. *Desalination, 176*(1–3), 91–101.

Vasanthi, S. S., & Kumar, S. (2019). Application of artificial neural network techniques for predicting the water quality index in the Parakai Lake, Tamil Nadu. *India. Applied Ecology and Environmental Research, 17*(2), 1947–1958.

Westerhoff, P., Chao, P., & Mash, H. (2004). Reactivity of natural organic matter with aqueous chlorine and bromine. *Water Research, 38*(6), 1502–1513.

Wu, W., Dandy, G., & Maier, H. (2011). Application of artificial neural networks to forecasting water quality in a chloraminated water distribution system.

Xu, Z., Shen, J., Qu, Y., Chen, H., Zhou, X., Hong, H., & Wu, F. (2022). Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. *Chemosphere, 286*, 131586.