Arabian Journal of Chemistry (2022) 15, 103794

King Saud University

Arabian Journal of Chemistry

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Comparative analysis of adaptive neuro-fuzzy inference system (ANFIS) and RSRM models to predict DBP (trihalomethanes) levels in the water treatment plant

Comfort N. Okoji [b], Anthony I. Okoji [a,c,*], Musa S. Ibrahim [b], Okpoko Obinna [b]

[a] *Landmark University, Chemical Engineering Department, SDG 7 (Affordable and Clean Energy Research Group), Nigeria*
[b] *Department of Biology and Forensic, Admiralty University, Ibusa, Delta State, Nigeria*
[c] *Department of Chemical Engineering, Landmark University, Omu-Aran, Kwara State, Nigeria*

Received 24 November 2021; accepted 10 February 2022
Available online 14 February 2022

KEYWORDS

Trihalomethane;
Adaptive neuro-fuzzy inference system;
Disinfection byproduct;
Response surface regression model;
Water treatment plants;
Neural networks

Abstract Since disinfection by-products are a growing concern, it is important to know their quantities in water treatment plants before they are released to the public. As a result, there is a requirement for constant monitoring of disinfection by-products (DBPs), which can have major consequences for human health and productivity. Consequently, in previous studies, several models for predicting disinfection byproduct formation in drinking water have been developed which were either linear/log-linear, hybrid or neural network (radial basis function). In this study, an adaptive neuro-fuzzy inference system (ANFIS) is proposed for predicting trihalomethane levels in real distribution systems. To train and verify the proposed model, 24 sets of data were used, including THMs levels (TCM, BDCM, DBCM and T-THM levels) and five parameters (pH, temperature, UVA_{254}, residual chlorine, and dissolved organic carbon). As compared to response surface modeling (RSRM) coefficient of determination, R^2 is between $0.727 < R^2 < 0.886$, average absolute deviation, AAD = 4.07–10.99 %), MAE = 0.01 – 0.978, and RMSE = 0.017 – 1.449. Further, ANFIS for THMs (T-THMs, TCM, BDCM, and DBCM) prediction consistently show higher regression coefficients between $0.956 < R^2 < 0.989$, average absolute deviation, AAD = 0.350 – 1.977 %), MAE = 0.002 – 0.133, and RMSE = 0.007 – 0.401, Consequently, based on the

* Corresponding author at: Landmark University, Chemical Engineering Department, Omu-Aran, Kwara State, Nigeria.
E-mail address: okoji.anthony@lmu.edu.ng (A.I. Okoji).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.arabjc.2022.103794

statistical indices obtained, ANFIS, on the other hand, proved to be effective for predicting the formation of THMs, and thus allowed improved DBPs monitoring in water treatment systems.

Nomenclature

AAD	Average absolute deviation	MFs	Membership functions
ANN	Artificial neural network	MRE	Mean regression error
ANFIS	Advanced neuro-fuzzy inference system	MS	Mass spectrometry
BDCM	Bromodichloromethane	MSE	Mean square error
BP	Backpropagation	R^2	Regression
DBCM	Dibromochloromethane	RBF	Radial basis function
DBPs	Disinfection by-products	RMSE	Root mean square error
DOC	Dissolved organic carbon	RSM	Response surface model
FIS	Fuzzy inference system	RSRM	Response surface regression model
FL	Fuzzy logic	THMs	Trihalomethanes
GC	Gas chromatography	TCM	Trichloromethane
HAAs	Haloacetic acids	TBM	Tribromomethane
HAN	Haloacetonitriles	UVA	Ultraviolet absorbance
MAE	Mean absolute error	WTPs	Water treatment plants

1. Introduction

When it comes to preventing waterborne infections, disinfection of water is crucial, but can also pose major health hazards, with its creation of disinfection by-products (DBPs) that could lead to birth defects, genotoxicity, and even cancer (Chaves, Guerreiro, Cardoso, Benoliel, & Santos, 2019, DeMarini, 2020, Kali et al., 2021, Kimura, Cuthbertson, Byer, & Richardson, 2019, Srivastav, Patel, & Chaudhary, 2020, C. Zhang and Lu, 2021). The most widely used disinfectants currently are chlorine and its compounds due to their availability, relatively low cost, and high oxidative strength (Chaukura et al., 2020, Ding et al., 2019, Du et al., 2017, Gopal, Tripathy, Bersillon, & Dubey, 2007). The vast majority of Nigeria's drinking water treatment plants (DWTPs) disinfect water with chlorine (Anthony, Ojemaye, Okoh, & Okoh, 2021).

Chlorine can form certain DBPs when it reacts with organic and inorganic precursors in water. In drinking water disinfection, there are a large number of DBPs (suspended solids), mostly chlorinated organic suspended solids, particularly trihalomethane (THMs such as trihalomethanes (T-THM) trichloromethane (TCM), bromodichloromethane (BDCM), and dibromochloromethane (DBCM)), haloacetic acids (HAA), and haloacetonitriles (HAN) (Ding et al., 2019, Kimura et al., 2019, Li and Mitch, 2018, Xiaoxiao Zhang et al., 2020, Zhou et al., 2019). Monitoring the levels of cytotoxic, genotoxic, and carcinogenic DBPs in drinking water is essential to the better control and protection of public health. Drinking water monitoring can be a challenging and time-consuming task that involves capital-intensive analysis with instruments such as gas chromatography (GC) and GC/mass spectrometry (MS) as well as complex pre-treatment procedures (Bond, Templeton, Rifai, Ali, & Graham, 2014, Li et al., 2016, Li and Mitch, 2018). Therefore, a special focus has been given to developing models that would estimate DBP formation, potentially providing a more effective alternative to standard monitoring (Chowdhury, 2009, Gougoutsa, Christophoridis, Zacharis, & Fytianos, 2016, Lin et al., 2018, Sohn, Amy, Cho, Lee, & Yoon, 2004, Uyak, Ozdemir, & Toroz, 2007). However, the conditions for DBPs formation in drinking water are rather complex, since water characteristics such as (dissolved organic carbon (DOC), ultraviolet absorbance at 254 nm (UVA_{254}), bromide ion concentration (Br pH, nitrite, ammonia, etc), and chlorination conditions (chlorine dose, temperature and reaction time) play a major role (Hong et al., 2020, Neale and Leusch, 2019, Sadiq and Rodriguez, 2004a, Sohn et al., 2004, Uyak et al., 2007). Since the complex connection makes it hard to predict DBPs formation with various factors, hence, this is the primary focus of DBPs studies. Models presented here were developed by chlorinating raw water, they were site-specific and based on careful laboratory evaluations (source water) or treated water from a treatment plant (Chowdhury, 2009, Singh and Gupta, 2012). SigmaPlot software was used to study the relationships between Br/DOC, Br/SUVA, Br/UVA, and BSF by Zheng et al. (2020). Water quality parameters and HAA formation were examined using Pearson correlations. Hence, linear models do not fit well the complex non-linear relationships among the various factors that influence the formation of DBPs (Kulkarni and Chellam, 2010, Singh and Gupta, 2012). Thus, alternative approaches to overcome these limitations are quite desirable. An artificial neural network (ANN) has been commonly regarded as a nonlinear estimator (Okoji et al., 2021a, 2021b, Singh and Gupta, 2012). Artificial intelligence may provide an attractive alternative to predict DBP formation in drinking water given the complex non-linear relationships between DBP formation and various factors, as well as the heterogeneity of drinking water contaminants (Xu et al., 2022). ANNs are artificial intelligence approaches that simulate the nervous system of a human brain through the use of the information processing network (Iliyas, Elshafei, Habib, & Adeniran, 2013). Artificial intelligence offers distinct advantages over linear regression since it can approximate any function with any accuracy, as well as its ability to learn, process parallel data, and resist noise (Iliyas et al., 2013)

Consequently, an adaptive neuro-fuzzy inference system (ANFIS) and the response surface method (RSM) are widely accepted as non-linear estimation techniques (Deng et al., 2021, Okoji et al., 2021a, 2021b, Sadiq and Rodriguez, 2004b, Singh and Gupta, 2012). As a result of the complicated nonlinear relationships between the formation of DBPs and various factors, with the heterogeneous nature of drinking water contaminants. Both ANFIS and RSM could provide an alternative approach to predict DBPs. In comparison with linear regression, ANFIS offers several important advantages, such as their ability to approximate any function accurately and learning, and data noisy resistance capabilities (Dolatabadi, Mehrabpour, Esfandyari, Alidadi, & Davoudi, 2018, Nabavi-Pelesaraei, Rafiee, Mohtasebi, Hosseinzadeh-Bandbafha, & Chau, 2019, Şahin and Erol, 2017). Despite the great potential of ANFIS and RSM for predicting DBPs formation owing to these advantages, only a few cases have been reported for DBPs prediction using neural networks. Based on the literature, six studies have been published on the prediction of DBPs with neural networks: an autoencoder-neural network (Peleato, Legge, & Andrews, 2018) and a hybrid genetic algorithm-based ANN (Moradi, Chow, Cook, Newcombe, & Amal, 2017) were applied in two, while BP ANNs were also used in the remaining four studies. (Deng et al., 2021, Kulkarni and Chellam, 2010, H. Lin et al., 2020, Singh and Gupta, 2012) and radial basis functions artificial neural network (Deng et al., 2021, Hong et al., 2020, H. Lin et al., 2020). The prediction of trihalomethane occurrence in tap water using RBF ANN considering the water quality parameters (Xu et al., 2022). The results of the study suggested that RBF ANNs are capable of establishing a relationship between THM production and the four parameters of water quality studied.

Despite the high potential of both ANFIS and RSM in DBP prediction, a review of the literature shows that little or no systematic studies have been conducted on their application. As a result of this study, adaptive neuro-fuzzy inference system (ANFIS) and response surface methods (RSM) have been applied to systematically explore the feasibility of predicting THMs such as trihalomethane (T-THM) which includes trichloromethane (TCM), bromodichloromethane (BDCM), and dibromochloromethane (DBCM) in actual production systems to develop new technology for predicting harmful substances in water. Hence, the present work contributes significantly to the fields of disinfectant by-product through prediction modeling and application of statistical regression with soft computing techniques to scientific problems: (ii) in addition, the interactive effects specific to different monitoring parameters were investigated via contour diagrams using ANFIS model code; and (iii) besides the application of ANFIS, as a modeling approach, an improved framework of RSRM modeling equation was also proposed for the disinfectant by-product (THMs).

2. Materials and method

These water treatment plants (WTPs) were built either through traditional methods (coagulation-sedimentation-filtration-chl orination) or through a chemical process (chlorine-chlorine, chlorine-ultraviolet). This study made use of codes to protect the corporate identities of the companies concerned. Water samples were collected from twenty-four (24) sources at each water treatment plant (WTP) in 2017 and 2018. To ensure there are no air bubbles and no headspace, the glass tubes labeled 50 mL were filled with water to overflowing and topped off with screw caps covered with Teflon septa. To limit the further production of disinfection by-products (DBPs) after collection, each vial was added 25 mg of ascorbic acid as a reducing agent (Benson et al., 2017).

The parameters of water quality such as pH, temperature, residual free chlorine, DOC, and UVA are replicated twice for each sample were determined based on a standard method (Apha, 1998). There is literature available for some characterization items that can assist in the operation of those items Liquid-liquidquid extraction and GC/ECD method were used according to EPA 551.1 (USEPA, 2006), to obtain the THMs (TCM, BDCM, DBCM, and TBM). The data of the DBPs and parameters of water quality are presented in Table S-1.

3. Modeling methodology

ANFIS and RSM predictive models were developed having the input variables such as reaction temperature, water pH (pH), UVA_{254}, residue chlorine (Cl_2), and DOC concentration. The output of the model was disinfection by-products (trihalomethanes – THMs) such as trihalomethanes (T-THM) trichloromethane (TCM), bromodichloromethane (BDCM), and dibromochloromethane (DBCM).

3.1. Response surface methods model (RSRM)

Considering various parameters for the prediction of DBP formations, the most representative conventional method is the multiple linear and log-linear regression (MLR). Consequently, response surface modeling (RSM) will be used to correct the independent variables with the dependent variables. The three most commonly used are Box-Behnken, central composite design (CCD), and factorial design methods. Five-level composite designs are used to merge axial points during experimental runs, while Box-Behnken and factorial composite designs are used on three-level designs. The Minitab 17 statistical software was used to evaluate the response surface method model. Additionally, the interactive impacts of the independent input variables on the (DBPs) were determined. Several independent variables were studied: reaction temperature (T), water pH (pH), UVA_{254}, residue chlorine (Cl_2), and DOC concentration.

According to Eq. (1), a quadratic model has been utilized to describe how the DBPs (Y) system responds to independent input variables. The input variables were reaction temperature (x_1), water pH (pH) (x_2), UVA_{254} (x_3), residue chlorine (Cl_2) (x_4), DOC concentration (x_5).

$$Y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \delta_4 X_4 + \delta_5 X_5 + \delta_{11} X_1^2 + \delta_{22} X_2^2 + \delta_{33} X_3^2 + \delta_{44} X_4^2 + \delta_{55} X_5^2 + \delta_{12} X_1 X_2 + \delta_{13} X_1 X_3 + \delta_{14} X_1 X_4 + \delta_{15} X_1 X_5 + \delta_{23} X_2 X_3 + \delta_{24} X_2 X_4 + \delta_{25} X_2 X_5 + \delta_{34} X_3 X_4 + \delta_{35} X_3 X_5 + \delta_{45} X_4 X_5 + \epsilon \quad (1)$$

where δ_0 is the offset term or model constant; $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$ are the linear or first-order terms; $\delta_{11}, \delta_{22}, \delta_{33}, \delta_{44}, \delta_{55}$ are the pure quadratic or squared terms; $\delta_{12}, \delta_{13}, \delta_{14}, \delta_{15}, \delta_{23}, \delta_{24}, \delta_{25}, \delta_{34}, \delta_{35}, \delta_{45}$ Here are the quadratic function's interactive terms; ϵ is the random error term that makes differences between experimental and predicted results uncertain. The correlation coefficient (R^2) value and was used to establish whether the quadratic model was acceptable.

3.2. ANFIS modeling of the DBPs

Fuzzy logic (FL) is a logical concept that converts linguistic terms into mathematical symbols by modifying logical operations. The FL method employs specific if-then rules and eliminates the need to apply crisp values in a physical process that involves qualitative or uncertain terms. However, obtaining a detailed and precise view of the problem is essential (Karkevandi-Talkhooncheh et al., 2017). Simulating a problem with the Fuzzy logic (FL) concept alone may not have resulted in the right result for several reasons, such as differences in expert opinions, lack of knowledge, or errors (Alahyar, Ansar, & Adel, 2017). Alternatively, artificial neural networks (ANNs) are capable of supervised learning based on data and can be utilized. Both FL and ANN techniques are combined into ANFIS. To perform the hybrid process, Jang (1993) proposes introducing membership functions (MFs) that are then optimized by applying ANNs and fuzzy if-then rules through a specific architecture called the Fuzzy Inference System (FIS). The ANFIS model for predicting DBPs in this study was proposed using the programming language of MATLAB 2013b and expected to give better outcomes when compared with RBF ANNs for nonlinearity purpose. For instance, reaction temperature (T), water pH, UVA_{254}, residue chloride (Cl_2), and DOC concentration were listed as input factors and THMs (TCM, BDCM, DBCM, and T-THM) as output factors. In addition, an adequate number of clusters must be provided for ANFIS to be able to provide sufficient prediction capabilities. The developed ANFIS model uses fuzzy c-means clustering to form a fuzzy inference system. The training method is dependent on how much data are assessed. To determine the number of rules and membership functions for the antecedents and consequents, the rule extraction process uses the Fuzzy c-means (FCM) clustering function or genfis3. Additionally, the Fuzzy c-means (FCM) clustering methods (genfis3) were utilized for optimizing the results. In this process, a set of rules has been extracted that model the data and generate an initial FIS for the ANFIS training procedure. MATLAB's genfis3 command creates a fuzzy C-means clustering structure that can be used to extract a set of rules and membership functions that model the training data of the fuzzy system using an ANFIS architecture. The number of clusters to use in modeling the data can be controlled with this function. In this study, 20 experimental datasets were used as training data (train data) for the neuron-based fuzzy inference system to train and test the ANFIS model. As a result of the model training process, the most accurate predictive model was discovered. The remaining 4 test data observations were used to create the best prediction model. Fig. 1 shows the ANFIS model architecture with five input independent variables and four output dependent variables.

3.3. Performance criteria

An analysis of the ANFIS and RSRM models was performed using statistical goodness-of-fit parameters. A good indicator for the correlation efficiency of the model is its coefficient of determination. Furthermore, some statistical models were used to estimate how much error there was between the anticipated and experimental values. As shown in Eqs. (2)–(6), they are the mean square error (MSE), the root mean square error (RMSE), the mean absolute error (MAE), and the average absolute deviation (AAD) (Jian, Hongdong, & Jingjing, 2011, Vu-Bac, Lahmer, Zhang, Zhuang, & Rabczuk, 2014).

4. Results and discussion

4.1. Predictions of DBPs with RSM models

Water quality models are built by selecting parameters by examining their correlations with THMs in step-wise regression. Consequently, in most cases, THMs are determined primarily by entering the input parameters into the models (Hong et al., 2016). Although DOC and UVA254 did not occur in raw water (water before disinfection), they appear to be closely related, according to the references (H. Lin et al., 2020). Organic and inorganic precursors for the formation of DBP still exist as indicators in the distribution system using DOC/UVA254.

Predictive model and response surface regression for each DBPs

$$T-THMs = -2234 - 25Temp + 763pH - 28956UVA_{254} - 807Cl_2 + 292DOC + 0.54Temp^2 - 53.4pH^2 + 2218UVA_{254}^2 + 13.3Cl_2^2 + 1.4DOC^2 - 0.9Temp*pH + 174Temp*UVA_{254} + 12.1Temp*Cl_2 - 0.87Temp*DOC + 3436pH*UVA_{254} + 67pH*Cl_2 - 37pH*DOC - 209UVA_{254}*Cl_2 - 290UVA_{254}*DOC + 3Cl_2*DOC \quad (7)$$

$$TCM = -2251 - 22Temp + 752pH - 27631UVA_{254} - 721Cl_2 + 270DOC + 0.52Temp^2 - 51.8pH^2 + 925UVA_{254}^2 + 12.9Cl_2^2 - 1.1DOC^2 - 1.3Temp*pH + 155Temp*UVA_{254} + 11.0Temp*Cl_2 - 0.60Temp*DOC + 3301pH*UVA_{254} + 59pH*Cl_2 - 35pH*DOC - 100UVA_{254}*Cl_2 - 120UVA_{254}*DOC + 3Cl_2*DOC \quad (8)$$

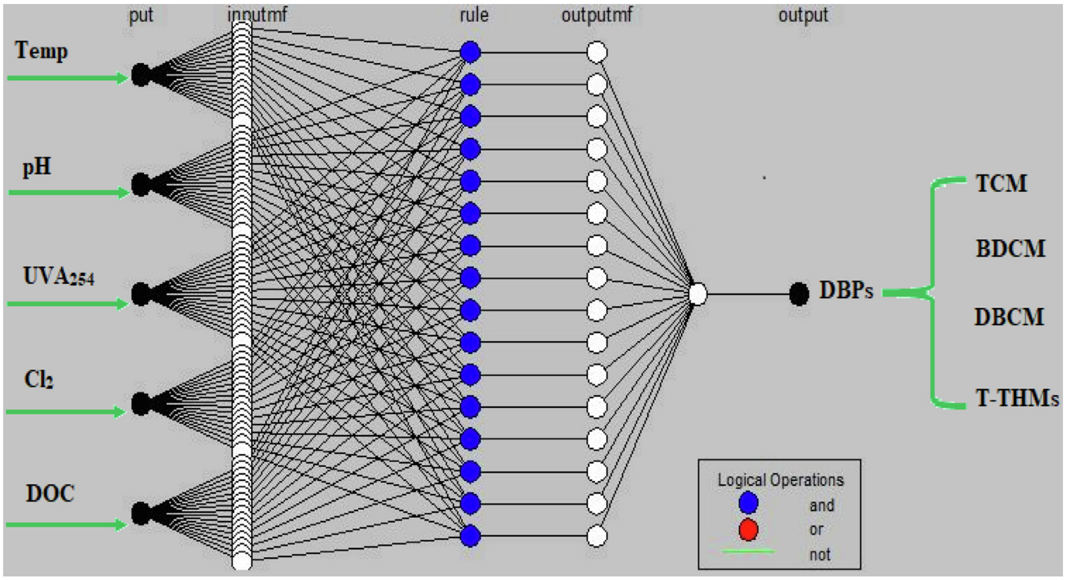

Fig. 1 ANFIS model architecture with five input variables (Okoji, Anozie, & Omoleye, 2022).

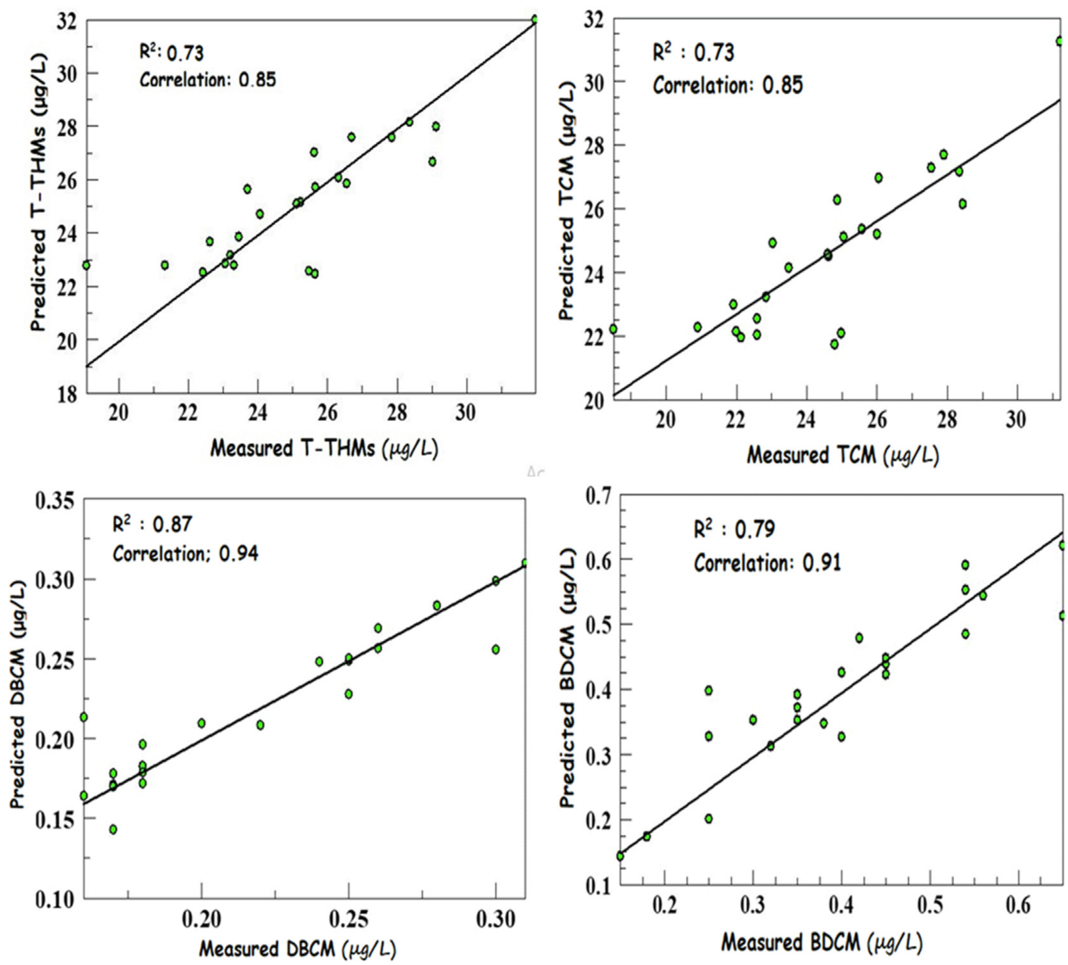

Fig. 2 R^2 and correlation for Model 1 to 4 of Disinfection by-product (DBPs) of water treatment plant.

$$DBCM = -33.6 - 2.37Temp + 19.5pH - 1013UVA_{254} - 36.9Cl_2 + 16.6DOC + 0.0141Temp^2 - 1.796pH^2 + 3550UVA_{254}^2 + 0.251Cl_2^2 + 0.936DOC^2 + 0.204Temp * pH + 14.16Temp * UVA_{254} + 0.467Temp * Cl_2 - 0.141Temp * DOC + 93.3pH * UVA_{254} + 3.48pH * Cl_2 - 1.84pH * DOC - 74.0UVA_{254} * Cl_2 - 120.8UVA_{254} * DOC - 1.23Cl_2 * DOC \quad (9)$$

$$BDCM = 51 - 1.09Temp - 8.1pH - 313UVA_{254} - 49.0Cl_2 + 5.5DOC + 0.0014Temp^2 + 0.23pH^2 - 2258UVA_{254}^2 + 0.136Cl_2^2 + 1.56DOC^2 + 0.123Temp * pH + 5.2Temp * UVA_{254} + 0.583Temp * Cl_2 - 0.121Temp * DOC + 42pH * UVA_{254} + 4.54pH * Cl_2 - 0.55pH * DOC - 35UVA_{254} * Cl_2 - 49UVA_{254} * DOC - 0.66Cl_2 * DOC \quad (10)$$

Based on the current research, TCM has a positive relationship with UVA_{254} aside from Temperature and pH considering model 2 (Eq. 3), while BDCM has a positive relationship with DOC aside from Temperature and pH with model 4 (Eq. 5). The result suggests that aromatic organic matter might play an important role in TCM formation. Conversely, the organic matter with UV absorbance has little effect on BDCM yields, and DOC is a more accurate marker for BDCM precursors. The significant high correlation between UVA_{254} and T-THM in model 1 (Eq. (7)) showed the dominant fraction of TCM to be responsible for this (Table 1). Additionally, TCM and T-THMs levels are negatively influenced by temperature (Temp), contrary to the prevailing perception that a high temperature will usually increase the yield of THMs (Model 1 and 5). (Hong et al., 2016). With respect to of this phenomenon, the temperature in this study captured both disinfection temperatures as well as seasonal changes. Therefore, the seasonal variation in organic precursors is likely responsible for the relationship between temperature and THMs.Table 2

Even though there is an apparent correlation of the input parameters and THMs level, the R^2 (the goodness fits) of the considered models is quite moderate (R^2 = 0.63–0.87, model 1–4) are moderately acceptable. Based on these results, it seems that the regression model may be suitable for predicting THM levels in water treatment system. However, natural treatment plant water disinfection is generally carried out under conditions in which some parameters, like temperature, do not change continuously, and disinfection time is difficult to measure. Additionally, water with high DOC and UVA_{254} levels may not be connected to DBPs as water with low DOC and UVA_{254} are often used in this type of laboratory simulation model. Consequently, regression is not a good method of predicting DBP levels in water treatment plants, making alternative methods fundamentally necessary.

Table 2 Evaluation models based on statistical analysis.

Equations	*Number*
$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_{i,pre} - Y_{i,exp}\right)^2}{\sum_{i=1}^{n}\left(Y_{i,exp} - Y_m\right)^2}$	(2)
$MSE = \frac{\sum_{i=1}^{n}\left(Y_{exp} - Y_{pre}\right)^2}{n}$	(3)
$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_{exp} - Y_{pre}\right)^2}{n}}$	(4)
$MAE = \frac{1}{n}\sum_{i=1}^{n}\lfloor\left(Y_{exp} - Y_{pre}\right)\rfloor$	(5)
$AAD = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{\lfloor\left(Y_{exp} - Y_{pre}\right)\rfloor}{Y_{exp}}\right)x100$	(6)

4.2. ANFIS models for predicting THMs level

The plots of the Gaussian Membership Functions (MF) for each of the five inputs (reaction temperature, water pH (pH), UVA254, residual chlorine (Cl_2), and DOC concentration) are displayed in Fig. 1. In Table S-2 are the results of the predictions of T-THMs, TCMs, BDCMs, DBCMs and ANFIS models. While, Fig. 3 presents a plot of the predicted values and measured values by ANFIS. The calculated R^2 for each DBPs from the model were 0.9780, 0.9894, 0.9553, and 0.9840 respectively. The near-unity values of R indicate agreement between predictions and experimental results. Consequently, R^2 values reflects 97.8 %, 98.9 %, 95.5 % and 98.4 %, an explanation can be given by the model for minimal vari-

Table 1 THM concentrations and corresponding water quality parameters in DWTP.

	Parameter	No. of data	Mean	Maximum	Minimum	SD	Mode
DBPs	TCM (μg/L)	24	24.58	31.21	18.5	2.82	22.59
	BDCM (μg/L)	24	0.4	0.65	0.15	0.13	0.35
	DBCM (μg/L)	24	0.22	0.31	0.16	0.051	0.18
	T-THMs (μg/L)	24	25.2	31.97	19.07	2.83	–
	Temperature (°C)	24	25.9	30.0	24.0	1.73	24.0
	pH	24	7.20	7.71	6.92	0.18	7.18
	UVA_{254} (μg/L)	24	0.015	0.038	0.001	0.01	0.005
	Cl_2 (μg/L)	24	0.33	1.51	0.05	0.32	0.28
	DOC (μg/L)	24	0.79	2.02	0.03	0.53	0.61

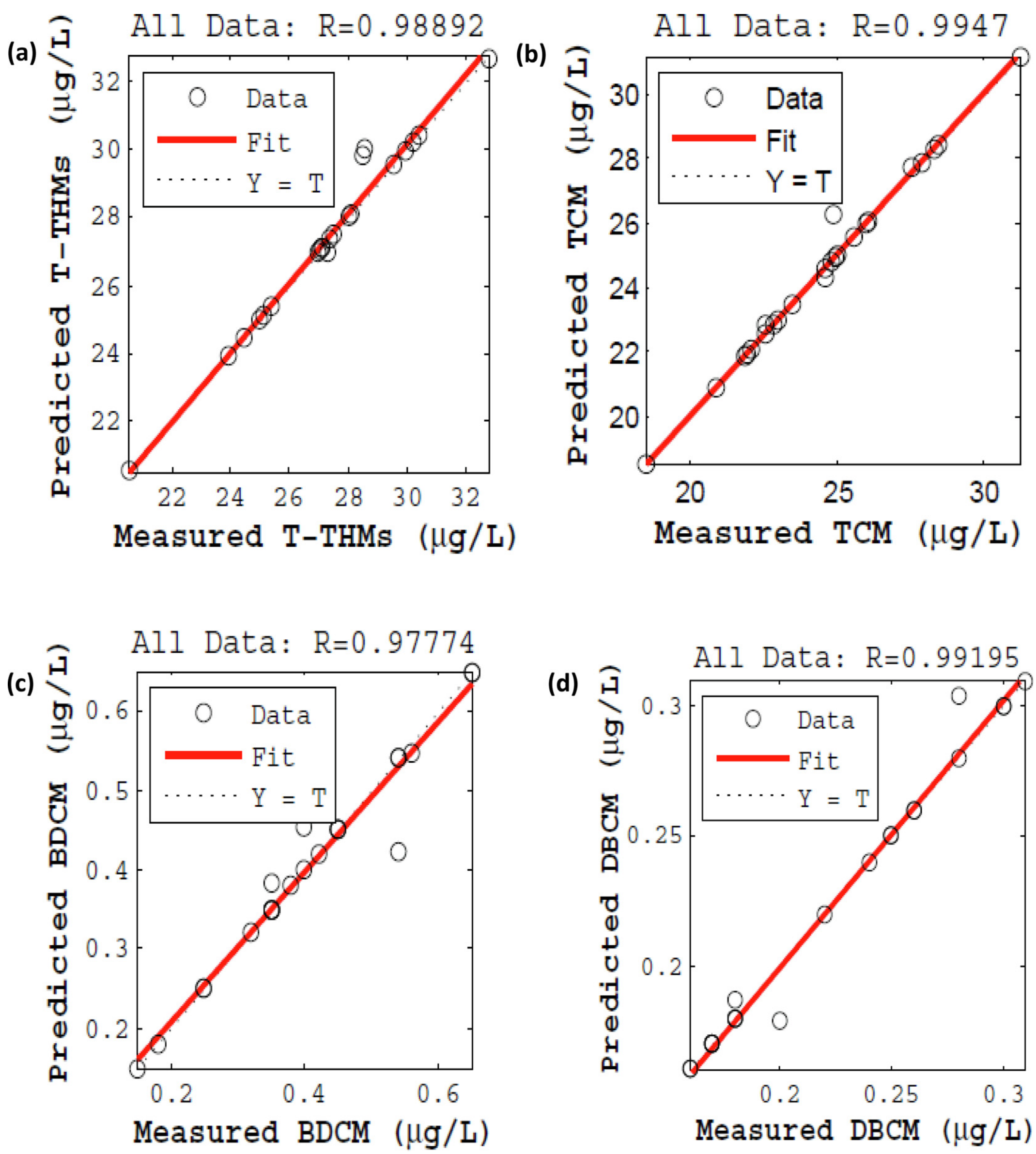

Fig. 3 Plot of predicted vs measured values (a) T-THMs (b) TCM (c) DBCM (d) BDCM.

ation between measured values and predictions. In addition, it can also be an indication that the model is well fitted with R^2 showing high value.

4.3. An evaluation of water quality characteristics

Consequently, most parameters of water quality showed seasonal variation. In general, temperatures, pH levels, DOC values, and UVA concentrations were higher in hot weather than in cold weather. Despite the fact that water treatment technique has not changed over time, the seasonal variation in above parameters in the water treatment plant is likely due mainly to variations in source water quality. Source water usually has a higher precipitation rate and more conducive conditions for bacteria and algae to grow at warm temperatures. According to the present study, water runoff and biogrowth in the plant are mostly responsible for the higher levels of DOC values, UVA_{254}, and pH. As a result of algal growth, both extracellular and intracellular organic matter can be released, contributing to DOC and UVA_{254} (Hong et al., 2013), a rise in water pH is caused by hydroxyl ions released by algal photosynthesis during the warm season.

4.4. Parameter interaction using ANFIS

Utilizing three dimensional surface and contour plots represented in Figs. 4 to 11, which allow visual observations, the interactions between the five parameters of water disinfection by-products (T-THM, TCM, BDCM and DBCM) were examined.

4.4.1. Effect of DOC and UVA_{254}

The contour diagram of the relationship between DOC and UVA_{254} on T-THMs and TCM are all significant as depicted in Figs. 4 and 5 respectively. The effects are consistent with

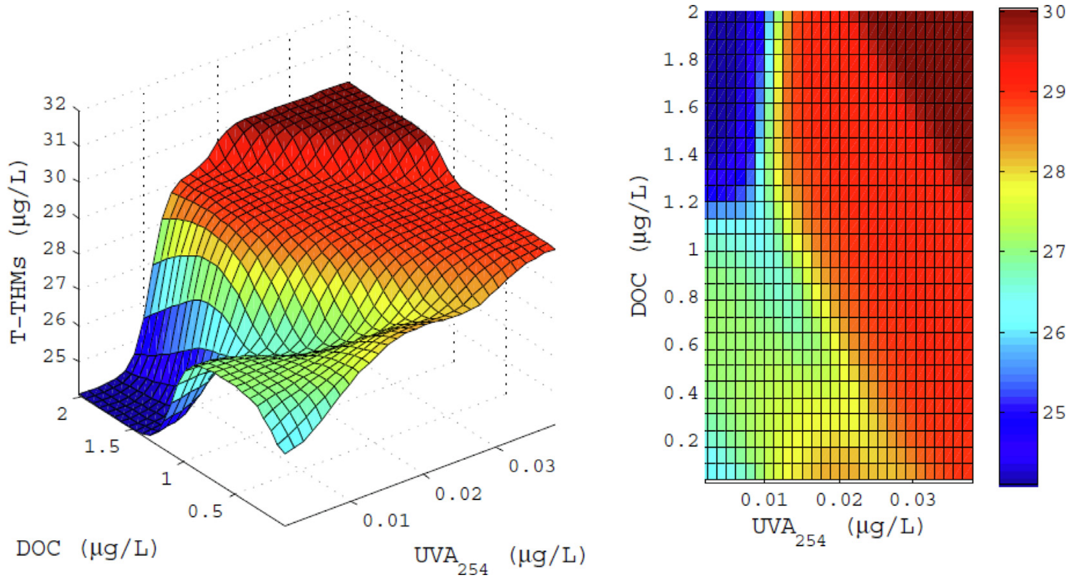

Fig. 4 The interaction effects between DOC (µg/L) and UVA (µg/L) on T-THMs (µg/L).

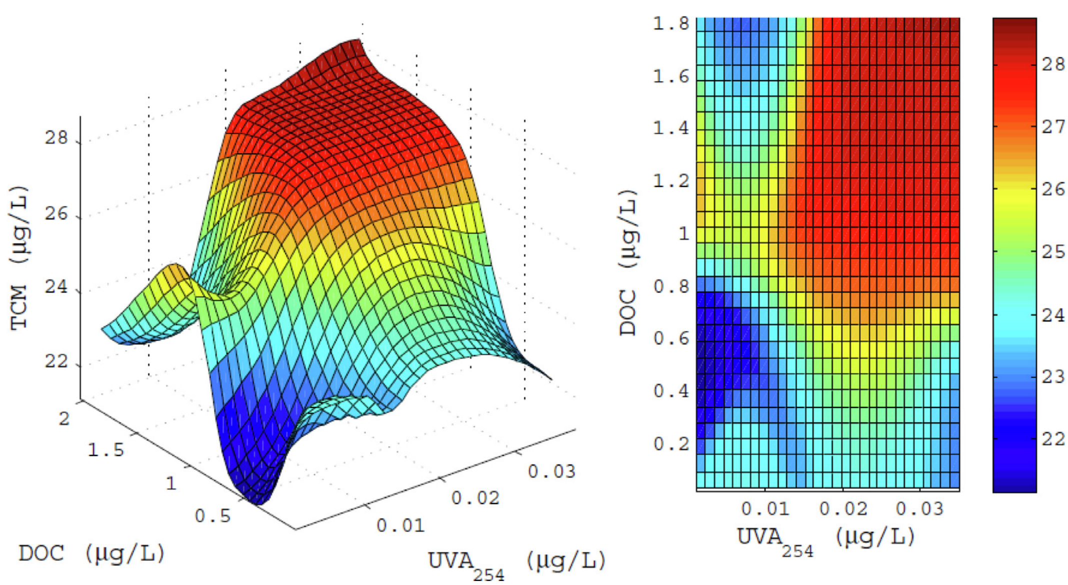

Fig. 5 The interaction effects between DOC (µg/L) and UVA (µg/L) on TCM (µg/L).

past works, (Hong et al., 2016). From displayed contour curve, it is very obvious that the organic matter (DOC) and UVA_{254} were the most important factors contributing to the presence of T-THMs and TCM in drinking water. Furthermore, Hong et al. (2020), found that DOC and UVA_{254} had equally important roles in the development of THMs in DWTP, particularly UVA254 as an indicator of the amount of aromatic carbon. However, according to studies conducted by (H. Lin et al. (2020)), UVA_{254} in DWTP showed a good sign for T-THMs or TCMs display. This claim also reflected on the contour diagram as showed in Figs. 6 and 7 respectively and which is rightly supported by the studies of Hu, Song, & Karanfil (2010) and Hong et al. (2013). Basically, bromide levels are considered to be an important factor in determining the formation of both BDCM and DBCM.

4.4.2. *Effect of pH and temperature (°C)*

Figs. 8 and 9 depicts the contour diagram of the relationship between pH and Temperature (°C) on T-THMs and TCM (µg/L) value. However, Fig. 3 shows how the parameters interacted, with a significant effect on the variation of T-THMs and TCM (µg/L) value of the DBPs. The T-THMs and TCM (µg/L) value of the DBPs were highest at the highest pH and Temperature (° C). As both the pH and Temperature (° C) increase, the DBPs value of the T-THMs and TCM (µg/L) increases as depicted in Figs. 8 & 9. This occurred when the temperature was higher than 25° C and increase in T-THMs and TCM were correlated with pH increase as well. Hence, the interaction between temperature and T-THMs and TCM emerged as a result of the collective effect of temperature and seasonal change in water quality on T-THMs and TCM values (Xiaoye

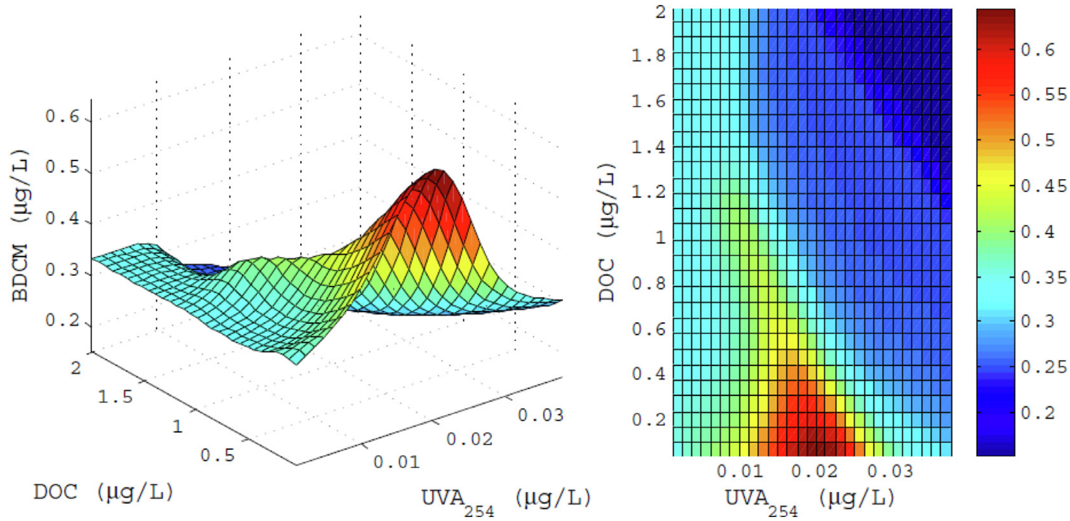

Fig. 6 The interaction effects between DOC (µg/L) and UVA (µg/L) on BDCM (µg/L).

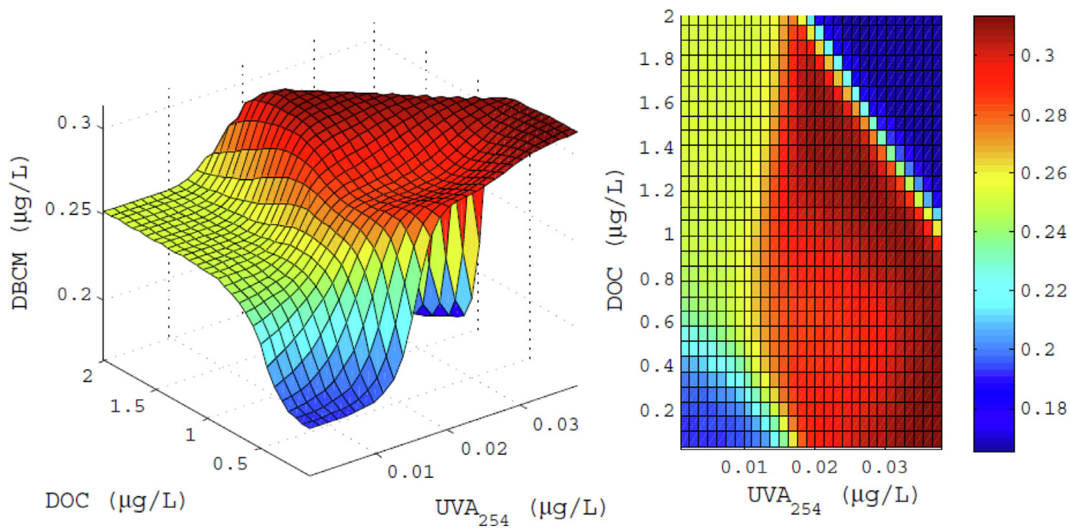

Fig. 7 The interaction effects between DOC (µg/L) and UVA (µg/L) on DBCM (µg/L).

Zhang, Tian, Zhang, Bai, & Zhang, 2019). As a result of water chlorination more commonly occurring in this dispensation, br- also has less significant effect on T-THMs and TCM concentrations. Consequently, Figs. 10 & 11 showed bromine THMs (BDCM and DBCM) were very low in quantities, whose formations are thought to be greatly affected by bromine concentrations. In general, the higher the Br/UVA_{254} ratio, the higher the bromine substitution factor (BSF). Considering that temperature can be used as a proxy for bromide level to some degree, and UVA is already included in the model, the key parameters influencing Br-THM formation were included in the ANFIS model. Hence, both ANFIS and RSMR models are able to predict BDCM and DBCM levels in drinking water treatment plant even without bromide levels.

Additionally to DOC and UVA_{254}, other environmental factors also affect DBP formation. THMs concentration was negatively impacted by temperature, while TCM concentration was positively influenced by it. On the other hand, this indicates that an increase in temperature may boost DBPs that are stable (such as THMs) and reduce those that are unstable. (HANs, HKs) (Hong et al., 2013). Therefore, the seasonal change in organic precursors may partly explain the connection among temperature and DBPs (THMs and DBCM). Despite prior studies indicating that the higher pH could increase the production of THMs and impede DBCM formation, there is no evidence that pH could affect concentration of most DBPs except BDCM (Fang, Ma, Yang, & Shang, 2010). However, this may be associated by an insignificant pH difference between the considered water samples (Table S-1).

4.5. Performance evaluation of developed models

The accuracy of RSRM and ANFIS models in predicting the DBPs value of the drinking water treatment plant was determined by assessing the R^2, mean square error (MSE), mean absolute error (MAE), average absolute deviation (AAD), and root mean square error (RMSE)

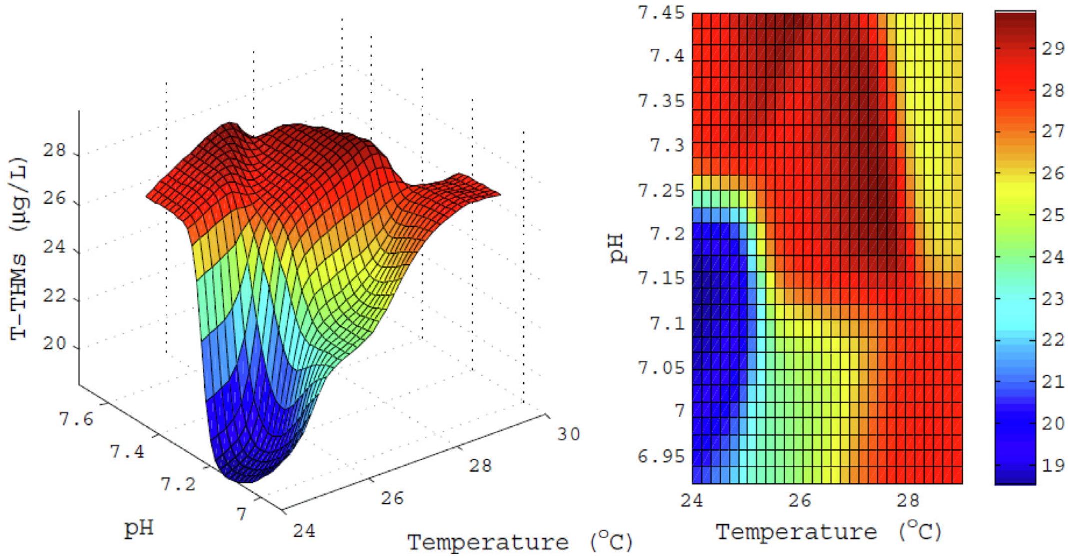

Fig. 8 The interaction effects between pH and Temp (° C) on T-THMs (μg/L).

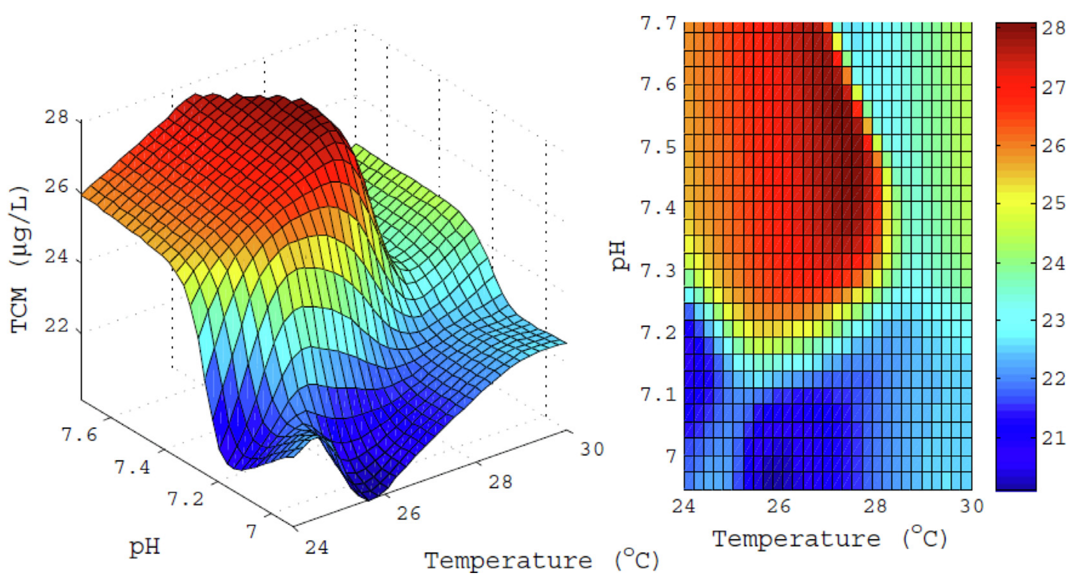

Fig. 9 The interaction effects between pH and Temp (° C) on TCM (μg/L).

Table 3 presents the results obtained, which supports the high values of R observed for RSMR and ANFIS. In addition, Figs. 2 and 3 demonstrate the correlation plots of experimental and predicted values for both models, with a R value close to 1 in both cases showing good correlation. Furthermore, both models had high R^2 values, each almost unity, and suggesting good fit with data. These two models were compared in terms of MSE and RMSE, which are measure of closeness between fitted lines and data points. In both MSE and RMSE, the values were low, indicating a good fit of the models. A model's accuracy and precision are measured using MAE and AAD. In general, the model performed better with lower statistical indicators. As a result, these values were evaluated for the developed models and are shown in Table 3. In light of the statistical indexes from Table 3, the ANFIS model outperformed RSRM, as indicated by its high AAD (2.067 to 10.989 %) for all the different DBPs evaluated. In the current study, it has been established that ANFIS provides better accuracy and precision in predictions compared to RSRM.

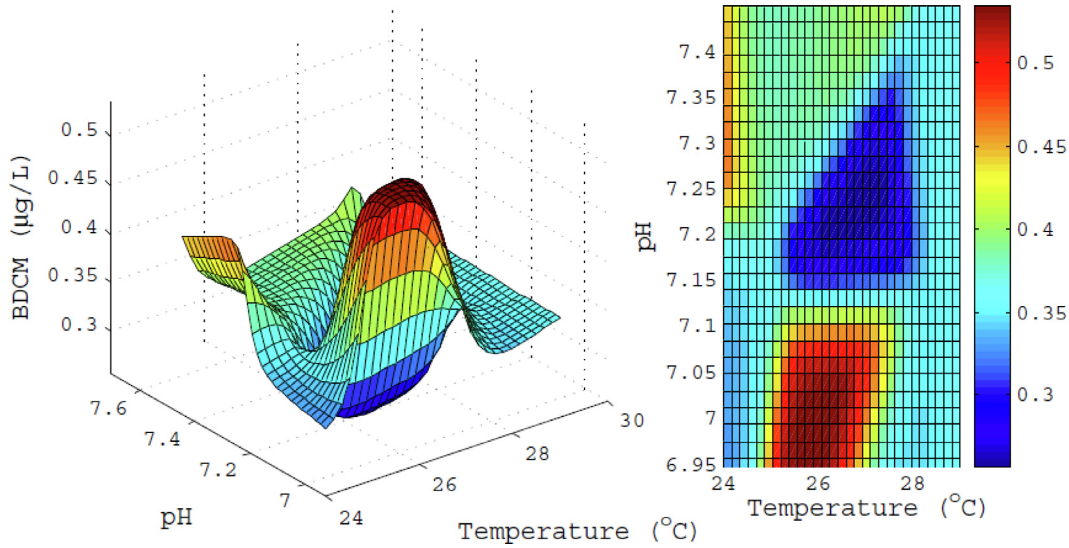

Fig. 10 The interaction effects between pH and Temp (° C) on BDCM (μg/L).

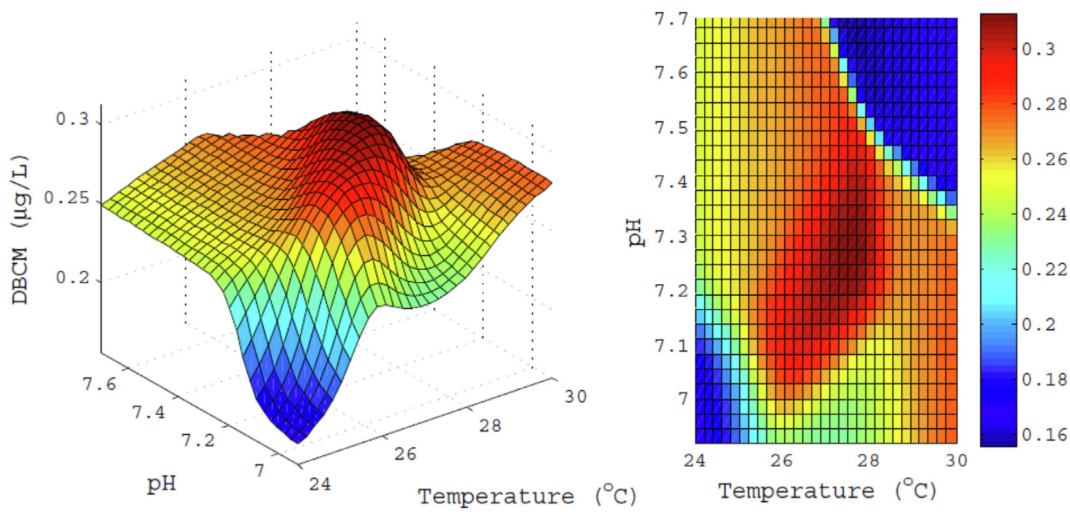

Fig. 11 The interaction effects between pH and Temp (° C) on DBCM (μg/L).

Table 3 Performance evaluation of RSRM and ANFIS models.

	RSMR				ANFIS			
	T-THMs	TCM	BDCM	DBCM	T-THMs	TCM	BDCM	DBCM
R	0.853	0.855	0.910	0.941	0.989	0.995	0.978	0.989
R^2	0.727	0.731	0.828	0.886	0.978	0.989	0.956	0.978
MSE	2.099	2.055	0.003	0.000	0.161	0.087	0.001	0.000
RMSE	1.449	1.433	0.055	0.017	0.401	0.294	0.028	0.007
MAE	0.978	0.977	0.040	0.010	0.133	0.087	0.009	0.002
AAD	4.067	4.171	10.989	4.982	0.468	0.350	1.977	0.973

5. Conclusions

In this study, the capability of the RSRM and ANFIS model in predicting DBPs value of the drinking water treatment plant was determined and has been evaluated. The performance analysis of the models are done by comparing the predicted value with the measured data. The accuracy of the model were determined using mean square error (MSE), mean absolute error (MAE), average absolute deviation (AAD), and root mean square error (RMSE). The main observations of the investigation are summarized as follows:

1. The evaluation of ANFIS model with site measured data indicates a good agreement with mean absolute error (MAE) (002 < MAE < 0.133) and (0.956 < R^2 < 0.989). The R2 tending to a unit (1), showed that ANFIS has the capability to predict a nonlinearity relationships between the inputs such as reaction temperature, water pH (pH), UVA_{254}, residue chlorine (Cl_2), and DOC concentration and Disinfectants by-products
2. The use of RSRM model with site measured data indicates an agreement with MAE (010 < MAE < 0.978) and (0.727 < R^2 < 0.886). The R^2 tending to a unit (1), showed that RSRM also has the capability to predict a nonlinearity relationships but with less efficient compared to the results of ANFIS model.
3. The results of the parameters interaction analysis show that the DOC and UVA_{254} on T-THMs and TCM are all significant, while T-THMs and TCM (μg/L) value of the DBPs were highest at the highest pH and Temperature (^{o} C). But low pH and temperatures impede the formation of DBCM.

It has been shown in this study that ANFIS is more capable of capturing complex and non-linear relationships regarding THMs and further enhancements in prediction accuracy compared to RSRM, also better compared to the previously studied RBF ANN (). ANFIS provided a significant opportunity for the evaluation of water treatment system functions, disinfection process controls and DBP monitoring in the actual water supply system.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.arabjc.2022.103794.

References

Alahyar, D., Ansar, H., Adel, E., 2017. Multiobjective ideal design of ANFIS for modeling and optimization of disinfectant substance consumed in water treatment. Desalin. Water Treat. 84, 85–92.

Anthony, E.T., Ojemaye, M.O., Okoh, A.I., Okoh, O.O., 2021. Potentials of low-cost methods for the removal of antibiotic-resistant bacteria and their genes in low budget communities: A review. J. Water Process Eng. 40, 101919.

Apha, A., 1998. Wef. Standard methods for the examination of water and wastewater, 21, 1378.

Benson, N.U., Akintokun, O.A., Adedapo, A.E., 2017. Chlorination by-products (CBPs) in drinking water and evaluation of potential health risks of long-term exposure in Nigeria.

Bond, T., Templeton, M.R., Rifai, O., Ali, H., Graham, N.J., 2014. Chlorinated and nitrogenous disinfection by-product formation from ozonation and post-chlorination of natural organic matter surrogates. Chemosphere 111, 218–224.

Chaukura, N., Marais, S.S., Moyo, W., Mbali, N., Thakalekoala, L. C., Ingwani, T., Nkambule, T.T., 2020. Contemporary issues on the occurrence and removal of disinfection byproducts in drinking water-A review. J. Environ. Chem. Eng. 8, (2) 103659.

Chaves, R.S., Guerreiro, C.S., Cardoso, V.V., Benoliel, M.J., Santos, M.M., 2019. Hazard and mode of action of disinfection by-products (DBPs) in water for human consumption: Evidences and research priorities. Comp. Biochem. Physiol. C: Toxicol. Pharmacol. 223, 53–61.

Chowdhury, A.R., 2009. Recent advances in heavy metals induced effect on male reproductive function-A retrospective. Al Ameen J. Med. Sci. 2 (2), 37–42.

DeMarini, D.M., 2020. A review on the 40th anniversary of the first regulation of drinking water disinfection by-products. Environ. Mol. Mutagen. 61 (6), 588–601.

Deng, Y., Zhou, X., Shen, J., Xiao, G., Hong, H., Lin, H., Liao, B.-Q., 2021. New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloketones in tap water. Sci. Total Environ. 772, 145534.

Ding, S., Deng, Y., Bond, T., Fang, C., Cao, Z., Chu, W., 2019. Disinfection byproduct formation during drinking water treatment and distribution: A review of unintended effects of engineering agents and materials. Water Res. 160, 313–329.

Dolatabadi, M., Mehrabpour, M., Esfandyari, M., Alidadi, H., Davoudi, M., 2018. Modeling of simultaneous adsorption of dye and metal ion by sawdust from aqueous solution using of ANN and ANFIS. Chemometrics Intell. Lab. Syst. 181, 72–78.

Du, Y., Lv, X.-T., Wu, Q.-Y., Zhang, D.-Y., Zhou, Y.-T., Peng, L., Hu, H.-Y., 2017. Formation and control of disinfection byproducts and toxicity during reclaimed water chlorination: a review. J. Environ. Sci. 58, 51–63.

Fang, J., Ma, J., Yang, X., Shang, C., 2010. Formation of carbonaceous and nitrogenous disinfection by-products from the chlorination of Microcystis aeruginosa. Water Res. 44 (6), 1934–1940.

Gopal, K., Tripathy, S.S., Bersillon, J.L., Dubey, S.P., 2007. Chlorination byproducts, their toxicodynamics and removal from drinking water. J. Hazard. Mater. 140 (1–2), 1–6.

Gougoutsa, C., Christophoridis, C., Zacharis, C.K., Fytianos, K., 2016. Assessment, modeling and optimization of parameters affecting the formation of disinfection by-products in water. Environ. Sci. Pollut. Res. 23 (16), 16620–16630.

Hong, H., Huang, F., Lin, H., Yu, H., Wang, F., Ding, L., 2013. Factors affecting formation of haloacetonitriles and haloketones during chlorination/monochloramination of Jinlan Reservoir water. Water Sci. Technol. Water Supply 13 (4), 1123–1129.

Hong, H., Song, Q., Mazumder, A., Luo, Q., Chen, J., Lin, H., Liang, Y., 2016. Using regression models to evaluate the formation of trihalomethanes and haloacetonitriles via chlorination of source water with low SUVA values in the Yangtze River Delta region, China. Environ. Geochem. Health 38 (6), 1303–1312.

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. J. Hydrol. 591, 125574.

Hu, J., Song, H., Karanfil, T., 2010. Comparative analysis of halonitromethane and trihalomethane formation and speciation in drinking water: the effects of disinfectants, pH, bromide, and nitrite. Environ. Sci. Technol. 44 (2), 794–799.

Iliyas, S.A., Elshafei, M., Habib, M.A., Adeniran, A.A., 2013. RBF neural network inferential sensor for process emission monitoring. Control Eng. Pract. 21 (7), 962–970.

Jian, M., Hongdong, Z., Jingjing, Y., 2011. Development and Application of Artificial Neural Networks. Electron. Des. Eng. 19, 24.

Kali, S., Khan, M., Ghaffar, M.S., Rasheed, S., Waseem, A., Iqbal, M. M., Zafar, M.I., 2021. Occurrence, influencing factors, toxicity, regulations, and abatement approaches for disinfection by-products in chlorinated drinking water: A comprehensive review. Environ. Pollut. 116950.

Karkevandi-Talkhooncheh, A., Hajirezaie, S., Hemmati-Sarapardeh, A., Husein, M.M., Karan, K., Sharifi, M., 2017. Application of adaptive neuro fuzzy interface system optimized with evolutionary algorithms for modeling CO2-crude oil minimum miscibility pressure. Fuel 205, 34–45.

Kimura, S.Y., Cuthbertson, A.A., Byer, J.D., Richardson, S.D., 2019. The DBP exposome: Development of a new method to simultaneously quantify priority disinfection by-products and comprehensively identify unknowns. Water Res. 148, 324–333.

Kulkarni, P., Chellam, S., 2010. Disinfection by-product formation following chlorination of drinking water: Artificial neural network models and changes in speciation with treatment. Sci. Total Environ. 408 (19), 4202–4210.

Li, C., Wang, D., Li, N., Luo, Q., Xu, X., Wang, Z., 2016. Identifying unknown by-products in drinking water using comprehensive two-dimensional gas chromatography–quadrupole mass spectrometry and in silico toxicity assessment. Chemosphere 163, 535–543.

Li, X.-F., Mitch, W.A., 2018. Drinking water disinfection byproducts (DBPs) and human health effects: multidisciplinary challenges and opportunities: ACS Publications.

Lin, H., Dai, Q., Zheng, L., Hong, H., Deng, W., Wu, F., 2020. Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: Taking haloacetic acids as a case study. Chemosphere 248, 125999.

Lin, J., Chen, X., Ansheng, Z., Hong, H., Liang, Y., Sun, H., Chen, J., 2018. Regression models evaluating THMs, HAAs and HANs formation upon chloramination of source water collected from Yangtze River Delta Region, China. Ecotoxicol. Environ. Saf. 160, 249–256.

Moradi, S., Chow, C.W., Cook, D., Newcombe, G., Amal, R., 2017. Estimating NDMA formation in a distribution system using a hybrid genetic algorithm. J.-Am. Water Works Assoc. 109 (6), E265–E272.

Nabavi-Pelesaraei, A., Rafiee, S., Mohtasebi, S.S., Hosseinzadeh-Bandbafha, H., Chau, K.-W., 2019. Comprehensive model of energy, environmental impacts and economic in rice milling factories by coupling adaptive neuro-fuzzy inference system and life cycle assessment. J. Cleaner Prod. 217, 742–756.

Neale, P.A., Leusch, F.D., 2019. Assessing the role of different dissolved organic carbon and bromide concentrations for disinfection by-product formation using chemical analysis and bioanalysis. Environ. Sci. Pollut. Res. 26 (17), 17100–17109.

Okoji, A.I., Anozie, A.N., Omoleye, J.A., 2021a. Evaluation of optimization techniques for predicting exergy efficiency of the cement raw meal production process. Cogent Eng. 8 (1), 1930493.

Okoji, A.I., Anozie, A.N., Omoleye, J.A., 2022. Evaluating the thermodynamic efficiency of the cement grate clinker cooler process using artificial neural networks and ANFIS. Ain Shams Eng. J.

Okoji, A.I., Anozie, A.N., Omoleye, J.A., Taiwo, A.E., Osuolale, F. N., 2021b. Energetic assessment of a precalcining rotary kiln in a cement plant using process simulator and neural networks. Alexandria Eng. J.

Peleato, N.M., Legge, R.L., Andrews, R.C., 2018. Neural networks for dimensionality reduction of fluorescence spectra and prediction of drinking water disinfection by-products. Water Res. 136, 84–94.

Sadiq, R., Rodriguez, M.J., 2004a. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Sci. Total Environ. 321 (1–3), 21–46.

Sadiq, R., Rodriguez, M.J., 2004b. Fuzzy synthetic evaluation of disinfection by-products—a risk-based indexing system. J. Environ. Manage. 73 (1), 1–13.

Şahin, M., Erol, R., 2017. A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games. Math. Computational Applications 22 (4), 43.

Singh, K.P., Gupta, S., 2012. Artificial intelligence based modeling for predicting the disinfection by-products in water. Chemometrics Intell. Lab. Syst. 114, 122–131.

Sohn, J., Amy, G., Cho, J., Lee, Y., Yoon, Y., 2004. Disinfectant decay and disinfection by-products formation model development: chlorination and ozonation by-products. Water Res. 38 (10), 2461–2478.

Srivastav, A.L., Patel, N., Chaudhary, V.K., 2020. Disinfection by-products in drinking water: Occurrence, toxicity and abatement. Environ. Pollut. 115474.

Usepa, 2006. National primary drinking water regulations: stage 2 disinfectants and disinfection byproducts rule: final rule. Fed. Reg. 71 (2), 388.

Uyak, V., Ozdemir, K., Toroz, I., 2007. Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs. Sci. Total Environ. 378 (3), 269–280.

Vu-Bac, N., Lahmer, T., Zhang, Y., Zhuang, X., Rabczuk, T., 2014. Stochastic predictions of interfacial characteristic of polymeric nanocomposites (PNCs). Compos. B Eng. 59, 80–95.

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.

Zhang, C., Lu, J., 2021. Optimizing disinfectant residual dosage in engineered water systems to minimize the overall health risks of opportunistic pathogens and disinfection by-products. Sci. Total Environ. 770, 145356.

Zhang, X., Chen, Z., Shen, J., Zhao, S., Kang, J., Chu, W., Wang, B., 2020. Formation and interdependence of disinfection byproducts during chlorination of natural organic matter in a conventional drinking water treatment plant. Chemosphere 242, 125227.

Zhang, X., Tian, Y., Zhang, X., Bai, M., Zhang, Z., 2019. Use of multiple regression models for predicting the formation of bromoform and dibromochloromethane during ballast water treatment based on an advanced oxidation process. Environ. Pollut. 254, 113028.

Zheng, L., Sun, H., Wu, C., Wang, Y., Zhang, Y., Ma, G., Hong, H., 2020. Precursors for brominated haloacetic acids during chlorination and a new useful indicator for bromine substitution factor. Sci. Total Environ. 698, 134250.

Zhou, X., Zheng, L., Chen, S., Du, H., Raphael, B.M.G., Song, Q., Hong, H., 2019. Factors influencing DBPs occurrence in tap water of Jinhua Region in Zhejiang Province, China. Ecotoxicol. Environ. Saf. 171, 813–822.