Chemosphere 248 (2020) 125999

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: Taking haloacetic acids as a case study

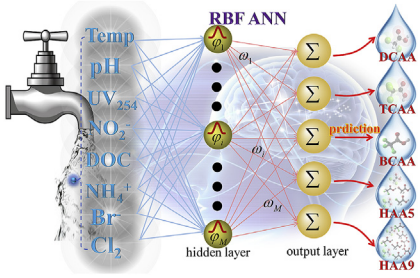
Hongjun Lin [a], Qunyun Dai [b], Lili Zheng [a], Huachang Hong [a, *], Wenjing Deng [c, **], Fuyong Wu [d]

[a] College of Geography and Environmental Sciences, Zhejiang Normal University, Jinhua, 321004, China
[b] Jinhua Maternal and Child Health Hospital, Jinhua, 321000, PR China
[c] Department of Science and Environmental Studies, The Education University of Hong Kong, Tai Po, N.T, Hong Kong
[d] College of Natural Resources and Environment, Northwest A&F University, Yangling, 712100, PR China

HIGHLIGHTS

- RBF ANN well captured the complex relationships between HAA and various factors.
- RBF ANN prediction showed high accuracy and allowed to further improvement.
- It is first report to systematically explore feasibility of RBF ANNs in DBPs prediction.
- The robust RBF ANN for HAA in this study paved a new way to predict DBPs in tap water.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 November 2019
Received in revised form
21 January 2020
Accepted 21 January 2020
Available online 22 January 2020

Handling Editor: Xiangru Zhang

Keywords:
Disinfection by-products
Multiple linear/log linear regression
Radial basis function
Artificial neural network
Haloacetic acids

ABSTRACT

Control of risks caused by disinfection by-products (DBPs) requires pre-knowledge of their levels in drinking water. In this study, a radial basis function (RBF) artificial neural network (ANN) was proposed to predict the concentrations of haloacetic acids (HAAs, one dominant class of DBPs) in actual distribution systems. To train and verify the RBF ANN, a total of 64 samples taken from a typical region (Jinhua region) in China were characterized in terms of water characteristics (dissolved organic carbon (DOC), ultraviolet absorbance at 254 nm (UVA_{254}), NO_2^--N level, NH_4^+-N level, Br^- and pH), temperature and the prevalent HAAs concentrations. Compared with multiple linear/log linear regression (MLR) models, predictions done by RBF ANNs showed rather higher regression coefficients and accuracies, indicating the high capability of RBF ANNs to depict complicated and non-linear relationships between HAAs formation and various factors. Meanwhile, it was found that, predictions of HAAs formation done by RBF ANNs were efficient and allowed to further improve the prediction accuracy. This is the first study to systematically explore feasibility of RBF ANNs in prediction of DBPs. Accurate predictions by RBF ANNs provided great potential application of DBPs monitoring in actual distribution system.

1. Introduction

As an inevitable part of drinking water preparation, disinfection

* Corresponding author.
** Corresponding author.
E-mail addresses: huachang2002@163.com (H. Hong), wdeng@ied.edu.hk (W. Deng).

https://doi.org/10.1016/j.chemosphere.2020.125999
0045-6535/

can kill pathogens and prevent waterborne diseases, but it may also cause serious problems to human health like birth defects, genotoxicty and even cancer risks due to formation of disinfection by-products (DBPs) (Li et al., 2014; Postigo et al., 2018; Regli et al., 2015; Sun et al., 2019; Wright et al., 2017; Wu et al., 2019; Zhang et al., 2019). At present, chlorine and its compounds are predominant disinfectants due to their low cost and strong oxidability (Ding et al., 2013; Du et al., 2017; Gopal et al., 2007; Sadiq et al., 2002). Chlorine can react with organic and inorganic precursors presented in water, so as to form manifold DBPs. Over 600 DBPs have been identified in disinfection of drinking water, a large part of which are chlorinated organic DBPs such as trihalomethane (THMs), haloacetic acids (HAAs), haloacetonitriles (HANs) and etc (Deng et al., 2014; Kimura et al., 2019; Li and Mitch, 2018; Zheng et al., 2020; Zhou et al., 2019). Considering their cytotoxicity, genotoxicity and cancer risk, monitoring DBPs in drinking water is essentially necessary for better control of them.

DBPs monitoring in drinking water is a time-consuming and laborious job, which involves expensive instruments analysis (e.g. gas chromatography (GC), GC/mass spectrometry (MS)) and complicated pre-treatment processes (Arbuckle Tye et al., 2002; Li et al., 2016; Li and Mitch, 2018). Hence, particular interest has been directed on development of models to estimate the formation of DBPs (Chowdhury et al., 2009; Hong et al., 2016; Lin et al., 2018; Sohn et al., 2004; Uyak et al., 2007), which may be an alternative for monitoring of DBPs in the field. However, DBPs formation in drinking water is rather complicated, which is mainly affected by water characteristics (dissolved organic carbon (DOC), ultraviolet absorbance at 254 nm (UVA_{254}), bromide ion concentration (Br^-), pH, nitrite, ammonia, and etc) and chlorination conditions (chlorine dose, temperature and reaction time) (Hong et al., 2017; Sadiq and Rodriguez, 2004; Uyak et al., 2007). The complicated relationship makes it difficult to predict DBPs formation with various factors although such a prediction represents the primary interest in DBPs research. A simplified solution to this problem in the literature was to adopt empirical models by linearly or log linearly regressing various involved factors (Lin et al., 2018; Sohn et al., 2004; Uyak et al., 2007). These proposed models were helpful to a better understanding of formation mechanisms of DBPs, and facilitated decision-making in DBPs control. However, most of these models were site-specific (Chowdhury et al., 2009; Singh and Gupta, 2012; Uyak et al., 2007), and developed based on chlorination of raw water (source water) or treated water from waterworks through careful design in laboratory. But in fact, the real drinking water needs to go through water treatment processes, disinfection and pipeline transportation. Concentrations and speciation of DBPs may be greatly changed as compared to those from simulated disinfection of source/treated water (Liu et al., 2011). It is necessary to develop their own predictive models for actual distribution systems. Moreover, most regression models included the parameter of “contact time” (Chowdhury et al., 2009; Hong et al., 2015, 2016; Lin et al., 2018). In laboratory, contact time is quite easy to be obtained. Yet for actual distribution systems, it is quite difficult to determine the contact time (i.e., residence time, duration from the beginning of disinfection in waterworks to the end of users), which usually needs consideration of distance, season and water supply data at sampling points. Therefore, it is quite difficult for these models to be applied to practice. Besides, linear models cannot exactly reflect the complicated non-linear relationships between various factors and DBPs formation (Kulkarni and Chellam, 2010; Singh and Gupta, 2012). It is, therefore, quite desirable to propose alternative methods to overcome these limitations.

Artificial neural networks (ANNs) have been generally deemed as standard non-linear estimators (Ghritlahre and Prasad, 2018; Singh and Gupta, 2012). Considering the complicated non-linear relationships of DBPs formation with various factors, and heterogeneity of drinking water contaminants, ANNs may provide an attractive alternative to predict DBPs formation. ANNs are information processing networks simulating nervous system of human brain (Bagheri et al., 2015; Iliyas et al., 2013). The distinct advantages of ANNs over linearly regressing models include their capabilities to approximate any functions to any accuracy, as well as their learning, parallel processing, and noise resistance abilities (Iliyas et al., 2013; You and Nikolaou, 1993). Curiously, in spite of great potential of ANNs in prediction of DBPs formation due to above-mentioned advantages, the cases regarding predictions of DBPs formation by using ANNs are still very limited. Pursuing the literature, a total of 6 studies are available on DBPs prediction through ANNs: an autoencoder-neural network (Peleato et al., 2018) and a hybrid genetic algorithm based ANN (Moradi et al., 2017) were respectively used in two studies, and back propagation (BP) ANNs were used in the other four studies (Kulkarni and Chellam, 2010; Park et al., 2018; Singh and Gupta, 2012; Ye et al., 2011). These studies well verified the feasibilities of ANNs in prediction of different DBPs. However, DBPs data in these studies was mostly (4 out of 6 studies) originated from simulated chlorination of raw/treated water, only two of which concerned about actual distribution systems (Moradi et al., 2017; Ye et al., 2011). Moreover, these two ANN models included “residence time”. Considering the difficulties in measuring residence time, developing prediction models of DBPs without “residence time” will be more convenient in practice. In addition, tests of different ANN models are still worth further investigating, and efficiencies of ANNs in applications of DBPs prediction should be further optimized. Radial basis function (RBF) ANN is a kind of typical feed forward neural networks, which possesses distinct advantages including universal approximation abilities, no local minimum problem and a faster learning algorithm over other ANNs (Jin and Bai, 2016; Zhao et al., 2019). In spite of high application potential of RBF ANNs, perusal of literatures shows lack of systematic studies regarding application of RBF ANNs in DBPs prediction.

As one of the most abundant DBPs occurring in drinking water, and a high risk to human health, HAAs have been limited by guidelines worldwide (Li and Mitch, 2018; Zhou et al., 2019). For example, the maximum contaminant level of 5 HAA species (HAA5, sum of chloro-(CAA), dichloro-(DCAA), trichloro(TCAA), bromo-(BAA) and dibromo-(DBAA)) in water has been set to be 60 μg/L by the United States Environmental Protection Agency (Richardson et al., 2007); In China, DCAA and TCAA, two most dominant HAA species (Sun et al., 2018; Wang et al., 2014; Weisel et al., 1999), are regulated and their maximum levels are set at 50 and 100 μg/L, respectively. Meanwhile, BCAA is also a frequently detected HAA species in drinking water (Ding et al., 2013; Gan et al., 2013). Therefore, DCAA, TCAA, BCAA, HAA5 and HAA9 (total HAA species, including DCAA, TCAA, BCAA, CAA, BAA, DBAA, BDCAA, DBCAA and TBAA (Yan et al., 2014)) were particularly concerned in this study.

The purpose of this study was, therefore, to predict levels of DCAA, TCAA, BCAA, HAA5 and HAA9 in actual distribution systems. Series of data regarding HAA concentrations and water characteristics was divided into training set and testing set for prediction by multiple linear (or log linear) regression (MLR) and RBF ANNs. Thereafter, prediction results were assessed, and advantages of proposed RBF ANNs for DBPs prediction over MLR were discussed. This study provided a first-hand report to systematically explore feasibilities of radial basis function (RBF) artificial neural network (ANN) in prediction of HAAs in actual distribution systems, which paved a new way to predict DBPs in tap water.

2. Materials and methods

2.1. Description of dataset and selection of water quality parameter

Dataset used to develop MLR and RBF ANN models in this paper was obtained from our previous study (Zhou et al., 2019). Water sampling, processing, and HAA analysis were briefly described as follows:

Tap water samples were collected from Jinhua region of Zhejiang Province, a representative region of south China, in which chlorine was used as the disinfectant. 64 representative sampling points consisting of 17 in summer, 23 in winter and 24 in spring were selected, and two replicate water samples were collected for each point. Details of sampling points can be referred to the literature (Zhou et al., 2019). Prior to tap water sampling, stabilized water temperature, pH and free chlorine residue of tap water were measured. Tap water samples for measurements of other water characteristics were collected in brown glass bottles (1000 mL). In contrast, tap water samples for HAAs analysis were filled into glass tubes (50 mL) with addition of dechlorination reagents (100 mg/L NH_4Cl). The tubes were then sealed by ground glass stopper. Ice-boxes were used to store all water samples, which were immediately transited to laboratory for further analysis.

The following standard methods were used to characterize common water quality parameters (APHA, 1998): pH, temperature, dissolved organic carbon (DOC) concentration and ultraviolet absorbance at 254 nm per cm path length (UVA_{254}) were respectively measured by a pH meter (Orion, model 420A), a Thermometer, a TOC analyzer (ELEMENTAR Liqui TOCII) and an UV−visible spectrophotometer (Shimadzu, UV-1601); residue chlorine, Br^-, NO_2^--N and NH_4^+-N were respectively measured by DPD titration method, ion chromatography (Dionex ICS-900), n-(1-naphthyl)-ethylenediamine dihydrochloride colorimetric method and salicylic acid−hypochlorous acid colorimetric method.

Measurement of HAAs was done according to standard method (USEPA, 2003): pH of water sample was first adjusted to below 0.5 using concentrated sulfuric acid. Thereafter, an appropriate amount of sodium sulfate was added into the sample followed by vigorous shake. After completely dissolving, methyl tert-butyl ether (with 1,2-dibromopropane as internal standard) was added into the sample which was then vigorously shaken for 2−3 min. HAAs in the supernatant were then subjected to methylation (addition of acidic methanol) followed by water bath heating for 2 h. At last, the methylated HAAs were separated by adding in Na_2SO_4 (150 g/L), which was then neutralized by $NaHCO_3$ before being analyzed by a GC-ECD system.

A point worth noting is that, though DOC, UVA_{254} and Br^- in tap water were not those in raw water (i.e. the water before chlorination), they were generally closely related with each other (see in supplement file, S-Table 1-2; S-Figs. 1 and 2). That is to say, DOC, UVA_{254} and Br^- in tap water can also be considered as good surrogate indicators for the organic and inorganic precursors of DBPs (including HAAs) formation, respectively.

According to knowledge available, HAAs formation is mainly controlled by water quality parameters: high level of organic matter (indicated by DOC and UVA_{254}) or low level of pH can facilitate HAAs formation (Liang and Singer, 2003; Nikolaou and Lekkas, 2001; Sun et al., 2018); high temperature might improve the production of di-HAAs, but decrease the yields of tri-HAAs (Hong et al., 2017); Increase of bromide level can shift the HAAs speciation to more brominated species (Hong et al., 2017); moreover, ammonia and nitrite might also be influencing factors to HAAs formation because they are chlorine consumer and might reduce the chlorine availability (Hu et al., 2010). Therefore, parameters of DOC, UVA_{254}, temperature, pH, Br^-, nitrite, ammonia and residue chlorine were all used to develop the prediction model in this study.

2.2. Multiple linear regression of DBPs production with various parameters

Multiple linear and log linear regression (MLR) is the most representative conventional method to predict DBPs production with various parameters. This method assumes linear or log linear relationships between DBPs mass concentration ([DBPs]) and each parameter of water quality including DOC concentration, UVA_{254}, bromide ion concentration ($[Br^-]$), reaction temperature (T), water pH (pH), residue chlorine ($[Cl_2]$), nitrite concentration ($[NO_2^-]$) and ammonia concentration ($[NH_4^+]$). Representative equations to describe DBPs production involved in this method can be expressed as follows: (1) for linear regression, (2) for log linear regression (Chowdhury et al., 2009; Hong et al., 2015, 2016; Kulkarni and Chellam, 2010; Lin et al., 2018):

$$[HAA] = K + a[DOC] + b[UVA_{254}] + c[Br^-] + d[T] + e[pH] + f[Cl_2] + g\left[NO_2^- - N\right] + h\left[NH_4^+ - N\right] \quad (1)$$

$$[HAA] = k \times [DOC]^a \times [UVA_{254}]^b \times [Br^-]^c \times [T]^d \times [pH]^e \times [Cl_2]^f \times \left[NO_2^- - N\right]^g \times \left[NH_4^+ - N\right]^h \quad (2)$$

where k, a, b, c, d, e, f, g and h are empirical constants. Because some water quality parameters will be eliminated during stepwise regression, parameters entering the model depend on actual situation.

Validation of this method was conducted by the data which was not included in the database used to simulate Eqs. (1) and (2) as inputs. Absolute relative error (E) of predicted DBPs concentration for each input set can be calculated through:

$$E = abs\left(\frac{[DBP_s]_{predicted} - [DBP_s]_{measured}}{[DBP_s]_{measured}}\right) \times 100\% \quad (3)$$

Herein, N_{25} was used to represent the percentage of predictions with E of less than 25%. Prediction quality of the method can be assessed according to the regression coefficient (r_p) and N_{25} value based on comparisons between the predicted and measured DBPs data (Kulkarni and Chellam, 2010). Linear regression was performed on a platform of SPSS 18 software, and values of r_p and N_{25} were accordingly determined.

2.3. Radial basis function (RBF) ANN

ANN, which is deemed as the most important embranchment of computational intelligence paradigms, is an information processing network simulating networks of neurons that make up human brain. Besides its self-learning, fault tolerance and distributed memory, a distinct advantage of ANN is its inherent ability to incorporate non-linear relationships into model, which avoid the complexity of conceptual models (Moradkhani et al., 2004; Zhao et al., 2019).

Feed forward neural networks including multilayer perceptron (MLP), back propagation (BP) networks and RBF ANN are much more popular than others in practical applications (Sreekanth et al., 2010). Distinct advantages of RBF ANN including universal approximation abilities, no local minimum problem and faster learning algorithm over other ANNs have been well documented in

the literature (Chen et al., 2019; Jin and Bai, 2016). As shown in Fig. 1, RBF ANNs generally possess a three-layer architecture: an input, an output, and a hidden layer. As implied by its name, radially symmetric basis functions (such as the Gaussian function $\phi_i(x)$) are adopted as activation functions of hidden nodes:

$$\phi_i(x) = \exp\left(-\frac{||x - c_i||}{\sigma_i^2}\right) \quad (4)$$

where c_i is the center, and σ_i is the spread of the *ith* RBF node, respectively. Transformation from input nodes to the hidden nodes is in non-linear form. In contrast, the network output (y) is approximated as a linear combination of the activation function output ($\phi_i(x)$) with the output layer weight (ω_i):

$$y = \sum_{i=0}^{n} \phi_i \omega_i \quad (5)$$

Training an RBF network is necessary for its practical application. Training processes can be done in two steps: one is to choose the centers from the training data (without training) or establish the centers by clustering the training data, and the other is to linearly estimate one weighting vector by using ordinary least squares. Selecting RBF centers can be achieved by using three learning strategies, among which, self-organized center selection is the most widely used one. As a self-organized technique, orthogonal least square (OLS) technique was adopted in this study due to its relatively high efficiency (Chen et al., 1991). This technique uses the gram-Schmidt algorithm to select and update the center, and uses adaptive gradient descent to adapt the weights (Iliyas et al., 2013). In this way, the network parameter values could be derived if the cost function is minimized:

$$\min J = \sum_{i=1}^{Q} \left(|Y_i - y_i|^2\right) \quad (6)$$

where Q, Y_i and y_i are the training pattern number, the network output and desire target output, respectively.

In this study, 64 representative sampling points throughout Jinhua region were selected, resulting in 64 representative DBPs data sets. More training data sets facilitate to obtain a more reliable network. Therefore, 80% (51 sets) and 20% (13 sets) of them were respectively used as training and verification samples. Meanwhile, an ANN better suited interpolation rather than extrapolation (Chen et al., 2020; Iliyas et al., 2013; Kulkarni and Chellam, 2010). Accordingly, these data sets were firstly ordered by sorting magnitudes of HAA, and then, data sets with evenly distributed, and the maximum as well as the minimum values were selected as training samples. While the remaining data, which were also evenly distributed and could fully represent the whole database, were selected as test samples.

MATLAB 2017b software provides function of NEWRB (P, T, Goal, Spread, n, DF), where P means input vector; T means target vector; n means the maximum number of neurons; and DF means number of neurons added between displays. This function will define an approximate RBF ANN where neurons in hidden layers are automatically determined. When performing this function, neurons will be automatically added till the target error square sum is achieved or the maximum number of neurons in hidden layers is reached.

Absolute relative errors (E) of RBF ANN predictions can be calculated according to Eq. (3). Prediction quality of RBF ANN can be assessed according to regression coefficient (r_p) and N_{25} value based on comparison between predicted and measured HAAs data (Kulkarni and Chellam, 2010). Values of r_p and N_{25} were analyzed through the method provided in Section 2.3, which were performed by SPSS 18 software.

3. Results and discussion

3.1. HAAs levels, water quality parameters and grouping

HAAs levels and water quality parameters of 64 water samples were originated from our previous studies (Zhou et al., 2019), which were divided into two groups (training and testing group). In view of the fact that an ANN is essentially of interpolation method rather than extrapolation method, the following grouping shown in Table 1 was made in this study. DCAA, TCAA, HAA5 and HAA9 are highly correlated with each other (values of R^2 ranged from 0.79 to 0.994, S-Table 3). That is to say, DCAA, TCAA, HAA5 and HAA9 have similar distribution patterns, and training data set and testing data set for DCAA, TCAA, HAA5 and HAA9 can be well arranged for one time. Sample No.1–51 serve as training samples, and Sample No.t1-t13 serve as testing samples. It can be seen that, levels of DCAA, TCAA, HAA5 and HAA9 in training samples are in range of 3.22–17.9 μg/L, 1.74–18.11 μg/L, 4.96–36.46 and 5.77–38.48 μg/L,

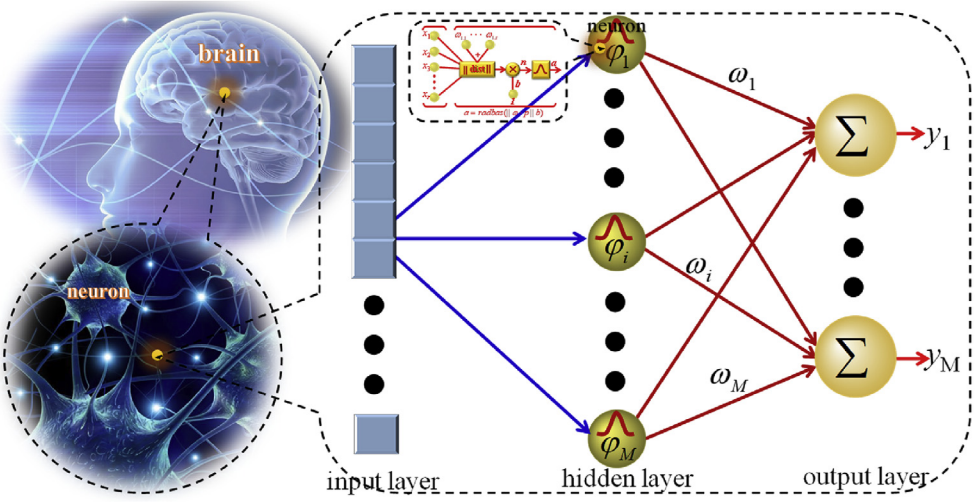


Fig. 1. Illustration of a typical RBF network model.

Table 1
Concentrations of HAAs and corresponding water quality parameters for tap water samples (data is originated from Zhou et al., 2019).

Sample Number[a]	HAAs Production					Water quality parameters							
	DCAA (μg/L)	TCAA (μg/L)	BCAA[b](μg/L)	HAA5 (μg/L)	HAA9 (μg/L)	Temp (°C)	pH	UVA_{254} (cm)	Cl_2 (mg/L)	NO_2^--N (μg/L)	DOC (mg/L)	NH_4^+-N (μg/L)	Br^- (μg/L)
1	6.40	5.92	0.65	13.66	15.89	24.1	7.24	0.032	0.06	4.28	1.83	4.38	6.05
2	5.70	4.39	0.78	10.09	11.16	8.5	7.18	0.020	0.15	0.00	0.93	29.65	2.85
3	5.03	3.20	0.84	8.23	9.29	9.0	7.11	0.019	0.28	0.00	0.96	32.35	2.62
4	5.28	3.92	0.72	9.20	10.19	10.8	7.05	0.016	0.24	0.30	0.88	33.69	2.54
5	3.99	2.82	0.55	6.81	7.36	19.0	6.99	0.008	0.14	0.86	0.75	27.48	2.38
6	4.30	3.38	0.57	7.73	8.30	20.0	7.02	0.016	0.09	1.71	0.76	29.60	3.38
7	4.24	2.95	0.61	7.19	7.79	18.5	6.92	0.007	0.14	1.14	0.53	30.66	1.75
8	4.76	5.04	0.88	10.56	12.39	23.5	7.23	0.008	0.17	0.91	1.22	2.19	5.64
9	4.86	6.04	1.11	12.20	14.59	21.8	7.71	0.019	0.20	1.21	1.12	6.56	20.64
10	5.06	3.96	0.91	9.90	11.96	23.0	7.31	0.013	0.28	0.91	1.07	9.84	6.70
11	4.81	3.31	0.82	8.12	9.22	12.0	7.00	0.009	0.36	0.00	0.61	33.69	2.15
12	4.90	3.87	0.89	8.77	9.95	10.5	7.07	0.013	0.28	0.61	0.50	29.65	1.15
13	4.88	3.37	0.88	8.24	9.40	13.5	7.08	0.011	0.36	0.30	0.61	26.95	1.23
14	3.72	1.95	0.67	5.67	6.33	16.0	7.37	0.001	0.40	0.57	0.07	26.43	2.00
15	3.44	1.78	0.62	5.22	5.84	17.0	7.16	0.002	0.43	0.57	0.03	14.80	2.13
16	4.35	2.60	0.87	7.87	9.35	24.8	7.32	0.020	0.33	1.82	0.80	0.00	2.86
17	5.42	4.36	0.87	9.78	11.08	11.5	7.23	0.015	0.11	0.00	0.60	22.91	1.31
18	5.84	4.63	1.05	10.47	12.01	12.0	6.99	0.017	1.04	0.00	0.59	32.35	1.23
19	3.39	2.96	0.86	6.35	7.28	17.0	7.43	0.005	0.31	0.86	0.14	17.97	1.75
20	3.22	1.74	0.82	4.96	5.77	15.2	7.45	0.005	0.33	0.86	0.08	29.60	2.50
21	7.80	6.02	1.23	13.90	15.25	17.8	7.13	0.005	1.51	0.86	0.20	8.46	1.75
22	16.84	18.11	1.02	36.46	38.48	24.2	7.3	0.038	0.4	1.80	2.02	6.51	3.9
23	15.61	17.58	0.80	33.19	34.53	7.0	7.26	0.035	0.05	0.30	1.54	21.56	3.08
24	15.75	14.21	1.43	30.19	32.20	9.0	7.18	0.022	0.24	0.00	1.10	6.74	1.77
25	11.80	9.01	0.82	20.93	21.76	18.9	6.85	0.016	0.50	0.86	1.03	10.57	2.06
26	17.90	7.77	1.29	25.74	27.11	14.1	7.07	0.010	0.54	1.43	0.52	7.40	1.75
27	11.49	13.44	0.98	26.27	29.16	24.9	7.45	0.030	0.28	2.12	1.32	4.37	5.14
28	12.10	13.09	1.26	26.43	29.16	24.6	7.28	0.029	0.40	1.52	1.73	2.19	4.50
29	10.26	11.58	1.05	21.84	23.45	10.0	7.00	0.028	0.19	0.30	1.03	24.26	3.54
30	12.28	9.20	1.42	21.79	23.30	16.1	7.26	0.011	0.31	0.57	0.97	23.26	2.75
31	11.44	12.75	1.07	25.78	28.02	24.4	7.51	0.030	0.46	1.52	2.14	5.46	4.21
32	8.50	10.40	1.61	19.97	24.10	24.5	7.44	0.024	0.02	1.52	1.41	1.09	10.71
33	10.41	9.41	1.34	19.82	21.66	12.0	7.11	0.025	0.24	0.61	1.48	31.00	3.69
34	12.44	9.90	2.16	22.35	25.38	12.0	7.07	0.023	0.45	0.30	0.85	1.35	3.23
35	10.74	10.86	2.16	21.80	24.97	13.0	7.13	0.019	0.40	0.00	0.88	26.95	2.46
36	8.08	4.87	1.64	13.27	14.92	17.0	7.12	0.015	0.47	0.57	1.22	27.48	3.13
37	7.43	5.58	0.98	13.62	14.61	15.8	7.15	0.010	0.54	0.86	0.79	3.17	1.50
38	9.37	7.72	1.50	17.09	18.72	18.1	7.21	0.009	0.17	1.14	0.67	25.37	3.63
39	9.28	10.87	1.07	21.30	23.74	24.6	7.61	0.029	0.08	2.76	1.01	8.78	5.4
40	11.16	7.46	3.00	18.92	22.86	12.5	7.36	0.028	0.50	0.30	1.55	29.65	1.23
41	6.86	6.34	0.91	13.20	14.46	11.0	7.20	0.019	0.31	0.00	0.86	24.26	1.23
42	6.93	7.13	0.88	14.05	15.36	10.5	7.23	0.022	0.09	0.00	0.94	20.22	2.92
43	5.02	3.98	1.50	9.06	10.75	18.5	7.27	0.014	0.05	1.43	1.20	25.37	5.50
44	7.94	4.06	1.42	12.06	13.48	15.2	7.30	0.008	0.38	1.14	0.55	23.26	1.63
45	8.02	4.22	1.45	12.31	13.84	15.5	7.31	0.008	0.38	1.71	0.73	19.03	2.06
46	7.84	7.37	0.89	16.36	18.02	25.2	7.26	0.031	0.29	1.83	0.96	0	4.89
47	8.72	7.14	0.87	15.86	16.98	10.0	7.10	0.021	0.57	0.00	0.75	13.48	1.23
48	8.21	7.23	1.01	15.43	16.75	5.0	6.93	0.018	0.50	0.30	0.70	10.78	1.15
49	6.92	5.54	0.94	12.52	13.46	14.9	7.12	0.010	0.50	0.86	0.37	8.46	1.56
50	6.74	5.22	0.96	12.03	13.09	13.8	6.85	0.008	0.54	0.57	0.46	14.80	1.94
51	6.52	5.09	0.93	11.69	12.70	14.9	6.92	0.009	0.59	2.00	0.23	17.97	1.81
t1	6.29	6.07	0.63	13.35	15.52	24.2	7.26	0.031	0.08	4.20	1.87	4.36	6.35
t2	4.23	2.09	0.74	6.32	7.06	16.0	7.09	0.004	0.45	0.86	0.20	21.14	2.13
t3	3.84	4.69	0.96	9.39	11.13	24.6	8.5	0.030	0.14	1.82	0.98	3.28	3.79
t4	3.81	3.49	0.63	7.30	8.36	9.0	6.86	0.021	0.06	0.61	0.36	28.30	1.85
t5	16.53	17.86	1.03	35.94	37.95	24.3	7.4	0.036	0.36	1.84	2.06	6.61	4.1
t6	15.23	17.10	1.14	32.33	33.96	14.0	7.15	0.030	0.24	0.30	1.46	10.78	2.92
t7	12.49	9.61	1.04	22.23	23.26	18.2	6.93	0.015	0.50	1.14	1.02	21.14	1.81
t8	9.78	10.88	1.13	20.67	22.40	13.5	7.02	0.025	0.38	0.61	1.25	31.00	2.77
t9	7.81	7.76	0.96	15.66	16.62	15.1	7.02	0.014	0.21	0.57	0.74	25.37	3.31
t10	7.85	7.54	1.09	16.13	17.22	16.8	7.05	0.011	0.59	0.57	0.99	29.60	2.56
t11	8.75	10.48	1.02	20.01	22.33	24.7	7.57	0.029	0.06	2.70	0.99	8.7	5.2
t12	7.75	7.23	0.87	15.14	16.81	25.0	7.24	0.033	0.27	1.81	0.94	0	4.83
t13	8.48	7.09	0.92	15.57	16.92	8.5	6.92	0.021	0.50	0.00	0.70	10.78	1.31

Note:
[a] Sample No.1−51 are training data sets for DCAA, TCAA, HAA5 and HAA9; Sample No. t1-t13 are testing data sets for DCAA, TCAA, HAA5 and HAA9.
[b] Data with pink refers to testing data sets for BCAA, while the remaining are used as training data sets for BCAA.

respectively; while the counterparts in testing samples range from 3.81 to 16.53, 2.09–17.10, 6.32–35.94 and 7.06–37.95 μg/L, respectively. It is clear that, all data levels of testing group are within the scope of training group and can represent the whole database.

However, for BCAA, its distribution pattern is quite different from that of DCAA, TCAA, HAA5 and HAA9, which can be concluded from correlation coefficients (S-Table 3, R^2 for BCAA & DCAA, BCAA & TCAA, BCAA & HAA5 and BCAA & HAA9 ranged from 0.085 to 0.187). That is to say, samples selected for training data sets for DCAA, TCAA, HAA5 and HAA9 are generally not suitable/representative for BCAA. For example, if grouping according to methods of DCAA, TCAA, HAA5 and HAA9, BCAA levels in testing sets will range between 0.63 and 1.14 μg/L, which is much lower as compared to those in training sets (0.55–3.0 μg/L), and thus, it is not suitable for developing models. Therefore, it is necessary to re-group the data based on the rule of data selection for ANN model. Accordingly, data marked with pink color (BCAA level: 0.57–2.16 μg/L) was defined as testing group, and the remaining data was set as training group (0.55–3.0 μg/L) for BCAA item modeling.

Based on the above grouping information, MLR and RBF-ANN models were developed, as is showed in Section 3.2 and 3.3.

3.2. Linearly and log linearly regression models

Training data set in Table 1 were used to simulate linear models and log linear models for HAAs prediction. When performing stepwise multiple regression procedure through SPSS software, independent variables (water quality parameters or their log values) enter the equation in order of their linear correlations with dependent variables (HAAs or their log values) (Lin et al., 2018; Sohn et al., 2004). Therefore, water quality parameters entering the model can be considered as the most important factors that influence HAAs formation (Hong et al., 2015). In this way, the following equations for DCAA (Model 1 and 6), TCAA (Model 2 and 7), BCAA (Model 3 and 8), HAA5 (Model 4 and 9) and HAA9 (Model 5 and 10) were generated, respectively. Analysis of variances shows that values of F for all models are larger than the critical F value at a confidence level of 99% or 95% (S-Table 4), indicating a significantly linear relationship between each HAA component and entering parameters.

3.3. Linear models

1) $DCAA = 2.361 + 251.722\ [UVA_{254}] + 3.801\ [Cl_2]$ ($R^2 = 0.376$, $p = 0.000$, $n = 51$);
2) $TCAA = 1.232 + 328.501[UVA_{254}]$ ($R^2 = 0.576$, $p = 0.000$, $n = 51$);
3) $BCAA = 0.811 + 0.265[DOC]$ ($R^2 = 0.103$, $p = 0.022$, $n = 51$);
4) $HAA5 = 8.823 + 527.103[UVA_{254}] - 0.153[NH_4^+\text{-N}]$ ($R^2 = 0.543$, $p = 0.000$, $n = 51$);
5) $HAA9 = 9.934 + 572.604[UVA_{254}] - 0.170\ [NH_4^+\text{-N}]$ ($R^2 = 0.573$, $p = 0.011$, $n = 51$).

Log linear models:

6) $DCAA = 10^{1.711} \times UVA_{254}^{0.412} \times Cl_2^{0.156}$ ($R^2 = 0.410$, $p = 0.000$, $n = 51$);
7) $TCAA = 10^{1.845} \times UVA_{254}^{0.583}$ ($R^2 = 0.521$, $p = 0.000$, $n = 51$);
8) $BCAA = 10^{0.018} \times DOC^{0.140}$ ($R^2 = 0.141$, $p = 0.007$, $n = 51$);
9) $HAA5 = 10^{2.007} * UVA_{254}^{0.475}$ ($R^2 = 0.465$, $p = 0.000$, $n = 51$);
10) $HAA9 = 10^{2.050} * UVA_{254}^{0.473}$ ($R^2 = 0.485$, $p = 0.000$, $n = 51$).

Models (Model 1–2, 6–7) show that DCAA and TCAA are positively related to UVA_{254}, suggesting that aromatic organic matters are important precursors in formation of DCAA and TCAA. It has been reported that aromatic organic matter may lead to the formation of aromatic halogenated DBPs during chlorination, which can be further decomposed to form haloacetic acids with presence of chlorine (Jiang et al., 2017; Pan and Zhang, 2013; Zhai and Zhang, 2011). HAA5 and HAA9 are also positively related to UVA_{254} (Model 4–5, 9–10), which can be attributed to the fact that DCAA and TCAA are two predominant HAA species for HAA5 and HAA9 (Table 1). While BCAA is positively related to DOC, indicating that precursor of BCAA may be different from those of DCAA and TCAA. This is similar to our previous studies that brominated HAAs were better correlated to DOC as compared with UVA_{254}, which indicated that precursors of brominated HAAs were more hydrophilic as compared to those of chlorinated HAAs (Zheng et al., 2020). As for water quality parameters, only chlorine (Model 1 and 6) and ammonia (Model 4 and 5) enter the models, which exert positive and negative influences on HAA formation, respectively. This result suggests that chlorine availability play an important role in HAAs formation in tap water. Ammonia is one of the chlorine consumers, which may decrease the amount of chlorine that is available for organic matters, and therefore played a negative role in HAAs formation (Yang and Shang, 2004). Though linear and log linear regression models could help to identify the potential key factors for HAA formation in tap water, values of R^2 in these models (Linear models: $R^2 = 0.103–0.576$; log linear models: $R^2 = 0.141–0.521$) were dominantly lower as compared to those obtained in laboratories ($R^2 = 0.800–0.959$) (Sohn et al., 2004; Song et al., 2017). This may be because 1) HAAs precursors are quite complex: aromatic carbon is not always a potent precursor for HAAs, and some aliphatic compounds can also greatly contribute to HAA yields (Hong et al., 2009), which may lead to reductions in relationship between DOC/UVA_{254} and HAA yields; 2) In the present study, we collected tap water from different plant in different season, so that chlorination conditions (e.g. chlorine dose, contact time) especially the reaction temperature were quite different among water samples. This may make relationship between HAAs formation and water quality parameters not as obvious as those in laboratories (under absolutely same chlorination conditions). Low coefficients of Model 1–10 indicated that these regression models might not be appropriate to describe HAAs formation in tap water, which could be further confirmed by the following prediction results.

Regression models (Model 1–10) were then used for prediction using the testing data sets in Table 1 as inputs. Fig. 2 a1-e1 (linear models) and Fig. 3 a1-e1 (log linear models) show comparisons between predicted HAAs levels and experimentally measured ones. It can be seen that, values of R^2 for predictions of DCAA, TCAA, BCAA, HAA5 and HAA9 are respectively within a range of 0.251–0.400, 0.268–0.270, 0.001–0.021, 0.181–0.219 and 0.213–0.254, indicating a low correlation between predicted and measured values. Meanwhile, values of N_{25} (the percentage of predictions within an absolute relative error of 25%) for DCAA, TCAA, BCAA, HAA5 and HAA9 predictions ranged from 46 to 54%, 31–38%, 46–54%, 38–46% and 46-46%, respectively, suggesting low quality of linear and log linear models in prediction of levels of DCAA, TCAA, BCAA, HAA5 and HAA9 in tap water.

To further explore prediction abilities of linear and log linear models, data sets used for simulate the models (51 groups) and all the data sets in Table 1 (64 groups) were used as model inputs, and comparisons between predicted HAAs production and experimentally measured ones are shown in Fig. 2 a2-e2, Fig. 3 a2-e2, Fig. 2 a3-e3 and Fig. 3 a3-e3. As is shown in Figs. 2 and 3, both R^2 and N_{25} for DCAA, TCAA, BCAA, HAA5 and HAA9 are not satisfied to get sound prediction results, indicating weak prediction abilities of linear and log linear models, which calls for alternative methods.

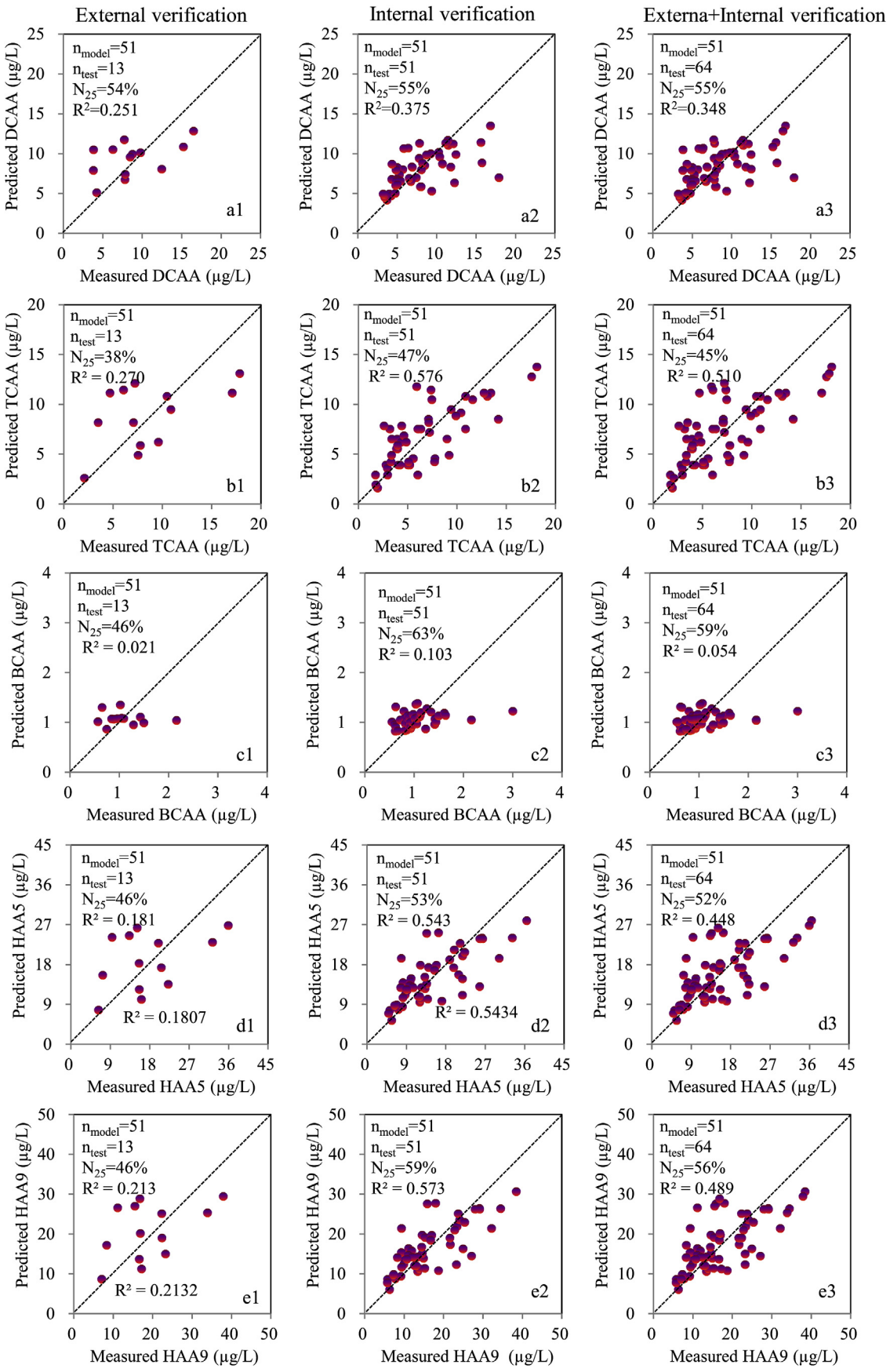


Fig. 2. Comparisons of predictions by linear models with experimental measurements (a, b, c, d and e refer to DCAA, TCAA, BCAA, HAA5 and HAA9 respectively, and the postfixes of 1, 2 and 3 refer to testing data number of 13, 51 and 64, respectively).

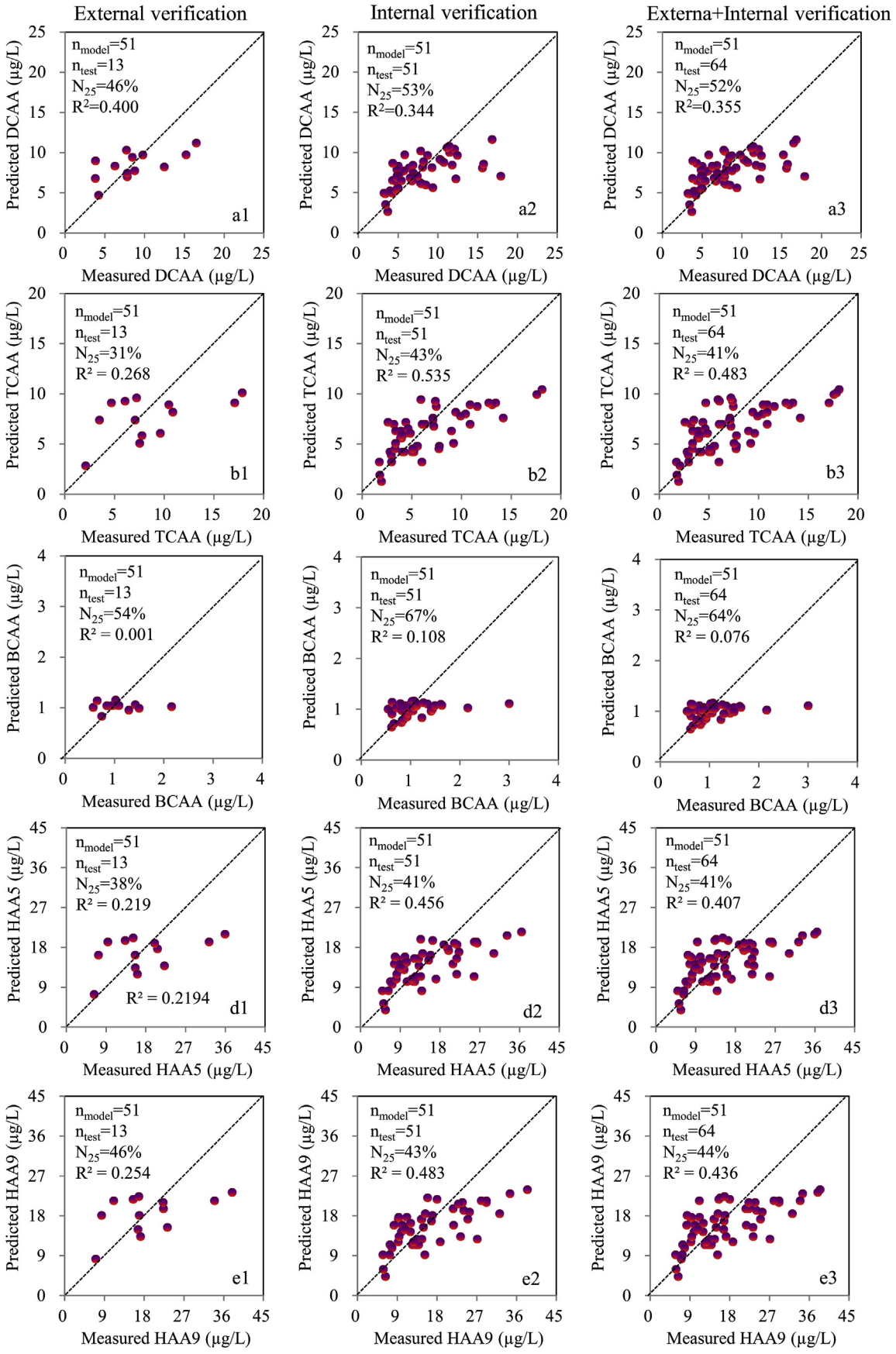


Fig. 3. Comparisons of predictions by log linear models with the experimental measurements (a, b, c, d and e refer to DCAA, TCAA, BCAA, HAA5 and HAA9 respectively, and the postfixes of 1, 2 and 3 refer to testing data number of 13, 51 and 64, respectively).

3.4. Prediction of HAAs levels with various parameters by RBF ANN

Function of newrb (P, T, GOAL, SPREAD, n, DF) in Matlab 2018b was used to yield RBF networks for prediction of DCAA, TCAA, BCAA, HAA5 and HAA9. All water quality parameters were included during modeling. It is generally reported that, two parameters, namely maximum neurons number (n) and spread (σ_i) of Gaussian function, critically determine the efficiency of an RBF network (Jin and Bai, 2016). In this study, the optimum neurons number (n) and spread of RBF network for each HAA were obtained through the following procedures.

1) HAA9 (can be a representative for DCAA, TCAA and HAA5 as they are highly related with HAA9 (S-Table 3)), and BCAA were selected as test samples to identify the neurons number (N). During modeling, DF was set at 5; neurons number (n) was set at 10, 20 and 30; spread was set at 10, 20, 30 … …90. That is to say, a total of 27 (3(N) × 9(Spread)) models were respectively developed for HAA9 and BCAA. Results showed that when n = 30 and spread = 30 (HAA9) or 40 (BCAA), the models had the best predictive performances, which can be seen from their N_{25} values and regression coefficients (R^2) (Supplement file, S-Figs. 3 and 4), from which it could be estimated that n = 30 might also be a good choice for other HAAs.

2) Then n = 30, spread = 10, 20, 30 … 90 were further used in developing DCAA, TCAA and HAA5. Results showed that when spread = 20 (for DCAA), 50 (TCAA) and 70 (HAA5), the models have the best predictive performances. Based on these results, spread can be further micro-adjusted until their predictive performances are ideal. Here, we take DCAA as an example to illustrate the effects of spread in Gaussian function on prediction ability of RBF networks (with a given n of 30). As shown in Fig. 4, among the three selected spread levels (10, 20, and 30), spread of 20 corresponds to the highest regression coefficients and N_{25} values under conditions of this study. Based on this, even better prediction models (a spread of 19.7) were obtained, which showed an R^2 value of 0.985 and an N_{25} of 92%. Spread of Gaussian function in RBF networks also showed significant effects on predictions of TCAA, BCAA, HAA5 and HAA9 (Supplement File, S-Fig. 5). These results indicate that RBF ANN method can be further optimized/improved to get more accurate predictions. In contrast, linear and log linear models assume the functional relationships, and their accuracies cannot be improved for the given input data. This is also a significant advantage of RBF ANN over MLR method for DBPs prediction.

On other hand, we also tried to develop RBF ANNs for HAAs with fewer water quality parameters (select 5–7 parameters according to their order of grey correlation coefficients, see supplement). However, their prediction performances were dominantly lower as compared to those networks using 8 water quality parameters (data was not shown), indicating that all water quality parameters were important for RBF ANN network of HAAs. Therefore, in this study, only the RBF ANNs using 8 parameters were presented here.

The final RBF ANNs with good predictive performance are presented in Fig. 5. Comparisons between predicted and measured HAAs (DCAA, TCAA, BCAA, HAA5 and HAA9) were carried out. It can be seen from Fig. 5 a1-e1 that, for all the 13 testing data sets, high regression coefficients (R^2 = 0.681–0.985) and high N_{25} values (85–92%) were obtained. As shown in Fig. 5 a2-e2 and Fig. 5 a3-e3, trained RBF networks showed high prediction abilities even when the training and all data sets were used as inputs. The consistently high regression coefficients and N_{25} values indicate that RBF ANNs are capable to accurately involve complicated non-linear relationships of DBPs formation with precursor characteristics and chlorination conditions. Whereas, MLR models can not take similar roles in DBPs prediction.

Overall, prediction accuracies of RBF ANNs were 21–47% higher than those of linear and log linear models (Figs. 2, 3 and 5, summarized in Table 2). The average absolute error of RBF ANN models for DCAA, TCAA, BCAA, HAA5 and HAA9 were 12%, 17%, 12%, 16%, and 14%, respectively, which were generally twice lower as compared to the corresponding linear and log linear models (Table 2). The high prediction accuracy and low absolute error indicated the superior prediction abilities of RBF networks over linear and log linear models in HAAs formation, and suggested that RBF networks had potential applications in DBPs monitoring and disinfection process optimizations.

Compared with reported BP ANN models for DCAA (R^2 = 0.867), TCAA (R^2 = 0.719) and HAA5 (R^2 = 0.895) in distribution systems (Ye et al., 2011), which included parameters of “residence time”, RBF ANN developed without “residence time (contact time)” in this study shows comparable or even better prediction performances (DCAA: R^2 = 0.908; TCAA: R^2 = 0.906; HAA5: R^2 = 0.896). This demonstrates that RBF ANNs are efficient in prediction of DBPs in actual distribution system and ANN models without “residence time” are still as good as those containing “residence time”.

4. Conclusions

This study provided a first-hand report on systematically explorations of feasibilities of RBF ANNs in DBP (e.g. HAAs) predictions in actual distribution systems. Totally 64 tap water samples taken from a typical region (Jinhua region) of China were used to train and verify RBF ANNs and MLR models. RBF ANN predictions showed rather higher regression coefficient and accuracies than MLRs, indicating high capabilities of RBF ANNs in capturing complex and non-linear relationships regarding HAAs formation. Moreover, RBF ANN predictions were more efficient and allowed further improvements in prediction accuracy. Accurate predictions of HAAs by RBF ANNs provided great potential benefits for DBPs

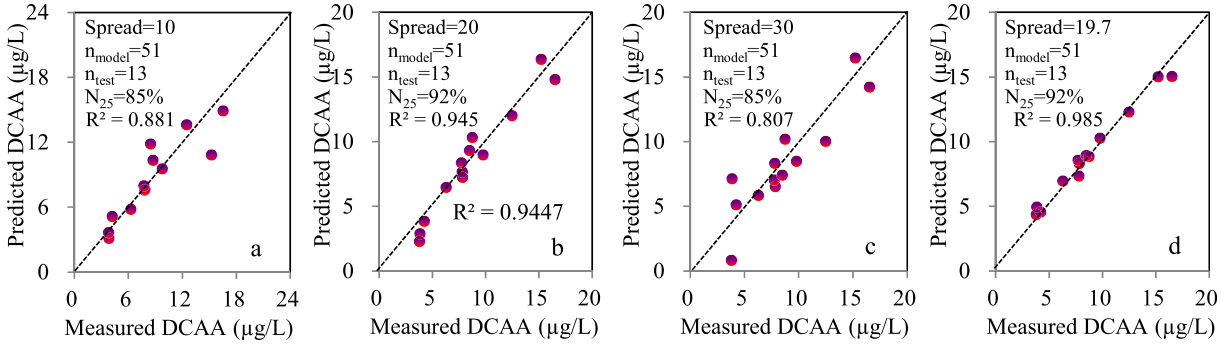


Fig. 4. Effects of spread of Gaussian function (a:10, b: 20, c: 30, d: 19.7) on DCAA predictions in RBF networks (n = 30).

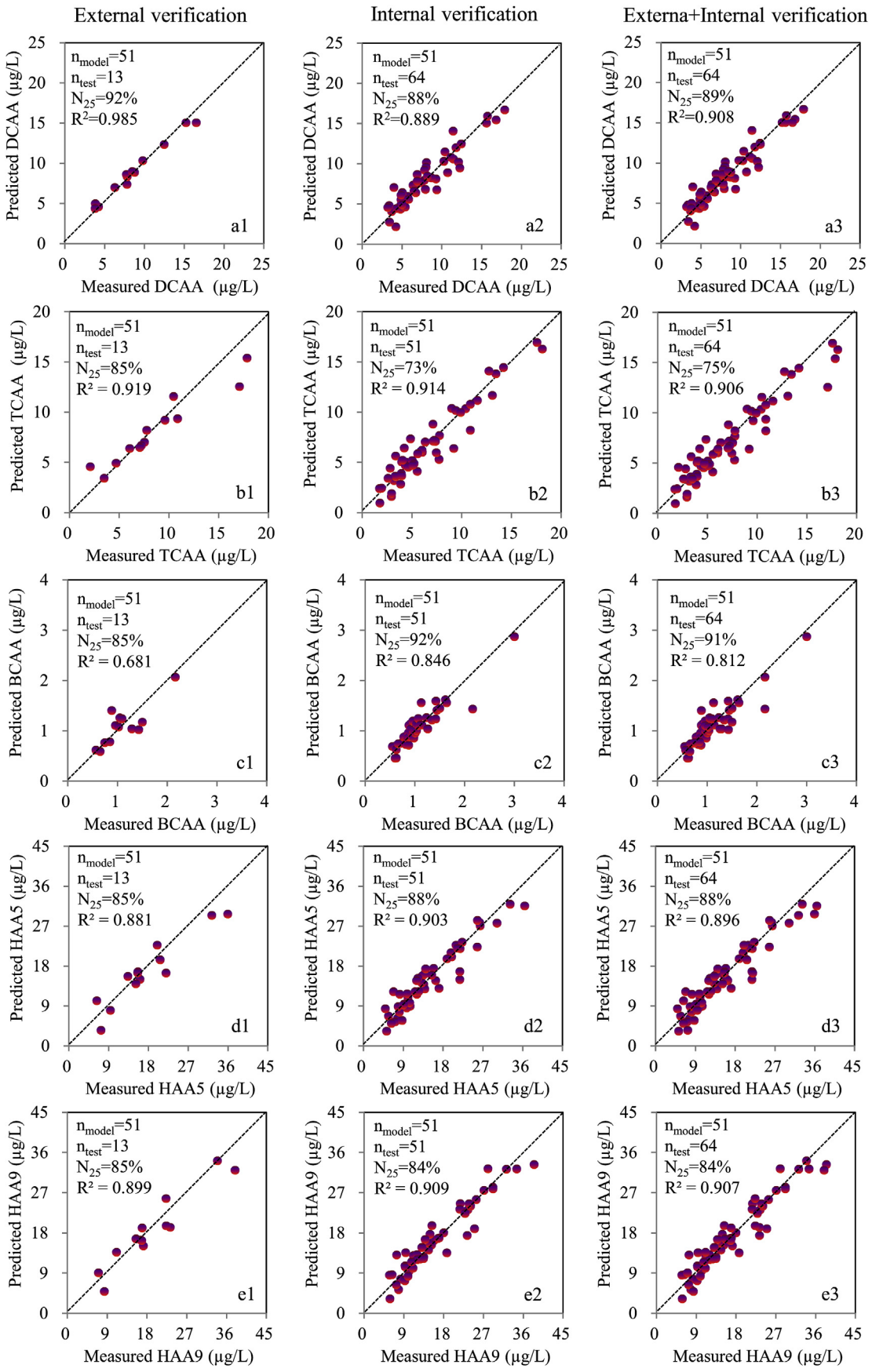


Fig. 5. Comparisons of RBF ANN′ predictions with experimental measurements for HAAs production (a, b, c, d and e refer to DCAA, TCAA, BCAA, HAA5 and HAA9, respectively, and the postfixes of 1, 2 and 3 refer to testing data number of 13, 51 and 64, respectively).

Table 2
Comparison of linear, log linear and RBF ANN models in terms of prediction accuracy and average absolute error.

Parameters	Model	DCAA	TCAA	BCAA	HAA5	HAA9
Prediction accuracy (%)	Linear	55	45	59	52	56
	Log linear	52	41	70	41	44
	RBF ANN	89	75	91	88	84
Average absolute error (%)	Linear	30	39	24	32	30
	Log linear	29	37	22	33	32
	RBF ANN	12	17	12	16	14

monitoring in actual water supply system, disinfection process controls and optimizations, showing a great application prospect.

CRediT authorship contribution statement

Hongjun Lin: Investigation, Methodology, Writing - original draft. **Qunyun Dai:** Investigation, Data curation, Formal analysis. **Lili Zheng:** Investigation, Data curation, Formal analysis. **Huachang Hong:** Conceptualization, Funding acquisition, Project administration, Writing - review & editing. **Wenjing Deng:** Investigation, Data curation. **Fuyong Wu:** Investigation, Data curation.

Acknowledgements

This study was financially supported by Public Welfare Project of the Science and Technology Department of Zhejiang Province (LGF18H260005), Foundation of Science and Technology Bureau of Jinhua, Zhejiang Province, China (No. 2014-3-030), Self-Design Project in Zhejiang Normal University (2019ZS05) and National Natural Science Foundation of China (No. 51978628).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2020.125999.

References

APHA, 1998. Standard Methods for the Examination of Water and Wastewater. American Public Health Association, Washington, DC, USA.
Arbuckle Tye, E., Hrudey Steve, E., Krasner Stuart, W., Nuckols Jay, R., Richardson Susan, D., Singer, P., Mendola, P., Dodds, L., Weisel, C., Ashley David, L., Froese Kenneth, L., Pegram Rex, A., Schultz Irvin, R., Reif, J., Bachand Annette, M., Benoit Frank, M., Lynberg, M., Poole, C., Waller, K., 2002. Assessing exposure in epidemiologic studies to disinfection by-products in drinking water: report from an international workshop. Environ. Health Perspect. 110, 53–60.
Bagheri, M., Mirbagheri, S.A., Ehteshami, M., Bagheri, Z., 2015. Modeling of a sequencing batch reactor treating municipal wastewater using multi-layer perceptron and radial basis function artificial neural networks. Process Saf. Environ. 93, 111–123.
Chen, S., Cowan, C.F.N., Grant, P.M., 1991. Orthogonal least-squares learning algorithm for radial basis function networks. IEEE Trans. Neural Network. 2 (2), 302–309.
Chen, Y., Shen, L., Li, R., Xu, X., Hong, H., Lin, H., Chen, J., 2020. Quantification of interfacial energies associated with membrane fouling in a membrane bioreactor by using BP and GRNN artificial neural networks. J. Colloid Interface Sci. 565, 1–10.
Chen, Y., Yu, G., Long, Y., Teng, J., You, X., Liao, B.-Q., Lin, H., 2019. Application of radial basis function artificial neural network to quantify interfacial energies related to membrane fouling in a membrane bioreactor. Bioresour. Technol. 293, 122103.
Chowdhury, S., Champagne, P., McLellan, P.J., 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: a chronological review. Sci. Total Environ. 407 (14), 4189–4206.
Deng, Z., Yang, X., Shang, C., Zhang, X., 2014. Electrospray ionization-tandem mass spectrometry method for differentiating chlorine substitution in disinfection byproduct formation. Environ. Sci. Technol. 48 (9), 4877–4884.
Ding, H., Meng, L., Zhang, H., Yu, J., An, W., Hu, J., Yang, M., 2013. Occurrence, profiling and prioritization of halogenated disinfection by-products in drinking water of China. Environ. Sci. Proc. Imp. 15 (7), 1424.
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. China 58, 51–63.
Gan, W.H., Guo, W.H., Mo, J.M., He, Y.S., Liu, Y.J., Liu, W., Liang, Y.M., Yang, X., 2013. The occurrence of disinfection by-products in municipal drinking water in China's Pearl River Delta and a multipathway cancer risk assessment. Sci. Total Environ. 447, 108–115.
Ghritlahre, H.K., Prasad, R.K., 2018. Investigation of thermal performance of unidirectional flow porous bed solar air heater using MLP, GRNN, and RBF models of ANN technique. Therm. Sci. Eng. Prog. 6, 226–235.
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), 1–6.
Hong, H., Qian, L., Xiong, Y., Xiao, Z., Lin, H., Yu, H., 2015. Use of multiple regression models to evaluate the formation of halonitromethane via chlorination/chloramination of water from Tai Lake and the Qiantang River, China. Chemosphere 119, 540–546.
Hong, H., Song, Q., Mazumder, A., Luo, Q., Chen, J., Lin, H., Yu, H., Shen, L., 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., Yan, X., Song, X., Qin, Y., Moradkhani, H., Lin, H., Chen, J., Liang, Y., 2017. Bromine incorporation into five DBP classes upon chlorination of water with extremely low SUVA values. Sci. Total Environ. 590–591, 720–728.
Hong, H.C., Wong, M.H., Liang, Y., 2009. Amino acids as precursors of trihalomethane and haloacetic acid formation during chlorination. Arch. Environ. Contam. Toxicol. 56 (4), 638–645.
Hu, J., Song, H., Addison, J.W., Karanfil, T., 2010. Halonitromethane formation potentials in drinking waters. Water Res. 44 (1), 105–114.
Iliyas, S.A., Elshafei, M., Habib, M.A., Adeniran, A.A., 2013. RBF neural network inferential sensor for process emission monitoring. Contr. Eng. Pract. 21 (7), 962–970.
Jiang, J., Zhang, X., Zhu, X., Li, Y., 2017. Removal of intermediate aromatic halogenated DBPs by activated carbon adsorption: a new approach to controlling halogenated DBPs in chlorinated drinking water. Environ. Sci. Technol. 51 (6), 3435–3444.
Jin, L., Bai, P., 2016. QSPR study on normal boiling point of acyclic oxygen containing organic compounds by radial basis function artificial neural network. Chemometr. Intell. Lab. Syst. 157, 127–132.
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, A., Zhao, X., Mao, R., Liu, H., Qu, J., 2014. Characterization of dissolved organic matter from surface waters with low to high dissolved organic carbon and the related disinfection byproduct formation potential. J. Hazard Mater. 271, 228–235.
Li, M., Xu, B., Liungai, Z., Hu, H.-Y., Chen, C., Qiao, J., Lu, Y., 2016. The removal of estrogenic activity with UV/chlorine technology and identification of novel estrogenic disinfection by-products. J. Hazard Mater. 307, 119–126.
Li, X.-F., Mitch, W.A., 2018. Drinking water disinfection byproducts (DBPs) and human health effects: multidisciplinary challenges and opportunities. Environ. Sci. Technol. 52 (4), 1681–1689.
Liang, L., Singer, P.C., 2003. Factors influencing the formation and relative distribution of haloacetic acids and trihalomethanes in drinking water. Environ. Sci. Technol. 37 (13), 2920–2928.
Lin, J., Chen, X., Ansheng, Z., Hong, H., Liang, Y., Sun, H., Lin, H., Chen, J., 2018. Regression models evaluating THMs, HAAs and HANs formation upon chloramination of source water collected from Yangtze River Delta Region, China. Ecotox. Environ. Safe. 160, 249–256.
Liu, W., Zhao, Y., Chow, C.W., Wang, D., 2011. Formation of disinfection byproducts in typical Chinese drinking water. J. Environ. Sci. China 23 (6), 897–903.
Moradi, S., Chow, C.W.K., 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.
Moradkhani, H., Hsu, K.-l., Gupta, H.V., Sorooshian, S., 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. J. Hydrol. 295 (1), 246–262.
Nikolaou, A.D., Lekkas, T.D., 2001. The role of natural organic matter during formation of chlorination by-products: a review. Acta Hydrochim. Hydrobiol. 29 (2–3), 63–67.
Pan, Y., Zhang, X., 2013. Four groups of new aromatic halogenated disinfection byproducts: effect of bromide concentration on their formation and speciation in chlorinated drinking water. Environ. Sci. Technol. 47 (3), 1265–1273.
Park, J., Lee, C.H., Cho, K.H., Hong, S., Kim, Y.M., Park, Y., 2018. Modeling trihalomethanes concentrations in water treatment plants using machine learning techniques. Desalin. Water Treat. 111, 125–133.
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.
Postigo, C., Emiliano, P., Barceló, D., Valero, F., 2018. Chemical characterization and relative toxicity assessment of disinfection byproduct mixtures in a large

drinking water supply network. J. Hazard Mater. 359, 166–173.
Regli, S., Chen, J., Messner, M., Elovitz, M.S., Letkiewicz, F.J., Pegram, R.A., Pepping, T.J., Richardson, S.D., Wright, J.M., 2015. Estimating potential increased bladder cancer risk due to increased bromide concentrations in sources of disinfected drinking waters. Environ. Sci. Technol. 49 (22), 13094–13102.
Richardson, S.D., Plewa, M.J., Wagner, E.D., Schoeny, R., DeMarini, D.M., 2007. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. Mutat. Res. 636 (1), 178–242.
Sadiq, R., Husain, T., Kar, S., 2002. Chloroform associated health risk assessment using bootstrapping: a case study for limited drinking water samples. Water, Air, Soil Pollut. 138 (1), 123–140.
Sadiq, R., Rodriguez, M.J., 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Sci. Total Environ. 321 (1), 21–46.
Singh, K.P., Gupta, S., 2012. Artificial intelligence based modeling for predicting the disinfection by-products in water. Chemometr. 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.
Song, Q., Ning, P., Sun, H., Lin, H., Chen, J., Hong, H., 2017. Regression models of HAAs formation upon chlorination of source water collected from Yangtze River Delta. Acta Sci. Circumstantiae 37 (6), 2048–2054 (in Chinese).
Sreekanth, S., Ramaswamy, H.S., Sablani, S.S., Prasher, S.O., 2010. A neural network approach for evaluation of surface heat transfer coefficient. J. Food Proc. 23 (4), 329–348.
Sun, H.-J., Zhang, Y., Zhang, J.-Y., Lin, H., Chen, J., Hong, H., 2019. The toxicity of 2,6-dichlorobenzoquinone on the early life stage of zebrafish: a survey on the endpoints at developmental toxicity, oxidative stress, genotoxicity and cytotoxicity. Environ. Pollut. 245, 719–724.
Sun, H., Song, X., Ye, T., Hu, J., Hong, H., Chen, J., Lin, H., Yu, H., 2018. Formation of disinfection by-products during chlorination of organic matter from phoenix tree leaves and Chlorella vulgaris. Environ. Pollut. 243, 1887–1893.
USEPA, 2003. Method 552.3: Determination of Haloacetic Acids and Dalapon in Drinking Water by Liquid-Liquid Microextraction, Derivatization, and Gas Chromatography with Electron Capture Detection. EPA 815-B-03-002. Revision 1.0.
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.
Wang, Y.-X., Zeng, Q., Wang, L., Huang, Y.-H., Lu, Z.-W., Wang, P., He, M.-J., Huang, X., Lu, W.-Q., 2014. Temporal variability in urinary levels of drinking water disinfection byproducts dichloroacetic acid and trichloroacetic acid among men. Environ. Res. 135, 126–132.
Weisel, C.P., Kim, H., Haltmeier, P., 1999. Exposure estimates to disinfection by-products of chlorinated drinking water. Environ. Health Perspect. 107 (2), 103–110.
Wright, J.M., Evans, A., Kaufman, J.A., Rivera-Núñez, Z., Narotsky, M.G., 2017. Disinfection by-product exposures and the risk of specific cardiac birth defects. Environ. Health Perspect. 125 (2), 269–277.
Wu, B., Zhang, Y., Hong, H., Hu, M., Liu, H., Chen, X., Liang, Y., 2019. Hydrophobic organic compounds in drinking water reservoirs: toxic effects of chlorination and protective effects of dietary antioxidants against disinfection by-products. Water Res. 166, 115041.
Yan, M., Korshin, G.V., Chang, H.-S., 2014. Examination of disinfection by-product (DBP) formation in source waters: a study using log-transformed differential spectra. Water Res. 50, 179–188.
Yang, X., Shang, C., 2004. Chlorination byproduct formation in the presence of humic acid, model nitrogenous organic compounds, ammonia, and bromide. Environ. Sci. Technol. 38 (19), 4995–5001.
Ye, B., Wang, W., Yang, L., Wei, J., Xueli, E., 2011. Formation and modeling of disinfection by-products in drinking water of six cities in China. J. Environ. Monit. 13 (5), 1271–1275.
You, Y., Nikolaou, M., 1993. Dynamic process modeling with recurrent neural networks. AIChE J. 39 (10), 1654–1667.
Zhai, H., Zhang, X., 2011. Formation and decomposition of new and unknown polar brominated disinfection byproducts during chlorination. Environ. Sci. Technol. 45 (6), 2194–2201.
Zhang, Y., Sun, H.-J., Zhang, J.-Y., Ndayambaje, E., Lin, H., Chen, J., Hong, H., 2019. Chronic exposure to dichloroacetamide induces biochemical and histopathological changes in the gills of zebrafish. Environ. Toxicol. 34 (7), 781–787.
Zhao, Z., Lou, Y., Chen, Y., Lin, H., Li, R., Yu, G., 2019. Prediction of interfacial interactions related with membrane fouling in a membrane bioreactor based on radial basis function artificial neural network (ANN). Bioresour. Technol. 282, 262–268.
Zheng, L., Sun, H., Wu, C., Wang, Y., Zhang, Y., Ma, G., Lin, H., Chen, J., 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., Gakoko Raphael, B.M., Song, Q., Wu, F., Chen, J., Lin, H., Hong, H., 2019. Factors influencing DBPs occurrence in tap water of Jinhua region in Zhejiang Province, China. Ecotox. Environ. Safe. 171, 813–822.