PERGAMON

Water Research 37 (2003) 4637–4644

WATER RESEARCH

www.elsevier.com/locate/watres

Formation, modeling and validation of trihalomethanes (THM) in Malaysian drinking water: a case study in the districts of Tampin, Negeri Sembilan and Sabak Bernam, Selangor, Malaysia

Md. Pauzi Abdullah[a,*], C.H. Yew[a], Mohamad Salleh bin Ramli[b]

[a] *Faculty of Science & Technology, School of Chemical Sciences and Food Technology, Universiti Kebangsaan, Bangi 43600, Malaysia*
[b] *Department of Chemistry, Jalan Sultan, 46661 Petaling Jaya, Selangor, Malaysia*

Received 9 January 2003; received in revised form 24 July 2003; accepted 31 July 2003

Abstract

A modeling procedure that predicts trihalomethane (THM) formation from field sampling at the treatment plant and along its distribution system using Tampin district, Negeri Sembilan and Sabak Bernam district, Selangor as sources of data were studied and developed. Using Pearson method of correlation, the organic matter measured as TOC showed a positive correlation with formation of THM ($r = 0.380, P = 0.0001$ for Tampin and $r = 0.478, P = 0.0001$ for Sabak Bernam). Similar positive correlation was also obtained for pH in both districts with Tampin ($r = 0.362, P = 0.0010$) and Sabak Bernam ($r = 0.215, P = 0.0010$). Chlorine dosage was also found to have low correlation with formation of THM for the two districts with Tampin ($r = 0.233, P = 0.0230$) and Sabak Bernam ($r = 0.505, P = 0.0001$). Distance from treatment plant was found to have correlation with formation of THM for Tampin district with $r = 0.353$ and $P = 0.0010$. Other parameters such as turbidity, ammonia, temperature and residue chlorine were found to have no correlation with formation of THM. Linear and non-linear models were developed for these two districts. The results obtained were validated using three different sets of field data obtained from own source and district of Seremban (Pantai and Sg. Terip), Negeri Sembilan. Validation results indicated that there was significant difference in the predictive and determined values of THM when two sets of data from districts of Seremban were used with an exception of field data of Sg. Terip for non-linear model developed for district of Tampin. It was found that a non-linear model is slightly better than linear model in terms of percentage prediction errors. The models developed were site specific and the predictive capabilities in the distribution systems vary with different environmental conditions.

Keywords: Trihalomethane; Formation; Modeling; Validation; Water treatment; Disinfection by-product

1. Introduction

Disinfection is the most important stage in the treatment of drinking water supplies, because it removes or inactivates pathogenic organisms responsible for waterborne diseases such as cholera, typhoid fever, and dysentery [1]. Most municipal water supply system in Malaysia use chlorination for water disinfections [2]. Chlorination is a widely used method of disinfections because it is extremely efficient and cost effective. Although, chlorination worked well, it was discovered that the use of chlorine posed potential health risks due to the formation of carcinogenic halo-organic compounds, as disinfections by-products [3,4] and major of which is trihalomethane (THM).

*Corresponding author. Tel.: +603-892-15447.
E-mail address: mpauzi@pkrisc.cc.ukm.my (Md.P. Abdullah).

0043-1354/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.watres.2003.07.005

As such the formation of THM during chlorination process is important and need to be monitored with the view to ensure the compliance of the guidelines set. A more effective approach can be obtained through the use of predictive models. The modeling of THM consists of establishing empirical or mechanistic relationships between THM levels in treated water, and the parameters of water quality and its operational control that can be linked to their formation.

This paper presents the development of THM predictive models with a particular focus on Tampin and Sabak Bernam districts' water treatment plants and its distributive system. The models developed were than validated using three sets of data obtained from own source and district of Seremban (Pantai and Sg. Terip), Negeri Sembilan.

Validation of a fitted regression is the demonstration or confirmation that the model is sound and effective for the purpose for which it was intended. Validation model requires assessing the effectiveness of the fitted equation against an independence set of data, and is essential if confidence in the model is to be expected.

2. Experimental

2.1. Determination of THM

A rapid and simple method for THM analysis by Purge and Trap coupled with capillary column gas chromatograph with electron capture detector was used to analyze THM level for the samples under studies. A Hewlett-Packard HP 5690 series II gas Chromatograph equipped with electron capture detector and HP chemstation was used to generate THM data.

THM and other volatile compounds are extracted (purged) from sample matrix by bubbling an inert gas (nitrogen) through the aqueous sample. Purged volatile components are trapped in a tube (trap) containing suitable sorbent materials. When purging is complete, the sorbent tuber (or trap) is heated and backflushed with an inert gas to desorb trapped sample components onto a gas chromatography (GC) column. The gas chromatograph is temperature-programmed to separate the method analytes, which are then detected with electron capture detector.

TOC analyzer (Shimadzu, model TOC-ASI-5000A) was used to analyze the value of total organic carbon (TOC). Ammonia was analyzed using Brab + Luebbe TRAACS 800 System.

2.2. Water treatment process

The objective of water treatment process in these districts was to produce water of acceptable quality suitable for human consumption complying with established standards. The process generally involves aeration, coagulation, flocculation, sedimentation, filtration and chlorination/fluoridization. Of interest to note is that the raw water at Pantai intake was found to be relatively unpolluted and its turbidity values were also relatively low (less than 5 NTU). Due to this good water quality the water treatment process in Pantai treatment plant, Seremban was even simpler compared to treatment plants in Sg. Terip, Seremban, ISO Dangi, Tampin and Bernam Headworks in Sabak Bernam where flocculation, sedimentation and filtration were waived.

2.3. Data collection and area of study for modeling and validation of THM formation model

Due to the rapid chemical changes that occur in water samples during transit and storage, certain parameters were measured on site, immediately after the sample is taken. These are temperature, pH, chlorine dosage and residual chlorine. The duration of this monitoring project were for 12 months period in the year of 2001. Based on the past data of THM level, the districts of Tampin of Negeri Sembilan and Sabak Bernam of Selangor were selected as both of them had a relatively high level of THM compared to other districts.

Our assumptions were based on common sense, literature statement and complimentary model with developed data generated by others. Working with field data obtained from Tampin and Sabak Bernam districts multivariate regression model for THM formation will be created using a step-by-step (stepwise) procedure with statistical package for social sciences (SPSS). This method consists of firstly classified the predictor variables according to their statistical significance. Pearson correlation coefficient was used to measure the strength of the relation between variables. The letter r represents the sample correlation coefficient and P represents the correlation coefficient of the population. We then include one variable at a time at different step. During the process of model development, several linear and non-linear regression structures were considered.

Independent sets of data from the same and different sources were collected for purposes of validating the model generated. In the case of Tampin and Sabak Bernam district, a total of 34 and 47 samples were collected, respectively, for the period of three months (January–March) in the year 2002. In addition to the above, Pantai and Sungai Terip distribution system in the districts of Seremban, Negeri Sembilan was also selected for this purpose. Duration of samples collection was for a period of three months (April–June) in the year of 2002. For purpose of validation, a total of 40 and 52 samples were collected from Sg. Terip and Pantai distributive system, respectively, in the district of Seremban. Due to the rapid chemical changes that occur in water samples during transit and storage, certain parameters were measured on site, immediately after the sample is taken. These are temperature, pH, chlorine dosage and residual chlorine.

3. Results and discussion

3.1. Modeling of THM formation

A total of 74 samples were collected for the district of Tampin for the year 2001. In the case of Sabak Bernam district, a total of 117 samples were collected for the same year. The drinking water quality parameters in both districts for the year 2001 are shown in Table 1. The mean value for ammonia, chlorine dosage and residue chlorine in Tampin were observed to be higher compared to Sabak Bernam. Other drinking water quality parameters such as pH and temperature did not show any significant difference between the two districts. However, the turbidity value in Tampin district is found to be higher compared to Sabak Bernam. This is due to the water quality at different point of intake and environment.

Three compounds (THM) were detected in majority of the samples. Only chloroform, dichlorobromomethane and dibromochloromethane could be quantified in all these samples. Bromoform was not detected for all samples under reviewed. Table 2 shows the mean value of chloroform and THM at each sampling station. Chloroform constitutes the major component in THM. The mean value of chloroform and THM at Tampin district range from 14.84 to 55.62 μg/l and 18.59 to 68.82 μg/l, respectively. The standard deviation for chloroform and THM at this district ranges from 10.99 to 34.40 and 12.87 to 28.48, respectively. It would appear the standard deviation for the field samples were high compared to the relative standard deviation set by IUPAC [5].

The mean value of chloroform and THM at Sabak Bernam district range from 42.92 to 81.00 μg/l and 54.64 to 89.83 μg/l, respectively. The standard deviations for chloroform and THM at this district range from 16.20 to 36.84 and 19.79 to 39.20, respectively. It would appear the standard deviation for the field samples were also high compared to the relative standard deviation set by IUPAC [5].

3.2. Effect of TOC

Using Pearson correlation method, a low but definite with small relationship ($r = 0.380$) was obtained between THM formation and TOC for Tampin district as shown in Table 3. However, moderate correlation with substantial relationship ($r = 0.478$) was obtained for Sabak Bernam district. This classification is in accordance to Guilford's rule of thumb [6]. Most investigators found that THM formation rose with increasing soluble humic material content in natural occurring water. The rate of THM formation is equal to that of the TOC consumption. A first-order reaction was reported with respect to TOC [7,8]. Indeed a higher available TOC will provide more THM if enough residue chlorine is available. Fulvic acid account for over 90% of the aquatic humic in many water sources. Babcock and Singer [8] found that relative contributions to THM production came from humic fraction than fulvic fraction since the former reacts more readily with chlorine. As a result of slow reaction between THM precursors with chlorination, THM formation with respect to TOC is of second-order reaction especially for long-term formation of THM [9]. THM production can thus be explained as a multi-stage process involving an initial fast reaction of chlorine with the TOC to produce chlorinated intermediates, which may then undergo further slow reaction by several possible pathways to produce THM and other products [7].

Table 1
Drinking water quality parameters for district of Tampin and Sabak Bernam for the year of 2001

Water quality parameter	Year 2001			
	Tampin		Sabak Bernam	
	Mean	Range	Mean	Range
1. pH	7.2	6.0–9.0	7.4	6.1–8.2
2. Turbidity (NTU)	3.85	1.36–16.00	0.68	0.21–2.83
3. Ammonia (mg/l)	0.0096	0.0000–0.0800	0.0086	0.0000–0.0835
4. TOC (mg/l)	2.160	1.244–3.731	2.223	1.705–3.330
5. Temperature (°C)	29.1	26.5–32.2	29.6	28.0–32.2
6. Cl_2 Dosage (mg/l)	4.262	3.481–4.773	2.493	2.260–2.730
7. Cl_2 residue (mg/l)	1.926	0.500–4.000	1.103	0.200–2.000
8. THM (μg/l)	0.0428	0.0024–0.1204	0.0699	0.0077–0.1365

Table 2
Mean values of concentration of THM and $CHCl_3$ at each sampling point in the distribution system for district of Tampin and Sabak Bernam for the year of 2001

Tampin sampling point[a]	THM (μg/l)		$CHCl_3$ (μg/l)	
	$x \pm s$	Min/max[b]	$x \pm s$	Min/max[b]
B002-Clear Water Tank (12)	31.85±12.87	8.50/49.30	34.34±10.99	6.30/37.50
B003-SRO (12)	18.59±26.77	2.40/54.50	14.84±17.75	1.50/42.30
A004-Kg. Keru (6)	68.82±38.05	25.40/120.40	55.62±34.40	18.40/106.00
A007-Bt. 1, Jln. Seremban (6)	50.28±27.75	3.96/67.70	41.87±24.10	3.02/58.10
A011-Pekan Air Kuning Selatan (5)	56.62±26.10	26.71/97.50	46.72±23.38	20.10/83.30
A012-Kg. Repah (7)	41.29±25.53	5.98/82.81	32.83±23.82	3.58/69.20
A014-Pekan Gemencheh (6)	42.52±20.62	23.30/77.80	41.05±16.21	21.70/60.30
A015-Pekan Btg. Melaka (6)	47.75±28.48	11.00/85.60	39.76±24.49	8.25/70.60
A016-Low Level SRO (7)	43.04±17.43	13.40/70.20	35.41±14.60	9.95/55.30
A019-Hospital Tampin (7)	41.34±21.16	12.90/68.30	31.86±16.65	9.60/54.70
Sabak Bernam sampling point[a]				
B002-Loji BRH Sg. Dusun (12)	54.64±36.53	12.30/119.80	42.92±30.26	8.80/108.00
B005-Tangki Balancing (8)	62.54±34.94	7.70/111.30	48.82±28.52	5.90/98.70
B009-Tangki Sg. Besar (9)	77.63±28.92	48.10/120.00	62.74±24.52	43.00/106.00
B013-Tangki Sabak Bernam (9)	84.96±27.77	52.40/135.70	51.30±16.20	42.70/121.00
B017-Tangki Spg. 4, BNO (9)	70.82±28.78	46.20/127.90	63.62±26.29	40.20/114.00
B020-Tangki Sg. Nibung (7)	72.22±33.82	30.30/134.30	63.90±30.84	23.90/120.00
B023-Tangki Tok Khalifah (10)	52.00±28.82	29.30/111.30	50.30±22.21	23.20/92.00
A010-Pekan Sg. Besar (4)	83.65±39.20	54.10/132.80	75.80±36.84	48.20/123.00
A011-Parit 6 (5)	69.40 ± 37.32	32.30/130.30	61.60 ± 36.21	25.20/121.00
A012-Bagan Terap (6)	71.07±28.73	54.90/127.10	63.17±27.93	45.50/118.00
A014-Hospital Sabak Bernam (6)	76.74±34.37	41.50/136.50	68.60±19.62	33.80/125.00
A015-Pekan Sabak Bernam (6)	75.25±33.35	41.50/133.50	66.83±31.54	33.70/122.00
A016-Sg. Pulai (7)	74.33±30.83	55.40/126.00	65.70±30.04	48.90/116.00
A018-Sg. Air Tawar (5)	77.80±29.79	38.90/115.50	69.48±28.48	31.30/106.00
A021-Pasir Panjang (5)	43.68±19.79	31.60/86.20	49.46±21.58	24.60/83.90
A022-Tanjung Medan (5)	66.95±36.41	25.70/124.70	58.88±34.06	19.60/113.00
A024-Kg. Baru (4)	89.83±30.57	55.00/129.00	81.00±28.29	49.50/118.00

[a] Numbers in parentheses are the number of measurements at each point.
[b] Minimum and maximum concentrations during the sampling periods.

Table 3
Relationship between formation of dependent variable THM with independent variables from water treatment plant and its distribution system in Tampin and Sabak Bernam district

THM formation with independent variable	Tampin ($n = 74$)		Sabak Bernam ($n = 117$)	
	Pearson r	P	Pearson r	P
TOC	0.380	0.0001	0.478	0.0001
pH	0.362	0.0010	0.215	0.0010
Chlorine dosage	0.233	0.0230	0.505	0.0001
Distance from treatment plant	0.353	0.0010	0.205	0.0130
Chlorine residue	−0.311	0.0030	−0.134	0.0740

3.3. Effect of pH

Simple regression analysis was used to examine the correlation of THM with respect to pH measured at all the sampling points in each distribution systems is as shown in Table 3. Pearson method of correlation was applied and a low correlation, definite with small relationship was obtained between THM formation and pH for both districts.

In general, the rate of THM production increases with pH [10]. Three-fold increase in the reaction rate per unit pH was reported by Kavanough [11]. According to Adin et al. [12], pH has two effects; decreased pH resulted in low THM formation and similarly increased pH results in high THM formation [12]. This is due to the fact that the initial attack is dependent on HClO concentration, which is related to pH. The lower the pH the higher HClO concentration resulting in a shift to higher concentration of humics. Formation of THM depends mainly on the last step of THM reaction pathway, which is base-catalyst as with the haloform reaction. This effect was also noted by other investigators [13,14], who reported decreased THM formation as a result of lowering the pH. Gracia-Villanova et al. [15] reported that there was a linear relationship between pH and THM formation. This parameter seems to be very important in controlling THM formation.

3.4. Effect of chlorine dosage

Attempts were made to determine the effect of chlorination dosage on the production of THM in the distribution system. Samples were obtained from the sampling points at two districts for the whole of year 2001. Using Pearson correlation method, a low but definite small relationship ($r = 0.233$) was obtained between THM formation and total chlorine dosage for Tampin district. However, moderate correlation with substantial relationship ($r = 0.505$) was obtained for Sabak Bernam district. This classification is in accordance to Guilford's rule of thumb [6]. Table 3 shows the regression test for correlation of THM formation with total chlorine dosage for these two distributive systems. The difference in the value of r may be due to the quality of water source, water temperature and also the material properties of the water pipes [16]. Two- and four-fold differences may occur within a water distribution system with different pattern of by-products from different sources [17] and this may vary the water quality significantly.

3.5. Effect of distance from treatment plant

A plot of distance against THM indicated that there was relationship between level of THM and distance for Tampin and also Sabak Bernam district. Table 3 shows simple regression test for THM level versus distance from treatment plant in the distribution system. Effect of distance from treatment plant in the distribution system is an indirect way of measuring the contact time between chlorine dosage and TOC. Using Pearson method of correlation, a low but definite with small relationship was obtained between THM formation and distance for both districts. Rajan et al. [18] reported that there was a direct distance-THM correlation in stimulated distribution system (SDS) [18]. However, Badawy [19] reported the distance-THM correlation to be an inverse correlation. No explanation was given for these phenomena by any of these authors. On of the possible cause of these differences phenomenon could be due to volatilization. In the case of Tampin and Sabak Bernam Districts, there were ventilated service reservoirs along the sampling line which resulted in volatilization. This is the cause for the low but definite small relationship between THM formation and its distance from the treatment plant.

3.6. Effect of residue chlorine

Attempts were made to determine the effect of residue chlorine on the production of THM in the distribution system. Using Pearson correlation method, a negative low but definite with small relationship ($r = -0.311$) was obtained between THM formation and residue chlorine for Tampin district. Addition of chlorine to water leads to the formation of hypochlorous acid (HOCl) and hypochloride ion (OCl^-). The formation of these species depends on the pH. In acidic solution HOCl is dominant, whereas in the alkaline solution formation of OCl^- dominates. In the case of Tampin, the pH value of the drinking water ranges from 6 to 9. As such hypochlorous acid is more prevalent chlorine specie which is responsible for the formation of THM. As such when THM concentration increases, the concentration of HOCl decreased, which in this case is also the residue chlorine. However, no correlation ($r = -0.134$) was obtained for Sabak Bernam district. Table 3 shows the regression test for dependence of production of THM with residue chlorine for the two distribution systems. The difference in the value of r may be due to the quality of the water source, water temperature and also the material properties of the water pipes [16].

3.7. Effect of temperature

Both simple regression tests indicated no correlation between temperature on the formation of THM for the district of Sabak Bernam and Tampin. As the sampling sites fall in the equatorial climate region, nearly all the samples under studies had constant values of temperature about 28–32°C. This effect may be the cause of no correlation between formation of THM and temperature.

In the studies on the effect of temperature on THM formation, Kohei et al. [20] and Peters et al. [13] found an Arrhenius dependency between the rate constant and temperature with activation energy of 10–20 kJ mol^{-1}. Gracia-Villanova et al. [15] reported that there was no linear relationship between temperature and THM formation. However, he found that the critical temperature value (T_c) for THM formation was 18.97°C. THM level was reduced drastically when the temperature was increased above T_c value. This could be uncounted for in terms of a shift in the extent of two phenomena:

Table 4
Results of statistical regression with field scale data for Tampin and Sabak Bernam districts

	Tampin		Sabak Bernam	
	Linear	Non-linear	Linear	Non-linear
Coefficient of correlation (r)	0.667	0.605	0.735	0.662
Model significance	$p<0.0001$	$p<0.0001$	$p<0.0001$	$p<0.0001$
Statistical coefficients:				
a	−0.155	10−6.777	−0.472	10−6.243
b	0.017	1.171	0.035	1.342
c	0.001	—	—	—
d	0.010	4.469	0.029	3.952
e	0.017	1.765	0.101	2.902
f	—	—	—	—

temperature increases the rate of formation up to T_c, at which the rate of removal of THMs, most likely owing to their volatility, becomes higher than their formation rate in the open system.

3.8. Effect of turbidity and ammonia

Simple regression tests indicated no correlation between turbidity and ammonia on the formation of THM for the district of Sabak Bernam and Tampin. As for turbidity factor, the values obtained were negligible and this may result in no correlation between the two variables. No correlation between ammonia and THM in the above districts may be due to the constant zero value of ammonia present in almost all the samples. However, ammonia present in the raw water or added during the treatment process can attenuate disinfection by-product (DBP) formation and result in the preferential formation of certain DBPs [21].

3.9. Suggested model

Using a SPSS statistical stepwise procedure, models for THM were developed for both districts. THM at both sites were predicted from raw water TOC, raw water as well as filtered water pH, distance from treatment plant and total chlorine dosage used at treatment plant. Using a non-linear and linear regression, THM was predicted from the same variables for both districts. The regression model obtained for Tampin district are as follows:

$$\text{THM} = a(\text{TOC})^b(\text{pH})^d(\text{D})^e$$

and

$$\text{THM} = a + b(\text{TOC}) + c(t) + d(\text{pH}) + e(D),$$

where TOC is the total organic carbon expressed in mg/l; t the contact time expressed in h; D the chlorine dose expressed in mg/l; and a, b, c, d, e the estimated values of statistical coefficients.

Same regression model was also obtained for Sabak Bernam district except its linear model where distance from the treatment plant (indirect measure of contact time) as an additional variable was found to be insignificant. The model accuracy is shown in Table 4. This model was found to be statistically significant for all four variables. Analyses shown that there is no collinearity among the four variable mentions in the above equation. In terms of coefficient of correlation, it appears that the linear correlation model is slightly better than the non-linear correlation model. However, the difference is not very significant.

The models were developed with a relatively small data sample because not all values of predictive parameters were available for the entire database ($n = 74$ for Tampin and $n = 117$ for Sabak Bernam). As can be seen from Table 4, all the variables are statistically significant. The coefficients of correlation of these models are less impressive in comparison with the performance of the model created from bench-scale data reported by previous literature [16]. The difference in the value of r may be due to the quality of the source water, water temperature and also the material properties of the water pipes [17]. In addition to the above, the estimated reaction time of chlorine in these utilities is not possible because each system may have specific hydraulic and flow rate conditions (present or absence of storage tank, water demand patterns, etc.).

3.10. Model validation

Validation of a fitted regression is the demonstration or confirmation that the model is sound and effective for the purpose for which it was intended. Validation model requires assessing the effectiveness of the fitted equation against an independence set of data, and is essential if confidence in the model is to be expected. Analyses were

Table 5
Tampin and Sabak Bernam model validation

Validation parameters		Tampin		Sabak Bernam	
		Linear	Non-linear	Linear	Non-linear
Same source		$N = 34$	$N = 34$	$N = 47$	$N = 47$
% Error of prediction		80.4	13.3	87.4	14.9
	%	74.0	2.3	0.4	2.3
Prediction bias	$t_{calculated}$	8.839	0.758	0.537	0.728
	$t_{critical}$	2.032	2.032	2.010	2.010
	Significant	Yes	No	No	No
Sg. Terip (Seremban)		$N = 40$	$N = 40$	$N = 40$	$N = 40$
% Error of prediction		55.0	13.0	290.0	51.8
	%	43.0	5.8	286.0	50.4
Prediction bias	$t_{calculated}$	4.262	1.561	28.008	13.567
	$t_{critical}$	2.021	2.021	2.021	2.021
	Significant	Yes	No	Yes	Yes
Pantai (Seremban)		$N = 52$	$N = 52$	$N = 52$	$N = 52$
% Error of prediction		48.7	34.3	308.0	53.5
	%	22.8	13.0	288.0	43.0
Prediction bias	$t_{calculated}$	3.094	7.111	3.917	10.296
	$t_{critical}$	2.000	2.000	2.000	2.000
	Significant	Yes	Yes	Yes	Yes

done to determine the mean square error of prediction (MSEP). MSEP is defined as the average square difference between independent observation and prediction from the fitted equation for the corresponding values of the independent variable [22]. It incorporates both the variance of prediction and the square of the bias of the prediction. A t-test was done on the predicted models and to determine its biasness by calculating the t_{value} for all these models compared to the $t_{critical}$ value. If $t_{value} < t_{critical}$, then its biasness is considered to be not significant and vice versa. Table 5 shows model validation results of Tampin and Sabak Bernam.

Using independent set of data obtained from Tampin and Sabak Bernam and based on the calculation of percentage error in prediction, it was found that linear model have higher percentage value of prediction error (error in prediction 80.4% for Tampin and 87.4% for Sabak Bernam) compared to the non-linear models (error in prediction 13.3% for Tampin and 14.9% for Sabak Bernam). Using the data obtained from Sg. Terip (different source) it was found that linear model have higher percentage value of prediction error (error in prediction 55.0% for Tampin and 290.0% for Sabak Bernam) compared to the non-linear models (error in prediction 13.0% for Tampin and 51.8% for Sabak Bernam). Similarly using the data obtained from Pantai (different source) it was found that linear model have higher percentage value of prediction error (error in prediction 47.8% for Tampin and 308.0% for Sabak Bernam) compared to the non-linear models (error in prediction 34.3% for Tampin and 53.5% for Sabak Bernam).

Based on the results of validation, it was found that non-linear model of Tampin is better compared to the linear model. Similarly non-linear model of Sabak Bernam appears to be better than its linear model. It was also found that independent set of data from different source, i.e. Sg. Terip has produced no significant error of prediction bias on the non-linear model generated from Tampin district. This could be explained by the fact that both Tampin and Sg. Terip, Seremban has similar water treatment process and distributive systems with only two service reservoirs at each system. In the case of Pantai, its distributive system is similar to Tampin but its water treatment process is different where process of flocculation, sedimentation and filtration were waived. This may result in the prediction error obtained during validation of Tampin's model to be significant. In the case of Sabak Bernam's model, result indicates that different source of data produced significant error of prediction and has significant prediction bias. This can be explained by the fact that the water quality in Sabak Bernam is different from the one in district of Seremban where the raw water has high quantity of TOC especially during raining season. In addition to the above the distributive system of Sabak Bernam has many service reservoirs along its distribution line. This may result in the significant prediction bias during model evaluation.

4. Conclusion

The models described above can be used to estimate the THM concentration for different water quality and operational conditions in distribution system where chlorination is the only treatment in the process.

Based on the validation results obtained, it would appear that the model developed is site specific and the predictive capabilities in the distribution systems vary with different environmental conditions. Non-linear model developed appeared to be better than the linear model in terms of percentage error of prediction. This indicated that the non-linear models developed could be used to estimate the THM concentration for different water quality and operational conditions where chlorination is the only treatment in the process. Due to the fact that very little data about THM are currently available for distribution system in Malaysia, the generation of data using a modeling approached is extremely useful. However, with given complexity of water quality, evolution in distribution system, modeling results must be interpreted with caution.

Acknowledgements

The authors gratefully acknowledge permission from Department of Chemistry Malaysia to use the laboratory facilities as part of their water quality research programme. The financial support of Malaysian Government under R&D grant, IRPA programme 08-02-02-0042 is gratefully acknowledged. Public Service Department is also acknowledged for financial support to C.H. Yew. We wish to express our thanks to SC Soh for typing the report.

References

[1] White GC. Handbook of chlorination and alternative disinfections, 3rd Ed. New York: Van Nostrand Reinhold Publisher, Inc; 1992.
[2] Sukiman S, Pauzi A. Chemical quality of Malaysian drinking water sources. Drinking Water Quality: Microbiological and Public Health Aspects, UKM; 1993. p. 63–9.
[3] Rook JJ. Formation of haloforms during chlorination of natural water. Water Treat Exam 1974;23:234–43.
[4] Bellar TA, Lichtenberg JJ, Kroner RC. The occurrence of organohalogens in chlorinated drinking waters. J AWWA 1974;66(11):703–6.
[5] IUPAC. Acceptable RSD for repeatability. Pure Appl Chem 1990;62:149–58.
[6] Guildford JP. Fundamental statistics in psychology and education. New York: McGraw-Hill; 1956.
[7] Trussell AR, Umphries MD. The formation of trihalomethanes. J AWWA 1978;70:604–12.
[8] Babcock D, Singer P. Chlorination and coagulation of humic and fulvic acids. J AWWA 1979;71:149–52.
[9] Gallard H, Gunter UV. Chlorination of natural organic matter: kinetics of chlorination and of THM formation. Water Res 2002;36:65–74.
[10] Steven AA, Slocum CJ, Seeger DR, Robeck GC. Chlorination of organics in drinking water. J AWWA 1976;68:615–20.
[11] Kavanaugh MC, Trussell AR, Cromer J, Trussell RR. An empirical kinetic model for trihalomethanes formation: applications to meet the proposed THM standard. J AWWA 1980;72:578–82.
[12] Adin A, Katzhendler J, Alkaslassy D, Rav AC. Trihalomethanes formation in chlorinated drinking water: a kinetic model. Water Res 1991;25:797–805.
[13] Peters CJ, Young RJ, Perry R. Factors influencing the formation of Haloforms in the chlorination of humic materials. Environ Sci Technol 1980;14:1391–5.
[14] Sandler R. Untersuchungen sur buildung von organischen chlorrerbindungen bei der chlorung huminsaurehaltiger wasser. Thesis Ph.D. Engler-Bunte Institute, University of Karlsruhe, 1977.
[15] Garcia-Villanova RJ, Garcia C, Gomez JA, Garcia MM, Ardanny R. Formation, evolution and modelling of trihalomethanes in the drinking water of a town: I. At the municipal treatment utilities. J Water Res 1997;41:251–5.
[16] Rodriguez MJ, Serodes J, Morin M. Estimation of water utility compliance with trihalomethane regulations using a modeling approach. J Water Supply Res Technol -Aqua 2000;49:57–73.
[17] Clark RM, Adams JQ. Control of disinfection by-products: economic and technological considerations. Safety of water disinfection: balancing chemical and microbial risk. Washington, DC: ILSI Press; 1993.
[18] Rajan S, Azariah J, Bauer U. Trihalomethane levels in Madras public drinking water supply system and its impact on public health. Zentralbl Hyg Umwletmed 1990;189: 312–32.
[19] Badawy MI. Trihalomethane in drinking water supplies and reused water. Bull Environ Contam Toxicol 1992;48:157–62.
[20] Kohei U, Hiroshi W, Takao T. Empirical rate equation for THM formation with chlorination of humic substances in water. Water Res 1983;17:1797–802.
[21] Clark RM, Sivaganesan M. Predicting chlorine residuals and formation of TTHMs in drinking water. J Environ Eng 1998;124:1203–10.
[22] Rawlings JO. Applied regression analysis: a research tool. California: Wadsworth & Brooks Inc; 1988.