Environ Monit Assess (2009) 149:429–436
DOI 10.1007/s10661-008-0219-4

Modeling the formation of trihalomethanes in drinking waters of Lebanon

Lucy Semerjian · John Dennis · George Ayoub

Received: 18 September 2007 / Accepted: 23 January 2008 / Published online: 6 February 2008
© Springer Science + Business Media B.V. 2008

Abstract The current research aims at developing predictive models for trihalomethane (THM) formation in Lebanon based on field-scale investigations as well as laboratory controlled experimentations. Statistical analysis on field data revealed significant correlations for TTHM with chlorine dose, Specific UV-A, and UV_{254} absorbing organics. Simulated distribution system-THM tests showed significant correlations with applied chlorine dose, total organic carbon, bromides, contact time, and temperature. Predictive models, formulated using multiple regression approaches, exhibiting the highest coefficients of determination were quadratic for the directly after chlorination (DAC; r^2=0.39, p<0.036) and network (r^2=0.33, p<0.064) THM databases, and logarithmic for the laboratory simulated THM database (r^2=0.70, p<0.001). Computed r^2 values implied low correlations for the DAC and network THM database, and high correlation for the laboratory simulated THM database. Significance of the models were at the 0.05 level for DAC database, 0.10 level for the network database, and very high (<0.01 level) for the laboratory simulated THM database. It is noteworthy to mention that no previous attempts to assess, monitor, and predict THM concentrations in public drinking water have been reported for the country although a large fraction of the population consumes chlorinated public drinking water.

Keywords Chlorination · Lebanon · Regression modeling · Trihalomethanes

L. Semerjian (✉) · G. Ayoub
Department of Civil and Environmental Engineering,
American University of Beirut,
Beirut, Lebanon
e-mail: ls07@aub.edu.lb

J. Dennis
Department of Community Health Sciences,
University of Calgary,
Calgary, Alberta, Canada

Introduction

In view of the discovery of trihalomethanes (THM) as undesirable by-products of chlorination, several countries have specified maximum allowable levels for such disinfection by-products (DBPs) in drinking waters. Consequently, an interest in developing mathematical models to describe or predict THM formation in chlorinated waters was sparked. It is clear that properly developed quality models to simulate the temporal and spatial variations of THM in water treatment plant (WTP) outlets and distribution systems can serve as planning tools that potentially assist utility operators in complying with the new strict quality rules by adopting proper source water management strategies. In fact, numerous empirical models have emerged during the last decade exhibiting varying levels of success in their predictive

capabilities (Kavanaugh et al. 1980; Urano et al. 1983; Engerholm and Amy 1983; Morrow and Minear 1987; Amy et al 1987, 1998; Malcom Pirnie Inc. 1992; Montgomery Watson 1993; Ibarluzea et al. 1994; Rathbun 1996; Garcia-Villanova et al. 1997; Golfinopoulos et al. 1998; Rodriguez et al. 2000; Westerhoff et al. 2000; Golfinopoulos and Arhonditsis 2002a, b; Nikolaou et al. 2004). Such predictive models are based on data obtained from either field sampling or laboratory-scaled studies. The empirical modeling approach involves statistical analysis of the derived data to develop functional equations to predict THM concentrations based on source water characteristics and reaction conditions. Developed models are generally expressed as a function of organic precursor expressed as total organic carbon (TOC), dissolved organic carbon (DOC), chlorophyll a, UV_{254}, or fluorescence; bromide; chlorine; pH; temperature; and reaction time. Laboratory studies have been found more reliable than field-scaled studies for developing empirical models because of controlled conditions especially in terms of pH, chlorine dose, and contact time estimation. On the other hand, a major drawback of laboratory studies remains that they do not account for the effects of the distribution system also known as "biofilm and pipe effects" on residual disinfectant concentrations and DBP formation. A qualitative comparison of laboratory and field-scaled studies in the development of THM predictive models is listed in Table 1 (Rodriguez et al. 2000; Sadiq and Rodriguez 2004).

Because of the complex nature of the reactions between organic precursors and chlorine, most developed models are based on fitting mathematical equations to various empirical observations, rather than mechanistic and kinetic considerations. The complexity of DBP formation also makes it difficult to develop universally applicable models that can be employed with diverse array of natural water sources.

Table 1 A qualitative comparison of laboratory and field-scaled models

Characteristic	Laboratory-scaled	Field-scaled
Statistical significance	High	Medium
Applicability	General	Site-specific
Inclusion of biofilm and pipe material effects	No	Yes
Controllability over explanatory variables	High	Low
Predictability of actual human exposure	Medium	High
Ease in model development	High	Low
Cost and resources involved in model development	Low	High

This paper summarizes the outcomes of attempts to derive empirical THM predictive models specific to the natural waters of Lebanon by conducting proper statistical analyses on results generated from field-scaled observations as well as standard laboratory-controlled simulated distribution system-trihalomethane tests. It is noteworthy to mention that no previous attempts to assess, monitor, and predict THM concentrations in public drinking water have been reported for the country although a large fraction of the population consumes chlorinated public drinking water.

Materials and methods

Field-scale investigations

A seasonal sampling program (March–May 2003, August–September 2003, and February–March 2004) was initiated during which a total of 196 samples were collected randomly from various water sources and water distribution systems throughout Lebanon (Fig. 1). Three sampling points were selected for each sampling location to track the water from its source to the distribution network. Therefore, duplicate samples were properly collected from the raw water source, directly after chlorination (DAC), and from the distribution network, stored at 4°C in a cooler, and carried to the Environmental Engineering Research Center (EERC) at the American University of Beirut (AUB) for further analysis. Within specified allowable holding times, collected samples were processed and analyzed for total trihalomethanes (TTHM) and various physico-chemical water quality characteristics correlated with THM formation; namely, pH, temperature, TOC, UV_{254} absorbing organic constituents, bromides, and ammonia–nitrogen. Sample preservation and analytical tests were performed in accordance to the "Standard Methods for the Examination of Water and Wastewater" (Eaton et al. 2005) or United States Environmental Protection Agency (USEPA 1995). Operational factors correlated with THM formation such as applied

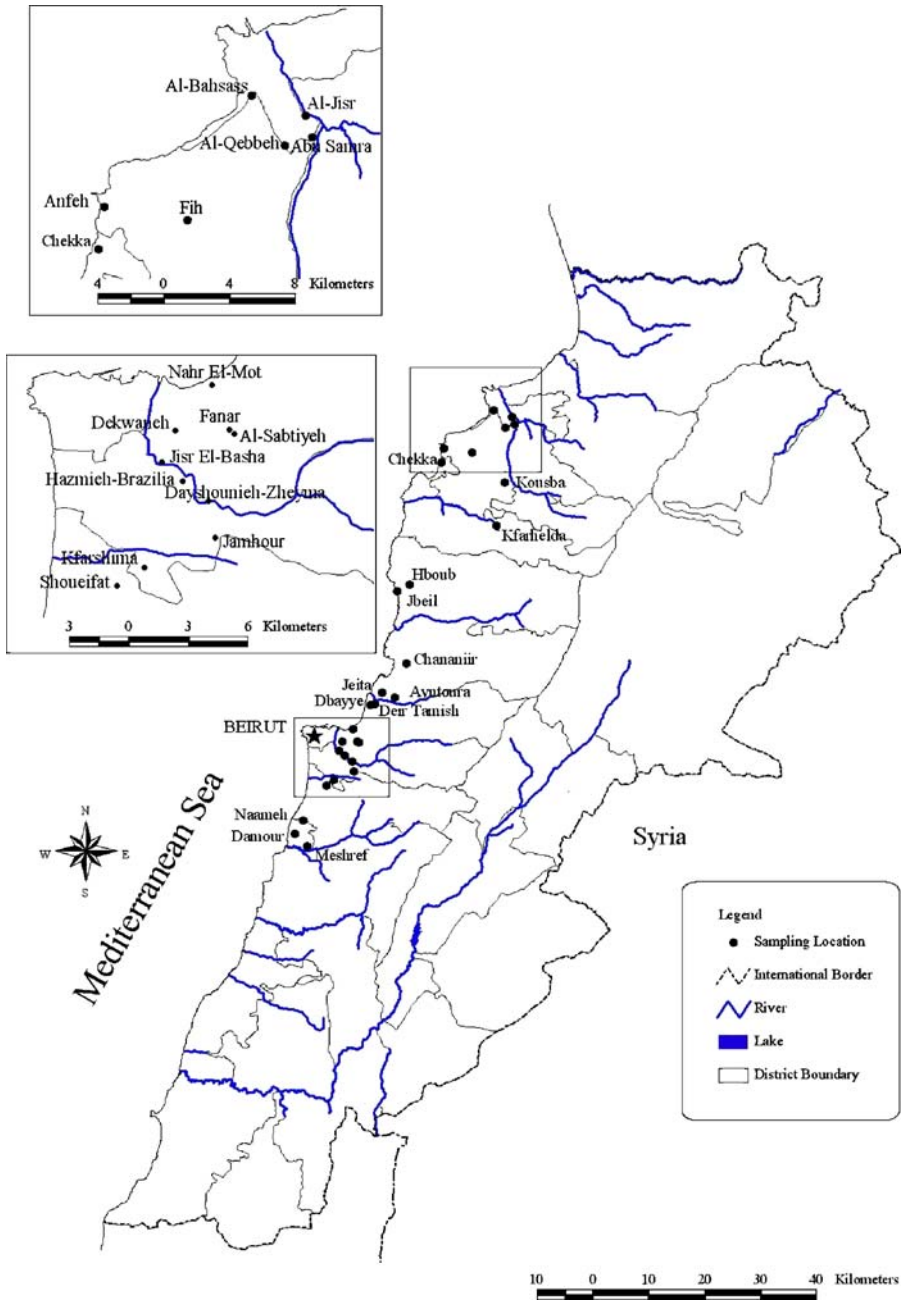

Fig. 1 Geographical distribution of investigated water distribution networks

chlorine dose as well as water residence time in networks were also recorded.

Varying yet tolerable seasonal concentrations of TTHMs were detected in the surveyed sampling locations. Also, concentrations of contributing water quality as well as operational characteristics varied with source water type, applied treatment scheme, and season (Semerjian et al. 2007a, b; Semerjian and Dennis 2007). Two datasets related to outcomes generated from the above mentioned analyses and recordings were produced. The first dataset (TTHM-DAC) relates TTHM concentrations detected in samples collected directly after chlorination to raw water source characteristics and operational variables, while the second dataset (TTHM-network) relates TTHM concentrations detected in samples collected from sampling points

within the network to raw water source characteristics and operational variables. Several approaches of statistical analyses were applied to the surveyed datasets. First, the degree of correlation between TTHM levels, and source water quality characteristics and operational variables contributing to THM formation was investigated by applying the bivariate correlation procedure to compute Pearson's correlation coefficient with a two-tailed test of significance using the statistical package for social sciences (SPSS 12, 2003). Also, several regression analysis strategies, namely, linear, logarithmic, exponential, and quadratic were applied to the field-derived datasets with the objective of formulating predictive mathematical models for THM formation specific to the chlorinated waters of Lebanon.

Laboratory-scale simulations

In addition to the field investigations, laboratory-scale simulated distribution system trihalomethane (SDS-THM) tests were conducted on selected public drinking water sources; namely the Kfar Helda WTP and Zheyma to predict as well as evaluate trihalomethane formations under controlled laboratory conditions. The SDS-THM test uses bench-scale techniques to provide an estimate of the concentration of TTHM in a sample that has been chlorinated comparably to finished drinking water, and under the same conditions and time as in a local water distribution system. Therefore, water samples were properly collected and stored at 4°C in a cooler and carried to the EERC at AUB for further analysis. On the day of collection, aliquots of the samples were tested for chlorine demand to produce a residual of 0.3–0.5 mg/L, and 0.5–1.0 mg/L free chlorine after a reaction time of 48 h. This parameter was necessary to determine the dosages of chlorine to be adopted later in the SDS-THM test to produce the desired free residual chlorine concentrations at the end of the incubation period. On the day of the SDS-THM test, samples were analyzed for pH, TDS, bromides, TOC, UV_{254} organics, UV scan, and ammonia–nitrogen. In this study, a comprehensive bench-top procedure was entailed where non-chlorinated samples were collected from the sampling locations, and in the laboratory all test variables, such as chlorine dose, residual concentrations, temperature, TOC, bromides, pH and storage time were adjusted to simulate local distribution system conditions. This technique allows the use of several concentrations or values for the same test variable thus mimicking water qualities experienced in different sampling seasons. For each of the selected sampling locations, collected samples were poured into 80 standard glass biochemical oxygen demand bottles having a capacity of 300 ml, and equipped with all-glass stoppers. These bottles were divided into five batches each of 16 bottles, and each bottle within a batch was modified differently in terms of pH adjustment, added chlorine dose, bromide concentrations, and incubation temperature. Selected pH values were 6.5, 7.0, 7.5, and 8.0 representing the range of pH values recorded in the investigated water sources. pH adjustment was done by using either 0.1 N HNO_3 or 0.1 N NaOH solutions, as required. Chlorine was added from a stock solution of commercial sodium hypochlorite in volumes dictated by the chlorine demand test performed previously to achieve free residual chlorine concentrations of 0.3–0.5 mg/L or 0.5–1.0 mg/L, thus imitating local residual chlorine levels encountered in the investigated networks. As for bromide levels, no additional bromide ions were added to the bottles to be incubated at winter conditions i.e. at a temperature of 15°C; whereas for bottles to be incubated at 28°C (summer conditions) an additional dose of 0.33 mg Br^-/L was added from a stock of potassium bromide (KBr) solution to achieve a final bromide concentration of 0.70 mg/L for Kfar Helda WTP sample, and 0.83 mg/L for Zheyma sample. TOC concentrations were kept constant for all temperatures, pH, and bromide levels, and at the level originally present in each source due to the unavailability of purified sources of humic substances. Each batch was incubated for a specific incubation period of 20 min, 1, 6, 24, or 48 h, thus covering all the reaction times encountered in the investigated sources. At the end of a specific incubation period, a sample aliquot was collected, preserved, and analyzed for THM by the liquid–liquid extraction gas chromatographic method. Also, another aliquot was taken from each glass bottle and analyzed for free residual chlorine to ensure the availability of free chlorine in the required concentrations at the end of each incubation period.

After completion of SDS-TTHM procedure and TTHM analyses, the SDS_T-TTHM dataset was generated relating recorded TTHM concentrations to simulated water source characteristics and operational variables. Similar to the datasets derived from field-

scale investigations, several approaches of statistical analyses using SPSS were applied to the SDS_T-TTHM dataset to investigate the degrees of correlation. Also, several regression analysis strategies, namely, linear, logarithmic, exponential, and quadratic were applied to the simulated datasets with the objective of formulating predictive mathematical models for THM formation specific to the chlorinated waters of Lebanon.

Results and discussion

Statistical correlation of surveyed and simulated TTHM levels with THM formation contributing factors

Using the dataset TTHM-DAC, TTHM levels recorded in samples collected directly after chlorination were correlated with raw water characteristics such as pH, temperature, bromide, TOC, UV_{254}, SUVA (UV_{254}/TOC), and ammonia as well as with operational parameters such as applied chlorine dose and DAC contact time. Using the TTHM-network dataset, TTHM levels recorded in samples collected from distribution networks were correlated with raw water characteristics such as pH, bromide, TOC, UV_{254}, SUVA, and ammonia as well as with network conditions and operational parameters such as temperature of water in the network, contact time of water with chlorine in the network, and applied chlorine dose. As for the dataset related to laboratory-scaled experiments, simulated TTHM levels were correlated with variables such as pH, TOC, bromides, contact time, incubation temperature, and applied chlorine dose. Table 2 summarizes the outcomes of bivariate Pearson correlations for the three datasets.

It can be noted that in the TTHM-DAC dataset, significant correlations are observed with chlorine dose

Table 2 Statistical correlations of surveyed and simulated TTHM with THM formation contributing factors

Independent variables		Dependent variables		
		TTHM-DAC	TTHM-network	SDS_T-TTHM
pH	Pearson correlation	0.031	0.091	0.081
	Significance	0.826	0.488	0.309
	N	53	60	160
Temperature	Pearson correlation	0.141	−0.052	0.344[a]
	Significance	0.313	0.693	0.000
	N	53	60	160
Cl_2 dose	Pearson correlation	0.341[b]	0.379[a]	0.170[b]
	Significance	0.012	0.003	0.032
	N	53	60	160
Contact time	Pearson correlation	0.079	−0.111	0.580[a]
	Significance	0.575	0.397	0.000
	N	53	60	160
Bromide	Pearson correlation	0.119	0.031	0.412[a]
	Significance	0.398	0.813	0.000
	N	53	60	160
TOC	Pearson correlation	0.087	0.185	0.249[a]
	Significance	0.537	0.158	0.002
	N	53	60	160
UV_{254}	Pearson correlation	0.438[a]	0.405[a]	ND
	Significance	0.001	0.001	
	N	53	60	
SUVA	Pearson correlation	0.276[b]	0.174	ND
	Significance	0.046	0.184	
	N	53	60	
Ammonia	Pearson correlation	0.235	0.028	ND
	Significance	0.090	0.833	
	N	53	59	

[a] Significant correlation at the 0.01 level

[b] Significant correlation at the 0.05 level

and SUVA at the 0.05 level, and with UV_{254} at the 0.01 level. As for the TTHM-network dataset, significant correlations are observed only with chlorine dose and SUVA at the 0.01 level. Finally, SDS_T-TTHM showed significant correlations with applied chlorine dose at the 0.05 level, and with TOC, bromides, contact time, and temperature at the 0.01 level.

Predictive model development for TTHM from surveyed and simulated TTHM levels

Working from the three databases, multivariate regression models for THM formation were created using the "forward" procedure with SPSS 12. THM concentrations were entered as the dependent variable while contributing source water quality characteristics and operational variables were considered as the independent variables. During model development, several regression structures were considered on each of the three datasets (DAC, network, and SDS_T), namely, linear, logarithmic, exponential, and quadratic in an attempt to formulate and adopt the equations exhibiting the highest coefficients of determination. Table 3 lists the models formulated by the different regression structures, and specifies the sample size (n), significance (p), and coefficient of correlation (r^2) for each model. It can be noted that formulated predictive models exhibiting the highest coefficients of determination are quadratic for the DAC (r^2=0.39, p<0.036) and network (r^2=0.33, p<0.064) THM databases, and logarithmic for the laboratory simulated THM database (r^2=0.70, p<0.001). Computed r^2 values imply low correlations for the DAC and network THM database, and high correlation for the laboratory simulated THM database. Significance of the models were at the 0.05 level for DAC database, 0.10 level for the network database, and very high (<0.01 level) for the laboratory simulated THM database. In previous research, laboratory-scaled studies have been found more reliable than field-scaled studies for developing empirical models because of controlled conditions especially in terms

Table 3 Formulated THM formation predictive models from surveyed and simulated TTHM levels in Lebanon

Origin of data	Formulated model	r^2	p	n
Field, DAC	Linear: $TTHM_{DAC}$=−9.22+360.97 (UV)+0.54(temp)+0.12(time)	0.38	<0.035	53
	Logarithmic: $\ln(TTHM_{DAC})$=3.70+0.41 ln(UV)+0.44ln(Br)	0.25	<0.036	
	Quadratic: $(TTHM_{DAC})^2$=17.31+10.52$(Cl_2)^2$+ 259728.60$(SUVA)^2$	0.39	<0.036	
	Exponential: $\ln(TTHM_{DAC})$=−0.21+ 62.43(UV)+0.06(temp)	0.29	<0.673	
Field, network	Linear: $TTHM_{net}$=4.01+319.88(UV)+ 2.42(Cl_2)	0.23	<0.038	60
	Logarithmic: $\ln(TTHM_{net})$=2.09+ 0.41ln(Cl_2)	0.12	<0.009	
	Quadratic: $(TTHM_{net})^2$=42.10+ 29.23$(Cl_2)^2$+353375.00$(UV)^2$	0.33	<0.064	
	Exponential: $\ln(TTHM_{net})$=1.53+ 42.21(UV)	0.11	<0.011	
Laboratory, SDS	Linear: SDS-TTHM=−12.66+0.48(time)+ 35.06(Br)+10.26(Cl_2)	0.54	<0.002	160
	Logarithmic: ln(SDS-TTHM)=6.11+ 0.211ln(time)+1.64ln(Br)+0.34ln(Cl_2)− 0.80ln(temp)	0.70	0.000	
	Quadratic: $(SDS\text{-}TTHM)^2$=−471.11+ 0.48$(time)^2$+1856.07$(Br)^2$+404.38$(Cl_2)^2$	0.31	<0.037	
	Exponential: ln(SDS-TTHM)=1.43+0.02 (time)−0.04(temp)+2.78(Br)+0.48(Cl_2)	0.61	<0.005	

TTHM (μg/L), UV (cm^{-1}), temp (°C), SUVA (L/mg cm), Br (mg/L), Cl_2 (mg/L), time (min for DAC and network datasets and h for SDS datasets)

of pH, chlorine dose, and contact time estimation (Rodriguez et al. 2000). This phenomenon is also evident in the current study.

Conclusions

Although multiple regression models developed by many researchers represent a rational framework for modeling THM formation in many sources, existing models including THM models developed in the current study may suffer from several limitations including calibration with a limited database, application to only a specific water source or condition, inadequate model validation, sometimes lack of terms to simulate important parameters, and unreliability if the value of any variable is outside the range found in the calibration data. On the other hand, such THM predictive models may have practical benefits and potential uses in Lebanon for the (1) identification of operational strategies by water utility operators that minimize THM formation; (2) selection of sampling points for water quality control within the distribution system when combined with residual disinfectant models; (3) application as decision-making tools by government officials when evaluating the technical feasibility of establishing more stringent regulations for THM, or estimating infrastructure needs for upgrading treatment facilities; and (4) improvement of decision-making for management of public health risks associated with THM (Rodriguez et al., 2000; Milot et al. 2002; Sadiq and Rodriguez 2004).

For the current study, significant correlations were observed for TTHM with chlorine dose, SUVA, and UV_{254} for the TTHM-DAC dataset. As for the TTHM-network dataset, significant correlations were observed only with chlorine dose and SUVA. SDS-$TTHM_T$ dataset showed significant correlations with applied chlorine dose, TOC, bromides, contact time, and temperature. Formulated THM formation predictive models exhibiting the highest coefficients of determination were quadratic for the DAC and network THM databases, and logarithmic for the laboratory simulated THM database.

Acknowledgment Special thanks are extended to the Environmental Engineering Research Center at the American University of Beirut, the Lebanese National Council for Scientific Research, and the concerned Water Establishments.

References

Amy, G. L., Minear, R. A., & Cooper, W. J. (1987). Testing and validation of a multiple nonlinear regression model for predicting trihalomethane formation potential. *Water Research*, *21*(6), 649–659.

Amy, G. L., Siddiqui, M., Ozekin, K., Zhu, H., Wang, C. (1998). Empirically based models for predicting chlorination and ozonation by-products: Haloacetic acids, chloral hydrate, and bromate, USEPA Report CX-819579.

Eaton, A. D., Clesceri L. S., Rice, E. W., Greenberg, A. E. (Eds.) (2005). *Standard methods for the examination of water and wastewater.* Washington, D. C: American Public Health Association, American Water Works Association, Water Environment Federation).

Engerholm, A., & Amy, G. L. (1983). Predictive model for chloroform formation from humic acid. *Journal of American Water Works Association*, *75*(8), 418–423.

Garcia-Villanova, R. J., Garcia, C., Gomez, A., Paz Garcia, M., & Ardanuy, R. (1997). Formation, evolution, and modeling of trihalomethanes in the drinking water of a town: I. At the municipal treatment utilities. *Water Research*, *31* (6), 1299–1308.

Golfinopoulos, S. K., & Arhonditsis, G. B. (2002a). Multiple regression models: A methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics. *Chemosphere*, *47*, 1007–1018.

Golfinopoulos, S. K., & Arhonditsis, G. B. (2002b). Quantitative assessment of trihalomethane formation using simulations of reaction kinetics. *Water Research*, *36*(11), 2856–2868.

Golfinopoulos, S. K., Xilourgidis, N. K., Kostopoulou, N., & Lekkas, T. D. (1998). Use of a multiple regression model for predicting trihalomethane formation. *Water Research*, *32*(9), 2821–2829.

Ibarluzea, J. M., Goni, F., & Santamaria, J. (1994). Trihalomethanes in water supplies in the San Sebastian Area, Spain. *Bulletin of Environmental Contamination and Toxicology*, *52*, 411–418.

Kavanaugh, I. C., Trussell, A. R., Cromer, J., & Rhodes, R. (1980). Empirical kinetic model of trihalomethane formation: Applications meet the proposed THM standard. *Journal of American Water Works Association*, *72*(10), 578–582.

Malcom Pirnie Inc. (1992). Water treatment plant simulation program, version 1.21, User's manual, USEPA, Office of Ground Water and Drinking Water Technology Transfer, p. 150.

Milot, J., Rodriguez, M. J., & Serodes, J. B. (2002). Contribution of neural networks for modeling trihalomethanes occurrence in drinking water. *Journal of Water Research and Planning-ASCE*, *128*(5), 370–376.

Montgomery Watson (1993). Final report: Mathematical modeling of the formation of THMs and HAAs in chlorinated waters. Denver, CO: American Water Works Association.

Morrow, C. M., & Minear, R. A. (1987). Use of regression models to link raw water characteristics to trihalomethane concentrations in drinking water. *Water Research*, *21*(1), 41–48.

Nikolaou, A. D., Lekkas, T. D., Golfinopoulos, S. K., & Kostopoulou, M. N. (2002). Application of different analytical methods for determination of volatile chlorination by-products in drinking water. *Talanta*, *56*, 717–726.

Rathbun, R. E. (1996). Regression equations for disinfection by-products for the Mississippi, Ohio and Missouri rivers. *Science of the Total Environment*, *191*, 235–244.

Rodriguez, M. J., Serodes, J., & Morin, M. (2000). Estimation of water utility compliance with trihalomethane regulations using a modeling approach. *Journal of Water Supply: Research and Technology-AQUA*, *49*(2), 57–73.

Sadiq, R., & Rodriguez, M. J. (2004). Disinfection by-products in drinking water and predictive models for their occurrence: A review. *Science of the Total Environment*, *321*, 21–46.

Semerjian, L. A., & Dennis, J. (2007). Multipathway risk assessment of trihalomethane exposure in drinking water of Lebanon. *Journal of Water and Health*, *5*(4), 511–522.

Semerjian, L. A., Dennis, J., & Ayoub, G. M. (2007a). Spatial and seasonal evolution of trihalomethane in distribution systems of Lebanon. *Journal of Water Supply: Research and Technology-AQUA*, *56*(4), 263–273.

Semerjian, L. A., Dennis, J., & Ayoub, G. M. (2007b). *Occurrence, formation, and speciation of trihalomethanes in drinking waters of Lebanon.* Paper presented at the 10th International Conference on Environmental Science and Technology, Kos, Greece.

SPSS (Statistical package for social studies) (2003). version 12.

Urano, K., Wada, H., & Takemasa, T. (1983). Empirical rate equation for trihalomethane formation with chlorination of humic substances in water. *Water Research*, *17*, 1797.

USEPA (Unites States Environmental Protection Agency) (1995). Determination of chlorination DBPs, chlorinated solvents, and halogenated pesticides/herbicides in drinking water by liquid–liquid extraction and gas chromatography with electron capture detector, Method 551.1, Office of Research and Development.

Westerhoff, P., Reckhow, D., Amy, G., Chowdhury, Z., McClellan, J., Dundorf, S. et al. (2000). *Role of five independent treatment processes NOM structure, DBP precursor removal, and DBP modeling parameters*. Paper presented at the American Water Works Association Annual Conference, Denver, CO.