PERGAMON

Chemosphere 47 (2002) 1007–1018

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Multiple regression models: A methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics

Spyros K. Golfinopoulos [a,*], George B. Arhonditsis [b]

[a] *Water and Air Quality Laboratory, Department of Environmental Studies, University of the Aegean, Karadoni 17, Mytilene 81100, Greece*

[b] *Department of Civil and Environmental Engineering, University of Washington, 313B More Hall, Box 352700, Seattle, Washington, USA*

Received 21 February 2001; received in revised form 5 September 2001; accepted 28 January 2002

Abstract

The presence of trihalomethanes (THMs) in drinking water has attracted the attention of both researchers and professionals, because of the harmful effects of these substances on human health. A multiple regression model was developed to estimate THM concentrations in finished drinking water, using data from the Menidi Treatment Plant of Athens. A number of routinely measured characteristics—including chlorine dose, chlorophyll a, temperature, pH and bromide—of raw water, were used to generate a reliable methodology for predicting both total THM and individual species concentrations. Seasonality effects were also considered during the analysis. In general, these models were found to give acceptable fits, estimating accurately lows and highs over the annual cycle. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Multiple regression model; Trihalomethanes; Drinking water; Chlorination

1. Introduction

Chlorination, the most widely technique used for drinking water disinfection, was reported to form trihalomethanes (THMs) by the reaction of halogens (most notably, chlorine or bromine) with an organic precursor (Rook, 1974). It has been demonstrated that humic substances, i.e. humic and fulvic acids, represent the predominant type of precursor found in natural waters and that chloroform is the principal THM species (Morrow and Minear, 1987). A variety of factors influence the THM formation including precursor type, precursor concentration (as total organic carbon (TOC)), chlorine dosage, temperature, pH, bromide concentrations and reaction time. These factors can affect the total yield of THMs as well as the rate at which THMs are formed (i.e. reaction kinetics). The effects (composite and singular) of these factors have been discussed in the literature (Engerholm and Amy, 1983; Elshorbagy et al., 2000).

The THMs formed during the chlorination process, representing between 5% and 20% of the chlorinated products, are chloroform ($CHCl_3$), dichlorobromomethane ($CHCl_2Br$), chlorodibromomethane ($CHClBr_2$) and bromoform ($CHBr_3$) (APHA, 1992; Jimenez et al., 1993). Their presence in nation's drinking waters is of concern from health-related aspect, since these compounds have been related to cancer and reproductive

---

[*] Corresponding author. Tel.: +30-2510-36000; fax: +30-2510-36099.

E-mail address: sgolf@env.aegean.gr (S.K. Golfinopoulos).

0045-6535/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.

PII: S0045-6535(02)00058-9

outcomes (Craun, 1991; Fayad, 1993; Simpson and Hayes, 1998; Nieuwenhuijsen et al., 2000). Consequently, new restrictive rules for maximum levels of total trihalomethanes (TTHMs) in distribution systems are imposed. For example, under the new disinfection/disinfectants by products (D/DBP) rule, maximum contaminant levels (MCL) of THM is lowered to 0.08 ppm instead of the previous limit of 0.1 ppm; whereas the possibility of specifying MCL for each individual THM species is addressed, since their health risks differ significantly (Elshorbagy et al., 2000). The European Community (EC) drinking water quality standard for TTHMs is 100 μg/l (EECD, 1997). Greece has no official guidelines or regulations for TTHMs (Golfinopoulos et al., 1998a; Golfinopoulos, 2000) and applies the EC standard.

It is evident that the development of quality models based on raw water characteristics and treatment practices can potentially assist the utilities' operators in abiding with the new strict rules. Numerous studies have emerged during the last decades using linear regression techniques to correlate TTHM formation potential with TOC, ultra-violet—254 or chlorine dosage (Harrington et al., 1992). Additional support was sought in nonlinear regression analysis, where THM values were associated successfully with laboratory chlorination data, such as pH, temperature, chlorine dose, bromide, and nonvolatile TOC level (Morrow and Minear, 1987). Moreover, the development of multiple regression models seems to be another promising method, since its application on field data from the Galatsi treatment plant (GTP) of Athens has resulted in accurate predictions of TTHM concentrations in the finished water leaving the plant (Golfinopoulos et al., 1998b).

Given the aforementioned context, the main objective of this work was to test and extrapolate the applicability of multiple regression analysis for estimating TTHM and individual THM species concentrations in drinking water. The model developed was based on experimental data from Menidi treatment plant (MTP), a plant with essential differences in raw water characteristics comparing with the rest treatment plants of Athens. Therefore, the potential generalisation of this methodology and its variation among different plants, revealing the effects of raw water characteristics and treatment practices, is another issue that is discussed.

2. Materials and methods

2.1. Samples

All data for this study were obtained from MTP of Athens, Greece. Water samples were collected from nine different points over the plant, which supplies water to nearly two million people. Lakes Mornos, Marathon and Iliki are the significant bodies of fresh water in Athens. The MTP with a design capacity of 380 000 m^3/d receives water mainly from Mornos lake and sometimes from Iliki lake. It consists of two separate units: the old (unit 1) and the new unit (unit 2) treating 250 000 and 130 000 m^3/d, respectively. Each unit is of conventional design: major features include coagulation, flocculation, sedimentation and filtration. The prechlorination is applied in the coagulation channels and the postchlorination is applied before the input of finished water in the distribution system. Fig. 1 shows each unit process and the location of sampling and chlorination points.

Sampling was carried out each month for two consecutives days, from June 1995 to August 1996, from January to October 1997 and from February to November 1998. Duplicate samples for THMs measurement were collected from each sampling location in 40-ml glass bottles and were capped with PTFE-faced silica septum (Pierce 13075), including samples containing raw water. Sample bottles were carefully filled just to overflowing, without passing air bubbles through sample or trapping air bubbles in sealed bottle. The bottles were prepared by washing with soap and water, rinsing with tap water, ultrapure water (Millipore: Milli-Ro 5 plus and Milli Q plus 185), acetone (Mallinckrodt Chemical Works St. Louis) and placing in an oven at 150 °C for 2 h. HCl (4 drops 6 N/40 ml) was added to each raw water sample to prevent biodegradation and dehydrohalogenation, while sodium thiosulfate (3 mg/40 ml) was also added to each sampling bottle as a reducing agent (APHA, 1989). After sampling, the bottles stored in the dark at temperatures between 0 and 4 °C, were carried by air to Mytilene for analysis in the Water and Air Quality Laboratory at the Department of Environmental Studies of the University of the Aegean.

2.2. Analytical procedure

All samples were analysed according to the respective procedure described in standard methods (APHA, 1989) and reported in previous works (Golfinopoulos et al., 1998a; Golfinopoulos, 2000). Especially, liquid–liquid extraction and gas chromatography were used to measure the concentration of THMs in the water samples.

The bromide ion concentrations in the raw and finished water were determined with the phenol red colorimetric method (APHA, 1989). The detection limit of bromide is 0.1 mg/l (APHA, 1989). The chlorophyll a (chla) measurements were carried out spectrophotometrically (Strickland and Parsons, 1972) (the values taken from the Department of Water Quality Control in the MTP). Typical data including, water temperature, pH, chlorine dose and free residual chlorine were collected in the plant.

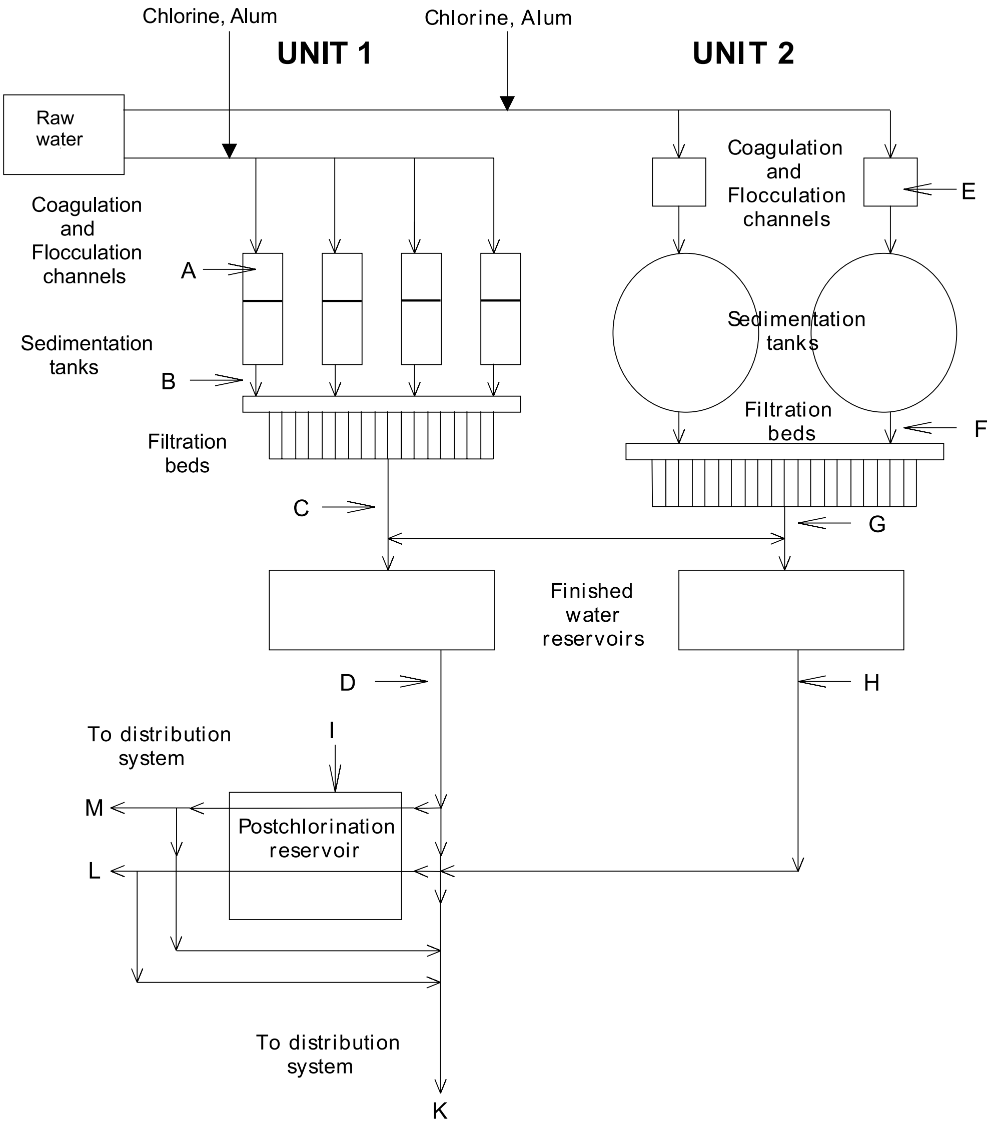

Fig. 1. Menidi Treatment Plant, sampling points given by the capitals A, E: sampling from coagualtion and flocculation channels; B, F: sampling from sedimentation tanks; C, G: sampling from filtration beds; D, H: sampling from finished water reservoir; I: sampling from postchlorination reservoir; K, L, M: sampling from distribution system.

2.3. Data set

The variables examined in the present work were temperature, pH, chlorine dose, bromide and chla. The concentrations of THMs are expressed as μg/l, the temperature as °C, the concentrations of chla as mg/m^3, while chlorine dose and bromide concentrations as mg/l. The development of the models was based on the THMs concentrations from the finished water reservoir, while for pH, temperature, bromide and chla concentrations were used the raw water values. The concentrations of chla in finished water reservoir were not included, since they were approaching zero. The sum of prechlorination and postchlorination was used for the chlorine dose, as the reaction is continued with the addition of chlorine during the latter step, contributing to the further formation of THMs. Moreover, this approach has improved the results, as the poor correlations attempted using solely prechlorination values were replaced by strong correlations based on the combined use of prechlorination and postchlorination values.

Seasonality trends in the formation of THM were also observed, as it is discussed in a following paragraph, and their inclusion in the model was deemed

Table 1
Range of variables, the means and the standard deviations

Variables	Number of observations	Min	Max	Mean	Standard deviation
TTHM	126	6	34	13	5
$CHCl_3$	126	4	18	8	3
$CHBrCl_2$	126	1	12	4	2
Chla	117	0.1	22.2	1.2	3.1
Temperature	126	8	23	15	3
pH	122	7.9	8.4	8.2	0.2
Chlorine dose	126	0.4	3.8	2.5	0.6
Bromide	126	0.1	0.4	0.2	0.2
Summer	126	0	1	0.41	0.47
Spring	126	0	1	0.25	0.45
Winter	126	0	1	0.14	0.35

necessary. So for the four distinct levels (seasons) affecting individually the dependent variables (TTHM, $CHCl_3$, $CHBrCl_2$), three (four minus one) "dummy" variables were defined (Draper and Smith, 1981). The seasons with the "dummy" variables (Winter (W), Spring (Sp), Summer (S)) and the respective summary statistics are given in Table 1; whereas their definition has been described elsewhere (Golfinopoulos et al., 1998b). Furthermore, there was a balance in the flow rate for the different parts of the plant. For that reason the retention time, which is also the chlorine reaction time with the organic substrate, was fairly constant for each step of the treatment plant and was not included as a variable in the model.

2.4. Data analysis

The variables were tested for normality using the Kolmogorov–Smirnov test, which is generally preferable to chi-square when dealing with continuous distributions (Conover, 1980). The relationships between the variables were examined by simple correlation (Vounatsou and Karydis, 1991). Multiple regression analysis was applied as a means for evaluating statistically significant variables of the system (Ott, 1988). The level of significance (α) for the inclusion of a variable in the model was 0.05. The t-criterion was used to eliminate statistically insignificant variables and for the coefficient b_j of the j variable, H_0: $b_j = 0$ and H_a: $b_j \neq 0$, this t statistic can be formed as (Vounatsou and Karydis, 1991)

$$t = \frac{b_j}{S_{b_j}} \qquad (1)$$

where S_{b_j} is the standard deviation of the respective coefficient b_j; if n is the total number of observations, k the number of independent variables used to describe the dependent variable in the model, then the criterion for the rejection of H_0 in favour of H_a is $|t| > t_{n-k-1,\alpha/2}$, indicating that the specific variable is statistically different from zero and should be incorporated in the regression model. Prior to the regression analysis, the database was subdivided into two parts containing 75% and 25% of the data, respectively. This split up was obtained in a random fashion, based on the ranking of the THM concentrations from highest to lowest and assigning to each case its rank value. The subsamples were drawn by choosing a set of these numbers at a random way, through the use of "random digits" tables (Zar, 1984). The bigger sample was the basis for the development of the mathematical equations, while the smaller one was used for the validation of these models.

3. Results and discussion

The ranges of variables, the means and the standard deviations, are shown in Table 1. These results concern both the old and the new unit, since no significant differences were detected between the two parts of the plant. Moreover, they were derived from the database that we used for both the development and the validation of the models, and not from the entire sampling network, illustrated in Fig. 1. The present study focuses on the characteristics of the data that were used for the regression analysis, which is the main issue of the present paper; whereas all the rest information (trends, levels and potential measures) is given elsewhere (Golfinopoulos, 2000). The raw water temperature in the plant ranged from 8 to 23 °C, while the pH varied from 7.9 to 8.4. The dose of chlorine (the sum of pre- and postchlorination) was lying in the range of 0.4–3.8 mg/l. The maximum values of chla concentrations in the raw water occurred on March 1997 (22.2 mg/m^3), whereas the plant was characterised occasionally by chla values beyond the detection limits (i.e. October and December 1995). Meanwhile, the phytoplanktonic biomass in the finished water was almost eliminated. The bromide concentrations in raw water were varying between undetectable levels to 0.4 mg/l. The average values for each of the dummy variables are the proportions of the observations that correspond to the individual seasons. Hence, the sum of those averages gives the overall proportion of the three seasons (winter, spring and summer) that is 0.8 (80%) and the remaining 20% corresponds to the fourth season (autumn). From all the above we have to note the wide range of (i) chla values (0.0–22.2 mg/m^3) with mean 1.18 mg/m^3, (ii) temperature (8–23 °C) with mean 15 °C, (iii) the chlorine dose (0.4–3.8 mg/l) with mean 2.5 mg/l, while the ranges of pH and bromide ions were rather narrow.

The results of the analyses show that THMs were detected in all the water samples. An increasing trend has been noticed during summer and autumn seasons, due to the increased dose of chlorine used for water disinfection and the higher water temperature. The TTHMs values ranged from 6 to 34 μg/l; the respective values for chloroform and dichlorobromomethane were 4–18 μg/l and 1–12 μg/l. The chlorinated products chlorodibromomethane and bromoform were excluded from the overall procedure, because the respective values of almost all the samples were lain beyond the detection limits. These results were attributed mostly to the above mentioned low level of bromide ions in raw water.

Generally, the TTHM values measured in MTP were almost three times lower than those observed in GTP, associated with the different quality of raw water. According to the data from the Department of Water Quality Control of Athens Water and Sewage Corporation the concentrations of organic compounds in Iliki lake (the main supplier of GTP) vary in high levels, while in Mornos lake (the main supplier of MTP) the organic content of water is lower. The main cause for this difference is the natural character of Iliki in contrast with Mornos, which is an artificial and well-protected lake from any source of contamination (Golfinopoulos, 2000). Another reason of the high THM concentrations in GTP is the high bromide level in the water of boreholes (0.3–1.3 mg/l), due to seawater intrusion to the groundwater and the possible contamination of the lake surface water from various anthropogenous and natural sources (Golfinopoulos et al., 1996).

The estimation of goodness-of-fit to the normal distribution showed that all the variables approach normality and therefore can be analysed with their raw forms by parametric statistical methods. The correlation matrix between all the variables is given in Table 2. It can be observed that both TTHM and individual THM species ($CHCl_3$, $CHBrCl_2$) concentrations exposed significant positive correlations with the temperature, the chlorine dose and the pH values indicating that these factors favour THM formation in water treatment plants. These results are in good accordance with what is frequently reported in the literature (Engerholm and Amy, 1983; Golfinopoulos et al., 1996). Another interesting point was the significant correlation between the temperature and the chlorine dose ($r = 0.53$), which is attributed mostly to the implementation of higher doses during the warm months in order to compensate losses due to dechlorination from solar rays (Veenstra and Schnoor, 1980). Also, the fact that the correlation between the temperature and chlorine dose was high, statistically means that an interaction may exist between the two independent variables. Hence, the relationship between the mean values of the dependent variables (TTHM, $CHCl_3$ or $CHBrCl_2$) and one of the independent variables (temperature or chlorine dose) is dependent upon the values of the other independent variable (Bowerman and O'Connell, 1990).

The role of seasonality, concerning the meteorological conditions (i.e. light intensity, photoperiod) aside from the temperature effects, was detected using the analysis of covariance; where the seasonal grouping was the categorical variable and the temperature was the covariate. The sources of differences among the seasons were sought by the least square difference test. The results of this statistical analysis are represented in Table 3. It can be seen that though most of the variability is attributed to the temperature (higher F values), THM concentrations also expose significant seasonal trends (influenced by the rest meteorological factors) characterised by summer highs and winter lows, while values during spring and autumn were lying at intermediate levels. Thus, it was deemed necessary to introduce seasonality effects in the model, using three nominal scale variables, the so-called "dummy" variables.

The least square regression coefficients, the standard errors, the *t*-values and the level of significance for rejecting null hypothesis for each variable included in the model are given in Tables 4–6, respectively. From these relationships, it is inferred that the regression analysis has led to the formulation of the following equations:

Table 2
Simple correlation between TTHM, $CHCl_3$, $CHBrCl_2$, Chla, temperature (T), pH, chlorine dose (Cl) and bromide (Br)

	TTHM	$CHCl_3$	$CHBrCl_2$	Chla	T	pH	Cl	Br
TTHM	1.00							
$CHCl_3$	0.96[a]	1.00						
$CHBrCl_2$	0.91[a]	0.79[a]	1.00					
Chla	−0.13	−0.13	−0.05	1.00				
T	0.34[a]	0.26[a]	0.41[a]	−0.25[b]	1.00			
pH	0.32[a]	0.32[a]	0.22[b]	0.04	0.33[a]	1.00		
Cl	0.33[a]	0.31[a]	0.30[a]	−0.36[a]	0.53[a]	0.19	1.00	
Br	0.07	−0.03	0.18	0.07	0.11	−0.16	0.05	1.00

[a] Significant value at 0.01.

[b] Significant value at 0.05.

Table 3
Analysis of covariance (ANCOVA) and least square difference test for detecting differences among the seasons of the year for TTHM, $CHCl_3$ and $CHBrCl_2$ concentrations[a]

Sources of variance	Degrees of freedom[b]	F ratio	Level of significance
TTHM			
Seasons	3/92	4.73[c]	0.01
Temperature	1/92	747.49[c]	0.00
Groups			
A	Winter–Autumn–Spring		
B	Spring–Summer		
$CHCl_3$			
Seasons	3/92	6.06[c]	0.00
Temperature	1/92	778.22[c]	0.00
Groups			
A	Winter–Autumn		
B	Autumn–Spring		
C	Spring–Summer		
$CHBrCl_2$			
Seasons	3/92	4.15[c]	0.02
Temperature	1/92	681.94[c]	0.00
Groups			
A	Winter–Autumn–Spring		
B	Summer		

[a] The overall adjusted r^2 values for these ANCOVAs were 0.89, 0.89 and 0.87, respectively.
[b] DF of groups/DF of error.
[c] Significant value at 0.05.

Table 4
Regression coefficients, standard errors, *t*-values and level of significance for the multiple regression model of TTHM concentrations

Variables	Regression coefficients	Standard errors	*t*-values	Level of significance
Chla	−0.26	0.15	−2.18	0.03
pH	1.57	0.25	6.30	0.00
[Br]	28.74	15.39	2.87	0.01
$[Br]^2$	−66.72	32.69	−2.04	0.04
S	−43.63	7.24	−6.03	0.00
Sp	1.13	1.18	2.96	0.01
T × S	2.62	0.41	6.34	0.00
T × Cl	−0.72	0.41	−2.76	0.02

$R^2 = 0.52$; $F = 152.51 > F_{8,90,0.05} \approx 2.34$.

$$\begin{aligned}\mathrm{TTHM} = {} & -0.26\,\mathrm{chla} + 1.57\,\mathrm{pH} + 28.74\,[\mathrm{Br}] \\ & - 66.72\,[\mathrm{Br}]^2 - 43.63\,\mathrm{S} + 1.13\,\mathrm{Sp} \\ & + 2.62\,\mathrm{T} \times \mathrm{S} - 0.72\,\mathrm{T} \times \mathrm{Cl} \end{aligned} \tag{2}$$

Table 5
Regression coefficients, standard errors, *t*-values and level of significance for the multiple regression model of $CHCl_3$ concentrations

Variables	Regression coefficients	Standard errors	*t*-values	Level of significance
Chla	−0.32	0.09	−2.07	0.04
pH	0.68	0.17	3.89	0.00
[Cl]	2.51	1.14	2.18	0.03
Sp	1.93	0.74	2.59	0.01
S	−22.07	4.89	−4.50	0.00
T × Cl	−0.12	0.06	−2.11	0.04
T × S	1.38	0.29	4.69	0.00

$R^2 = 0.51$; $F = 169.11 > F_{7,94,0.05} \approx 2.43$.

Table 6
Regression coefficients, standard errors, *t*-values and level of significance for the multiple regression model of $CHBrCl_2$ concentrations

Variables	Regression coefficients	Standard errors	*t*-values	Level of significance
Chla	−0.37	0.05	−2.192	0.02
pH	0.32	0.53	−3.53	0.00
[Br]	16.16	5.49	2.94	0.00
$[Br]^2$	−29.82	11.42	−2.6	0.01
[Cl]	−1.88	0.53	−3.53	0.00
S	5.17	1.57	3.27	0.00
T × Sp	−0.37	0.11	−3.25	0.00
T × Cl	0.12	0.02	5.85	0.00

$R^2 = 0.62$; $F = 119.64 > F_{8,93,0.05} \approx 2.34$.

$$\begin{aligned}\mathrm{CHCl_3} = {} & -0.32\,\mathrm{chla} + 0.68\,\mathrm{pH} + 2.51\,[\mathrm{Cl}] \\ & + 1.93\,\mathrm{Sp} - 22.07\,\mathrm{S} + 1.38\,\mathrm{T} \times \mathrm{S} \\ & - 0.12\,\mathrm{T} \times \mathrm{Cl} \end{aligned} \tag{3}$$

$$\begin{aligned}\mathrm{CHBrCl_2} = {} & -0.37\,\mathrm{chla} + 0.32\,\mathrm{pH} + 16.16\,[\mathrm{Br}] \\ & - 29.82\,[\mathrm{Br}]^2 + 1.88\,[\mathrm{Cl}] + 5.17\,\mathrm{S} \\ & - 0.37\,\mathrm{T} \times \mathrm{Sp} - 0.12\,\mathrm{T} \times \mathrm{Cl} \end{aligned} \tag{4}$$

All the multiple regression models are second-order, including first and second-degree terms in the independent variables. Generally, any independent variable x_i that appears in the regression model as x_i^p is called a *p*th-degree term or similarly, if two independent variables x_i and x_j appear together as $x_i^p x_j^q$, the term is called a $(p+q)$th-degree term (Ott, 1988). The three models are also statistically significant; since the total F criterion—based on the sum of squares due to regression over the sum of squares for error divided by the respective degrees of freedom—has provided evidence for rejecting H_0, whereas the separate *t*-values indicate that all the variables are significant at the probability level of 5%. Moreover, all the adjusted coefficients of determination ($R^2 > 0.5$) were satisfactory. We can also assume

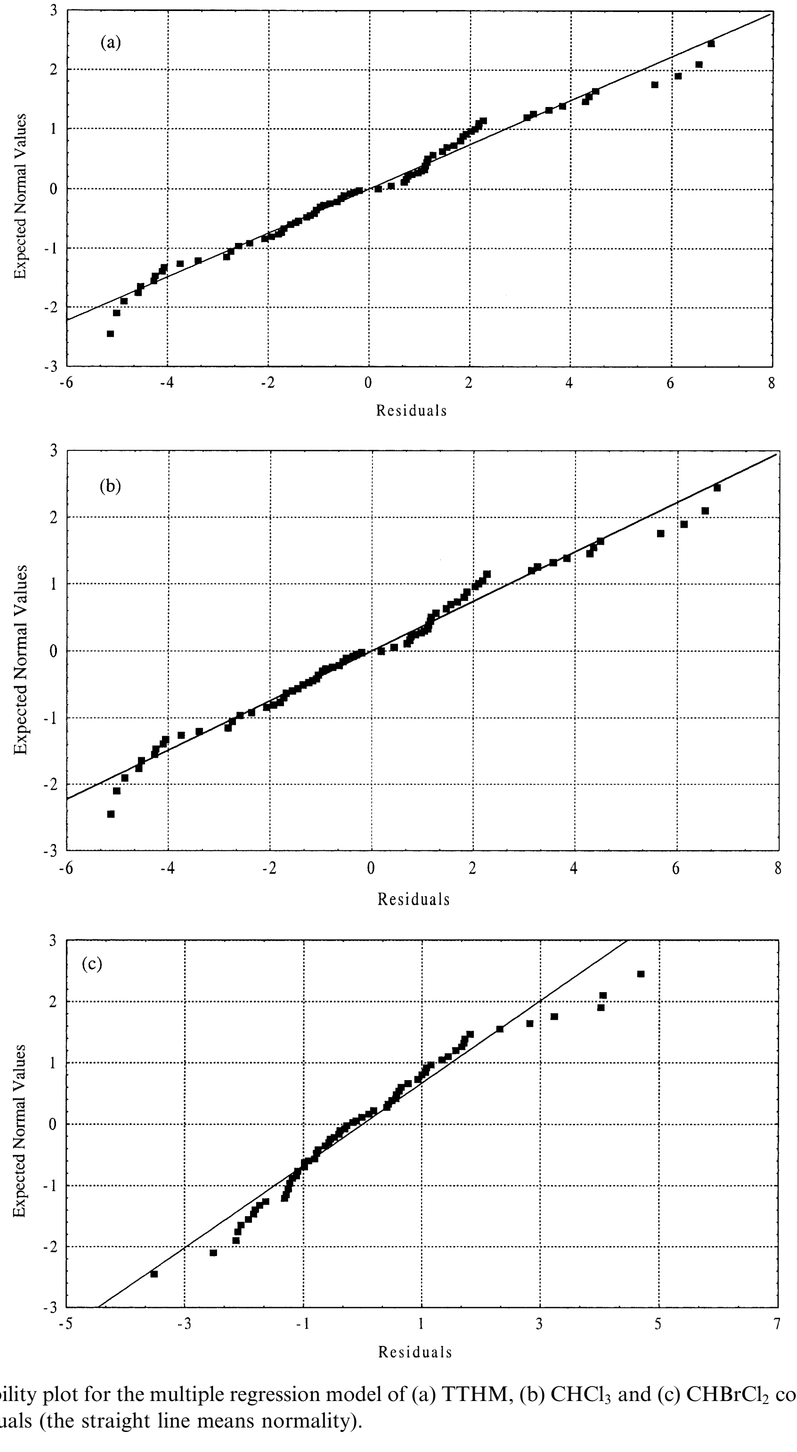

Fig. 2. Normal probability plot for the multiple regression model of (a) TTHM, (b) $CHCl_3$ and (c) $CHBrCl_2$ concentrations. Expected normal values vs residuals (the straight line means normality).

from the normal probability plot of Fig. 2 that the residuals are approaching normal distribution, as all the points are near the straight line which is the identical situation of normality (Draper and Smith, 1981). Finally, the residuals were examined for homoscedasticity, as illustrated in Fig. 3, where the residuals ($Y_{\mathrm{obs}}-Y_{\mathrm{pred}}$) are plotted as a function of their corresponding Y_{pred}'s. If homoscedasticity exists, then the residuals

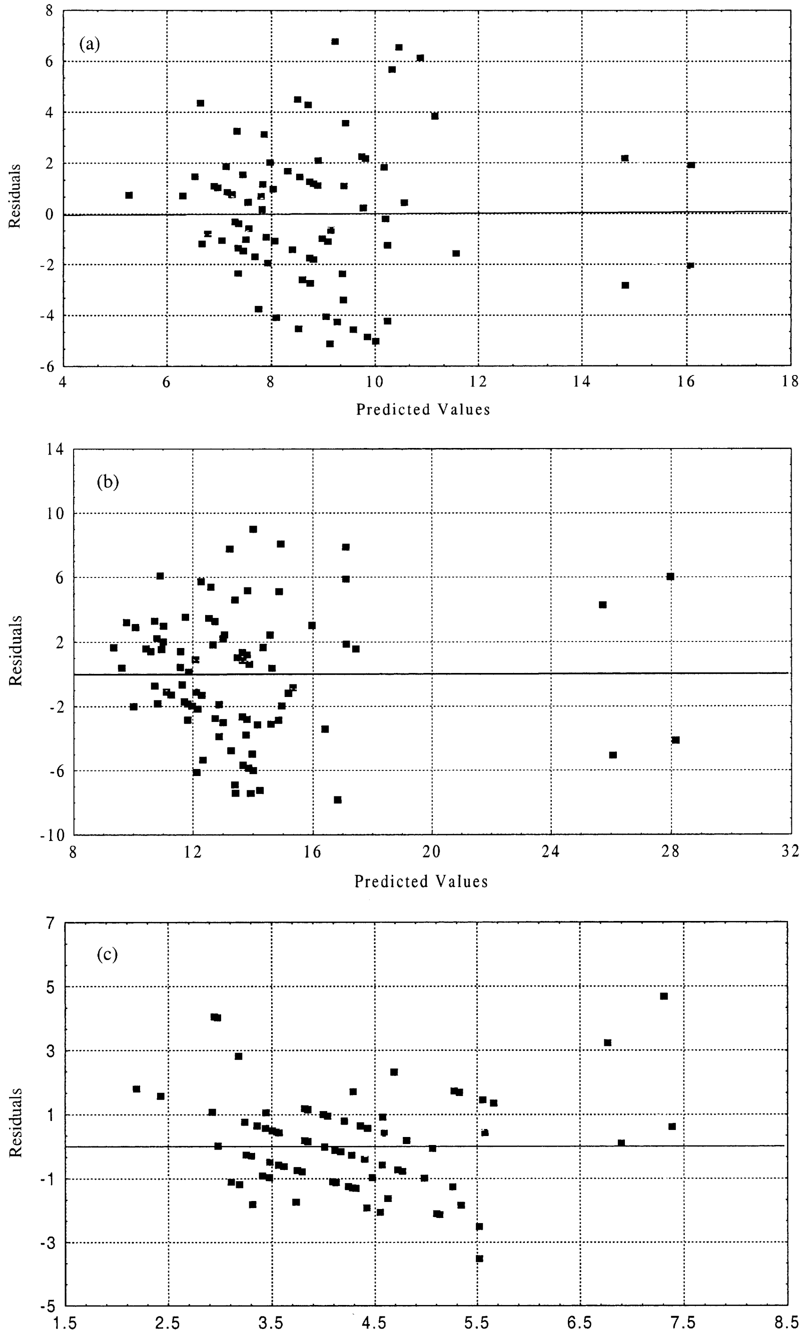

Fig. 3. Residuals vs predicted values of (a) TTHM, (b) $CHCl_3$ and (c) $CHBrCl_2$ concentrations.

should be attributed evenly above and below zero, otherwise we should suspect a calculation error or that an additional variable should be added to the regression model (Zar, 1984). In these particular cases, it can be seen that we have nearly constant variance and therefore the models are deemed valid to describe the experimental data.

The mathematical formulation developed for TTHMs of MTP is almost identical to the one constructed for GTP, though these two treatment plants are characterised by essential differences in terms of the quality of the raw water and the THMs levels (Golfinopoulos et al., 1998b). Both the equations include the variables chla (or ln chla), pH, [Br], $[Br]^2$, S, Sp, T × S, T × [Cl], and moreover the relative sizes of the respective regression coefficients are similar. This implies that the specific equation—after an appropriate calibration of the coefficients' values to account for differences in raw water characteristics—can be used for providing annual variations of TTHM concentrations in water treatment plants characterised by climatic conditions similar to those of the Greek area.

The statistical significance of the product T × [Cl] (great values of *t*-test) indicates that an interaction exists between the two independent variables influencing both total and individual THM species concentrations. This interaction was further discussed on the former modelling development and it was inferred that the fluctuations of [Cl] seem to affect the relationship (shape and slope of the respective lines) of TTHM and temperature (Golfinopoulos et al., 1998b). Meanwhile, the mathematical formulations of $CHCl_3$ and $CHBrCl_2$ values are characterised by the co-existence of the term [Cl]. This implies that [Cl] has a significant role via these equations, since both the direct effects of the chlorine dose and its interactions with the temperature are used for assessing $CHCl_3$ and $CHBrCl_2$ concentrations.

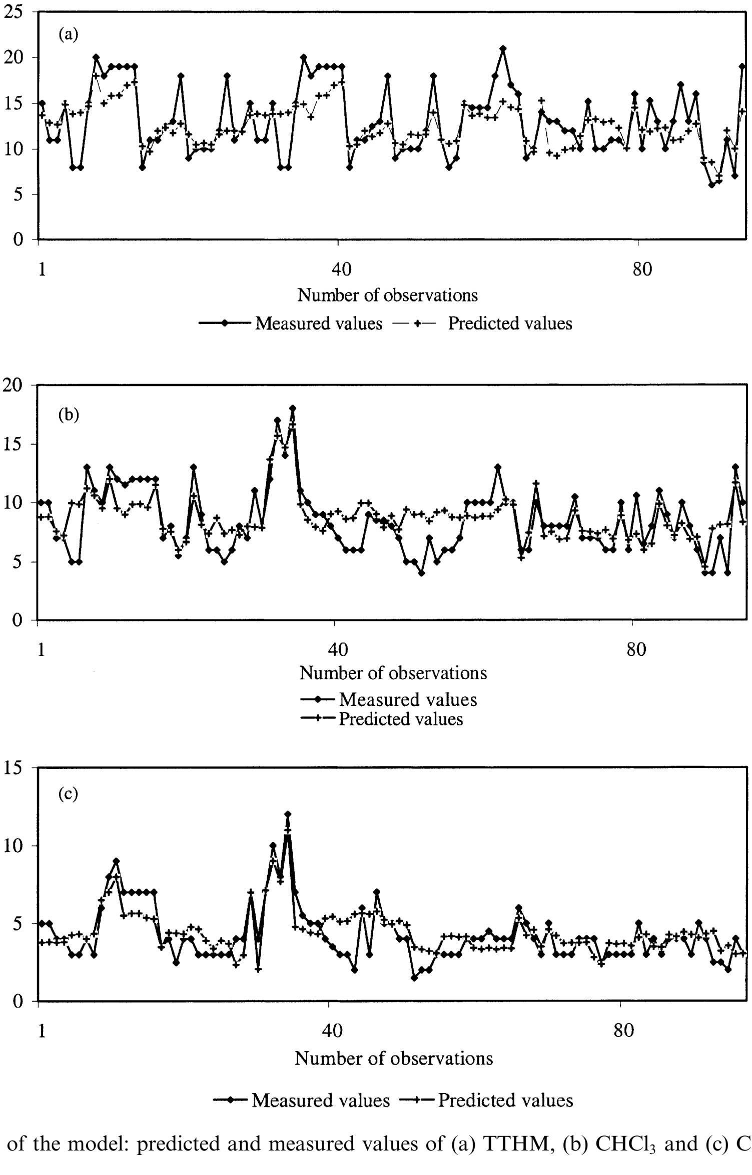

Fig. 4. Goodness of fit of the model: predicted and measured values of (a) TTHM, (b) $CHCl_3$ and (c) $CHBrCl_2$ concentrations.

The models for TTHM and $CHBrCl_2$ show that both the terms [Br] and $[Br]^2$ are statistically significant, meaning that there is a quadratic relationship between the respective concentrations and the bromides. The model without the term $[Br]^2$ resulted in a plot of residuals vs [Br], where the residuals for very low and very high values of [Br] were negative while the residuals for intermediate values were positive. The form of this plot indicated that the inclusion of a quadratic term ($[Br]^2$) in the model was necessary (Draper and Smith, 1981).

The comparison between experimental and predicted values of total and individual THM species concentra-

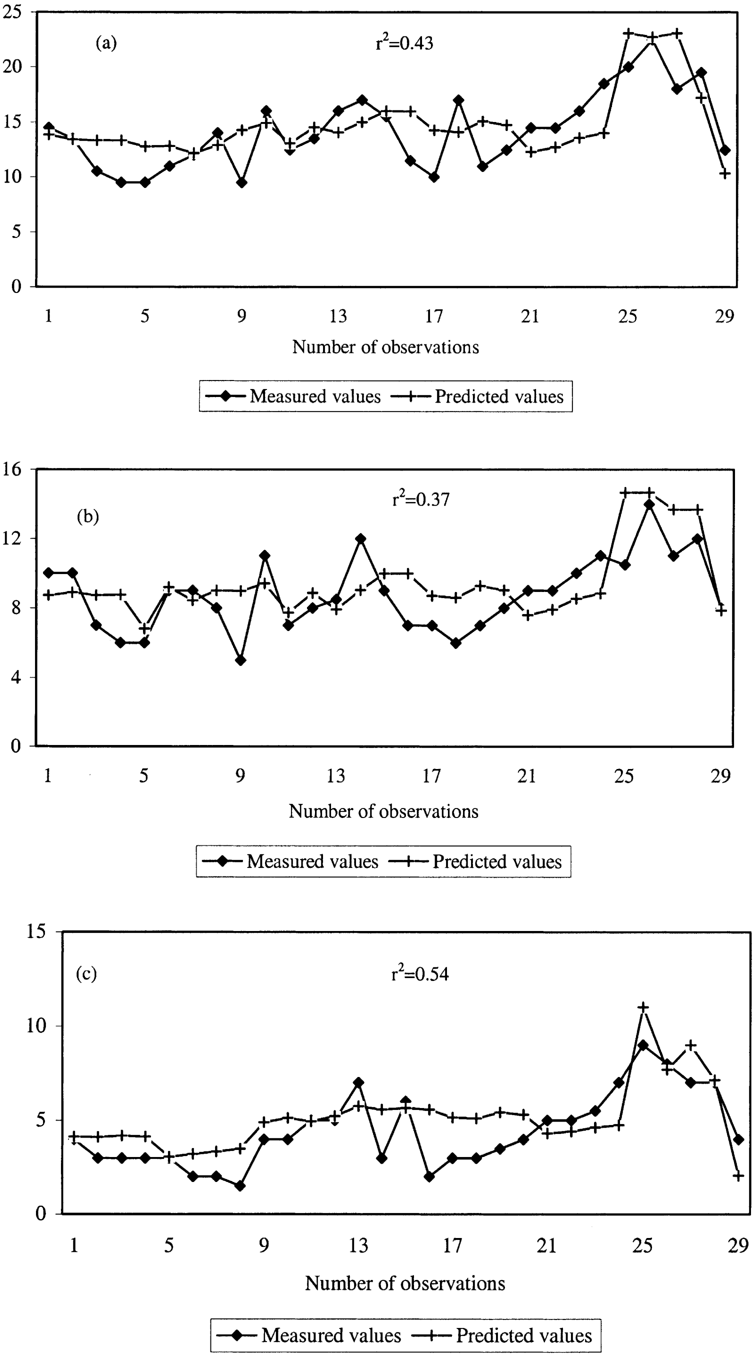

Fig. 5. Validation of the model: predicted and measured values of (a) TTHM, (b) $CHCl_3$ and (c) $CHBrCl_2$ concentrations.

tions is represented graphically in Fig. 4. It is observed that 80% and 47% of the predicted values fell within ±20% and ±10% of the measured TTHM values; these results are almost similar to those obtained from the former model (Golfinopoulos et al., 1998b). Meanwhile, 14% of the rest computed values under- or overestimate TTHM concentrations presenting relative high percentage difference (>35%). This trend concerned mostly the values that were relative high or low in total scale, indicating that the model has a lesser sensitivity to the extreme cases of TTHM formation occurring temporarily during the annual cycle. Similarly, the equations formulated for estimating $CHCl_3$ and $CHBrCl_2$ concentrations expose 86% and 79% agreement to within ±20% of the actual values, whereas the proportion of values lying in the zone of ±10% was 53 and 42%, respectively. With regard to the rest values, all are found near the limit of ±20% difference following more strictly annual lows and highs of sampled data. Thus, these multiple regression models provide accurate estimations of experimental $CHCl_3$ and $CHBrCl_2$ concentrations and present a better goodness-of-fit comparing with the TTHM model. Finally, the validation of these regression models showed a remarkable consistency with the previous described results (Fig. 5). It can be seen that 72, 79 and 76% of the predicted values fell within ±20% of the measured TTHM, $CHCl_3$ and $CHBrCl_2$ values, whereas only three cases for the $CHCl_3$ model and two cases for the $CHBrCl_2$ one presented high percentage difference, beyond the level of 35%.

4. Conclusions

The applicability of multiple regression analysis was tested for estimating TTHM and individual THM species concentrations in the finished water reservoirs. The mathematical formulation developed for TTHMs was almost identical to a former construction, based on different raw water characteristics. Therefore, it was inferred that the specific equation form could be used for providing annual variations of TTHM concentrations in water treatment plants characterised by climatic conditions similar to those of the Greek area. Meanwhile, the equations developed for estimating individual THM species concentrations expose 86% and 79% agreement to within ±20% of the actual values, following more strictly annual lows and highs and thus constituting more sensitive estimators of experimental $CHCl_3$ and $CHBrCl_2$ concentrations.

However, it must be noted that the complexity of TTHM formation reactions makes it difficult to develop universally applicable models. The present models are restricted to field data lying in the ranges of the specific data sets, whereas for more extensive uses it is essential to invoke recalibration processes and introduce optimisation techniques for modifying coefficients' values or even various terms of the model. As an alternative approach, the simulation of reaction kinetics based on hypothesized stoichiometric expressions is proposed as a useful and robust numerical tool that can handle the enormous variability of physical conditions encountered in the temporal or the spatial scale.

Moreover, it is important to recognize that this specific article has considered chla as the only index of the organic matter of water. Other researchers have observed that models based on different organic precursors result in varying yields and rates of THM formation (Elshorbagy et al., 2000). Therefore, given the demand for the development of methodologies with generalized applicability, future-modelling constructions should focus on the use of different precursor-types and investigate their relationships with raw water characteristics. Towards this orientation, the TOC seems to be a reliable surrogate parameter for the organic precursor term, since it has been proved to resulting in the best fit among experimental and simulated data. The specific issue in combination with the potential extrapolation of regression analysis or the simulation of reaction kinetics as methodological tools for predicting THM levels throughout a water distribution system—and not only in the finished reservoir of the water treatment plants—are aims for on-going research.

References

APHA, 1989. Standard Methods for the Examination of Water and Wastewater, seventeenth ed. American Public Health Association, Washington, DC.

APHA, 1992. Standard Methods for the Examination of Water and Wastewater, eighteenth ed. American Public Health Association, Washington, DC.

Bowerman, B.L., O'Connell, R.T., 1990. Linear Statistical Models: An Applied Approach, second ed. PWS-KENT Publishing Company, Boston.

Conover, W.J., 1980. Practical Nonparametric Statistics, second ed. John Wiley, New York.

Craun, G.F., 1991. Epidemiological studies of organic micropollutants in drinking water. In: Hutzinger, O. (Ed.), Handbook of Environmental Chemistry. Water Pollution, Part A, vol. 5. Springer, Berlin, pp. 1–44.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. John Wiley, New York.

EECD (European Economic Community Directive), 1997. Amended proposal for a Council Directive concerning the quality of water intended for human consumption—common position. In: Proceedings of the Council of the European Union, Directive 80/778/EEC, Com (97) 228 final 95/0010 SYN, Brussels.

Elshorbagy, W.E., Abu-Qdais, H., Elsheamy, K., 2000. Simulation of THM species in water distribution systems. Water Research 34 (13), 3431–3439.

Engerholm, B.A., Amy, G.L., 1983. A predictive model for chloroform formation from humic acid. Journal of American Water Works Association 75 (10), 418–423.

Fayad, N.M., 1993. Seasonal variations of THMs in Saudi Arabian drinking water. Journal of American Water Works Association 85 (1), 46–50.

Golfinopoulos, S.K., 2000. The occurrence of trihalomethane in the drinking water in Greece. Chemosphere 41 (11), 1761–1767.

Golfinopoulos, S.K., Kostopoulou, M.N., Lekkas, T.D., 1996. THM formation in the high bromide water supply of Athens. Journal of Environmental Science and Health A 31 (1), 67–81.

Golfinopoulos, S.K., Kostopoulou, M.N., Lekkas, T.D., 1998a. Volatile halogenated organics in the water supply system in Athens, Greece. Water Research 32 (6), 1811–1818.

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

Harrington, G.W., Chowdhury, Z.K., Owen, D.M., 1992. Developing a computer model to simulate DBP formation during water treatment. Journal of American Water Works Association 84 (11), 78–87.

Jimenez, M.C.S., Dominguez, A.P., Silverio, J.M.C., 1993. Reaction kinetics of humic acid with sodium hypochlorite. Water Research 27 (5), 815–820.

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, 41–48.

Nieuwenhuijsen, M.J., Toledano, M.B., Eaton, N.E., Elliott, P., 2000. Chlorination disinfection byproducts in water and their association with adverse reproductive outcomes: a review. Occupational and Environmental Medicine 57, 73–85.

Ott, L., 1988. An Introduction to Statistical Methods and Data Analysis, third ed. PWS-Kent Publising Company, Boston.

Rook, J.J., 1974. Formation of haloforms during chlorination of natural waters. Water Treatment and Examination 23, 234–243.

Simpson, K.L., Hayes, K.P., 1998. Drinking water disinfection by-products: an Australian perspective. Water Research 32 (5), 1522–1528.

Strickland, J.D.H., Parsons, T.R., 1972. A Practical Handbook of Seawater Analysis, second ed. Fish Res. Board of Canada, Buletin, p. 167.

Veenstra, J.N., Schnoor, J.L., 1980. Seasonal variation in trihalomethanes levels in an Iowa river water supply. Journal of American Water Works Association 72, 583–590.

Vounatsou, P., Karydis, M., 1991. Environmental characteristics in oligotrophic waters: data evaluation and statistical limitations in water quality studies. Environmental Monitoring and Assessment 18, 211–220.

Zar, J.H., 1984. Biostatistical Analysis, second ed. Prentice-Hall, Englewood Cliffs, New Jersey.