

Article

Developing THMs' Predictive Models in Two Water Supply Systems in Greece

Stavroula Tsitsifli * and Vasilis Kanakoudis

Civil Engineering Department, University of Thessaly, GR 38334 Volos, Greece; bkanakoud@civ.uth.gr

* Correspondence: tsitsifli@uth.gr; Tel.: +30-24210-74178

Received: 22 April 2020; Accepted: 14 May 2020; Published: 16 May 2020



Abstract: Disinfection is a very significant water treatment process for drinking water safety, as it inactivates pathogens from drinking water. However, disinfection-by-products (DBPs) are formed which are accused of contributing to cancer and reproductive/developmental effects. Research has provided many predictive models for the formation of DBPs based on various water quality parameters and following different methodologies. The present study aims at developing predictive models for the formation of DBPs in two drinking water supply systems in Greece. Data from the water supply systems are used. A statistical analysis took place to identify the predictive models for the formation of Total trihalomethanes (TTHMs). The results showed that some of the developed models are more reliable than others. However, further study is necessary in order to obtain more data on variables that are affecting trihalomethanes (THMs) formation. Such models can be used mainly locally.

Keywords: water supply system; drinking water quality; disinfection; disinfection-by-products

1. Introduction

Water is essential for life and therefore its safety is of paramount importance. Drinking water distribution networks (WDNs) supplying water to the consumers are extremely complex systems, comprised of hundreds of kilometers of pipes, storage tanks, pumps, valves, and other important assets for the operation of the WDN. It is well known that water quality deteriorates as water travels within the WDN, from the water intake to the tanks, the distribution systems and finally the consumers' taps. Physical, chemical, and biological risks are identified in WDNs. Several outbreaks connected to water quality have been recorded all over the world due to biological and chemical factors. Microbial load exists in natural water coming from surface or groundwater sources, but it can also be present under normal operating conditions, natural disasters, and malicious threats [1]. Natural disasters such as extreme weather phenomena (floods) or earthquakes might cause several kinds of damage to WDNs and may result in the entrance of micro-organisms at several parts of the network. Malicious threats might include terrorism attacks using biological or chemical compounds that may cause adverse effects to the health of water consumers. Even under normal operating conditions it is possible that contaminants might enter the network, after a scheduled or non-scheduled maintenance or even during normal operations (e.g., due to leaking pipes allowing substances from the ground to enter the network) [1]. As sometimes sewerage networks are located above water supply pipes, possible leakages in both systems might be the cause of water contamination. Trying to manage and reduce water losses in WDNs, pressure management among other measures is used. In these cases, as water pressure is reduced within the network, the water age (being the time water remains within the network) increases, especially at dead-ends of the network. Increased water age means that water quality deteriorates [1].

An important measure taken to reduce the risk of contaminated drinking water by micro-organisms is disinfection. Disinfection is a crucial water treatment method as it ensures that water is free of

pathogenic micro-organisms causing water borne diseases. It is worth noting that in the U.S. cholera incidence was reduced by 90%, typhoid by 80% and amoebic dysentery by 50% after introducing disinfection in water treatment [2,3]. Although disinfection is an extremely important water treatment process, disinfection-by-products (DBPs) are formed, being accused of various effects on health. The present paper aims to: (1) investigate the formation of DBPs in two water supply systems in Greece; (2) develop predictive models for the concentration of THMs in these systems; and (3) statistically evaluate the developed models and present limiting factors.

2. Disinfection Methods and Effects

It is known that disinfection is affected by many parameters such as water temperature, water pH, type of existing bacteria, type of disinfection, disinfectant dose, contact time and inorganic and organic material existing in water. DBPs formed are categorized as trihalomethanes (THMs) and haloacetic acids (HAAs).

Methods used for disinfection include chlorination, chloramination, use of chlorine dioxide, ozone, use of other chemical disinfectants (such as copper silver ionization and hydrogen peroxide) and non-chemical disinfection technologies such as UV radiation, etc. [4,5]. All disinfection methods have advantages and disadvantages. Some disinfectants are stable, and some are more effective than others (e.g., chlorine dioxide compared to “chlorine”). However, an important advantage is the existence of residual chlorine after chlorination. Some disinfectants (such as chlorine dioxide) do not form THMs when reacting with humic substances. However, the disinfectants’ cost, their generation on demand, taste and odor issues, their use in sites limited in space and the formation of THMs are some of their disadvantages [4]. Chlorination is the disinfection method most widely used, where liquefied chlorine gas or sodium hypochlorite solution (sometimes the term “chlorine” is used) is added in water [4]. Chlorination has the advantage of retaining chlorine residual throughout the water supply system.

One of the major concerns about disinfection processes is the formation of by-products that can be dangerous for human health. DBPs are formed due to disinfectant overdose or inappropriate use. Organic and inorganic compounds react with the disinfectant and form by-products, organochlorine and inorganic. Organic compounds include trihalomethane (THM) and haloacetic acids (HAAs). THMs include chloroform, bromodichloromethane, dibromochloromethane and bromoform [2]. The first researches on DBPs appeared in the 1970s, when Rook and others identified chloroform and other THMs in drinking water [2,6,7]. The formation of these DBPs is related to the existence of organic matter in water, water pH and temperature and the type of disinfectant used. Apart from, organic by-products, inorganic by-products are also formed such as chlorate and bromate, related to the type of disinfectant. Retention time is a critical factor for the formation of DBPs. Regarding chlorination, chlorine residual is considered as the most accepted and reliable indicator for real time control of bacteria. Research has not revealed any micro-organism that meets all criteria to become a reliable indicator for disinfection efficacy.

DBPs’ effects in humans have been studied by many researchers [2,8,9]. Studies record effects such as cancer and reproductive or developmental effects and negative impacts to organs such as liver, kidney, and the nervous system [2,8]. Studies also recorded infertility, teratogenicity, organs’ inefficiency, etc. [9]. Several epidemiological studies focus on the harmful effects of chlorine by-products and link their increased concentrations with an increased risk of various forms of cancer growth [9].

The literature review revealed that more than 600 DBPs are reported and only some of them are regulated, while others are considered as emerging, as they have lower occurrence levels and toxicological effects [2,10]. Several DBPs can be formed in drinking water and it is known that they have synergistic effects. Thus, as they are found in drinking water in different concentrations, their effects are difficult to be determined. Means of exposure include not only drinking but also bathing, cleaning, washing (dermal and inhalation exposure), etc.

Several organizations, as the World Health Organization (WHO), United States Environmental Protection Agency (US EPA) and the European Union (EU), have set regulations and guidelines to

control DBPs and minimize consumers' exposure. The maximum acceptable contaminant level set by U.S. EPA for total THMs (that is total concentration of all four THMs) is 0.08 mg/L, while this level set by WHO is 0.2 mg/L for chloroform, 0.1 mg/L for each dibromochloromethane and bromoform and 0.06 mg/L for bromodichloromethane. European Union guidelines set acceptable total THM concentration to 0.1 mg/L. The new updated European Directive for drinking water did not change the maximum acceptable THM concentration but encourage water utilities to pursue lower concentrations without affecting disinfection. In Greece, the maximum acceptable level of total THMs is set by the Joint Ministerial Decision (Y2/2600/2001) to 0.1 mg/L, complying with the EU Drinking Water Directive 98/83/EC.

3. DBP Predictive Models—A Review

The development of statistical models for DBPs is increasingly recognized as a methodological basis to predict the formation of DBPs. Such models are helpful to the water utility managers during decision making, for example, setting disinfectant dose, contact time, adjustment of pH, etc., in order to reduce the formation of DBPs and at the same time maintain the required disinfectant residual [5]. These models can also be used to identify the optimal locations for boosters in order to maintain the required levels of disinfectant residuals and reducing the DBPs' formation. Optimal water sampling points can be identified for water quality control using such models.

The levels of THMs in chlorinated water are associated with chlorine demand, pH, temperature and seasonal variability, chlorine contact time and organic materials or chemical elements [11–15] (Table 1). It is found in the literature that pH and temperature are proportional to THMs formation, but pH effects vary for different DBPs [8]. As micro-organisms increase as temperatures increases, higher disinfectant dose is applied during the summer period, resulting in high DBP concentrations. The conditions affecting the disinfection efficiency and the requirements to maintain disinfectant residuals simultaneously affect DBP formation (vicious cycle). Several studies [16] showed that THM concentrations are higher within the WDN compared to the storage tank. This is also due to the existence of organic matter in the biofilms located in the water pipes' walls. Organic matter in water is another parameter affecting proportionally DBP formation (Table 1).

Table 1. Disinfection-by-products (DBPs) predictive models' parameters (based on [8]).

	Parameter	Units
Br ⁻	Bromide ion	mg/L
Cl ₂	Initial chlorine concentration	
pH	pH	
T	temperature	°C
NVTOC	Non-volatile organic carbon	mg/L
TOC	Total organic carbon	mg/L
D	Chlorine dose	mg/L
t	Reaction time	hrs
UV	UV absorbance at 254 nm	cm ⁻¹
TTHMo	Initial total THM concentration	
Flu	fluorescence	%
C ₀	Residual chlorine at the treatment plant after chlorination	mg/L
α	Parameter depending on location which chloroform is predicted	
ε	Random error	
Ch-a	Chlorophyll-a	mg/m ³
DOC	Dissolved organic carbon	mg/L

The first models for chlorinated DBPs appeared in 1983 [17] predicting the formation of total THMs, while the first models for other DBPs appeared in 1994 [18]. DBP predictive models are data-driven statistical models or process-based ones. The first models are based on empirical relationships between dependent and independent variables [19]. On the other hand, process-based models are mainly

based on the assumed knowledge of the actual processes taking place in the water supply system [19]. Many studies are elaborated using data-driven statistical models as there are cases where parameter estimation within a process-based model is imprecise or difficult to obtain [20] or where the data required for the development of process-based models are not available [19]. Using such models, the researchers do not have to know the laws of chemistry and mathematics for the formation of DBPs in advance [19,21]. However, the factors affecting formation of chlorine decay should be known in advance. As data availability becomes stronger, statistical techniques are used more and more, using as much as possible the data already available.

Many researchers have proposed several empirical models for DBP formation [8,22–25]. Most of these models are functions of many parameters derived from linear and non-linear multi-regression analysis. The parameters taken into consideration in these models include: total organic carbon (TOC), ultraviolet light absorbance at 254 nm (UV254), temperature, chlorine dosage, bromide concentration, reaction time, and chlorination pH [8,22,26–29]. Most of the models developed in various studies are site-specific and cannot be used widely, as conditions vary. Models have been developed based on laboratory studies or real field data or both. Field studies compared to laboratory ones take into consideration the effects of the distribution system on residual disinfectant concentration and DBP formation as they measure or observe human exposure [8]. Another difference is that contact time, which can be easily estimated in laboratory studies, needs tracer studies or hydraulic simulation models in field studies. Prediction models are based on empirical relationships or kinetics involved during chlorination.

A few studies have been elaborated in Greece for the water supply network of Athens [30], for water from water treatment plants in Athens [31,32] and for river waters on the island of Lesbos [29,33].

4. Case Studies—THM Formation Models

This paper presents predictive models developed for WDNs of two cities in Greece. The water utility in city A supplies about 150,000 people, while the water utility in city B supplies about 55,000 people. The data are gathered from samplings that the water utilities elaborated, as this is their obligation according to the institutional framework. Depending on the water volume produced, the water utility has the obligation to elaborate water samplings and analysis of chemical and biological parameters. The water utilities investigate check monitoring and audit monitoring for the water they abstract and supply to the consumers. The sampling frequency depends on the water volume taken from the water source and the number of consumers supplied with water. For both cities, the data provided include both check and audit monitoring parameters (THM concentration is measured in audit monitoring). However, audit monitoring is done less frequently. In city A, samplings are taken from all over the water supply network including consumers' taps, while in city B only from the tanks after water intake. Both cities use groundwater sources to supply their consumers. Chlorination is performed in water tanks in both cases.

4.1. General Data

The data gathered and used in the analysis from the WDN of city A include pH, temperature (T), residual chlorine (mg/L Cl_2), conductivity ($\mu\text{S}/\text{cm}$), turbidity (NTU) and Total THMs (TTHMs in $\mu\text{g}/\text{L}$). The data refer to a period of six years (2013–2018) and samples are taken from various points of the networks including tanks and final consumers. Temperature values range from 13 to 30 °C and TTHMs values range from 0.39 to 22.84 $\mu\text{g}/\text{L}$, lower than the maximum allowable level.

For city B, the data gathered from the WDN include pH, total organic carbon concentration (TOC) (mg/L), conductivity ($\mu\text{S}/\text{cm}$), residual chlorine (mg/L), turbidity (NTU), total THMs ($\mu\text{g}/\text{L}$) and the concentrations of the four THMs (chloroform—CM, bromoform—BM, bromodichloromethane—DCBM and dibromochloromethane—DBCM) in $\mu\text{g}/\text{L}$. Data are gathered from 27 different water boreholes for a period of five years (2014–2018). TOC values range from 0.01 to 39.5 mg/L and TTHMs range from 0.48 to 68.35 $\mu\text{g}/\text{L}$, lower than the threshold. Both cities are supplied with water abstracted from

boreholes and this is why organic substances are not present in high concentrations in water. Table 2 presents the total number of values (N), average (AV), standard deviation (SD), minimum (MIN) and maximum (MAX) values of the parameters studied.

Table 2. Total number, average, standard deviation (SD), minimum and maximum values of the parameters in both water distribution networks (WDNs).

WDN A						WDN B					
Parameter	N	AV	SD	MIN	MAX	Parameter	N	AV	SD	MIN	MAX
pH	41	7.6707	0.1677	7.00	8.00	pH	64	7.6895	0.3931	6.92	8.90
T	41	20.122	3.816	13.0	30.0	Conductivity	64	715.5	179.3	419.0	1141.0
Conductivity	41	485.4	75.7	259.0	686.0	Turbidity	64	0.1983	0.3157	0.01	2.45
Turbidity	41	0.3993	0.138	0.21	0.82	TOC	64	5.227	7.52	0.01	39.5
Residual Chlorine	41	0.17	0.0729	0.05	0.40	Residual Chlorine	63	0.3313	0.1192	0.16	0.80
TTHMs	41	5.866	5.382	0.39	22.84	TTHMs	64	8.07	11.36	0.48	68.35

4.2. Statistical Analysis

The dependent variables (TTHM concentrations) were tested for normality using the Kolmogorov–Smirnov (K–S) test to check the goodness-of-fit to the normal distribution [34]. The results of the K–S tests for the estimation of goodness-of-fit of the dependent variables of the model to the normal distribution showed that all dependent variables followed the normal distribution at significance level 0.05 except for after ln-transformation steps [33] (Table 3). Independent variables were also tested for normality using K-S test. For city A, pH, turbidity and conductivity follow normal distribution at significance level 0.01 (Table 3). For city B, turbidity and TOC do not follow normal distribution at 0.05 significance level. Thus, log-transformation for turbidity took place (Table 3). Regarding TOC another transformation took place using the tool available from Minitab software. It was found that $(\text{TOC})^{-0.1}$ follows normal distribution at significance level 0.05.

Table 3. Kolmogorov-Smirnov (K-S) values for the parameters of both WDNs.

WDN A		WDN B	
Parameter	K-S	Parameter	K-S
pH	0.228	pH	0.077
T (°C)	0.165	Conductivity	0.099
Conductivity	0.212	Turbidity	0.275
Turbidity	0.225	TOC	0.244
Residual Chlorine	0.145	Residual Chlorine	0.120
TTHMs	0.224	TTHMs	0.282
ln (TTHMs)	0.153	Log turbidity	0.181
		Log (TTHMs)	0.127

The relationships between the variables were examined by Pearson correlation matrix. For city A, the results of Pearson correlation matrix (Table 4) show a moderate negative correlation between TTHMs and conductivity ($r = -0.406$) and a moderate correlation between TTHMs and residual chlorine ($r = 0.266$). The results for city B show a moderate positive correlation between TTHMs and turbidity ($r = 0.553$) and a moderate correlation between TTHMs and water pH ($r = 0.465$). Several studies have reported a linear relationship between TTHMs and pH [8]. Other variables show a low correlation to TTHMs (Table 4).

Table 4. Pearson correlation values for the parameters of both WDNs.

Parameter	WDN A					WDN B				
	pH	T	Residual Chlorine	Turbidity	Conductivity	Parameter	TTHMs	pH	Conductivity	TOC
T	-0.158					pH	0.465			
Res.chlorine	0.252	-0.362				Conductivity	-0.316	-0.397		
Turbidity	-0.524	0.105	-0.299			Turbidity	0.553	0.074	-0.173	
Conductivity	-0.136	-0.108	-0.307	0.049		TOC	0.049	0.097	0.145	0.115
TTHMs	-0.009	-0.125	0.266	-0.003	-0.406	Res. chlorine	0.301	0.218	-0.042	0.498

4.3. Multiple Regression Analysis

To perform a linear regression, the TTHM concentration (Y) is assumed to be a linear function of the inputs, X. The unknown parameters to be determined, a_i , are the coefficients, as given in (1):

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (1)$$

where n is the number of inputs used. The coefficients are chosen to minimize the sum of the squared differences between the predicted and actual values of Y. Multiple regression analysis is used to evaluate the statistically significant variables at a level of significance α .

5. Results and Discussion

5.1. Model Development

Based on the data, multiple regression analysis was applied at significance level α , which was 0.1 for WDN A and 0.05 for WDN B. Throughout the process of model development, several linear and non-linear regression analyses were performed. The inclusion of each variable in the proposed model was based on the t-criterion [29,35]. Methodological details about the model development are extensively discussed in past studies [25,29,31,36,37]. Multiple regression analysis tool from Minitab was used.

For the WDN A, all variables are initially used as independent variables (inputs). As the independent variables are not statistically significant ($p > 0.1$), they are excluded one-by-one from the model development process. Finally, only residual chlorine (ResChl) is found to be statistically significant ($p < 0.1$) (Table 5). Finally, the developed model is:

$$\ln\text{TTHMs} = 0.781 + 3.64 \text{ ResChl} \quad (2)$$

Table 5. Statistical analysis data for the 4 models developed for WDN A and WDN B.

Term	Coef	t-Value	p-Value	Model	R ²	Durbin Watson
Constant	0.781	2.23	0.032	$\ln\text{TTHMs} = 0.781 + 3.64 \text{ ResChl}$	8.63%	1.19478
ResChl	6.64	1.92	0.062			
Constant	1.096	4.33	0.000	$\log\text{TTHMs} = 1.096 + 0.602 \log\text{ResChl}$	9.13%	1.23129
logResChl	0.602	1.98	0.055			
Constant	-3.2	-1.79	0.079	$\ln\text{TTHMs} = -3.2 + 1.072 \text{ pH} - 3.658 (\text{TOC})^{-0.1}$	46.90%	2.14728
pH	1.072	4.59	0.000			
(TOC) ^{-0.1}	-3.658	-6.42	0.000			
Constant	-5.04	-3.18	0.002	$\log\text{TTHMs} = -5.04 + 8.17 \log\text{pH} - 1.595 (\text{TOC})^{-0.1}$	46.34%	2.14778
logpH	8.17	4.50	0.000			
(TOC) ^{-0.1}	-1.595	-6.40	0.000			

R² value is 8.63% for this model, which is very low, showing that this model is not very suitable for the prediction of TTHMs. All statistical analysis data are given in Table 5.

As the above model is not very effective at predicting TTHMs, another attempt was made for the model development. In this attempt the independent variable residual chlorine is log-transformed. Again, this is the only independent variable found to be statistically significant ($p < 0.1$) after eliminating the other variables one-by-one from the model (Table 5). The new model is:

$$\log\text{TTHMs} = 1.096 + 0.602 \log\text{ResChl} \quad (3)$$

For this model, all statistical analysis data are given in Table 5. The value of R^2 is 9.13% which is still very low but is slightly higher than the model's (3) R^2 value.

The same methodology is used for the WDN B. Several linear and non-linear regression analyses were performed. All variables are found not to be statistically significant ($p > 0.05$) except for pH and $(\text{TOC})^{-0.1}$. After eliminating the other independent variables one-by-one, pH and $(\text{TOC})^{-0.1}$ are found to be statistically significant ($p < 0.05$). The model developed for WDN B is:

$$\ln\text{TTHMs} = -3.2 + 1.072 \text{pH} - 3.658 (\text{TOC})^{-0.1} \quad (4)$$

The value of R^2 for this model is 46.90% which is moderate and shows that the model can be used (Table 5).

Instead of ln-transformation, log-transformation is used alternatively for TTHMs as normality hypothesis is valid. In this attempt $\log\text{TTHMs}$ is used as dependent variable and $\log\text{pH}$ and $(\text{TOC})^{-0.1}$ as independent variables, as they are found to be statistically significant ($p < 0.05$) (Table 5). The predictive model for TTHMs is the following:

$$\log\text{TTHMs} = -5.04 + 8.17 \log\text{pH} - 1.595 (\text{TOC})^{-0.1} \quad (5)$$

For this model R^2 value is 46.34% which is moderate and slightly lower than that of model (4) (Table 5).

For the models developed, Durbin Watson estimate, provided in Table 5, is used in order to check autocorrelation. The values of the Durbin–Watson statistic were found to be 1.19478 and 1.23129 for models 2 and 3, respectively, and 2.14728 and 2.14778 for models 4 and 5, respectively. The analysis showed that there is no autocorrelation.

ANOVA tests are elaborated. Statistical examination showed that the residuals of the models follow the normal distribution [38], and the mean value of the residuals should be zero. The residuals should be evenly attributed above and below zero, otherwise a calculation error should be suspected or an additional variable should be added to the regression model [34]. Figures 1–4 are provided for all four models. In all cases the analysis showed that the residuals are approaching normal distribution and the models are deemed valid to describe the experimental data.

5.2. Predicted and Observed Values Comparison

The developed models are used for the validation of the results. A comparison between predicted and observed values for all proposed models is presented in Figure 5a,b and Figure 6a,b. For WDN A, it can be observed that the models do not provide a satisfactory estimation of the THM concentrations formed. For the model (2), only 17% of the predicted values range in the $\pm 20\%$ of observed values, while for model (3), the percentage of the predicted values ranging in the $\pm 20\%$ of observed values is 14.6%. The regression coefficients (R^2) range from 0.0525 to 0.0774 (Figure 5a,b), which is not a satisfactory level of explanation of the observed variability.

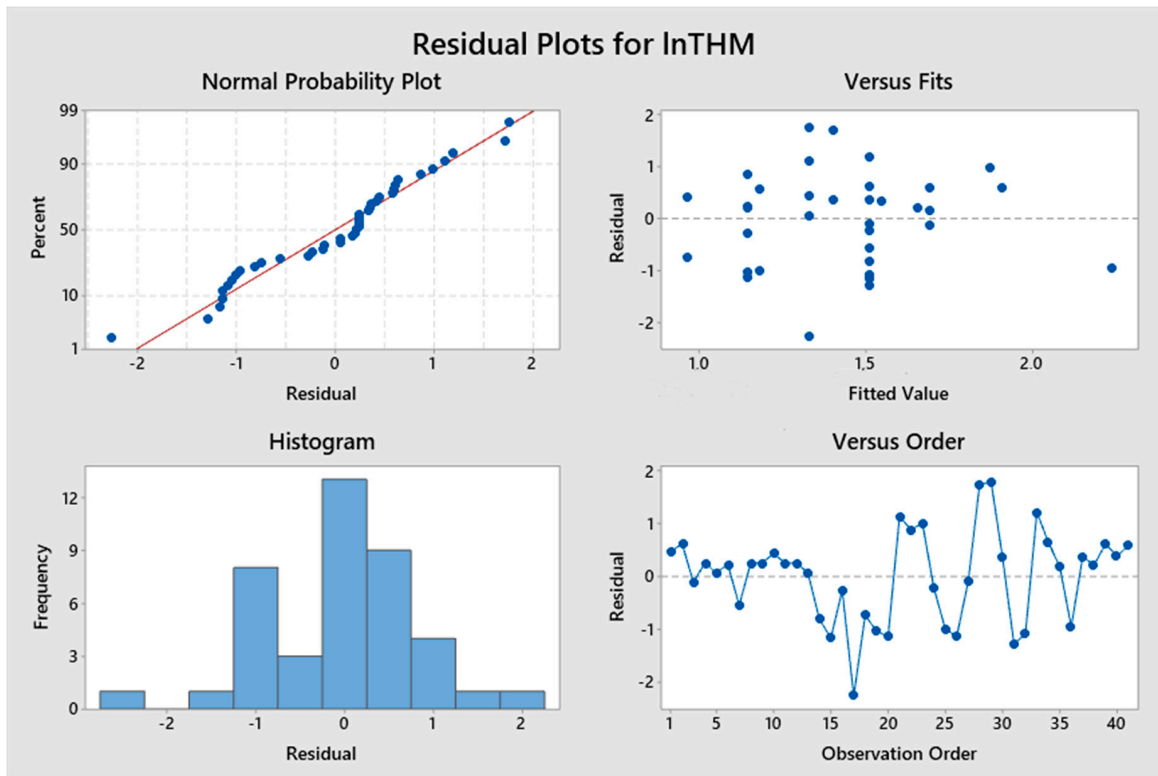


Figure 1. Residual plots for model (2).

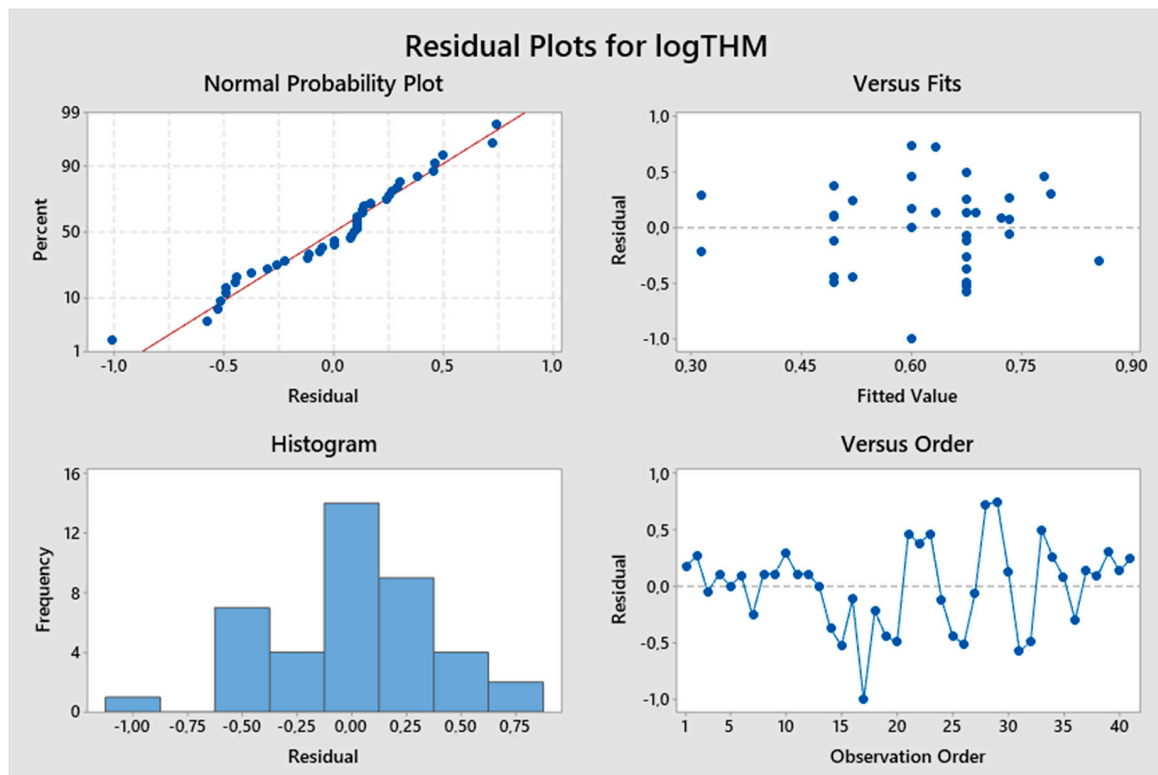


Figure 2. Residual plots for model (3).

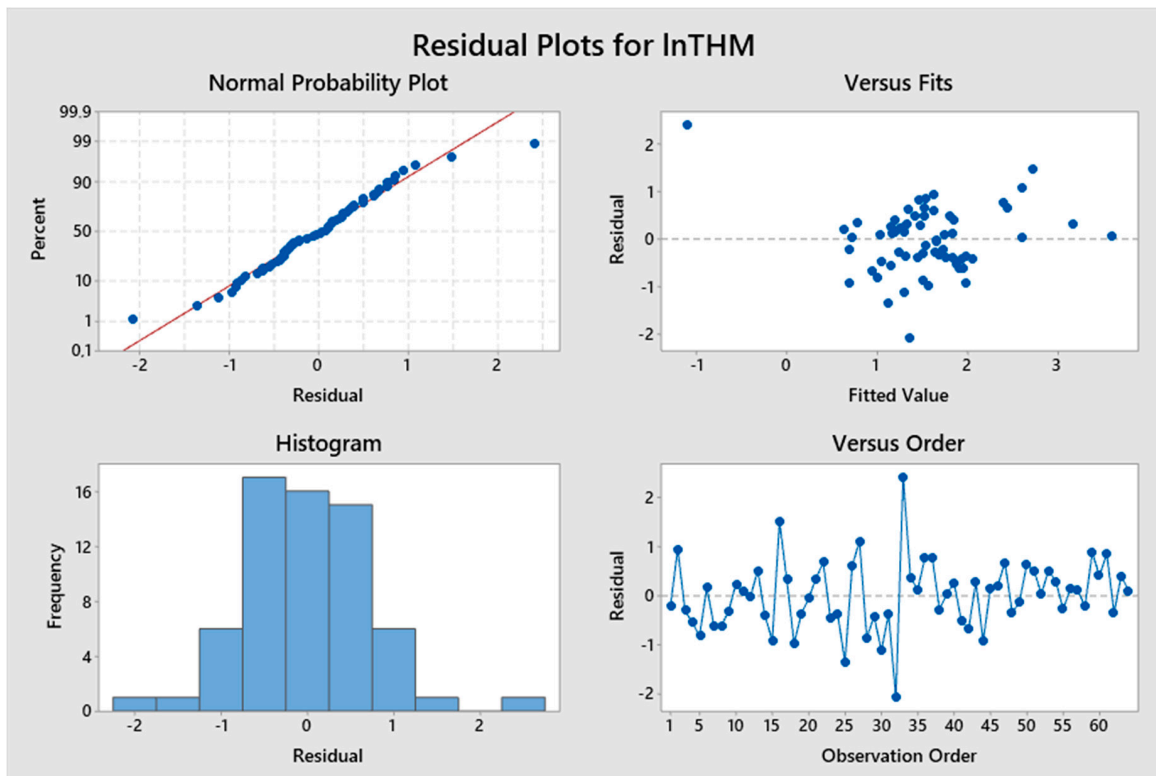


Figure 3. Residual plots for model (4).

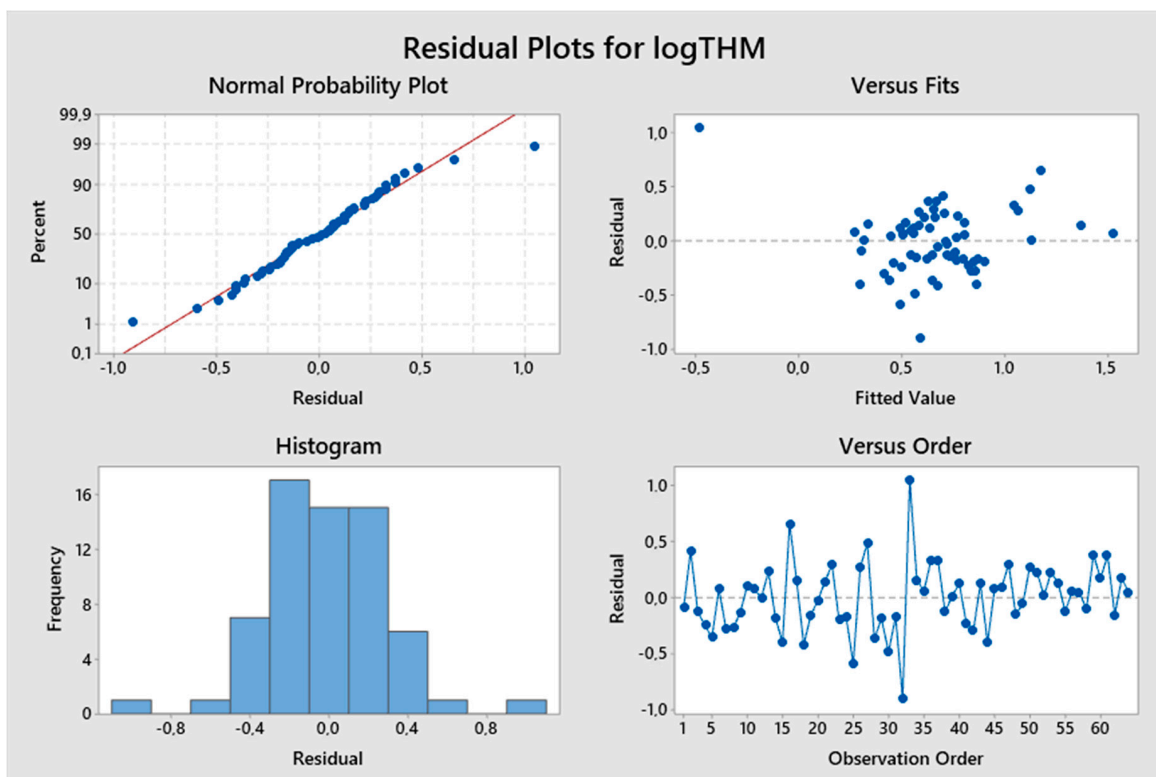


Figure 4. Residual plots for model (5).

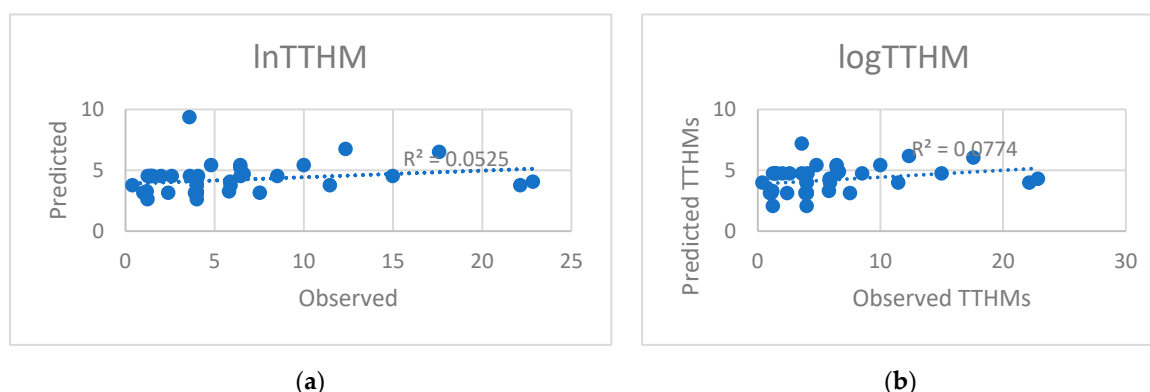


Figure 5. Predicted vs Observed values for (a) model (2); (b) model (3).

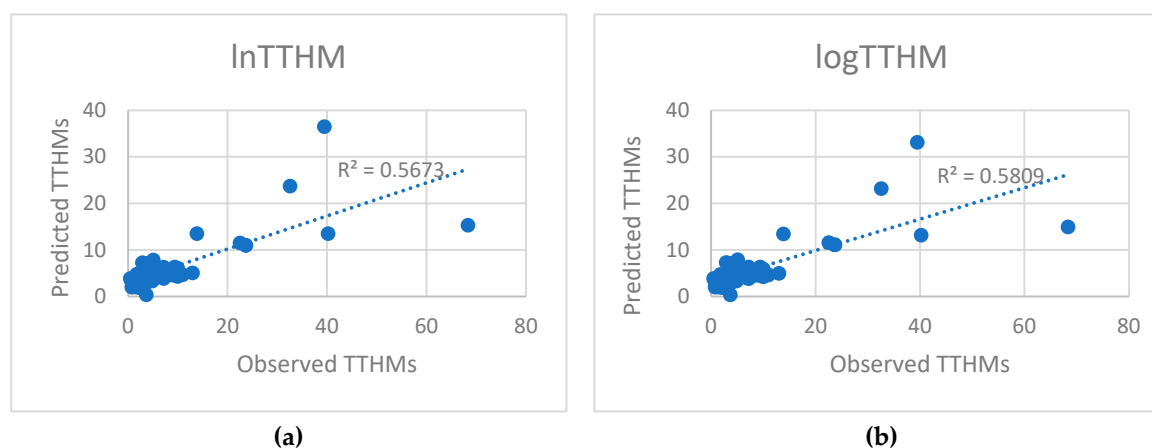


Figure 6. Predicted vs Observed values for (a) model (4); (b) model (5).

For the WDN B, the models provide a more satisfactory estimation of the TTHM concentrations formed. The percentage of observed values laying within $\pm 20\%$ of predicted values is 20.3% for both models (4) and (5). The regression coefficients (R^2) range from 0.5673 to 0.5809 (Figure 6a,b), which is a satisfactory level of explanation of the observed variability and can be compared to results reported in previous studies [29,31]. However, in other studies the regression coefficients are found to be much higher [25].

Performance evaluation was then used to test the accuracy of each calibrated model, which included the Mean Absolute Error (MAE), the Mean Square Error of Prediction (MSEP) and the maximum absolute error (MAX) produced by a prediction from the model. MAE is a measure of error expressed as the average of the absolute errors (difference between predicted and observed values). 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 [38]. MAX is the maximum absolute value of error and it is used as it gives an indication of the worst-case prediction made by the model. MAE values for models (2) and (3) are 3.43 and 3.40, respectively. MSEP values are 29.52 and 28.99, respectively, and MAX values are 18.79 and 18.55, respectively. For models 2 and 3 all the error indicators have slightly better (lower) values for model 3, but the difference is very small (Figure 7a). MAE values for models (4) and (5) are 3.80 and 3.88, respectively. MSEP values are 69.48 and 71.04, respectively, and MAX values are 53.07 and 53.45, respectively (Figure 7b). The error indicators show slightly better (lower) values for model 4, but the difference is small. According to Gibbs et al. [19], when comparing the prediction accuracy between two models the MSEP was considered the best indicator, as this error measurement penalizes larger prediction errors more harshly than the MAE.

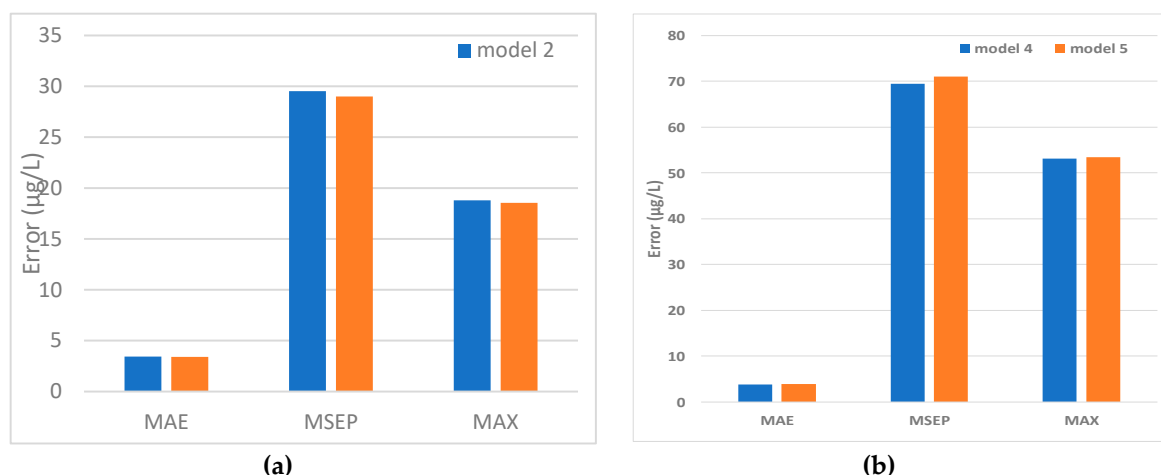


Figure 7. Error indicators for (a) models (2) and (3); (b) models (4) and (5).

Based on the analysis elaborated in this paper, the models developed for the WDN of city A are not reliable, as they fail to predict TTHM values accurately and cannot be used for the prediction of TTHMs in the network. A reason for this is that a full set of data with frequent measurements was not available from the water utility, as the water sampling frequency to detect TTHMs is very low. Another reason is that not all variables affecting the formation of TTHMs are measured and thus used in the model. As the variables monitored are set by legislation, the water utilities do not monitor other water quality variables. So, in order to develop reliable models for the prediction of TTHMs, enough data from the variables affecting TTHM formation is necessary.

The models developed for city B are more reliable as more data are available compared to city A. In this case also, the variables used are not all those affecting the formation of TTHMs and used in other models, as their measurements are not available from the water utility. It can be concluded from the analysis elaborated that the model (4) is more reliable and it can be used to predict the formation of TTHMs in the WDN of city B, if the variables' values are within the limits shown in Table 2. The models developed in these cases refer to water utilities using groundwater where, usually, organic matter is not present compared to surface water. Although organic matter needs to be present to form THMs, there are indications showing that THMs are formed even when groundwater is used. In the present cases, both WDNs use groundwater sources. However, from the data and the analysis, it is obvious that THMs are formed in low concentrations, probably because of the existence of organic matter which is present in this case. It is known that other factors such as the biofilm in the pipes' walls react with chlorine forming THMs. Comparing the models developed in this study with the models developed in other studies, it can be concluded that as pH values increase TTHM concentrations increase. However, there are models in the literature where pH effects on TTHM formation vary [8]. TOC increase results in increased TTHM concentrations in the present study, which is the case in other studies too [8]. However, other explanatory variables such as disinfectant dose, reaction time and others are not available in this study. The availability of reliable data is indicated by many studies as the factor motivating the choice of the explanatory variables. In the present study it must be noted that the model is site-specific and cannot be used extensively.

6. Conclusions

The paper presents an attempt to develop TTHM prediction models in two different WDNs in Greece. The paper aims at:

(a) Investigating the formation of THMs during chlorination of groundwaters taking into consideration the available variables. The effect of pH, and TOC was studied.

(b) Developing predictive models for the concentrations of total THMs formed during chlorination of these groundwaters. Modeling of THM formation can be helpful for the minimization of their

formation during water treatment, and therefore to comply with legislative measures and to protect human health.

(c) Statistically evaluating the developed models (correlation coefficient, Durbin–Watson estimate), in comparison to the models developed during previous studies for THMs [25,29,31,36] using the same modeling technique (multiple regression).

The study's results showed that the first two models developed for WDN A are not reliable and cannot be used for the prediction of TTHMs. Reasons for this include the lack of sufficient data and lack of data for explanatory variables affecting the formation of TTHMs. As groundwater is used, it is possible that the formation of TTHMs is affected by other factors, such as the pipes' biofilms existing in the water distribution pipes' network. The results showed that the models developed for WDN of city B can be used for the prediction of TTHMs, and more specifically model (4). However, as the models are not very reliable, further study should be conducted, including a larger dataset of the variables studied in this study and also explanatory variables such as disinfectant dose, retention time, temperature and variables indicating organic matter presence (such as chlorophyll a, dissolved organic carbon, etc.). Finally, as all samplings were done in autumn, it is suggested that seasonal variation should be taken into consideration elaborating more samplings all over the year, to study the effect of the season on the formation of TTHMs, which is found to be related in other studies [31].

Finally, it must be noted that the complexity of DBP formation reactions makes it difficult to develop universally applicable models. However, the existence of such models in regions and WDNs with the same characteristics could be useful, as they provide sufficient estimations of DBP concentrations that could minimize the need for complicated and expensive analysis of such compounds. DBP statistical models' development is a useful methodological basis to predict the formation of DBPs. Such models are helpful to the water utility managers during decision making, for example, setting disinfectant dose, the contact time, adjustment of pH, etc., in order to reduce DBP formation and at the same time maintain the required disinfectant residual. Prediction models can be used as a tool to select location for boosting chlorination residual levels to ensure complete removal of microbes and as well as minimization of DBP formation. The combination of DBP models with residual disinfectant models allow water utility managers to select the sampling points for water quality control within the distribution system [8]. Such models are used as the basis for epidemiological studies and health risk assessment [8].

Author Contributions: Conceptualization, methodology, statistical analysis, writing S.T.; supervision V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The research is elaborated within the framework of the invitation "Granting of scholarship for Post-Doctoral Research" of the University of Thessaly, which is implemented by the University of Thessaly and was funded by the Stavros Niarchos Foundation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kanakoudis, V.; Tsitsifli, S. Potable water security assessment – a review on monitoring, modelling and optimization techniques, applied to water distribution networks. *Destil. Water Treat.* **2017**, *99*, 18–26. [[CrossRef](#)]
2. Richardson, S.D.; Plewa, M.J.; Wagner, E.D.; Schoeny, R.; DeMarini, D.M. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutat. Res. Mutat. Res.* **2007**, *636*, 178–242. [[CrossRef](#)] [[PubMed](#)]
3. Ohanian, E.V.; Mullin, C.S.; Orme, J. Health effects of disinfectants and disinfection by-products: A regulatory perspective. *Water Chlorin. Chem. Environ. Impact Health Effects* **1989**, *6*, 75–86.
4. Environmental Protection Agency. *Water Treatment Manual: Disinfection*; Environmental Protection Agency: Wexford, Ireland, 2011.

5. Tsitsifli, S.; Kanakoudis, V. Disinfection Impacts to Drinking Water Safety—A Review. *Proceedings* **2018**, *2*, 603. [[CrossRef](#)]
6. Rook, J.J. Formation of haloforms during chlorination of natural waters. *Water Treat Exam* **1974**, *23*, 234–243.
7. Bellar, T.; Lichtenberg, J.; Kroner, R. The Occurrence of Organohalides in Chlorinated Drinking Waters. *J. Am. Water Work. Assoc.* **1974**, *66*, 703–706. [[CrossRef](#)]
8. Sadiq, R.; Rodríguez, M.J. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. *Sci. Total. Environ.* **2004**, *321*, 21–46. [[CrossRef](#)]
9. King, W.D.; Marrett, L.D. Case-control study of bladder cancer and chlorination by-products in treated water (Ontario, Canada). *Cancer Causes Control.* **1996**, *7*, 596–604. [[CrossRef](#)]
10. Boorman, G.A.; Dellarco, V.; Dunnick, J.K.; Chapin, R.E.; Hunter, S.; Hauchman, F.; Gardner, H.; Cox, M.; Sills, R.C. Drinking water disinfection byproducts: review and approach to toxicity evaluation. *Environ. Health Perspect.* **1999**, *107*, 207–217.
11. Chowdhury, S.; Champagne, P.; McLellan, P.J. Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review. *Sci. Total. Environ.* **2009**, *407*, 4189–4206. [[CrossRef](#)]
12. Nikolaou, A.D.; Arhonditsis, G.B.; Goulinopoulos, S.K.; Lekkas, T.D. Predicting the formation of trihalomethanes and haloacetic in surface waters by linear regression models. *Epidemiology* **2002**, *13*, 172.
13. Li, B.; Qu, J.; Liu, H.; Zhao, X. Formation and distribution of disinfection by-products during chlorine disinfection in the presence of bromide ion. *Sci. Bull.* **2008**, *53*, 2717–2723. [[CrossRef](#)]
14. Sohn, J.; Amy, G.; Yoon, Y. Bromide Ion Incorporation Into Brominated Disinfection By-Products. *Water Air Soil Pollut.* **2006**, *174*, 265–277. [[CrossRef](#)]
15. Ye, B.; Wang, W.; Yang, L.; Wei, J.; E, X. Factors influencing disinfection by-products formation in drinking water of six cities in China. *J. Hazard. Mater.* **2009**, *171*, 147–152. [[CrossRef](#)]
16. Abokifa, A.A.; Yang, Y.J.; Lo, C.S.; Biswas, P. Investigating the role of biofilms in trihalomethane formation in water distribution systems with a multicomponent model. *Water Res.* **2016**, *104*, 208–219. [[CrossRef](#)]
17. Minear, R.; Morrow, C.M. *Raw Water Bromide in Finished Drinking Water. Research Report 9*; Water Resources Research Center, University of Tennessee: Knoxville, TN, USA, 1983.
18. Ozekin, K. Modelling Bromate Formation During Ozonation and Assessing its Control. Ph.D. Thesis, University of Colorado, Denver, CO, USA, 1994.
19. Gibbs, M.S.; Morgan, N.; Maier, H.R.; Dandy, G.C.; Holmes, M.; Nixon, J.B. Use of artificial neural networks for modeling chlorine residuals in water distribution systems. In *Proceedings of the International Congress on Modeling and Simulation*; Modeling and Simulation Society of Australia and New Zealand Inc.: Townsville, Australia, 2003; Volume 2, pp. 789–794.
20. Rodríguez, M.J.; West, J.R.; Powell, J.; Sérodes, J.B. Application of two approaches to model chlorine residuals in Severn Trent Water Ltd (STW) distribution systems. *Water Sci. Technol.* **1997**, *36*, 317–324. [[CrossRef](#)]
21. Sérodes, J.B.; Rodríguez, M.J.; Ponton, A. Chlorcast (c): A methodology for developing decision-making tools for chlorine disinfection control. *Environ. Modell Softw.* **2001**, *16*, 53–62. [[CrossRef](#)]
22. Elshorbagy, W. Simulation of THM species in water distribution systems. *Water Res.* **2000**, *34*, 3431–3439. [[CrossRef](#)]
23. Abdullah, P.; Yew, C.; Bin Ramli, M.S. 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. *Water Res.* **2003**, *37*, 4637–4644. [[CrossRef](#)]
24. Amy, G.L.; Chadik, P.A.; Chowdhury, Z.K. Developing Models for Predicting Trihalomethane Formation Potential and Kinetics. *J. - Am. Water Work. Assoc.* **1987**, *79*, 89–97. [[CrossRef](#)]
25. Uyak, V.; Ozdemir, K.; Toroz, I. Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs. *Sci. Total. Environ.* **2007**, *378*, 269–280. [[CrossRef](#)] [[PubMed](#)]
26. Singer, P.C. Control of Disinfection By-Products in Drinking Water. *J. Environ. Eng.* **1994**, *120*, 727–744. [[CrossRef](#)]
27. Pourmoghaddas, H.; Stevens, A.A. Relationship between trihalomethanes and haloacetic acids with total organic halogen during chlorination. *Water Res.* **1995**, *29*, 2059–2062. [[CrossRef](#)]
28. Nikolaou, A.D.; Lekkas, T.D. The role of natural organic materials during formation of chlorination by-products: a review. *Acta Hydrochim. Hydrobiol.* **2001**, *2*, 63–77. [[CrossRef](#)]
29. Nikolaou, A.D.; Goulinopoulos, S.K.; Kostopoulou, M.N.; Lekkas, T.D. Decomposition of dihaloacetonitriles in water solutions and fortified drinking water samples. *Chemosphere* **2000**, *41*, 1149–1154. [[CrossRef](#)]

30. Golfinopoulos, S.K.; Nikolaou, A.D. Formation of DBPs in the drinking water of Athens, Greece: A ten-year study. *Glob. Nest J.* **2005**, *7*, 106–118.
31. Golfinopoulos, S.K.; Arhonditsis, G.B. Multiple regression models: A methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics. *Chemosphere* **2002**, *47*, 1007–1018. [[CrossRef](#)]
32. Farmaki, E.G.; Samios, S.A.; Thomaidis, N.S.; Golfinopoulos, S.; Efstathiou, C.E.; Lekkas, T.D. Artificial neural networks predictive models. A case study: carbon and bromine concentrations prediction based on chlorination time. *Global Nest J.* **2012**, *14*, 10–17.
33. Nikolaou, A.; Lekkas, T.; Golfinopoulos, S. Kinetics of the formation and decomposition of chlorination by-products in surface waters. *Chem. Eng. J.* **2004**, *100*, 139–148. [[CrossRef](#)]
34. Zar, J.H. *Biostatistical Analysis*, 2nd ed.; PrenticeHall: Englewood Cliffs, NJ, USA, 1984.
35. Ott, L. *An Introduction to Statistical Methods and Data Analysis*, 3rd ed.; PWS-Kent Publishing Company: Boston, MA, USA, 1988.
36. Golfinopoulos, S.K.; Xilourgidis, N.K.; Kostopoulou, M.N.; Lekkas, T.D. Use of a multiple regression model for predicting trihalomethane formation. *Water Res.* **1998**, *32*, 2821–2829. [[CrossRef](#)]
37. Uyak, V.; Toroz, I. Modeling the Formation of Chlorination By-Products During Enhanced Coagulation. *Environ. Monit. Assess.* **2006**, *121*, 503–517. [[CrossRef](#)] [[PubMed](#)]
38. Sprent, P.; Draper, N.R.; Smith, H. Applied Regression Analysis. *Biom. JSTOR* **1981**, *37*, 863. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).