A two-stage predictive model to simultaneous control of trihalomethanes in water treatment plants and distribution systems. Adaptability to treatment processes.

A. Domínguez-Tello [a,b,c], A. Arias-Borrego [a,b,c], T. García-Barrera [a,b,c] *, J.L. Gómez-Ariza [a,b,c] *

(a) Department of Chemistry and CC.MM, Faculty of Experimental Science, University of Huelva, Campus de El Carmen, 21007-Huelva, Spain.
(b) International Agrofood Campus of Excellence International CeiA3, University of Huelva, Spain.
(c) Research Center of Health and Environment (CYSMA), University of Huelva, Campus de El Carmen, 21007, Huelva, Spain.

Corresponding authors:
(*) J.L. Gómez-Ariza
Email: ariza@uhu.es
Tel. +34 959 219968
Fax: +34 959 219942
(*) Tamara García-Barrera
Email: tamara.garcia@dqcm.uhu.es
Tel. +34 959 219962
Fax: +34 959 219942

Abstract

The trihalomethanes (THMs) and others disinfection by-products (DBPs) are formed in drinking water by the reaction of chlorine with organic precursors contained in the source water, in two consecutive and linked stages, that starts at the treatment plant and continues in second stage along the distribution system (DS) by reaction of residual chlorine with organic precursors no removed. Following this approach, this study aimed at developing a two-stage empirical model for predicting the formation of trihalomethanes (TTHM) in the water treatment plant and subsequently their evolution along the water distribution system (WDS). The aim of the two-stage model was to improve the predictive capability for a wide range of scenarios of water treatments and distribution systems. The two-stage model was developed using multiple regression analysis from a database (Jan 2007 to Jul 2012) using three different treatment processes (conventional and advanced) in the water supply system of Aljaraque area (southwest of Spain). Then, the new model was validated using a recent database from the same water supply system (Jan 2011 to May 2015). The validation results indicated no significant difference in the predictive and observed values of TTHM (R^2: 0.874, analytical variance < 17 %). The new model was applied to three different supply systems with different treatment processes and different characteristics. Acceptable predictions were obtained in the three distribution systems studied, proving the adaptability of the new model to the boundary conditions. Finally the predictive capability of the new model was compared with seventeen other models selected from the literature, showing satisfactory results prediction and excellent adaptability to treatment processes.

Keywords

Trihalomethanes, two-stage predictive model, water treatment process, distribution system.

Introduction

In the last four decades, the chemical disinfection of drinking water has reduced significantly the incidence of infectious waterborne disease, but reactions of disinfectants such as chlorine with natural organic matter contained in source waters produce chemical mixtures of different undesirable compounds considered as disinfection byproducts (DBPs). Until now, more than 600 DBPs have been identified in drinking water (Richardson et al. 2007) and this number continues growing. Most drinking water treatment plants use chlorine for disinfection and therefore several types of chlorine containing DBPs are generated. Among them, trihalomethanes (THMs) and haloacetic acids (HAAs) are found at the highest concentrations in treated drinking water (Hamidin et al. 2008; Richardson 2003).

DBPs can enter the human body by multiple pathways, such as water ingestion, oral intake, inhalation through breathing and dermal contact through skin during regular indoor activities (showering, bathing and cooking). This chronic exposure to DBPs may pose risks to human health (Siddique et al. 2015), although inconsistent results have been reported across different epidemiological studies. In the case of THMs, the highest risk of total cancer in both males and females is associated to chloroform occurrence, mainly by inhalation (80-90% of the total risk), followed by oral exposure and dermal contact (Basu et al. 2011; Mishra et al. 2014). Several studies have reported associations between DBP exposure and increased risk of adverse developmental outcomes including low birth weight or small for gestational age births (Hoffman et al. 2008; Kumar et al. 2014), congenital anomalies, and birth defects such as cardiovascular and neural disorders (Grazuleviciene et al. 2013; Levallois et al. 2012; Nieuwenhuijsen et al. 2008). Others studies have found elevated rates of bladder, colon, rectum and brain cancers (Cantor et al. 2010; Melnick et al. 1994; Salas et al. 2013). For all these reasons, the different countries have regulated the permitted levels for the most prevalent DBPs in drinking water, in this way the United States Environmental Protection Agency established the limit of total concentration of four THMs to 80 μg/L and five HAAs to 60 μg/L (USEPA 2001) and the European Union has regulated the limit of total concentration of THMs to 100 μg/L from 2008 (98/83/EC 1998).

Since Johannes Rook (1974) discovered that THMs are formed by the reaction of chlorine with natural organic matter (NOM) in drinking water, hundreds of studies have been developed to determine the effects of THMs on health and its mechanisms of formation within–the treatment plant and its evolution along the distribution systems. The earliest models for predicting THMs and chloroform formation in drinking water were reported in 1983 (Engerholm and Amy 1983). To date more than 150 models have been developed through field and laboratory-scaled to predict DBPs formation in drinking water (Brown et al. 2011; Chowdhury et al. 2009).

These models have investigated the effects of different quality and operational parameters in controlling DBPs formation under a variety of environmental conditions, but mechanistic DBPs models are exceedingly difficult to derive due to seasonal, locational, and temporal variations in water quality, as well as the complexity of aquatic chemistry in terms of both disinfection kinetics and interactions in natural water matrices arising from heterogeneous natural organic matter (NOM) (Kulkarni and Chellam 2010).

Thus, most of these empirical models are site specific, and consequently their predictive capabilities in different water conditions remain inappropriate (Elshorbagy 2000). The development of a mathematical model that predict the formation of disinfection DBPs under different water quality and treatment conditions is of great interest and usefulness in the drinking water field (Lu et al. 2011). The great number of models proposed shows the challenge to get a universally applicable model (Golfinopoulos and Arhonditsis 2002).

In a recently study (Mayer et al. 2015), a large number of existing models were evaluated and overall poor performances were found. According to this study, most of the models are based on specific boundaries related to source water, water quality parameters and treatment conditions using untreated, coagulated or finished conventionally treated water. In this way, satisfactory model performance may be limited to a narrow range of treatment scenarios. Accordingly, the overall poor performance of the models tested, may be a function of applying them to datasets that did not satisfy all boundary conditions. Among the characteristics of distribution systems, piping materials (especially iron or copper) can affect the formation of DBPs.

A recent study showed that reactions between of certain organic precursors with zero-valent iron, may contribute to the formation of dichloroacetamide (DCAcAm) in distribution networks which contain cast iron pipes unlined, even in the absence of chlorinated disinfectants (Chu et al. 2016a). Metallic Cu alone did not affect HAcAm concentrations, but cooper increases reductive dehalogenation of haloacetamides by zero-valent iron in drinking water, reducing the integrated toxic risk (Chu et al. 2016b). A few studies focused on the consequences of suitable water treatment processes (Badawy et al. 2012; Mouly et al. 2010) concluding that DBPs formation and its spatial / seasonal variations depends largely on the efficient removal of NOM.

In a previous paper was demonstrates that most of THMs occur during the treatment process by reaction of chlorine with organic precursor compounds not removed during the preceding stages to the addition of the disinfectant. Trihalomethanes formation continues throughout the distribution system by reacting with the residual organic matter with the free residual chlorine and the chlorine applied in successive re-chlorination stages. This reaction is affected by environmental conditions as well as operational and morphological characteristics of the distribution system. The range of seasonal and spatial variation of trihalomethanes depends on the effectiveness of the removal of the organic matter during the treatment process (Domínguez-Tello et al. 2015), which is in good agreement with other authors in works to bench scale (Badawy et al. 2012) to real scale (Summerhayes et al. 2011), and laboratory scale (Sadrnourmohamadi and Gorczyca 2015) about a evaluation of the effect of ozonation on the total Trihalomethanes Formation Potential (THMFP) from river water samples.

The aim of this work is the development of a predictive model of formation of trihalomethanes with predictive capability in different scenarios of treatment and different distribution systems. To this end, the model was developed in two stages (treatment process and distribution system) using a wide range database. For this purpose, a historical database was used with water samples treated in Aljaraque WTP, with three different treatment processes. The model was developed using multiple regression analysis and validated using a recent database from the same distribution system. Subsequently, the model was tested on three water supply systems located in the surroundings of Huelva city (Southwest Spain), comparing the predicted and measured

TTHM values. The model fit well to different treatment scenarios with low prediction errors. The results showed good predictive capability, good adaptability to boundary conditions and low prediction errors. On the other hand, in this work seventeen models published were evaluated, analyzing comparatively the predictive capability of TTHM concentration of each of them, by variation the treatment process applied in the plant.

Materials and methods

Study sites

Water samples were collected from Jan 2011 to May 2015 through four water distribution systems (WDS) located at the province of Huelva (southwest Spain): Aljaraque (AL WDS), Lepe (Lep WDS), Riotinto (Rti WDS) and La Palma (Lpa WDS) (Fig. 1). All the systems worked with surface water source and used chorine in the form of sodium hypochlorite. Alj and Lep WDS supply water to a population of 60.000 and 150.000 inhabitants, respectively. Both plants operated with two conventional treatment processes in a seasonal scheme: (a) from October to May including pre-oxidation with potassium permanganate (pre-$KMnO_4$) coagulation-flocculation-sedimentation (CFS), rapid sand filtration (SF), second step filtration/adsorption with granular activate carbon (GAC) and disinfection; and (b) from May and September with the same treatment process but substituting the pre-oxidation by an advanced treatment with ozone (pre-O_3). The others two WDS, Lpa supply water to a population of 70.000 inhabitants using advanced treatment (pre-O_3, CFS, SF, GAC) and disinfection, and Rti for a population of 20.000 inhabitants using a conventional treatment process (pre-$KMnO_4$, CFS, SF, GAC) prior to the disinfection. All the WDS studied are affected by significant climatic and population variations, mainly the influence of seasonal coastal tourism. Table 1 summarizes the general characteristics of the water distribution systems studied.

Sampling strategy

An intensive monthly sampling campaign was performed in each supply system from January 2011 to May 2015. Sampling points were located at the water treatment plants (raw water and finished water) and two additional sample points in the reservoirs along of the distribution system. Thus, 720 samples were taken, 180 in each WDS under study (45 raw water, 45 finished water and 90 from distribution system). Samples were taken at the same day of the week and the first week of each month, using the same sampling route and the same sampling point in each selected location, ensuring the stability of hydraulic conditions and the representativeness of the samples in the studied system.

Samples were taken at the tap of each sampling point. In order to guarantee the representativeness of the sample, before sampling, it is necessary to renew the water contained in the section of pipeline between the sampling point and the reservoir or supply network to be sampled, for which the water was allowed to flow for at least five Minutes before filling the sample bottles. To analyze THMs, duplicate samples were collected in 125 mL amber glass bottles with teflon-lined screw caps, completely filling the bottle avoiding any headspace. A volume of 1.5 mL of 0.1 M sodium thiosulfate aqueous solution was added to eliminate any remaining residual chlorine quenching the sample to further THM formation. Temperature, pH, turbidity, conductivity and residual chlorine were *in situ* measured while dissolved organic carbon, UV_{254}, bromide, calcium and trihalomethanes species were determined in the laboratory. The samples were stored at 4 °C and analyzed within 2 days after collection. During each campaign, the operational parameters of the treatment plant and distribution system were collected to calculate the operational variables (chlorine dose in WTP and rechlorination, treatment flow, flow rate, water consumption and water level in the storage tanks).

Throughout the study, several actions have been implemented to reduce the effect of the uncertainties on the quality of the developed model: The coagulation / flocculation / sedimentation and oxidation processes were controlled and adjusted on the basis of jar-test. The sedimentation process was controlled to prevent flocs leakage. Through the sampling period, the quality of finished water was maintained at values of Fe, Mn and turbidity lower than 50 µg L^{-1}, 10 µg L^{-1} and 0.7 NTU, respectively. Stable process conditions were maintained, avoiding sampling during occasional process fluctuations. Sodium hypochlorite was used in the disinfection process; the concentration of chlorine in the solution (150–123 g/L) was used for calculating the accumulated dose of chlorine and chlorine dose. The contact time of water in reservoirs was calculated daily considering the flow of water supplied to each population nucleus by assuming complete mixing inside the reservoirs. The contact time considered for model development was the weekly average, in stable conditions.

Analytical methods

Water samples were analyzed for the regulated THMs using headspace-solid-phase microextraction (HS-SPME) coupled to gas chromatography-mass spectrometry (GC-MS), using a Varian CP-3800 gas chromatograph coupled to an ion trap mass spectrometer Varian Saturn 2000 MS (Varian, Sunnyvale, CA, USA). The analytical method has been published elsewhere (Domínguez-Tello et al., 2015). Briefly a DB-5 ms 30 m × 0.25 mm × 0.25 µm capillary column (Agilent Technologies) was used for the chromatographic separation of THMs, using the following temperature program: 40 °C for 4 min., ramped to 120 °C at 10 ºC/min. and hold for 1.5 min., and finally ramped to 250 °C at 25 ºC/min. for 5 min. (total run time: 23.7 min.). Injections were made in split mode (1:10) for 3 min. at 220 °C. Helium was used as carrier gas at a constant pressure of 29 kPa and a constant flow rate of 1 mL/min..

The HS-SPME extraction method has also been described in the previous publication (Domínguez-Tello et al. 2015). Briefly, the fiber used was made of carboxen/polydimethylsiloxane (CAR/PDMS 85 µm) purchased from Sigma-Aldrich. Before use each fiber was conditioned at 250 °C for 30 min. For HS-SPME extraction, 2 mL of sample was transferred to a 4 mL sample vial together with a magnetic bar, 250 µL of saturated sodium chloride solution and 2 µL of a 5 mg/L of internal standard solution (1,2-dibromopropane). The samples were sealed using screw cap, containing a PTFE-faced rubber septum. The analytes were extracted at 40 °C for 30 min. with stirring speed of 250 rpm. Then the fiber was introduced into the GC injection port at 270 °C during 4 min. for desorption.

Electron ionization mass spectra were recorded in scan mode using the m/z 29–300 at 3.5 scans per second. Each compound was quantified by comparing the relative area of the internal standard to the target analyte. The limits of detection (LOD) and quantification (LOQ), were 1.3, 0.8, 1.1, and 0.9 µg/L and 4.2, 2.5, 3.6, and 3.0 µg/L for $CHCl_3$, $CHCl_2Br$, $CHClBr_2$, and $CHBr_3$, respectively. The features of the method and validation dates were listed in Table S1 of Supplementary information. Milli-Q water was used through and was purified in a Gradient system (Millipore, Watford, UK). All the standards were of analytical grade and purchased from Sigma-Aldrich (Madrid Spain) Solvents were of HPLC grade and obtained from Sigma-Aldrich.

Sampling campaigns were monthly performed and quality control of the analytical method evaluated using four external standards with different concentrations of THMs, which confirm results reliability. Not less than 25% replicate samples were analyzed for THM to evaluate the method precision. Blanks were also used for background correction and error sources detection.

Conventional parameters for water quality control were analyzed using the following approaches: free residual chlorine was measured using a Hanna photometer HI-93711 following colorimetric method DPD according to Standard Method 4500-Cl-G. Turbidity was measured using a HACH 2100P turbidimeter. Bromide was analyzed using an ion-chromatograph (METROHM 861 Advanced compact IC) with chemical suppression and conductivity detector. A Metrohm 744 pH-meter equipped with a gel-filled electrode Water pH was used for measuring pH. Conductivity was measured with Crison CM35 conductivimeter. Samples were filtered using a 0.45 µm nylon membrane filter prior to the measurement of UV absorbance and DOC. UV_{254} absorbance spectra were measured using a Perkin-Elmer Lambda 18 spectrophotometer with 5 cm quartz cell and latterly spectra were normalized to a 1 cm cell length. DOC concentrations were obtained using a TOC-5000 Shimadzu analyzer, according to EPA Standard Method 5310C. SUVA was calculated by normalizing UV_{254} values with respect to DOC.

Modelling and validation

In this study, the predictive model of THMs was developed in two stages: (1) the treatment process, and (D70/2009) the distribution system, since as explained before, the THMs formation occur during the treatment process (first stages) starting with the addition of chlorine by its reaction with organic precursor compounds not removed during the earlier stages of treatment. As above commented, the TTHM variation in the water distribution system is influenced by trihalomethanes concentration and organic precursors in finished water, pH water and environmental variables, operational and morphological characteristics of the distribution system (Fig. 2).

Modelling

The approach tries to predict the formation of trihalomethanes in the water treatment process and subsequently their variation throughout the distribution system. The model was developed using a database (Domínguez-Tello et al. 2015) which considers 198 samples taken in a monthly sampling campaign conducted between January 2007 and July 2012 in Alj WTP, three reservoirs and networks of its distribution system. During the sampling campaign, the plant worked with four treatment processes from which only three were selected for the study. The database used provides a wide range of values of trihalomethanes (22.6-123.9 μg L^{-1}) which provides a high statistical potential and high range of application of the model. Subsequently the model was validated using the results from recent sampling campaign between February 2011 to July 2015, in Alj WDS.

A recent review from Ged at al. discusses the predictive capability and application of 87 DBPs models published over the last 30 years focused on chlorine disinfection. The results showed that multivariable power law models had the highest predictive capability for TTHM. Additionally, the best models for predicting TTHM were those including at least five of the seven explanatory variables (DOC, UV_{254}, Br^-, pH, chlorine dose, reaction time and temperature) (Ged et al. 2015). In accordance with this study, the direct explanatory variables initially considered in the predictive model were DOC, SUVA, UV_{254}, ion bromide (Br^-), pH, chlorine dose (d), accumulated chlorine dose (D), reaction time (t), temperature (T) and three composite explanatory variables associated with the disinfection reaction: R_{EF}, R_{DS} and δ, where R_{EF} is the product of the following variables: chlorine dose in WTP (d), reaction time (t), temperature (T) and UV_{254} absorbance in finished water. Similarly, the composite variable R_{DS} is the product of variables involved in the formation of TTHM along the distribution system (stage 2), including TTHM calculated in finished water ($TTHM_{Ef}$), which can be potentially produced by reaction of precursor unoxidized organic matter (UV_{254}), accumulated dose of chlorine (D), contact time (t_{DS}) and water temperature

(T_{DS}) in the point selected of distribution system. Finally, δ is an explanatory variable expressed as the difference between the dose of chlorine added in the treatment process and the value of free residual chlorine in the water finished, in relation to the contact time between the two points (Mouly et al. 2010). The detailed description of explanatory variables used in the two stages of the predictive model is shown in Fig. 2.

After verifying the statistical significance using the Pearson correlation matrix at 95% significance level: $p < 0.05$, the explanatory variables was selected. For that, the combinations of explicit variables were tested and were selected those in which the best statistical results and reproducibility of the model were obtained following criterion of maximum R^2, minimum standard error *s* and *Cp Mallows.*

For development of the model, multiple regression analysis of data was carried out. Polynomial: $Y = K + X_1 b_1 + X_2 b_2 + \ldots X_p b_p$ and logarithmic: $Y = K (X_1)^{b_1} (X_2)^{b_2} \ldots (X_p)^{b_p}$ forms were tested, where Y is the variable to be modeled (TTHM), X_i, $i = 1$ to p are the explanatory variables, b_i=1 to p represent the statistical coefficients to be estimated and K is a constant term. The model was developed according to the best combination of variables obtained in the statistical analysis.

Comparative statistical analyses of measured and predicted data from the model options, started with F-test, Student's T-test, linear correlation coefficient (R^2), analytical variance (AV): percentage of the absolute difference between the measured and predicted values, and standard error (SE) or root mean square error. F-test analysis determined the variance similarity between observed values and predicted values. If the F-test value was > 0.5, the Student's T-test with equal variance was conducted, and otherwise if F-test < 0.5 Student's T-test with unequal variance was conducted. If the Student's T-test result was < 0.5, the two data sets had no statistical similarity, and they were not equivalent. Instead if the Student's T-test result was > 0.5, the two data sets had no significant statistical differences, that is, they were equivalent and then uncertainty analyses were calculates: SE, AV, and linear correlation coefficient (R^2) (Chen and Westerhoff 2010). Both AV and SE reflected the deviation or uncertainty of predicted data relative to measured data; and R^2 indicated the correlation between predicted data and experimental data. According to the statistical analysis the model with higher value Student's T, higher R^2 and lower AV and SE was selected. Moreover the statistical significance of the selected model was checked by the F-value and Durbin-Watson estimate (1.5 – 2.5).

Validation and application

The purpose of validation is to measure the goodness of fit of the values predicted by the model in comparison with the experimental data measured. In order to validate the new two-step model developed, the $TTHM_{Ef}$ and $TTHM_{DS}$ values were calculated using the model equations for an independent set of additional database obtained from Aljaraque WDS between Jan 2011 to May 2015 (35 samples).

To evaluate the adaptability of the new model to different treatment processes and different conditions of the distribution system, the new model was applied to three different WDS (Lepe, La Palma and Rio Tinto).

The predicted values were compared with measured values calculating the difference between them, using AV, SE and R^2. A t-test was done on the predicted models and to determine the biasness by calculating the t_{value} for models compared to the $t_{critical}$ value (If $t_{value} < t_{critical}$ the biasness is considered to be not significant).

Results and discussion

Occurrence of THMs. Seasonal and spatial variation

Seasonal variations of the temperature and raw water quality cause variations of DBPs concentrations in the water supply. Additionally, seasonal changes in water consumption habits, human activities and environmental changes favor such variations (Fokmare and Musaddiq 2001; Karapinar et al. 2014). Therefore, to maintain suitable water quality according to the established regulations it is necessary to combine the water treatment processes with seasonal raw water conditions and the characteristics of the supply system. Considerable seasonal variations of drinking water quality has been reported in many drinking water systems including small water distribution system (Scheili et al. 2015).

In this work both the seasonal (from winter to summer) and spatial (from the water treatment plant to the end points of the distribution system) variations of TTHM were evaluated in the four WDS located at important areas of the southwest Spain (Aljaraque, Lepe, Riotinto and La Palma) during the period in which it was developed and validated the model. The seasonal and spatial variations of THMs are shown in Table 2.

TTHM levels were higher in summer followed by spring and lower in autumn and winter. The average levels of THMs measured in summer at the water treatment plant of Alj, Lep, Rti and Lpa were 1.41, 1.34, 1.41 and 1.15 times higher than the average levels in winter, respectively. The lower range of seasonal variation occurs in Lpa WTP where advanced treatment process with ozonation and GAC were used. The ranges of spatial variation in Alj, Lep, Rti and Lpa water distribution system were 1.1, 1.26, 1.34 and 1.13 times the concentration of TTHM in treated water, respectively.

Influence of the oxidation process. Ozonation test

The natural organic matter present in the source of water is the major precursor to the formation of DBPs. Water utilities need to apply treatment technology and optimize the treatment processes to remove organic precursors to effectively reduce the formation of DBPs (Hua et al. 2015).

To evaluate the influence of the treatment process and especially the oxidation process on the formation of THMs in distribution system, a real scale test was carried out in the Lepe WTP varying the ozone dose. A sampling campaign for two months with daily sampling of raw water, treated water and three sampling points of the distribution system (R1, R2 and R3) was performed. The results are shown in Fig. 3.

In agreement with other authors (Bond et al. 2014; Galapate et al. 2001; Sadrnourmohamadi and Gorczyca 2015), trihalomethanes formation depends on both oxidation kinetic and halogenation steps. The transformation of dissolved organic carbon during ozonation results in a higher reduction in TTHM (by conversion of hydrophobic fractions -main contributors to the formation of TTHM- to hydrophilic fractions).

As a result, it was found that increasing ozone dosage (1, 2 and 3 mg L^{-1}) reduce the content of DOC (31.2, 32.8 and 38.3%, respectively) and trihalomethanes (14, 34 and 48%, respectively) in the distribution system. Furthermore as shown in Fig. 3, a higher oxidation treatment contributes to reduce the variability of TTHM throughout the distribution system, suggesting less seasonal and spatial variation, achieving greater stability of supply water quality. The results suggest the importance of considers the effect of the treatment process in the development of DBPs predictive models.

Effect of natural organic matter

The natural organic matter present in the source water serves as the major precursor to the formation of DBPs. Aquatic NOM is a complex mixture of heterogeneous organic compounds varying in structure and functionality from source to source. Therefore, surrogate parameters are used to predict its removal through treatment, estimating its reactivity toward DBP formation, such as total organic carbon (TOC), dissolved organic carbon (COD), ultraviolet (UV) absorbance, and specific ultraviolet absorbance (SUVA) (Hua et al. 2015; Matilainen et al. 2011; Reckhow et al. 1990). SUVA is a good indicator of the formation of unknown DBPs, but generally the correlation between SUVA and THMs is poor, because THMs are produced from diverse types of precursors including UV and non-UV absorbing organic compounds. NOM with high SUVA values is rich in humic substances, hydrophobic compounds, and high molecular weight organic matter. (Ates et al. 2007; Kitis et al. 2002). The SUVA values of the raw waters from the four reservoirs used for this study were lower than 2 L mg^{-1} m^{-1}, suggesting the presence of low molecular weight compounds from NOM and low humic acids content.

In our study data from DOC, SUVA, UV_{254} were tested using the Pearson correlation test (Table 3). The highest correlation was found with the UV_{254} variable. Therefore, UV_{254} measurement was adopted as a critical variable of the model, which can be easily measured. In the individual study of each treatment process, positive correlations between the variable UV_{254} in treated water and the formation of TTHM were obtained (r: -0.696, -0.365 and -0.704 in TP1, TP2 and TP3, respectively), however a non-significant correlation (r: 0.140) was found in the pooled data from the three process schemes. The reason of this effect could be the different removal efficiency of NOM of each treatment scheme (Badawy et al. 2012).

The UV_{254} varies according to the treatment applied in the WTP and is a good indicator of the potential formation of THMs. High UV_{254} values in the finished water indicates poor oxidation and a high potential for formation of THMs throughout the distribution system. Thus the variable UV_{254} was applied in both stages of the model as significant indicator of trihalomethanes reaction formation.

Effect of pH

The concentration of THMs increases at high pH as a result of numerous hydrolysis reactions occurring in these compounds, and the increasing formation of hypochlorite ions at these pH, which reduce the effectiveness of chlorine disinfection. As a consequence, at higher pH values, more TTHM are formed (Hong et al. 2007; Zimoch et al. 2015). In this study a positive Pearson correlation was obtained between water pH and TTHM concentration (Table 3), both in global data (r: 0.562) and those from data groups of each treatment process (0.561, 0.874 and 0.877 in TP1, TP2 and TP3, respectively). Good correlation between water pH and the TTHM formation in distribution system (stage 2) was also found (r: 0.687).

Effect of water temperature

In the area under study, a clear relationship of water temperature with the formation of trihalomethanes was found (Table 3), which can be explained by the effect the temperature in the organic matter removal effectiveness during the treatment process, which was confirmed by the results obtained in the conventional treatment processes, used in this study, TP1 and TP3 (r: 0.944 and 0.918 respectively). However, a positive, but less marked correlation was observed in the advanced process TP2 (r: 0.220), which matches with low UV_{254} values in finished water due to the strong oxidation of organic matter by ozone. Despite the clear importance of temperature in the reaction of formation of TTHM, no direct correlation was found when Pearson test was applied to global data.

Effect of chlorine dose

Using the Pearson correlation method a strong relationship (r: 0.748) was obtained between THM formation and chlorine dose used in the treatment processes

($TTHM_{Ef}$). Also a strong relationship (r: 0.881) was obtained between TTHM formation in distribution system ($TTHM_{DS}$) with the accumulated dose of chlorine (Table 3).

Effect of contact time

In this study the contact time (t_{Ef}) for the reaction of chlorine with organic matter was measured from the chlorine dosing point to the finished water in the WTP. We also measured the contact time from the point of finished water to different sampling points in the distribution system (t_{DS}). The contact time in the studied water treatment plants (t_{EF}) were between 0.10 and 3.25 h. The contact time in the distribution systems (t_{DS}) were between 19.7 and 30.0 h.

In both cases a strong relationship was obtained between the contact time t_{Ef} (r: 0.951) and t_{DS} (r: 0.965) with the $TTHM_{Ef}$ and $TTHM_{DS,}$ respectively.

Effect of bromide

In the chlorination process, DBPs concentration increases with the level of bromide. This is because the bromide ion can be oxidized by free chlorine to produce hypobromous acid (HBrO) that reacts with NOM with more substitution ability than HClO. When the level of bromide increases, more bromide could be incorporated into DBPs, and consequently, the formation of chlorine-containing species decreases. Moreover, the weight of bromine atom is higher than chlorine, so DBPs formation increases more significantly with the increase of bromide (Bougeard et al. 2010; Hong et al. 2013; Watson et al. 2015).

In the present study a good relationship (r: 0.657) between the bromide in finished water with $TTHM_{Ef}$ was obtained. However, peak values of bromide in the raw water were observed that not always are neutralized during the treatment process, which can cause elevations of TTHM in the distribution system. Therefore, the bromide variable was included in the model, despite its low background in the water supply systems studied.

Reaction variables

TTHM formation behaves as a first order reaction with respect to chlorine dose and humic acid precursors. Therefore TTHM formation can be formulate as a function of the concentration of THMFP (humic acid precursor), residual chlorine, reaction time and reaction temperature (Li and Zhao 2006). Following this criterion, the variable reaction (R_{Ef}) was established as an indirect indicator of the reactivity of organic matter with chlorine in treated water in the plant. R_{Ef} represents the product of the dose of chlorine, contact time, water temperature and UV_{254} absorbance measured in finished water. Using the Pearson method (Table 3), a strong relationship (r: 0.965) was obtained between R_{Ef} and $TTHM_{Ef.}$ Similarly R_{DS} represents the product of the dose of chlorine, contact time, water temperature and UV_{254} absorbance measured in finished water ($R_{DS} = D \times t_{DS} \times T_{DS} \times UV_{254}$). A strong relationship (r: 0.749) was obtained between R_{DS} and $TTHM_{DS.}$

The TTHM in distribution system was inversely proportional to the variable δ. Good relationship was obtained between δ, $\delta \times T$ and $UV_{254} \times \delta \times T$ with $TTHM_{EF}$ (-0.861, -0.713 and -0.611 respectively). However, these variables were not selected in the model proposed since the correlation obtained between predicted and calculated values was lower (R^2: 0.789-0.878) and standard error higher (19.7-14.9) in respect to the variable *R*.

Modeling developed

After verifying the effect of different variables on the formation of trihalomethanes in the two stages, the statistical significance and their Pearson correlations, the explanatory variables were selected. Among the possible combinations of explicit variables (pH_{Ef}, d, t_{Ef}, T_{Ef}, Br^- UV_{254}, R_{Ef}, δ, δ T), pH_{Ef}, Br^- and R_{Ef}, were selected, because these variables obtained the best statistical results reproducibility model: R^2: 0.948, SE: 8.67 and Cp Mallows: 4.0 (Table 4).

Different options of models (linear and polynomial) were evaluated according to the best combinations of variables obtained in the statistical analysis and the accuracy of the predictions regarding the measured values. Based on the results obtained, the lineal model was selected. The summary of TTHMs models is shown in Table 5. The results of Student's T-test were > 0.5 (0.99 and 0.90 for Stage 1 and 2 respectively) shown no significant statistical difference between measured and predicted values. The analytical variance (AV), standard error (SE) and linear correlation coefficient (R^2) were: 13.6, 8.67, 0.948 and 9.9, 6.08 and 0.9 for Stage 1 and 2 models respectively. The new model is statistically significant and the value of the Durbin-Watson statistic were found to be 1.74 and 1.62 for Stage 1 and 2 respectively. The value of Durbin-Watson is preferred to be between 1.5 and 2.5 for statiscally best model (Kumari and Gupta 2015; Uyak et al. 2007). According to the comparative results, the lineal model was selected:

$$TTHM_{Ef} = 165 – 21.3\ pH_{Ef} + 0.232\ Br^{-} + 5.84\ R_{Ef}$$

$$TTHM_{DS} = 14.9 + 1.01\ THM_{Ef} + 0.20\ pH_{DS} - 0.104\ R_{DS}$$

where: $R_{Ef} = d \times t_{Ef} \times T \times UV_{254}$; $R_{DS} = D \times t_{DS} \times T_{DS} \times UV_{254}$

The analysis of variance (Garcia-Villanova et al.) showed that the model was statistically significant ($p < 0.05$). The examination of statistical residuals of model showed a normal distribution of data evently distribuited above and below the zero baselines and no visible trends. The range of application of the developed model is restricted by range of quality and operational variables taken as the basis of design and validation: $TTHM_{EF}$ (22.6–125.5μg L^{-1}), COD (1.40-5.09 μg L^{-1}), UV_{254} (0.017-0.076 cm^{-1}), pH_{Ef} (6.50–7.80), T (10.6-26.6 ºC), t (0.10-3.25 h), Br^{-} (20.0-176 μg L^{-1}), d (0.70-5.80 mg L^{-1}), $TTHM_{DS}$ (27.3–130.1 μg L^{-1}), t_{DS} (19.7-30.0 h), D (2.97-6.31 mg L^{-1}), pH_{DS} (6.73-7.75).

In developing the model, some effects that could be limiting and affect the quality of the predicted results can be observed. Therefore, the data were selected and those corresponding to unstable or anomalous behavior were removed. Thus, data which trihalomethanes in distribution system were lower than in finished water were discarded. This effect was observed in some reservoirs far away from the WTP, where the inlet water was cascading, affected probably by air-stripping. Furthermore, it is observed that quality measuring of certain operational variables such as contact time and cumulative dose of chlorine could be sources of uncertainties in the predictive model. However, it should be pointed out that the source water used in the development, validation and application of the model contains low levels of bromide and it would be advisable to check the effectiveness of the model with high levels of this ion.

Model Validation

The new predictive model was validated in the same WDS in which was developed (Aljaraque WDS) during period Jan 2011 to May 2015 (35 samples). During the validation period Aljaraque WTP operated with conventional treatment process from October to May (pre-$KMnO_4$, CFS, SF, GAC and disinfection) and advanced treatment between May and September (pre-O_3, CFS, SF, CAG and disinfection). The results of the validation analysis are shown in Table 6. A t-test was applied to determinate the biasness of the model. The t_{value} for the two stages (Alj WTP and Alj DS) were less than the $t_{critical}$ value and p values were also greater than 0.05, which indicates that the model biasness were no significant. The uncertainly analysis show a low deviation of predicted data relative to measured data. The standard errors (SE) of the two stages (Alj WTP and Alj DS) measured as root mean standard errors were 2.57 and 3.00 respectively. The analytical variance (AV) measured as average percentage of the absolute difference between measured and calculated by the model were 7.38% and 7.22% respectively. The maximum differences between measured and calculated TTHM in the two stages of the model were 17 and 12%, slightly lower in the DS indicating some adjustment in the second stage model. The analysis of variance of the first stage (WTP) obtained prediction errors (AV) greater than 10% in the values range studied (TTHM: 28.3-56.5 µg L^{-1}). Likewise in the second stage model (DS) 23% of the predictions obtained a variance > 10%. The two-stage model validation indicates very satisfactory predictions with R^2 of 0.91 and 0.87. The Fig. 4 shows measured vs. predicted TTHM values in water treatment plant (Stage 1), and distribution system (Stage 2).

Model Application to different water supply system

In order to check the suitability of the new two-stage model and its adaptability to different boundary conditions, the model developed was applied in three different water supply systems: Lepe, Riotinto and La Palma water distribution systems with substantial differences in treatment processes, water source and distribution system characteristics. The quality characteristics of raw water, treated water and water distribution system of Aljaraque, Lepe, Riotinto and La Palma are shown in Tables S2, S3 and S4 of supplementary information.

The numeric results of the application are shown in Table 6. Graphically, the measured and calculated values for the different distribution systems studied are represented in the Fig. 5. To analyze the results of applying of the model, similar statistical procedure to validation was followed. A t-test was used to determine the possible bias data. The predicted values measured values were compared with calculating the differences between them, using AV, SE and R^2.

The uncertainly analysis show low deviations of predicted data relative to measured data in the three WDS studied with SE and AV between 2.98-4.94 and 6.67-9.80 % respectively. The maximum differences between measured and calculated TTHM in the three WDS were between AV 14.0 to 26.6%. The largest variance was obtained in Lpa WDS, coinciding with the widest range of values TTHMs (14.8-64.8 µg L^{-1}). The R^2 obtained were between 0.82 to of 0.96. The new model was tested globally by adding all the data used from validation and application (Fig. 6) obtaining very good predictive capability of TTHM in the water treatment plant (R^2: 0.940, SE: 3.18, AV: 7.79%) and distribution system (R^2: 0.87, SE: 4.05, AV: 8.30%).

The results obtained shows good adaptability of the two-stage developed model to the boundary conditions of the three supply systems studied, despite their differences in source water, treatment processes and the characteristics of the distribution systems.

Comparison predictive capability of different models according to the treatment processes

Most empirical DPBs model proposed in the literature are based on databases from specific treatment conditions, water quality and distribution system. As a consequence, introduce specific values ranges related to boundary conditions and cannot be applied to any real situation (Amy et al. 1987). In the development of predictive models, the treatment process is a critical factor to be taken into account by using databases from different treatment scenarios. To demonstrate the influence of the treatment process on the accuracy and applicability of predictive models of THMs, a comparative study of the predictive efficiency of seventeen models was performed.

Mathematical expressions of the models were applied by replacing the explicit variables in the database Alj WDS, with groups according to the treatment process (TP1, TP2 and TP3). TP1: prechlorination with conventional treatment process, TP2: advanced treatment process with pre-ozonation, inter-chlorination, filtration and GAC, TP3: conventional process using potassium permanganate preoxidation and chlorination. The prediction results from the different models TTHM were compared with measured values. The comparative statistical analysis was performed using SE, AV and R^2. The predictive models evaluated are shown in Table 7 and the comparative results obtained, in Table 8.

Of seventeen predictive models evaluated, seven models obtained good prediction capability of TTHM with only one treatment process, specially the conventional process with prechlorination (TP1), however high errors were obtained with other treatment processes. Among them the models M11, M7, M3, M2 and M13 obtained small prediction errors, M14 and M4 obtained moderate errors. Also, other predictive models developed in WDS with conventional treatment processes, high concentrations of organic matter in water source and significant seasonal variations (Kumari and Gupta 2015), showed good predictive capacity for the conventional process TP1

Seven models (M5, M15, M16, M8, M10, M6 and M9) satisfy the boundary conditions with two treatment processes simultaneously, obtaining acceptable predictions for TTHM with the advanced treatment process TP2 and the conventional TP1. The M10 model obtained the best result with the conventional pre-chlorination process, and acceptable values with the conventional process with permanganate. The M12 model obtained low predictive capacity with the three treatment processes studied.

Of the seventeen models studied only two (M1 and M17) provided acceptable results in TTHM prediction levels for any treatment process. The M17 obtained the best overall results of predictive capacity in the three treatment models studied (SE: 18.66 AV: 27.36 and R^2: 0.86). The prediction results with M1 model are also good but with relatively high errors.

The new model developed in this work provides clearly the best results, with good individual predictive capability in each treatment process (SE: 7.97 6.37 and 11.51%, AV: 4.99%, 12.13%, 25.06% in TP1, TP2 and TP3 respectively) and overall good predictive capability (SE: 8.88, AV 16.63% and R^2: 0.94).

According to the results obtained, it was verified that most of the models are specific in their application and no satisfy the boundary conditions of all the treatment processes. Therefore, most of these models cannot be applied globally. The results suggest the need to develop predictive models of DBPs from databases that include different treatment scenarios, obtaining a wide range of application**.**

Conclusions

In this paper a predictive model of trihalomethanes formation in two stages (WTP and DS) was developed, which gets good predictive capability for a wide range of scenarios of water treatments and distribution systems. The two-stage model developed predicts with low error, the formation of TTHM in treatment process and water distribution system from quality and operational variables. The model developed links for the first time the formation of trihalomethanes in the distribution system with the effectiveness of the treatment process applied in the plant. Thus, the model can be used as a useful preventive tool for process treatment control, alerting about setting requirements that prevent high levels TTHM in drinking water.

The new predictive model includes two direct explanatory variables: pH and ion bromide (Br^-) and two composite variables R_{EF} and R_{DS} associated with the disinfection reaction in WTP and DS respectively. Both composite variables were calculated as the product of other direct variables: organic matter (UV_{254}), contact time (t_{Ef} and t_{DS}), chorine dose (d and D) and temperature (T and T_{DS}).

In this work it has been verified that the treatment processes applied in the WTP have a high influence on the predictive capability of TTHM in the distribution system. It was shown that an efficient oxidation treatment in the WTP contributes to reduce the range of TTHM values in the distribution system, thus reducing the effect of seasonal and spatial variation, achieving a higher stability of supply water quality. This result underscores the importance of considering the effect of the treatment process on the development of predictive models of DBPs, using databases that include different treatment scenarios.

The strategy of development of two-stage DBP predictive models using data from different treatment processes can contribute to improving the adaptability of future developments of models to different boundary conditions and to increase its range of application.

Acknowledgements

The authors wish to express their appreciation to the technicians and managers of the public entreprise GIAHSA (Gestión Integral del Agua de Huelva S.A.) for their help and cooperation in facilitating the collection of water samples and operating data during this real scale study. This work was supported by the project CTM2015-67902-C2-1-P from the Spanish Ministry of Economy and Competitiveness (MINECO), and by the project P12-FQM-0442 from the Regional Ministry of Economy, Innovation, Science and Employment (Andalusian Government, Spain). Finally, authors are grateful to FEDER (European Community) for financial support, Grants UNHU13-1E-1611 and UNHU15-CE-3140.

References

98/83/EC DW (1998) Council Directive 98/83/CE of 3 November on the Quality of Water Intended for Human Consumption.

Amy G, Chadik P, Chowdhury Z (1987) Developing models for predicting trihalomethanes formatin potential and kinetics Journal / American Water Works Association 79:89-97

Amy G, Siddiqui M, Ozekin K, Zhu H, Wang C (1998) Empirically based models for predicting chlorination and ozonation by-produts: Trihalomethanes, Haloacetic Acids, Chloral Hydrat, and Bromate US-EPA Raport number CX 819579

Ates N, Kitis M, Yetis U (2007) Formation of chlorination by-products in waters with low SUVA-correlations with SUVA and differential UV spectroscopy Water Research 41:4139-4148 doi:10.1016/j.watres.2007.05.042

Badawy MI, Gad-Allah TA, Ali MEM, Yoon Y (2012) Minimization of the formation of disinfection by-products Chemosphere 89:235-240 doi:http://dx.doi.org/10.1016/j.chemosphere.2012.04.025

Basu M, Gupta SK, Singh G, Mukhopadhyay U (2011) Multi-route risk assessment from trihalomethanes in drinking water supplies Environ Monit Assess 178:121-134 doi:10.1007/s10661-010-1677-z

Bond T, Huang J, Graham NJD, Templeton MR (2014) Examining the interrelationship between DOC, bromide and chlorine dose on DBP formation in drinking water - A case study Science of the Total Environment 470-471:469-479 doi:10.1016/j.scitotenv.2013.09.106

Bougeard CMM, Goslan EH, Jefferson B, Parsons SA (2010) Comparison of the disinfection by-product formation potential of treated waters exposed to chlorine and monochloramine Water Research 44:729-740 doi:10.1016/j.watres.2009.10.008

Brown D, Bridgeman J, West J (2011) Predicting chlorine decay and THM formation in water supply systems Rev Environ Sci Biotechnol 10:79-99 doi:10.1007/s11157-011-9229-8

Cantor KP et al. (2010) Polymorphisms in GSTT1, GSTZ1, and CYP2E1, Disinfection By-products, and Risk of Bladder Cancer in Spain Environmental Health Perspectives 118:1545-1550 doi:10.1289/ehp.1002206

Chen B, Westerhoff P (2010) Predicting disinfection by-product formation potential in water Water Research 44:3755-3762 doi:http://dx.doi.org/10.1016/j.watres.2010.04.009

Chowdhury S, Champagne P, McLellan PJ (2009) Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review Science of the Total Environment 407:4189-4206 doi:10.1016/j.scitotenv.2009.04.006

Chu W, Ding S, Bond T, Gao N, Yin D, Xu B, Cao Z (2016a) Zero valent iron produces dichloroacetamide from chloramphenicol antibiotics in the absence of chlorine and chloramines Water Research 104:254-261 doi:http://doi.org/10.1016/j.watres.2016.08.021

Chu W, Li X, Bond T, Gao N, Bin X, Wang Q, Ding S (2016b) Copper increases reductive dehalogenation of haloacetamides by zero-valent iron in drinking water: Reduction efficiency and integrated toxicity risk Water Research 107:141-150 doi:http://doi.org/10.1016/j.watres.2016.10.047

D70/2009 de 31 de marzo, por el que se aprueba el Reglamento de Vigilancia Sanitaria y Calidad del Agua de Consumo Humano de Andalucía. BOJA num 73 de 17/04/2009

Domínguez-Tello A, Arias-Borrego A, García-Barrera T, Gómez-Ariza JL (2015) Seasonal and spatial evolution of trihalomethanes in a drinking water distribution system according to the treatment process Environ Monit Assess 187 doi:10.1007/s10661-015-4885-8

Elshorbagy WA (2000) Kinetics of THM species in finished drinking water Journal of Water Resources Planning and Management 126:21-28 doi:10.1061/(ASCE)0733-9496(2000)126:1(21)

Engerholm BA, Amy GL (1983) A predictive model for chloroform formation from humic acid Journal / American Water Works Association 75:418-423

Fokmare AK, Musaddiq M (2001) Comparative studies of physico-chemical and bacteriological quality of surface and ground water at akola (Maharashtra) Pollution Research 20:651-655

Galapate RP, Baes AU, Okada M (2001) Transformation of dissolved organic matter during ozonation: Effects on trihalomethane formation potential Water Research 35:2201-2206 doi:10.1016/S0043-1354(00)00489-9

Garcia-Villanova RJ, Oliveira Dantas Leite MV, Hernández Hierro JM, de Castro Alfageme S, García Hernández C (2010) Occurrence of bromate, chlorite and chlorate in drinking waters disinfected with hypochlorite reagents. Tracing their origins Science of the Total Environment 408:2616-2620 doi:10.1016/j.scitotenv.2010.03.011

Ged EC, Chadik PA, Boyer TH (2015) Predictive capability of chlorination disinfection byproducts models Journal of Environmental Management 149:253-262 doi:http://doi.org/10.1016/j.jenvman.2014.10.014

Golfinopoulos SK, Arhonditsis GB (2002) Multiple regression models: A methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics Chemosphere 47:1007-1018 doi:http://dx.doi.org/10.1016/S0045-6535(02)00058-9

Grazuleviciene R, Kapustinskiene V, Vencloviene J, Buinauskiene J, Nieuwenhuijsen MJ (2013) Risk of congenital anomalies in relation to the uptake of trihalomethane from drinking water during pregnancy Occupational and Environmental Medicine 70:274-282 doi:10.1136/oemed-2012-101093

Hamidin N, Yu QJ, Connell DW (2008) Human health risk assessment of chlorinated disinfection by-products in drinking water using a probabilistic approach Water Research 42:3263-3274 doi:10.1016/j.watres.2008.02.029

Hoffman CS et al. (2008) Drinking Water Disinfection By-Product Exposure and Duration of Gestation Epidemiology 19:738-746 doi:10.1097/EDE.0b013e3181812beb

Hong H, Xiong Y, Ruan M, Liao F, Lin H, Liang Y (2013) Factors affecting THMs, HAAs and HNMs formation of Jin Lan Reservoir water exposed to chlorine and monochloramine Science of The Total Environment 444:196-204 doi:http://dx.doi.org/10.1016/j.scitotenv.2012.11.086

Hong HC, Liang Y, Han BP, Mazumder A, Wong MH (2007) Modeling of trihalomethane (THM) formation via chlorination of the water from Dongjiang River (source water for Hong Kong's drinking water) Science of The Total Environment 385:48-54 doi:http://dx.doi.org/10.1016/j.scitotenv.2007.07.031

Hua G, Reckhow DA, Abusallout I (2015) Correlation between SUVA and DBP formation during chlorination and chloramination of NOM fractions from different sources Chemosphere 130:82-89 doi:http://dx.doi.org/10.1016/j.chemosphere.2015.03.039

Karapinar N, Uyak V, Soylu S, Topal T (2014) Seasonal variations of NOM composition and their reactivity in a low humic water Environmental Progress & Sustainable Energy 33:962-971 doi:10.1002/ep.11878

Kitis M, Karanfil T, Wigton A, Kilduff JE (2002) Probing reactivity of dissolved organic matter for disinfection by-product formation using XAD-8 resin adsorption and ultrafiltration fractionation Water Research 36:3834-3848 doi:http://dx.doi.org/10.1016/S0043-1354(02)00094-5

Kulkarni P, Chellam S (2010) Disinfection by-product formation following chlorination of drinking water: Artificial neural network models and changes in speciation with treatment Science of The Total Environment 408:4202-4210 doi:http://dx.doi.org/10.1016/j.scitotenv.2010.05.040

Kumar S, Forand S, Babcock G, Hwang SA (2014) Total trihalomethanes in public drinking water supply and birth outcomes: A cross-sectional study Maternal and Child Health Journal 18:996-1006 doi:10.1007/s10995-013-1328-4

Kumari M, Gupta SK (2015) Modeling of trihalomethanes (THMs) in drinking water supplies: a case study of eastern part of India Environ Sci Pollut Res 22:12615-12623 doi:10.1007/s11356-015-4553-0

Levallois P, Gingras S, Marcoux S, Legay C, Catto C, Rodriguez M, Tardif R (2012) Maternal exposure to drinking-water chlorination by-products and small-for-gestational-age neonates Epidemiology 23:267-276 doi:10.1097/EDE.0b013e3182468569

Li X, Zhao H-b (2006) Development of a model for predicting trihalomethanes propagation in water distribution systems Chemosphere 62:1028-1032 doi:http://dx.doi.org/10.1016/j.chemosphere.2005.02.002

Lu O, Krasner SW, Liang S (2011) Modeling approach to treatability analyses of an existing treatment plant Journal - American Water Works Association 103:103-117

Matilainen A, Gjessing ET, Lahtinen T, Hed L, Bhatnagar A, Sillanpää M (2011) An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment Chemosphere 83:1431-1442 doi:http://dx.doi.org/10.1016/j.chemosphere.2011.01.018

Mayer BK, Daugherty E, Abbaszadegan M (2015) Evaluation of the relationship between bulk organic precursors and disinfection byproduct formation for advanced oxidation processes Chemosphere 121:39-46 doi:http://dx.doi.org/10.1016/j.chemosphere.2014.10.070

Melnick RL, Dunnick JK, Sandler DP, Elwell MR, Barrett JC (1994) Trihalomethanes and Other Environmental Factors That Contribute to Colorectal Cancer Environmental Health Perspectives 102:586-588

Mishra BK, Gupta SK, Sinha A (2014) Human health risk analysis from disinfection by-products (DBPs) in drinking and bathing water of some Indian cities Journal of Environmental Health Science and Engineering 12:73-73 doi:10.1186/2052-336X-12-73

Mouly D et al. (2010) Variations in trihalomethane levels in three French water distribution systems and the development of a predictive model Water Research 44:5168-5179 doi:http://dx.doi.org/10.1016/j.watres.2010.06.028

Nieuwenhuijsen MJ et al. (2008) Chlorination Disinfection By-Products and Risk of Congenital Anomalies in England and Wales Environmental Health Perspectives 116:216-222 doi:10.1289/ehp.10636

Reckhow DA, Singer PC, Malcolm RL (1990) Chlorination of humic materials: byproduct formation and chemical interpretations Environmental Science & Technology 24:1655-1664 doi:10.1021/es00081a005

Richardson SD (2003) Disinfection by-products and other emerging contaminants in drinking water TrAC Trends in Analytical Chemistry 22:666-684 doi:http://dx.doi.org/10.1016/S0165-9936(03)01003-3

Richardson SD, Plewa MJ, Wagner ED, Schoeny R, DeMarini DM (2007) Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research Mutation Research - Reviews in Mutation Research 636:178-242 doi:10.1016/j.mrrev.2007.09.001

Sadrnourmohamadi M, Gorczyca B (2015) Effects of ozone as a stand-alone and coagulation-aid treatment on the reduction of trihalomethanes precursors from high DOC and hardness water Water Research 73:171-180 doi:http://dx.doi.org/10.1016/j.watres.2015.01.023

Salas LA et al. (2013) Biological and Statistical Approaches for Modeling Exposure to Specific Trihalomethanes and Bladder Cancer Risk American Journal of Epidemiology 178:652-660 doi:10.1093/aje/kwt009

Scheili A, Rodriguez MJ, Sadiq R (2015) Seasonal and spatial variations of source and drinking water quality in small municipal systems of two Canadian regions Science of The Total Environment 508:514-524 doi:http://dx.doi.org/10.1016/j.scitotenv.2014.11.069

Siddique A, Saied S, Mumtaz M, Hussain MM, Khwaja HA (2015) Multipathways human health risk assessment of trihalomethane exposure through drinking water Ecotoxicology and Environmental Safety 116:129-136 doi:http://dx.doi.org/10.1016/j.ecoenv.2015.03.011

Summerhayes RJ et al. (2011) Spatio-temporal variation in trihalomethanes in New South Wales Water Research 45:5715-5726 doi:http://dx.doi.org/10.1016/j.watres.2011.08.045

USEPA (2001) National Primary Drinking Water Regulations. Washington D.C.

Uyak V, Ozdemir K, Toroz I (2007) Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs Science of The Total Environment 378:269-280 doi:http://dx.doi.org/10.1016/j.scitotenv.2007.02.041

Watson K, Farré MJ, Birt J, McGree J, Knight N (2015) Predictive models for water sources with high susceptibility for bromine-containing disinfection by-product formation: implications for water treatment Environ Sci Pollut Res 22:1963-1978 doi:10.1007/s11356-014-3408-4

Zimoch I, Szymura E, Moraczewska-Majkut K (2015) Changes of trihalomethanes (THMs) concentration in water distribution system Desalination and Water Treatment doi:10.1080/19443994.2015.1030115

Table 1. Description of water distribution systems

Water utility	Water (m3/d)	Raw Water Source	Treatment processes	Distribution system.	
Alj WTP	47,500	Surface Chanza Reservoir	Pre-$KMnO_4$,CFS, SF,GAC,NaClO 15Jun-15Sep: Pre-O_3,CFS,SF,CAG, NaClO	San Bartolomé DS1	D: 3.8 mg L^{-1} Cl t_{DS}: 20.4 h
Lep WTP	86,400	Surface Chanza Reservoir	Pre-$KMnO_4$,CFS, SF,GAC,NaClO 15Jun-15Sep:Pre-O_3, CFS,SF,CAG,NaClO	Ayamonte DS2	D: 5.1 mg L^{-1} Cl t_{DS}: 36.97 h
Rti WTP	8,640	Surface Jarrama Reserv.	Pre-$KMnO_4$, CFS, SF, GAC, NaClO	Fuente Corcha DS3	D: 2.9 mg L^{-1} Cl t_{DS}: 38.7 h
Lpa WTP	17,280	Surface CorumbelReserv	Pre-O_3, CFS, SF, GAC, NaClO	Niebla DS4	D: 4.7 mg L^{-1} Cl t_{DS}: 62.9 h

Pre-$KMnO_4$: Preoxidation with potassium permanganate; Pre-O_3 : Preoxidation with ozone. CFS: Coagulation + Flocculation + Sedimentation; SF: sand filters; GAC: Filtration with Granular Activate Carbon. NaClO: Disinfection with sodium hypochlorite.
DS1, DS2, DS3 and DS4: Sample points of San Bartolome, Ayamonte, Fuente Corcha and Niebla reservoirs
D: accumulate dose of chlorine (chlorine dose in WTP + chlorine dose in rechlorinations).
t_{DS}: contact time from finished water to sample point of the distribution system.
Pre-$KMnO_4$: preoxidation with potassium permanganate; CFS: Coagulation, Flocculation, Sedimentation; SF: Sand Filters; GAC: Granular Activate Carbon Filters; NaClO: Disinfection with sodium hypochlorite.

Table 2. Occurrence TTHMs formation. Seasonal variation

	Alj WDS		Lep WDS		Rti WDS		Lpa WDS	
	Alj WTP	DS1	LepWTP	DS2	Rti WTP	DS3	Lpa WTP	DS4
N	52	45	45	45	40	33	47	40
Total	34.6(12.0)	38.2(7.5)	35.1(8.4)	44.1(11.6)	34.0(12.9)	45.6(16.4)	35.3(11.5)	39.9(13.8)
Aut/Win	28.2(7.7)	34.1(4.9)	29.8(5.4)	42.6(9.1)	28.4(6.4)	39.0(19.4)	33.3(9.9)	36.5(14.1)
Spring	36.3(5.7)	39.2(2.7)	36.3(5.9)	43.9(12.6)	32.4(9.1)	46.5(15.7)	35.2(12.4)	39.4(13.6)
Summer	39.9(14.4)	41.8(9.3)	39.9(9.9)	46.3(14.9)	39.9(9.9)	50.3(14.3)	38.0(13.1)	44.2(13.4)

Average concentration of TTHM (µg L^{-1})
DS1: San Bartolomé Reservoir; DS2: Ayamonte Reservoir; DS3: Fuente la Corcha Reservoir; DS4: Niebla Reservoir.
Standard deviation in parentheses.

Table 3. Pearson correlation matrix. Explanatory variables

Stage 1		$TTHM_{Ef}$			$TTHM_{Ef}$	Stage 2		$TTHM_{DS}$			$TTHM_{DS}$
T	*r*	0.162	SUVA	*r*	0.094	$TTHM_{Ef}$	*r*	0.987	R_{DS}	*r*	0.749
	p	0.520		*p*	0.711		*p*	0.000		*p*	0.000
pH_{Ef}	*r*	0.562	d / COD	*r*	0.414	T_{DS}	*r*	0.097			
	p	0.015		*p*	0.050		*p*	0.701			
COD	*r*	0.102	δ	*r*	-0.861	pH_{DS}	*r*	0.687			
	p	0.687		*p*	0.000		*p*	0.002			
UV_{254}	*r*	0.140	$\delta \times T_{Ef}$	*r*	-0.713	t_{DS}	*r*	0.020			
	p	0.576		*p*	0.001		*p*	0.965			
t_{Ef}	*r*	0.951	d - Cl_2	*r*	0.748	D	*r*	0.881			
	p	0.000		*p*	0.000		*p*	0.000			
d	*r*	0.748	$UV_{254} \times \delta \times T_{Ef}$	*r*	-0.611						
	p	0.000		*p*	0.007						
Br^-	*r*	0.757	R_{Ef}	*r*	0.965						
	p	0.000		*p*	0.004						
			d-Cl_2)tT/UV_{254}	*r*	0.925						
				p	0.000						

Table 4. Model options

	Var	R^2	R^2 adj	Cp Mallows	SE	pH_{Ef}	d	t_{Ef}	T_{Ef}	Br^-	UV_{254}	R_{Ef}	δ	δT
A	2	0.935	0.927	5.3	8.70	x						x		
B	3	0.948	0.936	4.0	8.67	x				x		x		
C	5	0.962	0.945	5.2	9.10	x	x	x	x		x			
D	6	0.962	0.942	7	9.43	x	x	x	x	x	x			
E	3	0.878	0.852	4.0	14.9	x			x				x	
G	3	0.789	0.744	4.0	19.7	x				x				x

R^2, R^2 adj: correlation coefficient model and adjusted; SE: standard error; t_{Ef}: contact time; T_{Ef}: temperature; d: chlorine dose; $R_{Ef} = d \times t_{Ef} \times T_{Ef} \times UV_{254}$; $\delta = (d - Cl_2) / t_{Ef}$

Table 5. Statistical evaluation of TTHM models

Stage 1. Predictive Model of TTHM formation in WTP	N	F-test	T-test	R^2	R^2adj	SE	AV	Fvalue	Durbin Watson
$TTHM_{Ef} = 98.5 - 10.3\ pH_{Ef} + 6.90\ R_{Ef}$	33	0.89	0.98	0.935	0.927	8.70	13.9		
$TTHM_{Ef} = 1230.26\ (pH_{Ef})^{-1.84}\ (R_{Ef})^{0.470}$	33	0.62	0.88	0.894	0.880	9.48	15.5		
$\mathbf{TTHM_{Ef} = 165 - 21.3\ pH_{Ef} + 0.232\ Br^{-} + 5.84\ \boldsymbol{R}_{Ef}}$	33	0.91	0.99	0.948	0.936	8.67	13.6	126.9	1.74
$TTHM_{Ef} = 12022.64\ (pH_{Ef})^{-3.6}\ (Br^{-})^{0.299}\ (R_{Ef})^{0.389}$	33	0.67	0.89	0.914	0.896	9.08	14.3		
Stage 2. Predictive Model of TTHM formation in DS									
$\mathbf{TTHM_{DS} = 14.9 + 1.01\ THM_{Ef} + 0.20\ pH_{DS} - 0.104\ \boldsymbol{R}_{DS}}$	33	0.87	0,90	0.976	0.971	6.08	9.9	122.7	1.62
$TTHM_{DS} = 4.1115\ (THM_{Ef})^{0.914}\ (pH_{DS})^{-0.13}\ (R_{DS})^{-0.147}$	33	0.86	0.87	0.948	0.936	6.47	10.6		

$R_{EF} = (d \times t \times T \times UV_{254})$; $R_{DS} = (D \times T_{DS} \times t_{DS} \times UV_{254})$

Table 6. Model validation and application

	N	R^2	R^2adjust	SE	AV	AVmax	t_{value}	$t_{critical}$
Two-Stage Model VALIDATION								
Alj WTP	35	0.912	0.91	2.57	7.38	17.0	0.72	2.032
Alj DS	35	0.874	0.87	3.00	7.22	12.1	0.48	2.032
Two-Stage Model APPLICATION								
Lep WTP	27	0.965	0.963	2.98	7.68	15.8	0.53	2.056
Lep DS	27	0.884	0.879	3.67	7.05	14.0	0.65	2.056
Rti WTP	28	0.964	0.963	3.17	6.67	18.3	0.78	2.052
Rti DS	28	0.954	0.952	4.02	7.27	16.0	0.57	2.052
Lpa WTP	38	0.918	0.916	3.76	8.91	26.6	0.49	2.021
Lpa DS	38	0.825	0.820	4.94	9.80	26.6	0.90	2.042
GLOBAL WTP	128	0.940	0.939	3.19	7.79	26.6	0.49	1.980
GLOBAL DS	123	0.874	0.873	4.05	8.30	26.6	0.56	1.980

Table 7. Predictive models TTHMs evaluated

Nº	Model (TTHM)	Treatment proccess.	Ref
M1	$TTHM = 10^{0.518} (COD)^{0.801} (Cl_2)^{0.261} (Br^-)^{0.223} (t)^{0.264}$		(Amy et al. 1998)
M2	$TTHM = 10^{-1.385} (COD)^{1.098} (Cl_2)^{0.152} (Br^-)^{0.068} (T)^{0.609} (pH)^{1.601} (t)^{0.263}$	*(quality variables: raw water)*	(Amy et al. 1998)
M3	$TTHM = 0.044 (COD)^{1.030} (t)^{0.262} (pH)^{1.149} (Cl_2)^{0.277} (T)^{0.968}$		(Rodriguez et al. 2000)
M4	$TTHM = 16.9 + 16.0 (TOC) + 3.319(Cl_2) – 1.135(T) + 1.139(t)$	Charlesbourg	(Sérodes et al. 2003)
M5	$TTHM = 21.2 + 2.447(Cl_2) + 0.499(t)$	Sainte Foy	(Sérodes et al. 2003)
M6	$TTHM = 0.012 (COD\ UV_{254})^{0.47} (D)^{0.48}(T)^{1.10}(pH)^{2.38}(t)^{0.35}$		(Sohn et al. 2004)
M7	$TTHM = 11.967 (TOC)^{0.398} (T)^{0.158} (Cl_2)^{0.702}$	Istambul. Buyukcekmece WTP: Aer., pre-Cl_2, CFS-SF Dis.	(Toroz and Uyak 2005)
M8	$TTHM = 10^{-0.038} (Cl_2)^{0.654} (pH)^{1.322} (t)^{0.174} (SUVA)^{0.712}$	Istambul Turkey: Terkos WTP: Conv.Proc.; pre-Cl_2 Buyukcekmece WTP: Conv.Proc.; pre-Cl_2; Omerli WTP: Conv.Proc; pre-O_3.	(Uyak et al. 2007)
M9	$TTHM = 10^{-1.375} (t)^{0.258} (D/DOC)^{0.194} (pH)^{1.695} (T)^{0.507} (Br^-)^{0.218}$	Chlorination of the water from Dongjiang River (source water for Hong Kong's drinking water)	(Hong et al. 2007)
M10	$TTHM^2 = –471.11 + 0.48(t)^2 + 1856.07(Br^–)^2 + 404.38(D)^2$		(Semerjian et al. 2009)
M11	$TTHM = -51.408 + 8.449 (COD) + 13.529(Cl_2) + 2.997(pH) + 0.803(T) + 0.504(t) + 0.141(d-4.447)(T-15.03)$	Ontario Supply system.	(Chowdhury et al. 2010)
M12	$TTHM= e^{3.875} (TOC_{Ef}\ UV_{Ef})^{0.328} (Br+1)^{3.334} (Cl_2)^{0.151}\ t^{0.314}$	Mills WTP. AdjpH,preO3,CFS,SF,D. Disinfection: Cl_2 NH_3	(Lu et al. 2011)
M13	$TTHM = 1.58 (UV_{254}\ TOC)^{0.38} (Cl_2)^{1.14} (t)^{0.6}\ T^{0.5} (ph-2.6)^{0.96}\ Br^{0.6}$	Siouf Water Treatment Plant (WTP); Alejandría. Conventional proccess.	(Abdullah and Hussona 2013)
M14	$TTTHM = 0.0072\ pH^{2.60}\ T^{0.396}\ t^{0.475}\ TOC^{1.397}$	New York. Filtration Avoidance Determination garanted by the New York State Department of Health and the USEPA	(Mukundan and Van Dreason 2014)
M15	$TTHM = 33.436(pH)^{0.062} (T)^{0.069} (RCl_2)^{-0.048} (t/60)^{0.018} (TOC)^{0.079} (UV_{254})^{0.045}$	WTP: Maithon, Dhanbad, Raniganj, Barrackpore, and Ranchi. Conv.Proc;pre-Cl_2	(Kumari and Gupta 2015)
M16	$TTHM = 4,527(t*60)^{0,127} (Cl_2)^{0,595} (TOC)^{0,596} (Br/1000)^{0,103} (pH)^{0,66}$	Zai water treatment plant supplies water to west Amman,	(Semerjian et al. 2009)
M17	$TTHM = 10^{-6.777} (TOC)^{1.171} (pH)^{4.469} (Cl_2)^{1.765}$	Tampin WTP: Conv.Proc.; CFS SF Dis	(Abdullah et al. 2003)

Table 8. Predictive capability of models according to treatment processes

		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	*o.m.*
TP1	SE	20.54	11.09	7.85	22.03	69.63	10.15	7.25	29.59	77.38	2.69	7.21	29.88	87.88	17.83	77.46	33.50	23.52	*7.97*
	AV	11.98	7.63	4.43	12.91	49.24	16.83	4.04	20.62	54.60	1.63	4.28	16.45	59.17	10.38	54.77	22.80	15.44	*4.99*
	R^2	0.44	0.45	0.82	0.80	0.46	0.79	0.89	0.369	0.558	0.98	0.98	0.33	0.13	0.44	0.23	0.49	0.83	*0.82*
	Min (cal)	95.2	101.1	100.3	90.6	47.6	90.9	103.7	85.2	35.8	108.8	103.4	122.3	164.1	115.1	40.0	75.0	78.2	*101.8*
	Max(cal)	117.5	124.4	130.9	112.6	49.1	122.8	118.2	91.4	51.0	125.2	118.6	165.2	241.4	143.4	41.4	109.5	126.0	*123.3*
	Differ. Range %	-27 to 5	-14 to 11	-3 to 15	-30 to 5	-63 to -53	-10 to10	-11 to 0	-30 to 18	-70 to 61	-2 to -5	-9 to 0	-5 to 30	33 to 116	-8 to 28	-68 to 61	-40 to-27	-24 to-3	*-5 to 10 %*
TP2	SE	22.00	29.89	37.98	38.75	7.05	25.10	36.15	14.50	14.73	42.65	36.62	40.57	48.81	37.05	11.22	10.33	8.09	*6.37*
	AV	44.52	54.51	65.27	78.24	11.26	20.52	74.64	27.91	29.66	88.81	75.67	83.20	86.26	70.44	19.30	18.62	15.69	*12.13*
	R^2	0.40	0.48	0.59	0.50	0.62	0.33	0.30	0.40	0.38	0.38	0.49	0.10	0.58	0.20	0.55	0.67	0.49	*0.76*
	Min (cal)	58.5	54	50.7	72.7	43.2	36.0	74.0	55.1	88.3	89.2	77.05	72.4	58.2	60.5	37.4	30.4	35.0	*41.5*
	Max(cal)	73.6	92.7	107.6	97.1	46.9	91.0	92.3	67.6	96.2	96.2	92.6	95.3	127.4	105.1	40.0	64.0	52.8	*58.4*
	Differ.. Range %	23 to 79	13 to 88 %	12 to 96 %	32 to 139 %	-15 to 12 %	-18 to 66	55 to109	15 to 66 %	-43 to-33	62 to 119	62 to 109	52 to 134	22 to 132	33 to 132	-28 to -3	-36 to 17	-27 to -4	*-3 to 35 %*
TP3	SE	26.77	45.98	54.0	54.43	10.11	35.23	30.84	17.04	22.09	18.75	30.72	60.81	11.82	72.871	10.01	15.06	20.63	*11.51*
	AV	57.24	102.08	120.65	122.83	17.84	42.83	69.25	35.91	46.61	42.00	68.63	133.19	19.35	159.78	17.38	27.58	40.15	*25.06*
	R^2	0.73	0.54	0.17	0.71	0.59	0.31	0.10	0.59	0.51	0.75	0.21	0.63	0.41	0.43	0.87	0.30	0.47	*0.77*
	Min (cal)	55.5	70.4	83.9	77.5	39.8	18.1	69.9	48.8	59.3	16.2	67.4	76.6	27.2	86.1	40.4	37.8	22.2	*28.4*
	Max(cal)	80.1	104.2	111.5	120.8	43.1	29.1	82.8	65.8	70.6	31.3	79.9	129.4)	47.8	149.9	43.5	67.3	31.2	*46.4*
	Differ. Range %	-6 to 100	37 to 162 %	62 to 194 %	32 to 216 %	-24 to 11	-53 to 24	29 to 96 %	-9 to 65 %	-53 to-33	-69 to-22	27 to 101	30 to 224	-42 to18	101 to 275	-31 to14	-3 to 77	-49 to-29	*-7 to 29 %*
TOT	SE	23.26	32.30	38.39	41.52	40.82	20.36	27.75	21.42	47.23	26.95	27.21	45.59	58,44	48.31	45,56	22.03	18,66	*8,88*
	AV	44.17	67.95	80.93	86.37	20.41	30.03	61.86	30.01	42.08	56.28	62.11	98.29	58.70	102.08	24.70	23,81	27,36	*16.63*
	R^2	0.68	0.46	0.38	0.12	0.59	0.88	0.41	0.80	0.52	0.60	0.86	0.42	0.83	0.13	0.40	0.65	0.86	*0.94*

TTHM measured (Min – Max): TP1: (103.29 - 125.19); TP2: (40.7 - 55); TP3: (26.9 - 58.8).

TP1:Conventional treatment process: Pre-Cl_2, CFS, SF, Disinfection Cl_2.

TP2:Advanced treatment process: Pre-O_3, CFS, SF, GAC, Disinfection Cl_2

TP3: Conventional treatment process:Pre-$KMnO_4$,CFS, SF, GAC Disinfection Cl_2.

List of Figure Captions

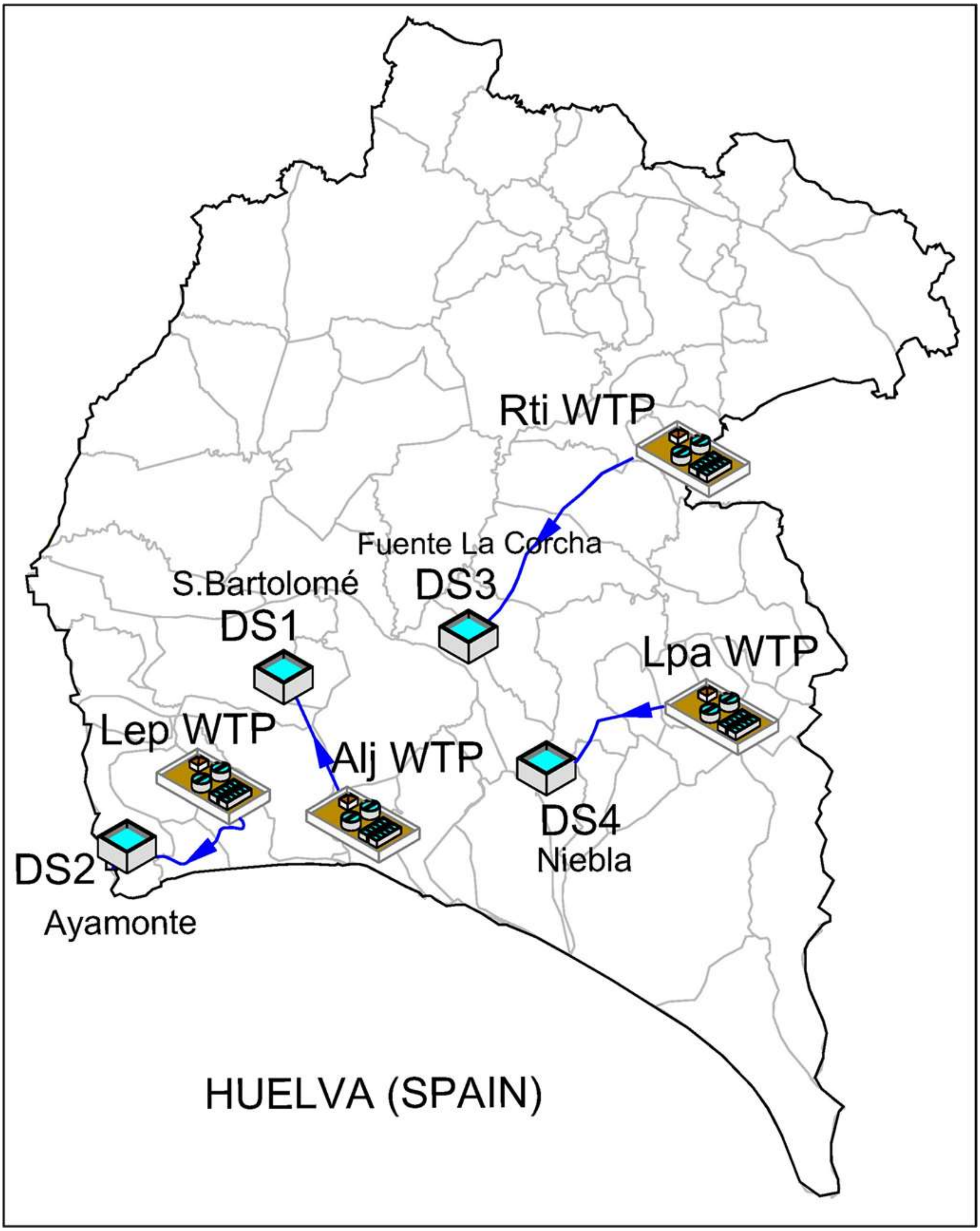

Figure 1. Study zone

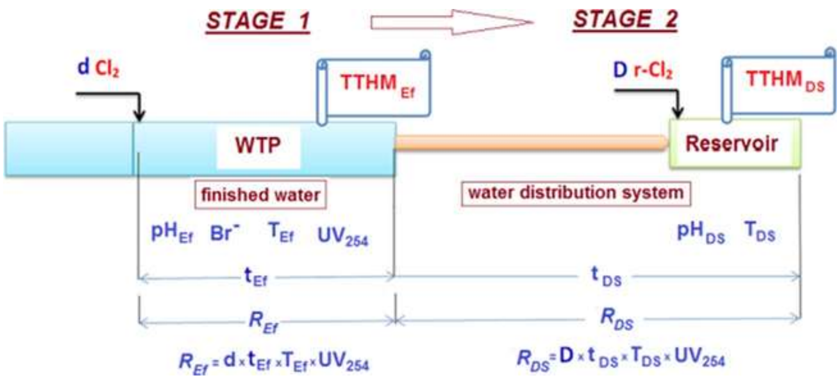

Figure 2. Model scheme

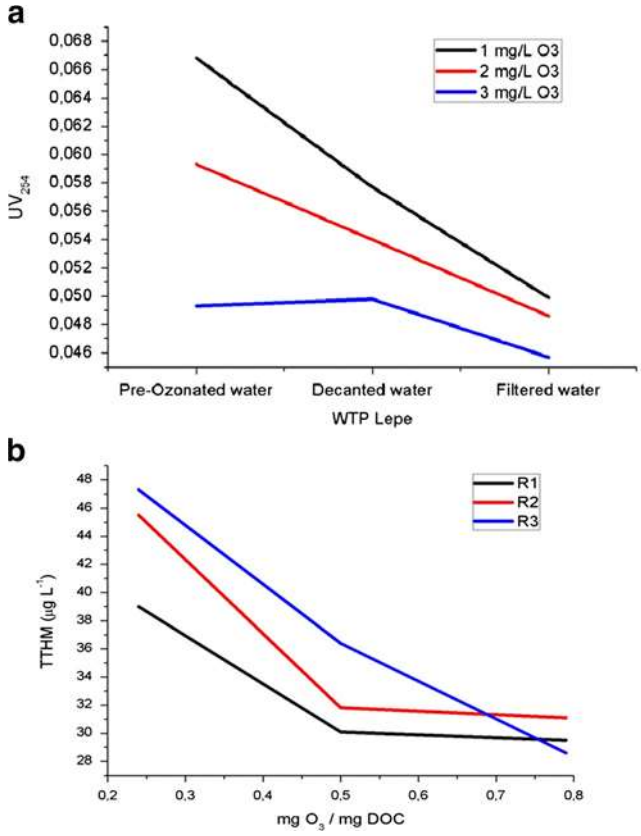

Fig. 3 Effect of ozonation

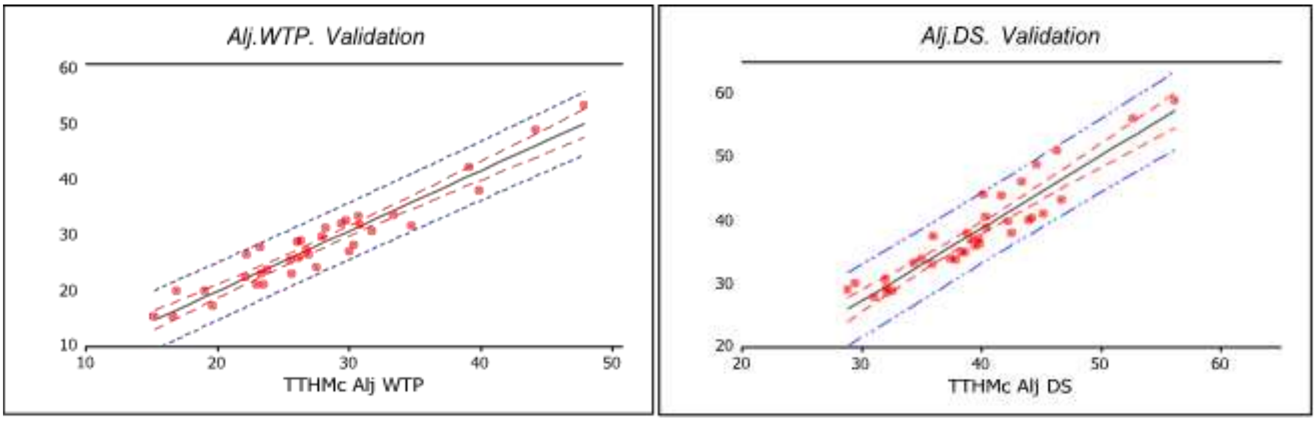

Fig. 4 Model validation. Aljaraque WTP and DS

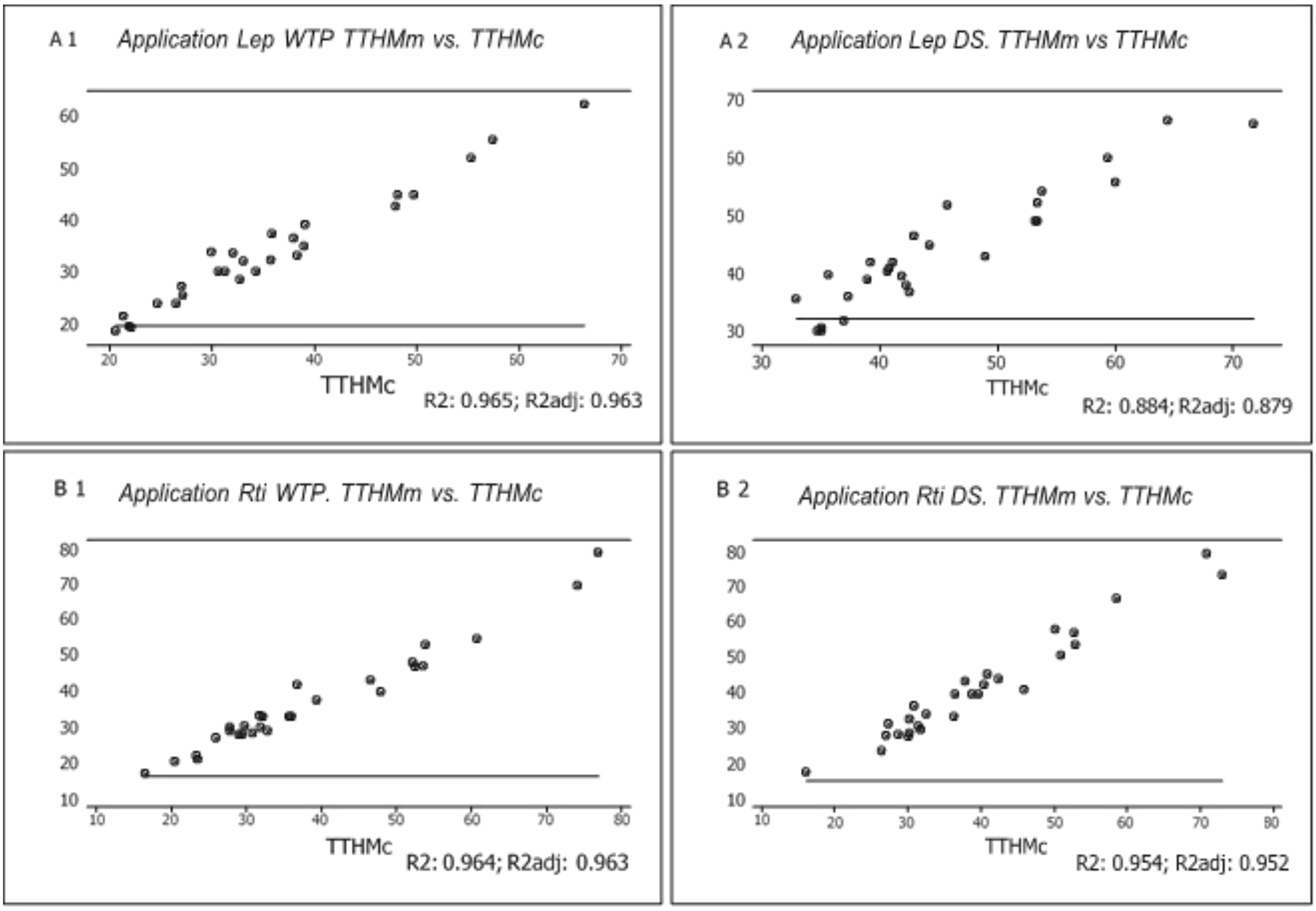

Figure 5. Model application in three WDSs

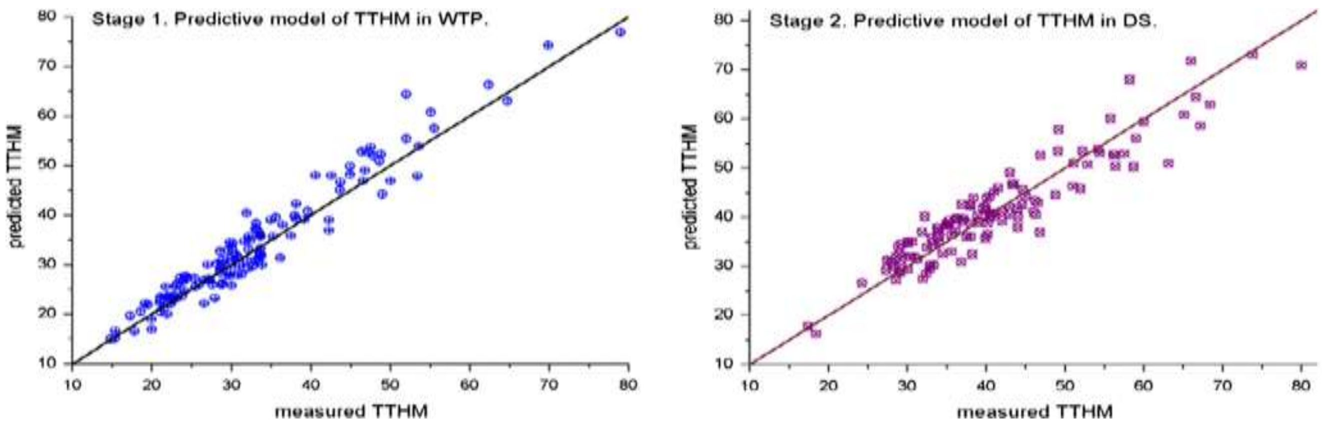

Fig. 6 Global predictive model. TTHM