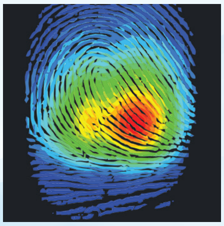
HORIBA
Scientific

Fluorescence

Trihalomethane Speciation Modeling in Drinking Water Using the A-TEEM™ Spectroscopic Technique and Multivariate Data Analysis



Application
Note
Environmental
Sciences
FL-2022-01-19

L. Chen*, A. Gilmore*, M. Theiler, D. Brogle****
*20 Knightsbridge Rd, HORIBA Scientific, Piscataway, NJ 08854
**1500 Ronson Rd, Middlesex Water Company, Iselin, NJ 08830

Abstract

This application note describes a rapid, sensitive non-destructive disinfection byproducts (DBP) prediction method using patented simultaneous Absorbance-Transmittance and fluorescence Excitation-Emission Matrix (A-TEEM) spectroscopy. The A-TEEM method provides Natural Organic Matter (NOM) composition and concentration information directly relating to key DBP precursors. In direct reference to a conventional purge and trap gas chromatography (GC) instrument method, over a period of approximately 188 days, we calibrated and validated a Multiple Linear Regression (MLR) model of three Trihalomethanes (THMs) simultaneously including chloroform (CHFM), bromodichloromethane (DCBM) and chlorodibromomethane (DBCM). The best correlation model yielded an adjusted R2 = 0.980 with a slope = 0.956 and close correspondence to each THM species concentration. All data processing can be made in real time using an automated chemometric engine to analyze and report data to an HTML-based dashboard. This A-TEEM method thus serves as an effective instantaneous THM formation and speciation tool for drinking water treatment systems with: 1) rapid data collection time (2-3 minutes per sample), 2) high sensitivity, 3) less expense than conventional chromatography-based analysis, and 4) compatibility with automated sampling.

Keywords: A-TEEM; DBPs precursors; THM speciation

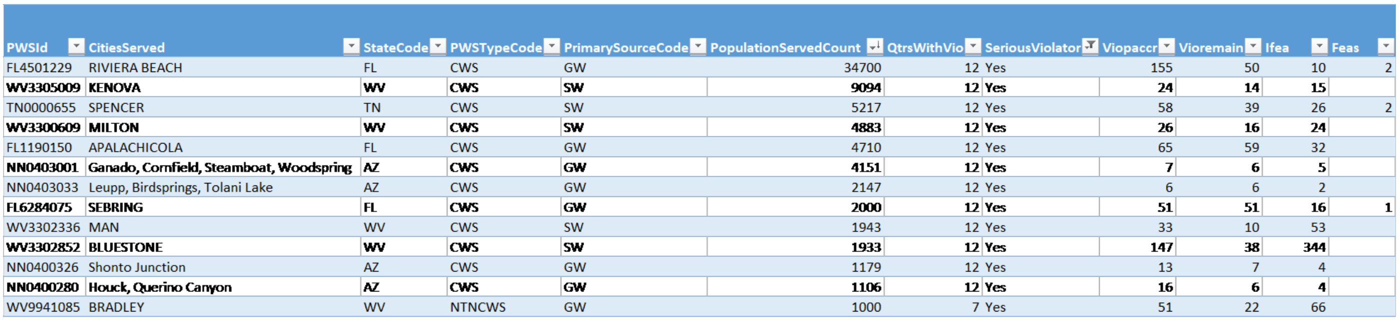
Introduction

DBPs including THMs and HAAs are regulated by many global environmental agencies for drinking water treatment because they are potential carcinogens. Because many NOM compounds present in drinking water sources react with halogenated disinfectants over time and are the precursors to DBPs, controlling NOM is critical for DBP control. Even properly designed and well-operated surface water filtration plants are not often able to remove all the source water NOM. As a result, chlorine or other halogenated disinfectants are used to kill the microorganisms. In most drinking water treatment schemes, the detection of DBPs is problematic because it is only focused in the distribution system. The main problems are that: (1) by the time DBPs form they are difficult to remove and treatments to prevent them can only be implemented retroactively, and (2) conventional methodology for DBP measurement is difficult, expensive, time consuming and difficult to apply in real time. Hence, a key goal of water treatment is to predict and measure DBPs in real time so treatment and prevention can be implemented in a forward manner.

DBPs are contaminants of concern nationwide. Nationwide (at the time this was written) there are 37 facilities violations reported for Stage 2 and 154 facilities for both Stage 1 and Stage 2 of the Disinfectants and Disinfection Byproducts Rules (D/DBPR) [US Environmental Protection Agency website]. Table 1 shows a sample of the search results by filtering only the serious violators serving populations > 1000.

The D/DBPR involves a monitoring program that monitors indicators of DBPs, i.e., precursors to byproducts in the source water. Because the DBP precursors are difficult and expensive to measure directly, most commonly the Total Organic Carbon (TOC) and/or Specific Ultraviolet Absorbance (SUVA, indicating organic matter aromaticity and reactivity) are often substituted as indicators of DBP precursors. The TOC analysis measures the overall organic carbon content in the water, and not just the specific organic carbon components that can form DBPs. In contrast, the A-TEEM and parallel factor analysis (PARAFAC) resolves information on the DBP precursor fractions.

The main purpose of the conventional coagulation and filtration treatment processes is to reduce the overall DOC, including the highly aromatic DBP precursor components. Importantly, the highly aromatic and hydrophobic DBP precursors that increase the THM formation potential can be identified and quantified using the A-TEEM method [Gilmore, A.M. 2017]. In fact, the A-TEEM method's multivariate analysis simultaneously resolves at least three NOM fractions, including the low-molecular weight low-aromatic fulvic-like fraction (C1), the highly aromatic



PWSId	CitiesServed	StateCode	PWSTypeCode	PrimarySourceCode	PopulationServedCount	QtrsWithVio	SeriousViolator	Viopaccr	Vioremain	Ifea	Feas
FL4501229	RIVIERA BEACH	FL	CWS	GW	34700	12	Yes	155	50	10	2
WV3305009	KENOVA	WV	CWS	SW	9094	12	Yes	24	14	15	
TN0000655	SPENCER	TN	CWS	SW	5217	12	Yes	58	39	26	2
WV3300609	MILTON	WV	CWS	SW	4883	12	Yes	26	16	24	
FL1190150	APALACHICOLA	FL	CWS	GW	4710	12	Yes	65	59	32	
NN0403001	Ganado, Cornfield, Steamboat, Woodspring	AZ	CWS	GW	4151	12	Yes	7	6	5	
NN0403033	Leupp, Birdsprings, Tolani Lake	AZ	CWS	GW	2147	12	Yes	6	6	2	
FL6284075	SEBRING	FL	CWS	GW	2000	12	Yes	51	51	16	1
WV3302336	MAN	WV	CWS	SW	1943	12	Yes	33	10	53	
WV3302852	BLUESTONE	WV	CWS	SW	1933	12	Yes	147	38	344	
NN0400326	Shonto Junction	AZ	CWS	GW	1179	12	Yes	13	7	4	
NN0400280	Houck, Querino Canyon	AZ	CWS	GW	1106	12	Yes	16	6	4	
WV9941085	BRADLEY	WV	NTNCWS	GW	1000	7	Yes	51	22	66	

Table 1: A Sample List of EPA Reported Drinking Water Facilities with DBPs MCL Violations Nationwide from 2017 - 2019
(Serious Violators and Population Served > 1000, See footnotes[1,2])

hydrophobic humic-like fraction (C2), and the tryptophan like fraction (C3). Most importantly, the A-TEEM method resolves the effects of coagulation to reduce all three fractions while increasing the C2:C1 ratio due to the higher affinity of C2 for coagulant adsorption, compared to C1.

We had earlier documented the A-TEEM's capacity to rapidly generate highly-correlated Simulated Distribution System THM Formation Potential (SDS-THMFP) values based on direct correlations with American Water Works Association (AWWA) standard 10-day tests for raw source and finished water [Gilmore, A.M. 2017]. In this application note, we further investigated the capacity of the A-TEEM to make instantaneous predictions of individual THM species formation as measured in the finished water based on the NOM composition of both raw source and finished water. This study compared reference calibration measurements of three observed THM species with GC and GC-MS to the A-TEEM predictions. The effectiveness of the THM species predictions, based on Multiple Linear Regression (MLR), was systematically evaluated with two aims, namely, 1) to identify the most significant independent variables, and 2) to restrict the variables to those generated directly by the A-TEEM instrument thus minimizing or eliminating the need for auxiliary measurements. The key variables evaluated included the C1-C3 fractions, the C2:C1 ratio, the absorbance at 254 nm (A254), the chlorine residual concentration in the finished water ($[Cl_2]$) and the source and finished water temperatures (T). We report an effective final model that serves as an early warning indicator of the finished water THMs based on the source water NOM composition.

Material And Methods

Daily raw source water, settled water, and effluent samples from a conventional surface water coagulation-sedimentation-filtration treatment plant located in New Jersey were collected. These samples were filtered by 0.45 micrometer (μm) membrane filters. A-TEEMs were then measured with a HORIBA Aqualog Model UV-800 with a fixed 5 nm optical bandpass. Automatic Sipper and Sampling unit (HORIBA W-10S and WS-10S) accessories were used for automatic sample extraction and measurement daily.

The Sipper demonstrated exceptional performance with a stable baseline absorbance throughout the entire one-year operation, without having to replace the flow cell in the sample cell.

The blank absorbance value was evaluated quarterly. It has shown an exceptional stability over the course of one year (Figure 1). Absorbance baseline values at 254 nm were compared and the standard deviation was 0.01052. This indicated the flow cell operation effectively prevented air bubbles and biofouling in the sample cell.

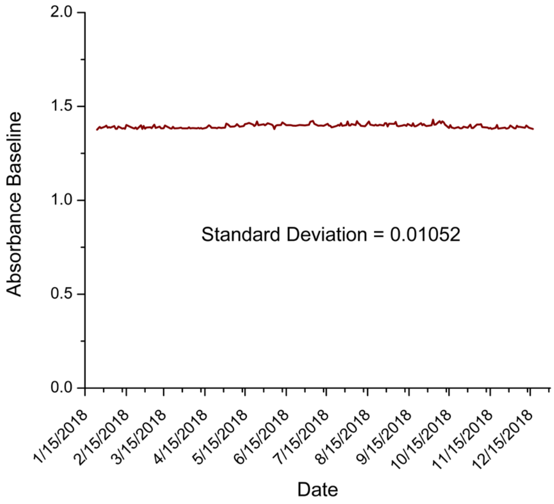


Figure 1: A stable absorbance baseline (A254nm) was achieved with a standard deviation = 0.01052, over a testing period of one year

This plant monitors TOC at source (raw), and treated (finished) water in accordance with the USEPA DBPR2. This plant applies Cl_2 as the primary disinfectant with an average residual at the plant exit of 1.7 ± 0.07 mg/L. Daily raw, settled, and finished samples were analyzed on site for DOC by a Siever's TOC Analyzer M5310 (Suez, France). Daily finished samples were also analyzed on site for THM species, i.e., CHFM, DCBM, DBCM, and BRFM, by a GC purge-n-trap THM analyzer PH-THM-1000 (Parker and Hannifin, USA). The THM samples were analyzed within 24 hours of collection time, or preserved with sodium thiosulfate at 4 °C or lower, and analyzed within 14 days of collection time.

---

[1] Data downloaded from the EPA's Enforcement and Compliance History online (ECHO) "Facility Search" database. Search criteria used: Search Type: Drinking water; Activity Status: Yes; Compliance Status: Health-based Violation; Contaminant(s) in Violation (3 years): Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules.

[2] Health-based violations are violations of maximum contaminant levels (MCLs) or maximum residual disinfectant levels (MRDLs), which specify the highest concentrations of contaminants or disinfectants, respectively, allowed in drinking water; or of treatment technique (TT) rules, which specify required processes intended to reduce the amounts of contaminants in drinking water. MCLs, MRDLs, and treatment technique rules are all health-based drinking water standards.

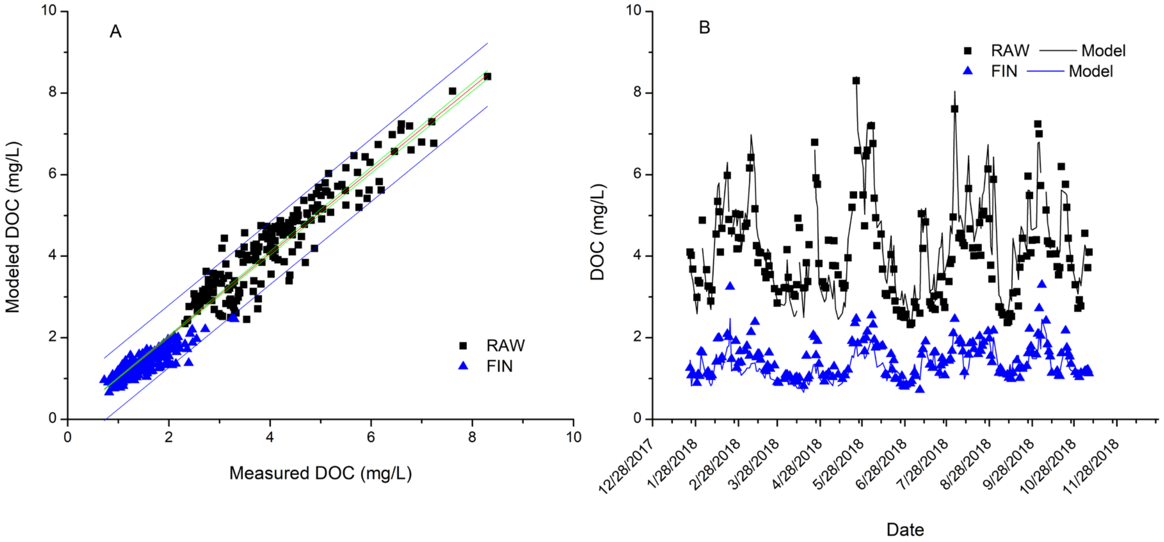


Figure 2: (A) Model linear correlation of the DOC calculated and the measured DOC values, and (B) the time series comparison of the model DOC calculation to actual daily DOC measurements from (A) for the raw (RAW) and finished (FIN) water samples.

PARAFAC Analysis

All EEMs were corrected for spectral excitation and emission response, dark offset and sealed water blank (Starna Q-10) signals before inner-filter effect correction was applied. The A-TEEM model of the NOM fractions (C1-C3) were analyzed using a split-half validated PARAFAC model (98.71% matching) using the Solo+MIA v8.7 (Eigenvector Inc., USA) package. All loadings were constrained to non-negativity and the pre-processing included Rayleigh masking of 16 nm and 32 nm, for first and second order, respectively, and normalization of the EEM area to unity.

DOC Method and THM Prediction Algorithm

The key classes of components resolved from the PARAFAC were directly correlated with independently measured DOC concentrations [Gilmore, A.M. and Tong, X. 2014; Gilmore, A.M. 2017] and THM concentrations, using the MLR functions in Origin v8.6 (OriginLab, USA).

Results And Discussion

There are several known factors that affect Total THM (TTHM) occurrence in the drinking water distribution system. Various engineering design factors in the disinfection treatment process should be considered, such as type and dosage of disinfectants, disinfection point, disinfectant contact time, and residence time in the distribution system [Saidan et al, 2013]. In the meantime, the type and level of DBP precursor in the source water is a key intrinsic factor for TTHM formation. The water temperature is another important factor, as seasonal variations have been observed between the TOC and byproducts formation, therefore, there is a strong correlation between the temperature and TTHM formation.

Characterization of DOC Measured During the Study

As shown in previous studies [Gilmore, A.M. 2017], the patented A-TEEM method using only the pH, A254 and C2:C1 ratio as linear coefficients, accurately predicts (Figure. 2A) the DOC for both the raw and finished water samples, with a slope = 1.02 and an adjusted R-square of 0.986. This is important for this water treatment plant because it reduces the frequency of the need for routine conventional DOC analysis. The time-series plot in Figure 2B shows the good correlation between the A-TEEM modeled DOC results and the actual measured DOC in daily raw and finished samples, over a course of approximately 9 months.

Evaluation of SUVA and Treatability with EPA Method 415.3

Figure 3 compares both the Aqualog SUVA and plant SUVA values for corresponding raw and finished water samples.

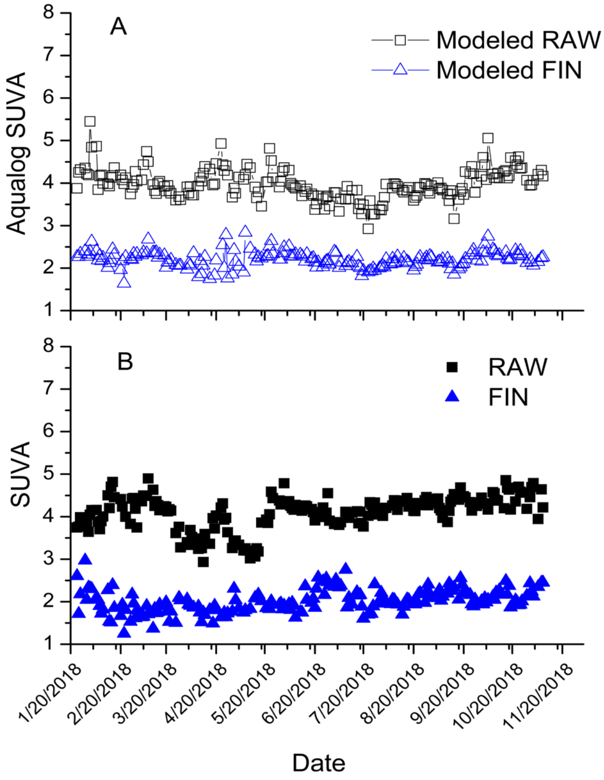


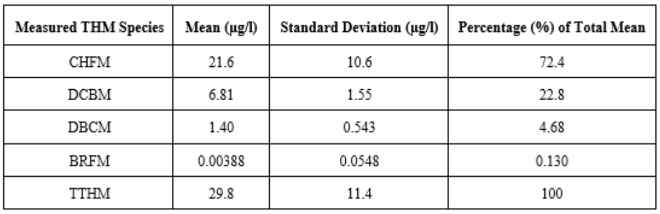
Figure 3: (A) Aqualog SUVA and (B) plant SUVA values for daily raw (RAW) and finished (FIN) samples. SUVA = 100 x A254 / TOC (L/mg-m).

The SUVA threshold for treatability is generally recognized to be at values < 2 and the raw water samples were significantly above this threshold. The finished samples were at, if not slightly below, this limit to indicate coagulation treatment was required.

SUVA is used as an indicator that estimates aromaticity of DOC [Weishaar et al, 2003; Krasner et al, 1996]. Conventionally, this requires side-by-side measurements from two instruments, thus the precision and accuracy are subject to errors due to different instrument responses. The Aqualog's single instrument SUVA determination method eliminates these errors and leads to a more precise estimate of aromaticity and treatability.

Evaluation of THM Species in the Finished Water

As shown in Table 2, based on the reference GC method measurements, CHFM was the most prominent THM with an average of 72.4% of the total mean THM, followed by DCBM with an average of 22.8%, DBCM with the average of 4.7%, and bromoform (BRFM) with a negligible average of 0.1%.



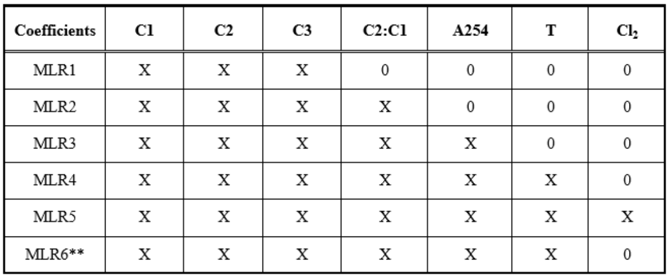
Measured THM Species	Mean (µg/l)	Standard Deviation (µg/l)	Percentage (%) of Total Mean
CHFM	21.6	10.6	72.4
DCBM	6.81	1.55	22.8
DBCM	1.40	0.543	4.68
BRFM	0.00388	0.0548	0.130
TTHM	29.8	11.4	100

Table 2: Percentage Distribution of THM Species

Evaluation of THM Multilinear Regression Correlations

Over a period of approximately 188 days, we collected reference GC and A-TEEM data to evaluate the Multiple Linear Regression (MLR) models of the three THMs independently, including CHFM, DCBM and DBCM. Notably because BRFM was present at very low to non-detectable limits, it was not evaluated further.

The mechanism and kinetics of THM formation, using the concentration of CHFM as the primary component, were systematically evaluated as a function of key independent variables (Table 3) associated with individual MLR regression coefficients.

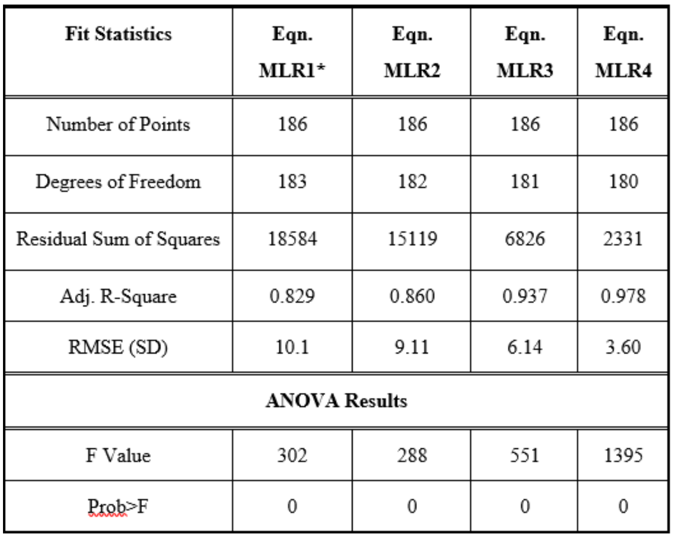


Coefficients	C1	C2	C3	C2:C1	A254	T	Cl_2
MLR1	X	X	X	0	0	0	0
MLR2	X	X	X	X	0	0	0
MLR3	X	X	X	X	X	0	0
MLR4	X	X	X	X	X	X	0
MLR5	X	X	X	X	X	X	X
MLR6**	X	X	X	X	X	X	0

Table 3: Coefficients* used for MLR Model Evaluation for Finished (MLR1 – 5) and Raw (MLR 6) Water THM Species Prediction

Five MLR models (MLR1-5) for CHFM for finished water samples only, were evaluated to test the following null hypotheses. For MLR1 we tested the hypothesis that the scores of C1-3 would not describe the variance in the observed CHFM concentration. For MLR2 we tested if variance would not be accounted for by also including the C2:C1 ratio since this is an indicator of aromaticity and NOM reactivity to Cl2. MLR3 additionally included A254 to test the influence of the aromatic NOM concentration. MLR4 further included the finished water temperature T to test if THM formation was not temperature dependent. MLR5 tested the [Cl2] assuming that THM formation requires chlorination. MLR6 for the raw source water included the same variables as MLR4 for the finished water, noting the source water was not chlorinated, even though MLR6 was regressed against the [CHFM] in the finished water.

The key goodness-of-fit parameters and Analysis of Variance (ANOVA) statistics results for MLR1-4 for CHFM in finished water are summarized in Table 4.



Fit Statistics	Eqn. MLR1*	Eqn. MLR2	Eqn. MLR3	Eqn. MLR4
Number of Points	186	186	186	186
Degrees of Freedom	183	182	181	180
Residual Sum of Squares	18584	15119	6826	2331
Adj. R-Square	0.829	0.860	0.937	0.978
RMSE (SD)	10.1	9.11	6.14	3.60
ANOVA Results				
F Value	302	288	551	1395
Prob>F	0	0	0	0

Table 4: Summary of Statistical Fits for MLR 1 – 4 using CHFM as an Example

It is clear that the adjusted R2, Residual Sum of Squares, Root Mean Square Error (RMSE) and ANOVA statistics (F-value) for predicting CHFM improved systematically during the sequential inclusion of C2:C1 ratio, A254 and especially the finished water temperature (T) coefficients. As shown in Figure 4A, T fluctuated from 3 degrees Celsius (^{o}C) to 28 ^{o}C, where the warmest months were July and August with temperatures above 25 ^{o}C. Figure 4B shows a comparison of regular residuals in CHFM predictions in finished water, without (MLR3) and with (MLR4) the T coefficient. This proves the fact that T is an important variable contributing to THM formation. The residuals curve also shares the same pattern with the T curve with the maximum in July and August.

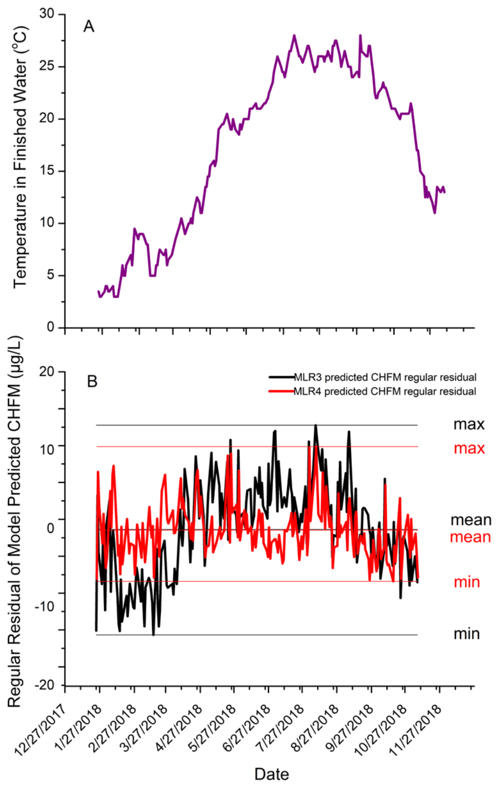


Figure 4: (A) Time-series plot of effluent temperature at the plant; (B) Comparison of time series plots of regular residuals of CHFM predictions in finished water without the temperature coefficient (MLR3, black lines), and with the temperature coefficient (MLR4, red lines).

Table 5 continues the analysis to show the goodness-of-fit and ANOVA summary for all THM species in finished water predicted by equation MLR5, which adds the [Cl_2] variable.

Fit Statistics	Eqn. MLR5 – Finished Water			
Parameters	**CHFM**	**DCBM**	**DBCM**	**TTHM**
Number of Points	186	186	186	186
Degrees of Freedom	179	179	179	179
Residual Sum of Squares	2305	218	29.3	3338
Adj. R-Square	0.978	0.977	0.930	0.983
RMSE (SD)	3.59	1.10	0.405	4.32
ANOVA Results				
F Value	1203	1154	356	1498
Prob>F	0	0	0	0

Table 5: Summary of Statistical Fits for Equation MLR 5 for All THM Species

As shown in Figure 5 the Cl_2 residual in the finished water as it leaves the plant is reasonably a constant (from 1.52 - 1.90 mg/L) during this study, as it was measured in the effluent only when exiting the plant. Noting this does not necessarily represent the chlorine residual at any point spatially or temporally in the distribution system network.

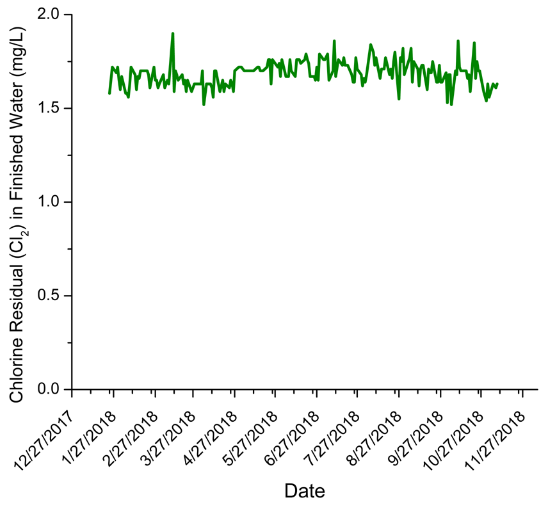


Figure 5: Time series plots of chlorine residual (Cl2) in finished water.

Table 6 shows that the predictive model using equation MLR6 for all THM species in raw water noting the variable [Cl_2] in the finished water was not considered in this case, since it may not be available when sampling source water. The overall correlation model yielded an adjusted R2 = 0.971 with a RMSE = 5.57 for TTHM. This prediction model adds tremendous value in that it can be deployed in an automated dashboard that predicts and alerts TTHM concentrations in real-time even before the water enters the treatment process, without any delay of lab analysis turnaround time.

Fit Statistics	Eqn. MLR6 – Raw Water			
Parameters	**CHFM**	**DCBM**	**DBCM**	**TTHM**
Number of Points	188	188	188	188
Degrees of Freedom	182	182	182	182
Residual Sum of Squares	4356	225	25.2	5637
Adj. R-Square	0.960	0.977	0.941	0.971
RMSE (SD)	4.89	1.11	0.371	5.57
ANOVA Results				
F Value	758	1336	499	1060
Prob>F	0	0	0	0

Table 6: Summary of Statistical Fits for Equation MLR 6 for All THM Species

Figure 6 shows the overall linear fits of THM species with the modeled THM species in finished water with a slope = 0.956, adjusted R2 = 0.980, and in raw water with a slope = 0.972, adjusted R2 = 0.955. As shown in Figure 7, the linear relationships allow accurate prediction of the THM species on a daily basis for both the finished and raw water samples.

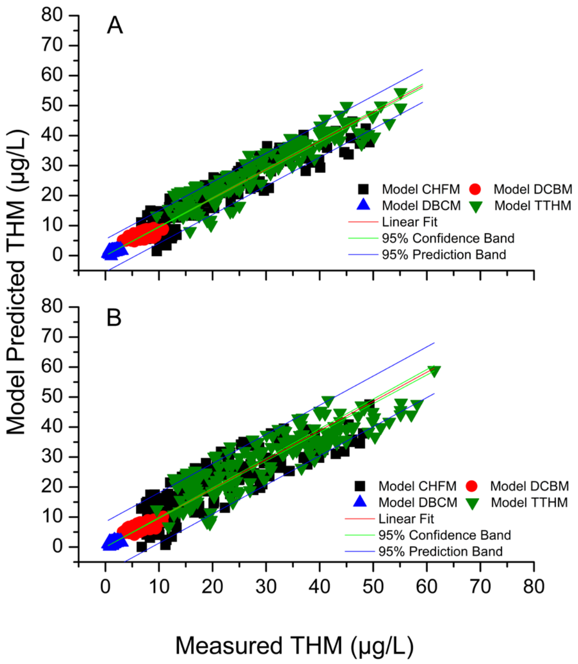


Figure 6: Overall linear fits of THM species with the modeled THM species in finished water (A) Slope = 0.956, adjusted R2 = 0.980; and raw water (B) Slope = 0.972, adjusted R2 = 0.955. The linear fit prediction (red line) shown is compared with the 95% confidence interval (green line) and 95% prediction interval (blue line).

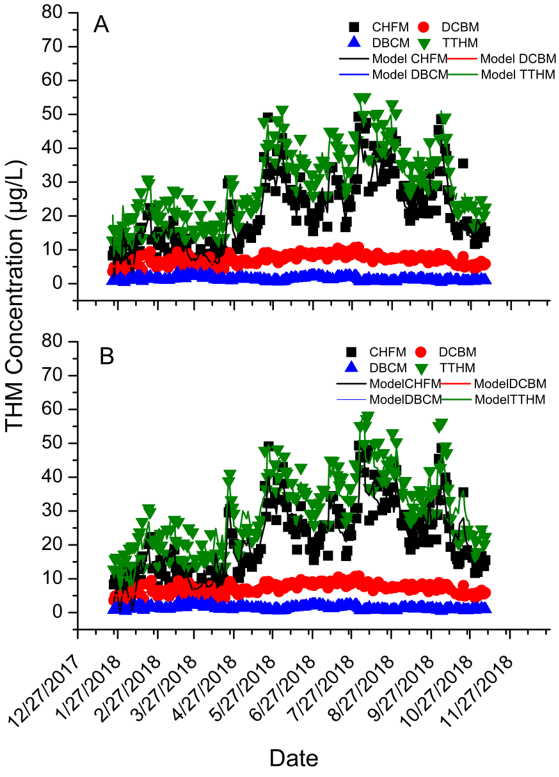


Figure 7: Comparison of the model and actual periodic measurements of THM species in finished water (A) and raw water (B).

Conclusions

The factors affecting TTHM formation and species from chlorination of water can be correlated by simple regression based on experimental data. The Aqualog system is compatible with an automated flow sampling device, and the output data can be reported instantaneously in an HTML-based dashboard. This allows for a reagent-free method that rapidly predicts THM and species concentrations. This new method has the potential to offer the cost, and operational advantages over conventional benchtop or online THM measurements using GC technology, which can require multiple units for effective plant and or distribution system monitoring. Subsequently, corrective or preventative actions to limit the occurrence of a water quality characteristic could provide a water treatment plant operator with rapid alternatives in water treatment practices to mitigate the potential for THM formation.

References

EPA's Enforcement and Compliance History Online (ECHO) database: https://echo.epa.gov/

Gilmore, A. M.; Tong, X. (2014). United States Patent No. US 8,901,513 B2. System and Method for Fluorescence and Absorbance Analysis.

Gilmore, A. M.; (2017). United States Patent No. US 9,670,072 B2. Determination of Water Treatment Parameters based on Absorbance and Fluorescence.

Saidan, M.; Rawajfeh, K.; & Fayyad, M., 2013. Investigation of Factors Affecting THMs Formation in Drinking Water. Am. Jour. of Envir. Engr., Vol. 3 No. 5, pp. 207-212.

Krasner, S.W.; Croue, J.P.; Buffle, J.; & Perdue, E.M., 1996. Three Approaches for Characterizing NOM. Jour. AWWA, 88:6:66.

Weishaar, J.L.; Aiken, G.R.; Bergamaschi, B.A.; Fram, M.S.; Fujii, R.; & Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. Envir. Sci. & Technol., 37:20:4702.

To learn more about the A-TEEM technique and applications go to **A-TEEM.com**

info.sci@horiba.com
www.horiba.com/scientific

HORIBA
Scientific

USA: +1 732 494 8660
UK: +44 (0)1604 542 500
China: +86 (0)21 6289 6060
Taiwan: +886 3 5600606
France: +33 (0)1 69 74 72 00
Italy: +39 06 51 59 22 1
India: +91 80 41273637
Brazil: +55 (0)11 2923 5400
Germany: +49 (0) 6251 8475 0
Japan: +81(75)313-8121
Singapore: +65 (0)6 745 8300
Other: +33 (0)1 69 74 72 00