2-1-2015

Evaluation of the Relationship between Bulk Organic Precursors and Disinfection Byproduct Formation for Advanced Oxidation Processes

Brooke Mayer
Marquette University, Brooke.Mayer@marquette.edu

Erin Daugherty
Arizona State University at the Tempe Campus

Morteza Abbaszadegan
Arizona State University at the Tempe Campus

Accepted version. *Chemosphere*, Vol. 121 (February 2015): 39-46. DOI.

Evaluation of The Relationship Between Bulk Organic Precursors and Disinfection Byproduct Formation for Advanced Oxidation Processes

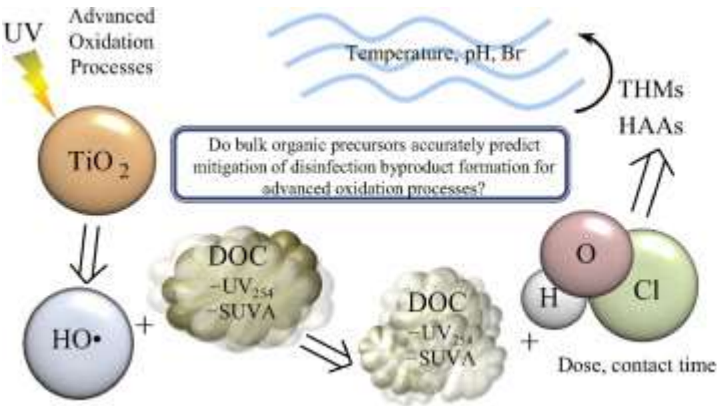
Brooke K. Mayer
Civil, Construction and Environmental Engineering,
Marquette University
Milwaukee, WI
Erin Daugherty
Arizona State University,
Tempe, AZ
Morteza Abbaszadegan
Arizona State University,
Tempe, AZ

Abstract: Advanced oxidation processes (AOPs) are gaining traction as they offer mineralization potential rather than transferring contaminants between media. However, AOPs operated with limited energy and/or chemical inputs

Chemosphere, Vol 121 (February 2015): pg. 39-46. DOI.

can exacerbate disinfection byproduct (DBP) formation, even as precursors such as dissolved organic carbon, UV_{254}, and specific UV absorbance (SUVA) decrease. This study examined the relationship between DBP precursors and formation using TiO_2 photocatalysis experiments, external AOP and non-AOP data, and predictive DBP models. The top-performing indicator, SUVA, generally correlated positively with trihalomethanes and haloacetic acids, but limited-energy photocatalysis yielded contrasting negative correlations. The accuracy of predicted DBP values from models based on bulk parameters was generally poor, regardless of use and extent of AOP treatment and type of source water. Though performance improved for scenarios bounded by conditions used in model development, only 0.5% of the model/dataset pairings satisfied all measured parameter boundary conditions, thereby introducing skepticism toward model usefulness. Study findings suggest that caution should be employed when using bulk indicators and/or models as a metric for AOP mitigation of DBP formation potential, particularly for limited-energy/chemical inputs.

Graphical abstract



Keywords: Disinfection byproduct; Advanced oxidation; Titanium dioxide photocatalysis; Trihalomethane; Haloacetic acid; Model

1. Introduction

Increasing implementation of advanced oxidation processes (AOPs) makes it imperative to improve understanding of their disinfection byproduct (DBP) mitigation potential, including the extent to which traditional DBP indicators and the predictive DBP formation models based on them reliably predict AOP performance. This study evaluated the relationship between traditional bulk DBP precursors and DBP formation following AOP treatment using pilot-scale TiO_2 photocatalysis experiments, external DBP data (both AOP and non-

AOP), and published predictive models for total trihalomethanes (TTHMs), haloacetic acids (HAA5s), and subspecies thereof.

Since their discovery in 1974 (Bellar et al., 1974 and Rook, 1974), the production of over 600 potentially harmful DBPs has attracted considerable attention (Krasner et al., 2006). Drinking water DBPs are widely regulated, as described in the Supplementary Information (SI), and increasing interest in water reclamation and reuse has sparked recognition of the occurrence and impacts of DBPs in wastewater (Wei et al., 2009). Chlorine disinfection alternatives such as chloramines, chlorine dioxide, ozone, or ultraviolet (UV) may reduce the formation of trihalomethanes (THMs) and haloacetic acids (HAAs); however, they may still form DBPs, and can shift DBP production toward species of greater health concern (Richardson, 2005). Additionally, alternative disinfectants may be ineffective against certain pathogens, increase nitrification and corrosion in the distribution system, and/or lack disinfectant residual (USEPA, 2006 and Zhang et al., 2008). Thus, the most common DBP reduction strategy is removal of precursor natural organic matter (NOM) prior to disinfection (Kulkarni and Chellam, 2010). The best available techniques – enhanced coagulation or softening, granular activated carbon (GAC), or membrane filtration (Liu et al., 2008b and USEPA, 2006) – are often sufficient to control DBP formation; however, alternatives such as AOPs are also being investigated (Mayer et al., 2014 and WHO, 2000).

The highly reactive, nonspecific radical species, such as $HO\cdot$, produced by AOPs are capable of mineralizing NOM, oxidizing trace emerging contaminants, and inactivating microorganisms (Gerrity et al., 2009). Given their potential to mineralize NOM rather than capturing and transferring it to another phase, implementation of AOPs is becoming increasingly attractive (Gerrity et al., 2009 and Mayer et al., 2014), making it critically important to examine their effect on DBP formation (Dotson et al., 2010). Photocatalysis provides one pathway by which $HO\cdot$ radicals are produced using ultraviolet (UV) light ($\lambda < 387$ nm) to irradiate semiconductors such as TiO_2. However, AOPs such as TiO_2 photocatalysis are unlikely to be operated at sufficiently high energy and/or chemical inputs to mineralize NOM. Instead, under more practical operating conditions, AOPs are likely to incompletely oxidize organics, resulting in a shift toward smaller, less aromatic, and more hydrophilic moieties (Liu et al., 2008a, Mayer et al., 2014 and Sarathy and Mohseni, 2010), which can exacerbate DBP formation beyond initial levels (Dotson et al., 2010, Gerrity et al., 2009, Liu et al., 2008a and Mayer et al., 2014). For example, 5 kW h m^{-3} TiO_2/UV AOP treatment increased TTHMs up to 119% and HAA5s up to 299% when applied to waters of varying quality, ranging from untreated to finished drinking water (Gerrity et al., 2009 and Mayer et al., 2014). When higher AOP energy/chemical inputs are used, more complete NOM mineralization follows, resulting in decreased DBP formation, e.g., reductions of up to 95% TTHMs and 96% HAA5s (Gerrity et al., 2009 and Mayer et al., 2014). The initial trend of increasing DBP production using limited-input AOP treatments contrasts with decreases in traditional bulk DBP precursors such as dissolved organic carbon (DOC), ultraviolet absorbance at a wavelength of 254 nm (UV_{254}), and specific UV absorbance (SUVA). This suggests that traditional bulk indicators of DBPs may not provide suitable performance metrics for AOPs, which fundamentally augment NOM characteristics rather than physically removing it.

2. Materials and methods

2.1. TiO_2 photocatalysis

Samples were collected from several locations in Central Arizona drinking water treatment trains using a mixture of surface and groundwater, including following sedimentation (SW), following GAC filtration (GAC), and from the distribution system (DS). Details of the treatment trains are provided in the SI. The samples were used as the influent water for pilot-scale TiO_2 photocatalysis experiments.

Photocatalysis was performed using the Purifics (London, ON) Photo-Cat® Lab reactor at an optimized concentration of 1 g L^{-1} (Gerrity et al., 2009 and Liu et al., 2008b) suspended reagent-grade Degussa P25 TiO_2 (Dusseldorf, Germany). The reactor was operated in batch configuration, as described previously (Gerrity et al., 2009 and Mayer et al., 2014). For each source water, dark adsorption (no UV) samples were collected to provide a measure of physical removal due to adsorption and retention on the Photo-Cat's®

Chemosphere, Vol 121 (February 2015): pg. 39-46. DOI. This article is © Elsevier and permission has been granted for this version to appear in e-Publications@Marquette. Elsevier does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Elsevier.

submicron-pore-size ceramic membrane filter. Additionally, three levels of treatment were tested for each water: limited (5 kW h m^{-3}), medium (80 kW h m^{-3}), and extended photocatalysis (160 kW h m^{-3}), providing a spectrum of partial, incomplete, and near-complete NOM oxidation (Mayer et al., 2014).

2.2. Water quality parameters

Following photocatalysis, turbidity was measured using a Hach (Loveland, CO) model 2100P turbidimeter, and a Mettler meter (Columbus, OH) was used for pH. Traditional DBP indicators quantified included DOC, UV_{254}, SUVA, and bromide (Br^-). Particulate matter was removed from untreated samples using 0.45 μm Pall (Port Washington, NY) Acrodisc® GHP membrane filters. DOC samples were acidified with 1 M HCl, and analyzed using a Shimadzu (Kyoto, Japan) 5050A Total Organic Carbon Analyzer. A Hach DR 5000 spectrophotometer was used to measure UV_{254}, and SUVA was calculated by normalizing UV_{254} values with respect to DOC. A Dionex (Sunnyvale, CA) DX-120 ion chromatograph was used to measure Br^-.

2.3. Disinfection byproduct formation

Treated samples were chlorinated in accordance with a modified simulated distribution system test (SDS-THM), which targeted a chlorine residual of 1 mg L^{-1} after 24 h at 28 °C (APHA, 2005 and Mayer et al., 2014). A Hach D4/4000U spectrophotometer and DPD free chlorine reagent powder pillows were used to analyze free chlorine. The dominant weight-based classes of chlorine-DBPs, TTHMs and HAA5s (Krasner et al., 2006), were measured as described by Mayer et al. (2014). To gauge the evolution of DBP formation over time, samples were collected at 10–15 min, 1 h, 2 h, and 6 h post-chlorination. 24 h DBP formation was reported by Mayer et al. (2014).

2.4. External DBP datasets

In addition to internal DBP data from this study, external AOP and non-AOP datasets of TTHMs, HAA5s, and/or subspecies were collected from published literature to provide additional data to investigate the relationship between precursors and DBPs and to evaluate DBP model performance. This external data also functioned as a control to ensure that study findings were not artifacts of the laboratory protocol used in the current study. Twenty external datasets encompassing a range of experimental parameters were selected, as detailed in the SI.

2.5. DBP models

Internal and external DBP data were compared with DBP predictions using 149 published chlorine-DBP models, as described in the SI. These models predicted mass concentrations of TTHMs, HAA5s, and/or subspecies using traditional DBP indicators: DOC, UV_{254}, SUVA, chlorine dose (D), pH, temperature (T), Br^-, and disinfection contact time (t). Models incorporating other parameters were not included as this information was not readily available for the datasets analyzed here. Additionally, this approach to model selection supported the study objective of evaluating the relationship between common bulk indicators of DBPs and measured DBP formation. This dictates that the predictive results reported here are purely empirically-derived due to the exclusion of semi-mechanistic models.

2.6. Statistics

GraphPad Prism was used to compute all statistics at a significance level of $\alpha = 0.05$. Normality was tested using the D'Agostino-Pearson omnibus test. The relationship between DBP indicators and DBP formation was assessed using the Pearson product-moment coefficient of correlation for normally distributed datasets, while the nonparametric Spearman rank correlation coefficient was used for non-normal distributions. To assess similarities between the means of groups, multiplicity-adjusted P values were calculated using one-way ANOVA for normally distributed groups, while the nonparametric Kruskal–Wallis one-way analysis of variance was used for non-normal distributions. For groups with statistical differences, post-hoc tests were used to identify dissimilar datasets by comparing each mean to all other means using Tukey's Honestly Significant Difference (HSD) test following ANOVA or the nonparametric Dunn's test following Kruskal–Wallis tests.

3. Results and discussion

Initial source water quality differed among samples, reflecting the range of treatments applied prior to photocatalysis, as shown in Table 1.

Table 1. Water quality parameters for TiO_2 photocatalysis experiments.

Water source	**Treatment level**	**pH**	**Turbidity (NTU)**	**DOC (mg L^{-1})**	**UV$_{254}$ (cm^{-1})**	**SUVA (L(mg m)$^{-1}$)**	**Br^{-} (μg L^{-1})**	**24-h TTHMs (μg L^{-1})**	**24-h HAA5s (μg L^{-1})**	**24-h Specific TTHMs (μg L^{-1})**	**24-h Specific HAA5s (μg L^{-1})**
CAP[a]	Untreated	8.50	2.20	5.70	0.046	0.81	109	–	–	–	–
SR[a]	Untreated	8.02	2.86	4.85	0.086	1.78	25	–	–	–	–
SW	Control	8.08	0.90	3.13	0.039	1.26	10	52.5	19.2	16.8	6.12
	Dark adsorption	7.75	0.20	2.70	0.029	1.06	0	62.4	28.9	23.1	10.72
	5 kWh m^{-3}	7.73	0.13	2.51	0.019	0.74	10	109	44.4	43.4	17.71
	80 kWh m^{-3}	8.43	0.23	0.54	0.003	0.61	10	11.0	4.04	20.4	7.48
	160 kWh m^{-3}	7.49	0.07	0.79	0.003	0.34	10	3.43	3.00	4.3	3.80
GAC	Control	8.31	0.39	2.09	0.032	1.51	32	43	18.7	20.6	8.93
	Dark adsorption	8.02	0.21	1.56	0.014	0.90	0	33.8	15.9	21.6	10.21
	5 kWh m^{-3}	8.06	0.21	1.56	0.009	0.55	2	60.3	29.0	38.7	18.60
	80 kWh m^{-3}	8.31	0.20	0.22	0.000	0.00	0	7.20	2.71	32.7	12.32
	160 kWh m^{-3}	8.44	0.16	0.45	0.000	0.00	0	1.92	0.00	4.3	0.00
DS	Control	8.42	0.09	3.02	0.013	0.44	0	29.6	8.97	9.8	2.97
	Dark adsorption	8.32	0.10	1.59	0.015	0.97	0	57.4	28.3	36.1	17.80
	5 kWh m^{-3}	8.28	0.13	1.94	0.014	0.73	0	64.8	35.8	33.9	18.74
	80 kWh m^{-3}	8.51	0.11	0.60	0.003	0.50	0	36.8	18.8	61.3	31.35
	160 kWh m^{-3}	8.38	0.09	0.44	0.002	0.45	0	17.6	8.76	40.0	19.91

[a] Untreated water from the Central Arizona Project Canal (CAP) and the Salt River (SR) were not tested, but provide a basis of comparison for water quality (and level of treatment) of the SW, GAC, and DS samples.

3.1. DBP precursors

Increasing removal of the bulk organic DBP precursors DOC, UV$_{254}$, and SUVA was observed following photocatalytic oxidation, as shown in Table 1. In comparison to DOC, faster and greater overall removal of UV$_{254}$, and hence SUVA, occurred. This signifies a progressive shift in NOM toward smaller, less aromatic species, whereas DOC reduction is only achieved following physical removal (dark adsorption), extended treatment due to more complete mineralization, or not at all due to refractory compounds, in agreement with previous reports (Gerrity et al., 2009, Huang et al., 2008, Liu et al., 2008a and Mayer et al., 2014). Changes in the NOM size distribution (quantified using high-performance size exclusion chromatography) and spectra (quantified using regional integration of excitation–emission matrix fluorescence spectroscopy) in response to limited energy photocatalytic treatment demonstrated shifts toward smaller, less aromatic, and less humic moieties, as reported by Mayer et al. (2014).

Concentrations in excess of 100 μg L^{-1} Br^{-} trigger concern about the formation of brominated DBPs (Zhang et al., 2011), which exhibit higher toxicity compared to their chlorinated counterparts (Plewa et al., 2002). However, chlorinated species tend to dominate in low Br^{-} waters (Richardson et al., 1999). In this study, a maximum of 32 μg L^{-1} Br^{-} was observed, effectively limiting the formation of brominated DBPs.

3.2. DBP formation and relation to precursors

As illustrated for the SW in Fig. 1, DBP evolution is characterized by rapid initial formation followed by a slower rate of generation several hours post chlorination (Gallard and von Guten, 2002). Fig. 1 also illustrates that dark adsorption and limited photocatalysis exacerbate DBPs beyond untreated control samples. Adsorption of higher molecular weight, hydrophobic NOM to TiO_2 surfaces can shift organic profiles toward more reactive species (Mayer et al., 2014), as shown by the increase in the specific DBP values shown in Table 1. Similar findings using industry-relevant energy and chemical inputs have been observed for TiO_2 photocatalysis (Gerrity et al., 2009, Liu et al., 2008b and Mayer et al., 2014) and other AOPs (Dotson et al., 2010, Kleiser and Frimmel, 2000 and Metz et al., 2011). This appears to stem from HO. radicals, and is not observed in non-AOP systems, e.g., UV alone (Kleiser and Frimmel, 2000). Incomplete NOM oxidation resulting from limited-input AOPs opens the rings of aromatic structures and cleaves double-bonded carbon structures, thereby yielding more reactive sites, which can increase DBP formation (Metz et al., 2011). Given sufficient energy/chemical inputs, more complete oxidation of NOM significantly reduces DBP concentrations.

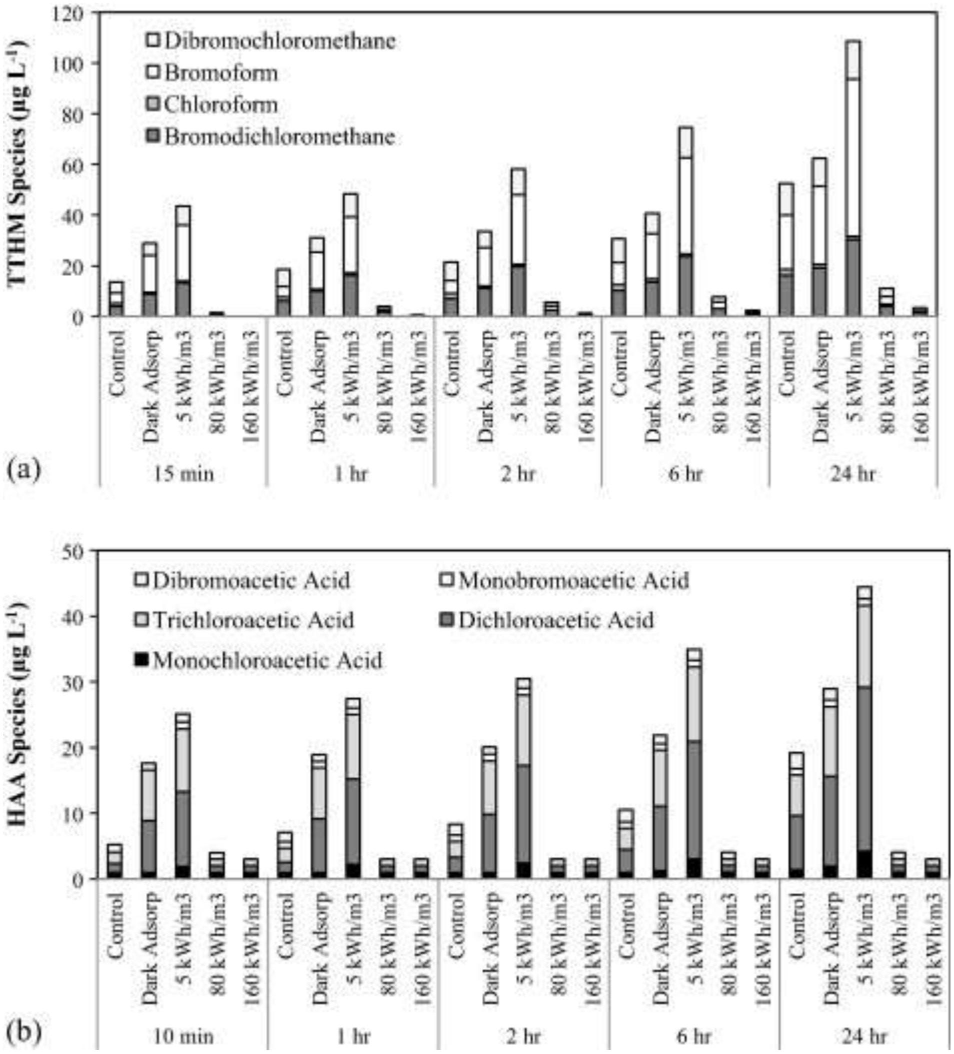


Fig. 1. Time evolution of (a) TTHM and (b) HAA5 species following chlorine disinfection in the SW as a function of time and extent of photocatalytic treatment. As shown in the SI, the GAC and DS data followed similar patterns, albeit with lower overall DBP formation.

The contrasting trend of decreasing bulk organic DBP precursors and increasing DBP formation using limited photocatalysis is apparent in Table 1. Limited-energy inputs correspond to spikes in specific TTHMs and HAA5s, which are normalized to DOC. This indicates that although the concentration of DOC in the system did not decrease substantially using 5 kW h m^{-3}, its DBP formation potential increased, signifying fundamental shifts toward more reactive moieties. Although contrasting DBP precursor/formation trends were observed for limited photocatalysis, linear correlation coefficients indicated that DOC, UV_{254}, and SUVA generally exhibited positive correlations with DBP formation, as shown in Fig. 2. Of these NOM-related parameters, SUVA correlated positively with DBPs most consistently, with the major exception of the 5 kW h m^{-3} dataset. For this dataset, the correlations between precursors and DBPs varied widely, even within individual parameter groupings. Notably, the majority of the DBP species were negatively correlated with SUVA at 5 kW h m^{-3} (though CF, DBCM, and DBA demonstrated strong positive correlations), whereas nearly all SUVA correlations were positive for the other treatment conditions. This supports the hypothesis that the incomplete oxidation of organics caused by limited-input AOPs alters NOM character such that traditional bulk measures such as SUVA are less reliable indicators of DBP formation compared to their effectiveness in physical removal scenarios such as enhanced coagulation or GAC filtration. The altered NOM composition resulting from limited-input AOPs exhibits different chlorine reactivity, which is not adequately conveyed by bulk properties such as SUVA, suggesting that caution is warranted when using bulk NOM precursors as indicators of AOP performance. With additional energy input, e.g., 80 and 160 kW h m^{-3}, organics are more completely oxidized, DOC and UV_{254} decrease, and SUVA demonstrates a strong positive correlation to DBP formation.

Chemosphere, Vol 121 (February 2015): pg. 39-46. DOI.

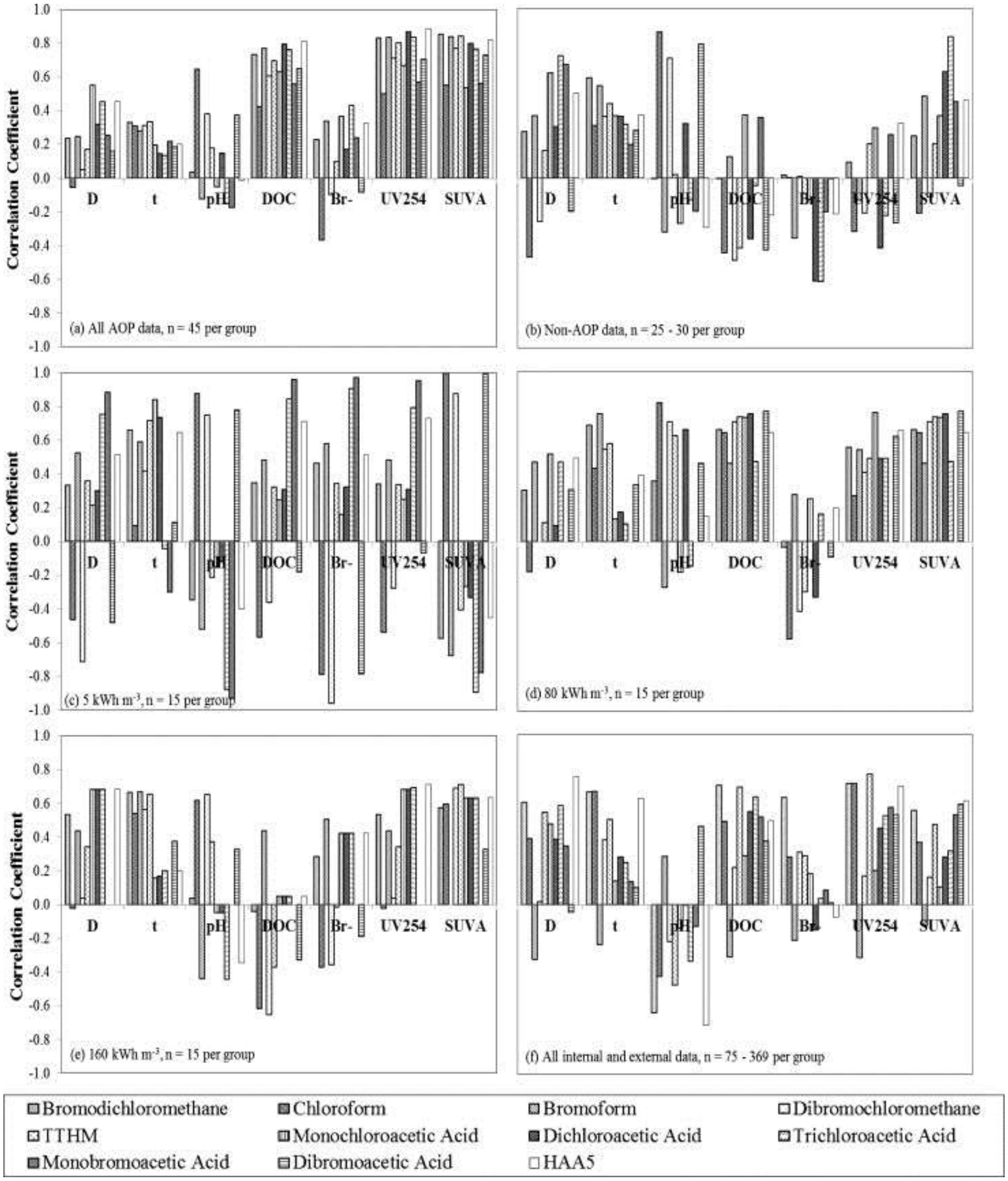


Fig. 2. Linear correlation coefficients (Pearson for normally distributed datasets, otherwise Spearman) between DBP formation and precursor measurements. Values closer to +1 indicate positive correlations, while values closer to −1 indicate negative correlations. Panels (a) through (e) are data from the current study, grouped by extent of photocatalytic treatment, while panel (f) illustrates correlations for all datasets combined, both internal and external.

Fig. 2 shows that DBP formation was also generally positively correlated with D and *t*, while Br^- and pH demonstrated mixed positive and negative associations with DBPs. A positive correlation between increasing pH and THMs is commonly reported, while HAAs are often negatively correlated (WHO, 2000). Full statistical results are provided in the SI.

3.3. DBP modeling assessment

3.3.1. Internal data

Comparisons of internal data and predicted TTHMs (combined results of 47 models) and HAA5s (combined results of 8 models) are illustrated in Fig. 3. Linear regression equations were used to gauge the overall accuracy of model predictions. The low R^2 values of 0.17 and 0.16 for TTHMs and HAA5s, respectively, signify poor overall model accuracy. Notably, these regressions are the combined measure of model fit, whereas the average R^2 values for individual TTHM models applied to this dataset was 0.58, with a range of 0.00–0.87. The average R^2 from individual HAA5 models was 0.15, with a range of 0.01–0.27. The R^2 values for all assessments are provided in the SI.

Chemosphere, Vol 121 (February 2015): pg. 39-46. DOI. This article is © Elsevier and permission has been granted for this version to appear in e-Publications@Marquette. Elsevier does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Elsevier.

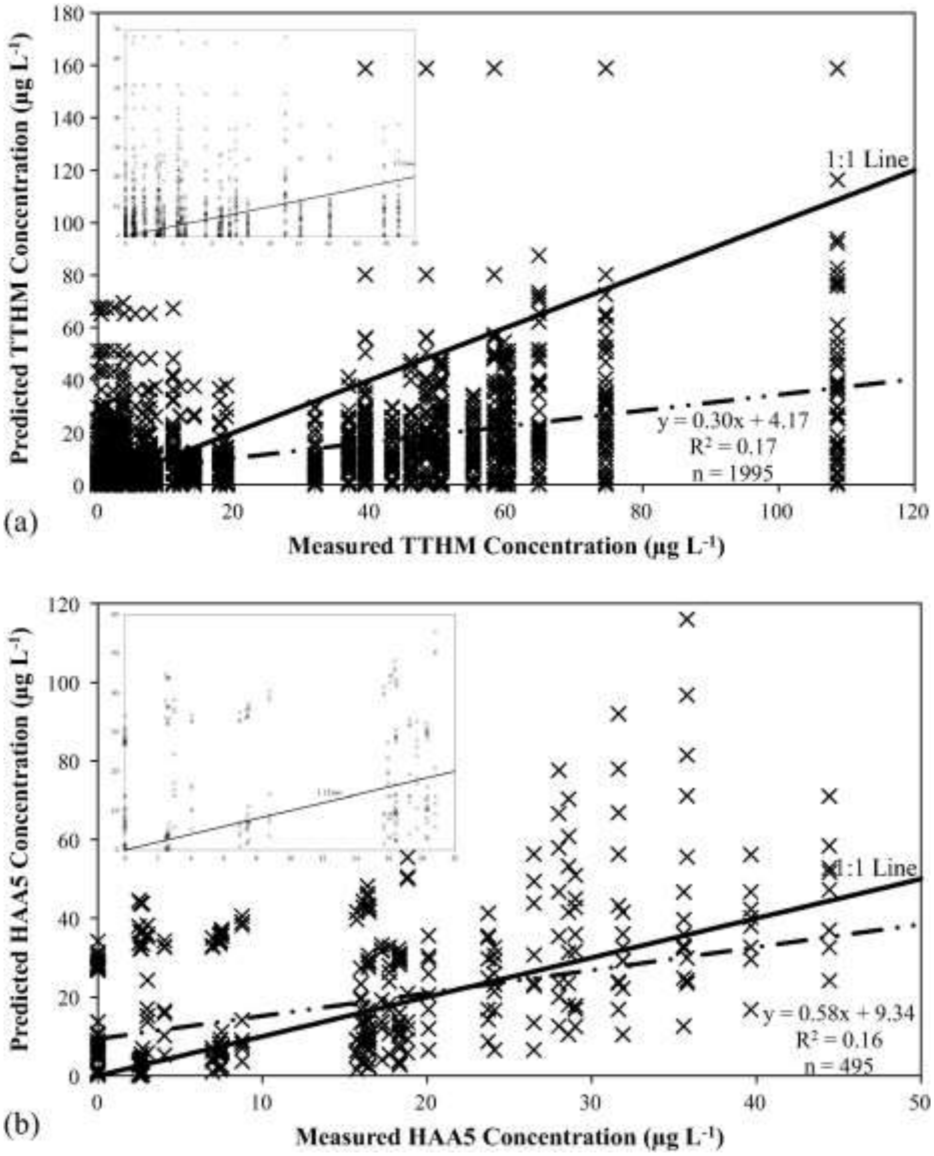


Fig. 3. Predicted versus measured DBP concentrations for all photocatalytic data from the current study. The predicted values are the results of (a) 47 independent published TTHM models and (b) 8 independent published HAA5 models. The insets show an enlarged view of data points up to 20 μg L^{-1} measured DBPs.

In addition to R^2 = 1.0, a perfect fit would produce slope = 1.0 and intercept = 0. Fig. 3 indicates that the models tend to over-predict at low DBP concentrations and under-predict at high concentrations (slopes <1 and intercepts >0). This suggests that empirical models commonly over-predict for conditions least conducive to DBP formation and under-predict for those most conducive, consistent with individual modeler reports (Amy et al., 1987, Amy et al., 1998 and Sohn et al., 2004). One possible explanation for this over/under-estimate trend is that these "extreme" formation conditions are more likely to extend beyond model boundary conditions, producing inherently less accurate predictions due to extrapolation, a practice which is typically discouraged (Amy et al., 1987).

The coefficients of determination (R^2) from comparisons of predicted DBP model values versus measured DBP formation for the internal data, grouped by level of photocatalytic treatment, are shown in Fig. 4a. ANOVA and post-hoc tests determined that model accuracy differed across all levels of treatment with the exception of the 80 and 160 kW h m^{-3} pairing. This suggests that model accuracy (a gauge of the predictive relationship between bulk precursors and DBP formation) differs as a function of photocatalytic treatment, i.e., the degree of NOM oxidation. Contrary to the hypothesis, however, incomplete oxidation at the 5 kW h m^{-3} treatment level, which is characterized by contradictory trends between bulk DBP indicators and DBP formation, exhibited the highest degree of overall model accuracy. The same trend was apparent with the HAA dataset, although statistics indicated that only the Dark Adsorption and 5 kW h m^{-3} pair differed significantly in model accuracy. Full statistical results are provided in the SI.

Chemosphere, Vol 121 (February 2015): pg. 39-46. DOI. This article is © Elsevier and permission has been granted for this version to appear in e-Publications@Marquette. Elsevier does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Elsevier.

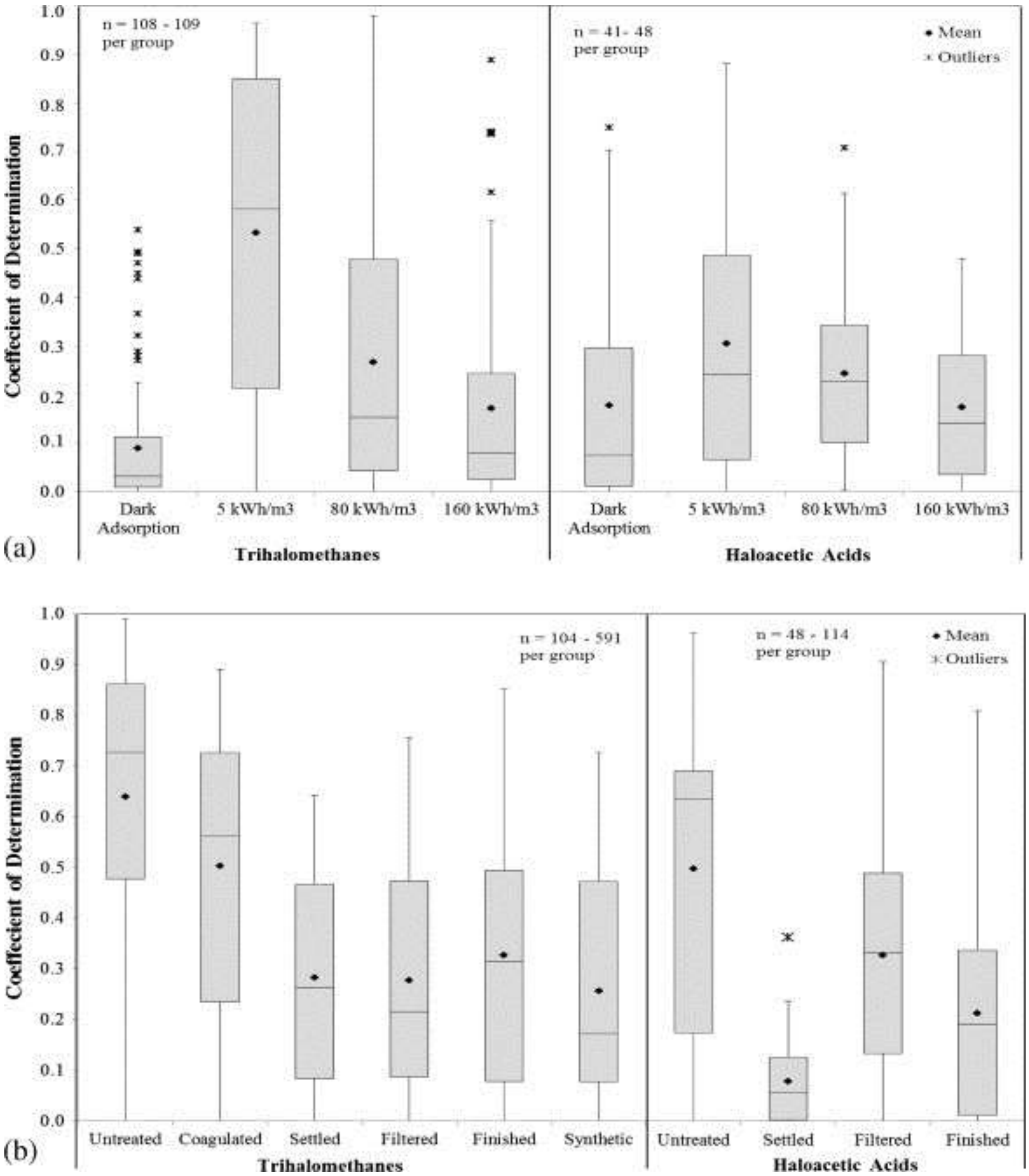


Fig. 4. Box and whisker plots showing the coefficients of determination (R^2) for predicted versus measured DBP values (including TTHM, HAA5, and subspecies thereof) for (a) photocatalysis-treated samples from the current study and (b) all internal and external datasets grouped by source water type. The whiskers denote the maximum and minimum values in the dataset exclusive of outliers, which were identified as data points located more than 1.5 times the interquartile range (IQR) beyond the IQR.

3.3.2. Source water

The internal and external DBP datasets analyzed here included water from varying sources, including untreated, partially treated, fully treated, and synthetic waters. These differences may impact model performance as the majority of DBP models were developed using untreated water, and as such may be less accurate for partially to fully-treated source waters, which vary in quantity and character of NOM.

The R^2 values from predicted DBP model values versus internal and external DBP measurements, grouped by type of source water, are shown in Fig. 4b. Statistical analyses showed that the accuracy of the predictive THM and HAA models differed as a function of source water. Post-hoc tests determined that THM model accuracy for the untreated and coagulated waters was significantly different from all other source waters. For HAAs, model accuracy differed across all water types with the exception of the settled and finished pair and filtered and finished pair. These findings suggest greater model accuracy is achieved when using untreated or coagulated waters, which are also the water sources used to develop the majority of the models assessed in this study. Additionally, modeling capabilities using synthetic samples were quite poor, possibly the result of more homogenous chemical compositions than those typically encountered in natural waters, for which the models were developed. Full statistical analyses are provided in the SI.

3.3.3. AOP versus non-AOP

The consolidated results of the 33 348 THM and 4773 HAA pairs of predicted DBPs versus internal and external data are shown in Fig. 5. Each box depicts the distribution of R^2 values from the predicted versus measured DBPs for a given dataset. Not all models yielded meaningful predictions for all datasets as a result of missing input data, inability to calculate R^2 due to model returning identical values for all scenarios, etc. Analysis of variance and post hoc tests identified differences in HAA model performance between the TiO_2 photocatalysis and non-TiO_2 AOP datasets. No statistical difference was found in the accuracy of model predictions among the THM datasets. Full statistical analyses are provided in the SI.

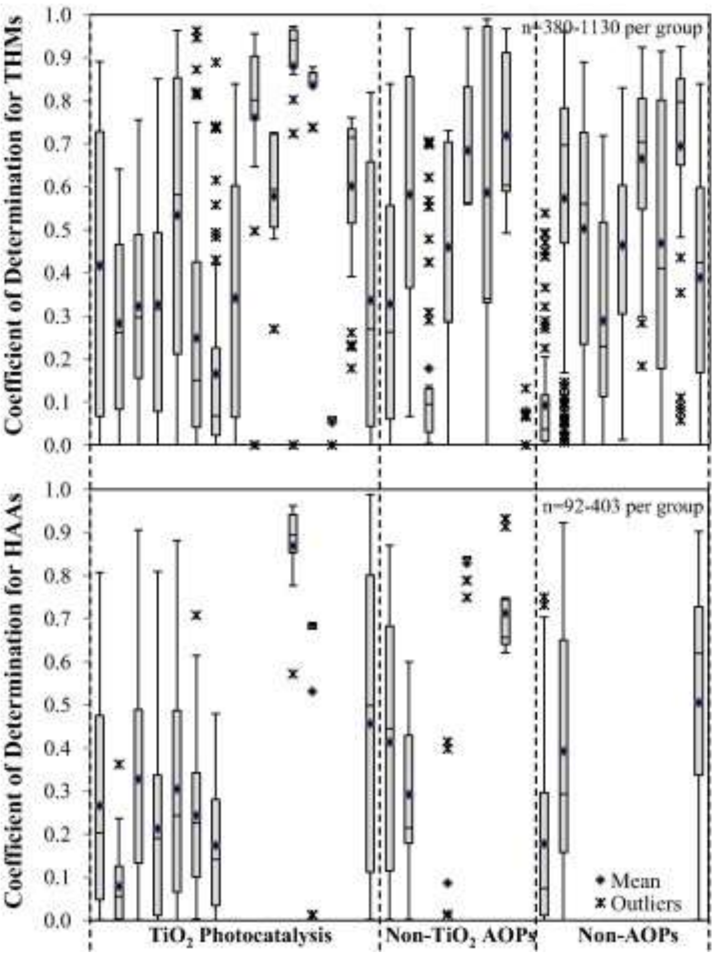


Fig. 5. Box and whisker plots of the coefficients of determination (R^2) for predicted versus measured THM and HAA values (including TTHM, HAA5, and subspecies thereof). Values are calculated from a maximum of 105 THM models (48 HAA models) for 31 different AOP and non-AOP THM datasets (16 HAA data sets). Data points located more than 1.5 times the interquartile range (IQR) outside the IQR are identified as outliers. The whiskers denote the maximum and minimum values in the dataset exclusive of outliers.

Fig. 3, Fig. 4 and Fig. 5 demonstrate generally poor performance by the predictive DBP models, regardless of the source water and/or treatment applied. This alludes to deeper challenges in the application of DBP models to external datasets beyond simply suggesting that caution be used when employing bulk indicators of DBP formation as a metric of AOP performance. Limitations of some existing models include calibration using a limited database, applicability to a specific water source or group thereof, disregard of important parameters such as reaction time, inadequate validation using independent external data, and/or inclusion of terms not typically measured such as site-specific parameters (Amy et al., 1987).

One of the key implications of the empirical nature of most models is the introduction of specific boundaries related to source water, water quality parameters, and treatment conditions. Most models were developed using untreated, coagulated, or finished conventionally-treated water. As such, satisfactory model performance may be limited to a narrow range of treatment scenarios. Accordingly, the overall poor performance of the models tested here may be a function of applying them to datasets that did not satisfy all boundary conditions. Unfortunately, strict agreement with all model boundary conditions is infrequently encountered (Amy et al., 1987). Of the 2085 THM and 557 HAA model/dataset combinations evaluated, only 12 (0.5%) combinations satisfied all constraints of model parameter boundary conditions. For these 12 datasets, the average R^2 value was 0.59 (ranging from 0.06 to 0.87), which supports the expectation that improved model accuracy is achieved for data within the boundary conditions. Notably, the majority of the datasets satisfying the boundary conditions were for models that did not report all conditions used to develop specific parameters, particularly Br^-.

Based on this analysis, it appears that many models are so specific in their application, based on the boundary conditions implemented during model development, that extrapolation is necessary for the vast majority of water systems. The inaccuracy this introduces brings into question the usefulness of many of these models outside of the specific scenario for which they were derived. Furthermore, this study suggests that caution should be employed when using models and/or bulk organic indicators such as DOC, UV_{254}, and SUVA to assess AOP performance, even in cases of qualitative comparison, and specifically when operating under limited-input energy and/or chemical conditions. More accurate predictions of AOP performance can be expected through development of site-specific mechanistic (or at least semi-mechanistic) models which directly account for hydroxyl radical interactions with organic matter.

Acknowledgements

This research was supported by the National Science Foundation Water & Environmental Technology Center at Arizona State University and the Arizona Water Institute. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of others. The authors gratefully acknowledge the laboratory assistance and insight provided by Paul Westerhoff and Chao-An Chin from Arizona State University. Thanks are also extended to City of Scottsdale personnel including Susan Butler, Laura McCasland, Kathy Gettens, Carie Wilson, Binga Talabi, and Don Henderson. Finally, we would like to thank Tony Powell of Purifics for his support of the project and use of the Photo-Cat® Lab reactor.

References

Amy et al., 1987. G.L. Amy, P.A. Chadik, Z.K. Chowdhury. Developing models for predicting trihalomethane formation potential and kinetics. *J. Am. Water Works Assoc*., 79 (7) (1987), pp. 89–97

Amy et al., 1998. Amy, G., Siddiqui, M., Ozekin, K., Zhu, H.W., Wang, C., 1998. Empirically Based Models for Predicting Chlorination and Ozonation by-products: Trihalomethanes, Haloacetic Acids, Chloral Hydrate, and Bromate. USEPA, EPA 815-R-98, Cincinnati, OH.

APHA, 2005. APHA. A.D. Eaton, L.S. Clescen, E.W. Rice, A.E. Greenberg (Eds.), Standard Methods for the Examination of Water and Wastewater, American Public Health Association, Washington, D.C. (2005)

Bellar et al., 1974. T.A. Bellar, J.J. Lichtenberg, R.C. Kroner. The occurrence of organohalids in chlorinated drinking water. *J. Am. Water Works Assoc.,* 66 (12) (1974), pp. 703–706

Dotson et al., 2010. A.D. Dotson, V.O.S. Keen, D. Metz, K.G. Linden. UV/H_2O_2 treatment of drinking water increases post-chlorination DBP formation. *Water Res.,* 44 (12) (2010), pp. 3703–3713

Gallard and von Guten, 2002. H. Gallard, U. von Guten. Chlorination of natural organic matter: kinetics of chlorination and of THM formation. *Water Res.,* 36 (1) (2002), pp. 65–74

Gerrity et al., 2009. D. Gerrity, B. Mayer, H. Ryu, J. Crittenden, M. Abbaszadegan. A comparison of pilot-scale photocatalysis and enhanced coagulation for disinfection byproduct mitigation. *Water Res.,* 43 (6) (2009), pp. 1597–1610

Huang et al., 2008. X. Huang, M. Leal, Q. Li. Degradation of natural organic matter by TiO_2 photocatalytic oxidation and its effect on fouling of low-pressure membranes. *Water Res.,* 42 (4–5) (2008), pp. 1142–1150

Kleiser and Frimmel, 2000. G. Kleiser, F.H. Frimmel. Removal of precursors for disinfection by-products (DBPs) – differences between ozone- and OH-radical-induced oxidation. *Sci. Total Environ.,* 256 (1) (2000), pp. 1–9

Krasner et al., 2006. S.W. Krasner, H.S. Weinberg, S.D. Richardson, S.J. Pastor, R. Chinn, M.J. Sclimenti, G.D. Onstad, A.D. Thruston. Occurrence of a new generation of disinfection byproducts. *Environ. Sci. Technol.,* 40 (23) (2006), pp. 7175–7185

Kulkarni and Chellam, 2010. P. Kulkarni, S. Chellam. Disinfection by-product formation following chlorination of drinking water: artificial neural network models and changes in speciation with treatment. *Sci. Total Environ.,* 408 (19) (2010), pp. 4202–4210

Liu et al., 2008a. S. Liu, M. Lim, R. Fabris, C. Chow, K. Chiang, M. Drikas, R. Amal. Removal of humic acid using TiO_2 photocatalytic process – fractionation and molecular weight characterisation studies. *Chemosphere*, 72 (2) (2008), pp. 263–271

Liu et al., 2008b. S. Liu, M. Lim, R. Fabris, C. Chow, M. Drikas, R. Amal. TiO_2 photocatalysis of natural organic matter in surface water: impact on trihalomethane and haloacetic acid formation potential. *Environ. Sci. Technol.,* 42 (16) (2008), pp. 6218–6223

Mayer et al., 2014. B.K. Mayer, E. Daugherty, M. Abbaszadegan. Disinfection byproduct formation resulting from settled, filtered, and finished water treated by titanium dioxide photocatalysis. *Chemosphere*, 117 (2014), pp. 72–78

Metz et al., 2011. D.H. Metz, M. Meyer, A. Dotson, E. Beerendonk, D.D. Dionysiou. The effect of UV/H_2O_2 treatment on disinfection by-product formation potential under simulated distribution system conditions. *Water Res.,* 45 (13) (2011), pp. 3969–3980

Plewa et al., 2002. M.J. Plewa, Y. Kargalioglu, D. Vankerk, R.A. Minear, E.D. Wagner. Mammalian cell cytotoxicity and genotoxicity analysis of drinking water disinfection by-products. *Environ. Mol. Mutagen.,* 40 (2) (2002), pp. 134–142

Richardson, 2005. S.D. Richardson. New disinfection by-product issues: emerging DBPs and alternative routes of exposure. *Glob. NEST*, 7 (1) (2005), pp. 43–60

Richardson et al., 1999. S.D. Richardson, A.D. Thruston, T.V. Caughran, P.H. Chen, T.W. Collette, T.L. Floyd, K.M. Schenck, B.W. Lykins, G. Sun, G. Majetich. Identification of new drinking water disinfection byproducts formed in the presence of bromide. *Environ. Sci. Technol*., 33 (19) (1999), pp. 3378–3383

Rook, 1974. J.J. Rook. Formation of halogens during the chlorination of natural water. *Water Treat. Exam*., 23 (2) (1974), pp. 234–243

Sarathy and Mohseni, 2010. S. Sarathy, M. Mohseni. Effects of UV/H_2O_2 advanced oxidation on chemical characteristics and chlorine reactivity of surface water natural organic matter. *Water Res.,* 44 (14) (2010), pp. 4087–4096

Sohn et al., 2004. J. Sohn, G. Amy, J. Cho, Y. Lee, Y. Yoon. Disinfectant decay and disinfection by-products formation model development: chlorination and ozonation by-products. *Water Res.,* 38 (10) (2004), pp. 2461–2478

USEPA, 2006. USEPA, 2006. National Primary Drinking Water Regulations: Stage 2 Disinfectants and Disinfection Byproducts Rule. 40 CR Parts 9, 141 and 142. Federal Register 71(2), 38.

Wei et al., 2009. L.-L. Wei, Q.-L. Zhao, S. Xue, C.-C. Chang, F. Tang, G.-L. Liang, T. Jia. Reduction of trihalomethane precursors of dissolved organic matter in the secondary effluent by advanced treatment processes. *J. Hazard. Mater*., 169 (1–3) (2009), pp. 1012–1021

WHO, 2000. WHO, 2000. Environmental Health Criteria 216: Disinfectants and Disinfectant By-Products. Geneva. <http://whqlibdoc.who.int/ehc/WHO_EHC_216.pdf>.

Zhang et al., 2008. Y. Zhang, A. Griffin, M. Edwards. Nitrification in premise plumbing: role of phosphate, pH and pipe corrosion. *Environ. Sci. Technol*., 42 (12) (2008), pp. 4280–4284

Zhang et al., 2011. J. Zhang, J. Yu, W. An, J. Liu, Y. Wang, Y. Chen, J. Tai, M. Yang. Characterization of disinfection byproduct formation potential in 13 source waters in China. *J. Environ. Sci*., 23 (2) (2011), pp. 183–188

Corresponding author at: Department of Civil, Construction and Environmental Engineering, Marquette University, P.O. Box 1881, Milwaukee, WI 53201-1881, United States. Tel.: +1 414 288 2161; fax: +1 414 288 7521.

Appendix A. Supplementary material

Supplementary Information
for:
Evaluation of the relationship between bulk organic precursors and disinfection byproduct formation for advanced oxidation processes

Brooke K. Mayer[1*], Erin Daugherty[2], and Morteza Abbaszadegan[2]

[1]Marquette University, P.O. Box 1881, Milwaukee, WI 53201-1881
[2]Arizona State University, P.O. Box 875306, Tempe, AZ 85287-5306

*Corresponding Author: Department of Civil, Construction and Environmental Engineering, Marquette University, P.O. Box 1881, Milwaukee, WI 53201-1881, Phone: 1-414-288-2161, Fax: 1-414-288-7521, E-mail: Brooke.Mayer@marquette.edu

The supplementary information consists of 67 pages, including 3 figures and 18 tables.

Supplementary Information

The Supplementary Information provides additional relevant details of this study, including:

- List of several significant disinfection byproduct (DBP) regulations
- Description of treatment trains for source waters used in photocatalysis experiments
- Identification of all external DBP datasets mined from the literature and any assumptions employed when using them
- List of all DBP models and the conditions under which they were developed
- Illustrations of the evolutions of DBPs as a function of time and extent of photocatalytic treatment
- Linear correlation coefficients for datasets grouped by type of source water
- Table of all coefficient of determinations (R^2) calculated for model/dataset pairings
- Full numerical statistics for tests of normal distributions, analysis of variance, and post hoc analyses.

1. Regulated Disinfection Byproducts

The 20th century heralded substantial public health improvements through the advent of widespread drinking water disinfection. Yet, in the last several decades, concern over the production of disinfection byproducts (DBPs) formed through the interactions of oxidizing disinfectants and natural organic matter (NOM) has stimulated increasing regulations on levels of residual disinfectants in distribution systems as well as maximum contaminant levels for DBPs. Table S1 lists several current DBP regulations established by regulatory agencies.

Table S1 – Regulated disinfection byproducts (all concentrations in μg L^{-1})

Species	Chemical Formula	Shorthand Notation	Main Causative Disinfectant	U.S. Environmental Protection Agency (USEPA)[a]	Health Canada (HC)[b]	European Union (EU)[c]	Australia[d]	World Health Organization (WHO)[e]
Total Trihalomethanes	--	TTHMs	Chlorine	80	100	100	250	[f]
Chloroform	$CHCl_3$	CF	Chlorine	--	--	--	--	300
Bromodichloromethane	$CHBrCl_2$	BDCM	Chlorine	--	16	--	--	60
Dibromochloromethane	$CHBr_2Cl$	DBCM	Chlorine	--	--	--	--	100
Bromoform	$CHBr_3$	BF	Chlorine	--	--	--	--	100
Haloacetic Acids	--	HAA5s	Chlorine	60	80	--	--	--
Monochloroacetic Acid	$C_2H_3ClO_2$	CAA	Chlorine	--	--	--	150	20
Dichloroacetic Acid	$C_2H_2Cl_2O_2$	DCA	Chlorine	--	--	--	100	50
Trichloroacetic Acid	$C_2HCl_3O_2$	TCA	Chlorine	--	--	--	100	200
Monobromoacetic Acid	$C_2H_3BrO_2$	MBA	Chlorine	--	--	--	--	--
Dibromoacetic Acid	$C_2H_2Br_2O_2$	DBA	Chlorine	--	--	--	--	--
Bromate	BrO_3	--	Ozone	10	10	10	20	10
Chlorite	ClO_2^-	--	Chlorine dioxide	1,000	1,000	--	800	700
Chlorate	ClO_3^-	--	Chlorine dioxide	--	1,000	--	--	700
Chloral hydrate (Trichloroacetaldehyde)	$C_2H_3Cl_3O_2$	CH	Chlorine	--	--	--	20	--
Chlorophenols	--	--	Chlorine	--	--	--	0.1	--
2-Chlorophenol	C_6H_5ClO	--	Chlorine	--	--	--	300	--
2,4-Dichlorophenol	$C_6H_4Cl_2O$	--	Chlorine	--	--	--	200	--
2,4,6-Trichlorophenol	$C_6H_3Cl_3O$	--	Chlorine	--	5	--	20	200
Cyanogen Chloride (as cyanogen)	CNCl	--	Chloramines	--	--	--	80	70
Formaldehyde	CH_2O	--	Chlorine, Ozone	--	--	--	500	--
Haloacetonitriles	--	HANs	Chlorine	--	--	--	--	--
Dichloroacetonitrile	C_2HCl_2N	DCAN	Chlorine	--	--	--	--	20
Dibromoacetonitrile	C_2HBr_2N	DBAN	Chlorine	--	--	--	--	70
N-Nitrosodimethylamine	$C_2H_6N_2O$	NDMA	Chloramines	--	0.04	--	--	0.1

[a] USEPA, 2009

[b] Health Canada, 2012

[c] EU, 1998

[d] NHMRC and NRMMC, 2011

[e] WHO, 2008

[f] The sum of the ratio of the concentration of each to its respective guideline value should not exceed 1

2. Source Water for Pilot-scale TiO_2 Photocatalysis Experiments

As described by Mayer et al. (2014), the settled water (SW) was collected following coagulation using approximately 12 mg L^{-1} alum ($Al_2(SO_4)_3$), flocculation, and sedimentation with a hydraulic residence time of 97 min. The granular activated carbon-filtered water (GAC) was collected from a direct filtration system consisting of microfiltration followed by GAC filtration operated at a loading rate of 2.4 gpm ft^{-2} and an empty bed contact time of 17 min. The distribution system water (DS) was collected from a point in the direct filtration plant's distribution system with a water age of approximately 45 hr and a residual chlorine concentration of 0.3 mg L^{-1}.

3. DBP Datasets

The sources for the 20 external DBP datasets mined from the literature are described in Tables S2 and S3. The data included TiO_2 photocatalysis, non-TiO_2 advanced oxidation processes (AOPs), and non-AOP treatments. The source water and water quality parameters characterizing the datasets encompass a wide range of values in an effort to evaluate effectiveness of model predictions over a wide range of scenarios. As noted by Amy et al. (1987), the ideal literature source in which all relevant modeling parameters, operating conditions, and resulting DBP concentrations were clearly identified and quantified, was rarely encountered. Thus, rational assumptions were made in an attempt to faithfully represent the data. Notations are made in the table regarding all data-related assumptions. These assumptions introduced inherent inaccuracies through estimates of graphical presentations, assumptions for parameters and operating conditions not clearly identified and/or quantified, etc. Notably, bromide concentrations were often unreported. When possible, external literature searches were used to quantify typical bromide content of the source water. In several cases, the bromide concentration was reportedly negligible (or assumed to be based on limited influence of saline water). In these scenarios, the bromide ion concentration was recorded as 1 µg L^{-1}, which is negligible, but enables model computations that would otherwise return a null output using bromide ion concentrations of zero. Similar approaches are reported by Amy et al. (1998) and Chen and Westerhoff (2010).

Table S2 – Description of advanced oxidation process (AOP) DBP datasets used in this comparative study

Source	Process Description	Source Water	DBP	Comments
Current Study	TiO_2 photocatalysis using low-pressure UV, 1 g L^{-1} TiO_2	Settled, GAC-filtered, and distribution system water from central AZ, U.S.	TTHM, HAA5, subspecies	For UV_{254} below the detection limit, a value of 0.00001 cm^{-1} was used to avoid null model outputs.[b]
(Chin and Bérubé, 2005)	O_3, UV, UV/ O_3 (low-pressure UV lamp)	Seymour Reservoir, BC, Canada	CF, HAA5	Based on reported 8-d chlorine residual of 10 – 20 mg L^{-1}, estimated excess D = 25 mg L^{-1}.[a, b, c]
(Dotson et al., 2010)	Low-pressure and medium-pressure UV with H_2O_2	Sand and GAC-filtered water from Cincinnati, OH, U.S.	TTHM	Br^- and pH were assumed to stay constant through AOP treatment.[a,c]
(Gerrity et al., 2009)	TiO_2 photocatalysis using low-pressure UV, 1 g L^{-1} TiO_2	Raw surface water from central Arizona, U.S.	TTHM, subspecies	For UV_{254} below the detection limit, a value of 0.00001 cm^{-1} was used to avoid null model outputs.[b]
(Kent et al., 2011)	TiO_2 photocatalysis using low-pressure UV, 1 g L^{-1} sputtered or suspended TiO_2	French River Water, NS, Canada	TTHM, HAA5, subspecies	Residual chlorine after 24 hr target of 1 mg L^{-1}, so D was estimated as being equal to DOC. Temperature not given; assumed to be 20°C. Brookman et al. (2011) reported no Br^- in source water.[a, b]
(Kleiser and Frimmel, 2000)	O_3/H_2O_2, UV/H_2O_2 (low-pressure UV lamp)	River Ruhr, Germany	TTHM	Temperature not given; assumed to be 20°C.[a, b]
(Lamsal et al., 2011)	UV, O_3/H_2O_2, UV/H_2O_2, O_3/UV using low-pressure UV	French River, Nova Scotia, Canada	TTHM, HAA5	Residual 24 hr chlorine target of 1 mg L^{-1}, so D was estimated as being equal to DOC. Two non-reported DOC values were estimated using nearby data. Brookman et al. (2011) reported no Br^- in source water.[b]
(Liu et al., 2008a)	TiO_2 photocatalysis using blacklight blue fluorescent lamps, 0.1 g L^{-1} TiO_2, and combined UV TiO_2/H_2O_2 treatment	Synthetic – humic acid	CF	Followed Method 5710b. Reported a target residual of 3 – 5 mg L^{-1} after 7 d. Liu et al. (2008b, 2007) used D ≈ 25 mg/L, so this value was assumed.[a, b]
(Liu et al., 2008b)	TiO_2 photocatalysis using using blacklight blue fluorescent lamps, 0.1 g L^{-1} TiO_2, and combined UV TiO_2/H_2O_2 treatment	Myponga Reservoir in South Australia	TTHM, HAA5	D reported was 20 – 30 mg L^{-1}, so used average of 25 mg L^{-1}. Br^- not reported, so used value reported by Chow et al. (2004) for the Myponga Reservoir.[a]
(Liu et al., 2010)	TiO_2 photocatalysis using blacklight blue fluorescent lamps, 0.1 g L^{-1} TiO_2	Surface water from the Myponga Reservoir and Copi Hollow, Australia	TTHM	D reported was 20 – 30 mg L^{-1}, so used average of 25 mg L^{-1}. Br^- not reported, so used value reported by Chow et al. (2004) for the Myponga Reservoir.
(Mayer et al., 2014)	TiO_2 photocatalysis using low-pressure UV, 1 g L^{-1} TiO_2	Settled, GAC-filtered, and distribution system water from central AZ, U.S.	TTHM, HAA5, subspecies	For UV_{254} below the detection limit, a value of 0.00001 cm^{-1} was used to avoid null model outputs.[b]

Source	Process Description	Source Water	DBP	Comments
(Philippe et al., 2010a)	TiO_2 photocatalysis using medium-pressure UV, 1 g L^{-1} TiO_2	Raw water from the Severn Trent region of the UK	TTHM	D was equal to 5xs DOC, which was read from plots. No Br^- was reported, so value reported by Brown (2009) was used.[a]
(Philippe et al., 2010b)	TiO_2 photocatalysis using medium-pressure UV, 1 g L^{-1} TiO_2	Synthetic – amino acids	CF	D was equal to 5xs DOC, which was read from plots.[a, b]
(Liu et al., 2007)	UV/H_2O_2, UV/Fe^{+3}, UV/H_2O_2/Fe^{+3} using blacklight blue fluorescent lamps	Synthetic – humic acid	TTHM	Reported excess chlorine dose. Liu et al. (2008b, 2007) used D = 25 mg L^{-1}, so this value was assumed.[a, b]
(Sarathy and Mohseni, 2010)	Low-pressure, high output lamp for UV/H_2O_2 with 15 mg L^{-1} H_2O_2	Surface water from the Capilano Reservoir, BC, Canada	TTHM, HAA5	[a, b, c]

[a] At least some values estimated from graphical data presentations.

[b] Bromide reported as zero or assumed to be negligible, so concentration of 1 μg L^{-1} used to avoid null model outputs.

[c] TOC concentrations reported rather than DOC.

Table S3 – Description of non-AOP DBP datasets used in this comparative study

Source	Process Description	Source Water	DBP	Comments
(Chowdhury et al., 2010)	Untreated water	Synthetic water samples of differing characteristics	TTHM, subspecies	a
(Elshorbagy, 2000)	Finished (desalinated) water	Field samples from Abu-Dhabi, UAE	TTHM	a, c
(Fabbricino and Korshin, 2009)	Untreated water	Seawater from Naples, Italy	HAA5, subspecies	Reported in μmol L-1 and converted to μg L-1.a
(Gang et al., 2003)	Raw and treated water	Surface water from the Mississippi River, Garden City, Maysville, and Lake Vandalia, U.S.	TTHM	a,b
(Hua and Reckhow, 2008)	Raw water	Surface water from Cambridge, MA, U.S.	TTHM	a
(Zhang et al., 2011)	Raw water	Surface water from China	TTHM	

a At least some values estimated from graphical data presentations.

b Bromide reported as zero or assumed to be negligible, so concentration of 1 μg L-1 used to avoid null model outputs.

c TOC concentrations reported rather than DOC.

4. DBP Models

In recent decades, hundreds of predictive DBP formation models have been published (refer to reviews by Chowdhury et al. (2009) and Sadiq and Rodriguez (2004)). They have been developed to qualify the relative significance of water quality and operational disinfection parameters, investigate kinetics, provide alternatives to field monitoring of DBPs, aid in assessments of exposure and health risk, and support comparative economic/performance decisions (Amy et al., 1987; Hong et al., 2007; Sadiq and Rodriguez, 2004). The sheer number of models exemplifies the challenge of attempting to develop a universally applicable model (Golfinopoulos and Arhonditsis, 2002). Mechanistic DBP 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 NOM (Kulkarni and Chellam, 2010). Accordingly, most modelers have attempted to establish statistically-based empirical and/or semi-mechanistic linear or nonlinear regression models using commonly measured water quality parameters and operating conditions (Sadiq and Rodriguez, 2004).

The 149 published chlorine-DBP models identified for use in this study are listed in Table S4. The models all used variations of common DBP indicators such as dissolved organic carbon (DOC), ultraviolet absorbance at a wavelength of 254 nm (UV_{254}), specific UV absorbance (SUVA), chlorine dose (D), pH, temperature (T), bromide ion concentration (Br^-), and disinfection contact time (t) to predict mass-based concentrations of total trihalomethanes (TTHMs), haloacetic acids (HAA5s), and subspecies thereof. Models incorporating other parameters, such as chlorophyll (Golfinopoulos and Arhonditsis, 2002; Golfinopoulos et al., 1998); fulvic acid (Rodrigues et al., 2007); specific rate constants (Chang et al., 2006; Fabbricino and Korshin, 2009); nitrogen, ammonia, or combined chlorine (Espigares et al., 2003); or site-specific parameters such as DBP yield coefficients or THM formation potential (Amy et al., 1998; Gang et al., 2002; Li and Zhao, 2006) were not included. This dictates that the predictive results reported here are purely empirically-derived due to the exclusion of semi-mechanistic models incorporating kinetic rate constants, species-specific reactivity parameters, artificial neural network modeling, etc. (Kulkarni and Chellam, 2010; Milot et al., 2002; Westerhoff et al., 2000). Exclusion of these more advanced models was obligatory since site-specific rates were not readily available for the DBP datasets and artificial neural networks are not generalizable models that can be expressed in mathematical form, but rather are designed to adapt to specific datasets through model training.

As part of the modeling effort, total organic carbon (TOC) and DOC were treated synonymously as they typically differ by less than 5%, and can be used interchangeably with little error (Owen et al., 1993; Rathbun, 1996a). Table S4 lists the NOM parameter used to develop the model (e.g., TOC or DOC), but when data was substituted into the model's NOM placeholder, DOC (as reported in the data sets) was treated as equivalent to TOC and vice versa. The conditions for model development are listed in Table S5. Empirical models such as these are typically recommended for application within the boundary conditions defined by the parameters from each specific study.

5. Time Evolution of DBPs

The 24-hr evolution of TTHMs and HAA5s is illustrated in Figure S1 as a function of the level of photocatalytic treatment and type of source water. Figure S2 shows the time evolution of the 4 subspecies of TTHMs and 5 subspecies of HAA5s for the SW, GAC, and DS, as a function of the level of photocatalytic treatment.

Table S4 – Published DBP models considered as part of this study

#	Source	Model
Total Trihalomethanes = TTHM = CF + BDCM + DBCM + BF		
1	(Al-Omari et al., 2004)	$\text{TTHM} = 4.527(t/60)^{0.127}(D)^{0.595}(\text{TOC})^{0.596}(\text{Br}^{-}/1000)^{0.103}(\text{pH})^{0.66}$
2	(Amy et al., 1998)	$\text{TTHM} = 10^{-1.385}(\text{DOC})^{1.098}(D)^{0.152}(\text{Br}^{-})^{0.068}(T)^{0.609}(\text{pH})^{1.601}(t)^{0.263}$
3	(Amy et al., 1998)	$\text{TTHM} = 10^{0.651}(\text{DOC})^{0.752}(D)^{0.246}(\text{Br}^{-})^{0.185}(t)^{0.258}$
4	(Amy et al., 1998)	$\text{TTHM} = 10^{0.387}(\text{DOC})^{0.839}(D)^{0.287}(\text{Br}^{-})^{0.259}(t)^{0.270}$
5	(Amy et al., 1998)	$\text{TTHM} = 10^{0.518}(\text{DOC})^{0.801}(D)^{0.261}(\text{Br}^{-})^{0.223}(t)^{0.264}$
6	(Boyalla, 2004)	$\text{TTHM} = 0.0001(D)^{3.14}(\text{pH})^{1.56}(\text{TOC})^{0.69}(t)^{0.175}$
7	(Chang et al., 1996)	$\text{TTHM} = 108.80(\text{TOC})^{0.2466}(t)^{0.2956}(\text{UV}_{254})^{0.9919}(D)^{0.126}$
8	(Chang et al., 1996)	$\text{TTHM} = 12.7254(\text{TOC})^{0.2910}(t)^{0.2706}(D)^{-0.0719}$
9	(Chang et al., 1996)	$\text{TTHM} = 131.7492(t)^{0.2931}(\text{UV}_{254})^{1.075}(D)^{0.1064}$
10	(Chen and Westerhoff, 2010)	$\text{TTHM} = 1147(\text{DOC})^{0.00}(\text{UV}_{254})^{0.83}(\text{Br}^{-}+1)^{0.27}$
11	(Harrington et al., 1992)	$\text{TTHM} = [0.00309(\text{DOC}*\text{UV}_{254})^{0.440}(D)^{0.409}(t)^{0.265}(T)^{1.06}(\text{pH}-2.6)^{0.715}(\text{Br}^{-}/1000+1)^{0.036}] * [106(\text{Br}^{-}/1000+1)^{0.48}(\text{UV}_{254})^{-0.089}]$
12	(Hong et al., 2007)	$\text{TTHM} = 10^{-1.375}(t)^{0.258}(D/\text{DOC})^{0.194}(\text{pH})^{1.695}(T)^{0.507}(\text{Br}^{-})^{0.218}$
13	(McBean et al., 2008)	$\text{TTHM} = 10^{0.715}(\text{DOC}_{\text{raw}})^{0.322}(\text{DOC}_{\text{treated}})^{0.761}(D_{\text{pre}})^{0.206}(D_{\text{post}})^{0.184}(T)^{0.204}$
14	(Milot et al., 2002)	$\text{TTHM} = 0.032(\text{DOC})^{1.053}(t)^{0.263}(\text{pH})^{1.168}(D)^{0.275}(T)^{1.040}$
15	(Rathbun, 1996a)	$\text{TTHM} = 14.6(\text{pH}-3.8)^{1.01}(D)^{0.206}(\text{UV}_{254}/2)^{0.849}(t)^{0.306}$
16	(Rodriguez et al., 2000)	$\text{TTHM} = 0.044(\text{DOC})^{1.030}(t)^{0.262}(\text{pH})^{1.149}(D)^{0.277}(T)^{0.968}$
17	(Rodriguez et al., 2000)	$\text{TTHM} = 1.392(\text{DOC})^{1.092}(\text{pH})^{0.531}(T)^{0.255}$
18	(Semerjian et al., 2009)	$\text{TTHM} = -9.22+360.97(\text{UV}_{254})+0.54(T)+0.12(t)$
19	(Semerjian et al., 2009)	$\text{TTHM} = 40.44(\text{UV}_{254})^{0.41}(\text{Br}^{-}/1000)^{0.44}$
20	(Semerjian et al., 2009)	$\text{TTHM} = \text{sqrt}[17.31+10.52(D)^{2}+259728.60(\text{SUVA})^{2}]$
21	(Semerjian et al., 2009)	$\text{TTHM} = \exp[-0.21+62.43(\text{UV}_{254})+0.06(T)]$
22	(Semerjian et al., 2009)	$\text{TTHM} = 4.01+319.88(\text{UV}_{254})+2.42(D)$
23	(Semerjian et al., 2009)	$\text{TTHM} = 8.08(D)^{0.41}$
24	(Semerjian et al., 2009)	$\text{TTHM} = \text{sqrt}[42.10+29.23(D)^{2}+353375.0(\text{UV}_{254})^{2}]$
25	(Semerjian et al., 2009)	$\text{TTHM} = \exp[1.53+42.21(\text{UV}_{254})]$
26	(Semerjian et al., 2009)	$\text{TTHM} = -12.66+0.48(t)+35.06(\text{Br}^{-}/1000)+10.26(D)$
27	(Semerjian et al., 2009)	$\text{TTHM} = 450.3(t)^{0.21}(\text{Br}^{-}/1000)^{1.64}(D)^{0.34}(T)^{0.801}$
28	(Semerjian et al., 2009)	$\text{TTHM} = \text{sqrt}[-471.11+0.48(t)^{2}+1856.07(\text{Br}^{-}/1000)^{2}+404.38(D)^{2}]$
29	(Semerjian et al., 2009)	$\text{TTHM} = \exp[1.43+0.02(t)-0.04(T)+2.78(\text{Br}^{-}/1000)+0.48(D)]$
30	(Sérodes et al., 2003)	$\text{TTHM} = 16.9+16.0(\text{TOC})+3.319(D)-1.135(T)+1.139(t)$
31	(Sérodes et al., 2003)	$\text{TTHM} = -0.101(\text{TOC})^{3.914}(t)^{0.117}$
32	(Sérodes et al., 2003)	$\text{TTHM} = 21.2+2.447(D)+0.449(t)$
33	(Sohn et al., 2004)	$\text{TTHM} = 75.7(\text{UV}_{254})^{0.593}(D)^{0.332}(\text{Br}^{-})^{0.060}(t)^{0.264}$
34	(Sohn et al., 2004)	$\text{TTHM} = 23.9(\text{DOC}*\text{UV}_{254})^{0.403}(D)^{0.225}(\text{Br}^{-})^{0.141}(t)^{0.264}$
35	(Sohn et al., 2004)	$\text{TTHM} = [10^{0.518}(\text{DOC})^{0.801}(D)^{0.261}(\text{Br}^{-})^{0.223}(t)^{0.264}]*(1.156)^{(\text{pH}-7.5)}(1.0263)^{(T-20)}$

#	Source	Model
36	(Sohn et al., 2004)	$TTHM = 0.00253(DOC)^{1.22}(D)^{0.442}(T)^{1.34}(pH)^{1.75}(t)^{0.34}$
37	(Sohn et al., 2004)	$TTHM = 0.012(DOC*UV_{254})^{0.47}(D)^{0.48}(T)^{1.10}(pH)^{2.38}(t)^{0.35}$
38	(Toroz and Uyak, 2005)	$TTHM = 11.967(TOC)^{0.398}(T)^{0.158}(D)^{0.702}$
39	(Urano et al., 1983)	$TTHM = 8.2x10^{-4}(pH-2.8)(TOC)(D)^{0.25}(t)^{0.36}$
40	(Uyak et al., 2005)	$TTHM = 0.0707(TOC+3.2)^{1.314}(pH-4.0)^{1.496}(D-2.5)^{-0.197}(T+10)^{0.724}$
41	(USEPA, 2001)[b]	$TTHM = [23.9(DOC*UV_{254})^{0.403}(D)^{0.225}(Br^{-})^{0.141}(t)^{0.264}]*(1.1560)^{(pH-7.5)}(1.0263)^{(T-20)}$
42	(USEPA, 2001)	$TTHM = 17.7(DOC*UV_{254})^{0.475}(D)^{0.173}(Br^{-})^{0.246}(1.316)^{(pH-8.0)}(1.036)^{(T-20)}(t)^{0.366}$
43	(Zhu, 1995)[a]	$TTHM = -480.51+36.67(DOC)+4.92(D)+0.16(Br^{-})+5.70(T)+33.45(pH)+0.92(t)$
44	(Zhu, 1995)	$TTHM = -826.40+36.76(DOC)+4.96(D)+0.15(Br^{-})+261.1*\log(T)+33.27(pH)+66.75(t)^{0.267}$
45	(Zhu, 1995)	$TTHM = 10^{-1.385}(DOC)^{1.098}(D)^{0.152}(Br^{-})^{0.068}(T)^{0.609}(pH)^{1.601}(t)^{0.263}$
46	(Zhu, 1995)	$TTHM = 10^{-0.377}(UV_{254})^{0.482}(D)^{0.339}(Br^{-})^{0.023}(T)^{0.617}(pH)^{1.609}(t)^{0.261}$
47	(Zhu, 1995)	$TTHM = 10^{-0.549}(DOC*UV_{254})^{0.421}(D)^{0.145}(Br^{-})^{0.041}(T)^{0.614}(pH)^{1.606}(t)^{0.261}$
Individual Trihalomethane Species		
48	(Amy et al., 1998)	$CF = 10^{-1.205}(DOC)^{1.617}(D)^{-0.094}(Br^{-})^{-0.175}(T)^{0.607}(pH)^{1.403}(t)^{0.306}$
49	(Amy et al., 1998)	$BDCM = 10^{-2.874}(DOC)^{0.901}(D)^{0.017}(Br^{-})^{0.733}(T)^{0.498}(pH)^{1.511}(t)^{0.199}$
50	(Amy et al., 1998)	$DBCM = 10^{-5.649}(DOC)^{-0.226}(D)^{0.108}(Br^{-})^{1.81}(T)^{0.512}(pH)^{2.212}(t)^{0.146}$
51	(Amy et al., 1998)	$BF = 10^{-7.83}(DOC)^{-0.983}(D)^{0.804}(Br^{-})^{1.765}(T)^{0.754}(pH)^{2.139}(t)^{0.566}$
52	(Amy et al., 1998)	$CF = 10^{1.331}(DOC)^{1.11}(D)^{0.324}(Br^{-})^{-0.532}(t)^{0.341}$
53	(Amy et al., 1998)	$BDCM = 10^{-0.203}(DOC)^{0.504}(D)^{0.126}(Br^{-})^{0.474}(t)^{0.187}$
54	(Amy et al., 1998)	$DBCM = 10^{-5.398}(DOC)^{0.943}(D)^{-0.228}(Br^{-})^{2.678}(t)^{0.175}$
55	(Amy et al., 1998)	$BF = 10^{-9.6}(DOC)^{0.203}(D)^{-0.639}(Br^{-})^{4.487}(t)^{0.319}$
56	(Amy et al., 1998)	$CF = 10^{1.092}(DOC)^{1.179}(D)^{0.378}(Br^{-})^{-0.454}(t)^{0.326}$
57	(Amy et al., 1998)	$BDCM = 10^{-0.416}(DOC)^{0.599}(D)^{0.125}(Br^{-})^{0.533}(t)^{0.205}$
58	(Amy et al., 1998)	$DBCM = 10^{-5.127}(DOC)^{0.194}(D)^{0.433}(Br^{-})^{2.427}(t)^{0.294}$
59	(Amy et al., 1998)	$BF = 10^{-9.427}(DOC)^{-0.329}(D)^{-0.035}(Br^{-})^{4.335}(t)^{0.307}$
60	(Amy et al., 1998)	$CF = 10^{1.211}(DOC)^{1.149}(D)^{0.345}(Br^{-})^{-0.492}(t)^{0.333}$
61	(Amy et al., 1998)	$BDCM = 10^{-0.311}(DOC)^{0.556}(D)^{0.121}(Br^{-})^{0.505}(t)^{0.196}$
62	(Amy et al., 1998)	$DBCM = 10^{-5.248}(DOC)^{0.55}(D)^{0.105}(Br^{-})^{2.549}(t)^{0.234}$
63	(Amy et al., 1998)	$BF = 10^{-9.5}(DOC)^{-0.075}(D)^{-0.34}(Br^{-})^{4.409}(t)^{0.313}$
64	(Chen and Westerhoff, 2010)	$CF = 1805(DOC)^{0.11}(UV_{254})^{1.22}(Br^{-}+1)^{-2.19}$
65	(Chen and Westerhoff, 2010)	$BDCM = 137(DOC)^{0.16}(UV_{254})^{0.94}(Br^{-}+1)^{3.66}$
66	(Hong et al., 2007)	$BDCM = 10^{-3.201}(t)^{0.297}(pH)^{2.878}(t)^{0.414}(Br^{-})^{0.371}$
67	(Hong et al., 2007)	$CF = 10^{-0.748}(t)^{0.210}(D/DOC)^{0.221}(pH)^{1.374}(T)^{0.532}(Br^{-})^{-0.184}$
68	(Malcom Pirnie Inc, 1992)	$CF = 0.078(TOC*UV_{254})^{0.616}(D)^{0.391}(t)^{0.265}(T)^{1.15}(pH-2.6)^{0.8}$
69	(Malcom Pirnie Inc, 1992)	$BDCM = 0.863(TOC*UV_{254})^{0.177}(D)^{0.309}(t)^{0.271}(T)^{0.72}(pH-2.6)^{0.925}(Br^{-}+1)^{0.0722}$
70	(Malcom Pirnie Inc, 1992)	$DBCM = 2.57(UV_{254}/TOC)^{-0.184}(D)^{-0.0746}(t)^{0.57}(pH-2.6)^{1.35}(Br^{-}+1)^{2.08}$
71	(Malcom Pirnie Inc, 1993)	$CF = 0.997(TOC)^{0.580}(UV_{254})^{0.580}(D)^{0.814}(t)^{0.278}(Br^{-}+1)^{-4.27}(T)^{0.569}(pH-2.6)^{0.759}$
72	(Montgomery Watson, 1993)	$CF = 0.064(TOC)^{0.329}(UV_{254})^{0.874}(Br-+0.01)^{0.404}(pH)^{1.161}(D)^{0.561}(t)^{0.269}(T)^{1.018}$
73	(Montgomery Watson, 1993)	$BDCM = 0.0098(Br^{-})^{0.181}(pH)^{2.55}(D)^{0.497}(t)^{0.256}(T)^{0.519}$

#	Source	Model
74	(Montgomery Watson, 1993)	$BDCM = 1.325(TOC)^{-0.725}(Br^{-})^{0.794}(D)^{0.632}(t)^{0.204}(T)^{1.441}$
75	(Montgomery Watson, 1993)	$DBCM = 14.998(TOC)^{-1.665}(Br^{-})^{1.241}(D)0.729(t)^{0.261}(T)^{0.989}$
76	(Montgomery Watson, 1993)	$DBCM = 0.028(UV_{254})^{-1.175}(TOC)^{-1.078}(Br^{-})^{1.573}(pH)^{1.956}(D)^{1.072}(t)^{0.2}(T)^{0.596}$
77	(Montgomery Watson, 1993)	$BF = 6.533(TOC)^{-2.031}(Br^{-})^{1.388}(pH)^{1.603}(D)^{1.057}(t)^{0.136}$
78	(Rathbun, 1996b)	$CF = 0.442(pH)^{2}(D)^{0.229}(DOC)^{0.912}(Br^{-}/1000)^{-0.116}$
79	(Rathbun, 1996b)	$BDCM = 17.5(pH)^{1.01}(D)^{0.0367}(DOC)^{0.228}(Br^{-}/1000)^{0.513}$
80	(Rathbun, 1996b)	$DBCM = 26.6(pH)^{1.80}(D)^{-0.0928}(DOC)^{-0.758}(Br^{-}/1000)^{1.20}$
81	(Rathbun, 1996b)	$BF = 0.290(pH)^{3.51}(D)^{-0.347}(DOC)^{-0.330}(Br^{-}/1000)^{1.84}$
82	(USEPA, 2001)[b]	$CF = [266(DOC*UV_{254})^{0.403}(D)^{0.424}(Br^{-})^{-0.679}(t)^{0.333}]*(1.1322)^{(pH-7.5)}(1.0179)^{(T-20)}$
83	(USEPA, 2001)[b]	$BDCM = [1.68(DOC*UV_{254})^{0.260}(D)^{0.114}(Br^{-})^{0.462}(t)^{0.196}]*(1.0977)^{(pH-7.5)}(1.0260)^{(T-20)}$
84	(USEPA, 2001)[b]	$DBCM = [8*10^{-3}(DOC*UV_{254})^{-0.056}(D)^{-0.157}(Br^{-})^{1.425}(t)^{0.148}]*(1.1271)^{(pH-7.5)}(1.0212)^{(T-20)}$
85	(USEPA, 2001)[b]	$BF = [4.4*10^{-5} (DOC*UV_{254})^{-0.300}(D)^{-0.221}(Br^{-})^{2.134}(t)^{0.143}]*(1.3907)^{(pH-7.5)}(1.0374)^{(T-20)}$
86	(USEPA, 2001)	$CF = [101.0(DOC*UV_{254})^{0.615}(D)^{0.699}(Br^{-})^{-0.468}(t)^{0.336}]*(1.099)^{(pH-7.5)}(1.035)^{(T-20)}$
87	(USEPA, 2001)	$BDCM = [7.57(DOC*UV_{254})^{0.443}(D)^{0.563}(Br^{-})^{0.0739}(t)^{0.281}]*(1.355)^{(pH-7.5)}(1.030)^{(T-20)}$
88	(USEPA, 2001)	$DBCM = [3.99(DOC*UV_{254})^{0.535}(D)^{0.125}(Br^{-})^{365}(t)^{0.322}]*(1.436)^{(pH-7.5)}(1.037)^{(T-20)}$
89	(USEPA, 2001)	$BF = [1.47*10^{-1} (DOC*UV_{254})^{0.408}(D)^{-0.115}(Br^{-})^{0.961}(t)^{0.324}]*(1.438)^{(pH-7.5)}(1.048)^{(T-20)}$
90	(Zhu, 1995)	$CF = -366.42+31.03(DOC)+5.27(D)-0.23(Br^{-})+4.50(T)+25.99(pH)+0.72(t)$
91	(Zhu, 1995)	$BDCM = -71.05+6.66(DOC)-0.65(D)+0.19(Br^{-})+0.80(T)+4.15(pH)+0.12(t)$
92	(Zhu, 1995)	$DBCM = -33.88-0.41(DOC)+0.16(D)0.16(Br^{-})+0.33(T)+2.55(pH)+0.06(t)$
93	(Zhu, 1995)	$BF = -8.67-0.63(DOC)+0.14(D)+0.05(Br^{-})+0.07(T)+0.75(pH)+0.02(t)$
94	(Zhu, 1995)	$CF = -762.77+31.20(DOC)+5.12(D)+143.53\exp(-0.0029(Br^{-}))$ $+210.09*\log(T)+25.87(pH)+41.05(t)^{0.300}$
95	(Zhu, 1995)	$BDCM = -135.54+6.41(DOC)-3.93(D)^{0.310}+1.04*(3.20+0.35(Br^{-})-0.00039(Br^{-})^{2})$ $+34.16*\log(T)+4.12(pH)+16.06(t)^{0.194}$
96	(Zhu, 1995)	$DBCM = -54.52-0.39(DOC)+0.11(D)+ 1.01*(-2.99+0.10(Br^{-})-0.00011(Br^{-})^{2})$ $+13.16*\log(T)+2.53(pH)+11.73(t)^{0.154}$
97	(Zhu, 1995)	$BF = -7.51+0.62(DOC)+0.069(D)+0.99(1.34-0.23(Br^{-})-0.00015(Br^{-})^{2})$ $+2.76*\log(T)+0.75(pH)+0.062(t)^{0.764}$
98	(Zhu, 1995)	$CF = 10^{0.96}(UV_{254})^{0.604}(D)^{0.286}(Br^{-})^{-0.242}(T)^{0.619}(pH)^{1.416}(t)^{0.304}$
99	(Zhu, 1995)	$BDCM = 10^{-1.708}(UV_{254})^{0.590}(D)^{-0.022}(Br^{-})^{0.697}(T)^{0.505}(pH)^{1.519}(t)^{0.196}$
100	(Zhu, 1995)	$DBCM = 10^{-5.694}(UV_{254})^{-0.005}(D)^{-0.024}(Br^{1.820}(T)^{0.510}(pH)^{2.211}(t)^{0.146}$
101	(Zhu, 1995)	$BF = 10^{-9.267}(UV_{254})^{-0.722}(D)^{0.926}(Br^{-})^{1.804}(T)^{0.741}(pH)^{2.158}(t)^{0.571}$
102	(Zhu, 1995)	$CF = 10^{-0.040}(DOC*UV_{254})^{0.578}(D)^{-0.038}(Br^{-})^{-0.217}(T)^{0.615}(pH)^{1.412}(t)^{0.304}$
103	(Zhu, 1995)	$BDCM = 10^{-2.067}(DOC*UV_{254})^{0.422}(D)^{-0.111}(Br^{-})^{0.714}(T)^{0.502}(pH)^{1.515}(t)^{0.196}$
104	(Zhu, 1995)	$DBCM = 10^{-5.763} (DOC*UV_{254})^{-0.049}(D)^{0.051}(Br^{-})^{1.817}(T)^{0.511}(pH)^{2.211}(t)^{0.146}$
105	(Zhu, 1995)	$BF = 10^{-8.776}(DOC*UV_{254})^{-0.496}(D)^{0.996}(Br^{-})^{1.784}(T)^{0.744}(pH)^{2.146}(t)^{0.569}$
Haloacetic Acids = HAA5 = MCA + DCA + TCA + MBA + DBA		
106	(McBean et al., 2008)	$HAAs = 10^{3.333}(DOC_{raw})^{0.380}(DOC_{treated})^{0.774}(D_{pre})^{0.102}(pH_{raw})^{-2.599}$
107	(Sérodes et al., 2003)	$HAAs = 2.720+(TOC)^{0.653}+(D)^{0.458}+(t)^{0.295}$
108	(Sérodes et al., 2003)	$HAAs = 1.33+(TOC)^{2.612}+ (D)^{0.102}+ (T)^{0.255}+ (t)^{0.102}$

#	Source	Model
109	(Sérodes et al., 2003)	$HAAs = -8.202+4.869(TOC)+1.053(D)+0.364(t)$
110	(USEPA, 2001)	$HAA5 = 30.0(TOC)^{0.997}(D)^{0.278}(Br^{-})^{-0.138}(T)^{0.341}(pH)^{-0.799}(t)^{0.169}$
111	(USEPA, 2001)	$HAA5 = 30.7(DOC*UV_{254})^{0.302}(D)^{0.541}(Br^{-})^{-0.012}(0.932)^{(pH-7.5)}(1.021)^{(T-20)}(t)^{0.161}$
112	(USEPA, 2001)	$HAA5 = 41.2(DOC*UV_{254})^{0.498}(D)^{0.388}(Br^{-})^{-0.156}(0.867)^{(pH-7.5)}(1.021)^{(T-20)}(t)^{0.263}$
Individual Haloacetic Acid Species		
113	(Amy et al., 1998)	$MCA = 0.45(t)^{-0.009}(T)^{0.573}(pH)^{-0.279}(D)^{0.397}(DOC)^{0.173}(Br^{-})^{0.029}$
114	(Amy et al., 1998)	$MBA = 6.21*10^{-5}(t)^{0.090}(T)^{0.707}(pH)^{0.604}(D)^{0.754}(DOC)^{-0.584}(Br^{-})^{1.100}$
115	(Amy et al., 1998)	$DCA = 0.30(t)^{0.218}(T)^{0.456}(pH)^{0.200}(D)^{0.379}(DOC)^{1.396}(Br^{-})^{-0.149}$
116	(Amy et al., 1998)	$TCA = 92.68(t)^{0.180}(T)^{0.299}(pH)^{-1.627}(D)^{0.331}(DOC)^{1.152}(Br^{-})^{-0.229}$
117	(Amy et al., 1998)	$DBA = 3.59*10^{-5}(t)^{0.095}(T)^{0.380}(pH)^{-0.001}(D)^{0.673}(DOC)^{-1.086}(Br^{-})^{2.052}$
118	(Amy et al., 1998)	$MCA = 12.82(t)^{-0.066}(DOC)^{-0.377}(Br^{-})^{-0.303}(D)^{0.671}$
119	(Amy et al., 1998)	$MBA = 3.97*10^{-3}(t)^{0.132}(DOC)^{0.409}(Br^{-})^{0.834}(D)^{0.095}$
120	(Amy et al., 1998)	$DCA = 10.96(t)^{0.230}(DOC)^{0.704}(Br^{-})^{-0.514}(D)^{0.751}$
121	(Amy et al., 1998)	$TCA = 6.22(t)^{0.164}(DOC)^{0.900}(Br^{-})^{-0.267}(D)^{0.697}$
122	(Amy et al., 1998)	$DBA = 4.84*10^{-5}(t)^{0.077}(DOC)^{-0.424}(Br^{-})^{2.222}(D)^{0.379}$
123	(Amy et al., 1998)	$MCA = 11.49(t)^{-0.021}(DOC)^{-0.668}(Br^{-})^{-0.415}(D)^{1.061}$
124	(Amy et al., 1998)	$MBA = 3.33*10^{-3}(t)^{-0.020}(DOC)^{0.211}(Br^{-})^{0.925}(D)^{0.313}$
125	(Amy et al., 1998)	$DCA = 6.31(t)^{0.213}(DOC)^{0.846}(Br^{-})^{-9.416}(D)^{0.742}$
126	(Amy et al., 1998)	$TCA = 3.97(t)^{0.163}(DOC)^{1.063}(Br^{-})^{-0.215}(D)^{0.680}$
127	(Amy et al., 1998)	$DBA = 3.92*10^{-5}(t)^{0.068}(DOC)^{-0.318}(Br^{-})^{2.256}(D)^{0.397}$
128	(Amy et al., 1998)	$MCA = 12.30(t)^{-0.043}(DOC)^{-0.522}(Br^{-})^{-0.355}(D)^{0.859}$
129	(Amy et al., 1998)	$MBA = 3.68*10^{-3}(t)^{0.077}(DOC)^{0.299}(Br^{-})^{0.877}(D)^{0.206}$
130	(Amy et al., 1998)	$DCA = 8.38(t)^{0.222}(DOC)^{0.777}(Br^{-})^{-0.466}(D)^{0.744}$
131	(Amy et al., 1998)	$TCA = 4.98(t)^{0.163}(DOC)^{0.988}(Br^{-})^{-0.240}(D)^{0.683}$
132	(Amy et al., 1998)	$DBA = 4.41*10^{-5}(t)^{0.072}(DOC)^{-0.374}(Br^{-})^{2.237}(D)^{0.387}$
133	(Chen and Westerhoff, 2010)	$DCA = 188.5(DOC)^{0.57}(UV_{254})^{0.73}(Br^{-}+1)^{-2.42}$
134	(Chen and Westerhoff, 2010)	$TCA = 1282(DOC)^{0.19}(UV_{254})^{1.44}(Br^{-}+1)^{-2.85}$
135	(Montgomery Watson, 1993)	$MCA = 1.634(TOC)^{0.753}(Br^{-}+0.01)^{-0.085}(pH)^{-1.124}(D)^{0.509}(t)^{0.300}$
136	(Montgomery Watson, 1993)	$DCA = 0.606(TOC)^{0.291}(UV_{254})^{0.726}(Br^{-}+0.01)^{-0.568}(D)^{0.48}(t)^{0.239}(T)^{0.665}$
137	(Montgomery Watson, 1993)	$TCA = 87.182(TOC)^{0.355}(UV_{254})^{0.901}(Br^{-}+0.01)^{0.679}(pH)^{1.732}(D)^{0.881}(t)^{0.264}$
138	(Montgomery Watson, 1993)	$MBA = 0.176(TOC)^{1.664}(UV_{254})^{-0.624}(Br^{-})^{-0.795}(pH)^{-0.927}(t)^{0.145}(T)^{0.45}$
139	(Montgomery Watson, 1993)	$DBA = 84.945(TOC)^{-0.62}(UV_{254})^{0.651}(Br^{-})^{1.073}(D)^{-0.2}(t)^{0.12}(T)^{0.657}$
140	(USEPA, 2001)	$MCA = 4.58(DOC*UV_{254})^{-0.090}(D)^{0.662}(Br^{-})^{-0.224}(1.042)^{(pH-7.5)}(1.024)^{(T-20)}(t)^{0.043}$
141	(USEPA, 2001)	$DCA = 60.4(DOC*UV_{254})^{0.397}(D)^{0.665}(Br^{-})^{-0.558}(1.034)^{(pH-7.5)}(1.017)^{(T-20)}(t)^{0.222}$
142	(USEPA, 2001)	$TCA = 52.6(DOC*UV_{254})^{0.403}(D)^{0.749}(Br^{-})^{-0.416}(0.8739)^{(pH-7.5)}(1.014)^{(T-20)}(t)^{0.163}$
143	(USEPA, 2001)	$MBA = 2.06*10^{-2}(DOC*UV_{254})^{0.358}(D)^{-0.101}(Br^{-})^{0.812}(0.6526)^{(pH-7.5)}(1.162)^{(T-20)}(t)^{0.043}$
144	(USEPA, 2001)	$DBA = 9.42*10^{-5}(DOC*UV_{254})^{0.0590}(D)^{0.182}(Br^{-})^{2.109}(1.210)^{(pH-7.5)}(1.007)^{(T-20)}(t)^{0.070}$
145	(USEPA, 2001)	$MCA = 1.31*10^{-1}(DOC*UV_{254})^{0.202}(D)^{0.275}(Br^{-})^{-0.958}(0.124)^{(pH-7.5)}(1.036)^{(T-20)}(t)^{0.923}$
146	(USEPA, 2001)	$DCA = 38.4(DOC*UV_{254})^{0.503}(D)^{0.421}(Br^{-})^{-0.393}(0.867)^{(pH-7.5)}(1.019)^{(T-20)}(t)^{0.293}$

#	Source	Model
147	(USEPA, 2001)	$TCA = 47.8(DOC*UV_{254})^{0.627}(D)^{0.729}(Br^{-})^{-0.425}(0.602)^{(pH-7.5)}(1.011)^{(T-20)}(t)^{0.174}$
148	(USEPA, 2001)	$MBA = 3.0*10^{-1}(DOC*UV_{254})^{0.093}(D)^{0.964}(Br^{-})^{-0.408}(0.134)^{(pH-7.5)}(1.054)^{(T-20)}(t)^{0.554}$
149	(USEPA, 2001)	$DBA = 3.96*10^{-1}(DOC*UV_{254})^{0.509}(D)^{-0.251}(Br^{-})^{0.689}(1.302)^{(pH-7.5)}(1.019)^{(T-20)}(t)^{0.310}$

TTHM=total trihalomethanes (□g L^{-1}); CF=chloroform (□g L^{-1}); BDCM=bromodichloromethane (□g L^{-1}); DBCM=dibromochloromethane (□g L^{-1}); BF=bromoform (□g L^{-1}); HAA=haloacetic acid (□g L^{-1}); MCA=monochloroacetic acid (□g L^{-1}); DCA=dichloroacetic acid (□g L^{-1}); TCA=trichloroacetic acid (□g L^{-1}); BAA=bromoacetic acid (□g L^{-1}); MBA=monobromoacetic acid (□g L^{-1}); TOC=total organic carbon (mg L^{-1}); DOC=dissolved organic carbon (mg L^{-1}); UV_{254}=ultraviolet absorbance at a wavelength of 254 nm (cm^{-1}); D=chlorine dose (mg L^{-1}); T=temperature (°C); t=reaction time (hr); Br^{-}=bromide ion concentration (□g L^{-1}); SUVA=specific ultraviolet absorbance at a wavelength of 254 nm (L $(mg\text{-}m)^{-1}$)

[a] Zhu (1995) published additional models, including predictive power functions, which gave the best fit, and are reported in Amy et al. (1998).

[b] Reported in USEPA (2001) based on Sohn et al. (2004) and Amy et al. (1998).

Table S5 – Conditions for model development

Source	Model #[k]	Source Water	Organic Content (mg L^{-1})	Temperature (°C)	Time (hr)	Chlorine Dose (mg L^{-1})	pH	Bromide (µg L^{-1})
(Al-Omari et al., 2004)	1	Raw and settled waters from Amman, Jordan	0.7≤TOC≤3.1	20	0.5≤t≤24	1≤D≤5	7.65≤pH≤8.35	$920 \leq Br^{-} \leq 2,140$
(Amy et al., 1998)	2, 48–51	Raw surface and groundwaters from the U.S.	1.2≤DOC≤10.6	15≤T≤25	2≤t≤168	1.51≤D≤33.55	6.5≤pH≤8.5	$7 \leq Br^{-} \leq 600$
(Amy et al., 1998)	3, 4, 5, 52–63	Alum or iron-coagulated waters from the U.S.	1.00≤DOC≤7.77	20	2≤t≤168	1.11≤D≤24.75	7.5	$36 \leq Br^{-} \leq 308$
(Amy et al., 1998)	113–117	Raw surface and groundwaters from the U.S.	1.2≤DOC≤10.7	15≤T≤25	2≤t≤168	2.11≤D≤26.4, 0.5≤D:DOC≤3 mg mg^{-1}	6.5≤pH≤8.5	$7 \leq Br^{-} \leq 560$
(Amy et al., 1998)	118–122	Alum-coagulated waters from the U.S.	1≤DOC≤4.6	20	2≤t≤168	1.11≤D≤14.19, 1≤D:DOC≤3 mg mg^{-1}	7.5	$36 \leq Br^{-} \leq 308$
(Amy et al., 1998)	123–127	Iron-coagulated waters from the U.S.	1.03≤DOC≤4.20	20	2≤t≤168	5≤D≤24, 1≤D:DOC≤3 mg mg^{-1}	7.5	$37 \leq Br^{-} \leq 308$

Source	**Model #[k]**	**Source Water**	**Organic Content (mg L^{-1})**	**Temperature (°C)**	**Time (hr)**	**Chlorine Dose (mg L^{-1})**	**pH**	**Bromide (µg L^{-1})**
(Amy et al., 1998)	128–132	Alum or iron-coagulated waters from the U.S.	$1.0 \leq DOC \leq 4.6$	20	$2 \leq t \leq 168$	$1.11 \leq D \leq 24$, $1 \leq D:DOC \leq 3$ mg mg^{-1}	7.5	$36 \leq Br^- \leq 308$
(Boyalla, 2004)	6	Raw water from Newfoundland, Canada	$5.58 \leq TOC \leq 27$ $4.01 \leq DOC \leq 10.88$	23	$1 \leq t \leq 120$	$22.36 \leq D \leq 31.25$	$3.95 \leq pH \leq 9.08$	
(Chang et al., 1996)	7, 8, 9	Raw water from Taiwan	$0.61 \leq TOC \leq 1.24$	$21 \leq T \leq 27$	$0.5 \leq t \leq 168$	$3.5 \leq D \leq 5.0$	$7.8 \leq pH \leq 8.0$	$<10 \leq Br^- \leq 20$
(Chen and Westerhoff, 2010)	10, 64, 65, 133, 134	Wastewater, groundwater, surface water, and coagulated waters from the U.S.	$0.6 \leq DOC \leq 23$	25	$24 \leq t \leq 72$	$D=3(DOC)+8(NH_3\text{-}N)+10$	~8.2	$0 \leq Br^- \leq 1{,}000$
(Harrington et al., 1992)	11	Raw waters from the U.S.	$3 \leq TOC \leq 13.8$	$10 \leq T \leq 30$	$0.1 \leq t \leq 168$	$1.5 \leq D \leq 69$	$4.6 \leq pH \leq 9.8$	$10 \leq Br^- \leq 245$
(Hong et al., 2007)	12, 66, 67	Raw water from Dongjiang River, Hong Kong		$15 \leq T \leq 25$	$6 \leq t \leq 168$	$0.5 \leq D/DOC \leq 3$ mg mg^{-1}	$6.5 \leq pH \leq 8.5$	$10 \leq Br^- \leq 310$
(Malcom Pirnie Inc, 1992)	68, 69, 70	Raw water	$2.8 \leq TOC \leq 11$[g]	$13 \leq T \leq 20$[g]	$0.1 \leq t \leq 105$[g]	$3.0 \leq D \leq 25.0$[g]	$5.6 \leq pH \leq 9.0$[g]	$5 \leq Br^- \leq 43$[g]
(Malcom Pirnie Inc, 1992)	71							
(McBean et al., 2008)	13, 106	Treated waters from Ontario, Canada	$1.3 \leq DOC \leq 13.2$[f] $0.8 \leq DOC \leq 5.6$	$1.0 \leq T \leq 25.4$		$0.42 \leq D \leq 5.5$ $0.15 \leq D_{pre} \leq 4.7$ $0.03 \leq D_{post} \leq 2.89$	$7.14 \leq pH \leq 8.57$[f] $7.02 \leq pH \leq 8.34$	
(Milot et al., 2002)	14	Raw and treated waters from the U.S.	$2.2 \leq DOC$ or $TOC \leq 13.8$	$10 \leq T \leq 30$	$0.1 \leq t \leq 168$	$1.5 \leq D \leq 69.0$	$4.6 \leq pH \leq 10$	
(Montgomery Watson, 1993)	72, 77, 135–139		$2.8 \leq TOC \leq 11$[b]	$10 \leq T \leq 20$[b]	$0.1 \leq t \leq 96$[b]	$3.0 \leq D \leq 25.3$[b]	$5.6 \leq pH \leq 9$[b]	

Source	**Model #^{k}**	**Source Water**	**Organic Content (mg L^{-1})**	**Temperature (°C)**	**Time (hr)**	**Chlorine Dose (mg L^{-1})**	**pH**	**Bromide (μg L^{-1})**
(Montgomery Watson, 1993)	73		2.8≤TOC≤11^{b}	10≤T≤20^{b}	0.1≤t≤96^{b}	3.0≤D≤25.3^{b}	5.6≤pH≤9^{b}	For D/Br^{-} <75^{j}
(Montgomery Watson, 1993)	74		2.8≤TOC≤11^{b}	10≤T≤20^{b}	0.1≤t≤96^{b}	3.0≤D≤25.3^{b}	5.6≤pH≤9^{b}	For D/Br^{-} >75^{j}
(Montgomery Watson, 1993)	75		2.8≤TOC≤11^{b}	10≤T≤20^{b}	0.1≤t≤96^{b}	3.0≤D≤25.3^{b}	5.6≤pH≤9^{b}	For D/Br^{-} <50^{j}
(Montgomery Watson, 1993)	76		2.8≤TOC≤11^{b}	10≤T≤20^{b}	0.1≤t≤96^{b}	3.0≤D≤25.3^{b}	5.6≤pH≤9^{b}	For D/Br^{-} >50^{j}
(Rathbun, 1996b)	78–81	Untreated Mississippi, Missouri, and Ohio River water from the U.S.	2.2≤DOC≤12	25	168	15≤D≤50	6.85≤pH≤9.75^{c}	10≤Br^{-} ≤85^{c}
(Rathbun, 1996b)	15	Untreated Mississippi, Missouri, and Ohio River water from the U.S.	2.2≤DOC≤13^{c}	25	2≤t≤168	15≤D≤50	5.5≤pH≤10	
(Rodriguez et al., 2000)	16	Untreated waters	2.2≤DOC≤7.7	10≤T≤30	0≤t≤168	1.5≤D≤50	4.6≤pH≤10	10 ≤Br^{-} ≤420
(Rodriguez et al., 2000)	17	Distribution system in Quebec, Canada	1.2≤DOC≤7.7^{i}	6≤T≤20^{i}			6.4≤pH≤8.0^{i}	
(Semerjian et al., 2009)	18–29	Treated waters from Lebanon		15≤T≤28	0.33≤t≤48		6.5≤pH≤8.5	700 ≤Br^{-} ≤830
(Sérodes et al., 2003)	30, 107	Raw water from Charlesbourg, Canada	1.5≤TOC≤1.8	8≤T≤20	1≤t≤48	2≤D≤4	7.5	
(Sérodes et al., 2003)	31, 108	Treated water from Quebec City, Canada	1.6≤TOC≤2.2	8≤T≤20	1≤t≤48	1≤D≤3	7.5	

Source	**Model #^{k}**	**Source Water**	**Organic Content (mg L^{-1})**	**Temperature (^{o}C)**	**Time (hr)**	**Chlorine Dose (mg L^{-1})**	**pH**	**Bromide (µg L^{-1})**
(Sérodes et al., 2003)	32, 109	Treated water from Sainte Foy, Canada	$2.1 \le TOC \le 2.6$	$8 \le T \le 20$	$1 \le t \le 48$	$1 \le D \le 3$	7.5	
(Sohn et al., 2004)	33, 34	Raw and coagulated waters from the U.S.	$1 \le DOC \le 10.6^d$	$15 \le T \le 25^d$	$2 \le t \le 168^d$	$1.11 \le D \le 33.5^d$	$6.5 \le pH \le 8.5^d$	$7 \le Br^- \le 600^d$
(Sohn et al., 2004)	35	Temp. and pH correction for Amy et al., 1987	$1 \le DOC \le 10.6^d$	$15 \le T \le 25^d$	$2 \le t \le 168^d$	$1.11 \le D \le 33.5^d$	$6.5 \le pH \le 8.5^d$	$7 \le Br^- \le 600^d$
(Sohn et al., 2004)	36, 37	Raw waters from the U.S. (short-term models)	$3.0 \le DOC \le 11.7$	$10 \le T \le 20$	$0.1 \le t \le 2$	$1.5 \le D \le 1.7$	$4.6 \le pH \le 8.3$	$10 \le Br^- \le 152$
(Toroz and Uyak, 2005)	38	Distribution system in Istanbul, Turkey	$2.98 \le TOC \le 3.41$	$11.07 \le T \le 26.74$		$3.2 \le D \le 6.2$	$7.34 \le pH \le 8.50$	$145 \le Br^- \le 255$
(Urano et al., 1983)	39	Humic acid	$2.1 \le TOC \le 9^c$	20	$1 \le t \le 48^c$	50	$4 \le pH \le 10^c$	
(Uyak et al., 2005)	40	Raw water from Istanbul, Turkey	$4.20 \le TOC \le 6.20$	$7.2 \le T \le 22.7$		$2.82 \le D \le 6.75$	$7.10 \le pH \le 7.90$	
(USEPA, 2001)	110	Raw surface and groundwaters from the U.S.	$1.2 \le DOC \le 10.6$	$15 \le T \le 25$	$2 \le t \le 168$	$1.51 \le D \le 33.55$	$6.5 \le pH \le 8.5$	$7 \le Br^- \le 600$
(USEPA, 2001)	41, 82–85, 111, 140–144	Coagulated waters from the U.S.	$1.00 \le DOC \le 7.77$	20	$2 \le t \le 168$	$1.11 \le D \le 24.75$	7.5	$23 \le Br^- \le 308$
(USEPA, 2001)	42, 86–89, 112, 145–149	GAC-filtered waters from the U.S.	$0.14 \le DOC \le 2.0$	$3 \le T \le 33$	$2 \le t \le 168$	$0.5 \le D \le 3.0$	$6.7 \le pH \le 10$	$10 \le Br^- \le 570$
(Zhu, 1995)	43–47, 90–105	Raw waters from the U.S.	$1.2 \le DOC \le 10.6$	$15 \le T \le 25$	$2 \le t \le 168$	$1.51 \le D \le 33.55$	$6.5 \le pH \le 8.5$	$7 \le Br^- \le 600$

TOC=total organic carbon (mg L^{-1}); DOC=dissolved organic carbon (mg L^{-1}); UV$_{254}$=ultraviolet absorbance at a wavelength of 254 nm (cm^{-1}); D=chlorine dose (mg L^{-1}); T=temperature (^{o}C); t=reaction time (hr); Br$^-$=bromide ion concentration (□g L^{-1}); SUVA=specific ultraviolet absorbance at a wavelength of 254 nm (L (mg-m)$^{-1}$)

Source	Model #[k]	Source Water	Organic Content (mg L^{-1})	Temperature (°C)	Time (hr)	Chlorine Dose (mg L^{-1})	pH	Bromide (µg L^{-1})

[a] Boundary conditions for model provided in Amy et al. (1987)
[b] As reported in Milot et al. (2002)
[c] Estimated from data presented in graphical form
[d] Boundary conditions for model provided in Amy et al. (1998)
[e] Reported in USEPA (2001) based on Amy et al. (1998)
[f] The model differentiates between raw and treated water. The first entry is for raw water DOC and pH, while the second is for treated water DOC and pH
[g] As reported in Gang et al. (2002)
[h] Bench-scale data collected from Amy et al. (1987), Montgomery Watson (1991), and Rathbun (1996a)
[i] Reported boundaries are 10th and 90th percentile values from the entire field-scale database
[j] As reported in Sadiq and Rodriguez (2004)
[k] Model # refers to number assigned in Table S4

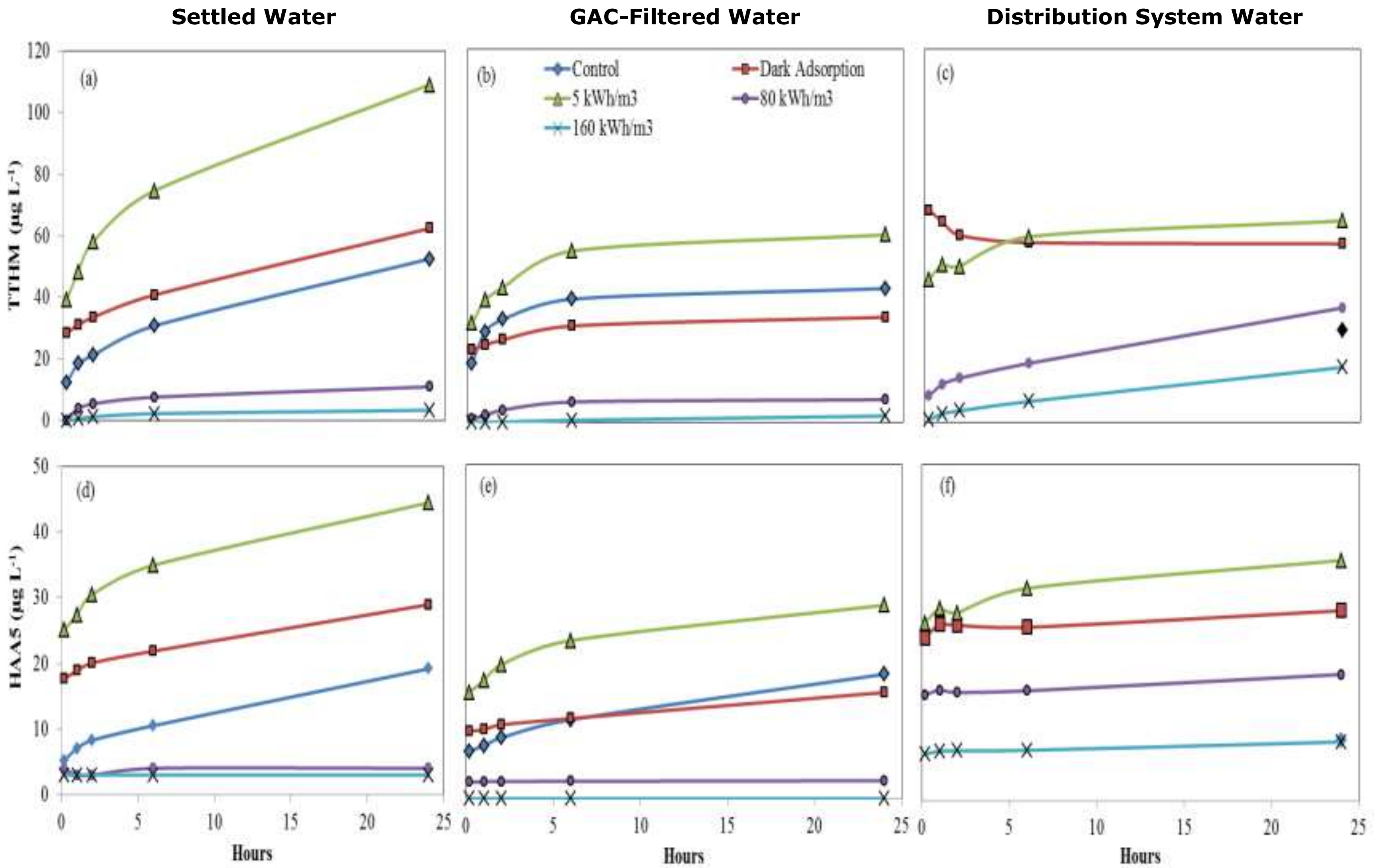


Figure S1 – Time evolution of TTHMs and HAA5s following chlorine disinfection. Panels (a) and (d) illustrate the formation of TTHMs and HAA5s, respectively, for the SW as a function of time and extent of photocatalytic treatment; (b) and (e) illustrate the same for the GAC-filtered water; (c) and (f) illustrate the same for the DS water.

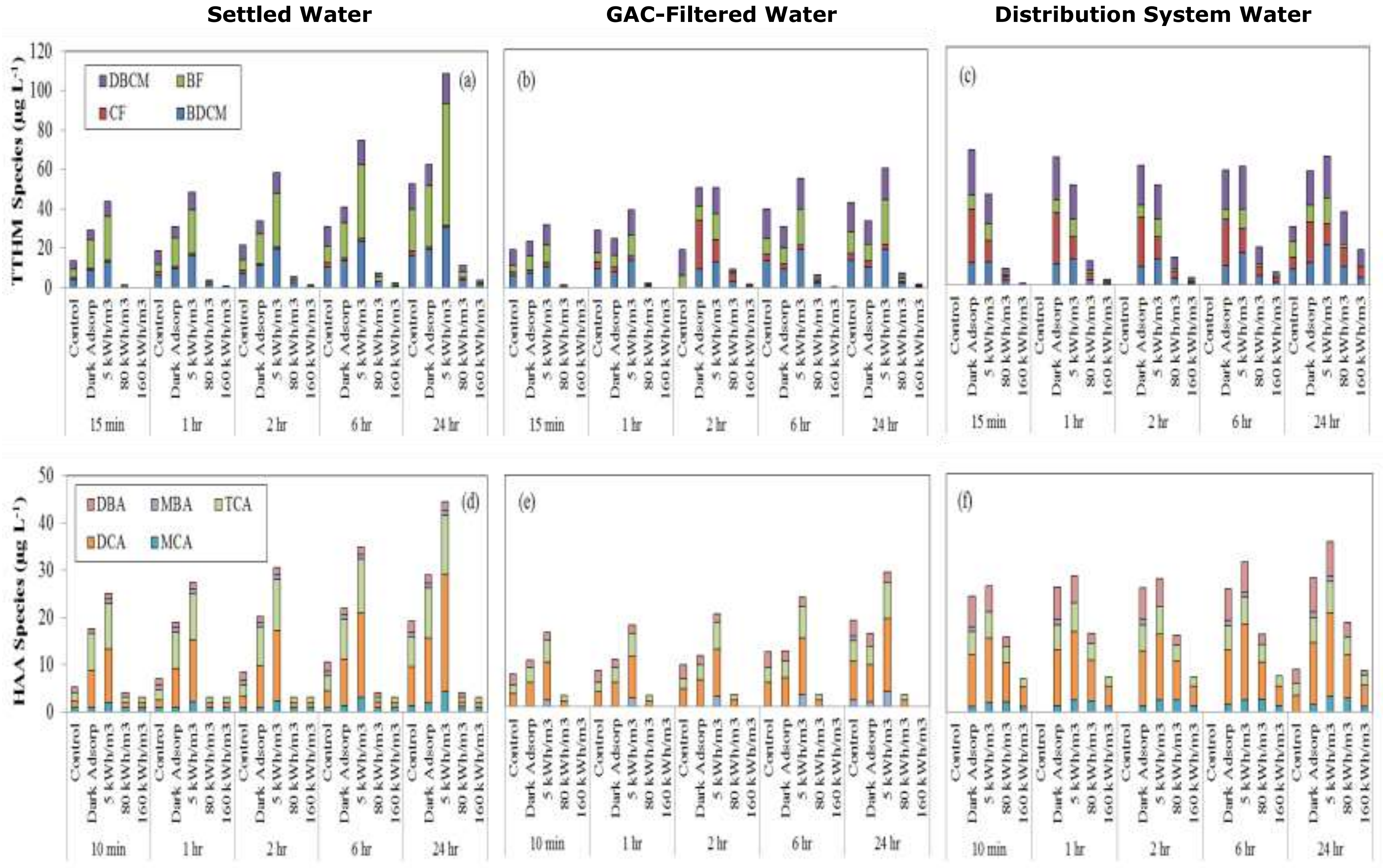


Figure S2 – Time evolution of TTHM and HAA5 species following chlorine disinfection. Panels (a) and (d) illustrate the formation of THMs and HAAs, respectively, for the SW as a function of time and extent of photocatalytic treatment; (b) and (e) illustrate the same for the GAC-filtered water; (c) and (f) illustrate the same for the DS water.

6. Correlation Analyses of Empirical Data Parameters and DBP Formation

6.1 Normal Distribution Testing

Prior to testing correlations, each data grouping (e.g., extent of photocatalytic treatment, type of source water) was tested for normal distribution using the D'Agostino-Pearson omnibus test (α = 0.05). The results of these tests are reported in Tables S6 – S8, with normally distributed datasets shaded in gray.

Table S6 – P values from the D'Agostino-Pearson omnibus test for normality of internal DBP data grouped by extent of photocatalytic treatment

Data Group	All AOPs	Non-AOPs	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}
D	0.0004	0.4086	0.0369	0.0256	0.0329
t	0.0029	0.0115	0.0416	0.0416	0.0416
pH	<0.0001	0.0002	0.0551	0.0551	0.0551
DOC	0.0271	<0.0001	0.0678	0.0344	0.0292
Br^-	<0.0001	0.0027	0.0316	0.0291	0.0291
UV_{254}	0.0216	0.0002	0.0762	0.0291	0.0568
SUVA	0.0606	0.4687	0.0396	0.0311	0.0338
BDCM	0.0115	0.0579	0.0266	<0.0001	<0.0001
CF	<0.0001	0.0031	0.0386	0.0002	<0.0001
BF	<0.0001	<0.0001	0.0002	0.3090	0.0218
DBCM	0.0335	0.1987	0.8709	<0.0001	<0.0001
TTHM	0.0033	0.7868	0.0007	<0.0001	<0.0001
MCA	0.0527	0.0010	0.1757	0.1270	0.0461
DCA	0.1050	0.2076	0.0418	0.0066	0.0034
TCA	0.1705	0.1330	0.2957	0.3600	0.1770
MBA	<0.0001	0.0051	0.0309	[a]	[a]
DBA	0.0003	0.0780	0.1271	0.0058	<0.0001
HAA5	0.2060	0.0764	0.9224	0.0002	0.3540
n =	45	25 – 30	15	15	15

[a] Values did not change, so no P value could be computed

Table S7 – P values from the D'Agostino-Pearson omnibus test for normality of all THM data (internal and external) grouped by type of source water

Data Group	Untreated	Coagulated	Settled	Filtered	Finished	Synthetic	All Data
D	<0.0001	<0.0001	<0.0001	0.0082	0.2394	<0.0001	<0.0001
t	<0.0001	[a]	0.0177	<0.0001	0.0177	<0.0001	<0.0001
T	<0.0001	<0.0001	[a]	<0.0001	[a]	<0.0001	<0.0001
pH	<0.0001	0.0021	[a]	<0.0001	[a]	<0.0001	<0.0001
DOC	<0.0001	0.0311	<0.0001	0.0184	0.0002	0.0721	<0.0001
Br^-	<0.0001	<0.0001	0.0065	<0.0001	0.0066	<0.0001	<0.0001
UV_{254}	<0.0001	0.0153	0.0018	0.0345	0.1553	0.0794	<0.0001
SUVA	<0.0001	0.1111	0.0041	0.0618	0.0886	<0.0001	<0.0001
BDCM	0.0059	0.0187	0.1268	0.1913	0.1913	[b]	<0.0001
CF	<0.0001	0.0175	0.5105	0.0168	0.0168	<0.0001	<0.0001
BF	<0.0001	0.0085	0.0002	0.0250	0.0250	[b]	<0.0001
DBCM	<0.0001	0.2021	0.2468	0.0003	0.0003	[b]	0.1998
TTHM	<0.0001	0.1932	0.0165	0.4758	0.2382	0.1256	<0.0001
n =	127 – 211	24	25	25 – 39	25	15 – 79	198 - 369

[a] Values did not change, so no P value could be computed

[b] No values reported in datasets analyzed

Table S8 – P values from the D'Agostino-Pearson omnibus test for normality of all HAA data (internal and external) grouped by type of source water

Data Group	Untreated	Settled	Filtered	Finished	All Data
D	<0.0001	<0.0001	0.2394	0.1088	<0.0001
t	0.0123	0.0179	0.0179	0.0179	<0.0001
T	0.0089	[b]	[b]	[b]	0.0389
pH	0.0217	[b]	[b]	[b]	<0.0001
DOC	<0.0001	<0.0001	0.0002	0.1076	<0.0001
Br^-	0.0001	0.0065	0.0066	[b]	<0.0001
UV_{254}	<0.0001	0.0041	0.1531	<0.0001	<0.0001
SUVA	0.1123	0.0513	0.0771	0.0906	<0.0001
MCA	[a]	<0.0001	0.0030	0.9891	0.0071
DCA	[a]	0.0217	0.1618	0.2585	<0.0001
TCA	[a]	0.2757	0.3805	0.0293	<0.0001
MBA	[a]	<0.0001	<0.0001	0.0090	<0.0001
DBA	[a]	<0.0001	0.0067	0.0037	0.0013
HAA5	0.0609	0.9440	0.3489	0.0653	<0.0001
n =	4 – 31	25	25	25	75 – 11

[a] n too small to compute P value

[b] Values did not change, so no P value could be computed

6.2 Analyses of Linear Correlation Grouped by Type of Source Water

Figure S3 shows the correlation analyses between parameters commonly associated with DBP formation and the DBPs themselves, grouped by type of source water. Linear correlation coefficients (either the Pearson product-moment coefficient of correlation for normally distributed data or the Spearman rank correlation coefficient for non-normal distributions) were computed for datasets. All data, both internal and external, were considered in the statistical analysis. Figure S3 includes varying numbers of data pairs per group, dependent upon reports from individual literature sources, datasets in which the parameters were varied, etc. (for example, if contact time was held constant for a particular dataset, no correlation can be computed).

Although the data matrices are sparse for some groups, it appears that D, T, t, DOC, UV_{254}, and SUVA are generally positively correlated with DBP formation. The relationship of THMs and HAAs with pH and Br^- is less straightforward as both positive and negative correlations are observed. The treated samples show less variation within parameter groups, while the untreated samples show the greatest variation. This may stem from untreated samples containing more complex NOM, whereas following different stages of drinking water treatment (e.g., coagulation, sedimentation, and filtration), less NOM is present. Moreover, the NOM composition is likely more homogenized as physicochemical treatments such as coagulation tend to target certain NOM moieties, i.e., the aromatic fraction, which is most closely linked to DBP formation (Weishaar et al., 2003).

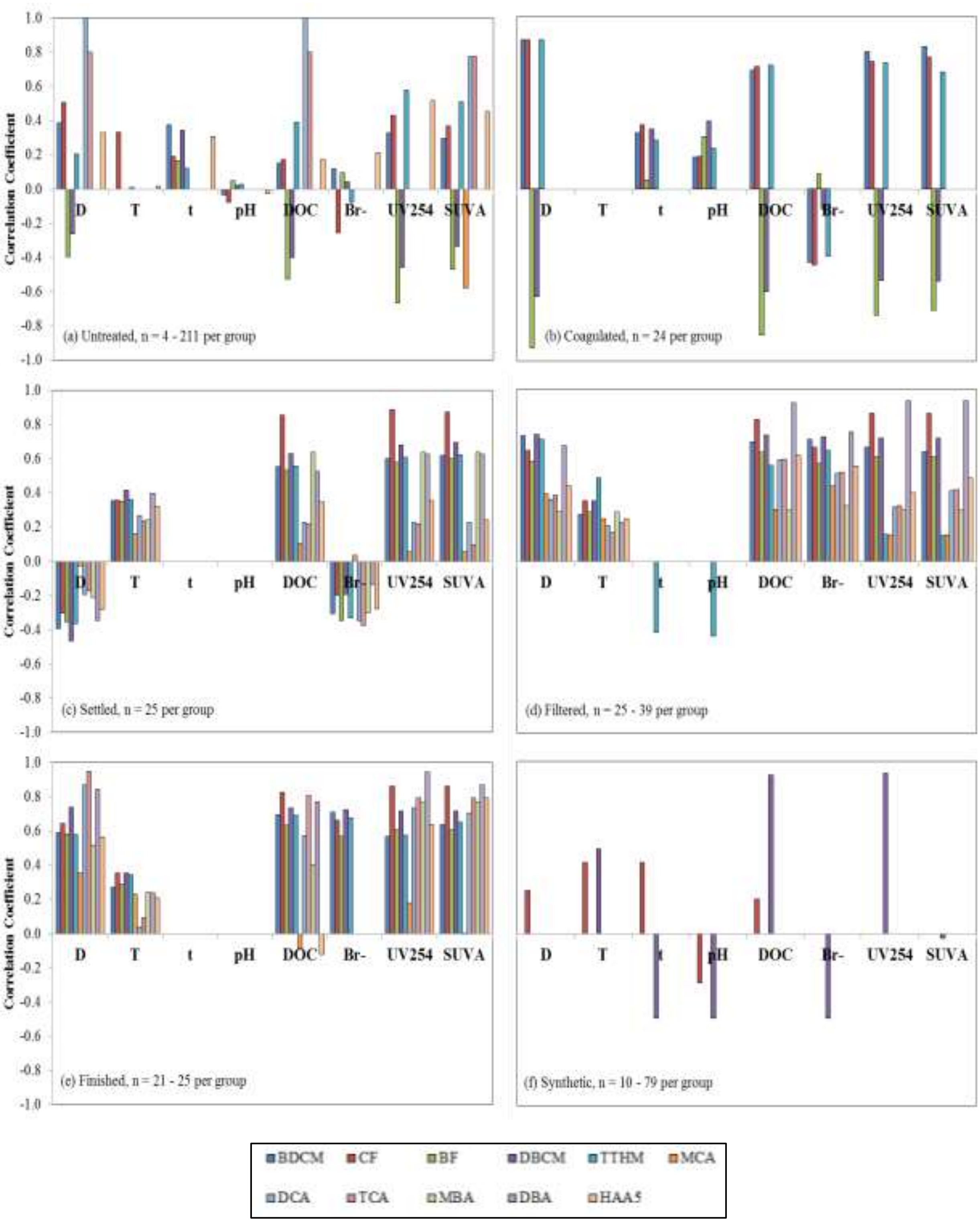


Figure S3 – Linear correlation coefficients (Pearson for normally distributed datasets, otherwise Spearman) between DBP formation and precursor measurements for different types of source water. Values closer to +1 indicate positive correlations, while values closer to -1 indicate negative correlations. Data includes all applicable internal and external datasets.

7. DBP Model Performance

Coefficients of determination (R^2) were computed for each of the 2,642 datasets of THM and HAA model/data pairings analyzed in this study. All R^2 values, representing the model's ability to accurately predict DBP formation using bulk precursors such as DOC, UV_{254}, SUVA, D, t, T, and Br^-, are provided in Tables S9 and S10 for THMs and HAAs, respectively. The values ranged from 0.00 – 0.99, with an overall average of 0.39. The R^2 values shown in bold, underlined font are for the datasets which satisfied all boundary conditions under which the model was developed. A total of 12 (0.5%) of the THM and HAA model/data datasets satisfied all model boundary conditions. Notably, the majority of the datasets satisfying the boundary conditions were for models that did not report all conditions used to develop specific parameters, particularly Br^-. Moreover, many of the DBP models included in this study were bounded by time periods ≥2 hr, which may adversely impact model performance for the rapid reactions characteristic of short-term DBP formation for internal DBP measurements at 10 – 15 min and 1 hr.

While it is recommended that models be applied to datasets satisfying the boundary conditions, based on this analysis, it appears that many models are so specific in their application, due to the boundary conditions implemented during model development, that extrapolation is necessary for the vast majority of water systems. The inaccuracy this introduces brings into question the usefulness of many of these models outside of the specific scenario from which they were derived, even in cases of qualitative comparisons.

Table S9 – Average correlation coefficients (R^2 values) for pairs of predictive THM models versus measured THM datasets

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
1	(Al-Omari et al., 2004)[a]	TTHM		0.94 (NR[k])	0.29	0.00	0.59	0.60	0.57	0.11	0.70	0.08	0.00	0.28	0.78	0.78	0.55
2	(Amy et al., 1998)	TTHM	786 (80)	0.90 (0.84[d])	0.33	0.01	0.84	0.60	0.61	0.04	0.91	0.49	0.04	0.34	0.72	0.65	0.79
48	(Amy et al., 1998)	CF	786 (80)	0.87	0.02	0.04	0.25	0.64	0.70	0.03	0.04	0.81	0.07	0.27	0.62	0.64	0.84
49	(Amy et al., 1998)	BDCM	786 (80)	0.90	0.17	0.07	0.47	0.33	0.17	0.44	0.65	0.02	0.04	0.14	0.57	0.50	0.45
50	(Amy et al., 1998)	DBCM	786 (80)	0.89	0.01	0.02	0.06	0.08	0.22	0.04	0.25	0.10	0.06	0.01	0.17	0.15	0.00
51	(Amy et al., 1998)	BF	786 (80)	0.61	0.00	0.00	0.00	0.07	0.03	0.10	0.95	0.58	0.74	0.07	0.05	0.02	0.08
3	(Amy et al., 1998)	TTHM	143[e]	0.87	0.31	0.02	0.67	0.57	0.47	0.08	0.88	0.17	0.01	0.11	0.73	0.67	0.61
52	(Amy et al., 1998)	CF	143[e]	0.90	0.19	0.45	0.21	0.22	0.41	0.48	0.00	0.16	0.14	0.26	0.85	0.71	0.61
53	(Amy et al., 1998)	BDCM	143[e]	0.84	0.22	0.09	0.39	0.31	0.24	0.48	0.73	0.04	0.06	0.06	0.51	0.47	0.43
54	(Amy et al., 1998)	DBCM	143[e]	0.31	0.04	0.02	0.01	0.42	0.22	0.24	0.22	0.09	0.06	0.01	0.49	0.32	0.00
55	(Amy et al., 1998)	BF	143[e]	0.73	0.00	0.01	0.34	0.07	0.02	0.07	0.89	0.50	0.74	0.00	0.13	0.00	0.00
4	(Amy et al., 1998)	TTHM	143[e]	0.88	0.28	0.02	0.63	0.54	0.39	0.07	0.30	0.96	0.00	0.10	0.73	0.66	0.59
56	(Amy et al., 1998)	CF	143[e]	0.92	0.23	0.49	0.29	0.25	0.48	0.49	0.01	0.21	0.33	0.24	0.85	0.74	0.64
57	(Amy et al., 1998)	BDCM	143[e]	0.84	0.21	0.09	0.41	0.33	0.22	0.46	0.73	0.04	0.05	0.05	0.49	0.48	0.42
58	(Amy et al., 1998)	DBCM	143[e]	0.34	0.03	0.03	0.01	0.05	0.21	0.23	0.13	0.09	0.04	0.00	0.59	0.31	0.00
59	(Amy et al., 1998)	BF	143[e]	0.72	0.00	0.01	0.04	0.01	0.02	0.00	0.89	0.50	0.74	0.01	0.02	0.00	0.02
5	(Amy et al., 1998)	TTHM	286[e]	0.87 (0.98[d])	0.29	0.02	0.64	0.55	0.42	0.07	0.86	0.12	0.01	0.10	0.73	0.66	0.60

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
60	(Amy et al., 1998)	CF	286[e]	0.91	0.21	0.47	0.25	0.23	0.45	0.48	0.00	0.18	0.22	0.25	0.85	0.73	0.63
61	(Amy et al., 1998)	BDCM	286[e]	0.83	0.22	0.09	0.41	0.32	0.23	0.47	0.73	0.04	0.06	0.06	0.49	0.47	0.42
62	(Amy et al., 1998)	DBCM	286[e]	0.32	0.04	0.02	0.00	0.27	0.22	0.23	0.17	0.08	0.05	0.00	0.66	0.32	0.00
63	(Amy et al., 1998)	BF	286[e]	0.72	0.00	0.01	0.15	0.01	0.02	0.00	0.89	0.50	0.74	0.01	0.04	0.00	0.00
6	(Boyalla, 2004)	TTHM	51	0.77	0.08	0.05	0.12	0.15	0.52	0.22	0.57	0.00	0.01	0.24	0.50	0.78	0.20
7	(Chang et al., 1996)	TTHM	120	0.97	0.28	0.09	0.86	0.36	0.36	0.27	0.89	0.40	0.19	0.07	0.76	0.55	0.29
8	(Chang et al., 1996)	TTHM	120	0.94	0.29	0.08	0.44	0.47	0.41	0.09	0.68	0.53	0.31	0.09	0.14	0.52	0.53
9	(Chang et al., 1996)	TTHM	120	0.95	0.28	0.11	0.87	0.35	0.36	0.33	0.88	0.39	0.25	0.07	0.76	0.53	0.26
10	(Chen and Westerhoff, 2010)	TTHM	210	0.87	0.14	0.00	0.68	0.19	0.22	0.38	0.20	0.02	0.01	0.08	0.75	0.57	0.32
64	(Chen and Westerhoff, 2010)	CF	210	0.88	0.08	0.37	0.02	0.04	0.09	0.82	0.33	0.01	0.09	0.12	0.85	0.52	0.11
65	(Chen and Westerhoff, 2010)	BDCM	210	0.69	0.01	0.01	0.32	0.08	0.08	0.71	0.26	0.02	0.00	0.01	0.56	0.54	0.01
11	(Harrington et al., 1992)	TTHM		NR (NR[b,c,k])	0.38	0.09	0.81	0.51	0.60	0.19	0.86	0.37	0.12	0.41	0.72	0.68	0.67
12	(Hong et al., 2007)	TTHM	72	0.90	0.01	0.11	0.00	0.00	0.07	0.02	0.85	0.06	0.14	0.02	0.15	0.18	0.02
66	(Hong et al., 2007)	BDCM	72	0.90	0.10	0.20	0.09	0.04	0.20	0.08	0.96	0.15	0.25	0.06	0.04	0.11	0.00
67	(Hong et al., 2007)	CF	72	0.85	0.02	0.22	0.06	0.02	0.05	0.01	0.00	0.11	0.03	0.01	0.06	0.10	0.02
68	(Malcom Pirnie Inc, 1992)	CF			0.09	0.49	0.02	0.07	0.16	0.69	0.18	0.03	0.03	0.30	0.69	0.56	0.35

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
69	(Malcom Pirnie Inc, 1992)	BDCM			0.16	0.14	0.29	0.23	0.20	0.43	0.81	0.06	0.10	0.30	0.70	0.57	0.45
70	(Malcom Pirnie Inc, 1992)	DBCM			0.02	0.02	0.03	0.00	0.22	0.70	0.15	0.08	0.05	0.00	0.29	0.21	0.01
71	(Malcom Pirnie Inc, 1993)	CF			0.09	0.54	0.01	0.07	0.07	0.70	0.17	0.03	0.04	0.09	0.71	0.11	0.17
13	(McBean et al., 2008)	TTHM	180	0.73	0.24	0.12	0.72	0.40	0.67	0.03	0.20	0.15	0.02	0.26	0.95	0.82	0.80
14	(Milot et al., 2002)	TTHM	1483	0.87	0.33	0.01	0.84	0.59	0.68	0.05	0.88	0.55	0.05	0.35	**0.63**	**0.54**	0.67
72	(Montgomery Watson, 1993)	CF			0.21	0.10	0.89	0.21	0.06	0.73	0.86	0.66	0.74	0.29	0.79	0.84	0.65
73	(Montgomery Watson, 1993)	BDCM			0.11	0.44	0.07	0.03	0.22	0.20	0.92	0.21	0.24	0.11	0.50	0.32	0.01
74	(Montgomery Watson, 1993)	BDCM			0.03	0.08	0.04	0.26	0.06	0.18	0.69	0.00	0.05	0.07	0.37	0.09	0.24
75	(Montgomery Watson, 1993)	DBCM			0.10	0.17	0.07	0.05	0.36	0.29	0.80	0.00	0.48	0.10	0.25	0.47	0.21
76	(Montgomery Watson, 1993)	DBCM			0.14	0.02	0.13	0.53	0.50	0.45	0.20	0.09	0.09	0.10	0.72	0.56	0.19
77	(Montgomery Watson, 1993)	BF			0.10	0.01	0.10	0.06	0.28	0.76	0.30	0.15	0.23	0.00	0.01	0.45	0.00
78	(Rathbun, 1996b)	CF		0.97	0.33	0.32	0.49	0.25	0.55	0.83	0.00	0.12	0.38	0.33	0.61	0.45	0.69
79	(Rathbun, 1996b)	BDCM		0.86	0.07	0.00	0.14	0.02	0.16	0.71	0.26	0.01	0.00	0.07	0.02	0.23	0.18
80	(Rathbun, 1996b)	DBCM		0.94	0.02	0.00	0.18	0.50	0.16	0.68	0.48	0.17	0.09	0.02	0.68	0.38	0.09

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
81	(Rathbun, 1996b)	BF		0.78	0.00	0.06	0.02	0.16	0.00	0.85	0.52	0.15	0.39	0.09	0.71	0.18	0.07
15	(Rathbun, 1996a)	TTHM	685	0.99	0.34	0.14	0.84	0.41	0.43	0.32	0.85	0.42	0.29	0.22	0.75	0.74	0.31
16	(Rodriguez et al., 2000)	TTHM	1483	0.90	0.33	0.01	0.84	0.59	0.68	0.05	0.88	0.55	0.05	0.35	0.64	0.56	0.69
17	(Rodriguez et al., 2000)	TTHM	113	0.34	0.22	0.28	0.79	0.39	0.66	0.01	0.21	0.25	0.04	0.19	**0.82**	0.79	0.81
18	(Semerjian et al., 2009)	TTHM	53	0.38	0.26	0.02	0.83	0.26	0.37	0.45	0.62	0.43	0.43	0.14	0.96	0.61	0.45
19	(Semerjian et al., 2009)	TTHM	53	0.25	0.08	0.00	0.36	0.14	0.19	0.37	0.17	0.01	0.00	0.12	0.73	0.61	0.42
20	(Semerjian et al., 2009)	TTHM	53	0.39	0.26	0.13	0.45	0.14	0.41	0.81	0.16	0.09	0.21	0.13	0.79	0.46	0.16
21	(Semerjian et al., 2009)	TTHM	53	0.29	0.06	0.00	0.75	0.05	0.17	0.42	0.21	0.13	0.05	0.00	0.86	0.10	0.19
22	(Semerjian et al., 2009)	TTHM	60	0.23	0.17	0.01	0.35	0.13	0.33	0.43	0.19	0.00	0.01	0.10	0.79	0.56	0.30
23	(Semerjian et al., 2009)	TTHM	60	0.12	0.02	0.03	0.01	0.26	0.35	0.41	0.17	0.00	0.01	0.13	0.56	0.76	0.03
24	(Semerjian et al., 2009)	TTHM	60	0.33	0.07	0.01	0.08	0.01	0.20	0.46	0.19	0.01	0.02	0.07	0.78	0.54	0.19
25	(Semerjian et al., 2009)	TTHM	60	0.11	0.10	0.00	0.77	0.08	0.21	0.41	0.21	0.13	0.05	0.00	0.74	0.12	0.14
26	(Semerjian et al., 2009)	TTHM	160	0.54	0.05	0.11	0.03	0.01	0.18	0.35	0.59	0.03	0.01	0.10	0.54	0.76	0.02

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
27	(Semerjian et al., 2009)	TTHM	160	0.70	0.01	0.00	0.01	0.00	0.11	0.11	0.48	0.03	0.01	0.01	0.47	0.17	0.00
28	(Semerjian et al., 2009)	TTHM	160	0.31	0.02	0.04	0.01	0.21	0.34	0.48	0.24	0.00	0.01	0.12	0.53	0.76	0.02
29	(Semerjian et al., 2009)	TTHM	160	0.61	0.01	0.00	0.00	0.10	0.10	0.34	0.34	0.06	0.02	NA[f]	0.18	0.14	NA[f]
30	(Sérodes et al., 2003)[a]	TTHM	53	0.78	0.30	0.00	0.67	0.53	0.59	0.06	0.71	0.36	0.20	0.03	0.20	0.39	0.25
31	(Sérodes et al., 2003)[a]	TTHM	55	0.89	0.02	0.27	0.62	0.18	0.35	0.11	0.21	0.50	0.11	0.01	0.53	0.62	0.42
32	(Sérodes et al., 2003)[a]	TTHM	51	0.56	0.09	0.17	0.07	0.07	0.08	0.12	0.64	0.23	0.22	0.12	0.54	0.76	0.02
33	(Sohn et al., 2004)	TTHM		0.90	0.36	0.13	0.76	0.46	0.48	0.33	0.89	0.24	0.15	0.17	0.78	0.62	0.41
34	(Sohn et al., 2004)	TTHM		0.92	0.38	0.10	0.83	0.45	0.69	0.19	0.69	0.63	0.34	0.16	0.77	0.66	0.43
35	(Sohn et al., 2004)	TTHM	288[g]	0.92	0.30	0.02	0.66	0.55	0.43	0.07	0.88	0.15	0.01	0.33	0.80	0.74	0.77
36	(Sohn et al., 2004)	TTHM		0.91	0.30	0.02	0.79	0.57	0.61	0.04	0.89	0.52	0.04	0.35	0.63	0.54	0.64
37	(Sohn et al., 2004)	TTHM		0.91	0.35	0.11	0.78	0.47	0.51	0.18	0.85	0.42	0.14	0.42	0.76	0.73	0.63
38	(Toroz and Uyak, 2005)[a]	TTHM	13	0.91	0.18	0.00	0.40	0.30	0.64	0.13	0.19	0.01	0.02	0.37	0.77	0.86	0.58
39	(Urano et al., 1983)	TTHM			0.33	0.04	0.77	0.57	0.60	0.06	0.86	0.59	0.10	0.23	0.59	0.59	0.59
40	(Uyak et al., 2005)[a]	TTHM		0.96 (0.86)	0.45	0.18	0.80	0.45	NA[f]	NA[f]	NA[f]	NA[f]	NA[f]	0.39	0.63	0.47	0.67
41	(USEPA, 2001)	TTHM	288	0.92	0.33	0.06	0.79	0.49	0.44	0.19	0.90	0.23	0.08	0.35	0.93	0.88	0.70
82	(USEPA, 2001)	CF	288	0.90	0.20	0.08	0.43	0.17	0.33	0.58	0.08	0.95	0.89	0.41	0.96	0.83	0.60

Model #^{j}	Source	Species	Source Model^i		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
83	(USEPA, 2001)	BDCM	288	0.85	0.14	0.01	0.46	0.34	0.25	0.56	0.65	0.01	0.00	0.31	0.83	0.81	0.55
84	(USEPA, 2001)	DBCM	245	0.94	0.01	0.00	0.00	0.02	0.11	0.04	0.39	0.05	0.02	0.02	0.01	0.00	0.00
85	(USEPA, 2001)	BF	171	0.93	0.00	0.00	0.04	0.25	0.10	0.60	0.37	0.06	0.03	0.13	0.07	0.05	0.08
42	(USEPA, 2001)	TTHM	2946	0.83	0.26	0.06	0.73	0.44	0.31	0.16	0.92	0.19	0.07	0.35	0.88	0.86	0.67
86	(USEPA, 2001)	CF	2960	0.73	0.21	0.09	0.61	0.23	0.49	0.63	0.27	0.87	0.49	0.31	0.90	0.80	0.68
87	(USEPA, 2001)	BDCM	2959	0.58	0.34	0.08	0.78	0.48	0.47	0.53	0.88	0.25	0.06	0.42	0.92	0.89	0.46
88	(USEPA, 2001)	DBCM	2948	0.73	0.20	0.03	0.70	0.39	0.24	0.26	0.85	0.10	0.03	0.31	0.88	0.86	0.00
89	(USEPA, 2001)	BF	2948	0.77	0.06	0.01	0.43	0.29	0.14	0.55	0.68	0.00	0.00	0.24	0.61	0.72	0.06
43	(Zhu, 1995)	TTHM	786	0.81	0.29	0.12	0.85	0.48	0.68	0.03	0.54	0.60	0.18	0.28	0.76	0.70	0.58
90	(Zhu, 1995)	CF	786	0.79	0.28	0.10	0.85	0.46	0.76	0.80	0.52	0.62	0.17	0.32	0.80	0.72	0.64
91	(Zhu, 1995)	BDCM	786	0.67	0.26	0.18	0.86	0.47	0.53	0.60	0.49	0.66	0.19	0.15	0.72	0.71	0.40
92	(Zhu, 1995)	DBCM	786	0.79	0.00	0.03	0.01	0.05	0.07	0.01	0.69	0.11	0.20	0.02	0.11	0.07	0.03
93	(Zhu, 1995)	BF	786	0.68	0.10	0.16	0.51	0.27	0.06	0.60	0.21	0.01	0.21	0.00	0.02	0.05	0.20
44	(Zhu, 1995)	TTHM	786	0.83	0.30	0.00	0.56	0.52	0.47	0.07	0.77	0.47	0.34	0.28	0.72	0.67	0.53
94	(Zhu, 1995)	CF	786	0.81	0.31	0.00	0.62	0.53	0.52	0.48	0.76	0.52	0.35	0.33	0.75	0.68	0.58
95	(Zhu, 1995)	BDCM	786	0.71	0.27	0.00	0.59	0.53	0.44	0.38	0.79	0.39	0.30	0.11	0.71	0.67	0.29
96	(Zhu, 1995)	DBCM	786	0.80	0.06	0.15	0.04	0.08	0.09	0.02	0.59	0.34	0.37	0.01	0.10	0.02	0.01
97	(Zhu, 1995)	BF	786	0.75	0.09	0.29	0.37	0.20	0.28	0.51	0.00	0.29	0.42	0.01	0.09	0.03	0.02

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2
45	(Zhu, 1995)	TTHM	786		0.33	0.01	0.84	0.60	0.61	0.04	0.91	0.49	0.04	0.34	0.72	0.65	0.79
46	(Zhu, 1995)	TTHM	786	0.70	0.40	0.16	0.73	0.47	0.57	0.31	0.89	0.33	0.23	0.49	0.82	0.83	0.67
98	(Zhu, 1995)	CF	786	0.61	0.38	0.20	0.70	0.35	0.60	0.60	0.42	0.75	0.62	0.45	0.87	0.82	0.64
99	(Zhu, 1995)	BDCM	786	0.88	0.07	0.01	0.44	0.25	0.16	0.62	0.55	0.00	0.00	0.27	0.88	0.87	0.37
100	(Zhu, 1995)	DBCM	786	0.89	0.01	0.00	0.01	0.00	0.11	0.02	0.38	0.05	0.02	0.00	0.11	0.12	0.01
101	(Zhu, 1995)	BF	786	0.61	0.04	0.02	0.06	0.11	0.00	0.08	0.76	0.01	0.00	0.14	0.02	0.02	0.11
47	(Zhu, 1995)	TTHM	786	0.81	0.38	0.08	0.86	0.50	0.55	0.19	0.89	0.41	0.18	0.40	0.86	0.83	0.68
102	(Zhu, 1995)	CF	786	0.73	0.34	0.09	0.82	0.40	0.61	0.62	0.55	0.82	0.56	0.32	0.91	0.84	0.65
103	(Zhu, 1995)	BDCM	786	0.90	0.08	0.00	0.50	0.28	0.16	0.57	0.54	0.00	0.00	0.22	0.82	0.86	0.41
104	(Zhu, 1995)	DBCM	786	0.89	0.05	0.00	0.18	0.36	0.01	0.47	0.38	0.16	0.07	0.16	0.17	0.14	0.13
105	(Zhu, 1995)	BF	786	0.61	0.02	0.02	0.05	0.12	0.03	0.13	0.76	0.01	0.00	0.16	0.06	0.00	0.07
Number of Samples					**75**	**30**	**45**	**25**	**25**	**25**	**15**	**15**	**15**	**124**	**6**	**24**	**76**
Minimum				**0.11 (0.84)**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
Maximum				**0.99 (0.98)**	**0.45**	**0.54**	**0.89**	**0.64**	**0.76**	**0.85**	**0.96**	**0.96**	**0.89**	**0.49**	**0.96**	**0.89**	**0.84**
Average				**0.75 (0.89)**	**0.17**	**0.09**	**0.42**	**0.28**	**0.32**	**0.33**	**0.53**	**0.25**	**0.17**	**0.17**	**0.57**	**0.50**	**0.34**
Standard Deviation				**0.22 (0.08)**	**0.13**	**0.12**	**0.32**	**0.20**	**0.21**	**0.25**	**0.31**	**0.25**	**0.20**	**0.14**	**0.29**	**0.28**	**0.27**

[a] Model developed using TOC rather than DOC

[b] R^2 not reported for validation data set

[c] Independent external validation performed by Greiner et al. (1992)

[d] External validation reported by Zhu (1995) using data from Montgomery Watson (1991)

[e] Reported by Zhu (1995)

Model #[j]	Source	Species	Source Model[i]		Current Study									(Gerrity et al., 2009)			
			# of Samples	Validation	All Data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	All Data	Raw Water	Coagulated Water	TiO_2

[f] NA - Not applicable; model does not return meaningful results. Possible reasons include all predictions are identical (so no R^2 value), input data missing (e.g., no UV_{254}), etc.

[g] Reported by USEPA (2001)

[h] Some models could not be calculated for some data sets due to lack of UV_{254} data. For this subset of data/models, n = 181 for the AOP data set, n = 201 for non-AOP data, n = 143 for TiO_2 data set, and n = 38 for all other types of AOP data

[i] Values reported are internal modeling values for # of sources and validation, while values shown in parentheses are for external validation of model

[j] Model # refers to number assigned in Table S4

[k] NR – Not reported; although validation was performed, R^2 was not provided

Table S9 (Continued) – Average correlation coefficients (R^2 values) for the predicted THMs versus measured THMs

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
1	(Al-Omari et al., 2004)[a]	TTHM	0.56	0.73		0.95	0.08		0.73	0.29	0.29	0.87	0.32
2	(Amy et al., 1998)	TTHM	0.56	0.81		0.97	0.08		0.73	0.04	0.29	0.87	0.34
48	(Amy et al., 1998)	CF			0.52			0.06				0.84	
49	(Amy et al., 1998)	BDCM										0.74	
50	(Amy et al., 1998)	DBCM										NA[f]	
51	(Amy et al., 1998)	BF										NA[f]	
3	(Amy et al., 1998)	TTHM	0.56	0.76		0.96	0.08		0.74	0.09	0.29	0.86	0.33
52	(Amy et al., 1998)	CF			0.72			0.06				0.84	

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
53	(Amy et al., 1998)	BDCM										0.74	
54	(Amy et al., 1998)	DBCM										NA[f]	
55	(Amy et al., 1998)	BF										NA[f]	
4	(Amy et al., 1998)	TTHM	0.56	0.78		0.96	0.08		0.73	0.10	0.29	0.86	0.33
56	(Amy et al., 1998)	CF			0.73			0.07				0.84	
57	(Amy et al., 1998)	BDCM										0.74	
58	(Amy et al., 1998)	DBCM										NA[f]	
59	(Amy et al., 1998)	BF										NA[f]	
5	(Amy et al., 1998)	TTHM	0.56	0.77		0.96	0.08		0.73	0.09	0.29	0.86	0.33
60	(Amy et al., 1998)	CF			0.72			0.06				0.84	
61	(Amy et al., 1998)	BDCM										0.74	
62	(Amy et al., 1998)	DBCM										NA[f]	
63	(Amy et al., 1998)	BF										NA[f]	
6	(Boyalla, 2004)	TTHM	0.57	0.75		0.96	0.08		0.57	0.55	0.29	0.88	0.33
7	(Chang et al., 1996)	TTHM	0.91	0.88		0.86	NA[f]		0.42	0.07	0.70	NA[f]	0.97
8	(Chang et al., 1996)	TTHM	0.56	0.65		0.93	0.07		0.76	0.01	0.29	0.86	0.31
9	(Chang et al., 1996)	TTHM	0.90	0.88		0.86	NA[f]		0.39	0.10	0.70	NA[f]	0.97

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
10	(Chen and Westerhoff, 2010)	TTHM	0.85	0.93		0.89	NA[f]		0.41	0.12	0.71	NA[f]	0.98
64	(Chen and Westerhoff, 2010)	CF			NA[f]			NA[f]				NA[f]	
65	(Chen and Westerhoff, 2010)	BDCM										NA[f]	
11	(Harrington et al., 1992)	TTHM	0.81	0.94		0.95	NA[f]		0.58	0.01	0.73	NA[f]	0.99
12	(Hong et al., 2007)	TTHM	NA[f]	0.50		0.89	0.07		NA[f]	0.31	0.29	NA[f]	0.30
66	(Hong et al., 2007)	BDCM										NA[f]	
67	(Hong et al., 2007)	CF			0.27			0.00				NA[f]	
68	(Malcom Pirnie Inc, 1992)	CF			NA[f]							NA[f]	
69	(Malcom Pirnie Inc, 1992)	BDCM										NA[f]	
70	2(Malcom Pirnie Inc, 1992)	DBCM										NA[f]	
71	(Malcom Pirnie Inc, 1993)	CF			NA[f]							NA[f]	
13	(McBean et al., 2008)	TTHM	**0.56**	0.68		0.96	0.08		0.74	**0.06**	0.29	**0.86**	0.33
14	(Milot et al., 2002)	TTHM	0.56	0.81		0.97	0.08		**0.72**	0.08	**0.29**	**0.87**	0.34
72	(Montgomery Watson, 1993)	CF			NA[f]							NA[f]	
73	(Montgomery Watson, 1993)	BDCM										0.74	

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
74	(Montgomery Watson, 1993)	BDCM										0.74	
75	(Montgomery Watson, 1993)	DBCM										NA[f]	
76	(Montgomery Watson, 1993)	DBCM										NA[f]	
77	(Montgomery Watson, 1993)	BF										NA[f]	
78	(Rathbun, 1996b)	CF			0.48			0.06				0.84	
79	(Rathbun, 1996b)	BDCM										0.74	
80	(Rathbun, 1996b)	DBCM										NA[f]	
81	(Rathbun, 1996b)	BF										NA[f]	
15	(Rathbun, 1996a)	TTHM	0.87	0.92		0.89	NA[f]		0.43	0.08	0.71	NA[f]	0.98
16	(Rodriguez et al., 2000)	TTHM	0.56	0.80		0.97	0.08		0.72	0.08	0.29	0.87	0.34
17	(Rodriguez et al., 2000)	TTHM	**0.56**	0.81		0.97	0.08		0.73	0.01	0.29	**0.86**	**0.34**
18	(Semerjian et al., 2009)	TTHM	0.88	0.90		0.87	NA[f]		0.39	0.12	0.70	NA[f]	0.97
19	(Semerjian et al., 2009)	TTHM	0.74	0.95		0.94	NA[f]		0.46	0.12	0.71	NA[f]	0.99
20	(Semerjian et al., 2009)	TTHM	0.73	0.81		0.88	NA[f]		0.18	0.48	0.68	NA[f]	0.97
21	(Semerjian et al., 2009)	TTHM	0.94	0.65		0.72	NA[f]		0.23	0.11	0.42	NA[f]	0.89
22	(Semerjian et al., 2009)	TTHM	0.89	0.90		0.87	NA[f]		0.58	0.01	0.70	NA[f]	0.97
23	(Semerjian et al., 2009)	TTHM	0.56	0.00		0.00	NA[f]		0.76	0.70	NA[f]	0.86	NA[f]

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
24	(Semerjian et al., 2009)	TTHM	0.92	0.78		0.80	NA[f]		0.66	0.06	0.67	NA[f]	0.92
25	(Semerjian et al., 2009)	TTHM	0.97	0.65		0.72	NA[f]		0.23	0.11	0.52	NA[f]	0.92
26	(Semerjian et al., 2009)	TTHM	0.56	NA[f]		NA[f]	NA[f]		0.74	0.70	NA[f]	0.86	0.00
27	(Semerjian et al., 2009)	TTHM	0.56	0.00		0.00	NA[f]		0.76	0.57	0.00	0.86	NA[f]
28	(Semerjian et al., 2009)	TTHM	0.56	0.00		NA[f]	NA[f]		0.74	0.71	NA[f]	0.86	NA[f]
29	(Semerjian et al., 2009)	TTHM	0.56	NA[f]		NA[f]	0.00		0.26	0.12	NA[f]	0.88	0.00
30	(Sérodes et al., 2003)[a]	TTHM	0.56	0.80		0.97	0.08		0.74	0.14	0.29	0.86	0.34
31	(Sérodes et al., 2003)[a]	TTHM	0.57	0.76		0.88	0.13		0.56	0.01	0.28	0.88	0.43
32	(Sérodes et al., 2003)[a]	TTHM	0.56	NA[f]		NA[f]	NA[f]		0.74	0.70	NA[f]	0.86	0.00
33	(Sohn et al., 2004)	TTHM	0.84	0.95		0.92	NA[f]		0.48	0.02	0.71	NA[f]	0.98
34	(Sohn et al., 2004)	TTHM	0.82	0.94		0.94	NA[f]		0.56	0.00	0.73	NA[f]	0.99
35	(Sohn et al., 2004)	TTHM	0.56	0.77		0.96	0.08		0.73	0.09	0.29	0.86	0.33
36	(Sohn et al., 2004)	TTHM	0.56	0.82		0.97	0.09		0.71	0.11	0.29	0.87	0.34
37	(Sohn et al., 2004)	TTHM	0.84	0.93		0.93	NA[f]		0.54	0.00	0.73	NA[f]	0.99
38	(Toroz and Uyak, 2005)[a]	TTHM	0.56	0.68		0.94	0.07		0.73	0.42	0.29	0.86	0.32
39	(Urano et al., 1983)	TTHM	0.56	0.80		0.97	0.08		0.73	0.07	0.29	0.87	0.34
40	(Uyak et al., 2005)[a]	TTHM	NA[f]	0.82		0.97	NA[f]		0.73	0.62	0.29	0.87	0.34

Model #^{j}	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
41	(USEPA, 2001)	TTHM	0.82	0.94		0.94	NA^f		0.56	0.00	0.73	NA^f	0.99
82	(USEPA, 2001)	CF			NA^f			NA^f				NA^f	
83	(USEPA, 2001)	BDCM										NA^f	
84	(USEPA, 2001)	DBCM										NA^f	
85	(USEPA, 2001)	BF										NA^f	
42	(USEPA, 2001)	TTHM	0.83	0.93		0.93	NA^f		0.53	0.01	0.73	NA^f	0.99
86	(USEPA, 2001)	CF			NA^f			NA^f				NA^f	
87	(USEPA, 2001)	BDCM										NA^f	
88	(USEPA, 2001)	DBCM										NA^f	
89	(USEPA, 2001)	BF										NA^f	
43	(Zhu, 1995)	TTHM	0.56	0.80		0.97	0.08		0.74	0.10	0.29	0.86	0.34
90	(Zhu, 1995)	CF			0.59			0.05				0.84	
91	(Zhu, 1995)	BDCM										0.74	
92	(Zhu, 1995)	DBCM										NA^f	
93	(Zhu, 1995)	BF										NA^f	
44	(Zhu, 1995)	TTHM	0.56	0.80		0.97	0.08		0.74	0.10	0.29	0.86	0.34
94	(Zhu, 1995)	CF			0.59			0.05				0.84	
95	(Zhu, 1995)	BDCM										0.74	
96	(Zhu, 1995)	DBCM										NA^f	
97	(Zhu, 1995)	BF										NA^f	
45	(Zhu, 1995)	TTHM	0.56	0.81		0.97	0.08		0.73	0.04	0.29	0.87	0.34
46	(Zhu, 1995)	TTHM	0.81	0.96		0.93	NA^f		0.51	0.01	0.71	NA^f	0.99
98	(Zhu, 1995)	CF			NA^f			NA^f				NA^f	
99	(Zhu, 1995)	BDCM										NA^f	

Model #[j]	Source	Species	External AOP Data										
			(Lamsal et al., 2011)	(Liu et al., 2010)	(Liu et al., 2008a)	(Liu et al., 2008b)	(Liu et al., 2007)	(Philippe et al., 2010b)	(Philippe et al., 2010a)	(Dotson et al., 2010)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Kleiser and Frimmel, 2000)
100	(Zhu, 1995)	DBCM										NA[f]	
101	(Zhu, 1995)	BF										NA[f]	
47	(Zhu, 1995)	TTHM	0.82	0.94		0.94			0.55	0.01	0.73	NA[f]	0.99
102	(Zhu, 1995)	CF			NA[f]		NA[f]	NA[f]				NA[f]	
103	(Zhu, 1995)	BDCM										NA[f]	
104	(Zhu, 1995)	DBCM										NA[f]	
105	(Zhu, 1995)	BF										NA[f]	
Number of Samples			**5**	**12**	**10**	**10**	**5**	**54**	**6**	**14**	**9**	**4**	**14**
Minimum			**0.56**	**0.00**	**0.27**	**0.00**	**0.00**	**0.00**	**0.18**	**0.00**	**0.00**	**0.74**	**0.00**
Maximum			**0.97**	**0.96**	**0.73**	**0.97**	**0.13**	**0.07**	**0.76**	**0.71**	**0.73**	**0.88**	**0.99**
Average			**0.68**	**0.76**	**0.58**	**0.88**	**0.08**	**0.05**	**0.60**	**0.18**	**0.46**	**0.84**	**0.59**
Standard Deviation			**0.15**	**0.23**	**0.16**	**0.21**	**0.02**	**0.02**	**0.17**	**0.22**	**0.22**	**0.05**	**0.35**

[a] Model developed using TOC rather than DOC

[b] R^2 not reported for validation data set

[c] Independent external validation performed by Greiner et al. (1992)

[d] External validation reported by Zhu (1995) using data from Montgomery Watson (1991)

[e] Reported by Zhu (1995)

[f] NA - Not applicable; model does not return meaningful results. Possible reasons include all predictions are identical (so no R^2 value), input data missing (e.g., no UV$_{254}$), etc.

[g] Reported by USEPA (2001)

[h] Some models could not be calculated for some data sets due to lack of UV$_{254}$ data. For this subset of data/models, n = 181 for the AOP data set, n = 201 for non-AOP data, n = 143 for TiO$_2$ data set, and n = 38 for all other types of AOP data

[i] Values reported are internal modeling values for # of sources and validation, while values shown in parentheses are for external validation of model

[j] Model # refers to number assigned in Table S4

[k] NR – Not reported; although validation was performed, R^2 was not provided

Table S9 (Continued) – Average correlation coefficients (R^2 values) for pairs of predictive THM models versus measured THM datasets

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
1	(Al-Omari et al., 2004)[a]	TTHM		0.26	0.47	0.70	0.87	0.76	0.25	0.59	0.29	0.36
2	(Amy et al., 1998)	TTHM		0.17	0.62	0.70	0.79	0.88	0.48	0.41	0.64	0.56
48	(Amy et al., 1998)	CF	0.63	0.43					0.08	0.00	0.08	0.63
49	(Amy et al., 1998)	BDCM		0.12					0.70	0.61	0.70	
50	(Amy et al., 1998)	DBCM		0.19					0.09	0.07	0.09	
51	(Amy et al., 1998)	BF							0.01	0.17	0.01	
3	(Amy et al., 1998)	TTHM		0.56	0.80	0.75	0.41	0.83	0.38	0.38	0.49	0.37
52	(Amy et al., 1998)	CF	0.60	0.59					0.07	0.07	0.06	0.59
53	(Amy et al., 1998)	BDCM		0.62					0.75	0.55	0.75	
54	(Amy et al., 1998)	DBCM		0.72					0.04	0.01	0.04	
55	(Amy et al., 1998)	BF							0.00	0.23	0.00	
4	(Amy et al., 1998)	TTHM		0.51	0.80	0.75	0.43	0.78	0.23	0.38	0.42	0.32
56	(Amy et al., 1998)	CF	0.60	0.60					0.07	0.08	0.05	0.60
57	(Amy et al., 1998)	BDCM		0.62					0.74	0.55	0.74	
58	(Amy et al., 1998)	DBCM		0.70					0.07	0.00	0.07	
59	(Amy et al., 1998)	BF							0.00	0.25	0.00	

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO$_2$	Other AOPs
5	(Amy et al., 1998)	TTHM		0.72	0.80	0.75	0.42	0.80	0.32	0.37	0.45	0.34
60	(Amy et al., 1998)	CF	0.60	0.60					0.07	0.08	0.06	0.59
61	(Amy et al., 1998)	BDCM		0.62					0.75	0.55	0.75	
62	(Amy et al., 1998)	DBCM		0.72					0.06	0.00	0.06	
63	(Amy et al., 1998)	BF							0.00	0.24	0.00	
6	(Boyalla, 2004)	TTHM		0.11	0.28	0.57	0.76	0.35	0.10	0.44	0.07	0.22
7	(Chang et al., 1996)	TTHM		0.60	NA[f]	0.92	0.27	0.74	0.72	0.58	0.67	0.86
8	(Chang et al., 1996)	TTHM		0.61	0.18	0.47	0.06	0.58	0.55	0.16	0.44	0.78
9	(Chang et al., 1996)	TTHM		0.60	NA[f]	0.92	0.25	0.71	0.72	0.60	0.65	0.84
10	(Chen and Westerhoff, 2010)	TTHM		0.02	NA[f]	0.60	0.00	0.81	0.34	0.53	0.66	0.07
64	(Chen and Westerhoff, 2010)	CF	0.88	0.04					0.41	0.47	0.07	0.87
65	(Chen and Westerhoff, 2010)	BDCM		0.51					0.04	0.50	0.04	
11	(Harrington et al., 1992)	TTHM		0.28	NA[f]	0.87	0.92	0.87	0.57	0.69	0.66	0.90
12	(Hong et al., 2007)	TTHM		0.16	0.83	0.30	0.83	0.11	0.00	0.12	0.02	0.15
66	(Hong et al., 2007)	BDCM		0.01					0.49	0.49	0.49	
67	(Hong et al., 2007)	CF	0.49	0.37					0.02	0.01	0.01	0.49

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
68	(Malcom Pirnie Inc, 1992)	CF		0.24					0.45	0.42	0.06	0.95
69	(Malcom Pirnie Inc, 1992)	BDCM		0.13					0.69	0.70	0.69	
70	(Malcom Pirnie Inc, 1992)	DBCM		0.23					0.02	0.06	0.02	
71	(Malcom Pirnie Inc, 1993)	CF		0.18					0.46	0.38	0.05	0.95
13	(McBean et al., 2008)	TTHM		0.00	0.30	0.36	0.19	0.85	0.39	0.57	0.52	0.32
14	(Milot et al., 2002)	TTHM		0.26	0.44	0.71	0.86	0.84	0.53	0.60	0.58	0.46
72	(Montgomery Watson, 1993)	CF	0.91	0.24					0.55	0.19	0.75	0.91
73	(Montgomery Watson, 1993)	BDCM		0.02					0.39	0.62	0.39	
74	(Montgomery Watson, 1993)	BDCM		0.69					0.00	0.45	0.00	
75	(Montgomery Watson, 1993)	DBCM		0.34					0.14	0.04	0.14	
76	(Montgomery Watson, 1993)	DBCM		0.17					0.03	0.26	0.03	
77	(Montgomery Watson, 1993)	BF							0.01	0.25	0.01	
78	(Rathbun, 1996b)	CF	0.58	0.07					0.03	0.01	0.03	0.58
79	(Rathbun, 1996b)	BDCM		0.11					0.67	0.50	0.67	
80	(Rathbun, 1996b)	DBCM		0.15					0.01	0.08	0.01	
81	(Rathbun, 1996b)	BF							0.03	0.22	0.03	
15	(Rathbun, 1996a)	TTHM		0.12	NA[f]	0.92	0.81	0.77	0.71	0.74	0.67	0.85
16	(Rodriguez et al., 2000)	TTHM		0.26	0.46	0.72	0.86	0.84	0.54	0.60	0.60	0.46

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO$_2$	Other AOPs
17	(Rodriguez et al., 2000)	TTHM		0.00	0.09	0.41	0.32	0.85	0.48	0.37	0.70	0.38
18	(Semerjian et al., 2009)	TTHM		0.56	NA[f]	0.80	0.11	0.69	0.76	0.64	0.68	0.93
19	(Semerjian et al., 2009)	TTHM		0.02	NA[f]	0.57	0.00	0.44	0.21	0.44	0.61	0.28
20	(Semerjian et al., 2009)	TTHM		0.00	NA[f]	0.18	0.13	0.06	0.47	0.57	0.49	0.36
21	(Semerjian et al., 2009)	TTHM		0.00	NA[f]	0.46	0.00	0.08	0.05	0.22	0.09	0.42
22	(Semerjian et al., 2009)	TTHM		0.00	NA[f]	0.61	0.16	0.84	0.23	0.66	0.27	0.70
23	(Semerjian et al., 2009)	TTHM		NA[f]	0.21	0.54	0.18	0.71	0.11	0.60	0.07	0.23
24	(Semerjian et al., 2009)	TTHM		0.00	NA[f]	0.61	0.13	0.85	0.10	0.64	0.12	0.53
25	(Semerjian et al., 2009)	TTHM		0.00	NA[f]	0.53	0.00	0.11	0.08	0.24	0.13	0.47
26	(Semerjian et al., 2009)	TTHM		0.58	0.55	0.81	0.37	0.64	0.14	0.62	0.04	0.47
27	(Semerjian et al., 2009)	TTHM		0.17	0.55	0.71	0.51	0.09	0.00	0.29	0.00	0.21
28	(Semerjian et al., 2009)	TTHM		0.49	0.46	0.65	0.17	0.64	0.09	0.56	0.03	0.32
29	(Semerjian et al., 2009)	TTHM		0.10	NA[f]	0.55	0.34	NA[f]	NA[f]	0.00	NA[f]	0.08
30	(Sérodes et al., 2003)[a]	TTHM		0.56	0.01	0.68	0.24	0.83	0.61	0.56	0.53	0.73
31	(Sérodes et al., 2003)[a]	TTHM		0.61	0.31	0.36	0.14	0.72	0.20	0.50	0.48	0.17
32	(Sérodes et al., 2003)[a]	TTHM		0.56	0.37	0.50	0.33	0.64	0.36	0.62	0.10	0.74
33	(Sohn et al., 2004)	TTHM		0.62	NA[f]	0.92	0.47	0.88	0.76	0.64	0.78	0.87
34	(Sohn et al., 2004)	TTHM		0.34	NA[f]	0.88	0.39	0.70	0.78	0.61	0.73	0.91

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
35	(Sohn et al., 2004)	TTHM		0.28	0.75	0.75	0.85	0.81	0.23	0.42	0.39	0.39
36	(Sohn et al., 2004)	TTHM		0.19	0.47	0.71	0.92	0.80	0.47	0.64	0.50	0.44
37	(Sohn et al., 2004)	TTHM		0.12	NA[f]	0.87	0.91	0.83	0.66	0.72	0.72	0.92
38	(Toroz and Uyak, 2005)[a]	TTHM		0.00	0.25	0.52	0.17	0.73	0.22	0.67	0.02	0.30
39	(Urano et al., 1983)	TTHM		0.22	0.51	0.74	0.80	0.85	0.68	0.39	0.71	0.74
40	(Uyak et al., 2005)[a]	TTHM		0.00	NA[f]	0.28	0.38	0.48	0.00	0.05	0.10	0.09
41	(USEPA, 2001)	TTHM		0.29	NA[f]	0.88	0.83	0.93	0.53	0.61	0.81	0.81
82	(USEPA, 2001)	CF	0.97	0.53					0.45	0.40	0.01	0.97
83	(USEPA, 2001)	BDCM		0.26					0.78	0.76	0.78	
84	(USEPA, 2001)	DBCM		0.46					0.10	0.04	0.10	
85	(USEPA, 2001)	BF							0.00	0.19	0.00	
42	(USEPA, 2001)	TTHM		0.17	NA[f]	0.88	0.80	0.91	0.43	0.57	0.77	0.60
86	(USEPA, 2001)	CF	0.95	0.58					0.52	0.46	0.60	0.95
87	(USEPA, 2001)	BDCM		0.01					0.70	0.84	0.70	
88	(USEPA, 2001)	DBCM		0.04					0.18	0.00	0.18	
89	(USEPA, 2001)	BF							0.05	0.24	0.05	
43	(Zhu, 1995)	TTHM		0.07	0.39	0.65	0.56	0.85	0.35	0.59	0.28	0.65
90	(Zhu, 1995)	CF	0.59	0.23					0.06	0.05	0.04	0.58

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
91	(Zhu, 1995)	BDCM		0.14					0.78	0.62	0.78	
92	(Zhu, 1995)	DBCM		0.17					0.17	0.11	0.17	
93	(Zhu, 1995)	BF							0.03	0.04	0.03	
44	(Zhu, 1995)	TTHM	0.59	0.25	0.56	0.69	0.59	0.85	0.44	0.57	0.41	0.59
94	(Zhu, 1995)	CF		0.40					0.07	0.06	0.06	0.58
95	(Zhu, 1995)	BDCM		0.32					0.76	0.62	0.76	
96	(Zhu, 1995)	DBCM		0.02					0.27	0.09	0.27	
97	(Zhu, 1995)	BF							0.04	0.42	0.04	
45	(Zhu, 1995)	TTHM		0.17	0.62	0.70	0.79	0.88	0.48	0.41	0.64	0.56
46	(Zhu, 1995)	TTHM		0.17	NA[f]	0.91	0.91	0.87	0.77	0.75	0.80	0.90
98	(Zhu, 1995)	CF	0.91	0.43					0.63	0.54	0.61	0.91
99	(Zhu, 1995)	BDCM		0.11					0.66	0.78	0.66	
100	(Zhu, 1995)	DBCM		0.19					0.11	0.05	0.11	
101	(Zhu, 1995)	BF							0.00	0.18	0.00	
47	(Zhu, 1995)	TTHM		0.17	NA[f]	0.88	0.79	0.90	0.67	0.64	0.82	0.90
102	(Zhu, 1995)	CF	0.95	0.43					0.84	0.71	0.72	0.95
103	(Zhu, 1995)	BDCM		0.11					0.68	0.73	0.68	
104	(Zhu, 1995)	DBCM		0.19					0.01	0.15	0.01	

Model #[j]	Source	Species	External Non-AOP Data						Grouped Data			
			(Chin and Bérubé, 2005)	(Chowdhury et al., 2010)	(Elshorbagy, 2000)	(Gang et al., 2002)	(Hua and Reckhow, 2008)	(Zhang et al., 2011)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
105	(Zhu, 1995)	BF							0.00	0.17	0.00	
Number of Samples			**13**	**32**	**40**	**56**	**20**	**13**	**243[h]**	**261[h]**	**191[h]**	**52[h]**
Minimum			**0.49**	**0.00**	**0.01**	**0.18**	**0.00**	**0.06**	**0.00**	**0.00**	**0.00**	**0.07**
Maximum			**0.97**	**0.72**	**0.83**	**0.92**	**0.92**	**0.93**	**0.84**	**0.84**	**0.82**	**0.97**
Average			**0.72**	**0.29**	**0.46**	**0.67**	**0.47**	**0.69**	**0.33**	**0.39**	**0.34**	**0.58**
Standard Deviation			**0.18**	**0.23**	**0.23**	**0.19**	**0.32**	**0.25**	**0.28**	**0.24**	**0.30**	**0.27**

[a] Model developed using TOC rather than DOC

[b] R^2 not reported for validation data set

[c] Independent external validation performed by Greiner et al. (1992)

[d] External validation reported by Zhu (1995) using Montgomery Watson (1991)

[e] Reported by Zhu (1995)

[f] NA - Not applicable; model does not return meaningful results. Possible reasons include all predictions are identical (so no R^2 value), input data missing (e.g., no UV_{254}), etc.

[g] Reported by USEPA (2001)

[h] Some models could not be calculated for some data sets due to lack of UV_{254} data. For this subset of data/models, n = 181 for the AOP data set, n = 201 for non-AOP data, n = 143 for TiO_2 data set, and n = 38 for all other types of AOP data

[i] Values reported are internal modeling values for # of sources and validation, while values shown in parentheses are for external validation of model

[j] Model # refers to number assigned in Table S4

[k] NR – Not reported; although validation was performed, R^2 was not provided

Table S10 – Average correlation coefficients (R^2 values) for pairs of predictive HAA models versus measured HAA datasets

Model #[j]	Source	Species	Source Model[i]		Current Study									External Non-AOP
			# of Samples	Validation	All data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	(Fabbricino and Korshin, 2009)
	(Amy et al., 1998)	HAA5[d]	738 (60)		0.06	0.13	0.14	0.00	0.34	0.02	0.84	0.43	0.06	0.12
113	(Amy et al., 1998)	MCA	738 (60)	0.14	0.08	0.09	0.21	0.00	0.09	0.00	0.13	0.20	0.17	0.07
114	(Amy et al., 1998)	MBA	738 (60)	0.43	0.00	0.03	0.00	0.36	0.27	0.17	0.32	0.00	0.04	0.00
115	(Amy et al., 1998)	DCA	738 (60)	0.83 (0.86)	0.22	0.04	0.81	0.22	0.64	0.10	0.22	0.02	0.00	0.69
116	(Amy et al., 1998)	TCA	738 (60)	0.87 (0.91)	0.33	0.41	0.56	0.24	0.64	0.19	0.09	0.18	0.14	0.16
117	(Amy et al., 1998)	DBA	738 (60)	0.77	0.00	0.00	0.03	0.07	0.42	0.81	0.48	NA[f]	NA[f]	0.19
	(Amy et al., 1998)	HAA5[d]	288		0.03	0.06	0.05	0.00	0.13	0.00	0.22	0.46	0.02	0.24
118	(Amy et al., 1998)	MCA	288	0.3	0.02	0.01	0.14	0.01	0.13	0.01	0.03	0.22	0.24	0.23
119	(Amy et al., 1998)	MBA	288	0.33	0.03	0.01	0.33	0.02	0.33	0.24	0.30	0.01	0.03	0.15
120	(Amy et al., 1998)	DCA	288	0.84	0.21	0.30	0.48	0.12	0.39	0.19	0.25	0.10	0.00	0.59
121	(Amy et al., 1998)	TCA	288	0.93	0.34	0.55	0.47	0.23	0.64	0.27	0.01	0.32	0.11	0.14
122	(Amy et al., 1998)	DBA	288	0.85	0.00	0.00	0.03	0.05	0.44	0.46	0.50	NA[f]	NA[f]	0.29
	(Amy et al., 1998)	HAA5[d]	288		0.03	0.08	0.01	0.00	0.00	0.00	0.34	0.50	0.05	0.25
123	(Amy et al., 1998)	MCA	288	0.44	0.02	0.02	0.13	0.00	0.14	0.04	0.13	0.25	0.28	0.74
124	(Amy et al., 1998)	MBA	288	0.39	0.02	0.02	0.21	0.01	0.15	0.26	0.30	0.01	0.03	0.03
125	(Amy et al., 1998)	DCA	288	0.9	0.23	0.25	0.60	0.14	0.54	0.19	0.24	0.00	0.00	0.61
126	(Amy et al., 1998)	TCA	288	0.94	0.35	0.49	0.56	0.23	0.66	0.24	0.03	0.26	0.13	0.16
127	(Amy et al., 1998)	DBA	288	0.92	0.01	0.00	0.03	0.06	0.44	0.33	0.51	NA[f]	NA[f]	0.30

Model #[j]	Source	Species	Source Model[i]		Current Study									External Non-AOP
			# of Samples	Validation	All data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	(Fabbricino and Korshin, 2009)
	(Amy et al., 1998)	HAA5[d]	288		0.03	0.07	0.04	0.00	0.05	0.00	0.27	0.48	0.03	0.25
128	(Amy et al., 1998)	MCA	288	0.36	0.02	0.02	0.14	0.01	0.14	0.03	0.07	0.24	0.27	0.40
129	(Amy et al., 1998)	MBA	288	0.35	0.03	0.01	0.27	0.00	0.26	0.28	0.30	0.01	0.03	0.15
130	(Amy et al., 1998)	DCA	288	0.88 (0.95)	0.22	0.28	0.54	0.13	0.47	0.19	0.00	0.71	0.23	0.60
131	(Amy et al., 1998)	TCA	288	0.93 (0.97)	0.35	0.52	0.52	0.23	0.65	0.26	0.33	0.08	0.00	0.15
132	(Amy et al., 1998)	DBA	288	0.84	0.00	0.00	0.03	0.06	0.44	0.41	0.06	0.41	0.48	0.29
133	(Chen and Westerhoff, 2010)	DCA	207	0.84	0.15	0.36	0.30	0.04	0.15	0.48	0.86	0.10	0.15	0.44
134	(Chen and Westerhoff, 2010)	TCA	207	0.81	0.18	0.75	0.05	0.12	0.09	0.65	0.23	0.61	0.40	0.62
106	(McBean et al., 2008)	HAA	180	0.72	0.12	0.32	0.16	0.00	0.13	0.00	0.87	0.10	0.28	0.36
135	(Montgomery Watson, 1993)	MCA			0.07	0.10	0.17	0.00	0.33	0.00	0.19	NA[f]	NA[f]	0.18
136	(Montgomery Watson, 1993)	DCA			0.20	0.34	0.59	0.10	0.42	0.32	0.88	0.25	0.30	0.71
137	(Montgomery Watson, 1993)	TCA			0.03	0.05	0.57	0.00	0.02	0.36	0.03	0.17	0.09	0.16
138	(Montgomery Watson, 1993)	MBA			0.02	0.00	0.02	0.12	0.02	0.03	0.07	0.28	0.01	0.23
139	(Montgomery Watson, 1993)	DBA			0.29	0.07	0.42	0.05	0.16	0.51	0.01	0.10	0.14	0.83

Model #[j]	Source	Species	Source Model[i]		Current Study									External Non-AOP
			# of Samples	Validation	All data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	(Fabbricino and Korshin, 2009)
107	(Sérodes et al., 2003)[a]	HAA5	53	0.89	0.13	0.01	0.21	0.12	0.52	0.01	0.74	0.22	0.24	0.63
108	(Sérodes et al., 2003)[a]	HAA5	55	0.8	0.03	0.29	0.18	0.10	0.91	0.07	0.65	0.30	0.43	0.92
109	(Sérodes et al., 2003)[a]	HAA5	51	0.92	0.12	0.08	0.24	0.10	0.67	0.00	0.05	0.34	0.24	0.68
110	(USEPA, 2001)	HAA5	738	0.87	0.07	0.15	0.15	0.00	0.33	0.01	0.20	0.61	0.41	0.77
111	(USEPA, 2001)	HAA5	288	0.94	0.13	0.27	0.27	0.00	0.15	0.00	0.01	0.25	0.35	0.75
140	(USEPA, 2001)	MCA	288	0.27	0.01	0.05	0.20	0.00	0.12	0.31	0.52	NA[f]	NA[f]	0.02
141	(USEPA, 2001)	DCA	288	0.84	0.23	0.35	0.56	0.10	0.49	0.30	0.30	0.01	0.03	0.65
142	(USEPA, 2001)	TCA	288	0.9	0.32	0.73	0.38	0.18	0.55	0.43	0.76	0.23	0.28	0.17
143	(USEPA, 2001)	MBA	288	0.4	0.05	0.01	0.53	0.14	0.21	0.40	0.38	0.01	0.02	0.70
144	(USEPA, 2001)	DBA	288	0.82	0.01	0.00	0.02	0.01	0.44	0.51	0.46	0.48	0.33	0.56
112	(USEPA, 2001)	HAA5	2807	0.59	0.07	0.17	0.17	0.00	0.18	0.00	0.02	0.27	0.40	0.76
145	(USEPA, 2001)	MCA	2841	0.29	0.00	0.00	0.03	0.02	0.11	0.00	0.01	NA[f]	NA[f]	0.15
146	(USEPA, 2001)	DCA	2881	0.54	0.22	0.26	0.68	0.14	0.52	0.22	0.16	0.12	0.00	0.65
147	(USEPA, 2001)	TCA	2891	0.61	0.29	0.70	0.47	0.16	0.49	0.39	0.84	0.43	0.06	0.24
148	(USEPA, 2001)	MBA	2892	0.19	0.05	0.05	0.02	0.00	0.01	0.14	0.13	0.20	0.17	0.11
149	(USEPA, 2001)	DBA	2886	0.75	0.02	0.00	0.02	0.06	0.46	0.36	0.32	0.00	0.04	0.79
Number of Samples					**75**	**30**	**45**	**25**	**25**	**25**	**15**	**15**	**15**	**14**
Minimum				**0.14**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

Model #[j]	Source	Species	Source Model[i]		Current Study									External Non-AOP
			# of Samples	Validation	All data	Non-AOP	TiO_2	SW	GAC	DS	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}	(Fabbricino and Korshin, 2009)
Maximum				**0.94**	**0.35**	**0.75**	**0.81**	**0.36**	**0.91**	**0.81**	**0.88**	**0.71**	**0.48**	**0.92**
Average				**0.67**	**0.11**	**0.18**	**0.27**	**0.08**	**0.33**	**0.21**	**0.30**	**0.24**	**0.17**	**0.39**
Standard Deviation				**0.26**	**0.11**	**0.21**	**0.22**	**0.09**	**0.22**	**0.20**	**0.27**	**0.19**	**0.15**	**0.27**

[a] Model developed using TOC rather than DOC

[b] R^2 not reported for validation data set

[c] Independent external validation performed by Greiner et al. (1992)

[d] No model number provided as HAA5 was calculated as the sum of 5 HAA

[e] Reported by Zhu (1995)

[f] NA - Not applicable; model does not return meaningful results. Possible reasons include all predictions are identical (so no R^2 value), input data missing (e.g., no UV_{254}), etc.

[g] Reported by USEPA (2001)

[h] Some models could not be calculated for some data sets due to lack of UV_{254} data. For this subset of data/models, n = 181 for the AOP data set, n = 201 for non-AOP data, n = 143 for TiO_2 data set, and n = 38 for all other types of AOP data

[i] Values reported are internal modeling values for # of sources and validation, while values shown in parentheses are for external validation of model

[j] Model # refers to number assigned in Table S4

Table S10 (Continued) – Average correlation coefficients (R^2 values) for pairs of predictive HAA models versus measured HAA datasets

Model #[j]	Source	Species	External AOP Data					Grouped Data			
			(Lamsal et al., 2011)	(Liu et al., 2008b)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Chin and Bérubé, 2005)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
	(Amy et al., 1998)	HAA5[d]	0.84	0.57	0.02	0.68	0.66	0.70	0.80	0.79	0.21
113	(Amy et al., 1998)	MCA				0.01		0.21	0.09	0.21	
114	(Amy et al., 1998)	MBA				NA[f]		0.00	0.33	0.00	
115	(Amy et al., 1998)	DCA				0.69		0.84	0.72	0.84	
116	(Amy et al., 1998)	TCA				0.68		0.84	0.61	0.84	

Model #[j]	Source	Species	External AOP Data					Grouped Data			
			(Lamsal et al., 2011)	(Liu et al., 2008b)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Chin and Bérubé, 2005)	All AOPs	All Non-AOP	All TiO$_2$	Other AOPs
117	(Amy et al., 1998)	DBA				NA[f]		0.02	0.52	0.02	
	(Amy et al., 1998)	HAA5[d]	0.84	0.85	0.02	0.68	0.63	0.58	0.56	0.65	0.44
118	(Amy et al., 1998)	MCA				0.01		0.12	0.00	0.12	
119	(Amy et al., 1998)	MBA				NA[f]		0.33	0.35	0.33	
120	(Amy et al., 1998)	DCA				0.69		0.82	0.73	0.82	
121	(Amy et al., 1998)	TCA				0.68		0.85	0.67	0.85	
122	(Amy et al., 1998)	DBA				NA[f]		0.02	0.63	0.02	
	(Amy et al., 1998)	HAA5[d]	0.84	0.87	0.02	0.68	0.75	0.58	0.66	0.77	0.38
123	(Amy et al., 1998)	MCA				0.01		0.12	0.01	0.12	
124	(Amy et al., 1998)	MBA				NA[f]		0.21	0.34	0.21	
125	(Amy et al., 1998)	DCA				0.69		0.83	0.76	0.83	
126	(Amy et al., 1998)	TCA				0.68		0.86	0.70	0.86	
127	(Amy et al., 1998)	DBA				NA[f]		0.02	0.62	0.02	
	(Amy et al., 1998)	HAA5[d]	0.84	0.86	0.02	0.68	0.66	0.59	0.62	0.72	0.42
128	(Amy et al., 1998)	MCA				0.01		0.12	0.00	0.12	
129	(Amy et al., 1998)	MBA				NA[f]		0.27	0.35	0.27	
130	(Amy et al., 1998)	DCA				0.69		0.82	0.74	0.82	
131	(Amy et al., 1998)	TCA				0.68		0.85	0.69	0.85	
132	(Amy et al., 1998)	DBA				NA[f]		0.02	0.62	0.02	
133	(Chen and Westerhoff, 2010)	DCA				NA[f]		0.30	0.77	0.30	
134	(Chen and Westerhoff, 2010)	TCA				NA[f]		0.05	0.90	0.05	
106	(McBean et al., 2008)	HAA	0.84	0.95	0.02	0.69	0.63	0.51	0.50	0.81	0.00

Model #[j]	Source	Species	External AOP Data					Grouped Data			
			(Lamsal et al., 2011)	(Liu et al., 2008b)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Chin and Bérubé, 2005)	All AOPs	All Non-AOP	All TiO$_2$	Other AOPs
135	(Montgomery Watson, 1993)	MCA				0.01		0.09	0.00	0.09	
136	(Montgomery Watson, 1993)	DCA				NA[f]		0.59	0.85	0.59	
137	(Montgomery Watson, 1993)	TCA				NA[f]		0.57	0.03	0.57	
138	(Montgomery Watson, 1993)	MBA				NA[f]		0.02	0.11	0.02	
139	(Montgomery Watson, 1993)	DBA				NA[f]		0.42	0.20	0.42	
107	(Sérodes et al., 2003)[a]	HAA5	0.84	0.95	0.02	0.69	0.62	0.64	0.56	0.69	0.21
108	(Sérodes et al., 2003)[a]	HAA5	0.84	0.78	0.01	0.68	0.74	0.32	0.42	0.90	0.09
109	(Sérodes et al., 2003)[a]	HAA5	0.84	0.94	0.02	0.69	0.65	0.54	0.57	0.77	0.16
110	(USEPA, 2001)	HAA5	0.84	0.94	0.02	0.68	0.65	0.68	0.64	0.80	0.20
111	(USEPA, 2001)	HAA5	0.79	0.96	0.40	NA[f]	0.93	0.87	0.70	0.99	0.60
140	(USEPA, 2001)	MCA				NA[f]		0.20	0.01	0.20	
141	(USEPA, 2001)	DCA				NA[f]		0.56	0.79	0.56	
142	(USEPA, 2001)	TCA				NA[f]		0.38	0.73	0.38	
143	(USEPA, 2001)	MBA				NA[f]		0.53	0.36	0.53	
144	(USEPA, 2001)	DBA				NA[f]		0.02	0.64	0.02	
112	(USEPA, 2001)	HAA5	0.75	0.89	0.41	NA[f]	0.91	0.70	0.74	0.90	0.49
145	(USEPA, 2001)	MCA				NA[f]		0.03	0.01	0.03	
146	(USEPA, 2001)	DCA				NA[f]		0.68	0.84	0.68	
147	(USEPA, 2001)	TCA				NA[f]		0.47	0.89	0.47	
148	(USEPA, 2001)	MBA				NA[f]		0.02	0.03	0.02	
149	(USEPA, 2001)	DBA				NA[f]		0.02	0.81	0.02	
Number of Samples			**6**	**5**	**9**	**4**	**13**	**77[h]**	**49**	**52[h]**	**25**

Model #[j]	Source	Species	External AOP Data					Grouped Data			
			(Lamsal et al., 2011)	(Liu et al., 2008b)	(Sarathy and Mohseni, 2010)	(Kent et al., 2011)	(Chin and Bérubé, 2005)	All AOPs	All Non-AOP	All TiO_2	Other AOPs
Minimum			**0.62**	**0.57**	**0.01**	**0.01**	**0.62**	**0.00**	**0.00**	**0.00**	**0.00**
Maximum			**0.93**	**0.96**	**0.41**	**0.69**	**0.93**	**0.87**	**0.90**	**0.99**	**0.60**
Average			**0.71**	**0.87**	**0.09**	**0.53**	**0.71**	**0.41**	**0.51**	**0.46**	**0.29**
Standard Deviation			**0.11**	**0.11**	**0.16**	**0.29**	**0.11**	**0.31**	**0.29**	**0.34**	**0.19**

[a] Model developed using TOC rather than DOC

[b] R^2 not reported for validation data set

[c] Independent external validation performed by Greiner et al. (1992)

[d] No model number provided as HAA5 was calculated as the sum of 5 HAA

[e] Reported by Zhu (1995)

[f] NA - Not applicable; model does not return meaningful results. Possible reasons include all predictions are identical (so no R^2 value), input data missing (e.g., no UV_{254}), etc.

[g] Reported by USEPA (2001)

[h] Some models could not be calculated for some data sets due to lack of UV_{254} data. For this subset of data/models, n = 181 for the AOP data set, n = 201 for non-AOP data, n = 143 for TiO_2 data set, and n = 38 for all other types of AOP data

[i] Values reported are internal modeling values for # of sources and validation, while values shown in parentheses are for external validation of model

[j] Model # refers to number assigned in Table S4

8. Statistical Analyses on Grouped Modeling Results

Coefficients of determination (R^2) were calculated for each model/dataset pairing, although not all models yielded meaningful predictions for all datasets as a result of missing input data, inability to calculate R^2 due to model returning identical values for all scenarios, etc. The R^2 values were grouped in accordance with level of photocatalytic/AOP treatment, type of source water, etc. in order to further explore model accuracy (i.e., the relationship between precursor measures and DBP formation) using statistical analyses, all of which were performed at a significance level of $\alpha = 0.05$.

8.1 Predictive Model Accuracy as a Function of Extent of Photocatalytic Treatment

The D'Agostino-Pearson omnibus test revealed that at least one THM and HAA dataset did not follow a normal distribution, as shown in Table S11, with normally distributed datasets shaded in gray.

Table S11 – P values from the D'Agostino-Pearson omnibus test for normality for R^2 values of DBP model predictions versus internal data grouped by extent of photocatalytic treatment

Data Group	THMs	HAAs
Dark Adsorption	<0.0001	0.0016
5 kWh m^{-3}	<0.0001	0.0700
80 kWh m^{-3}	0.0004	0.2516
160 kWh m^{-3}	<0.0001	0.0138
n =	108 – 109	41 – 48

The Kruskal-Wallis nonparametric test was used to test for differences in group means. The tests revealed significant differences in the accuracy (R^2) of DBP model predictions among the levels of photocatalytic treatment tested in this study (P <0.0001 for THMs and P = 0.0256 for HAAs). Dunn's post hoc test was used to ascertain which groups were significantly different. The results are shown in Tables S12 and S13, with significantly different pairings shaded in gray.

Table S12 – Dunn's post-hoc analysis for THM model accuracy using internal data grouped by extent of photocatalytic treatment

Data Group	Dark Adsorption	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}
Dark Adsorption				
5 kWh m^{-3}	<0.0001			
80 kWh m^{-3}	<0.0001	<0.0001		
160 kWh m^{-3}	0.0267	<0.0001	0.1716	
n per group = 108 – 109				

Table S13 – Dunn's post-hoc analysis for HAA model accuracy using internal data grouped by extent of photocatalytic treatment

Data Group	Dark Adsorption	5 kWh m^{-3}	80 kWh m^{-3}	160 kWh m^{-3}
Dark Adsorption				
5 kWh m^{-3}	0.0411			
80 kWh m^{-3}	0.2455	>0.9999		
160 kWh m^{-3}	>0.9999	0.2632	0.9509	
n per group = 41 – 48				

8.2 Predictive Model Accuracy as a Function of Type of Source Water

The D'Agostino-Pearson omnibus test revealed that at least one THM and HAA dataset did not follow a normal distribution, as shown in Table S14, with normally distributed datasets shaded in gray.

Table S14 – P values from the D’Agostino-Pearson omnibus test for normality for R^2 values of THM and HAA model predictions versus internal data grouped by type of source water

Data Group	THMs	HAAs
Untreated	<0.0001	<0.0001
Coagulated	<0.0001	[a]
Settled	<0.0001	0.0034
Filtered	<0.0001	0.1910
Finished	<0.0001	0.0728
Synthetic	<0.0001	[a]
n =	104 – 591	48 – 114
[a] No HAA measurements for this type of source water reported in the datasets analyzed.		

The Kruskal-Wallis nonparametric test was used to test for differences in group means. The tests revealed significant differences in the accuracy (R^2) of DBP model predictions among the levels of photocatalytic treatment tested in this study ($P < 0.0001$ for THMs and HAAs). Dunn’s post hoc test was used to ascertain which groups were significantly different. The results are shown in Tables S15 and S16, with significantly different pairings shaded in gray.

Table S15 – Dunn’s post-hoc analysis for THM model accuracy for aggregated internal and external data grouped by source water type

Data Group	Untreated	Coagulated	Settled	Filtered	Finished	Synthetic
Untreated						
Coagulated	0.0001					
Settled	<0.0001	<0.0001				
Filtered	<0.0001	<0.0001	>0.9999			
Finished	<0.0001	0.0006	>0.9999	>0.9999		
Synthetic	<0.0001	<0.0001	>0.9999	>0.9999	>0.9999	
n = 104 – 591						

Table S16 – Dunn’s post-hoc analysis for HAA model predictions versus aggregated internal and external data grouped by source water type

Data Group	Untreated	Settled	Filtered	Finished
Untreated				
Settled	<0.0001			
Filtered	0.0317	<0.0001		
Finished	<0.0001	0.0512	0.1842	
n = 48 – 114				

8.3 DBP Model Accuracy for TiO_2 Photocatalysis, Other AOP, and Non-AOP Treatment

The D’Agostino-Pearson omnibus test revealed that at least one THM and HAA dataset did not follow a normal distribution, as shown in Table S17, with normally distributed datasets shaded in gray.

Table S17 – P values from the D’Agostino-Pearson omnibus test for normality for R^2 values of THM and HAA model predictions versus internal and external data grouped by type of AOP treatment

Data Group	THMs	HAAs
TiO_2 Photocatalysis	<0.0001	<0.0001

Non-TiO_2 AOP	<0.0001	<0.0001
Non-AOP	<0.0001	<0.0001
n =	380 – 1130	92 – 403

The Kruskal-Wallis nonparametric test was used to test for differences in group means. The tests revealed significant differences in the accuracy (R^2) of DBP model predictions among the levels of photocatalytic treatment tested in this study ($P < 0.0001$ for HAAs). However, no significant difference was observed for THMs ($P = 0.0568$). Dunn's post hoc test was used to ascertain which HAA groups were significantly different. The results are shown in Table S18, with significantly different pairings shaded in gray.

Table S18 – Dunn's post-hoc analysis for HAA model accuracy of data grouped by type of AOP treatment

Data Group	TiO_2 Photocatalysis	Non-TiO_2 AOPs	Non-AOPs
TiO_2 Photocatalysis			
Non-TiO_2 AOPs	0.0001		
Non-AOPs	0.09204	0.1424	
n = 92 – 403			

References

Al-Omari, A., Fayyad, M., Abdel, A., 2004. Modeling trihalomethane formation for Jabal Amman water supply in Jordan. Environ. Model. Assess. 9, 245–252.

Amy, G., Siddiqui, M., Ozekin, K., Zhu, H.W., Wang, C., 1998. Empirically based models for predicting chlorination and ozonation by-products: trihalomethanes, haloacetic acids, chloral hydrate, and bromate. USEPA, EPA 815-R-98, Cincinnati, OH.

Amy, G.L., Chadik, P.A., Chowdhury, Z.K., 1987. Developing models for predicting trihalomethane formation potential and kinetics. J. Am. Water Work. Assoc. 79(7), 89–97.

Boyalla, R.L., 2004. Formation and modeling of disinfection by-products in Newfoundland communities. Thesis, Memorial University of Newfoundland.

Brookman, R.M., McClintock, A.M., Gagnon, G.A., 2011. A method for the detection of bromate in brackish water. Aquac. Eng. 45(1), 9–12.

Brown, D., 2009. The Management of Trihalomethanes in Water Supply Systems. Thesis, University of Birmingham.

Chang, E.E., Chao, S.H., Chiang, P.C., Lee, J.-F., 1996. Effects of chlorination on THMs formation in raw water. Toxicol. Environ. Chem. 56(1-4), 211–225.

Chang, E.E., Chiang, P.C., Chao, S.H., Lin, Y.L., 2006. Relationship between chlorine consumption and chlorination by-products formation for model compounds. Chemosphere 64(7), 1196–1203.

Chen, B., Westerhoff, P., 2010. Predicting disinfection by-product formation potential in water. Water Res. 44(13), 3755–3762.

Chin, A., Bérubé, P.R., 2005. Removal of disinfection by-product precursors with ozone-UV advanced oxidation process. Water Res. 39(10), 2136–2144.

Chow, C., Fabris, R., Cook, D., Drikas, M., 2004. Evaluation of Treatment Systems for Removal of Natural Organics. Research Report 23, Cooperative Research Centre for Water Quality and Treatment, South Australia.

Chowdhury, S., Champagne, P., McLellan, P.J., 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: a chronological review. Sci. Total Environ. 407(14), 4189–4206.

Chowdhury, S., Champagne, P., McLellan, P.J., 2010. Investigating effects of bromide ions on trihalomethanes and developing model for predicting bromodichloromethane in drinking water. Water Res. 44(7), 2349–2359.

Dotson, A.D., Keen, V.O.S., Metz, D., Linden, K.G., 2010. UV/H_2O_2 treatment of drinking water increases post-chlorination DBP formation. Water Res. 44(12), 3703–3713.

Elshorbagy, W.A., 2000. Kinetics of THM species in finished drinking water. J. Water Resour. Plan. Manag. 126(1), 21–28.

Espigares, M., Lardelli, P., Ortega, P., 2003. Evaluating trihalomethane content in drinking water on the basis of common monitoring parameters: regression models. J. Environ. Health 66(3), 9–13.

EU, 1998. Council Directive [WWW Document]. URL http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1998:330:0032:0054:EN:PDF

Fabbricino, M., Korshin, G. V, 2009. Modelling disinfection by-products formation in bromide-containing waters. J. Hazard. Mater. 168(2-3), 782–786.

Gang, D., Clevenger, T.E., Banerji, S.K., 2003. Relationship of chlorine decay and THMs formation to NOM size. J. Hazard. Mater. 96(1), 1–12.

Gang, D.D., Robert, L., Thomas, E., Shankha, K., 2002. Using chlorine demand to predict TTHM and HAA9 formation. J. Am. Water Work. Assoc. 94(10), 76–86.

Gerrity, D., Mayer, B., Ryu, H., Crittenden, J., Abbaszadegan, M., 2009. A comparison of pilot-scale photocatalysis and enhanced coagulation for disinfection byproduct mitigation. Water Res. 43(6), 1597–1610.

Golfinopoulos, S.K., Arhonditsis, G.B., 2002. Multiple regression models: a methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics. Chemosphere 47(9), 1007–1018.

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

Greiner, A.D., Obolensky, A., Singer, P.C., 1992. Technical note: Comparing predicted and observed concentrations of DBPs. J. Am. Water Work. Assoc. 84(11), 99–102.

Harrington, G.W., Chowdhury, Z.K., Owen, D.M., 1992. Developing a computer model to simulate DBP formation in drinking water treatmen. J. Am. Water Work. Assoc. 84(11), 78–87.

Health Canada, 2012. Guidelines for Canadian Drinking Water Quality - Summary Table. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Canada, Ottawa, Ontaria, Canada.

Hong, H.C., Liang, Y., Han, B.P., Mazumder, A., Wong, M.H., 2007. Modeling of trihalomethane (THM) formation via chlorination of the water from Dongjiang River (source water for Hong Kong's drinking water). Sci. Total Environ. 385(1-3), 48–54.

Hua, G., Reckhow, D.A., 2008. DBP formation during chlorination and chloramination: Effect of reaction time, pH, dosage, and temperature. J. Am. Water Work. Assoc. 100(8), 82–95.

Kent, F.C., Montreuil, K.R., Brookman, R.M., Sanderson, R., Dahn, J.R., Gagnon, G.A., 2011. Photocatalytic oxidation of DBP precursors using UV with suspended and fixed TiO_2. Water Res. 45(18), 6173–6180.

Kleiser, G., Frimmel, F.H., 2000. Removal of precursors for disinfection by-products (DBPs)--differences between ozone- and OH-radical-induced oxidation. Sci. Total Environ. 256(1), 1–9.

Kulkarni, P., Chellam, S., 2010. Disinfection by-product formation following chlorination of drinking water: artificial neural network models and changes in speciation with treatment. Sci. Total Environ. 408(19), 4202–4210.

Lamsal, R., Walsh, M.E., Gagnon, G.A., 2011. Comparison of advanced oxidation processes for the removal of natural organic matter. Water Res. 45(10), 3263–3269.

Li, X., Zhao, H., 2006. Development of a model for predicting trihalomethanes propagation in water distribution systems. Chemosphere 62(6), 1028–1032.

Liu, S., Lim, M., Chiang, K., Amal, R., Fabris, R., Chow, C., Drikas, M., 2007. A study on the removal of humic acid using advanced oxidation processes. Sep. Sci. Technol. 42(7), 1391–1404.

Liu, S., Lim, M., Fabris, R., Chow, C., Chiang, K., Drikas, M., Amal, R., 2008a. Removal of humic acid using TiO_2 photocatalytic process--fractionation and molecular weight characterisation studies. Chemosphere 72(2), 263–271.

Liu, S., Lim, M., Fabris, R., Chow, C., Drikas, M., Amal, R., 2008b. TiO_2 photocatalysis of natural organic matter in surface water: impact on trihalomethane and haloacetic acid formation potential. Environ. Sci. Technol. 42(16), 6218–6223.

Liu, S., Lim, M., Fabris, R., Chow, C.W.K., Drikas, M., Korshin, G., Amal, R., 2010. Multi-wavelength spectroscopic and chromatography study on the photocatalytic oxidation of natural organic matter. Water Res. 44(8), 2525–2532.

Malcom Pirnie Inc, 1992. Water Treatment Plant Simulation Program, Version 1.21, User's Manual. Washington, D.C.

Malcom Pirnie Inc, 1993. Bay-Delta water quality modeling report No. 15-041. Metropolitan Water District of Southern California, Los Angeles, CA.

Mayer, B.K., Daugherty, E., Abbaszadegan, M., 2014. Disinfection byproduct formation resulting from settled, filtered, and finished water treated by titanium dioxide photocatalysis. Chemosphere 117, 72–78.

McBean, E., Zhu, Z., Zeng, W., 2008. Systems analysis models for disinfection by-product formation in chlorinated drinking water in Ontario. Civ. Eng. Environ. Syst. 25(2), 127–138.

Milot, J., Rodriguez, M.J., Se, J.B., 2002. Contribution of neural networks for modeling trihalomethanes occurrence in drinking water. J. Water Resour. Plan. Manag. 128(5), 370–377.

Montgomery Watson, 1991. Disinfection/disinfection by-products database and model project. Denver, CO.

Montgomery Watson, 1993. Mathematical modeling of the formation of THMs and HAAs in chlorinated natural waters. American Water Works Association, Denver, CO.

NHMRC, NRMMC, 2011. Australian drinking water guidelines paper 6 national water quality management strategy. National Health and Medical Research Council, National Resource Management Ministerial Council, Commonwealth of Australia, Canberra.

Owen, D.M., Amy, G.L., Chowdhury, Z.K., 1993. Characterization of natural organic matter and its relationship to treatability, in: AWWA. Denver, CO.

Philippe, K.K., Hans, C., MacAdam, J., Jefferson, B., Hart, J., Parsons, S.A., 2010a. Photocatalytic oxidation, GAC and biotreatment combinations: An alternative to the coagulation of hydrophilic rich waters? Environ. Technol. 31(13), 1423–1434.

Philippe, K.K., Hans, C., MacAdam, J., Jefferson, B., Hart, J., Parsons, S.A., 2010b. Photocatalytic oxidation of natural organic matter surrogates and the impact on trihalomethane formation potential. Chemosphere 81(11), 1509–1516.

Rathbun, R.E., 1996a. Regression equations for disinfection by-products for the Mississippi, Ohio and Missouri rivers. Sci. Total Environ. 191(3), 235–244.

Rathbun, R.E., 1996b. Speciation of trihalomethane mixtures for the Mississippi, Missouri, and Ohio Rivers. Sci. Total Environ. 180(2), 125–135.

Rodrigues, M.S.M., Silva, J.C.G., Anutenes, M.C.G., 2007. Factorial analysis of the trihalomethane formation in water disinfection using chlorine. Anal. Chim. Acta 595(1-2), 266–274.

Rodriguez, M.J., Serodes, J., Morin, M., 2000. Estimation of water utility compliance with trihalomethane regulations using a modeling approach. J. Water Supply Res. Technol. AQUA 49, 57–73.

Sadiq, R., Rodriguez, M.J., 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: A review. Sci. Total Environ. 321(1-3), 21–46.

Sarathy, S., Mohseni, M., 2010. Effects of UV/H2O2 advanced oxidation on chemical characteristics and chlorine reactivity of surface water natural organic matter. Water Res. 44(14), 4087–4096.

Semerjian, L., Dennis, J., Ayoub, G., 2009. Modeling the formation of trihalomethanes in drinking waters of Lebanon. Environ. Monit. Assess. 149(1-4), 429–436.

Sérodes, J.-B., Rodriguez, M.J., Li, H., Bouchard, C., 2003. Occurrence of THMs and HAAs in experimental chlorinated waters of the Quebec City area (Canada). Chemosphere 51(4), 253–263.

Sohn, J., Amy, G., Cho, J., Lee, Y., Yoon, Y., 2004. Disinfectant decay and disinfection by-products formation model development: chlorination and ozonation by-products. Water Res. 38(10), 2461–2478.

Toroz, I., Uyak, V., 2005. Seasonal variations of trihalomethanes (THMs) in water distribution networks of Istanbul City. Desalination 176(1-3), 127–141.

Urano, K., Wada, H., Takemasa, T., 1983. Empirical rate equation for trihalomethane formation with chlorination of humic substances in water. Water Res. 17(12), 1797–1802.

USEPA, 2001. WTP Model v. 2.0 Manual.

USEPA, 2009. National Primary Drinking Water Regulations [WWW Document]. EPA 816-F-09-0004. URL http://water.epa.gov/drink/contaminants/index.cfm.

Uyak, V., Toroz, I., Meriç, S., 2005. Monitoring and modeling of trihalomethanes (THMs) for a water treatment plant in Istanbul. Desalination 176(1-3), 91–101.

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. Environ. Sci. Technol. 37(20), 4702–4708.

Westerhoff, P., Debroux, J., Amy, G.L., Gatel, D., Mary, V., Cavard, J., 2000. Applying DBP models to full-scale plants. J. Am. Water Work. Assoc. 92(3), 89–102.

WHO, 2008. Guidelines for drinking water quality, 3, V. 1. ed. World Health Organization, Geneva.

Zhang, J., Yu, J., An, W., Liu, J., Wang, Y., Chen, Y., Tai, J., Yang, M., 2011. Characterization of disinfection byproduct formation potential in 13 source waters in China. J. Environ. Sci. 23(2), 183–188.

Zhu, H.W., 1995. Modeling the effects of coagulation on chlorination by-product formation. Thesis, University of Colorado.