

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/41120920>

Bayesian Statistical Modeling of Disinfection Byproduct (DBP) Bromine Incorporation in the ICR Database

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · FEBRUARY 2010

Impact Factor: 5.33 · DOI: 10.1021/es9028606 · Source: PubMed

CITATIONS

11

READS

48

3 AUTHORS:



[Royce A Francis](#)

George Washington University

13 PUBLICATIONS 113 CITATIONS

[SEE PROFILE](#)



[Jeanne Vanbriesen](#)

Carnegie Mellon University

83 PUBLICATIONS 1,740 CITATIONS

[SEE PROFILE](#)



[Mitchell J. Small](#)

Carnegie Mellon University

189 PUBLICATIONS 2,686 CITATIONS

[SEE PROFILE](#)

Bayesian Statistical Modeling of Disinfection Byproduct (DBP) Bromine Incorporation in the ICR Database

ROYCE A. FRANCIS,^{*,†}
JEANNE M. VANBRIESEN,[‡] AND
MITCHELL J. SMALL^{‡,§}

*Department of Geography and Environmental Engineering,
Johns Hopkins University, 313 Ames Hall, 3400 N. Charles
Street, Baltimore, Maryland 21218, Department of Civil and
Environmental Engineering, and Department of Engineering
and Public Policy, Carnegie Mellon University, 5000 Forbes
Avenue, Pittsburgh, Pennsylvania 15213*

*Received September 22, 2009. Revised manuscript received
December 17, 2009. Accepted January 7, 2010.*

Statistical models are developed for bromine incorporation in the trihalomethane (THM), trihaloacetic acids (THAA), dihaloacetic acid (DHAA), and dihaloacetonitrile (DHAN) subclasses of disinfection byproducts (DBPs) using distribution system samples from plants applying only free chlorine as a primary or residual disinfectant in the Information Collection Rule (ICR) database. The objective of this study is to characterize the effect of water quality conditions before, during, and post-treatment on distribution system bromine incorporation into DBP mixtures. Bayesian Markov Chain Monte Carlo (MCMC) methods are used to model individual DBP concentrations and estimate the coefficients of the linear models used to predict the bromine incorporation fraction for distribution system DBP mixtures in each of the four priority DBP classes. The bromine incorporation models achieve good agreement with the data. The most important predictors of bromine incorporation fraction across DBP classes are alkalinity, specific UV absorption (SUVA), and the bromide to total organic carbon ratio (Br:TOC) at the first point of chlorine addition. Free chlorine residual in the distribution system, distribution system residence time, distribution system pH, turbidity, and temperature only slightly influence bromine incorporation. The bromide to applied chlorine (Br:Cl) ratio is not a significant predictor of the bromine incorporation fraction (BIF) in any of the four classes studied. These results indicate that removal of natural organic matter and the location of chlorine addition are important treatment decisions that have substantial implications for bromine incorporation into disinfection byproduct in drinking waters.

Introduction

Disinfection byproducts (DBPs) in drinking water have been of great regulatory concern since their identification in the

1970s. Current DBP regulations address the occurrence of four trihalomethanes (THM) and five haloacetic acids (HAA5) in distributed drinking waters (1, 2), primarily because these were considered the most prevalent in drinking water treated with free chlorine.

The regulatory tools in place to address these potential risks have been questioned by new concerns over the influence of bromine and iodine substitution on byproduct toxicity (3), extrapolation of single-chemical studies to mixtures relevant to human exposures (4), and interpretation of mass-based occurrence and exposure assessment information (5). Recent laboratory evidence demonstrating the differences in toxicity of unregulated DBPs (6), in addition to the development of analytical methods for quantifying halogen-based total organic halogen (TOX) fractions (7), and new statistical evidence that bromine substitution patterns are correlated across classes (8), present a need for modeling bromine substitution in DBP mixtures.

Modeling and predictive approaches to DBP formation and occurrence have been of particular importance for evaluating treatment alternatives with respect to public health consequences. Most approaches employ a logarithmic power statistical model for predicting TTHM or HAA5 (9). These approaches typically include explanatory variables such as pH, TOC, UV254, SUVA, temperature, distribution system residence time, chlorine residual, bromide concentration, and chlorine demand. While most studies have been formulated to focus on predicting concentrations of individual DBP species, or sum concentrations of species within particular DBP subclasses, others have been designed to explain DBP speciation and bromine incorporation (8, 10–14). Statistical models have occasionally achieved good agreement with observed data, especially when class-sums (e.g., TTHM, HAA5) are used as the dependent variable. This, however, has rendered some models particularly vulnerable to lack of generalizability, especially when developed from low-bromide; low-TOC; or low-bromide and low TOC; data sets (15). Most studies focusing on bromine incorporation have been laboratory-based (10, 11, 13, 14, 16–18), while at least two recent studies describe bromine substitution patterns in the Information Collection Rule database (8, 19).

A nationwide statistical modeling study of DBP mixtures in terms of the bromine incorporation fraction (BIF) using the Information Collection Rule (ICR) database is important for several reasons. First, as previously mentioned, most modeling approaches have focused on aggregated subgroups of DBPs, not individual species. These aggregated subgroups may be representative of waters with low bromide content (e.g., <0.1 mg/L); however, they are unlikely to be representative of waters with higher bromide. Brominated DBPs are generally not correlated, or are negatively correlated, with fully chlorinated DBPs. Consequently, when the DBP mixture is expected to be dominated by fully chlorinated byproducts, existing models provide representative results and are useful to inform risk models. However, when brominated DBP concentrations are a substantial fraction of the DBPs, existing models may be insufficient to predict DBPs and their associated risks. Thus, new models are needed for the potentially more important conditions under which brominated DBPs are expected.

Moreover, the bromine incorporation fraction has been suggested as an important metric for assessing toxicological similarity among DBP mixtures (20). Recent observations of a high degree of BIF correlation among DBP classes (8, 19) makes modeling the BIF more important. Successful models for bromine incorporation could be used in

* Corresponding author phone: +1 412.780.6453; fax: E-mail: +1 410.516.8996; e-mail: royce.francis@jhu.edu.

[†] Johns Hopkins University.

[‡] Department of Civil and Environmental Engineering, Carnegie Mellon University.

[§] Department of Engineering and Public Policy, Carnegie Mellon University.

TABLE 1. Summary Statistics for Explanatory Variables Considered in This Analysis^a

	<i>N</i>	min	median	max	average	std. dev.	MRL	% <MRL
pre-Cl	3381	N/A	N/A	N/A	0.783	N/A	N/A	N/A
time _{DBP} [h]	3381	1	20	408	30.1	35.58	N/A	N/A
Cl _{Plant} [mg/L]	3381	0.2	2.75	12.7	3.22	1.90	N/A	N/A
Cl _{DS_Free} [mg/L]	3381	0.1	0.8	6	0.878	0.594	0.1	5.03
pH _{DS}	3381	1.16	7.67	9.97	7.79	0.639	N/A	N/A
pH _{clpt}	3381	5	7.5	10.8	7.50	0.823	N/A	N/A
ALK _{clpt} [mg/L-CaCO ₃]	3381	2	59	500	73.84	65.880	2	0.30
TURB _{clpt} [NTU]	3381	0.01	1.6	340	6.20	19.91	0.01	1.48
TEMP _{clpt} [C]	3381	0.6	16.6	30	16.45	6.85	N/A	N/A
TOC _{clpt} [mg/L-C]	3381	0.7	1.95	13.2	2.15	1.20	0.7	10.1
Br _{inf} [mg/L]	3381	0.02	0.023	0.6	0.047	0.063	0.02	45.0
UV254 _{inf}	3381	0.009	0.0475	0.88	0.064	0.061	0.009	4.67
Br:TOC _{clpt} [mg/L-Br:mg/L-C]	3381	0.000299	0.00211	0.106	0.00	0.009	N/A	N/A
SUVA _{clpt} [mg/L·cm]	3381	0.00408	0.0181	0.293	0.02	0.02	N/A	N/A
Cl:TOC _{plant} [mg/L-Cl:mg/L-C]	3381	0.236	1.28	15	1.64	1.41	N/A	N/A
Br:Cl _{plant} [mg/L-Br:mg/L-Cl]	3381	0.00199	0.0141	0.706	0.03	0.060	N/A	N/A

^a For all measurements below the minimum reporting level (MRL), the minimum reporting level is substituted. Pre-Cl is a binary variable indicating whether a plant applies chlorine before a clarification or filtration step.

planning epidemiological and treatment process studies, for example, matching newly studied systems with those archived in U.S. Environmental Protection Agency databases intended to assess the toxicological similarity of DBP mixtures (21, 22).

Finally, modeling bromine incorporation is important for increasing understanding of the trade-offs between natural organic matter (NOM) removal and brominated DBP formation. Bromide concentrations are not removed through the treatment process, while NOM removal is designed to decrease DBP formation potential and enhance microbial protection. On the contrary, source water bromide is oxidized to aqueous bromine (HOBr) when chlorine is applied (the term chlorine refers to the applied oxidant in the form of HOCl and/or Cl₂). Thus, the interaction between NOM removal decisions and source water bromide concentrations has important implications for DBP mixture composition. When DBP mixture composition is studied in terms of the bromine incorporation fraction (11, 14), mixture composition is controlled by initial water quality measures such as the bromide to chlorine ratio or bromide to total organic carbon ratio. If similar patterns exist in nationwide distribution system data, such findings would have important implications for the discussion of future distribution system regulations and regulatory revisions, especially if enhanced control of brominated DBPs is desired. Our approach may also facilitate further discussion of the trade-offs between such treatment goals as TOC removal and brominated DBP reduction in predictive modeling for treatment process planning and risk assessment purposes.

Statistical Methods

In this study, we use Bayesian Markov Chain Monte Carlo (MCMC) methods to estimate the conditional and marginal posterior distributions of the mean vector and covariance matrix of the multivariate distribution of 16 DBPs in five priority subclasses, and the parameter vector and variance of linear models for BIF in four priority subclasses. The general steps of the modeling study include: imputation of missing and censored data in the DBP database (censored explanatory variables are replaced with values at the detection limit); estimation of the posterior distribution of the parameters for the joint distribution on the 16 DBPs; calculation of the (appropriately transformed) bromine incorporation fraction; and estimation of the posterior distribution for the coefficients of the linear

models developed. A Gibbs' sampling algorithm is used to implement the MCMC estimation.

Study Design. This study analyzes distribution system DBP measurements from the publicly available ICR database for plants with surface or groundwater as a source, and using only free chlorine as a preoxidant, primary, and/or residual disinfectant. The ICR database is extensively described elsewhere (23, 24). The extraction of the data used in this analysis, including 3381 total records, follows the approach of Obolensky and Singer (8) including only those DBP samples where at least one subclass (e.g., THM, THAA, DHAA, DHAN, MHA) had no missing observations. The extracted data do not include DBP classes with more than one carbon atom (e.g., haloketones, halofuranones); thus inferences drawn from this study may not be applicable to bromine incorporation into these DBP classes. Censored and missing measurements in the ICR DBP data have been identified as an important problem in previous studies (8, 19). The imputation steps we have employed for missing and censored DBP data are described in the authors' previous work (19). In short, we use a multivariate model to predict missing and censored measurements given the information encoded in the observed measurements. The explanatory variables considered are listed in Table 1, and the DBPs are listed in Table 2. The former include influent UV254 absorbance; an indicator variable for prechlorination; distribution system residence time; distribution system free chlorine residual; total chlorine applied through the plant; pH in the distribution system and at the first point of chlorine addition; and alkalinity, turbidity, temperature, and TOC at the first point of chlorine addition. No ICR records are included with missing explanatory variables, and left-censored explanatory variable measurements are replaced with their detection limit. It is important to point out that we do not perform multiple imputation for the explanatory variables. Although the decision to include only completely observed cases (e.g., nonmissing explanatory variable data) may have introduced some bias to our inferences (25), this bias is comparable to the amount expected when using only completely observed responses (e.g., nonmissing DBP data). We decided that the low correlation between the independent variables in the analysis would make it difficult to reduce the bias through imputation of the independent variables, and not worth the computational cost. For each ICR plant, the bromide concentration is taken from the influent measurement. Applied chlorine is calculated as the sum of applied disinfectant through the plant. Because of concerns about

TABLE 2. Summary Statistics for Dependent Variables (DBPs) Included in This Analysis^a

	<i>N</i>	min	median	max	average	std. dev.	MRL	fraction missing	fraction <MRL	fraction imputed
CHCL3	3088	0.00838	0.201	1.340	0.2384	0.1922	0.00838	0.087	0.046	0.133
BDCM	3122	0.00610	0.0482	0.305	0.0580	0.0438	0.00610	0.077	0.047	0.124
DBCM	3117	0.00482	0.0106	0.227	0.0222	0.0282	0.00482	0.078	0.285	0.363
CHBR3	3115	0.00396	0.00396	0.155	0.0074	0.0126	0.00396	0.079	0.700	0.779
TCAA	3057	0.00612	0.0612	0.730	0.0792	0.0700	0.00612	0.096	0.107	0.203
BDCAA	1029	0.00481	0.0125	0.0866	0.0182	0.0156	0.00481	0.696	0.045	0.741
DBCAA	945	0.00793	0.00793	0.0329	0.00978	0.00452	0.00793	0.720	0.213	0.933
TBAA	744	0.01348	0.0135	0.0290	0.0136	0.00107	0.0135	0.780	0.216	0.996
DCAA	3078	0.00776	0.0853	0.666	0.0984	0.0825	0.00776	0.090	0.077	0.167
BCAA	3062	0.00577	0.0138	0.202	0.0182	0.0149	0.00577	0.094	0.210	0.304
DBAA	3082	0.00459	0.00459	0.0918	0.00699	0.00701	0.00459	0.088	0.667	0.756
DCAN	2976	0.00455	0.0173	0.235	0.0207	0.0180	0.00455	0.120	0.144	0.264
BCAN	2990	0.00324	0.00534	0.0855	0.00792	0.00677	0.00324	0.116	0.315	0.431
DBAN	2954	0.00259	0.00259	0.0586	0.00491	0.00560	0.00259	0.126	0.517	0.644
MCAA	2981	0.0212	0.0212	0.878	0.0284	0.0394	0.0212	0.118	0.656	0.774
MBAA	3062	0.00720	0.00720	0.0792	0.00794	0.00380	0.00720	0.094	0.819	0.913

^a Missing and below MRL values included in DBP observations and are imputed during analysis. Fraction missing and fraction <MRL columns have the same denominator (*N*), and fraction imputed is the sum of these two columns.

collinearity and consistency with relevant laboratory results and other previous studies, we do not include influent bromide concentrations, UV254, and total applied chlorine as covariates in the model. Instead, we include the bromide to applied (through the plant) chlorine ratio, bromide to TOC ratio at the first point of chlorine addition, chlorine applied (through the plant) to TOC ratio, and specific UV absorbance (SUVA) at the first point of chlorine addition. A cross-validation approach is used to assess the influence of collinearity among the variables, as discussed below.

Estimation of the Bromine Incorporation Fraction. The bromine incorporation fraction (BIF) (8, 11) is the proportion of total halogen positions with bromine-substituted atoms in a DBP mixture (example shown for THM class):

$$\text{BIF}_{\text{THM}} = \frac{[\text{THM} - \text{Br}]}{[\text{THM} - \text{X}]} = \frac{0 \times [\text{CHCl}_3] + 1 \times [\text{BDCM}] + 2 \times [\text{DBCM}] + 3 \times [\text{CHBr}_3]}{3 \times ([\text{CHCl}_3] + [\text{BDCM}] + [\text{DBCM}] + [\text{CHBr}_3])} \quad (1)$$

Equivalent information is provide by φ , the ratio of bromine-substituted to chlorine-substituted halogen positions in a DBP class. For the THM class this ratio is calculated from THM species micromolar concentrations as follows (17):

$$\varphi_{\text{THM}} = \frac{[\text{THM} - \text{Br}]}{[\text{THM} - \text{Cl}]} = \frac{0 \times [\text{CHCl}_3] + 1 \times [\text{BDCM}] + 2 \times [\text{DBCM}] + 3 \times [\text{CHBr}_3]}{3 \times [\text{CHCl}_3] + 2 \times [\text{BDCM}] + 1 \times [\text{DBCM}] + 0 \times [\text{CHBr}_3]} \quad (2)$$

This ratio is related to the BIF through the following:

$$\varphi_{\text{THM}} = \frac{\text{BIF}_{\text{THM}}}{1 - \text{BIF}_{\text{THM}}} \quad (3)$$

Due to its range from 0 to 1, we employ the logit transformation of the BIF:

$$\text{BF}_{\text{THM}} = \text{logit}(\text{BIF}_{\text{THM}}) = \ln(\varphi_{\text{THM}}) = \ln\left(\frac{\text{BIF}_{\text{THM}}}{1 - \text{BIF}_{\text{THM}}}\right) \quad (4)$$

Q-Q plots indicate the approximate normality of BF for each class studied (results not shown). Furthermore, the residuals for the linear models developed in this study are approximately normal (results not shown).

To estimate the model coefficients β , we employ a sequential Gibbs' sampler in the linear regression context. To develop the regression model for BF, we use the following factorization:

$$p(\text{BF}|X, \beta, \sigma_{\text{BF}}^2, Y) \propto \iint p(f|X, \beta, \sigma_{\text{BF}}^2) \times p(Y_{\text{mis}}, Y_{\text{cens}}, Y_{\text{obs}}|\mu_Y, \Sigma_Y) p(\mu_Y, \Sigma_Y) p(\beta, \sigma_{\text{BF}}^2) d\mu_Y d\Sigma_Y \quad (5)$$

where Y are concentrations of individual DBP species in $\mu\text{mol/L}$, X are explanatory variables, β are linear coefficients, σ_{BF}^2 is the variance of BF, μ_Y and Σ_Y are the parameters of the joint distribution of DBP concentrations, and $f|Y$ is the transformation from concentration data to the class BF. This factorization permits sequential updating of the parameters for the multivariate distributions on the dependent variables (e.g., DBPs), and the parameters of the linear regression within the Gibbs' sampler. This sequential updating also permits posterior estimation using complete-data methods. Given the complete data set, we write the overall data likelihood as a normal distribution, assuming the residuals are normally distributed with equal variance, conditionally independent given \mathbf{x} , the explanatory variables. Assuming an informationless prior distribution for $(\beta, \sigma_{\text{BF}}^2)$, the posterior distributions for the parameters are found as follows. Setting $y_i = \text{BF}_i$, we write

$$y_i \sim N(\mathbf{x}_i \beta, \sigma_{\text{BF}}^2) \quad (6)$$

The posterior distribution on β and σ_{BF}^2 is proportional to

$$p(\beta, \sigma_{\text{BF}}^2|y) \propto \sigma_{\text{BF}}^{-1} \prod_{i=1}^n [N(y_i|\mathbf{x}_i \beta, \sigma_{\text{BF}}^2)] \propto \sigma_{\text{BF}}^{-(n+1)} \exp\left[-\frac{(n-k)s^2}{2\sigma_{\text{BF}}^2} - \frac{(\beta - \hat{\beta})' \mathbf{x}(\beta - \hat{\beta})}{2\sigma_{\text{BF}}^2}\right] \quad (7)$$

where

$$\hat{\beta} = (\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' (y_{\text{BF}})$$

$$s^2 = \left(\frac{1}{n-k}\right) (y_i - \mathbf{x} \hat{\beta})' (y_i - \mathbf{x} \hat{\beta})$$

The marginal posterior distributions for $\beta|\sigma_{\text{BF}}^2$ and σ_{BF}^2 are given by

$$\begin{aligned} p(\sigma_{BF}^2) &\sim \text{Inv} - \chi^2(n - k, s^2) \\ p(\beta | \sigma_{BF}^2) &\sim N(\hat{\beta}, (\mathbf{x}'\mathbf{x})^{-1} \sigma_{BF}^2) \end{aligned} \quad (8)$$

For further details on Gibbs' sampling and Bayesian linear modeling, the reader is referred to Gelman et al (26) and Box and Tiao (27).

Model Selection Using PSR Cross-Validation. The method of cross-validation in this study, predictive sample reuse (PSR), was chosen specifically for its theoretical basis for comparing and selecting from among alternative statistical models based on their ability to predict observed data, rather than solely their statistical fit to the observed data. Four alternative models for explaining bromine incorporation in each of the four DBP classes studied were constructed by review of previous bromine incorporation studies (8, 10–14), covariance assessment, and preliminary linear regression of the BF onto the water quality parameters in Table 1. These model alternatives consider collinearity between the bromide to applied chlorine ratio, the ratio of chlorine applied through the plant to influent TOC, and bromide to TOC ratio at the first point of chlorine addition. The models chosen each include all of the following variables: distribution system residence time, free chlorine residual, and pH; alkalinity, turbidity, temperature, and SUVA at first point of chlorine addition; and the chlorine to TOC ratio through the plant. The four alternative models differed with respect to whether they included the bromide to chlorine applied ratio, bromide to TOC ratio, chlorine applied to TOC ratio, or all three: Alternative 1 (A1): Excludes chlorine applied through the plant to influent TOC (Cl:TOC) ratio. Alternative 2 (A2): Excludes influent bromide to chlorine applied through the plant (Br:Cl) ratio. Alternative 3 (A3): Excludes bromide to TOC (Br:TOC) ratio at the first point of chlorine addition. Alternative 4 (A4): Includes all independent variables.

The best model from this subset is chosen using 10-fold predictive sample reuse (PSR) cross-validation (28, 29) for the bromine incorporation fraction models for each class. In short, the following general approach is applied to each of the 4 DBP classes studied: (1) The data set is split into 10 partitions; (2) One of the 10 partitions is reserved as the “hold-out” partition; (3) Each of the four model alternatives for a DBP class is fit to the nine “training” partitions; (4) The posterior predictive likelihood of the “hold-out” partition is calculated given the parameters obtained by fitting the model to the “training” partitions. The product of the 10 posterior predictive likelihoods obtained through this process for each model is used as the selection criterion to choose the best model alternative in each class. This method incorporates model selection and validation into one step by selecting the model on the basis of its predictive ability. A brief outline of the method follows. Let $y = (y_1, y_2, \dots, y_N)$ be N independent and identically distributed observations (e.g., the BF for each observation). Let $y_{(j)} = (y_1, y_2, \dots, y_{j-1}, y_{j+1}, \dots, y_N)$ represent the data set with the j^{th} observation omitted. For each alternative model, M , we have parameters θ and explanatory variables x . If we have l models, the posterior predictive distribution for a future observation (\tilde{y}, x) when M_l is true, conditional on the observed data set and parameter estimates is

$$f(\tilde{y} | y, x, M_l) \propto \int f(\tilde{y} | \theta, x, M_l) \prod_{i=1}^N f(y_i | \theta, x_i, M_l) p(\theta) d\theta_l \quad (19)$$

Now, suppose the data set is split into J groups (e.g., $J = 10$). For J iterations, we exclude group j from parameter estimation, then compute the likelihood of observing group

TABLE 3. Conditional Posterior PSR Likelihood of Observed BF for each DBP Class under Each Model Alternative^a

model	THM	THAA	DHAA	DHAN
A1	−4776	−4685	−4692	−4631
A2	−4773*	−4673*	−4679*	−4638
A3	−4826	−4702	−4790	−4718
A4	−4934	−4692	−4721	−4630*

^a Observed BF calculated at each iteration of the Gibbs' sampler using imputed data set. $N = 3381$. The preferred model has the maximum (least negative) PSR likelihood.

TABLE 4. Correlation between Observed and Predicted (Both Averaged over MCMC chain) BF in Each Class under the Selected Model Alternative^a

THM	THAA	DHAA	DHAN
0.812 (0.810, 0.813)	0.594 (0.587, 0.598)	0.707 (0.698, 0.714)	0.661 (0.654, 0.667)

^a 95% credible intervals shown in parentheses and italics. $N = 3381$.

j given the parameters estimated from the other observations. This is a slight modification from formal leave- n -out cross-validation in that we restrict the possible training sets to contiguous groups of N/J observations, not every possible combination of N/J observations. The model selection criterion is the product of the likelihood of observing each excluded group, L_i :

$$L_i = \prod_{j=1}^J f_j(y_j | y_{(j)}, x_{(j)}, M_l) \quad (18)$$

The model maximizing L_i is selected. We subsequently evaluate the model fit through consideration of the correlation coefficient of averaged MCMC predictions with the observed bromine incorporation fraction.

Results and Discussion

Model Assessment. The observed bromine incorporation fraction for each observation is computed by averaging each BF observation over the MCMC chain, using simulated values for censored and missing data. Left-censored and missing measurements are imputed for each draw of (μ, Σ) from their posterior distribution. After computing BF for each observation using these “complete” data sets, the BF is averaged over these imputations. MCMC BF predictions are averaged over each iteration of the MCMC chain for each bromine incorporation fraction corresponding to the original data set by computing predicted BF for each draw of β from its conditional posterior distribution, then averaging over this posterior distribution.

For model selection, the alternative model likelihoods (L_i from eq 18) are reported in Table 3, supporting the choice of A2 for modeling BF in the THM, THAA, and DHAA classes, while the DHAN BF is best represented using model A4. While the absolute PSR differences are small, one must keep in mind that one log unit difference is equal to a factor of 2.72. In other words, model A4 compared with model A1 for the DHAN class is approximately 2.72 times more likely to explain the data based on the PSR log-likelihood values. Table 4 reports the correlation coefficient between predicted bromine incorporation fractions from MCMC output for each class and the observed bromine incorporation fraction for the selected models. Figure 1 shows the model fit using average MCMC predictions

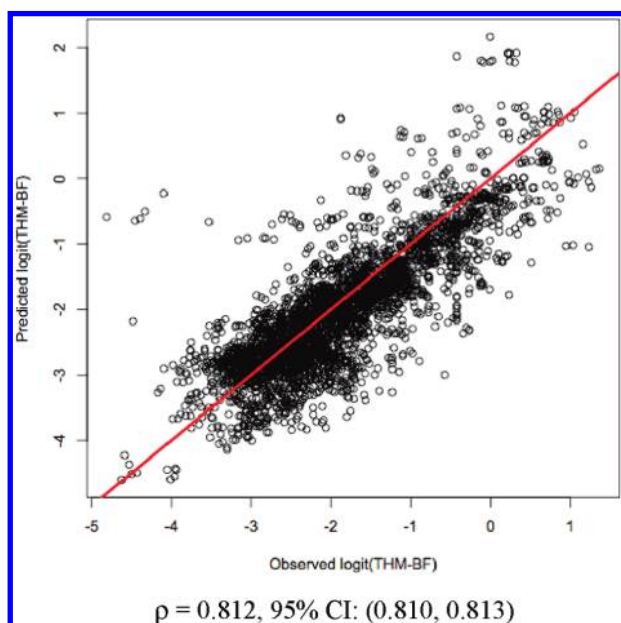


FIGURE 1. Model fit showing average (over all posterior parameter estimates obtained from multiple imputation method) $\text{logit}(\text{BIF})$ versus observed $\text{logit}(\text{BIF})$. Shown with imputed censored and missing measurements, for preferred model alternative in the THM class. Correlation coefficient between observed and predicted $\text{logit}(\text{BIF})$ with 95% credible intervals (CI) shown beneath the plot. Plots for remaining classes shown in the Supporting Information.

versus observed bromine incorporation fraction for the preferred model alternative for the THM class, with the 1:1 reference line in red. The model fit for the THAA, DHAA, and DHAN classes are illustrated in Supporting Information Figure S1. Table 4 shows that the selected models generate predictions fairly well correlated with observed bromine incorporation fraction data, with THMs having the highest levels of agreement between observed and predicted bromine incorporation fraction. The correlation coefficients in the remaining classes, especially the DHAN and THAA classes may be slightly lower due to the increased levels of bromine incorporation in these classes. Because the higher censoring levels increases the uncertainty associated with the more highly bromine-substituted compounds in these classes, our imputation technique may lead to a slightly lower correlation between observed and predicted BF values in these classes. Figure 1 and Supporting Information Figure S1 show that our models tend to slightly under-predict BF in each class studied.

Each model parameter is evaluated for convergence in the Gibbs' sampler using the potential scale reduction factor (PSRF). The authors have described the use of this metric for model evaluation elsewhere (19). For each bromine incorporation fraction model, high levels of convergence have been achieved (e.g., PSRFs sufficiently close to 1 for all parameters).

Bromine Incorporation Fraction Models. The results of the linear model alternatives described above for the bromine incorporation fraction in the THM, THAA, DHAA, and DHAN classes are presented in this section. SI Tables S1–S4 report the model coefficients, coefficient of variation, and 95% credible intervals for the preferred models in each class. The equations for each DBP class are presented in eqs 19–22, including only the statistically significant terms:

THM:

$$\begin{aligned} \text{logit}(\text{BF}_{\text{THM}}) = & -0.484 - 0.047 \ln(t_{\text{DS}}) + \\ & 0.117 \ln(\text{Cl}_{\text{DS_free}}) - 0.090 \ln(\text{pH}_{\text{DS}}) + 0.406 \ln(\text{ALK}_{\text{inf}}) - \\ & 0.007(\text{TEMP}_{\text{clpt}}) + 0.803 \ln(\text{Br:TOC}_{\text{clpt}}) - \\ & 0.329 \ln(\text{SUVA}_{\text{clpt}}) - 0.206 \ln(\text{Cl:TOC}_{\text{plant}}) + \\ & \varepsilon(\sigma_{\varepsilon} = 0.425) \end{aligned} \quad (19a)$$

THAA:

$$\begin{aligned} \text{logit}(\text{BF}_{\text{THAA}}) = & -1.222 - 0.082 \ln(t_{\text{DS}}) + \\ & 0.107 \ln(\text{pH}_{\text{DS}}) + 0.246 \ln(\text{ALK}_{\text{inf}}) - 0.032 \ln(\text{TURB}_{\text{clpt}}) - \\ & 0.015(\text{TEMP}_{\text{clpt}}) + 0.657 \ln(\text{Br:TOC}_{\text{clpt}}) - \\ & 0.347 \ln(\text{SUVA}_{\text{clpt}}) - 0.123 \ln(\text{Cl:TOC}_{\text{plant}}) + \\ & \varepsilon(\sigma_{\varepsilon} = 0.553) \end{aligned} \quad (20)$$

DHAA:

$$\begin{aligned} \text{logit}(\text{BF}_{\text{DHAA}}) = & -1.898 + 0.071 \ln(\text{Cl}_{\text{DS_free}}) + \\ & 0.409 \ln(\text{ALK}_{\text{inf}}) - 0.024 \ln(\text{TURB}_{\text{clpt}}) + \\ & 0.726 \ln(\text{Br:TOC}_{\text{clpt}}) - 0.327 \ln(\text{SUVA}_{\text{clpt}}) - \\ & 0.079 \ln(\text{Cl:TOC}_{\text{plant}}) + \varepsilon(\sigma_{\varepsilon} = 0.712) \end{aligned} \quad (21)$$

DHAN:

$$\begin{aligned} \text{logit}(\text{BF}_{\text{DHAN}}) = & -2.844 + 0.057 \ln(\text{Cl}_{\text{DS_free}}) + \\ & 0.246 \text{pH}_{\text{DS}} + 0.273 \ln(\text{ALK}_{\text{inf}}) + \\ & 0.723 \ln(\text{Br:TOC}_{\text{clpt}}) - 0.298 \ln(\text{SUVA}_{\text{clpt}}) + \\ & \varepsilon(\sigma_{\varepsilon} = 0.741) \end{aligned} \quad (22)$$

These models illustrate several interesting concepts concerning BIF in these four DBP classes. We can divide the models into three parts to facilitate the discussion: effects related to continuing reactions between NOM and free chlorine in the distribution system; effects related to NOM properties; and the interactions of bromide with free chlorine and NOM.

We deal first with effects of continuing reactions in the distribution system on BIF. For all four classes studied, BIF decreases as distribution system residence time increases. Although residence time is only selected in the THM and THAA classes, free chlorine residual is statistically significant for the THM and THAA classes, and free chlorine residual is statistically associated with THM, DHAA, and DHAN BIF. Because free chlorine residual is usually larger closest to points of chlorine addition, we can infer that this positive coefficient indicates an inverse relationship between BIF and residence time for each class studied. This, in turn, shows that continuing DBP reactions decrease BIF through production of more fully chlorinated DBPs in the distribution system. In our analysis, we also show that pH at the first point of chlorine addition is never selected into the predictive models; rather, pH in the distribution system is significant for THM, THAA, and DHAN BIF. The underlying reasons for significance of pH in the distribution system and not at the first point of chlorine addition are unclear. The role of pH in the HOBr/HOCl reaction system is well-known thus one might expect the pH at the first point of chlorine addition to be more important than the pH in the distribution system due to the effect of HOBr/HOCl competing substitution reactions on the BIF. However, our result suggests that hydrolysis and other intermediate distribution system reactions may be more important for distribution system BIF. For example, in the THM

class, the negative association between distribution system pH and BIF may be attributable to base-catalyzed hydrolysis reactions that may occur in the distribution system (30). These reactions may convert intermediate DBPs, dominated by chlorinated species, into THMs, reducing THM BIF. The opposite effect is observed for distribution system pH in the THAA and DHAN BIF models. The effect of pH on bromine incorporation in these classes has not been extensively evaluated, and may warrant additional research.

We turn next to effects of NOM properties on BIF in the four classes. Influent alkalinity has recently been associated with increased brominated DBP formation (15). We find that alkalinity at the first point of chlorine addition is positively associated with increases in BIF. This association is strongest for the THM and DHAA classes, while slightly lower for DHAN and THAA classes. Obolensky and Singer (15) were the first to identify the relationship between alkalinity and bromine incorporation. Citing results from Archer and Singer (31), they propose a hypothesis which takes into account the negative relationship between alkalinity and SUVA. Specifically, Obolensky and Singer suggest that hydrophobic NOM may be immobilized under high ionic strength and hardness conditions typical of high alkalinity source waters. They suggest that this, in turn, would reduce the number of DBP reactive sites since hydrophobic NOM is more DBP reactive than hydrophilic NOM. Consequently, they explain that higher alkalinity waters would have elevated bromide to DBP precursor concentrations, especially when overall concentrations of DBP precursors are low. However, the observation of this effect in the ICR database may be confounded by the influence of alkalinity on NOM removal via coagulation in drinking water plants. This process is dependent upon control to low pH (5.5–6.5), which can be difficult to achieve in waters with significant buffering (i.e., with high alkalinity). Future research into the relationship among alkalinity, BIF, and DBP formation is needed to understand the NOM properties that might be associated with alkalinity that contribute to increased bromine incorporation in DBP mixtures. Thus, in considering reasons for the effect of alkalinity on bromine incorporation, it is important to distinguish the role of *mineral alkalinity* in bromine substitution from the role of alkaline pH, which has been extensively evaluated, especially with respect to the competing hypobromous and hypochlorous acid reactions with NOM and their implications for individual DBP formation (10, 32).

After alkalinity, specific UV absorption (SUVA) at the first point of chlorine addition is also a strong predictor of BIF. While SUVA has been shown to be positively associated with DBP formation (12, 33), it has also been shown to be negatively associated with brominated DBP formation (12). In our analysis, we observe a negative association between BIF and SUVA in each class studied. Furthermore, its effect coefficient is consistent in magnitude for each of the four models, indicating a possible underlying NOM attribute closely related to BIF across DBP classes. As with alkalinity, this property's relationship to BIF has not been extensively studied.

The other variables that could indicate the effects of NOM on BIF, turbidity and temperature, are only marginally associated with BIF in the four classes. Turbidity is statistically significant only in the THM and DHAA models. Its negative coefficient supports the general finding that NOM is negatively associated with BIF, while temperature's negative coefficient is also consistent with the general notion that increased chlorine demand reduces bromine incorporation (11, 14). Krasner et al (11) find that when bromide concentration is an order of magnitude less than dissolved organic carbon (DOC), time and temperature have a marginal effect on DBP speciation.

Most bromine incorporation studies have addressed the interactions between bromide, NOM, and free chlorine

(10, 11, 13, 14, 16). These studies have focused on the bromide to chlorine applied (Br:Cl) ratio (10, 11, 16, 34) and the bromide to total organic carbon (Br:TOC) ratio (11, 14). In our analysis, Br:Cl through the plant is not selected in any of the predictive models. This may reflect the low bromide nature of this data set, as Symons et al (14) conclude that when DOC is not constant, Br:DOC controls speciation. In this nationwide data set, because TOC is not constant, and DBP formation reactions may not proceed to completion (10), the Br:Cl influence is not independent of TOC. Consequently, Br:TOC controls BIF, as discussed by Symons et al (14) and Cowman and Singer (10). Generally, since hypobromite is a more efficient oxidant and faster substitution agent than hypochlorite, when TOC is limited or reactions do not proceed to completion, the Br:TOC ratio will be more influential than the Br:Cl ratio for predicting BIF (32). The interaction between bromide and chlorine added is not lost, however, as the chlorine to TOC ratio through the plant (Cl:TOC) is selected as a moderate predictor of BIF in the THM, THAA, and DHAA classes. Cl:TOC has not been observed to control BIF or DBP speciation in previous studies; however, its negative coefficient and moderate magnitude indicate that bromide concentrations for this data set are sufficiently low that speciation is primarily controlled by Br:TOC.

Our findings, in conjunction with the laboratory studies referenced above, may suggest some implications for treatment considerations if brominated DBPs at current occurrence levels are found to be toxicologically important. Figure 2 illustrates the joint effect of Br:TOC and Cl:TOC on BIF using the THM class as a representative example. To construct this figure, the quantile of the average BIF computed over the MCMC chain for each observation ($N = 3381$) is graphed versus $\log(\text{Br:TOC})$ and $\log(\text{Cl:TOC})$. The size of the points indicates the quantile of the BIF distribution: the smallest points correspond to the lower quantiles, while the larger points corresponding to the upper quantiles. The colors indicate the quantiles of the THM concentrations: red circles indicate large total THM concentrations, while the blue circles indicate smaller total THM concentrations. The shape of the points indicate the source water type: circles indicate surface waters, triangles indicate ground waters, and squares indicate a mixture of the two.

This plot illustrates the risk–risk trade-off between total DBP concentration and bromine incorporation that is induced by the common approach of reducing TOC before chlorination. As expected, as TOC is reduced (move right on the figure), both Cl:TOC and Br:TOC are increased. While increases in Cl:TOC do not lead to increases in BIF (size of points shows no vertical pattern on the plot), increased Br:TOC shows a strong relationship to increased BIF (point size increases to the right on the plot), despite the moderate negative association between Cl:TOC and BIF. Obolensky and Singer (24) report that ICR plants with groundwater as their source water used higher Cl:TOC ratios relative to surface water plants, and free chlorine only plants used higher Cl:TOC ratios relative to plants employing chloramines. Thus, this trade-off is most important for plants with lower TOC concentrations, and groundwater chlorine-only plants relative to surface water chlorine-only plants. With this information, we can now split the plot vertically in half, corresponding generally to surface water classification and source water bromide concentration strata. On the left, most plants would be high TOC, low bromide surface waters, implied by low Br:TOC ratios. These plants would produce finished waters with moderate BIF. On the right side, these plants are generally high bromide groundwater plants (with a modest number of surface water plants), implied by the high Br:TOC ratios, producing finished waters with high BIF. These may be the most toxicologically important plants if bromine substitution is found to be toxicologically important. Fur-

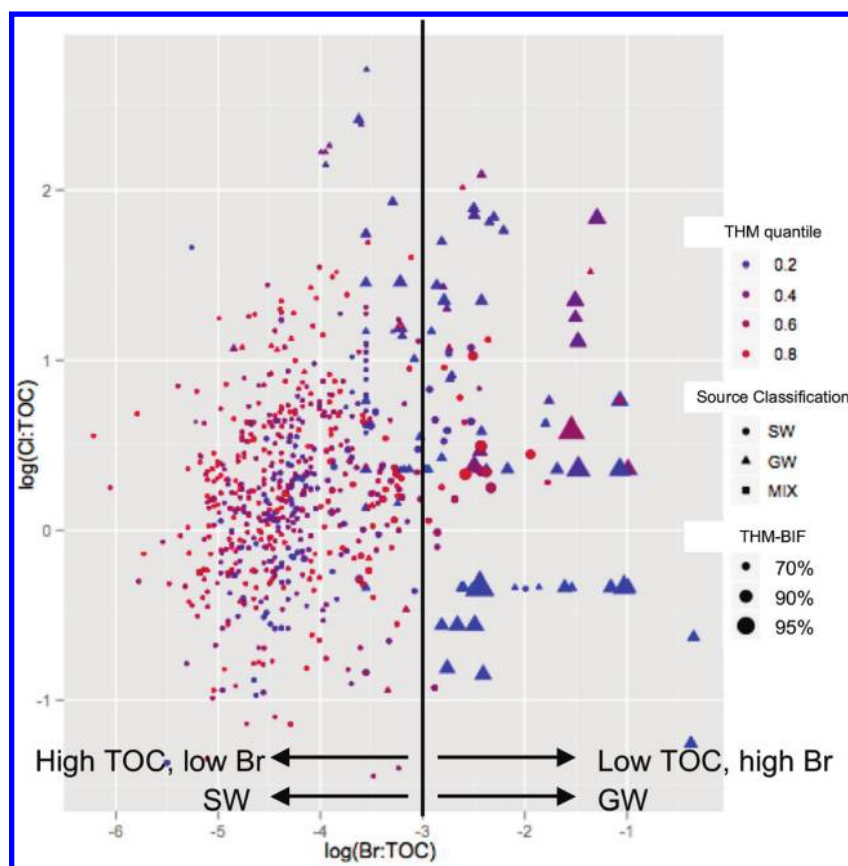


FIGURE 2. Interaction effects between Cl:TOC and Br:TOC for the THM class. Color indicates THM concentration quantile, size indicates BIF (in percentage). For example, large, red points correspond with large BIF and large THM concentration, small, blue points correspond with lower BIF and lower THM concentration. MCMC BIF shown versus water quality parameters, $N = 3381$.

thermore, plants indicated by moderate to large red points may be of the highest toxicological importance, as they have both the highest total DBP concentrations and BIF. Although the low-Br nature of this data set may limit the relative influence of Br:Cl on bromine incorporation, Figure 2, illustrating the oxidant-NOM relationships captured by this data set, shows the predominant influence of Br:TOC on bromine incorporation in chlorinated drinking waters, and its potential relevance for identifying toxicologically important water utilities. Indeed, methodological advances are being developed to employ the BIF as an important measure of toxicological similarity and an important metric for risk analysis (20, 35).

In summary, BIF may be successfully modeled using readily measured parameters such as those included here from the ICR database. Because DBP reactions may not proceed to completion under treatment conditions represented by this data set, bromine incorporation into chlorinated DBP mixtures is primarily controlled by alkalinity at the first point of chlorine addition, SUVA at the first point of chlorine addition, bromide to TOC ratio at the first point of chlorine addition, and chlorine applied through the plant to influent TOC ratio. Although the bromide to applied chlorine ratio has been observed in laboratory studies to control DBP bromine incorporation, these studies were conducted under conditions where the influence of the bromide to applied chlorine ratio is independent of TOC concentrations. This research shows that the principal treatment trade-off influencing bromine incorporation in DBP mixtures is the location of chlorine addition relative to processes removing natural organic matter. To best address potential treatment trade-offs, further research is needed in the following areas:

1. Greater examination of the influence of pH in the distribution system on the BIF in distributed waters.
2. Investigation into the association between distribution system free chlorine concentrations and BIF in distributed waters, especially for DBP classes expected to be significantly impacted by biotic degradation.
3. Evaluation of the interactions between Br:TOC, Cl:TOC, and Br:Cl, and bromine incorporation in treated waters, especially in waters with high Br:TOC ratios.
4. The influence of enhanced coagulation and alkalinity on bromine incorporation into chlorinated DBP mixtures.

Acknowledgments

This work was conducted under support from a National Science Foundation Graduate Research Fellowship, Carnegie Mellon Center for Water Quality in Urban Environmental Systems, and the H. John Heinz III Chair in Environmental Engineering.

Supporting Information Available

Tables S1–S4 contain summary statistics for the four linear models discussed in this paper. Figures S1 replicates Figure 1 for all four DBP classes studied in this analysis. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Literature Cited

- (1) U.S. Environmental Protection Agency. National primary drinking water regulations; monitoring requirements for public drinking water supplies; Final rule. *Fed. Regist.* **1996**; *61* (94), 24353–24388.

- (2) U.S. Environmental Protection Agency. National primary drinking water regulations: Stage 2 disinfectants and disinfection byproducts rule. *Fed. Regist.* **2006**; 71 (2), 387–493.
- (3) Richardson, S.; Thruston, A. D.; Chaim, R.; Groisman, L.; Popilevsky, I.; Juraev, O.; Glezer, V.; McKague, A. B.; Plewa, M.; Wagner, E. Tribromopyrrole, Brominated acids, and other disinfection byproducts produced by disinfection of drinking waters rich in bromide. *Environ. Sci. Technol.* **2003**, 37 (17), 3782–3793.
- (4) Teuschler, L. K.; Simmons, J. E. Approaching DBP toxicity as a mixtures problem. *J. Am. Water Works Assoc.* **2003**, 95 (6), 131–138.
- (5) Singer, P. C. DBPs in drinking water: Additional scientific and policy considerations for public health protection. *J. Am. Water Works Assoc.* **2006**, 98 (10), 73–80.
- (6) Muellner, M. G.; Wagner, E. D.; McCalla, K.; Richardson, S.; Woo, Y.-T.; Plewa, M. J. Haloacetonitriles vs. regulated haloacetic acids: Are nitrogen-containing DBPs more toxic. *Environ. Sci. Technol.* **2007**, 41 (2), 645–651.
- (7) Hua, G.; Reckhow, D. A. Determination of TOCl, TOBr and TOI in drinking water by pyrolysis and off-line ion chromatography. *Anal. Bioanal. Chem.* **2006**, 384, 495–504.
- (8) Obolensky, A.; Singer, P. C. Halogen substitution patterns among disinfection byproducts in the information collection rule database. *Environ. Sci. Technol.* **2005**, 39 (8), 2719–2730.
- (9) Sadiq, R.; Rodriguez, M. J. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: A review. *Sci. Total Environ.* **2004**, 321 (1–3), 21–46.
- (10) Cowman, G. A.; Singer, P. C. Effect of bromide ion on haloacetic acid speciation resulting from chlorination and chloramination of aquatic humic substances. *Environ. Sci. Technol.* **1996**, 30 (1), 16–24.
- (11) Krasner, S. W.; Scilimenti, M. J.; Chinn, R.; Chowdhury, Z. K.; Owen, D. M., The impact of TOC and bromide on chlorination by-product formation. In *Disinfection By-Products in Water Treatment: The Chemistry of their Formation and Control*; Minear, R. A., Amy, G. L., Eds.; CRC-Lewis Publishers: New York, 1996; pp 59–90.
- (12) Roccaro, P.; Chang, H.-S.; Vagliasindi, F. G. A.; Korshin, G. V. Differential absorbance study of effects of temperature on chlorine consumption and formation of disinfection by-products in chlorinated water. *Water Res.* **2008**, 42, 1879–1888.
- (13) Sohn, J.; Amy, G.; Yoon, Y. Bromide ion incorporation into brominated disinfection byproducts. *Water, Air, Soil Pollut.* **2006**, 174, 265–277.
- (14) Symons, J. M.; Krasner, S. W.; Scilimenti, M. J.; Simms, L. A.; Sorensen, H. W.; Speitel, G. E.; Diehl, A. C., Influence of bromide ion on trihalomethane and haloacetic acid formation. In *Disinfection By-Products in Water Treatment: The Chemistry of their Formation and Control*; Minear, R. A., Amy, G. L., Eds.; CRC-Lewis Publishers: New York, 1996; pp 91–130.
- (15) Obolensky, A.; Singer, P. C. Development and interpretation of disinfection byproduct formation models using the information collection rule database. *Environ. Sci. Technol.* **2008**, 42 (15), 5654–5660.
- (16) Chang, E. E.; Lin, Y. P.; Chiang, P. C. Effects of bromide on the formation of THMs and HAAs. *Chemosphere* **2001**, 43 (8), 1029–1034.
- (17) Mok, K. M.; Wong, H.; Fan, X. J. Modeling bromide effects on the speciation of trihalomethanes formation in chlorinated drinking water. *Global NEST J.* **2005**, 7 (1), 1–16.
- (18) Nokes, C. J.; Fenton, E.; Randall, C. J. Modelling the formation of brominated trihalomethanes in chlorinated drinking waters. *Water Res.* **1999**, 33 (17), 3557–3568.
- (19) Francis, R. A.; Small, M. J.; VanBriesen, J. M. Multivariate distributions of disinfection byproducts in chlorinated drinking water. *Water Res.* **2009**, 43, 3453–3468.
- (20) Bull, R. J.; Rice, G. E.; Teuschler, L. K.; Feder, P. I. Chemical measures of similarity among disinfection byproduct mixtures. *J. Toxicol. Environ. Health, Part A* **2009**, 72 (7), 482–493.
- (21) Rice, G. E.; Teuschler, L. K.; Bull, R. J.; Simmons, J. E.; Feder, P. I. Evaluating the similarity of complex drinking water disinfection byproduct mixtures: Overview of the issues. *J. Toxicol. Environ. Health, Part A* **2009**, 72 (7), 429–436.
- (22) Simmons, J. E.; Richardson, S. D.; Teuschler, L. K.; Miltner, R. J.; Speth, T. F.; Schenk, K. M.; Hunter III, E. S.; Rice, G. E. Research issues underlying the four-lab study: Integrated disinfection by-products mixtures research. *J. Toxicol. Environ. Health, Part A* **2008**, 71, 1125–1132.
- (23) McGuire, M. J.; McLain, J. L.; Obolensky, A. *Information Collection Rule Data Analysis*; AWWA Research Foundation, American Water Works Association: Denver, CO, 2002; p 600.
- (24) Obolensky, A.; Singer, P. C.; Shukairy, H. M. Information collection rule data evaluation and analysis to support impacts on disinfection by-product formation. *ASCE J. Environ. Eng.* **2007**, 133 (1), 53–63.
- (25) Chen, Q.; Ibrahim, J. G.; Chen, M.-H.; Senchaudhuri, P. Theory and inference for regression models with missing responses and covariates. *J. Multivariate Anal.* **2008**, 99, 1302–1331.
- (26) Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B., *Bayesian Data Analysis*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, 2004; p xxv.
- (27) Box, G. E. P.; Tiao, G. C., *Bayesian Inference in Statistical Analysis*; Addison-Wesley Publishing Company: Reading, MA, 1973.
- (28) Geisser, S. The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* **1975**, 70 (350), 320–328.
- (29) Geisser, S.; Eddy, W. F. A predictive approach to model selection. *J. Am. Stat. Assoc.* **1979**, 74 (365), 153–160.
- (30) Xie, Y. F., *Disinfection Byproducts in Drinking Water: Formation, Analysis, and Control*; CRC Press: Washington, DC, 2003.
- (31) Archer, A. D.; Singer, P. C. An evaluation of the relationship between SUVA and NOM coagulation using the ICR database. *J. Am. Water Works Assoc.* **2006**, 98 (7), 110–123.
- (32) Westerhoff, P.; Chao, P.; Mash, H. Reactivity of natural organic matter with aqueous chlorine and bromine. *Water Res.* **2004**, 38, 1502–1513.
- (33) Croue, J.-P.; Debroux, J.-F.; Amy, G. L.; Aiken, G. R.; Leenheer, J. A., Natural organic matter: Structural characteristics and reactive properties. In *Formation and Control of Disinfection By-products in Drinking Water*; Singer, P. C., Ed. AWWA: Denver, 1999.
- (34) Symons, J. M.; Bellar, T. A.; Carswell, J. K.; DeMarco, J.; Kropp, K. L.; Robeck, G. G.; Seeger, D. R.; Slocum, C. J.; Smith, B. L.; Stevens, A. A. National Organics Reconnaissance Survey for halogenated organics. *J. Am. Water Works Assoc.* **1975**, 67 (11), 634–647.
- (35) Feder, P. I.; Ma, Z. J.; Bull, R. J.; Teuschler, L. K.; Rice, G. E. Evaluating sufficient similarity for drinking-water disinfection by-product (DBP) mixtures with bootstrap hypothesis test procedures. *J. Toxicol. Environ. Health, Part A* **2009**, 72 (7), 494–504.

ES9028606