ANALYSIS OF INFORMATION COLLECTION RULE DATA TO ASSESS THE IMPACT OF WATER QUALITY AND TREATMENT ON DISINFECTION BYPRODUCT OCCURRENCE IN DRINKING WATER

Alexa Obolensky

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering, School of Public Health

Chapel Hill
2007

Approved by,

Advisor: Philip C. Singer

Reader: Douglas J. Crawford-Brown

Reader: Charles N. Haas

Reader: Lawrence L. Kupper

Reader: Hiba M. Shukairy

Reader: Howard S. Weinberg

ABSTRACT

Alexa Obolensky: Analysis of Information Collection Rule Data to Assess the Impact of Water Quality and Treatment on Disinfection Byproduct Occurrence in Drinking Water
(Under the direction of Philip C. Singer)

Information Collection Rule (ICR) data were analyzed to investigate relationships between water quality and treatment processes and occurrence of disinfection byproducts (DBPs) in drinking water. A new metric developed to quantify extent of halogen substitution in different byproduct classes indicated strong interclass correlations among bromine fractions in dihaloacetic acids (X_2AAs), trihalomethanes (THMs), trihaloacetic acids (X_3AAs), and dihaloacetonitriles (X_2ANs). These measurements were sensitive to censored data handling. Bromine fraction covariance properties were applied in a test for multivariate outliers to identify data entry or analytical errors which was used in database screening.

Database screening indicated a high level of ICR data quality. Recovery of categorical descriptors substantially amplified the data set. Data patterns showed expected relationships between source water quality and disinfection practices. Plants with high organic precursor concentrations preferentially employed chloramines and avoided prechlorination. Plants with high bromide levels also tended to employ chloramines although bromide occurrence did not impact prechlorination practice. Variability in applied chlorine dose among ICR plants diminished when dose was normalized to total organic carbon (TOC); the median chlorine to TOC ratio was 1.54 mg Cl_2/mg C.

Multiple linear regression models for finished water DBP concentrations at chlorine plants indicated significant shifts across compound classes in the direction and magnitude of

influence for bromide, alkalinity, pH, chlorine consumed, and organic precursor concentrations. Results suggested that alkalinity serves as an indicator of organic matter hydrophobicity and reactivity towards DBP formation. pH effects were in accord with current understanding, though observed large differential impacts across species within THM and X_3AA classes were not previously noted. Model results suggested that chlorine consumed after initial dose is less relevant for X_2AA formation than for other DBPs examined. Based on model projections, use of alternative disinfectants in combination with subsequent chlorination led to substantially lower DBP concentrations compared to use of free chlorine alone. Chloral hydrate, an exception, was enhanced under certain ozone treatment conditions. Model projections indicated that softening treatment led to substantial reductions of brominated THM and X_2AA species, and all X_3AA species, though total organic halogen was unaffected. Softening effects were attributed to improved organic precursor removal.

ACKNOWLEDGEMENTS

I express my deep gratitude to Dr. Philip Singer, my thesis advisor, for his wisdom, support, and endurance through the course of this work. I am very grateful to Dr. Hiba Shukairy for her ongoing encouragement and good humor, her hard work and reliability in support of my research, and for serving on my PhD committee. I also extend many thanks to the other members of my committee for their expertise and commitment: Professors Douglas Crawford Brown, Lawrence Kupper, and Howard Weinberg at UNC, and Professor Charles Haas at Drexel University. I greatly appreciate the valuable guidance provided by Richard Miltner at the Cincinnati EPA. I would not have been able to complete this work without the support of my long-time employer, the Philadelphia Water Department, allowing me a year leave of absence to complete my UNC residency. Howard Neukrug, Mike Pickel, Geoff Brock, and Gary Burlingame all facilitated my work-study efforts. I thank my comrades Gretchen Onstad and Vanessa Speight, who were great company at UNC and have continued to cheer me on from their places beyond the finish line. Finally, my heartfelt thanks go to all my family and friends for their support, most especially my husband Rick, who listened a lot and never complained.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

Aux 1	**primary auxiliary ICR database**
ANOVA	**analysis of variance**
BCAA	**bromochloroacetic acid**
BCAN	**bromochloroacetonitrile**
BDCAA	**bromodichloroacetic acid**
BDCM	**bromodichloromethane**
Br	**bromide**
BrClAA	**bromochloroacetic acid**
$BrCl_2AA$	**bromodichloroacetic acid**
Br_2AA	**dibromoacetic acid**
Br_2ClAA	**dibromochloroacetic acid**
Br_3AA	**tribromoacetic acid**
Br/TOC	**bromide to total organic carbon ratio**
$CHBrCl_2$	**bromodichloromethane**
$CHBr_2Cl$	**dibromochloromethane**
$CHBr_3$	**tribromomethane, or bromoform**
$CHCl_3$	**trichloromethane, or chloroform**
CL2	**free chlorine disinfection category**
Cl_2AA	**dichloroacetic acid**
Cl_2/N	**chlorine to nitrogen ratio**
Cl_2/TOC	**chlorine to total organic carbon ratio**
Cl_3AA	**trichloroacetic acid**
Cl_3AH	**trichloroacetaldehyde hydrate**
CLM	**chloramine disinfection category**
ClO_2^-	**chlorite**
ClO_3^-	**chlorate**
CL2/CLM	**free chlorine with chloramine disinfection category**
ClO2/Cl2	**chlorine dioxide with free chlorine disinfection category**
ClO2/Cl2/ClM	**chlorine dioxide with free chlorine and chloramine disinfection category**
ClO2/ClM	**chlorine dioxide with chloramine disinfection category**
CLR	**clearwell unit process type**
CNCl	**cyanogen chloride**
DBAA	**dibromoacetic acid**
DBAN	**dibromoacetonitrile**
DBCAA	**dibromochloroacetic acid**
DBCM	**dibromochloromethane**
DBP	**disinfection byproduct**
DCAA	**dichloroacetic acid**
DCAN	**dichloroacetonitrile**
DCB	**disinfectant contact basin unit process type**
D/DBP	**disinfectant and disinfection byproduct**

DHAA	**dihaloacetic acid**
DHAA_Br/X	**dihaloacetic acid bromine fraction**
DHAN	**dihaloacetonitrile**
DHAN_Br/X	**dihaloacetonitrile bromine fraction**
DIS/GW	**disinfected ground water treatment plant type**
DIS/WHSALE	**disinfected wholesale treatment plant type**
DOC	**dissolved organic carbon**
DS_DIS	**distribution system disinfectant category**
EPA	**Environmental Protection Agency**
FIL	**filtration unit process type**
FLC	**flocculation unit process type**
GAC	**granular activated carbon unit process type**
GW	**ground water source type**
HAA	**haloacetic acid**
HAA5	**sum of five haloacetic acid species' concentrations**
HAA9	**sum of nine haloacetic acid species' concentrations**
HAN	**haloacetonitrile**
HRT	**hydraulic residence time**
ICR	**Information Collection Rule**
ICR FED	**ICR federal database system**
LT2	**Long-Term 2 Interim Enhanced Surface Water Treatment Rule**
MBAA	**monobromoacetic acid**
MCAA	**monochloroacetic acid**
MCL	**maximum contaminant level**
MHAA	**monohaloacetic acid**
MRL	**minimum reporting limit**
MSRC_CAT	**major source water category**
MWTPTYPE	**major water treatment plant type category**
NOM	**natural organic matter**
OCB	**ozone contact basisn unit process type**
O3/Cl2	**ozone with free chlorine disinfection category**
O3/Cl2/ClM	**ozone with free chlorine and chloramine disinfection category**
O3/ClM	**ozone with chloramine disinfection category**
OTHER/GW	**other groundwater treatment type**
PRE	**presedimentation unit process type**
Pre-Cl2	**prechlorination**
PUR	**purchased finished water source type**
Q	**flow rate**
QA	**quality assurance**
QC	**quality control**
RAP	**rapid mix unit process type**
RCB	**recarbonation basin unit process type**
RPD	**relative percent difference**
SCC	**solids contact clarifier unit process type**

SDS	**simulated distribution system**
SED	**sedimentation basin unit process type**
SDWA	**Safe Drinking Water Act**
SUVA	**specific ultraviolet absorbance**
SW	**surface water source type**
TBAA	**tribromoacetic acid**
TCAA	**trichloroacetic acid**
TCAN	**trichloroacetonitrile**
TCM	**trichloromethane, or chloroform**
TCP	**1,1,1-trichloropropanone**
THAA	**trihaloacetic acid**
THAA_Br/X	**trihaloacetic acid bromine fraction**
THM	**trihalomethane**
THM_Br/X	**trihalomethane bromine fraction**
THM4	**sum of four trihalomethane species' concentrations**
TOC	**total organic carbon**
TOX	**total organic halogen**
TSUVA	**specific ultraviolet absorbance calculated with total organic carbon**
TTHM	**total trihalomethane**
UNFILT/SW	**unfiltered surface water treatment type**
UV	**ultraviolet light irradiation treatment**
UV254	**ultraviolet absorbance at 254 nanometers**
V	**volume**
WTP	**water treatment plant**
WTP_DIS	**water treatment plant disinfectant category**
X_2AA	**dihaloacetic acid**
X_2AN	**dihaloacetonitrile**
X_3AA	**trihaloacetic acid**

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Disinfection byproducts (DBPs) in drinking water are a significant public health concern due to their ubiquity, limited chemical identification, and poorly understood health effects. Halogenated organic DBPs are formed when chlorine used in water treatment, for disinfection or other purposes, reacts with natural organic matter (NOM) and halide ions (bromide and possibly iodide) present in source waters. The phenomenon of DBP formation was first recognized in 1974 with the discovery of trihalomethanes (THMs) in drinking water, long after the practice of water chlorination was firmly established and credited with the elimination of waterborne cholera and typhoid epidemics in the United States (Bellar, Lichtenberg, and Kroner 1974; Rook 1974; Murphy and Craun 1999). Following many years of laboratory and plant-scale research, and different regulatory actions to control DBP formation, the Information Collection Rule (ICR) was mandated to provide a comprehensive survey of treatment and occurrence data related to DBPs and microbial pathogens at all large water utilities in the United States (U.S. EPA 1996a). The ICR generated the largest database of DBP field data ever assembled. Analyses of these data, which became fully available in 2002, are the subject of this dissertation.

1.1.1 DBP Formation

Since the discovery of THMs as chlorination byproducts, hundreds of chemical byproducts of disinfection have been detected in treated drinking water (Richardson 1998, 2003; Krasner et al. 2006). The most prevalent halogenated chlorination byproducts on a weight basis are THMs, followed by haloacetic acids (HAAs), chloral hydrate, haloacetonitriles, haloketones, and chloropicrin (Quimby et al. 1980; Christman et al. 1983; Krasner et al. 1989; Stevens et al. 1989). Taken together, these commonly identified compounds account for between 20 and 60% of organically bound halogen produced during chlorination of natural waters, as characterized by measurement of total organic halogen (TOX) concentration (Krasner et al. 1989; Singer and Chang 1989; Singer et al. 1995; Shukairy et al. 2002). On a weight basis, the THMs and HAAs typically comprise more than 50% of the identified DBPs in chlorinated drinking water (Richardson 1998).

Byproducts of drinking water chlorination are the result of oxidation and substitution reactions between chlorine added for disinfection and NOM present in virtually all source waters. Naturally occurring bromide ion in the source water also plays an important role as an inorganic DBP precursor because of its rapid oxidation to bromine by chlorine, and the subsequent participation of bromine in bromination reactions. Iodide is present in much lower concentrations than bromide in source waters and is oxidized to unreactive iodate by free chlorine (Bischel and von Gunten 1999). However, under certain treatment conditions, available iodide may become incorporated into halogenated organic DBPs since it reacts with monochloramine to yield free iodine, which can participate in iodination reactions with NOM (Hansson 1987; Bischel and von Gunten 1999; Plewa et al. 2004).

The structure and composition of NOM and its reactivity with chlorine have been the subject of extensive research (Christman et al. 1983, 1989; Barrett et al. 2000). NOM is not a

specific chemical compound but consists of a complex mixture of organic molecules ranging widely in molecular weight. It is comprised largely of heterogeneous macromolecular structures produced from condensation of terrestrial vegetative decay products such as lignins, tannins, polysaccharides, and other biogenic compounds (Shapiro 1957; Christman and Oglesby 1971; Leenheer and Croue 2003). These products can be regarded as precursors to fossil fuels and their "age" and the oxygen conditions of their formation environment influence their chemical composition, and thus their reactivity with chlorine. NOM in surface waters may also include algal exudates and cellular algal decay products to varying extents, depending on conditions. Thus, NOM composition and reactivity with chlorine (or bromine) varies by source water as well as across time for a given water.

A variety of types of reactive chemical centers in natural organic matter are likely to act as DBP precursors. Important centers for halogen substitution include electron-rich sites on aromatic rings, such as those activated by an ortho- or para-substituted hydroxy or methoxy substituent (Norwood et al. 1980; de Leer et al. 1985; Norwood et al. 1987; Christman et al. 1989; Reckhow, Singer, and Malcom 1990). Aliphatic structures with a beta-diketone moiety or other enolizable group are also highly labile towards oxidative halogen substitution (Reckhow and Singer 1985; Christman et al. 1989). For DBPs with more than one halogen substituent, the formation process involves sequential halogen substitution reactions. For all DBPs, the final product or intermediates must be separated from the larger parent molecule by oxidative and/or hydrolytic cleavage. Therefore DBP formation reactions are multi-step processes. The amorphous and varied nature of the organic substrate and the complexity of the formation reactions have precluded development of mechanistically-based kinetic models for DBP formation that might provide a deterministic approach to predicting

DBP levels under specific treatment conditions. Nevertheless, researchers studying model compounds considered representative of NOM substructures have suggested plausible mechanisms for the formation of various DBPs and provided rational explanations for observed product mixtures and pH effects (de Leer et al. 1985; Reckhow and Singer 1985).

Researchers have characterized and quantified NOM by a number of methods including analysis of oxidative degradation products, molecular weight distribution, acid- and base-soluble fractionation (e.g., humic acids, fulvic acids, hydrophilic and hydrophobic fractions), nuclear magnetic resonance spectroscopy, and elemental composition (Norwood et al. 1987; Christman et al. 1989; Amy et al. 1992; Krasner et al. 1996a; Croue et al. 1999; Leenheer et al. 2000). For water treatment process control, NOM is most routinely quantified by measurement of Total Organic Carbon (TOC) and Ultraviolet Absorbance at 254 nanometers (UV254). Because of its strong response to conjugated double bond systems that are generally the most prevalent reactive sites for chlorine attack, UV254 is a more specific measure of organic DBP precursor concentration than TOC. Dissolved Organic Carbon (DOC), defined as the portion of TOC passing through a 0.45 micron filter, typically accounts for 90% or more of the TOC in drinking water sources. The ratio UV254/DOC, termed Specific Ultraviolet Absorbance (SUVA), has become an established parameter for comparing the reactivity towards chlorine of NOM from different sources (Edzwald et al. 1985; Najm et al. 1994; Krasner et al. 1996a, 1997).

The basis of DBP formation control strategies is reduction of the concentrations and duration of NOM and chlorine in contact with one another. This can be achieved by replacing chlorine with an alternative disinfectant, moving some or all chlorine addition downstream of TOC removal processes in a treatment train, enhancing TOC removal processes, or some

combination of these approaches. Alternative disinfectants include monochloramine, ozone, chlorine dioxide, and ultraviolet light irradiation (UV). Though ozone and chlorine dioxide are effective in reducing DBPs associated with chlorine, they each generate other DBPs of concern and their implementation is considerably more complex and costly (Culp 1984; Werdehoff and Singer 1987; Myers 1990; Cavanagh et al. 1992; Miltner et al. 1992; Trussell 1992; Glaze et al. 1993). UV is a comparatively new technology for drinking water disinfection that has thus far not been associated with DBP formation (Zheng et al. 1999). It may become widely implemented in the U.S. over the next decade because of its effectiveness for inactivating protozoan organisms resistant to chlorine (e.g., *Cryptosporidium*), as well as viruses and bacteria, without yielding DBPs. However, UV does not provide a residual disinfectant and thus secondary disinfection with a chlorine-based product remains necessary where a residual is required.

Lowering TOC concentration prior to disinfectant addition limits the concentration of organic precursors available for DBP formation reactions and, in doing so, reduces the oxidant demand of the water and hence the disinfectant dose needed to achieve a target residual concentration at the end of subsequent treatment steps (which is the regulatory basis for disinfection credit). TOC removal during water treatment is a central area of research addressing DBP control and other water treatment objectives. Fortunately, the fraction of NOM most labile towards DBP formation reactions is also the fraction most easily removed during conventional coagulation, settling, and filtration (Chadik and Amy 1983; Krasner et al. 1997; Randtke 1999). However, for most treatment plants, it is not possible to simply move chlorine addition downstream of these clarification processes without implementing significant additional changes in treatment. Chlorine application upstream of coagulation

may be employed for iron and manganese oxidation, for algae and/or biofilm control in basins and pipes, and for disinfection credit (where adequate residence time is not available in downstream basins). To control THMs, many conventional drinking water treatment plants have substantially lowered the concentration of chlorine applied to raw water or, where feasible, moved their primary chlorine addition point downstream of the clarification basins (Singer 1994; AWWA 2000).

As mentioned, monochloramine is one alternative to chlorine for disinfection. Chlorine and ammonia react together to form monochloramine (also called "combined" chlorine, in contrast to "free" chlorine). THM formation is largely arrested when ammonia is added to chlorinated water to produce a combined chlorine residual (Symons et al. 1981; Speitel 1999). However, monochloramine is a much weaker biocide than free chlorine and thus its use for primary disinfection in treatment plants is usually limited by the long contact times or high disinfectant doses needed to obtain required disinfection credits (Haas 1999). Chloramine application is more difficult to control because of the complex pH-dependent chemistry involved in the equilibria between chlorine, ammonia, and the three chloramine species: mono-, di-, and trichloramine (Morris 1967; Wajon and Morris 1980; Jafvert and Valentine 1992; Speitel 1999). However, because monochloramine is less reactive than free chlorine, it can provide a more persistent residual in water moving through a distribution system. For this reason, and for the benefit of controlling THMs, monochloramine has come into widespread use for secondary disinfection in finished water leaving the treatment plant (Duke et al. 1980; Singer et al. 1982; Speitel 1999; Seidel et al. 2005). Unfortunately recent research indicates that carcinogenic nitrosamines may represent a previously unrecognized class of chloramination byproducts, raising a new specter of concern about the use of this

alternate disinfectant (Najm and Trussell 2001; Choi and Valentine 2002; Mitch and Sedlak 2002; Gerecke and Sedlak 2003; Valentine et al. 2005).

THMs have been studied extensively since their discovery as drinking water DBPs (Rook 1974; Bellar, Lichtenberg, and Kroner 1974). THMs are terminal byproducts which tend to be stable in drinking water distribution systems and continue to form as long as organic substrates are present and a free chlorine residual persists. HAAs and other halogenated DBPs have not been studied as extensively as THMs and their formation kinetics and stability are less well characterized. HAAs tend to form faster than THMs (Reckhow and Singer 1984), some of the HAA species are known to decompose at elevated pH levels (Reckhow and Singer 1985; Krasner et al. 1989), and some of the species are subject to biodegradation in the absence of a chlorine residual (Williams et al. 1994; Baribeau et al. 2006). Recent research has shown that trihalogenated haloacetaldehyde and HAA species decompose to yield corresponding THM compounds as products (Xie and Reckhow 1996; Zhang and Minear 2002). Decomposition rates for these compounds increase rapidly with temperature and with the number of bromine substituents on the individual species. After THMs, HAAs and haloacetaldehydes are the most prevalent chlorination byproducts (Blank et al. 2002; Richardson 2003; Krasner et al. 2006). Other DBPs, such as di- and trichloropropanone and di- and trichloroacetonitrile, generally present in lower concentrations, are also known to decompose and the trihalogenated classes may yield THMs under certain conditions (Trehy and Bieber 1981; Bieber and Trehy 1983; Croue and Reckhow 1989). Thus, especially in warm temperatures or long distribution system transit and/or storage times, these higher molecular weight metastable DBP species may convert to

THMs. This is an important consideration in studying distributions of DBP species both within and between different compound classes.

Interest in DBP halogen substitution patterns has increased due to research indicating that brominated DBP compounds may have more adverse health effects than their fully chlorinated counterparts (Zavaleta et al. 1999). Controlled laboratory studies have shown that bromine incorporation into HAAs parallels that of bromine incorporation into THMs (Pourmaghaddas et al. 1993; Cowman and Singer 1996). Though this has not been thoroughly demonstrated using field data, a recent study by Roberts et al. (2002) using a subset of ICR data, suggests that bromine incorporation into trihaloacetic acids parallels that of bromine incorporation into THMs. Similar parallels in bromine substitution patterns between the dihaloacetic acids and dihaloacetonitriles have been observed in field data collected at the Philadelphia Water Department (Obolensky 1998). In controlled laboratory experiments, several researchers have examined the effects of water quality and treatment factors on the extent of bromine substitution in THMs and HAAs (Minear and Bird 1980; Oliver 1980; Amy et al. 1991; Summers et al. 1993; Shukairy, Miltner, and Summers 1994; Krasner et al. 1996b; Shukairy and Summers 1996; Symons et al. 1996). These studies have established that the degree of bromine substitution in DBPs is driven by source water bromide concentration, the ratio between bromide and organic precursor concentrations (which can be altered through treatment), and the ratio between bromide and chlorine dose, reaction time, and temperature.

In summary, factors that need consideration in studying the formation, occurrence, and relative distribution of DBP compounds in drinking water include: water temperature; pH of chlorination and distribution; source water bromide concentration; TOC concentration

at the point of chlorine contact; chemical characteristics of NOM in the water (e.g., SUVA, hydrophilic vs. hydrophobic DOC distribution); type of disinfection scenario (chlorination only, chloramination only, chlorination/chloramination, or ozonation/chlor(am)ination); points of disinfectant addition relative to precursor removal and other treatment steps; disinfectant dose, contact time, and residual concentration; and conditions affecting biodegradation and/or hydrolytic loss of DBPs in the distribution system. Interactions between these factors may be expected. For example, the importance of biodegradation may depend on pH, temperature, and chlorine residual. Similarly, hydrolytic loss pathways can be pH dependent. Many of the reactions are under kinetic control and thus are strongly impacted by temperature.

1.1.2 Regulatory History and the ICR

Since the discovery of DBPs in chlorinated drinking water in 1974, water treatment practices have adjusted in response to evolving Safe Drinking Water Act (SDWA) regulations and their ammendments addressing DBPs (Singer 1994). Soon after the identification of chloroform and other THM species as chlorination DBPs in drinking water, the U.S. Environmental Protection Agency (EPA) set an interim Maximum Contaminant Level (MCL) for total THMs (TTHM, the sum of four chlorinated and brominated THMs) (U.S. EPA 1979). Addressing the potential chronic risk from bladder cancer, this rule limited the annual average of quarterly system-wide TTHM concentrations to 100 μg/L for community water systems that added a disinfectant to their treatment process and served 10,000 or more persons. Following this relatively quick action, the complexity of the DBP issue became increasingly apparent, and it took almost twenty years for the next step in DBP regulations to be enacted.

During the 1980's and 1990's many new (non-THM) DBPs were discovered, including those produced from alternative disinfectants, the effects of treatment processes on DBP formation were becoming better-understood, and health effects research was progressing (Singer 1994; Symons 1999). However, large uncertainties surrounding DBP health risks contrasted with the comparatively well understood acute risks from waterborne microbial pathogens (Murphy and Craun 1999). Given the intrinsic conflict between water treatment goals for controlling DBP formation and inactivating microbial pathogens, and the consequent potential for adverse unintended consequences from treatment modifications implemented to address one or the other issue in isolation, a carefully balanced regulatory approach was needed to address these risks in concert (Craun, Hauchman, and Robinson 2001). This need prompted linkage of SDWA microbial and DBP rule development tracks into a two-staged cluster of paired regulations (U.S. EPA 1994a). In support of this phased rulemaking, the ICR was designed to obtain field data needed to characterize plant-specific tradeoffs associated with simultaneous control of DBP and microbial risks (U.S. EPA 1994b, 1996a). ICR monitoring was completed in December 1998 and the data started to become available the following year.

The Stage 1 Disinfectants and Disinfection Byproducts Rule (Stage 1 D/DBP Rule), promulgated in 1998, expanded applicability of DBP provisions to small systems (i.e., systems serving < 10,000 persons), lowered the TTHM MCL, established new MCLs for the sum of five HAAs and for specific byproducts of ozone (bromate) and chlorine dioxide (chlorite), set limits on residual disinfectant concentrations, and mandated NOM (TOC) removal as a treatment technique for DBP control (U.S. EPA 1998a). The Interim Enhanced Surface Water Treatment Rule was issued concurrently with the Stage 1 D/DBP Rule to

address microbial pathogens, in particular epidemic risk from the protozoan parasite *Cryptospridium* (U.S. EPA 1998b).

The Stage 2 D/DBP Rule, promulgated in 2006, further expands the applicability of DBP provisions to all systems that deliver disinfected water, encompassing consecutive systems that distribute purchased treated water without adding additional disinfectant (U.S. EPA 2000a, 2003, 2006a; Scharfenaker 2001). This new rule does not establish MCLs for additional DBPs or change the numerical values of existing MCLs, but substantially modifies DBP monitoring programs, compliance calculations, and reporting requirements in a manner that will further drive down DBP occurrence levels for many utilities. Under previous rules, quarterly results from locations with high DBP levels could be offset by corresponding results from other locations with low concentrations in the same system, because only the quarterly mean across all monitoring sites was considered in the reported annual average. Under the new provisions, annual average THM and HAA MCLs must be met for each individual monitoring location in each public water system. Moreover, under the Initial Distribution System Evaluation provision, siting of these monitoring locations must be newly optimized to ensure that monitoring captures high occurrence levels. Additionally, the Stage 2 Rule requires reporting and follow-up operational evaluation if any individual DBP monitoring result (i.e., not the annual average) exceeds a particular threshold value. This latter provision was a compromise to address concern over possible acute reproductive/developmental health risks, that is, risks stemming from short-duration exposure. These risks were the subject of extensive discussion during the negotiated rulemaking for Stage 2 (U.S. EPA 2006a). Although data supporting acute health effects from DBPs were not considered adequate to warrant changing the compliance basis of

enforceable MCLs from annual average values to individual sample values, it was agreed that drawing attention to elevated discrete DBP results would heighten awareness, leading to better understanding and control of distribution system operations, and helping to avoid future MCL violations. In practice, water utilities will most likely work hard to avoid these exceedances, both to escape reporting requirements and to avert negative public perceptions. The Long Term 2 Enhanced Surface Water Treatment Rule (LT2), addressing the issue of endemic risk from *Cryptosporidium*, was promulgated in concert with the Stage 2 D/DBP Rule (U.S. EPA 2006b).

ICR data were used extensively to support development of the companion Stage 2 and LT2 Rules. However, analyses conducted in that context focused on national data distributions, development of treatment and occurrence baselines, and model projections for regulatory impact forecasting. One of the motivations for the research behind this dissertation was the rich source of untapped information available from the ICR survey.

Over the past three decades, laboratory studies conducted under controlled conditions have provided important insights into factors affecting DBP formation (Rook 1978; Babcock and Singer 1979; Minear and Bird 1980; Oliver 1980; Christman et al. 1983; Reckhow and Singer 1984; Stevens et al. 1989; Cowman and Singer 1996; Shukairy and Summers 1996; Symons et al. 1996). In order to establish effects, many laboratory studies have employed conditions outside the range of general applicability for water treatment (e.g., extreme chlorine doses or bromide levels, buffered synthetic waters). Various findings from these research efforts have been borne out in limited field monitoring studies (Arguello et al. 1979; Brett and Calverly 1979; Veenstra and Schnoor 1980; Otson, Williams, and Bothwell 1981; Nieminski, Chaudhuri, and Lamoreaux 1993; Singer, Obolensky, and Greiner 1995; Arora,

LeChevallier, and Dixon 1997), and isolated field observations have often been corroborated by laboratory investigations (Williams, Rindfleisch, and Williams 1994; Williams, Williams, and Gordon 1996). However, previous field monitoring studies have been too limited in scope and scale to permit extensive use of statistical methods for drawing inferences from observational data sets. For a complex dynamic system such as formation of DBPs during water treatment, this type of analysis requires simultaneous control for numerous variables. It is also important that data be structured, collected in a consistent format, and acquired with comparable analytical methodologies and quality assurance. ICR data provide the scale and consistency needed to perform valid statistical analyses of observational data. The data cover all seasons and represent water quality, treatment, and distribution system characteristics across a nationally representative set of source waters, treatment plants, and distribution conditions.

1.2 RESEARCH OBJECTIVES

The overall objective of this research was to use standard statistical methodologies and the power of the large ICR dataset to isolate the effects of water quality and treatment on observed DBP concentrations, including individual compound concentrations, class and subclass total concentrations, and relative concentrations and halogen speciation patterns within and between compound classes. The research also encompassed review, screening, and descriptive summary of relevant information in the database.

Specific goals of the research included application of multivariate analyses to characterize intra- and interclass bromine substitution patterns in four classes of DBPs monitored under the ICR: THMs, dihaloacetic acids (DHAAs or X_2AAs), trihaloacetic acids (THAAs or X_3AAs), and dihaloacetonitriles (DHANs). As a component of the database

screening effort, findings from the analysis of comparative speciation patterns were applied to identify anomalous ICR results for individual analytes and thereby improve the database for subsequent analyses. Insights about speciation patterns also supported formulation of models to estimate brominated THAA species concentrations that were not reported, so that the option of using this expanded database would be available for subsequent analyses. A second main goal of the research was to apply multiple linear regression models as tools to elucidate the influences of water quality and treatment factors on DBP formation and occurrence.

Findings from this research are expected to improve understanding of DBP formation and occurrence and the effects of water quality characteristics and treatment practices under true field conditions. This information should allow utilities to better evaluate control strategies that limit overall DBP formation as well as the formation of selected DBPs, and thereby assist them in complying with regulations and reducing consumer exposure to DBPs. Because ICR source waters, treatment plants, and distribution systems capture the spectrum of conditions across the United States very well, this information should be relevant and useful for a wide audience of utilities, consultants, and regulatory agencies. This research was also expected to demonstrate how better information could be obtained from monitoring data to assist in future decision-making for utilities, regulators, and public health professionals.

1.3 ORGANIZATION OF THESIS

This thesis contains four papers that are in different stages of preparation or publication. These papers are presented as Chapters 2 through 5. Chapter 2, published in *Environmental Science and Technology* in 2005, is entitled “Halogen substitution patterns

among disinfection byproducts in the Information Collection Rule database," and deals with the correlation of halogen substitution patterns in different DBP compound classes and the impact of censored data handling. Employing a new metric developed in this research, the bromine fractions of halogen in four DBP classes were treated as a multivariate response variable and a quality assurance application of a test for multivariate outliers was demonstrated. Chapter 3, published in the *Journal of Environmental Engineering* in 2007, is entitled "Information Collection Rule data evaluation and analysis to support impacts on disinfection by-product formation," and describes the database screening and review effort undertaken for this research and their results, and provides summaries and general analysis of key ICR water quality and treatment data relevant to DBP formation. Included is a demonstration of the dependence between water quality conditions and disinfection practices. Chapter 4, entitled "Development and interpretation of models to describe the impacts of water quality and treatment on DBP formation using the ICR database," describes the development and interpretation of multiple linear regression models to describe relationships between water quality and treatment factors and finished water DBP concentrations. The emphasis is on use of these models as analytical rather than predictive tools. Many of the findings were consistent with current understanding of factors affecting DBP formation, such as pH and bromide effects. Results suggested that alkalinity, not previously included in DBP regression models, was an indicator of NOM characteristics such that higher alkalinity waters preferentially yielded higher concentrations of certain DBP species. Chapter 5, entitled "Applications of models to describe the impacts of water quality and treatment on DBP formation using the ICR database," demonstrates applications of the models to evaluate the impact of various treatment factors on finished water DBP concentrations. Specifically, point

of chlorine addition (at raw or settled/filtered water), use of alternative disinfectants (monochloramine, ozone, or chlorine dioxide), and softening were examined. Chapters 4 and 5 are draft manuscripts planned for submission to *Environmental Science and Technology* as companion papers.

The overall findings of the research are summarized in Chapter 6, along with recommendations for further applications of the tools developed, and some discussion of the advantages and limitations of the regression model as a tool for studying observational field data. Appendices A and B provide supplemental information for Chapters 2 and 4, respectively.

CHAPTER 2: HALOGEN SUBSTITUTION PATTERNS AMONG DISINFECTION BYPRODUCTS IN THE INFORMATION COLLECTION RULE DATABASE

2.1 INTRODUCTION

Because of limited understanding of health effects for individual disinfection byproduct (DBP) compounds, analytical obstacles, and lack of water treatment and occurrence data for most non-trihalomethane (THM) species, U.S. Environmental Protection Agency (EPA) regulations include drinking water maximum contaminant levels only for sums of trihalomethane and haloacetic acid (HAA) species concentrations (U.S. EPA 1998, 2003). Including chlorinated and brominated homologues, THMs comprise four species while HAAs comprise nine species, which include two monohaloacetic acids (MHAAs), three dihaloacetic acids (DHAAs), and four trihaloacetic acids (THAAs). THMs are regulated as "total THM", the sum of all four brominated and chlorinated species' concentrations, and HAAs are regulated as HAA5, the sum of five of the nine brominated and chlorinated HAA species' concentrations. HAA5 includes two MHAA species, two DHAA species, and one THAA species, excluding three brominated THAA species and one brominated DHAA species. Available data suggest that brominated DBP compounds may have more adverse health effects than their fully chlorinated counterparts (Bull et al. 2001;

Richardson 2003). Recent research has also drawn attention to the occurrence and potential health effects of iodinated byproducts (Weinberg et al. 2002; Plewa et al. 2004). Thus, insight into halogen substitution patterns in drinking water DBP data is important for improving understanding of DBP formation, occurrence, and exposure in drinking water as part of the overall goal of public health protection through optimized treatment and appropriate regulation.

Many researchers have demonstrated how water quality and treatment factors influence levels of bromine substitution in THM and HAA byproduct classes (Minear and Bird 1980; Oliver 1980; Gould, Fitchhorn, and Urheim 1983; Krasner et al. 1989, 1990, 1996, 2002; Poumoghaddas et al. 1993; Symons et al. 1993, 1996; Shukairy et al. 1994; Cowman and Singer 1996; McLain, Obolensky, and Shukairy 2002). Concentrations of brominated THM and HAA species have been shown to increase with source water bromide concentration, the ratio between bromide and total organic carbon (TOC) concentrations at the point of chlorination, and the ratio between bromide and chlorine. Shifts from chlorinated to brominated species concentrations with increasing source water bromide levels have also been observed in non-THM/HAA byproduct classes such as haloacetonitriles (Krasner et al. 1990, 1996, 2002), cyanogen halides (Krasner et al. 1990), halonitromethanes (Krasner et al. 2001, 2002), haloaldehydes (Krasner et al. 2002), haloketones (Krasner et al. 2001), and halofuranones (Suzuki and Nakanishi 1995; Onstad 2003). Fewer studies have attempted to characterize or quantify interrelationships among halogen substitution patterns across different DBP classes (Krasner et al. 1989, 1990; Suzuki and Nakanishi 1995; Roberts et al. 2002; Onstad 2003). This paper reports on characterization of interrelationships among halogen substitution patterns within and across DBP compound classes and demonstrates

how these interdependencies can be utilized for analytical quality assurance purposes. Further work will apply the findings to DBP occurrence modeling and will examine the influence of water quality and treatment factors on inter- and intra-class bromine substitution patterns.

The Information Collection Rule (ICR) provides the largest database of drinking water DBP monitoring data collected to date and reflects the wide variation in water quality and treatment conditions across large U.S. utilities (those serving 100,000 or more persons) for the 1997-1998 monitoring period (U.S. EPA 1996a; McGuire, McLain, and Obolensky 2002). These data include species concentrations representing 10 classes or subclasses of chlorination byproducts and, considering brominated and chlorinated compounds, include complete homologue series concentrations for 5 of these groups: MHAAs, DHAAs, THAAs, THMs, and dihaloacetonitriles (DHANs). Generally speaking the mono-, di-, and tri-haloacetic acid groups (i.e. MHAAs, DHAAs, and THAAs) are subclasses of the HAA class. However, their distinct behavior with respect to factors influencing formation and decay in water treatment and distribution systems necessitate their separate consideration in most analyses, including the present one. To make the current discussion less cumbersome, the HAA subclasses will be referred to as distinct classes henceforth. ICR DBP data were collected quarterly. Corresponding source water quality data (collected monthly), including bromide and TOC concentrations, are available for each DBP sampling period for each treatment plant. The rigor and consistency of quality assurance standards and analytical methodologies employed during ICR data collection, together with the large size of the database, render these data especially suitable for statistical analysis of speciation patterns within and across DBP compound classes. Wysock et al. (2002) provide a summary ICR

implementation and data collection procedures and Fair et al. (2002) describe analytical quality assurance programs and precision and accuracy results.

2.2 METHODS

This work utilized data collected quarterly under the ICR from July 1997 through December 1998 by 297 large U.S. drinking water utilities comprising 500 water treatment plants and their distribution systems (U.S. EPA 1996a; McGuire, McLain, and Obolensky 2002). Data were obtained from the U.S. EPA Auxiliary 1 Database Version 5.0 (Aux 1), a Microsoft Access relational database designed for public access (U.S. EPA 2000b). Records retained for this analysis were distribution system and simulated distribution system (SDS) samples from all non-blending ICR treatment plants that used only free chlorine for primary or secondary disinfection (no ozone, chlorine dioxide, or chloramine) and added no additional source water in the treatment train beyond the plant influent. In the ICR, "non-blending" plants are those for which distribution system samples could be confidently ascribed to the finished water from one particular treatment plant. ICR SDS samples were treatment plant finished water samples incubated at ambient water conditions for a time equivalent to the estimated distribution detention time of a specific ICR sample location. Obolensky and Frey (2002) discuss DBP results for ICR distribution system and SDS samples. The analysis centers on reported analytical results for THM, DHAA, THAA, and DHAN species as listed in 2.1 (because of consistently low concentrations MHAA results were not included). The resulting data set contained records for a total of 6565 sample observations with non-missing values for all species within at least one of the four DBP classes considered. Table 2.1 shows the number of records available for each class individually. Out of the total number of records, 4641 had complete data for both THM and

DHAA classes; 2961 had complete data for THM, DHAA, and DHAN classes; and 948 had complete data for all four classes. The smaller number of records having complete THAA data is because ICR monitoring of three of the brominated THAA species was optional. The data utilized in this analysis represent 283 of the 500 ICR treatment plants overall, including 68 plants with complete data for the THAA class. The overall median source water bromide concentration for applicable plant/sampling periods was 22 μg/L. Approximately 25% of the samples were associated with source water bromide exceeding 50 μg/L while 10% of the samples were associated with source water bromide exceeding 100 μg/L. This compares to median, 75th, and 90th percentiles of 35 μg/L, 83 μg/L, and 160 μg/L, respectively, for the ICR dataset overall. For each sampling period at each treatment plant, the Br/TOC ratio was computed using influent water bromide concentration and TOC concentration at the first point of chlorine addition in the treatment train. Boundaries of Br/TOC quartiles were determined based on the entire ICR data set and were used to group data for trend analysis.

Table 2.1 DBP species included in analysis, associated ICR minimum reporting levels (MRLs), total numbers of sample records, and numbers of records with below MRL results

DBP Class	DBP Species	Abbreviation	MRL[a] (μg/L)	MRL[a] (μmol/L) x 10^3	N[b]	N below MRL	% below MRL
THM	chloroform	$CHCl_3$	1.0	8.4	5858	505	9%
	bromodichloromethane	$CHBrCl_2$	1.0	6.1	5858	511	9%
	dibromochloromethane	$CHBr_2Cl$	1.0	4.8	5858	1669	28%
	bromoform	$CHBr_3$	1.0	4.0	5858	4301	73%
THAA	trichloroacetic acid	Cl_3AA	1.0	6.1	1080	94	9%
	bromodichloroacetic acid	$BrCl_2AA$	1.0	4.8	1080	88	8%
	dibromochloroacetic acid	Br_2ClAA	2.0	7.9	1080	789	73%
	tribromoacetic acid	Br_3AA	4.0	13.5	1080	1056	98%
DHAA	dichloroacetic acid	Cl_2AA	1.0	7.8	5268	261	5%
	bromochloroacetic acid	BrClAA	1.0	5.8	5268	1093	21%
	dibromoacetic acid	Br_2AA	1.0	4.6	5268	3529	67%
DHAN	dichloroacetonitrile	Cl_2AN	0.5	4.6	4880	789	16%
	bromochloroacetonitrile	BrClAN	0.5	3.2	4,880	1,438	29%
	dibromoacetonitrile	Br_2AN	0.5	2.5	4,880	2,355	48%

[a] Minimum reporting level, MRLs were uniform across ICR laboratories.
[b] Count of records with non-missing values for all species in class with at least one species > MRL.

Queries were written in Aux 1 to tabulate data for export to SAS (SAS Institute, Cary NC). All data processing and analysis was conducted in the SAS software environment. The handling of data with below minimum reporting level (MRL) results was an important consideration because of the high prevalence of these censored results in the dataset. In this case the data were left-censored, meaning the numerical values were unknown except for their being below the fixed reporting limit cutoff. Table 2.1 shows that within each DBP class the fraction of censored records increased sharply with the number of bromine substituents on a species. Moreover, MRLs were not uniform within or between classes. MRLs for DHAN species, 0.5 μg/L, were lower than those for most other DBPs (1.0 μg/L) while MRLs for brominated THAA species increased with the number of bromine substituents. In the most extreme case, the very high tribromoacetic acid (Br_3AA) MRL of

4.0 μg/L compounded with the generally lower occurrence concentration levels for brominated compounds resulted in 98% of Br_3AA analytical results being censored. To date, most analyses and summaries of ICR DBP data have employed uniform replacement values of zero for below MRL results because interest was generally focused on upper ranges of occurrence distributions (McGuire, Mclain, and Obolensky 2002). For example, tabulated class-sum DBP concentrations in the Aux 1 database (e.g., TTHM and HAA5) were computed using zero in place of any below MRL component species result. However, because MRLs differed within and between classes and affected large fractions of data, this uniform zero replacement method could result in under-representation of species with higher MRL values and thereby introduce differential bias into halogen composition estimates that are of central interest for this analysis.

To assess the impact of censored data handling for this research, results were compared using three approaches which are termed here "Zero", Half-MRL", and Level". The Zero method has been described above, that is all left-censored analytical results were assigned a replacement concentration value of zero. With the exception of Br_2ClAA and Br_3AA, the Half-MRL method used replacement values equal to half the compound-specific MRL (see Table 2.1). Because of the comparatively high MRLs for Br_3AA (4.0 μg/L) and Br_2ClAA (2.0 μg/L) and the large percentage of samples with below MRL entries, censored data replacement values for these two analytes were not uniformly set to half the MRL but were assigned on the basis of reported $BrCl_2AA$ concentrations for the same sample as follows. Censored Br_2ClAA values were set to 0.5 μg/L if $BrCl_2AA$ for the same sample was below MRL. Otherwise, censored Br_2ClAA values were set to 1.0 μg/L (half the MRL). Censored Br_3AA values were set to 0.5 μg/L if both $BrCl_2AA$ and Br_2ClAA for the same

sample were below MRL. Otherwise, censored Br_3AA values were set to 1.0 μg/L unless Br_2ClAA for the same sample was ≥ 2.0 μg/L in which case censored Br_3AA values were set to 2.0 μg/L (half the MRL).

The third censored data handling approach, the Level method, was designed to equalize censoring levels between analogous THAA and THM species and between analogous DHAA and DHAN species. Accordingly, censoring cutoffs for $CHBr_2Cl$ and $CHBr_3$ were increased to 2.0 and 4.0 μg/L, respectively to match the MRL-censoring levels of corresponding THAA species (i.e., Br_2ClAA and Br_3AA). Similarly, censoring cutoffs for the three DHAN species were increased from 0.5 μg/L to 1.0 μg/L to match MRL-censoring levels for the corresponding DHAA species. Replacement concentration values of zero were then applied to all censored data.

The extent of bromine substitution in a DBP class, for an individual sample, was characterized by bromine fraction of total molar halogen in the class, as described by Equation 2.1. For convenience, this measure is termed bromine fraction. To obtain the bromine fraction value for any class, molar bromine concentration and total molar halogen concentration were computed for each compound in the class. After summing molar bromine and total molar halogen concentrations across all species in the class, bromine fraction was determined as the ratio between class-sum bromine and class-sum total halogen. Table 2.2 provides illustrative calculations of bromine fraction for THAA and DHAA sample data. Similar calculations apply for THM and DHAN classes, respectively.

$$\left(\mathrm{Br}/\mathrm{X}\right)_{\mathrm{Class}} = \frac{\sum_{\mathrm{Species}} \left(\text{molar conc.}\right) \times \left(\#\,\text{bromine}\right)}{\sum_{\mathrm{Species}} \left(\text{molar conc.}\right) \times \left(\#\,\text{halogen}\right)} \tag{2.1}$$

Table 2.2 Outline of THAA and DHAA bromine fraction calculations for individual sample concentration data

			THAA		
	Cl_3AA	**$BrCl_2AA$**	**Br_2ClAA**	**Br_3AA**	
Conc. (μg/L)	**6.5**	**12.0**	**5.8**	**4.0**	
Mol. Wt. (g/mol)	163.4	207.8	252.4	296.7	
Conc. (μmol/L)	0.0398	0.0577	0.0230	0.0135	
# Br	0	1	2	3	
# Halogen (X)	3	3	3	3	
μmol-Br/L	0.0000	0.0577	0.0460	0.0404	$\Sigma = 0.1442$
μmol-X/L	0.1193	0.1732	0.0689	0.0404	$\Sigma = 0.4020$
Br/X (mol/mol)					**0.359 (36 %)**
			DHAA		
	Cl_2AA	**BrClAA**	**Br_2AA**		
Conc. (μg/L)	**7.2**	**14.0**	**5.0**		
Mol. Wt. (g/mol)	128.9	173.35	217.8		
Conc. (μmol/L)	0.0559	0.0808	0.0230		
# Br	0	1	2		
# Halogen (X)	2	2	2		
μmol-Br/L	0.0000	0.0808	0.0459		$\Sigma = 0.1267$
μmol-X/L	0.1117	0.1615	0.0459		$\Sigma = 0.3192$
Br/X (mol/mol)					**0.397 (40 %)**

In an earlier approach to quantifying the extent of bromine substitution in THMs, Gould and coworkers (1983) developed the bromine incorporation factor, η, which ranges from 1, for 100% $CHCl_3$, to 3, for 100% $CHBr_3$. This metric, describing moles bromine per mole THM, has been used by many researchers studying the influence of water quality and treatment conditions on bromine incorporation in THMs (Shukairy, Miltner, and Summers 1994; Krasner et al. 1996b; Symons et al. 1996; McLain, Obolensky, and Shukairy 2002).

Some studies have also adapted η to measure bromine incorporation in other byproduct classes or subclasses (Shukairy, Miltner, and Summers 1994). However the range of η depends on the number of halogens in the compound class or subclass, making comparisons across classes difficult. Straightforward interclass comparisons of extents of bromine substitution are possible using the bromine fraction metric presented here because it is normalized in the same manner for all byproduct classes regardless of the number of halogen substituents in the class, and all values range from zero to one.

For each ICR DBP sample, bromine fraction was determined for the THM, DHAA, THAA, and DHAN classes as data permitted (i.e., non-missing data were required for all species in the class) after making appropriate replacements for censored concentration values. Because bromine fraction is undefined when the class-sum total halogen concentration (denominator of bromine fraction) is zero, such data were counted as null observations for the class in question and could not be included in subsequent analyses. This only impacted the data set compiled using the Zero method for censored data handling and involved relatively few sample records.

Nonparametric correlation was used in the analysis of bromine fraction data. Because of their strongly skewed distributions, least-squares regression and Pearson correlation (standard R-squared), which assume normally distributed data, are not valid procedures for these variables. Spearman's Rank correlation was used to characterize pairwise associations between class bromine fractions. This statistic, equivalent to a Pearson correlation computed from ranks of the observation values rather than from the values themselves (differences may stem from ties in data ranks, that is, repeated values), indicates the presence and strength of a

monotonic association between two variables, as opposed to a linear association indicated by the Pearson correlation (Townend 2002).

For each sample, the relative extent of bromine substitution in two DBP classes was characterized by the ratio between the two corresponding bromine fraction values. Thus, a class pair ratio of 1 would indicate equivalent extents of bromine substitution in the two classes, a ratio above 1 would indicate a greater extent of bromine substitution in the class represented by the ratio numerator, and a ratio below 1 would indicate a greater extent of bromine substitution in the class represented by the ratio denominator. Bromine fraction ratios were determined for each of the six possible class pairs (considering the four DBP classes examined). Since ratios are undefined in the case of a zero-valued denominator and the choice of denominator for any particular class pair ratio was arbitrary, sample observations for a class pair were excluded from this particular analysis if either bromine fraction value was zero.

The four or fewer measures of bromine incorporation on an individual water sample (i.e., bromine fractions for THM, DHAA, DHAN, and THAA classes) were treated statistically as a multivariate response. A test for multivariate outliers developed by Caroni and Prescott (1992) was performed to identify outliers on the basis of their effects on the covariance structure of the data. This test was applied according to the approach demonstrated by Pennell (2002) in an application to environmental data. The method employs sequential application of Wilk's test for a single outlier using a conservative Bonferroni approach to controlling the Type I error rate for a pre-specified maximum number of outliers while controlling for swamping and masking effects. Swamping may occur when observations are tested as a group and some observations are erroneously identified as

outliers due to the extremeness of other points in the group. Masking describes the failure to identify single outliers due to the presence of less extreme outliers still present in the data set. The method assumes that a *p*-dimensional variate follows a multivariate normal distribution with homogeneous variance (homoscedasticity). In this case p=4 or less. Because the bromine fraction measures exhibit strong departures from normality and homoscedasticity, an arcsine transformation was applied to each of the four measures to improve adherence to distributional assumptions of the statistical test. The arcsine transformation, commonly employed for proportion data, involves converting the original proportion variable, Y_i, to Y_i' according to Equation 2.2.

$$Y_i' = 2 \times m_i^{0.5} \times \sin^{-1}\left(Y_i^{0.5}\right) \tag{2.2}$$

In Equation 2.2, m_i is the sample size, or denominator of the proportion, and Y_i' is given in radians (Draper and Smith 1996). Since the proportion represented by Y_i for this application is molar bromine divided by total molar halogen, m_i is the total molar halogen for the sample observation, for the class in question (see Equation 2.1).

Four-way, three-way, two-way, and one-way outlier tests were performed on subsets of observations having bromine fraction values correspondingly for four, three, two, or one of the four compound classes. The four-way test was performed on the subset of observations with data for all four classes. The three-way test was performed on the four subsets of observations having data for THM, DHAA, and DHAN; THM, DHAA, and THAA; THM, THAA, and DHAN; and DHAA, THAA, and DHAN. The two-way test was performed on the six subsets of observations having data for THM and DHAA; THM and DHAN; THM

and THAA; DHAA and DHAN; THAA and DHAN; and DHAA and THAA. Finally, the one-way test was performed on the four subsets of observations having data for THM, DHAA, DHAN, and THAA, respectively. Using this approach, a single observation could be included in several different outlier tests. The analysis was conducted this way so that, when evaluating any particular combination of classes for the presence of outliers, the largest possible data set would be utilized. One-way tests were included to help discern which outliers may have been identified on the basis only of extremeness in the univariate distribution rather than on discordance with respect to interclass bromine substitution patterns. Test results were merged by sample observation identification number and observations identified as outliers in any test were tagged, retaining information about which tests they were identified in. Bromine fraction data were plotted and outlier results were reviewed by graphic inspection and by examination of underlying DBP species concentration data. SAS programming code for conducting the multivariate outlier tests is provided in Appendix A (Obolensky and Singer 2005, Supplemental Information).

2.3 RESULTS

2.3.1 Extents of Bromine Substitution

Unless otherwise stated, results presented here are from analysis employing the Zero censored data handling method. Results using the alternative approaches are shown for comparison to demonstrate the impact of censored data handling on measures of absolute and relative extents of bromine substitution. Figure 2.1 shows box and whisker plots illustrating distributions of bromine fraction values for the four DBP compound classes considered. For each class, the extent of bromine substitution varied widely over the whole data set,

reflecting the broad range of water quality and treatment conditions for six quarters of data from 283 treatment plants (68 plants for THAA). Roberson (2002) discusses hydrological conditions for the ICR monitoring period in historical context. Unusually wet conditions across much of the U.S. during this time period, compared to average conditions over the past century, could mean that bromide levels for some ICR source waters were lower than what might normally be expected. This would have a downward influence on extents of bromine substitution in DBPs from affected utilities. The greatest variability in extent of bromine substitution was observed in the DHAN class. This is a function, in part, of consistently low absolute DHAN species concentration levels such that small shifts in composition translate to large differences in bromine fraction. Using the Zero censored data handling method, median bromine fraction values for the THM, DHAA, THAA, and DHAN classes were 10.1%, 7.9%, 8.7%, and 17.2%, respectively. All of the distributions were strongly right-skewed, that is, skewed toward higher values. The bromide to TOC ratio (Br/TOC) associated with each sampling event explained much of the bromine incorporation variability for each class, as expected on the basis of previous research (Shukairy et al. 1994; Krasner et al. 1996; Symons et al. 1996). Figure 2.2 shows the distributions of THM bromine fraction within approximate Br/TOC quartiles, demonstrating the strong trend of increasing extent of bromine substitution with higher Br/TOC levels. It is evident from Figure 2.2 that a large amount of variation in bromine fraction occurs within the higher Br/TOC ranges, indicating that the Br/TOC ratio alone does not fully explain the extent of bromine substitution in THMs. Similar patterns were observed for the other three DBP classes examined. In laboratory experiments, the degree of bromine substitution in THMs has also been found to depend on the ratio between source water bromide and applied chlorine dose

(Oliver 1980; Symons et al. 1993, 1996; Krasner et al. 1996b) and on characteristics of natural organic matter comprising the TOC (Liang and Singer 2003). These relationships were not further examined in the present analysis which centers on interdependencies among extents of bromine substitution in different byproduct classes rather than their absolute levels. Further work currently underway will address the influence of water quality and treatment factors on extents of bromine substitution in different DBP classes.

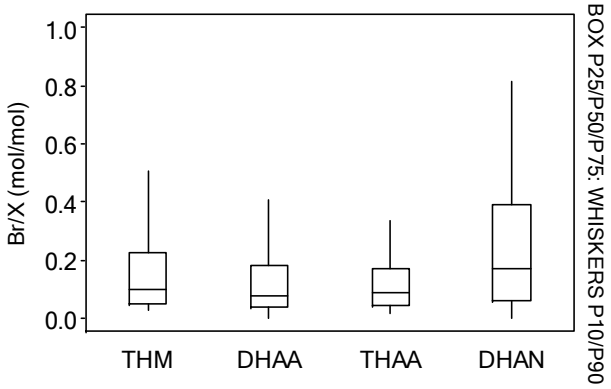


Figure 2.1 Distributions of bromine fraction in THM, DHAA, THAA, and DHAN classes (N = 5858, 5268, 1080, and 4880, respectively) showing 10th, 25th, 50th, 75th, and 90th percentiles for each class.

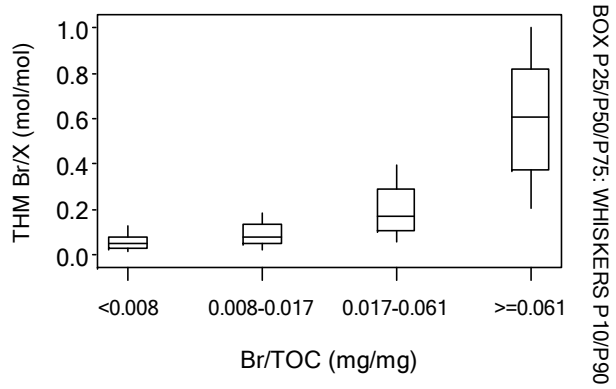


Figure 2.2 Distributions of THM bromine fraction for different ranges of Br/TOC (ratio between influent bromide and TOC at the first point of chlorine addition) showing 10th, 25th, 50th, 75th, and 90th percentiles for each range; N = 1524, 1323, 1303, and 802 for increasing Br/TOC ranges.

2.3.2 Relative Extents of Bromine Substitution and InterClass Bromine Fraction Correlations

Although extents of bromine substitution in each DBP class varied widely across the data set overall, bromine fractions in the different classes were strongly correlated on a sample-by-sample basis. Figure 2.3 illustrates this for a subset of 106 sample records covering data for five treatment plants (Group 1). Data were plotted for 5 plants at a time, grouped by sequential ID numbers, because this allowed adequate resolution of individual observations. The group shown in Figure 2.3 was selected for illustration because it encompassed a wide range in bromine fractions for the different DBP classes (due to a wide range in bromide levels) and included plants reporting data for all four THAA species. The x-axis positions in Figure 2.3 and subsequent similar plots correspond to individual water samples ordered by plant, sampling period, and sampling location. Figure 2.3a shows bromine fraction results for THM and DHAA classes and illustrates the extremely close correspondence and near equivalence between extents of bromine substitution in these two

classes, for any particular sample. The same sample observations are shown in Figures 2.3a–2.3c. Figure 2.3b adds results for the THAA class, showing that the extent of bromine incorporation in THAAs closely follows that in THMs and DHAAs, at equal or slightly lower levels. Adding DHAN results to the plot, Figure 2.3c shows that the correspondence between bromine fractions in different classes is maintained with inclusion of a fourth DBP class. The samples represented on the far right side of the plots have missing bromine fraction results for the DHAA and THAA classes. In contrast to the THAA class, the extent of bromine incorporation in DHANs tended to consistently exceed that in THMs and DHAAs. These data show how interclass correlations persist among bromine fraction results for all four classes examined while levels of bromine incorporation fluctuate widely. Plant influent bromide concentrations for these data ranged from 30 to 240 μg/L (median 72 μg/L), TOC at the first point of chlorination ranged from 0.4 to 4.1 mg/L (median 2.2 mg/L), and Br/TOC ratios ranged from 0.012 to 0.686 mg/mg (median 0.033 mg/mg). The phenomenon of interclass bromine fraction correspondence prevailed across the entire data set. Additional examples showing data for two more five-plant groups with different water quality conditions are provided in Figure 2.4 (Group 2) and Figure 2.5 (Group 3).

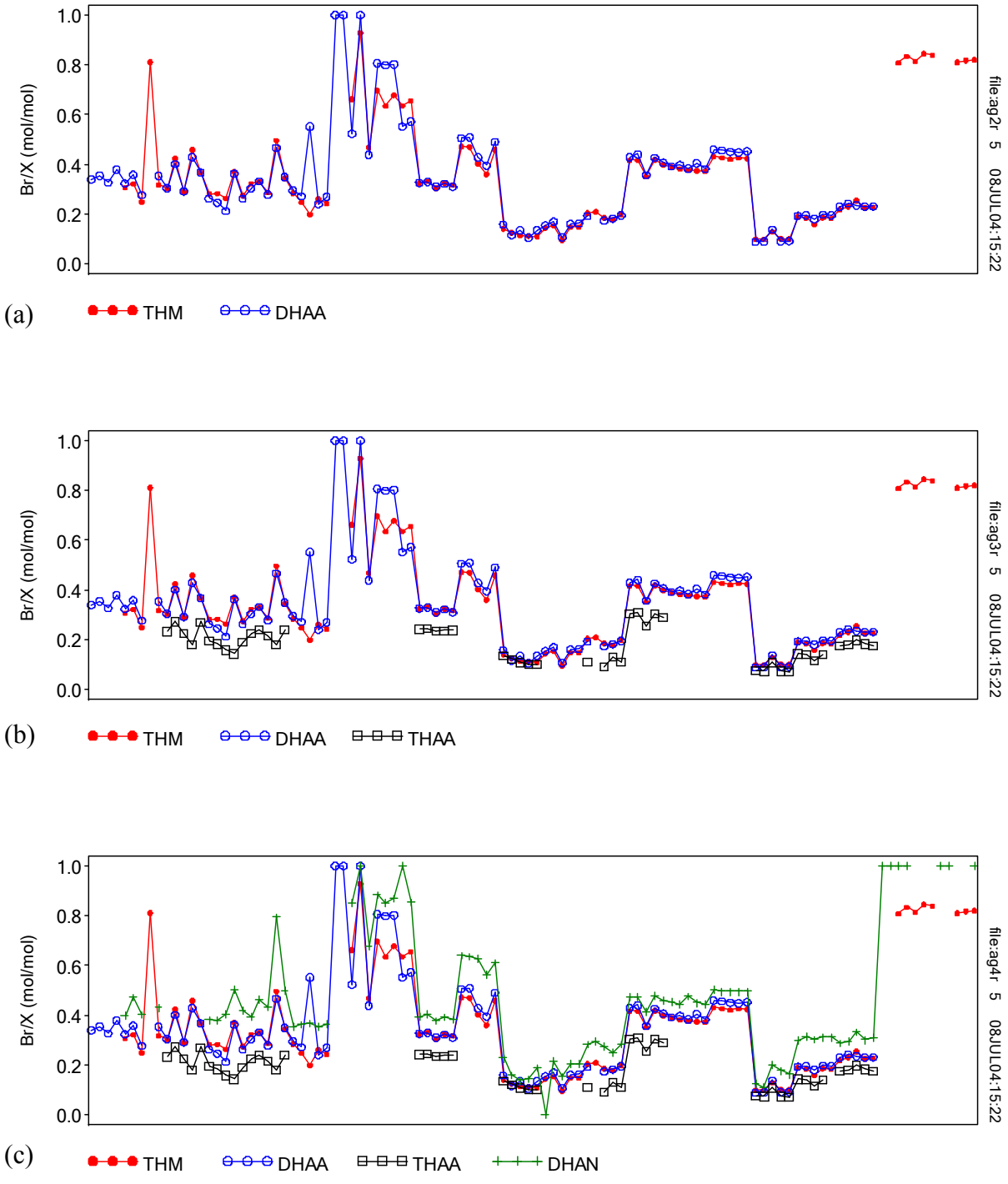


Figure 2.3 Bromine fraction data for five ICR plants (Group 1) showing all data for (a) THM and DHAA classes, (b) THM, DHAA, and THAA classes, and (c) THM, DHAA, THAA, and DHAN classes; N = 106, Br 30–240 μg/L (median = 72 μg/L), TOC <0.7–4.1 mg/L (median = 2.2 mg/L), Br/TOC 0.012–0.686 mg/mg (median = 0.033 mg/mg).

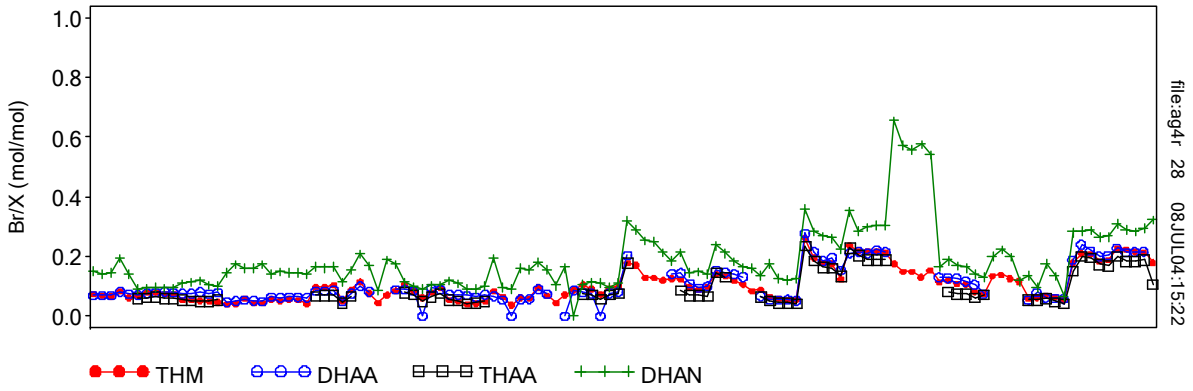


Figure 2.4 Bromine fraction data for five ICR plants (Group 2) showing all data for THM, DHAA, THAA, and DHAN classes; N = 120, Br <20–120 μg/L (median = 24 μg/L), TOC 1.5–7.1 mg/L (median = 2.4 mg/L), Br/TOC 0.004–0.036 mg/mg (median = 0.013 mg/mg).

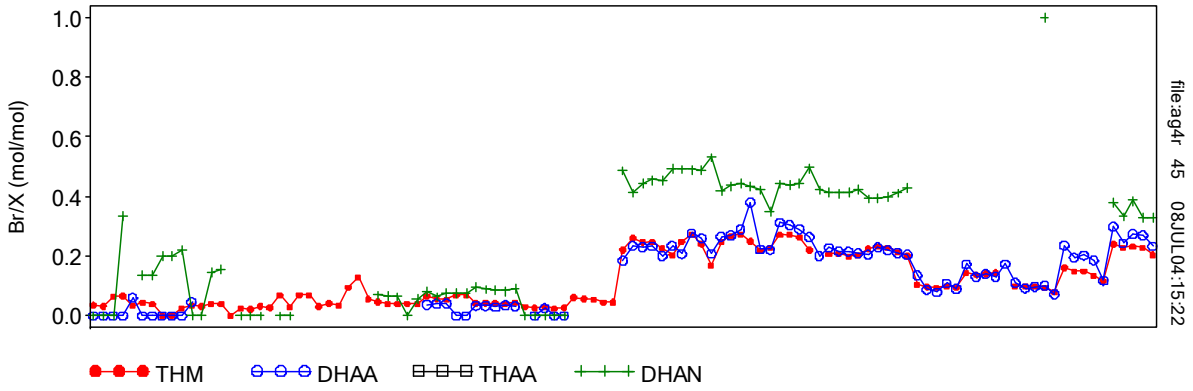


Figure 2.5 Bromine fraction data for five ICR plants (Group 3) showing all available data for THM, DHAA, THAA, and DHAN classes; N = 109, Br <20–100 μg/L (median = 25 μg/L), TOC 0.9–3.6 mg/L (median = 2.0 mg/L), Br/TOC 0.003–0.041 mg/mg (median = 0.041 mg/mg).

The scatterplot array in Figure 2.6 includes results for the entire dataset, plotted to illustrate the six possible pairwise associations among the bromine fraction measures for the

four DBP classes. Dashed diagonal lines are provided as reference points for equivalence between bromine fractions for the class pair and do not signify statistical or modeling output. The number of observations (*N*) and Spearman's Rank correlation coefficient (*SR*) specific to each bromine fraction class pair are indicated on the individual plots in Figure 2.6. Correlation coefficients, ranging from 0.745 to 0.939 for the six pairwise associations, were all statistically significant ($p < 0.0001$). The strongest correlations were observed for the THAA class with coefficients of 0.938, 0.939, and 0.862 for associations with THM, DHAA, and DHAN classes, respectively.

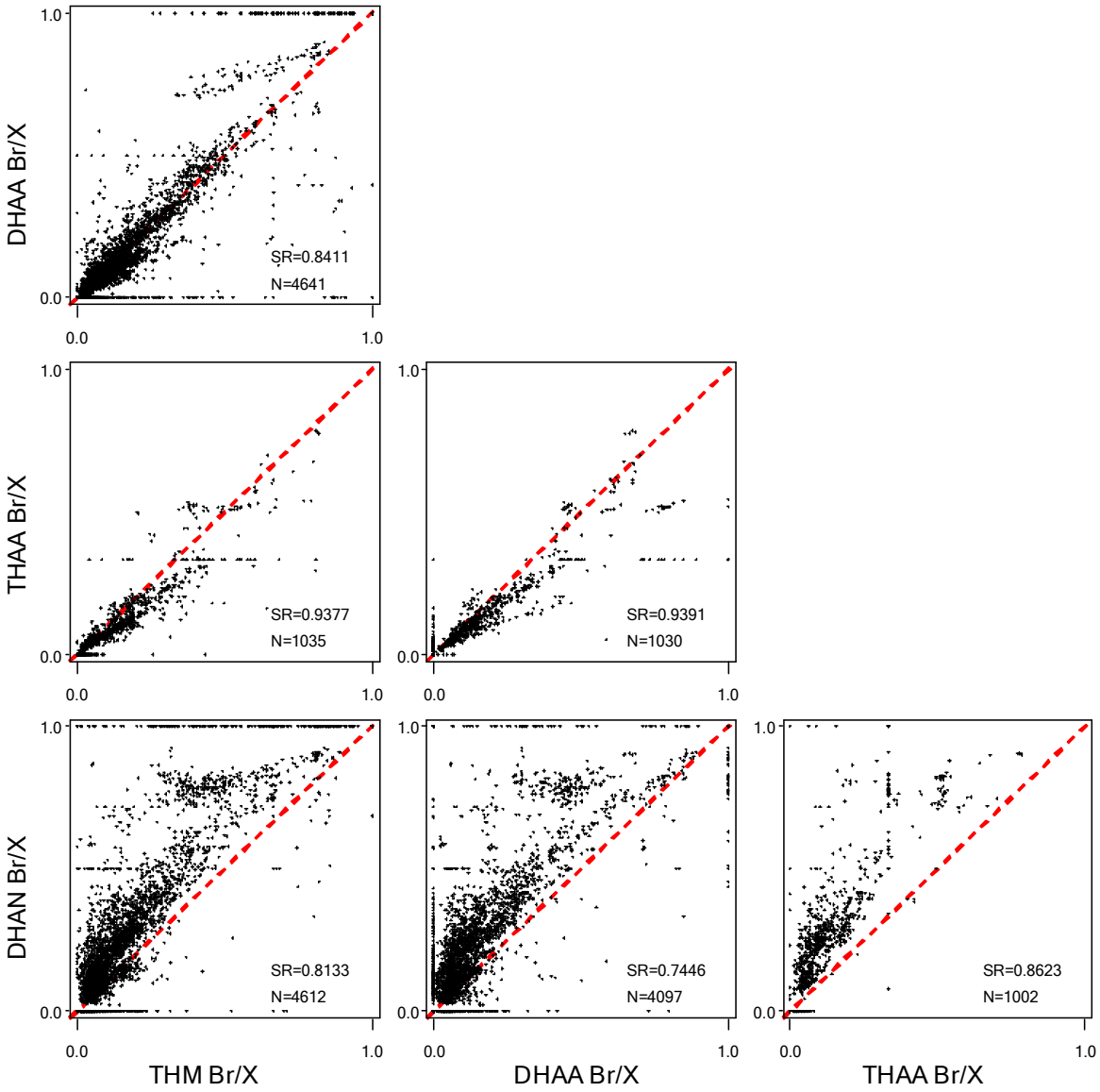


Figure 2.6 Scatterplot array of bromine fraction in four DBP classes showing all data (number of observations [*N*] and Spearman's Rank correlation coefficient [*SR*] shown on each plot); dashed line indicates $y = x$ for reference

Empirical distributions of class pair bromine fraction ratios are illustrated with box and whisker plots in Figure 2.7. For the class pairs DHAA/THM, THAA/THM, THAA/DHAA and DHAA/DHAN median values of bromine fraction ratios were 0.96, 0.79, 0.80, and 0.60, respectively. This indicates that bromine incorporation in DHAAs tended to

be nearly equivalent to that in THMs (96%) whereas bromine incorporation in THAAs tended to somewhat lower than in THMs (79%) or in DHAAs (80%). Bromine incorporation in DHANs was substantially higher than in any other DBP class examined. As shown in Figure 2.7, the median bromine fraction of halogen in DHAAs was only 60% that in DHANs or, equivalently, bromine incorporation in DHANs was 167% that in DHAAs at the median level. Similarly, bromine incorporation in DHANs was 160% that in THMs at the median (not shown in Figure 2.7). Furthermore, DHAN bromine fraction exceeded THM bromine fraction in 89% of samples.

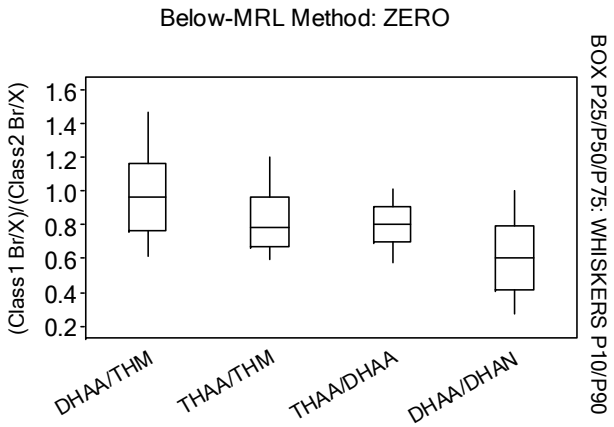


Figure 2.7 Distributions of ratios between bromine fractions for four class pairs (N = 3627, 946, 852, 2964 for DHAA/THM, THAA/THM, THAA/DHAA, and DHAA/DHAN, respectively).

2.3.3 Effect of Censored Data Handling

Measures of absolute and relative extent of bromine incorporation in different DBP classes were significantly impacted by the method employed for censored data handling. Data shown in Figures 2.1–2.7 reflect use of the Zero censored data replacement method.

Figure 2.8 compares empirical distributions of the ratio [THAA_Br/X]/[THM_Br/X] for three different censored data replacement methods. Both alternatives to the Zero method resulted in higher relative extents of bromine incorporation in THAAs compared to THMs. For the Level and Half-MRL methods, median [THAA_Br/X]/[THM_Br/X] values were 0.88 and 1.17, respectively, compared to 0.79 for the Zero method. The Half-MRL method also resulted in much greater variability in [THAA_Br/X]/[THM_Br/X], evidenced by the wider interquartile range and more strongly skewed distribution in Figure 2.8. This is attributable to the variety of replacement values employed for brominated THAA species in the Half-MRL method, compared to fixed replacement values in each of the other methods. The median values of [DHAA_Br/X]/[DHAN_Br/X] increased from 0.60 for the Zero method to 0.67 for the Level method, a similar effect to that seen for [THAA_Br/X]/[THM_Br/X]. The Half-MRL method, yielding a median value of 0.63, had much less effect on [DHAA_Br/X]/[DHAN_Br/X] than on [THAA_Br/X]/[THM_Br/X]. This is attributable to the lower censored data replacement values and the smaller portion of censored data for DHAAs compared to THAAs.

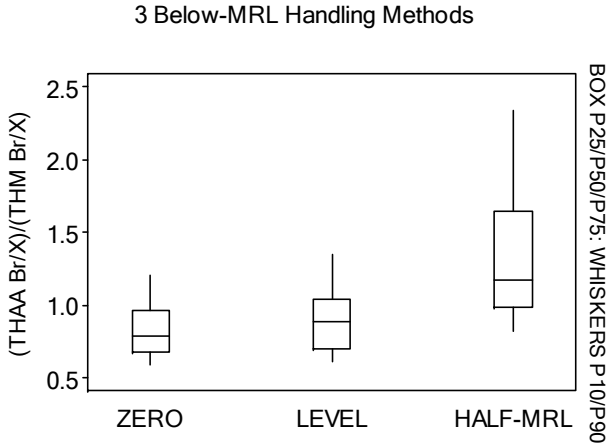


Figure 2.8 Distributions of the ratio between THAA and THM bromine fractions using three different methods for censored data replacement: Zero, Level, and Half-MRL.

Given the high MRLs for Br_2ClAA (2.0 μg/L) and Br_3AA (4.0 μg/L) in the ICR database and the fact that greater fractions of sample observations were censored for these analytes compared to the corresponding THM species which had MRLs of 1.0 μg/L (see Table 2.1), it is reasonable to conclude that uniform replacement of censored concentration values with zero imposed a downward bias on THAA bromine fractions compared to THM bromine fractions. Conversely, the Half-MRL method probably exerts an upward bias on THAA bromine fractions compared to THM bromine fractions because of higher replacement values for censored brominated HAA concentrations. For instance, using the Half-MRL method, 75% of censored Br_2ClAA results were replaced with 1.0 μg/L whereas all censored $CHBr_2Cl$ results were replaced with 0.5 μg/L. Similarly, 80% of censored Br_3AA results were replaced with 1.0 μg/L (58%) or 2.0 μg/L (22%) whereas all censored $CHBr_3$ results were replaced with 0.5 μg/L. The remaining 25% of censored Br_2ClAA results and 20% of censored Br_3AA results were replaced with 0.5 μg/L.

With respect to relative extents of bromine incorporation across byproduct classes, the Level method of censored data replacement was expected to remove differential censoring bias from comparisons between THM and THAA classes and between DHAA and DHAN classes. Although only eight records had measurable concentrations of all THAA and THM species, the median value of [THAA_Br/X]/[THM_Br/X] for those records was 0.94, closest to that obtained for the whole data set using the Level method for censored data handling. This supports the premise that the true ratio between THAA and THM bromine fractions is probably closer to 1.0, as it is for the ratio between DHAA and THM bromine fractions. Lower estimates may be biased by high MRLs for brominated THAA species and the practice of replacing left-censored analytical results with zero values in data analyses. A much larger number of records (778) were available with measurable concentrations of all DHAA and DHAN species. The median [DHAA_Br/X]/[DHAN_Br/X] value for these records was 0.68, in very close agreement with the value of 0.67 obtained for the entire data set using the Level method.

With consideration for censored data handling, this analysis shows that, while extents of bromine incorporation in DHAAs, THMs, and THAAs (to a somewhat lesser extent) tend to be comparable, significantly more bromine is taken up in DHANs. Censored data replacement is shown here to have a nontrivial impact on evaluations of DBP halogen composition when data are subject to differential censoring (because of unequal MRLs). In an earlier analysis of ICR THM and HAA data, Roberts et al. (2002) used the expectation of parallel interclass bromine substitution patterns to develop predictive models for brominated THAA species on the basis of measured THM species and Cl_3AA concentrations. The authors found that ratios between $BrCl_2AA$ and Cl_3AA sample concentrations (i.e.,

$BrCl_2AA/Cl_3AA$) tended to be equivalent to corresponding THM species ratios for the same sample (i.e., $CHBrCl_2/CHCl_3$) but that Br_2ClAA/Cl_3AA ratios were consistently lower than corresponding $CHBr_2Cl/CHCl_3$ ratios. Negative bias on Br_2ClAA concentration values stemming from use of the zero censored data replacement approach probably accounted for some of the observed differences.

Pourmoghaddas et al. (1993) and Cowman and Singer (1996) raised the issue of relative instability of the brominated THAA species. This may be a contributing factor to lower extents of bromine substitution in distribution system THAAs compared to THMs. First-order decay rates for THAA species in aqueous solution increase with the number of bromine substituents (i.e., $Br_3AA > Br_2ClAA > BrCl_2AA$) with decay being negligible for Cl_3AA (Zhang and Minear 2002). Moreover, THAA species decomposition proceeds through decarboxylation to yield the corresponding THM species so that bromine fraction loss in the THAA class is accompanied by bromine fraction gain in the THM class. For five days detention at 23°C, Zhang and Minear's decay rates translate to approximate losses of 20, 3.5, and 0.5% for Br_3AA, Br_2ClAA, and $BrCl_2AA$, respectively (Zhang and Minear 2002). Taking into consideration accompanying increases in brominated THMs and the lower typical temperatures (mean 18.8 °C) and estimated detention times (mean 27 h) for the ICR samples examined here, it is plausible that brominated THAA decay contributed to the somewhat lower extent of bromine substitution in THAAs compared to THMs observed in this work and by Roberts et al. (2002).

With regard to the higher extent of bromine substitution in DHANs compared to THMs, DHAAs, or THAAs, one explanation could involve differences in formation kinetics. Researchers have established that bromine reacts faster than chlorine in halogen substitution

reactions (Oliver 1980; Gould, Fitchhorn, and Urheim 1983; Symons et al. 1996; Westerhoff, Chao, and Mash 2004). For most ICR plant-sampling periods under study, bromide concentrations were limited relative to other reactants involved in DBP formation (i.e., organic precursor levels and applied chlorine dose). Therefore kinetic competition among labile halogen substitution sites on organic precursors would influence apportionment of available bromine. If DHANs form quickly compared to the other DBP classes they would take up more of the limited available bromine. The finding of substantially higher extents of bromine substitution in DHANs compared to THMs or HAAs suggests that the presence of nitrogen as an activating agent in the aromatic ring structures of organic precursors facilitates halogen substitution more so than other activating agents. This is consistent with the findings of Reckhow, Singer, and Malcom (1990).

Results show that the Level method reduced differential bias between bromine fraction measures for different DBP classes and thereby allowed for a truer picture of relative extents of bromine incorporation across classes. However, because the approach involves discarding perfectly good analytical results, it is undesirable for most applications. The Half-MRL method produced positive bias in THAA bromine fraction and increased its variance substantially. Because the remaining analyses discussed here concern interdependence among DBP class bromine fraction measures rather than their absolute or relative levels, the Zero method for censored data handling was employed to retain precision in the covariance structure of the data. The comparison of simple censored data handling methods was conducted here primarily as a sensitivity analysis to demonstrate the potential impact such handling can have on conclusions drawn from data analysis. This issue, not commonly addressed in DBP studies, is especially important to consider in data analysis involving

classes of DBPs that occur predominantly near reporting limits. More sophisticated approaches to filling in censored data may be useful in future studies (Helsel and Gilliom 1986; El-Shaarawi 1989; Haas and Jacangelo 1993; Liu, Kolpin, and Meeker 1997; Taylor et al. 2001; Lubin et al. 2004).

2.3.4 Multivariate Outlier Detection

As discussed, the test for multivariate outliers is premised on assumptions of normally distributed variables with homogeneous variance. Figures 2.9 and 2.10 illustrate the effect of the arcsine transform on conditioning THM bromine fraction data to better meet these assumptions (see Equation 2.2). Comparison of the histograms in Figure 2.9 shows that the transformation substantially improved the THM bromine fraction distribution insofar as it could be better approximated by a normal distribution. Skewness was reduced from 1.98 to 1.04 and kurtosis was reduced from 3.38 to 1.52. Figure 2.10, which describes distributions of THM bromine fraction for sequential deciles of the data, shows how the arcsine transformation reduced variance inhomogeneity. In the untransformed data there was a strong tendency towards increasing variability at higher bromine fraction values, as evidenced by progressively increasing interquartile ranges (represented by the box heights). Although not entirely removed, this tendency is reduced in the transformed data which has more consistent variability throughout the data range. Similar improvements in data condition, with respect to meeting distributional assumptions, through application of the arcsine transformation were obtained for the DHAA, THAA, and DHAN class bromine fraction variables.

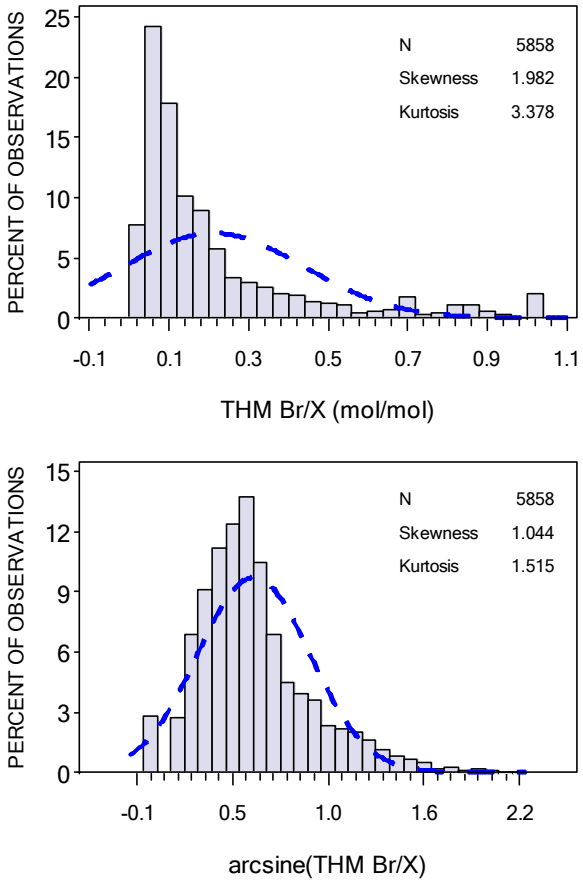


Figure 2.9 THM bromine fraction distributions before and after arcsine transform; dashed line indicates best-fit normal distribution.

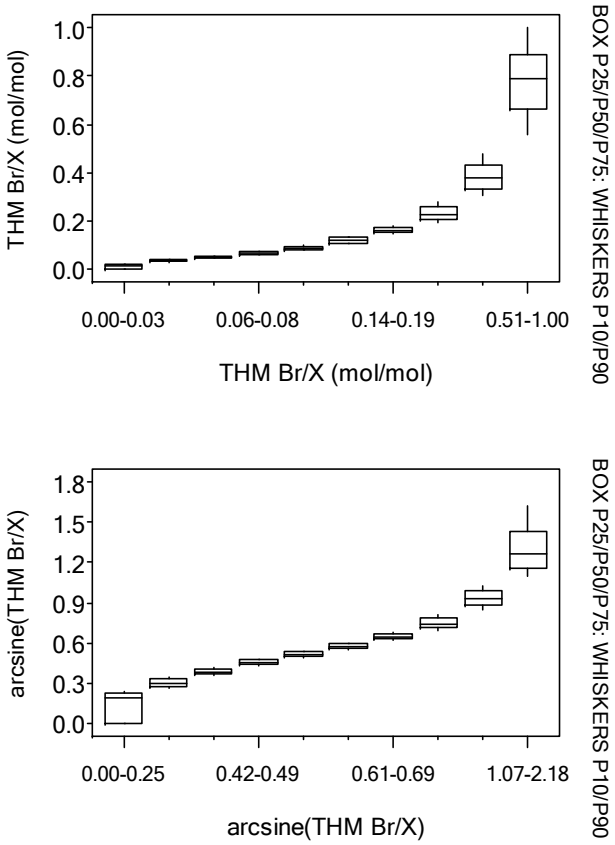


Figure 2.10 THM bromine fraction distributions within data deciles before and after arcsine transform.

Figure 2.11 provides an example of multivariate outlier detection test results, showing transformed bromine fraction data for samples from 5 treatment plants (Group 3). The untransformed bromine fraction data for this group of samples were shown in Figure 2.5. One sample among the 98 observations, indicated by the star plotting symbol, has a clearly discordant DHAA bromine fraction value and was successfully identified as a multivariate outlier by the test. The discordant nature of this sample is less clearly evident from graphic inspection of the untransformed data in Figure 2.5. The transformation's pre-multiplier term more heavily weights data for samples with higher absolute molar halogen concentrations

while the arcsine function magnifies the lower end of the proportion scale (see Equation 2.2). This reshaping of the data dampens differences between bromine fraction values for extremely low concentration samples, emphasizes differences for higher concentration samples, and helps distinguish differences in the low bromine fraction range. Thus, the line traces in Figure 2.11 are quite different than for the corresponding untransformed data in Figure 2.5. Most notably, the line trace for THM has moved up, due to higher absolute concentrations, that for DHAN has moved down, due to much lower absolute concentrations, and the discordant DHAA value stands out more prominently. The discordant point was identified as an outlier in both the three-way multivariate outlier test involving THM, DHAA, and DHAN and in the two-way test involving THM and DHAA. However, it was not identified as an outlier in the two-way test involving DHAA and DHAN, probably because the large amount of scatter in DHAN bromine fraction rendered this test less sensitive. Inspection of species data for the identified sample revealed a reported BrClAA result (35.0 μg/L) that was an order of magnitude higher than reported BrClAA concentrations for the other four samples from the same plant/sampling period (range 5.1-6.7 μg/L). No other DBP species' results for the sample in question were unusual compared to corresponding data for other samples from the same plant/sampling period. The anomalous result was assumed to reflect a decimal place data entry error (i.e., the analytical result was probably 3.5 μg/L). Additional examples of the use of the arcsine transform and multivariate outlier detection test for identifying discordant values are presented in Appendix A (Obolensky and Singer 2005, Supplemental Information).

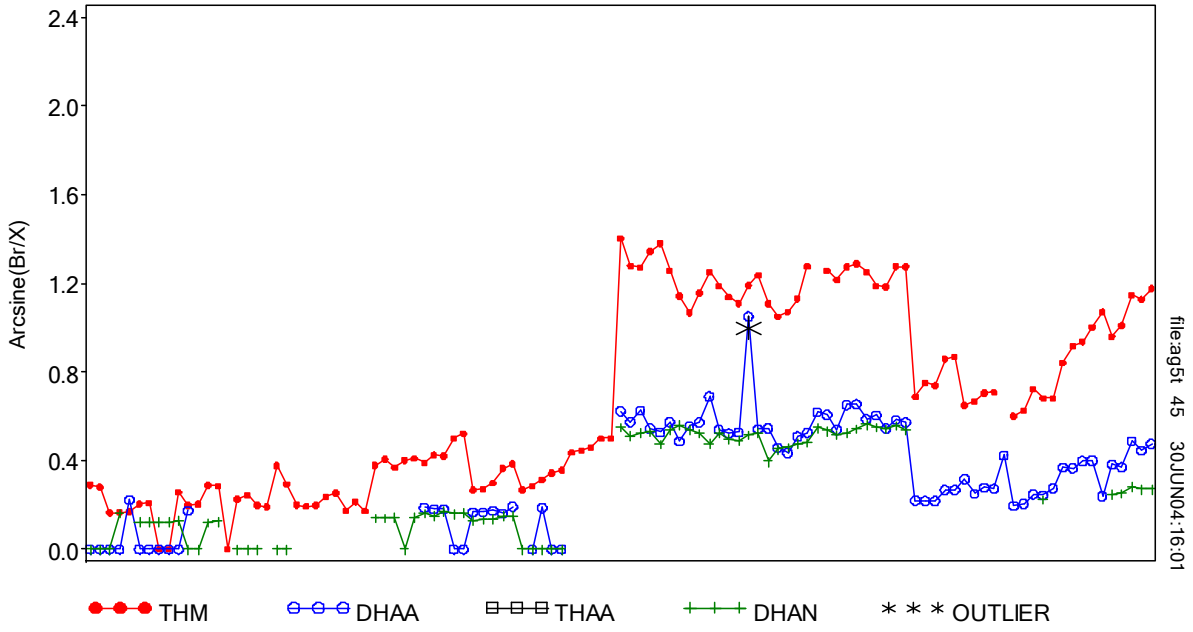


Figure 2.11 Arcsine transformed bromine fraction data for 5 plants (Group 3) showing detected outliers; $N = 109$

Figure 2.12 provides an overview of multivariate outlier test results. Included are scatterplots of transformed bromine fraction data for DBP class pairs (DHAA vs. THM, THAA vs. THM, THAA vs. DHAA, and DHAN vs. THM), showing all sample observations with identified multivariate outliers circled. The plots show that the majority of identified outliers were points in the periphery of the data cloud, indicating that the method was successful at capturing observations manifesting discordancy with respect to interclass bromine substitution patterns. Because such plots can only show two class-dimensions at a time, some outlier points may fall within the expected distribution (the data cloud) because they were discordant with respect to a DBP class not shown on that particular plot. For example, there are five THM bromine fraction values in the plot of DHAN vs. THM (lower left quadrant) that fall within the data cloud (points labeled a through e in Figure 2.12).

However, these samples were outliers based on discordancy between THM and DHAA bromine fraction values, not between THM and DHAN bromine fraction value. The five points can be found in the periphery of the data cloud in the plot of DHAA vs. THM (upper left quadrant). This is a shortcoming of the two-dimensional data presentation.

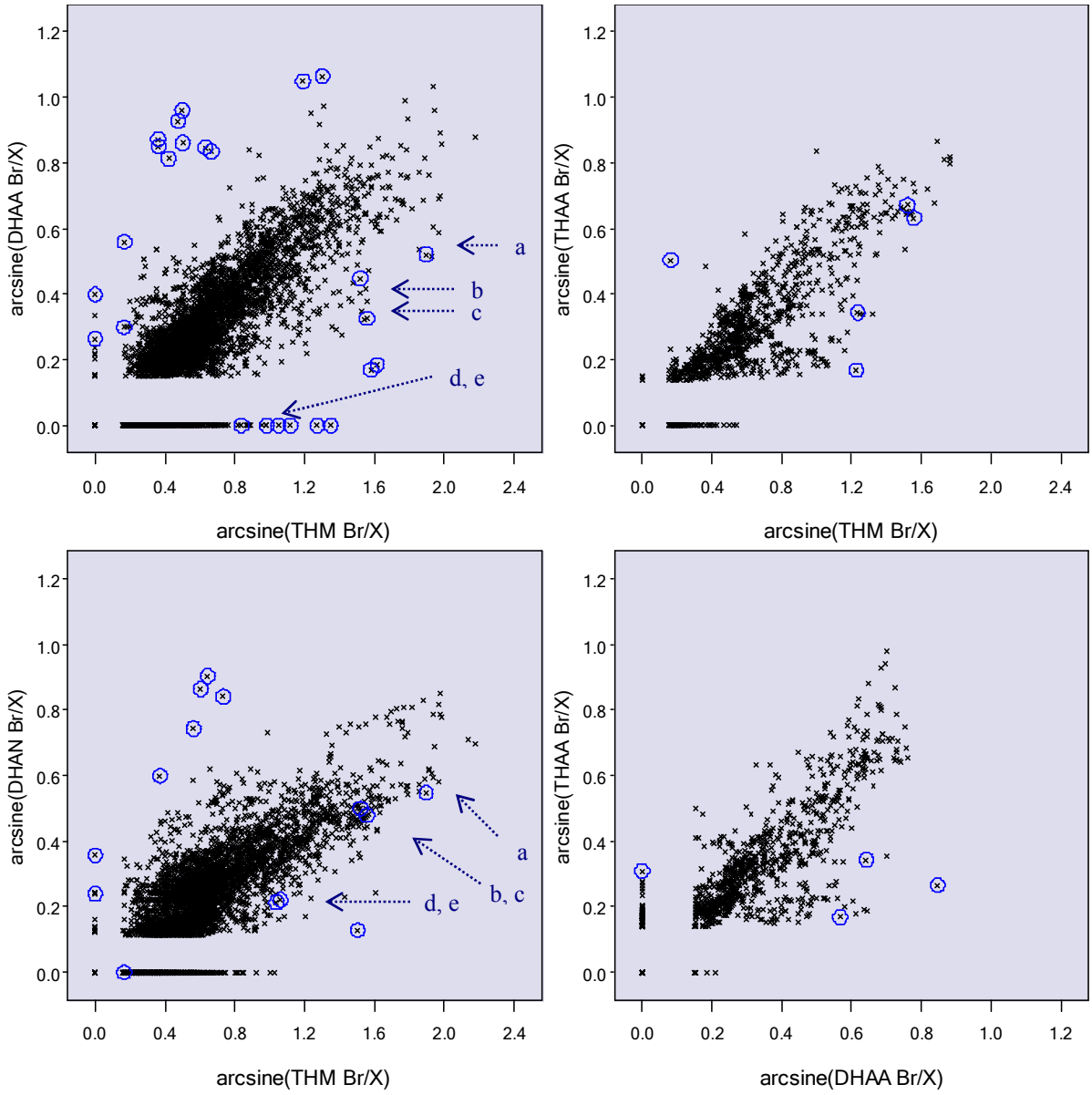


Figure 2.12 Scatterplot array of arcsine-transformed bromine fraction data for four DBP class pairs showing all data with identified multivariate outliers circled: clockwise from top left DHAA vs. THM (N=4641); THAA vs. THM (N=1035); DHAN vs. THM (N=4612); THAA vs. DHAA (N=1030): points labeled a through e in the upper left plot represent identical sample observations to corresponding labeled points in the lower left plot

Overall, 48 of the 6,565 sample observations in the data set were identified as outliers in this analysis. Reported species concentration data for these samples were carefully reviewed in the context of concurrent data for the same plant/sampling period, as exemplified

by discussion of Figure 2.11. Twelve of these observations were associated with extreme values of transformed THM bromine fraction for samples having elevated THM concentrations dominated by brominated species but negligible concentrations of other measured DBPs. Half of these samples were from a single water utility. These 12 observations are not indicated as outliers in Figure 2.12. Of the remaining 36 identified outliers, 7 were associated with lower than expected DHAA bromine fraction values in samples with low DHAA concentrations and censored data for brominated DHAA species. These results may be related to DHAA degradation. Fourteen identified outliers were associated with fairly obvious data entry errors such as misplaced decimal points or interchanged species entries. The remaining 15 outlier samples had one or more DBP species with an anomalously high or low concentration value that would suggest a need for chromatogram review. Finally, graphic inspection of all results in the form of plots similar to Figure 2.11 indicated that six observations exhibiting discordancy with respect to interclass bromine substitution patterns were not identified by the outlier tests.

The finding that patterns of bromine substitution in different DBP compound classes are highly correlated has important implications for DBP research beyond its potential applicability for quality assurance data review purposes. Given that these patterns were found to be consistent across all four compound classes examined here, it is likely that they extend to other DBP classes such as haloketones, halopicrins, haloacetaldehydes, and halofuranones (MX and brominated homologues) where brominated species are not routinely measured. Indeed, earlier studies have noted correspondences between species with analogous halogen substitution in different DBP classes. Using data from a field survey Krasner et al. (1990) noted that dihaloacetic acid concentrations correlated with chloroform concentrations but not

with any other of the THM species. Suzuki and Nakanishi (1995) observed good correlation between analogous compounds in the THM and trihalo-5-hydroxy-2(5H)-furanone (i.e. MX and BMX1-3) classes. Onstad (2003) found brominated HAAs, but not brominated THMs, to be good predictors for brominated furanones. The high degree of interdependence between extents of bromine substitution in different DBP classes indicates the presence of much redundant information in the set of discrete concentration measurements for individual species. The current work demonstrates how this phenomenon can be applied to augment traditional QA/QC procedures through the identification of discordant samples for further analytical review. Many other applications relevant for water treatment and health effects studies can be envisioned, including development of predictive relationships for DBP species that are difficult to measure, and optimization of monitoring strategies. For example, for certain objectives, measurement results for a limited subset of species from a group of DBP compound classes might provide aqequate information about concentration levels of all members of the classes.

CHAPTER 3: INFORMATION COLLECTION RULE DATA EVALUATION AND ANALYSIS TO SUPPORT IMPACTS ON DISINFECTION BY-PRODUCT FORMATION

3.1 BACKGROUND AND INTRODUCTION

The Information Collection Rule (ICR) was designed to obtain water quality, water treatment, and occurrence information needed for development of Safe Drinking Water Act regulations governing control of pathogenic microorganisms and disinfection by-products (DBPs) in drinking water (U.S. EPA 1994a,b, 1996a). The rule required all large public water systems in the United States (those serving populations of 100,000 or more) to report sampling results for water quality and microbiological and DBP parameters, and to provide detailed information about treatment plant design and operations. The 296 affected utilities, ranging across 49 states and territories, reported ICR data from 500 water treatment plants and their source waters and distribution systems for 18 monthly monitoring periods from July 1997 through December 1998. Additional data from DBP precursor removal studies (U.S. EPA 1996a, 2000c; Hooper and Allgeier 2002; DiGiano and Bond 2004) and protozoan method performance studies (Connell et al. 2000; Messner and Wolpert 2002) augment ICR monitoring results.

Preparation for and implementation of the ICR was a very large and complex undertaking on the part of the U.S. Environmental Protection Agency (U.S. EPA), the

American Water Works Association, participating utilities, and analytical laboratories. Extensive requirements governing sample collection, analytical methods, laboratory certification, and quality control applied to data collection and reporting activities. The ten guidance manuals and five instructional videos published by the U.S. EPA for ICR implementation reflect the effort involved (Wysock et al. 2002). A data acquisition and reporting system was designed to ensure that validated ICR data met rigorous quality assurance standards developed for the program. The overall process involved data submission and validation, U.S. EPA summary report generation, utility and laboratory review of summary reports, and resubmission to correct errors. These review cycles controlled data entry and omission errors to the extent feasible. Validated ICR data were housed in a centralized Oracle database system (ICR Fed) (Oracle Corporation, Redwood Shores, CA) and a primary auxiliary Microsoft Access database (Aux 1) (Microsoft Corporation, Redmond WA) was subsequently developed to facilitate ICR data retrieval and analysis by the public (U.S. EPA 2000b). Fair et al. (2002) discuss the quality assurance program for ICR analytical data and present precision and accuracy results for chemical analytes.

ICR sampling locations included treatment plant influents, points immediately downstream of treatment train unit processes, finished waters, entry points to the distribution system for blended systems (i.e., systems with water blended from different plants), and distribution system sites. ICR samples and plant information were collected monthly unless stated otherwise. Finished water samples were collected at a point after all treatment processes were complete, including clearwell storage and the final point of disinfectant addition. Standard water quality variables (pH, alkalinity, hardness, temperature, and

turbidity), total organic carbon (TOC), ultraviolet absorbance at 254 nm (UV254), and residual disinfectant concentrations were monitored at plant influents and key points through treatment plant process trains, depending on process configurations and disinfectant addition points. Bromide and ammonia were also monitored at treatment plant influents. DBP samples were collected quarterly at filter effluent locations if chorine was added upstream, at all finished water locations, at points of entry to the distribution system (for blended systems), and at four distribution system locations for each treatment plant. Standard water quality variables and free and total chlorine were monitored concurrently at all DBP sample collection points.

For each unit process, the basin (or pipe) volume and flow rate for the sampling date were reported together with the identity, formula, and dose of any chemicals or disinfectants added. Sequence numbers associated with each unit process indicated the upstream-downstream order of processes in each treatment train. Additional reported plant and utility information included influent and effluent flow rates, wholesale and retail population served, flow and type of any additional waters blended into the treatment train after the influent, sludge processing data, and information about the water resources serving the plant intakes.

The ICR provides the largest comprehensive drinking water database collected to date. It represents water quality, treatment, and distribution system characteristics across a nationally representative set of source waters, treatment plants, and distribution conditions for the 1997−1998 monitoring period. ICR data were used extensively to support development of the Stage 2 Disinfection Byproducts Rule and the Long Term 2 Enhanced Surface Water Treatment Rule through construction of treatment and occurrence baselines for regulatory impact forecasting (U.S. EPA 2000a; Scharfenaker 2001; U.S. EPA 2006a,b).

McGuire, McLain, and Obolensky (2002) report on the data analyses conducted for that effort.

The work reported herein is part of a study being conducted to extract and analyze valuable and previously untapped information about DBPs from the ICR database. The large scale of the database, together with the rigor and consistency of quality assurance standards and analytical methodologies employed for data collection, render the data especially suitable for a statistically based analysis of observational data. A thorough study of these data is expected to contribute to an improved understanding of DBP formation and occurrence and the effects of water quality characteristics and treatment practices on DBP formation under field conditions. This information should help utilities better evaluate control strategies that limit overall DBP formation, as well as the formation of selected DBPs, and assist them in complying with regulations and reducing consumer exposure to DBPs. Because ICR source waters, treatment plants, and distribution systems capture the spectrum of conditions across the United States, results of this study should be relevant and useful for a wide audience of utilities, consultants, and regulatory agencies.

3.2 OBJECTIVES

The primary objectives of this paper are to present sampling distributions for important water quality and treatment parameters influencing DBP formation, and to provide documentation of ICR data handling and analysis methods in support of this objective. This characterization of the database was needed for ongoing statistical analysis of interrelationships among the data and should be of general benefit to investigators utilizing these data. For example, Obolensky and Singer (2005) provide an analysis of ICR DBP halogen speciation patterns and Archer and Singer (2005) address relationships between

source water quality and the removal of TOC and UV absorbing substances. ICR DBP occurrence data are not presented here but have been summarized, along with microbiological data and other ICR information, in McGuire, McLain, and Obolensky (2002).

Underlying this characterization of the ICR database are results of data-screening and review efforts conducted from an end-user's perspective. Confidence in conclusions drawn from analysis of any data set depends on the soundness of the database and the transparency of data-handling and analysis methodologies. Despite the extensive ICR quality assurance program, the sheer quantity of ICR descriptive information and analytical results suggests the inevitability that some of these data may be faulty. For example, utility data entry errors that remained uncorrected during the formal ICR data review processes generally were beyond the control or remedy of database management or technical intervention. Anomalous data entries, occasionally encountered in the course of this study, had the potential to distort analysis results. Additionally, some missing data appeared to be recoverable. Hence, a secondary objective of the current research was to systematically review DBP-related water quality and treatment data in order to identify, flag, and possibly correct outliers or missing results in a rational and justifiable manner before conducting further analyses. This additional quality assurance effort allows subsequent analyses of ICR data to be based on the best possible data set and avoids conclusions based on faulty or incomplete data. Methodologies employed for this review are documented here together with the findings.

3.3 ANALYSIS PROCEDURES

ICR data were obtained from the U.S. EPA Auxiliary 1 Database Version 5.0 (Aux 1, U.S. EPA 2000b). The "AUX 1 Tables/Entity Types Details Report," included in the

database, documents the requirements involved in populating Aux 1, the public database, from validated data in the original Oracle data repository, ICR Fed. Data from Aux 1 tables were joined to create an analytical data set of ICR treatment plant process train and water quality information for export to and subsequent analysis in the SAS software environment (SAS Institute, Cary, N.C.). The analytical data set was modified (see the following) to enable use of censored data (analytical results below the minimum reporting level, or MRL) and to reflect data screening and review efforts undertaken for the current research.

Because statistical distributions for many water quality variables tend to be lognormal and/or strongly positively skewed, logarithmic plotting scales are employed here and the median is used to characterize central tendency (mean values are provided in tabular summaries to permit comparisons with other data sets). Box and whisker plots, used to illustrate and compare data distributions, indicate the median (central horizontal line), interquartile range (25th–75th percentile, box boundary), 10th percentile (lower whisker) and 90th percentile (upper whisker) values, with results outside the 10th–90th percentile levels shown as discrete points.

Analysis of variance (ANOVA) methods were used to test hypotheses about population means of water quality and treatment variables among different data subgroups defined by levels of classification variables (e.g., source water type, disinfectant type, etc.). Because ICR data are generally unbalanced (i.e., numbers of observations differ among subgroups), least-squares means were used in comparisons for multiway ANOVA tests. The Tukey-Kramer multiple comparison adjustment for *p* values was used when more than one pair of means was compared in a particular analysis (SAS Institute 1999). With the exception of pH, ANOVA was conducted on log-transformed response variables in order to better meet

the method's assumption of normally distributed data. The nonparametric Spearman's rank correlation on untransformed data was used to measure correlation where needed.

3.3.1 Censored Data

Aux 1 analytical results that were below the MRL were replaced with one-half the ICR analyte-specific MRL value (MRLs are listed in the Aux 1 table "TUXANLYT" and discussed in U.S. EPA 1996b). Accordingly, below MRL results for bromide, TOC, and UV254 (represented by "−999" in Aux 1) were replaced with 0.01, 0.35 mg/L, and 0.0045 cm^{-1}, respectively. TSUVA (L mg^{-1} m^{-1} units), a variant of specific ultraviolet absorbance (SUVA) using TOC in place of dissolved organic carbon (because the latter parameter was not measured as part of the ICR program), was derived as 100×UV254/TOC. ICR MRL values for chlorine residual and ammonia concentrations were not specified as such but the data entry software did not allow input of numbers below 0.1 mg/L. Censored values for these analytes (i.e., values <0.1 mg/L), represented by "−333" entries in Aux 1, were uniformly replaced with 0.05 mg/L in the analytical data set, following the half-MRL approach used for other analytes and treating the minimum 0.1 mg/L entry as the MRL.

3.3.2 Process Train Logic

Several treatment parameters related to DBP formation were derived using SAS data step programming logic. At the unit process level, these parameters included hydraulic residence time (HRT) and process effluent chlorine residual concentration. Theoretical HRT, in minutes, was calculated as $V/(694.44 \times Q)$ using reported process volume (V, in gal.) and flow rate (Q, in million gallons per day, or MGD) data. In the case of missing or zero-valued data for volume (less than 0.5% of records), HRT was set equal to the reported T50 value

(the time required for 50% of the water to pass through the basin in question). For unit processes downstream of chlorine or ammonia addition, chlorine contact time was set equal to the HRT and the associated chlorine residual concentration was determined as the nearest available downstream result prior to further disinfectant or blended water addition.

Derived parameters at the plant/month level included: prechlorination practice (yes or no); number of chlorine addition points; water quality results at each point of chlorine addition; influent bromide/TOC ratio (mg Br/mg TOC); bromide/chlorine ratio (mg Br/mg Cl_2, using first chlorine dose); total chlorine dose; chlorine/TOC ratio (mg Cl_2/mg TOC, using total chlorine dose and TOC concentration at the first chlorination point); total ammonia dose; chlorine/ammonia ratio (mg Cl_2/mg N, using first ammonia dose and sum of existing free chlorine residual and concurrent chlorine dose, if necessary, at the point of ammonia addition); and overall disinfection category (see the following). A plant-month was classified as practicing prechlorination unless a clarification process preceded the first point of chlorine addition. Flocculation, sedimentation, filtration, and the use of solids contact clarifiers, adsorption clarifiers, membrane filtration, slow sand filtration, and granular activated carbon filtration were considered clarification processes for this purpose. Nine overall disinfection categories were determined on the basis of Aux 1 categories for treatment plant disinfection type (*WTP_DIS*: CL2=free chlorine; CL2_CLM=free chlorine and chloramines; CLM=chloramines; CLX=chlorine dioxide; or O3=ozone) and distribution system disinfection type (*DS_DIS*: CL2 or CLM), with consideration for whether ozone and chlorine dioxide plants also used free chlorine during treatment.

3.3.3 Data Screening

The data review conducted for this work addressed both questionable and missing records, encompassing Aux 1 monthly plant and treatment data, and water quality results for influent through finished water samples. Questionable data were flagged and missing data for categorical variables were replaced (see the following) with appropriate values where possible. In a few cases questionable numerical data were replaced with corrected values using methods described in the Results section. Resulting flags, replacement values, and explanations were compiled in a project quality assurance table.

Due to the presence of a few unrealistically high reported bromide concentrations, all bromide results from plants having any bromide result above 1 mg/L were screened. TOC data were examined for all plant/months where reported TOC concentration at the first point of chlorine addition was ≥1 mg/L higher than at the plant influent. Similarly, UV254 data were examined where this difference exceeded 0.5 cm^{-1}. Aux 1 values for both UV254 and TOC are averages of reported duplicate sample results and validated ICR data had to meet a 20% sample pair relative percent difference limit for these analytes. Assuming this would reduce the chance of utility data entry error, anomalous Aux 1 data for these parameters were flagged only in cases of extreme discrepancy. All Aux 1 pH results below 4.0 or above 11.0 were examined to determine their plausibility. pH values below 2.0, considered implausible for full-scale treatment, were flagged. Otherwise, results were compared with upstream and downstream values, and with results for the same location in other sampling periods. Unit process flow and volume records were screened as a quality assurance measure for derived chlorine contact time estimates. For each distinct process at each plant, the relative ranges for flow and volume were calculated as (maximum value - minimum value)/(average value).

Data were examined further if the relative range for flow or volume exceeded 1.0 for a unit process.

Null (missing) values for categorical variables describing source water type (*MSRC_CAT*), plant disinfectant type (*WTP_DIS*), and distribution system disinfectant type (*DS_DIS*) were traceable to deviations from various Aux 1 requirements for underlying primary information. Appropriate replacement values for these entries could usually be determined with confidence based on underlying water resource and process information. Additionally, some Aux 1 *WTP_DIS* and *DS_DIS* entries that were incorrect due to faulty underlying disinfectant process information were replaced with corrected values. Data were also examined wherever *MSRC_CAT*, *WTP_DIS*, *DS_DIS*, or *MWTPTYPE* (treatment plant type) values varied across the 18-month ICR period for an individual plant.

Graphical analysis was used to aid in reviewing water quality and treatment information. Plots of monthly data were scanned for unusual fluctuations in water quality or disinfectant dose across sampling months and data were examined more closely as needed. Plots of individual plant/month data, illustrating patterns of water quality data and disinfectant doses through the process train, were used to assess data consistency within and between sampling months.

3.4 RESULTS

3.4.1 Data Screening

Results of the data screening and review effort are summarized in Table 3.1, which lists the total number of Aux 1 records and the number and percent of entries flagged or replaced for each parameter considered.

Table 3.1 Aux 1 Data Screening and Review Results

Parameter	Number of Aux 1 records	Number of records replaced[a]	Number of records flagged[b]	% of records affected
Bromide	8,720	12	—	0.13
pH	50,350	—	20	0.04
TOC	31,895	—	37	0.12
UV254	31,930	1	32	0.10
MWTPTYPE	8,953	3	—	0.03
MSRC_CAT	8,953	482	—	5.4
WTP_DIS	8,953	1,797	—	20.1
DS_DIS	8,953	227	—	2.5
Chlorine dose	14,489	—	296	2.0
Ammonia dse	3,030	—	25	0.8
Residual free Cl_2	30,139	—	164	0.5
Residual total Cl_2	31,490	—	101	0.3
Unit process volume	39,595	—	217	0.5
Unit process flow	43,000	—	491	1.1

[a]Missing or questionable value replaced in analytical data set.
[b]Value tagged as questionable in analytical data set.

Very few data were flagged for bromide, TOC, UV254, or pH. Seven questionable bromide results for one plant were replaced after determining that decimal placement data entry errors had been made (this was verified with utility personnel at the plant in question). Five additional bromide results from four other plants were replaced based on decimal placement data entry errors. Flagged TOC and UV254 results involved substantial and

implausible increases or decreases in the parameter value across a treatment train for a plant/month or across sampling months at the same plant. For example, a plant having flocculation tank effluent TOC concentrations typically in the 2−3 mg/L range reported 30 mg/L at this location in one month, although influent TOC was below 5 mg/L. In another case, an extremely high flocculation tank effluent UV254 result of 0.810 cm^{-1}, far outside the expected range for this parameter, was five times greater than the plant influent result. Besides the 14 pH results below 2.0, six other pH results were flagged based on inconsistencies with data for upstream/downstream samples or with samples for the same location in other sampling periods.

As indicated in Table 3.1, replacement values were determined for substantial numbers of *MSRC_CAT* and *WTP_DIS* entries (5 and 20% of ICR records, respectively). These were mostly replacement of missing values. Some corrections to *WTP_DIS* (42) and *DS_DIS* (51) entries were made after discovering errors in process train disinfection information. Three *MWTPTYPE* values of "OTH" (other) for one treatment plant were corrected after verifying with the plant that the same conventional process was used throughout the ICR period.

Reasons for flagged flow and volume data included apparently mistaken loss or gain of digits in data entry, substantial change in the process flow across plant-months without corresponding change in volume (or vice versa), intermittent zero flow entry for a major process with normal volume entry (or vice versa), and unit process flow values substantially higher than a plant's finished water flow for the same month. Unit process flow and volume data were employed to estimate chlorine contact times. Less than 1% of calculated chlorine contact times for unit processes downstream of chlorine addition were affected by flagged

flow or volume records. Reasons for flagging chlorine and ammonia dose and chlorine residual records included large discrepancies in doses across sampling months at a given plant (e.g., 50 mg/L Cl_2 at a plant normally dosing 5 mg/L), chlorine doses of zero reported with concurrent measurable downstream chlorine residual levels for processes where a plant normally applied chlorine, and anomalous chlorine residual patterns across treatment trains within a given plant/month. Overall, 2% of chlorine dose values and less than 1% of ammonia dose values and chlorine residual measurements were flagged (see Table 3.1).

3.4.2 ICR Data Summary

ICR data discussed in the remainder of this paper reflect the above-described review effort. Statistical analyses and data summaries are based on an analytical data set that excludes flagged records and incorporates replacement values as noted in Table 3.1.

3.4.2.1 Influent Water Quality

Influent water quality data for TOC, bromide, UV254, and TSUVA are summarized in Table 3.2. Side-by-side box and whisker plots in Figure 3.1 show distributions of TOC results after grouping data by source water type (Figure 3.1a: "GW" for ground water and "SW" for surface water) and type of secondary disinfectant (Figure 3.1b: "CL2" for free chlorine, "CLM" for combined chlorine). Surface water sources (N = 5429) had significantly higher TOC concentrations than ground water sources (N = 1941; p <0.0001). Fifty-five percent of influent sample TOC concentrations were below the MRL of 0.7 mg/L for ground waters, compared to less than 2% for surface waters. Conversely, only 21% of ground water TOC concentrations exceeded 2.0 mg/L, compared to 71% for surface waters. However, the distribution of influent TOC data for ground waters was skewed such that extreme values

(>10 mg/L) were actually more prevalent in ground waters (5.4%) than in surface waters (1.3%). It is noteworthy that these very high ICR ground water TOC sources were all in the State of Florida.

Table 3.2 Summary of Influent Water Quality

		TOC (mg/L as C)			Bromide (mg/L)			UV254 (cm^{-1})			TSUVA (L/mg m)		
Category		*N*	Mean	Median	*N*	Mean	Median	*N*	Mean	Median	*N*	Mean	Median
All Categories		7,522	2.82	2.35	7,981	0.072	0.035	7,536	0.088	0.060	6,873	2.91	2.54
Source[a]													
GW		1,941	1.83	0.35	2,137	0.103	0.065	1,915	0.061	0.012	1,759	2.74	1.75
SW		5,429	3.18	2.65	5,772	0.060	0.030	5,472	0.097	0.073	4,980	2.97	2.63
Residual[b]													
CL2		5,146	2.04	1.85	5,479	0.051	0.029	5,136	0.060	0.041	4,681	2.81	2.33
CLM		2,375	4.53	3.60	2,501	0.117	0.073	2,399	0.147	0.102	2,191	3.11	2.83
Source	Residual												
GW	CL2	1,644	0.81	0.35	1,806	0.086	0.050	1,612	0.022	0.005	1,486	2.63	1.29
	CLM	297	7.47	8.00	331	0.195	0.140	303	0.272	0.321	273	3.38	3.42
SW	CL2	3,414	2.62	2.35	3,638	0.034	0.021	3,439	0.078	0.056	3,112	2.90	2.54
	CLM	2,014	4.15	3.50	2,133	0.105	0.061	2,032	0.130	0.098	1,867	3.09	2.79
Residual	Pre-Cl_2[c]												
CL2	Yes	3,524	1.63	1.50	3,802	0.056	0.030	3,564	0.048	0.032	3,245	2.74	2.13
	No	1,517	3.01	2.70	1,624	0.040	0.027	1,472	0.091	0.074	1,338	2.99	2.65
CLM	Yes	1,558	3.56	3.18	1,655	0.115	0.071	1,584	0.112	0.088	1,441	3.04	2.70
	No	775	6.54	4.95	817	0.121	0.076	781	0.221	0.142	716	3.30	3.08

[a]Data with source type "MIX" or "PUR" (purchased finished water) not included.
[b]1 plant/month could not be classified.
[c]148 plant/months could not be classified.

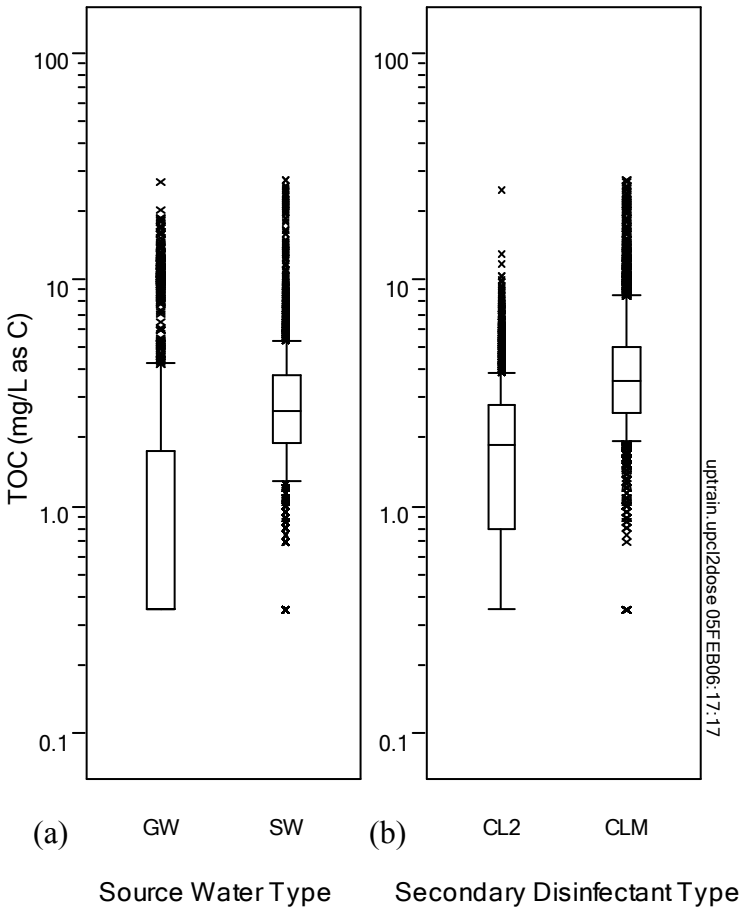


Figure 3.1 Box and whisker plots of influent TOC concentrations subset by major source water type (a) and type of secondary disinfectant (b)

ICR data indicate that influent TOC levels strongly impacted disinfection practices. As illustrated by Figure 3.1b, TOC levels were significantly lower at plants using free chlorine for secondary disinfection (N = 5146; median TOC = 1.85 mg/L) than at plants using chloramines (N = 2375; median TOC = 3.60 mg/L) (p <0.0001). Chloramine use was uncommon with influent TOC concentrations below 2.0 mg/L (only 252 out of 3,025 plant/months) but almost universal where influent TOC concentrations exceeded 10 mg/L (170 out of 176 plant/months). Virtually all ground water plants with even moderate TOC concentrations used chloramines for secondary disinfection. For ground water plants using

free chlorine, influent TOC concentrations were below 2 mg/L in more than 90% of plant/months whereas the median TOC concentration for ground water plants using chloramines was 8 mg/L. Overall, the median ICR plant influent TOC level was 2.35 mg/L and chloramines were used in 32% of ICR plant/months (see Table 3.2).

The practice of prechlorination was also associated with influent waters having lower TOC levels. Distributions of influent TOC concentration for prechlorinating and nonprechlorinating plants are compared in Figure 3.2 which plots data separately for all plants (Figure 3.2a), plants using free chlorine as a secondary disinfectant (Figure 3.2b), and plants using combined chlorine as a secondary disinfectant (Figure 3.2c). Influent TOC levels were significantly lower at plants practicing prechlorination than at plants applying the first chlorine dose downstream of a clarification process. This was true for the data set overall, for free chlorine and combined chlorine plants considered separately, and for ground and surface water plants considered separately (p <0.0001 in all cases) (see Table 3.2). Overall, the median influent TOC concentration was 2.0 mg/L for prechlorinating plants, compared to 3.2 mg/L for nonprechlorinating plants. Prechlorination was practiced in 69% of ICR plant/months overall, but only in 22% of plant/months with very high TOC concentrations (>10 mg/L). However, all of these prechlorinating high-TOC plants used chloramines for secondary disinfection and almost half also used either chlorine dioxide or ozone in the treatment plant (compared to <5% of nonprechlorinating high-TOC plants).

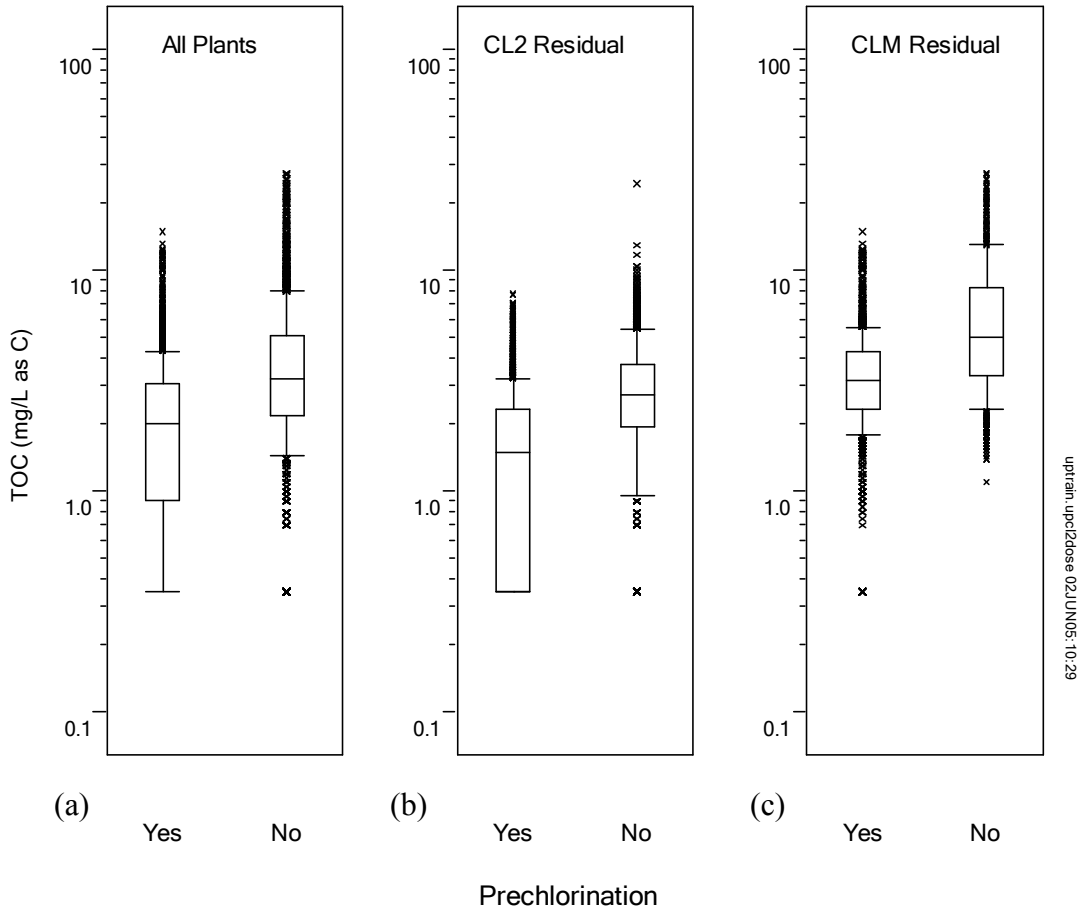


Figure 3.2 Box and whisker plots of influent TOC concentrations for all plants (a), plants using free chlorine for secondary disinfection (b) and plants using combined chlorine for secondary disinfection (c), with data in each group subset by prechlorination practice

As an indicator of organic DBP precursor concentration, TOC is a key factor influencing disinfection practices. Effective DBP control requires limiting free chlorine contact with DBP precursors, and thus regulations include mandated precursor removal levels in the treatment plant. The data in Figures 3.1 and 3.2 show that, in order to limit DBP formation, plants with higher TOC concentrations tend to use combined chlorine as a secondary disinfectant and/or must practice chlorination downstream of clarification, after appreciable TOC has been removed. In an earlier analysis of ICR data addressing the

prevalence of alternative disinfectant use in relation to both source water TOC and bromide levels, McGuire and Hotaling (2002) demonstrated some trends similar to those observed here. Chen and Regli (2002) also noted the general dependence of ICR disinfectant use on source water quality.

Figure 3.3 illustrates distributions of influent bromide concentration data according to source water type (Figure 3.3a) and secondary disinfectant type (Figure 3.3b). Ground waters had significantly higher bromide concentrations than surface waters (p <0.0001), with median values of 0.065 and 0.030 mg/L for ground and surface waters, respectively (see Table 3.2). Bromide concentrations exceeding 0.1 mg/L were twice as prevalent in ground waters (33%) than in surface waters (15%), though the highest bromide concentration, 2.23 mg/L, was observed in a surface water impoundment. Higher influent bromide levels were also seen at plants using chloramines for residual disinfection (median 0.073 mg/L) compared to plants using free chlorine (median 0.029 mg/L), and this difference was statistically significant for the data set overall and for both ground and surface water plants considered separately (p <0.0001 in all cases). Thus, the data indicate that, in addition to organic DBP precursor level, source water bromide level influences secondary disinfection practices. This was also observed by McGuire and Hotaling (2002). Like TOC, bromide is a key DBP precursor, so contact of free chlorine with waters containing high concentrations of bromide in the presence of organic precursors must be limited in order to maintain regulatory compliance. Hence, waters with elevated bromide concentrations are more likely to use combined chlorine as a secondary disinfectant, as evidenced by Figure 3.3b. It should be noted that, although bromide and TOC data have opposing tendencies with respect to source water type, the two parameters are uncorrelated (Spearman correlation coefficient = 0.148).

Thus, their influences on disinfection practices can be considered to be independent and additive.

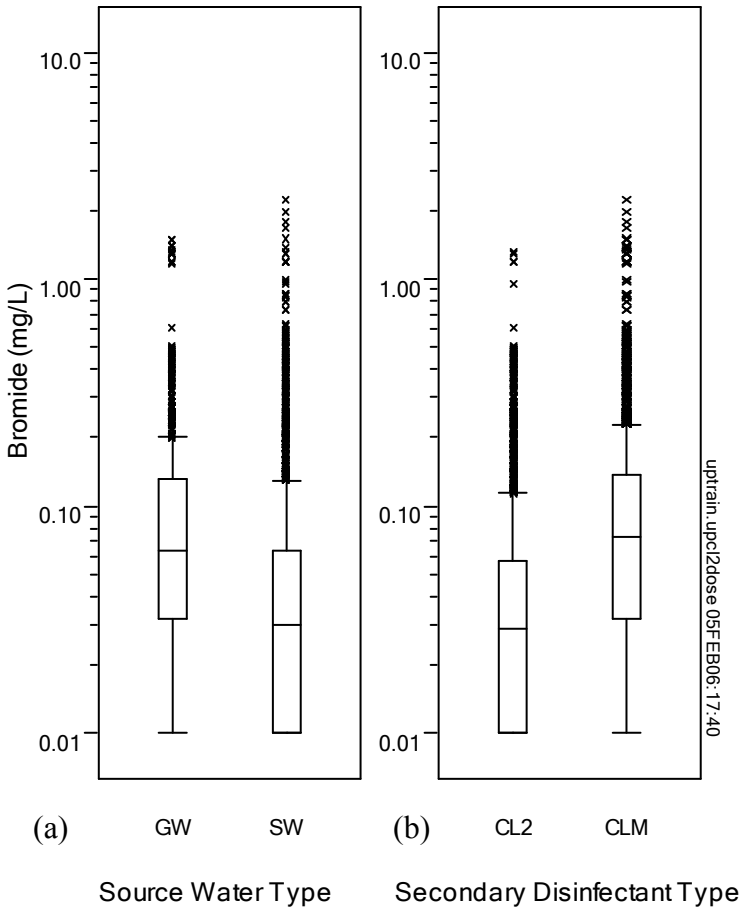


Figure 3.3 Box and whisker plots of influent bromide concentrations subset by major source water type (a) and type of secondary disinfectant (b)

Unlike the pattern observed for TOC, there was little difference in bromide levels on the basis of prechlorination practice (see Table 3.2). The difference was marginally significant for ground water plants ($p = 0.027$) and for plants using chloramines ($p = 0.010$), with lower bromide concentrations at prechlorinating plants in each of these subgroups, but not significant for plants using surface water or plants using free chlorine. This makes sense

because clarification processes, which are capable of removing appreciable amounts of TOC, generally have no effect on bromide concentrations, and thus high bromide levels alone would not be expected to influence prechlorination practices.

As illustrated by Figure 3.4, higher bromide concentrations coupled with lower TOC concentrations translated to significantly higher Br/TOC ratios for ground waters compared to surface waters (p <0.0001). The median Br/TOC ratio for ground waters (0.085 mg/mg) was almost an order of magnitude higher than that for surface waters (0.011 mg/mg). This is important from the standpoint of formation and occurrence of individual brominated DBP compounds because the Br/TOC ratio strongly influences the extent of bromine substitution in DBPs (Symons et al. 1993, 1996; Shukairy and Summers 1996; Obolensky and Singer 2005). Although ground waters generally have lower TOC concentrations than surface waters, they may in some cases yield higher concentrations of certain brominated DBP compounds due to their much higher Br/TOC ratios. Br/Cl_2 ratios (not shown) exhibited patterns similar to those found for Br/TOC except that the contrast between ground and surface water data was less pronounced. The median Br/Cl_2 ratio for ground waters was 0.030 mg/mg, compared to 0.013 mg/mg for surface waters.

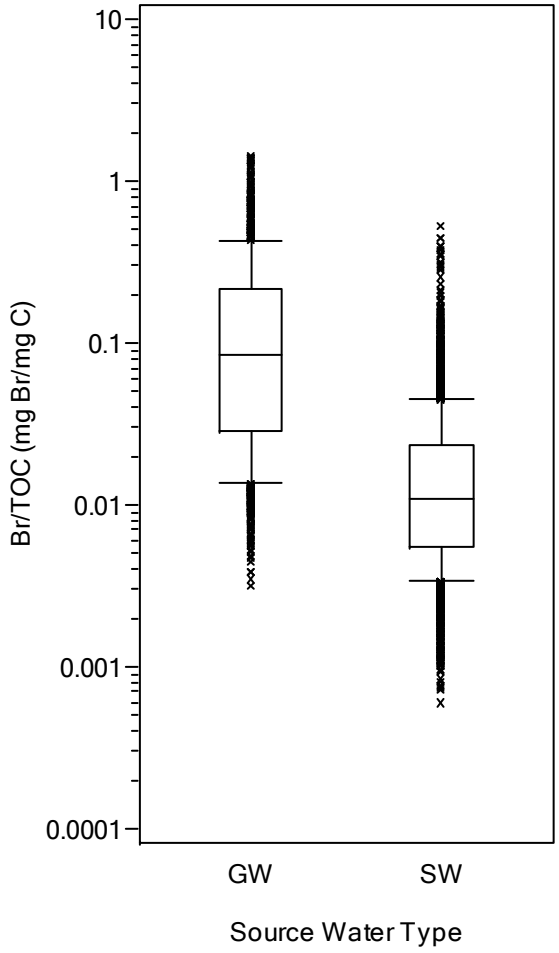


Figure 3.4 Box and whisker plots of influent bromide/TOC ratio, subset by major source water type

Patterns of influent UV254 data were very similar to those observed for TOC with respect to source water type and disinfection practices (see Table 3.2). This is not surprising considering that UV254 absorbance, like TOC concentration, is a good indicator, perhaps better, of DBP formation potential (Croue et al. 1999; Liang and Singer 2003). UV254 absorbance values for ground waters, free chlorine plants, and prechlorinating plants were substantially lower than for surface waters, combined chlorine plants, and nonprechlorinating plants, respectively, with all differences statistically significant at the 0.001 level.

Influent TSUVA values differed to a much lesser extent with respect to source water type and disinfection practices than did corresponding absolute TOC and UV254 values (see Table 3.2). There was no statistically significant difference between influent TSUVA levels for ground and surface waters. Slightly lower TSUVA levels at free chlorine plants, compared to chloramine plants, were statistically significant at the 0.0001 level. Differences with respect to prechlorination practice were statistically significant for free chlorine plants, with slightly lower TSUVA levels at prechlorinating plants compared to nonprechlorinating plants (p <0.0001), but not for plants using postchloramination (see Table 3.2). Thus, the data indicate that, on a nationwide basis, NOM characteristics, as measured by TSUVA, are less of a driver for disinfection choices than the absolute TOC concentration or UV254 absorbance.

3.4.2.2 Treatment Information

Chlorine and ammonia doses were analyzed from several standpoints, including number of chlorine addition points, pH at the first point of chlorine addition, total applied chlorine and ammonia doses, Cl_2/TOC ratio, Br/Cl_2 ratio, and Cl_2/N ratio. Free and combined chlorine contact times and residual concentrations were examined up to the point of entry to the distribution system (finished water). Chlorine dose data were available for 7236 plant/months, encompassing 422 ICR treatment plants. Plants without chlorine dose information usually had no process train information and were of the type categorized in Aux 1 as disinfected ground water ("DIS/GW"), unfiltered surface water ("UNFILT/SW"), or disinfected purchased finished water ("DIS/WHSALE"). These plants, including almost half of the ICR plants served by ground water (57 out of 132), are not represented in any analyses addressing disinfectant dose or other process train data.

Total applied chlorine dose results are summarized in Table 3.3. The overall median total chlorine dose was 3.6 mg/L, with a dose of 7.5 mg/L or less employed in 90% of ICR plant/months. Up to four points of chlorine addition were used at ICR treatment plants, one or two locations being the most common pattern. Overall, the fractions of plant/months having one, two, three, or four applied chlorine locations were 36, 52, 11, and 1%, respectively, and there was no meaningful difference between these frequencies on the basis of secondary disinfectant type. Total chlorine dose increased significantly with the number of dosing points in plants using free chlorine as a secondary disinfectant but not in plants using combined chlorine for secondary disinfection.

Table 3.3 Summary of ICR Plant/Month Chlorine Doses by Disinfection Practices

Category	*N* plants [a]	Total chlorine dose (mg/L as Cl_2)			[Total Cl_2 dose]/[TOC] ($mg\ Cl_2$)/(mg C)			Cl_2/NH_3 ($mg\ Cl_2$/mg N)[b]		
		N	Mean	Median	*N*	Mean	Median	*N*	Mean	Median
All categories	422	7,236	4.2	3.6	6,259	2.13	1.54	2,321	5.96	4.80
Overall disinfection type										
CLM	30	379	5.7	5.9	268	1.55	1.27	236	8.00	5.46
CLO2/CLM	9	115	6.0	6.0	93	1.61	1.65	114	4.43	4.68
O3/CLM	7	97	3.9	3.9	92	1.14	1.14	96	5.19	4.83
Cl2/ClM	124	1,919	5.5	5.0	1,681	1.95	1.60	1,731	5.69	4.66
CLO2/CL2/CLM	4	54	4.5	4.3	54	1.68	1.71	50	6.54	6.50
O3/CL2/CLM	7	96	6.7	4.6	85	1.41	1.35	94	8.02	6.34
CL2	283	4,211	3.5	2.8	3,684	2.37	1.60			
CLO2/CL2	16	232	3.6	3.1	198	1.54	1.35			
O3/CL2	10	133	2.7	2.5	104	1.28	1.14			
Secondary disinfection										
CL2	298	4,576	3.5	2.8	3,986	2.30	1.55			
CLM	162	2,660	5.5	5.0	2,273	1.83	1.52			

[a]Total number of plants (overall or for CL2 and CLM groups) is less than sum across disinfection types because plants that varied practices are represented in more than one category.
[b]Chlorine value in calculation is the sum of the existing chlorine residual and concurrent chlorine dose at point of ammonia addition.

Generalizations about differences in total chlorine dose with respect to overall disinfection category are constrained by the sparseness of data for many of the categories (see Table 3.3). Five of the nine disinfection categories are represented by ten or fewer treatment plants and most of the data fall into just two categories ("CL2" and "CL2/CLM"). Nevertheless, certain comparisons of interest can be made by considering the nine disinfection categories in three groups: no free chlorine used (chloramines either alone or in combination with chlorine dioxide or ozone); both free chlorine and chloramines used (with or without the conjunctive use of chlorine dioxide or ozone); and no chloramines used (free chlorine alone or in combination with chlorine dioxide or ozone). Considering plant/months with no free chlorine contact (first three categories in Table 3.3), chlorine dose was significantly lower (p <0.01) when ozone was employed in conjunction with chloramines ("O3/CLM") than when chloramines were used alone ("CLM") or in combination with chlorine dioxide ("CLO2/CLM"). Median total chlorine doses for these three categories were 3.9, 5.9, and 6.0 mg/L, respectively (see Table 3.3). In contrast, considering plant/months using free chlorine in conjunction with chloramines, there was no significant difference in chlorine dose based on additional disinfectant use ("CL2/CLM," "CLO2/CL2/CLM," or "O3/CL2/CLM"). The same was true for plant/months with no chloramine use ("CL2," "CLO2/CL2," or "O3/CL2"). Unlike ozone plants that also utilized free chlorine, ozone plants that used only combined chlorine (i.e., "O3/CLM") were probably using ozone for primary disinfection as part of their DBP control strategy, satisfying a substantial portion of the organic oxidant demand through ozone contact, and lessening the total chlorine dose required to achieve their target residual concentrations. Plants using both ozone and free chlorine were most likely using ozone for alternate purposes such as preoxidation or taste and

odor control, without employing high enough doses to significantly affect organic oxidant demand.

Overall, plants using chloramines applied significantly higher chlorine doses than plants without any ammonia addition (p <0.0001). The median and 90th percentile total chlorine doses for chloramine plants were 5.0 and 9.1 mg/L, respectively. Corresponding values for free chlorine plants were 2.8 and 6.0 mg/L (see under "Secondary disinfection" heading in Table 3.3). Higher doses at chloramine plants are in accord with their higher TOC concentrations, as discussed earlier in connection with Figure 3.1, and their consequent higher chlorine demands, necessitating the use of chloramines as a measure to limit DBP formation.

Cl_2/TOC ratios were much less variable than absolute chlorine doses, with median Cl_2/TOC values for the nine disinfection categories ranging from 1.14 to 1.71 mg/mg (see Table 3.3). The coefficient of variation among these nine median values was 16%, compared to 30% for the absolute dose values. This stems from the fact that total chlorine dose and TOC at the first point of chlorine addition were correlated (Spearman rank correlation coefficient 0.576), an indication that chlorine dose is driven in large part by the concentration of organic material. When normalized to TOC, chlorine doses (i.e., Cl_2/TOC ratio) at free chlorine plants were actually higher than at chloramine plants (p <0.0001), although this difference was small in magnitude (see Table 3.3). Among chloramine plants, there were no significant differences in Cl_2/TOC ratio based on additional oxidant use (i.e., "CLM" versus "O3/CLM" or "CLO2/CLM," "CL2/CLM" versus "O3/CL2/CLM" or "CLO2/CL2/CLM"). However, significantly lower Cl_2/TOC ratios were applied at plants using chlorine dioxide (p

= 0.0002) or ozone (p <0.0001) in conjunction with free chlorine, compared to plants using free chlorine alone.

Very high Cl_2/TOC ratios (5 mg/mg or more) were utilized in a small fraction of plant/months (4%). These were almost exclusively at plants with very low TOC concentrations (75% <1 mg/L). About half of these data represented disinfected ground water ("DIS/GW") and "OTHER/GW" plants. These high Cl_2/TOC ratios are consistent with the use of chlorine for nondisinfection purposes, such as iron and manganese removal, as indicated by unit process information for many of these plants. The high Cl_2/TOC ratios at plants with low TOC concentrations account, in part, for the significantly higher (p <0.0001) Cl_2/TOC ratios observed at ground water plants (compared to surface water plants), at free chlorine pants (compared to chloramine plants), and at prechlorinating plants (compared to nonprechlorinating plants).

Table 3.3 also provides statistics for Cl_2/N ratio at the point of ammonia addition for plants using chloramines, grouped by disinfection category. For the data set overall, the median Cl_2/N value was 4.80 mg/mg, corresponding closely to a 1:1 molar Cl_2/N ratio. However, Cl_2/N ratios were highly variable, with an overall coefficient of variation of 71% (within-category coefficients of variation ranged from 18 to 83%). Use of free chlorine in conjunction with chloramines ("CL2/CLM") was the predominant method of disinfection for plants employing chloramines, accounting for three quarters of all plant/months (124 plants). The next most prevalent chloramine disinfection category was chloramines used alone ("CLM," 30 plants). The latter plants had significantly higher Cl_2/N ratios than "CL2/CLM" plants (p <0.0001), with a median value of 5.46 mg/mg. Due to the small numbers of plants involved, there is no statistical significance to the apparently higher Cl_2/N ratios for

CLO2/CL2/CLM and O3/CL2/CLM plants, compared to other disinfection categories. There were a substantial number of plant/months with extreme Cl_2/N ratios. About 10% of the results showed Cl_2/N molar ratios in excess of 2:1. Extreme values may stem, in part, from inaccuracies in ammonia dose, chlorine dose, or chlorine residual values, with the further possibility of compounded errors. Chen and Regli (2002) also noted the substantial variation in Cl_2/N ratios and suggested it might reflect the challenge ICR plants faced in maintaining appropriate ratios in response to fluctuating water quality.

Aside from the minor categories of slow sand filtration and membrane filtration which were employed at a limited number of ICR plants, pH conditions were very similar among the five nonsoftening treatment categories (unfiltered surface water, disinfected and "other" groundwater, conventional, and direct or in-line filtration), with median chlorination pH values between 7.4 and 7.6 and interquartile ranges all within pH 6.8–8.0. Significantly higher pH values at softening plants compared to all other treatment categories (p <0.0001) reflects use of lime or caustic soda for precipitative softening processes. The median pH value at the first point of chlorination for softening plants was 9.1, with values ranging as high as 12.1, and fewer than 1% of plant/months below pH 7.0. For the purposes of this discussion, softening plants include all 5 ICR softening treatment categories: single- and two-stage softening, split- and complex parallel treatment softening; and coagulation-sedimentation softening.

Table 3.4 summarizes unit process HRT and chlorine contact results. Considering major processes and referring to median HRT, chlorine contact time was shortest in rapid mix basins (1 min), followed by filters (27 min), flocculation basins (50 min), and sedimentation basins and clearwells (259 and 275 min, respectively). Overall, 60% of rapid

mix processes had upstream chlorine addition, compared to 62% for flocculation basins, 67% for sedimentation basins, 88% for filtration processes, and 98% for clearwells. For chloramine plants, the fractions of unit processes having upstream ammonia addition were 29% for rapid mix basins, 35% for flocculation basins, 41% for sedimentation basins, 43% for filtration processes, and 80% for clearwells. Accordingly, almost one-third of plants using chloramines added ammonia at the head of the plant, another 14% added ammonia before the filters, and another 37% added ammonia after the filters, prior to the clearwell. The remaining 20% of plants added ammonia after the clearwell, at the point of entry to the distribution system.

Table 3.4 Summary of Residual Chlorine Data for Processes with Free or Combined Chlorine Contact

Process type[a]	*N* plants	Median HRT (min.)	*N* process/ months	*N* process/ months with chlorine contact	Free Cl_2 contact			Combined Cl_2 contact		
					N[b]	Median free Cl_2 (mg/L as Cl_2)	% <0.1	*N*[b]	Median total Cl_2 (mg/L as Cl_2)	% <0.1
PRE	49	303.4	801	151	129	0.40	17.05	—	—	—
RAP	311	1.0	6,022	3,613	2,583	1.10	3.52	649	3.20	0.89
FLC	291	50.1	5,512	3,424	2,283	0.90	3.37	774	3.00	1.12
SED	285	259.1	5,694	3,771	2,585	0.60	7.66	902	2.80	1.22
FIL	376	27.0	6,499	5,718	4,322	0.80	9.35	1,078	2.80	1.08
CLR	368	274.8	6,716	6,576	4,054	1.20	1.23	1,999	2.70	0.05
DCB	46	75.5	8,47	808	506	1.20	7.31	140	3.10	0.00
SCC	61	147.1	1,097	570	445	0.40	11.24	93	3.00	0.00
RCB	23	6.3	446	160	43	0.20	6.98	78	2.30	0.00
OCB	22	9.1	1,778	339	285	0.10	37.54	—	—	—
GAC	10	21.7	174	72	72	0.05	68.06	—	—	—

[a]PRE = presedimentation , RAP = rapid mix, FLC= flocculation basin, SED = sedimentation, FIL = filtration, CLR = clearwell, DCB = disinfection contact basin, SCC = solids contact clarifier, RCB = recarbonation basin, OCB = ozone chamber, GAC = granular activated carbon.
[b]*N* = number of process/months with chlorine residual data available.

As shown in Table 3.4, process effluent chlorine residuals were considerably lower and more variable for processes with free chlorine contact than for combined chlorine processes. Among major process types (RAP, FLC, SED, FIL, CLR, see Table 3.4), median total chlorine concentrations for combined chlorine processes ranged from 2.7 to 3.2 mg/L, while median free chlorine concentrations for free chlorine processes ranged from 0.6 to1.2 mg/L. For free chlorine processes, differences among residual free chlorine concentration levels by process type were all statistically significant (p <0.0001). For chloramine processes, none of the process types was distinct in this respect. For major process types, below-reportable chlorine residual concentrations (<0.1 mg/L) were much more common for free chlorine processes (1−9%) than for combined chlorine processes (<2%). These patterns are in accord with the greater stability of combined chlorine and the higher chlorine doses applied at plants utilizing chloramines (see Table 3.3).

The relative uniformity of process effluent chlorine concentrations reflects the fact that chlorine doses are designed to achieve target residual concentrations. Given this relative uniformity and the large variation in HRTs across process types, when $C_{eff}t$ values (i.e., the product of HRT and process effluent residual chlorine concentration) are computed for purposes of quantifying disinfection potential, the effect of HRT dwarfs that of chlorine residual so that $C_{eff}t$ patterns are very similar to those for HRT with respect to process type. Thus, the largest $C_{eff}t$ values were generally found for sedimentation basins and clearwells.

3.5 CONCLUSIONS

Results of the data review efforts conducted for this study demonstrate that the ICR data collection program succeeded in achieving a high level of data quality for parameters

related to DBP formation. For all parameters examined, 2% or fewer records were identified as potentially faulty. By filling in significant numbers of missing entries for categorical descriptors of source water and disinfection types in this research, a more complete database was available for further analysis compared to previous efforts. In keeping with a transparent, objective, and systematic process for database analysis, it was important that this review be undertaken prior to use of the data set for statistical analyses of interrelationships in the data, and that data-handling and analysis methodologies be properly documented.

Analysis of ICR water quality and treatment data revealed several dominant patterns. Driven by DBP control considerations, plants with higher precursor concentrations tended to use chloramines and to avoid prechlorination. Overall, chloramines were used in 32% of ICR plant/months and prechlorination was practiced in 69% of plant/months. Nationwide, absolute TOC concentration and UV254 absorbance were more important drivers for treatment choices than the ratio of the two parameters (TSUVA), which reflects NOM characteristics. Likewise, plants with elevated levels of bromide in their raw water tended to use chloramines. The influences of bromide and TOC concentrations on disinfection practices appear to be independent of one another. ICR data indicated that chlorine doses were generally set to achieve a target residual at unit process effluents, typically about 1 mg/L for free chlorine processes and 3 mg/L for combined chlorine processes. The ratio between applied chlorine dose and TOC fell within a relatively narrow range, regardless of disinfection scheme, with an overall median ratio of 1.5 mg Cl_2 per mg C. A subset of plants with very low TOC concentrations using relatively high chlorine doses for nondisinfection purposes comprised an exception. Significantly higher chlorine doses were employed at chloramine plants compared to free chlorine plants, but this distinction largely disappeared

when chlorine doses were normalized to TOC. Although ratios of chlorine to ammonia-nitrogen at chloramine plants varied widely, the overall median value of 4.8 mg Cl_2/mg N was very close to the theoretical 1:1 molar ratio required for monochloramine formation.

The sampling distributions for important water quality and treatment parameters presented in this paper, based on a screened and more complete ICR database than utilized previously (McGuire, McLain, and Obolensky 2002), provides an overview of comprehensive industry-wide statistics for water quality and treatment in the 1997–1998 time frame. This should be of general interest to other investigators and provides a basis for further research on important drinking water quality and treatment issues, supporting development of a better understanding of DBP formation and control.

CHAPTER 4: DEVELOPMENT AND INTERPRETATION OF MODELS TO DESCRIBE THE IMPACTS OF WATER QUALITY AND TREATMENT ON DBP FORMATION USING THE ICR DATABASE

4.1 BACKGROUND AND INTRODUCTION

Multiple linear regression models (regression models) were used to study relationships between water quality, treatment processes, and DBP formation employing full-scale water treatment plant data from the Information Collection Rule (ICR). This type of model provides an efficient statistical approach to isolating and quantifying interrelated effects on DBP formation from simultaneously varying factors in a single data set, and lends itself to straightforward interpretation. A long history exists of developing regression models based on data from controlled laboratory experiments. Chowdhury and Amy (1999) review these and other DBP modeling efforts. The best known regression models were developed by Amy and coworkers using data from bench chlorination experiments (Amy, Chowdhury, and Chadik 1987; Amy et al. 1998). These models became the basis for the U.S. EPA's Water Treatment Plant Simulation Program (EPA Model, Harrington, Chowdhury, and Owen 1992), used extensively to support regulatory development through forecasting DBP production under specified water quality and treatment conditions (Solarik et al. 2000; Swanson et al. 2002). Little work has been done developing this type of model from observational field data (Moore, Tuthill, and Polakoff 1979; Otson, Williams, and Bothwell

1981; Golfinpoulos et al. 1998) although, at least conceptually, it provides an excellent analytical tool for studying such data. Limitations include the need for a large data set that includes key water quality and treatment variables with consistent information structure. Availability of the ICR database (U.S. EPA 2000b), containing DBP data and associated water quality and treatment information from a comprehensive survey of large U.S water utilities, provided a unique opportunity to apply regression modeling techniques to the analysis of real water treatment plant data. Descriptive aspects of the ICR program and a number of data analyses are provided by McGuire, McLain, and Obolensky (2002). Obolensky, Singer, and Shukairy (2007) summarize specific data used in this research and outline associated data handling methodologies.

Models developed for the present research have a mathematical structure similar to that used for the EPA Model. However, the focus on using regression models as an investigative tool to examine complex relationships between water quality, treatment, and DBP occurrence, as is done in this research, differs fundamentally from previous modeling efforts aimed at predicting DBP concentrations. Although forecasting per se was not a goal of the current work, the models developed herein needed to perform adequately in a predictive capacity so that large error terms would not preclude detecting and comparing effects of interest. The size and structure of the ICR database was considered ample for developing models with sufficient performance to support the research goals. Like previous DBP regression models, the current models are developed empirically based on a rational framework and do not derive from kinetic rate expressions describing elemental chemical reaction steps. The molecular-level mechanistic understanding of DBP formation needed for such deterministic models does not currently exist, although some attempts have been made

to approach modeling from this perspective (Kavanaugh et al. 1980; Adin et al. 1981; Clark 1998; Gang et al. 2002; Sohn et al. 2004). Reaction pathways for DBP formation involve complex arrays of halogen addition, substitution, oxidation, hydrolysis, and decarboxylation steps in parallel and in series. Although studies with model compounds have shed much insight into these processes (Boyce and Hornig, 1983; Reckhow and Singer 1985), the varied and changing chemical nature of organic DBP precursors in natural waters have precluded identification or measurement of specific substrates needed to develop deterministic models. Acknowledging practical limitations to describing DBP formation at a fundamental level, the rational empirical modeling framework encompasses current knowledge of basic drivers for DBP formation in drinking water and includes the variables characterizing water quality and treatment processes that are easily obtained and commonly monitored. Accordingly, these empirical models can be used to gain a better understanding of how DBPs are related to water quality and process parameters in real, dynamic treatment systems.

4.2 SCOPE OF RESEARCH

ICR data were used to develop models for DBP concentrations in finished water from treatment plants using only free chlorine for disinfection prior to distribution (chlorine plants). Separate models were developed for twelve individual DBP compounds, total organic halogen (TOX), THM4 (sum of 4 trihalomethanes), X_2AA (sum of three dihaloacetic acids), X_3AA (sum of 4 trihaloacetic acids), and HAA9 (sum of 9 haloacetic acids) (see Table 4.1). The twelve DBP species modeled were chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$), bromoform ($CHBr_3$), dichloraoacetic acid (Cl_2AA), bromochloroacetic acid (BrClAA), dibromoacetic acid

(Br_2AA), trichloroacetic acid (Cl_3AA), bromodichloroacetic acid ($BrCl_2AA$), dibromochloroacetic acid (Br_2ClAA), tribromoacetic acid (Br_3AA), and chloral hydrate (trichloroacetaldehyde hydrate, Cl_3AH). DBP species and class sum variables were modeled in molar concentration units to facilitate interpretation. TOX was modeled in reported weight-based chlorine-equivalent concentration units. Although HAA9 models implicitly account for monochloroacetic acid and monobromoacetic acid occurrence, no attempt was made to model the individual monohaloacetic acid species due to their consistently low concentrations in the database and the relatively large uncertainties associated with their analysis (Weinberg 2000; Fair et al. 2002; Domino et al. 2004).

Table 4.1 Dependent variables: summary of data

Variable	DBP Class	*N*	Min [c]	Max	Mean	Median	90th %ile	CV %	MRL	% < MRL
$CHCl_3$ (µg/L)	THM	1118	0.5	180	22	17	45	84	1.0	0.5
$CHBrCl_2$ (µg/L)	THM	1131	0.5	48	7.7	6.0	16	85	1.0	6.6
$CHBr_2Cl$ (µg/L)	THM	1130	0.5	34	3.2	1.5	8.6	138	1.0	39.2
$CHBr_3$ (µg/L)	THM	1130	0.5	19	0.8	0.5	1.6	141	1.0	84.7
Cl_2AA (µg/L)	X_2AA	1096	0.5	72	12	9.9	22	75	1.0	1.8
BrClAA (µg/L)	X_2AA	1092	0.5	24	2.8	2.1	6.0	93	1.0	21.3
Br_2AA (µg/L)	X_2AA	1100	0.5	14	1.0	0.5	2.3	116	1.0	75.0
Cl_3AA (µg/L)	X_3AA	1083	0.5	74	11	8.4	21	86	1.0	5.8
$BrCl_2AA$ (µg/L) [a]	X_3AA	313	0.5	16	3.4	2.6	7.5	85	1.0	15.3
Br_2ClAA (µg/L) [a]	X_3AA	291	0.5	11	1.6	1.0	3.5	93	2.0	73.5
Br_3AA (µg/L) [a]	X_3AA	235	0.5	9.3	1.2	1.0	2.0	85	4.0	98.7
Cl_3AH (µg/L)	X_3AH	1012	0.5	46	3.8	2.7	8.2	97	0.5	19.3
TOX (µg/L as Cl)	n/a	1102	25	1010	147	122	280	69	50	9.8
THM4 (µg/L)	n/a	1111	2.0	214	33	28	64	71	n/a	n/a
X_2AA (µg/L)	n/a	1088	1.5	76	15	13	29	66	n/a	n/a
X_3AA (µg/L) [a]	n/a	234	2.0	62	15	13	27	66	n/a	n/a
HAA9 (µg/L) [a]	n/a	228	5.0	112	30	27	52	56	n/a	n/a
$BrCl_2AA$ (µg/L) [b]	X_3AA	998	0.4	22	3.5	2.8	7.1	84	n/a	n/a
Br_2ClAA (µg/L) [b]	X_3AA	975	0.5	11	1.4	1.0	2.6	80	n/a	n/a
Br_3AA (µg/L) [b]	X_3AA	920	0.5	9.3	1.1	1.1	1.2	47	n/a	n/a
X_3AA (µg/L) [b]	n/a	916	2.0	91	17	14	32	69	n/a	n/a
HAA9 (µg/L) [b]	n/a	844	5.0	150	34	29	63	61	n/a	n/a

a reported ICR analytical results
b including projected data for brominated X_3AA species
c minimum values reflect replacement of below MRL data with 0.5*MRL (further adjustment for Br_2ClAA and Br_3AA described in Appendix B)

Model independent variables, listed in Table 4.2, describe water quality and treatment characteristics of potential importance in determining finished water DBP levels. Models

were restricted to finished water DBP concentrations at chlorine plants in order to limit influences related to alternative disinfectant use and distribution system transit and transformations, thereby providing a base case for further study of these and other effects. ICR plants using only free chlorine tended to have lower DBP precursor levels (Obolensky, Singer, and Shukairy 2007) than those in which combined chlorine was used. This restriction affects the design matrix and thus, potentially, the ability of associated models to detect effects that might be discernible given a wider range of independent variable values. The data domain is a fundamental consideration in any modeling work and findings must be interpreted in this context.

Table 4.2 Independent variables: summary of data

Variable	*N*	Min	Max	Mean	Median	90^{th} %ile	CV %	MRL	% < MRL
Turbidity (NTU) [a]	1186	0.01	200	8.2	2.7	20	201	0.01	0
Bromide (mg/L) [a]	1150	0.01	0.96	0.03	0.02	0.07	130	0.02	45
Temperature (°C) [a]	1213	0.4	33.5	16.4	16	26	44	n/a	n/a
Alkalinity (mg/L as $CaCO_3$) [a]	1189	1.0	342	76	62	165	87	1	0
TOC (mg/L) [b]	1074	0.35	7.8	2.1	2.0	3.3	45	0.7	2.8
UV254 (cm^{-1}) [b]	1043	0.005	0.342	0.047	0.039	0.081	72	0.009	1.0
Cl_2 Consumed (mg/L as Cl_2) [c]	1095	0.09	11.8	1.98	1.6	3.5	75	n/a	n/a
Cl_2 Contact Time (h) [c]	1125	0.2	745	15.5	8.2	23	273	n/a	n/a
pH [d]	1213	6.1	10.6	7.8	7.7	8.8	9	n/a	n/a
Cl_2 Residual (mg/L as Cl_2) [d]	1190	0.3	5.5	1.6	1.4	2.7	49	0.1	0
Cl_2 Point [e]	1180	0	1	0.67	n/a	n/a	n/a	n/a	n/a

a plant influent
b at first point of chlorine addition in process train
c cumulative from plant influent
d finished water
e categorical variable representing first chlorine addition point at raw (1) or settled water (0)

4.3 METHODS AND PROCEDURES

4.3.1 Statistical Basis for Models

ICR data were collected by large U.S. utilities in 1997-1998 but the water quality and treatment domain encompassed can be presumed to be representative of the population of large utilities over time, more generally. Assumptions requiring validity for hypothesis testing with linear regression include independent random sampling and normally distributed errors with equal variance (Kleinbaum et al. 1998). Independence implies that individual observations are not influenced by one another such as might be the case for observations close in time or space, or related in some other way resulting in unaccounted for correlations. Random sampling implies lack of bias in the sampling method such that each member of the statistical population being sampled has an equal probability of being selected as an observation. Normally distributed errors with equal variance means that residual values (i.e. differences between model predictions and observed values) follow a gaussian distribution for any fixed combination of values for the independent variables (or equivalently, for a given predicted value) and that the variance of this distribution is homogeneous across the range of predicted values. Neither the dependent nor independent variables must be normally distributed to meet these assumptions. However, error terms are unlikely to be gaussian if the dependent variable distribution is skewed. A good spread in independent variable values facilitates detection of effects and helps increase parameter estimate precision (Chambers et al. 1983).

The assumption of independence should be valid because, although each ICR plant was sampled six times (six quarters in the 18-month monitoring period), several months

elapsed between sampling events at each of the plants. Quarterly DBP sampling was employed in the ICR to impose a random sampling design (and obtain seasonal representation) and there is no reason to suspect bias in the grab sampling method used for obtaining water samples, or in the manner that engineering and treatment data were reported. Although third and fourth calendar quarters are more heavily represented in the ICR data set than first and second calendar quarters, seasonal bias is presumed to be accounted for by inclusion of water quality and treatment information reflective of seasonal effects. Assumptions of normally distributed errors with equal variance were tested in the course of model development.

4.3.2 Design Matrix and Data Handling

ICR data were obtained from the U.S. EPA Auxiliary 1 Database Version 5.0 (Aux 1, U.S. EPA 2000b) and screened prior to use. Data screening involved recovery of missing classification data generated during Aux 1 production (e.g. descriptors of source water and disinfection types), flagging of questionable reported data, and corrections of obvious decimal place data entry errors, and is described elsewhere (Obolensky, Singer, and Shukairy 2007; Obolensky and Singer 2005). Flagged data were not used in modeling work reported here. Data handling and analysis were conducted in the SAS software environment (SAS Institute, Cary NC). To ensure meaningful association between finished water DBP concentrations and source water quality and treatment practices, data for plants treating purchased finished water or adding any additional blended source to the treatment train were excluded from the data set. Data for treatment plants classified in Aux 1 as *disinfected wholesale*, *disinfected groundwater*, *other groundwater*, or *other* were also excluded:

disinfected wholesale plants treated purchased finished water so treatment information is incomplete; *disinfected groundwater* plants generally had no treatment train (missing model input data); and the two other cited plant types utilized oxidants for non-disinfection purposes (e.g., iron or manganese control) which could obscure relationships of interest. There were 1,224 finished water DBP records meeting the stated criteria, representing 225 of the 500 ICR treatment plants and comprising 43% of Aux 1 finished water DBP records. Tables 4.1 and 4.2 summarize the data for dependent and independent model variables, respectively, where missing values account for N < 1,224.

4.3.3 Censored Data

ICR measurement results below the minimum reporting level (MRL), i.e. censored data, were replaced with one-half the analyte-specific MRL (see Tables 4.1 and 2). Exceptions were made for Br_2ClAA and Br_3AA because of their high MRLs (2.0 and 4.0 μg/L, respectively) and the belief that the half-MRL method would bias their occurrence data upwards. Instead, an approach based on intraclass halogen patterns was employed (see Appendix B). Tables 4.1 and 4.2 show the fractions of data affected by MRL censoring.

4.3.4 Brominated X_3AA Species Measurement Data and Estimation

All ICR plants were required to report data for 6 of the 9 chlorinated and brominated HAA species. Approximately one third of the plants optionally reported data for the remaining three brominated X_3AA species that make up HAA9 ($BrCl_2AA$, Br_2ClAA, and Br_3AA, see Table 4.1). To augment these reported data, missing $BrCl_2AA$, Br_2ClAA, and Br_3AA values were estimated using simple projection models that depended on reported THM species and Cl_3AA data according to Equations 4.1-4.3, where molar concentrations

are used. These models, based on the concept of parallel halogen speciation patterns in different DBP classes (Obolensky and Singer 2005), are modifications of the models of Roberts, Singer, and Obolensky (2002) by addition of slope and intercept fitting parameters which were assumed to be one and zero, respectively, in the previous work.

$$[BrCl_2AA] = C + k*[Cl_3AA]*[CHBrCl_2]/[CHCl_3] \quad (4.1)$$

$$[Br_2ClAA] = C + k*[\ Cl_3AA]*[CHBr_2Cl]/[\ CHCl_3] \quad (4.2)$$

$$[Br_3AA] = C + k*[\ Cl_3AA]*[CHBr_3]/[\ CHCl_3] \quad (4.3)$$

Slope, k, and intercept, C, parameters were obtained by least-squares regression using data for all ICR samples except those from plants treating purchased finished water or adding an additional blended source to the treatment train. Accordingly, more than 2,500 records from approximately 140 treatment plants were used to calibrate Equations 4.1-4.3. The fitted equations were used to estimate concentrations of $BrCl_2AA$, Br_2ClAA, and Br_3AA for samples without reported results and thereby increase the available data for subsequent multiple linear regression modeling of these species and the aggregate DBP measures that include them (i.e., X_3AA and HAA9).

4.3.5 Model Independent Variables

Independent variables used for model building, listed in Table 4.2, included water quality descriptors (turbidity, bromide, temperature, alkalinity, total organic carbon (TOC), ultraviolet absorbance at 254 nm (UV254), pH, and total chlorine residual), chlorine consumed, free chlorine contact time, and a coded variable to describe the location of first

chlorine addition in the treatment train (1 = raw water, 0 = settled/filtered water). Plant influent sample results were used for turbidity, bromide, temperature, and alkalinity. Results at the first point of chlorine addition in the treatment train were used for TOC, and UV254. Chlorine consumed was calculated as the difference between total applied chlorine dose through the plant and total residual chlorine in the finished water. Free chlorine contact time was calculated as the sum of unit process hydraulic residence times from the first point of chlorine addition. Finished water sample results were used for pH and chlorine residual. Extraction of these data from the Aux 1 database and relevant computations are discussed elsewhere (Obolensky, Singer, and Shukairy 2007).

Adjustment of chlorine consumed to account for ammonia demand was rejected after initial consideration because it resulted in poorer model results, perhaps due to inaccuracy in reported influent ammonia data. The decision to use finished water pH values was based on an analysis of lime and caustic soda addition in ICR plants and a determination that the finished water location would, overall, provide the best single representation of pH conditions during chlorine contact. Other independent variables initially considered for inclusion in the models were found to be redundant with those listed in Table 4.2 because of collinearity or model construction. These included calcium hardness, total hardness, chlorine dose, and the ratio variables UV254/TOC, bromide/TOC, and bromide/Cl_2. Calcium hardness and total hardness were linearly dependent with alkalinity ($r > 0.9$). Alkalinity was retained as a variable because it is the most commonly reported of the three related measurements. Chlorine consumed and chlorine dose were linearly dependent by construction ($r > 0.9$): chlorine consumed equals dose minus residual and dose is driven by a target residual goal which is approximately constant relative to much larger differences in chlorine dose

(Obolensky, Singer, and Shukairy, 2007). Chlorine consumed was retained as an independent variable instead of dose because it has been used extensively in previous laboratory studies and modeling work.

Because distributions of turbidity and chlorine contact time data were strongly positively skewed and had large coefficients of variation (see Table 4.2), these variables were log-transformed for scale adjustment. After determining that models were optimized by log transformation of dependent variables (see below), the benefits of transforming remaining independent variables were evaluated with respect to linearity of relationships with the log-transformed dependent variables (determined by pearson correlation and graphic inspection) and/or uniformity in the spread of values (determined by graphic inspection). Accordingly, bromide, chlorine consumed, TOC, UV254, and chlorine residual were log-transformed. Thus, information provided by the ratio variables cited above was implicit in individual model terms (e.g., log(UV254/TOC) = log(UV254)-log(TOC), etc.).

4.3.6 Correlation Matrices

Table 4.3 provides the matrix of Pearson correlation coefficients for continuous independent variables after applicable transformation. As expected, TOC and UV254, both of which are measures of natural organic matter (NOM) concentration before oxidant addition, were fairly strongly correlated (r = 0.655). Chlorine consumed was moderately correlated with temperature (r = 0.411) because chlorine reactions are accelerated at higher temperature and chlorine doses are increased to account for this; Chlorine consumed was also moderately correlated with both TOC (r = 0.430) and UV254 (r = 0.453) due to substantial chlorine demand exerted by NOM. Weaker but noteworthy correlations existed between bromide and

alkalinity (r = 0.370), bromide and temperature (r = 0.351), UV254 and turbidity (r = 0.306), and TOC and turbidity (r = 0.286). In many U.S. surface waters, bromide tends to increase with temperature as hot weather coincides with dry conditions and greater opportunities for salt water intrusion and groundwater infiltration into surface supplies (ground waters generally have higher bromide concentrations than surface waters). The latter would tend to also increase alkalinity in surface waters, which may account in part for a relationship between bromide and alkalinity. In surface waters impacted by wastewater flows and associated constant bromide loading, decreased dilution during hot dry periods can result in higher bromide concentrations. Turbidity may be an indicator of runoff which introduces fresh organic material into surface waters, accounting for the positive correlations with UV254 and TOC.

Table 4.4 provides the matrix of Pearson correlation coefficients between dependent and independent variables after applicable log transformations. The aggregate DBP measures (THM4, X_2AA, X_3AA, HAA9, and TOX) were all moderately correlated with chlorine consumed, TOC, and UV254 ($r > 0.4$). THM4 was more strongly correlated with temperature (r = 0.48) than were the other aggregate measures ($r < 0.34$). The absence of correlation between pH and THM4 as well as the negative correlation between X_3AA and pH (r = −0.35) are also noteworthy.

Table 4.3 Correlation matrix of independent variables after log transformation as needed for model development

Variable [a]	turbidity	bromide	temperature	alkalinity	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual
turbidity	1.000									
bromide	0.156	1.000								
temperature	0.099	0.351	1.000							
alkalinity	0.013	0.370	0.044	1.000						
TOC	0.286	0.275	0.194	0.024	1.000					
UV254	0.306	0.240	0.172	-0.008	0.655	1.000				
Cl_2 consumed	0.217	0.262	0.411	0.073	0.430	0.453	1.000			
Cl_2contact time	-0.046	-0.096	-0.168	-0.025	-0.088	-0.062	0.064	1.000		
pH	-0.121	-0.031	-0.096	0.099	-0.070	-0.025	-0.133	0.022	1.000	
Cl_2residual	0.240	0.214	0.244	0.081	0.215	0.161	0.254	-0.162	-0.040	1.000

[a]all variables log- transformed except alkalinity, temperature, and pH

Table 4.4 Pearson correlation coefficients between independent and dependent variables after log transformation as needed for model development

Independent Variables[a] / Dependent Varibles[a]	turbidity	bromide	temperature	alkalinity	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual
$CHCl_3$	0.243	0.007	0.391	-0.279	0.480	0.515	0.596	0.062	0.016	0.213
$CHBrCl_2$	0.319	0.611	0.463	0.181	0.510	0.349	0.542	-0.019	-0.170	0.279
$CHBr_2Cl$	0.138	0.784	0.347	0.413	0.200	0.059	0.204	0.022	-0.060	0.185
$CHBr_3$	0.028	0.558	0.191	0.312	0.060	0.066	0.046	0.015	0.054	0.058
Cl_2AA	0.210	-0.081	0.268	-0.306	0.470	0.455	0.470	-0.060	-0.060	0.230
BrClAA	0.309	0.606	0.399	0.258	0.420	0.250	0.414	-0.051	-0.140	0.331
Br_2AA	0.056	0.663	0.272	0.364	0.130	0.133	0.143	-0.026	0.015	0.163
Cl_3AA	0.179	-0.132	0.217	-0.390	0.380	0.424	0.512	0.084	-0.290	0.199
$BrCl_2AA$	0.249	0.489	0.301	0.091	0.440	0.314	0.463	-0.001	-0.380	0.301
Br_2ClAA	0.079	0.665	0.258	0.279	0.190	0.111	0.214	-0.012	-0.180	0.239
Br_3AA	0.021	0.412	0.137	0.143	0.270	0.186	0.161	-0.029	-0.190	0.232
Cl_3AH	0.189	0.036	0.452	-0.187	0.395	0.317	0.565	0.180	-0.132	0.174
THM4	0.272	0.277	0.477	-0.126	0.520	0.522	0.647	0.062	-0.002	0.255
X_2AA	0.240	0.103	0.339	-0.197	0.510	0.466	0.510	-0.064	-0.080	0.277
X_3AA	0.192	0.067	0.281	-0.266	0.410	0.434	0.545	0.077	-0.350	0.230
HAA9	0.241	0.092	0.309	-0.251	0.510	0.478	0.546	0.007	-0.190	0.278
TOX	0.257	0.174	0.281	-0.173	0.550	0.560	0.487	-0.066	-0.120	0.279

[a]all variables log- transformed except alkalinity, temperature, and pH

Several strong trends were evident across species within the THM, X_2AA, and X_3AA classes. With increasing bromine substitution in each class, correlations with chlorine consumed, TOC, and UV254 tended to drop sharply while correlations with alkalinity and bromide tended to increase sharply. Each of the fully chlorinated species (i.e., $CHCl_3$, Cl_2AA, Cl_3AA, Cl_3AH) were either uncorrelated with or weakly inversely correlated with both bromide and alkalinity. Cl_3AA and $BrCl_2AA$ were the only species having any notable correlation with pH (r = −0.290 and −0.380, respectively). Correlations for aggregate variables reflect the domination of chlorinated DBP species in this data set, which stems from the relatively low bromide conditions at the chlorine plants under study (median bromide concentration = 0.02 mg/L; see Table 4.2). ICR plants with high bromide concentrations tended to use chloramines and thus were not included in this analysis (Obolensky et al. 2007).

Variable selection results for the regression models will be strongly dependent on relationships shown in these correlation matrices. Independent variables having strong simple correlations with dependent variables are likely to be selected as statistically significant unless their effects are accounted for by other terms in the model. However, the absence of strong simple correlation with the dependent variable does not preclude an independent variable from being an important predictor. For example, although the influence of pH on THM formation is well known (Oliver and Thurman 1983; Reckhow and Singer 1985; Stevens et al. 1989; Liang and Singer 2001), no simple correlation existed between pH and THM4 or any individual THM species, as shown in Table 4.4. A weakly correlated variable may have discernible influence after controlling for other factors in the model; the ability to

elucidate such relationships is a primary reason for developing and employing the regression model in this research.

4.3.7 Model Building Procedure

Regression models were specified using a cross-validation approach (Muller and Fetterman 2003). Previous DBP models have either lacked validation or been validated retrospectively, so that any associated acceptance criteria were not considered during model development (Edzwald, Becker, and Wattier 1985; Amy, Chadik, and Chowdhury 1987; Amy et al. 1998; Golfinopoulos et al. 1998; Gang et al. 2002; Sohn et al. 2004). Formally testing exploratory results on an independent data set increases confidence that the chosen model reflects underlying relationships among variables and is not merely a best fit to the particular data at hand. This approach also controls Type I error inflation caused by performing multiple tests for significance with the same dataset. For each model, a portion of the data (usually half) was used in an exploratory process to identify significant variables and develop the regression model, while the remaining data were reserved for model validation. The following nine steps were taken to select and validate models: 1) identify candidate independent variables, adjust their scale and/or location, generate polynomial and/or interaction terms; 2) randomly split data set and reserve validation data; 3) assess collinearity among independent variables and intercept, discarding collinear terms; 4) identify optimal dependent variable transform from Box-Cox series; 5) conduct residual diagnostics to assess validity of assumptions and model specification, and check for outliers and influential observations; 6) select model using stepwise regression, adjusted R^2, and Mallows' *Cp* criteria (Daniel and Wood 1971); 7) perform cross-validation with reserved data; 8) pool data

to recalibrate parameter estimates for validated model and repeat residual diagnostics; 9) if validation fails, repeat from step 2 with larger exploratory data fraction.

Based on preliminary results for THM4 and X_2AA, second-order terms did not increase model precision meaningfully and were thus not utilized. Box-Cox analysis indicated use of log transformation as the optimum choice for all dependent variables. The SAS *Reg* procedure was used to select model independent variables while model validation computations were carried out using the SAS *Interactive Matrix Language* procedure. Using parameter estimates from the preferred model, adjusted R^2 and predicted values of the independent variable were computed for both the exploration and validation data subsets. Bias was considered acceptable if R^2 for the validation data was within ten percent of that for the exploration data (Muller and Fetterman 2003). Further details of the model development procedure are provided in the Supplemental Information (see Appendix B).

4.4 RESULTS AND DISCUSSION

4.4.1 Brominated X3AA Species Estimation

Table 4.5 lists the fitting parameters and R^2 values for Equations 4.1-4.3. The fit for $BrCl_2AA$ was very good ($R^2 = 0.876$), while that for Br_2ClAA was fair ($R^2 = 0.671$) and that for Br_3AA was very poor ($R^2 = 0.229$). Lower precision for Br_2ClAA and Br_3AA estimates may be partly attributable to measurement uncertainty associated with low occurrence levels and high MRLs for these compounds (67 and 97% of results, respectively, were below MRL). However, model precision was not improved by dropping censored data. Low recovery and accuracy in quantitative analysis of the brominated X_3AA species (Munch et al. 2000; Domino et al. 2004) probably depressed reported concentrations and contributed

substantially to scatter in the models. These analytical methodology problems are known to affect Br_2ClAA and Br_3AA more than $BrCl_2AA$ (Weinberg 2000), commensurate with the model fitting results.

Table 4.5 Brominated X_3AA species data and fitting parameters for Equations 4.1-4.3

Variable	*N*	MRL (μg/L)	*N* > MRL	MRL replacement	Model Results			
					N [a]	R^2	Intercept	Slope
$BrCl_2AA$	4122	1	3408	0.5	3943	0.8756	0.422	0.804
Br_2ClAA	3767	2	1230	0.5-1.0[b]	3600	0.6712	0.770	0.418
Br_3AA	2773	4	79	0.5-3.0[b]	2663	0.2295	1.014	0.270

[a] records used to parameterize model (data sets having required $CHCl_3$, Cl_3AA and brominated THM data)
[b]see SI

As expected, intercepts for Equations 4.1-4.3 corresponded to censored data replacement values. Model slopes decreased sharply with the number of bromine substituents: from 0.804 for $BrCl_2AA$ to 0.418 for Br_2ClAA to 0.270 for Br_3AA (see Table 4.5). Thus, the ratios of brominated X_3AA species to Cl_3AA tended to be lower than analogous THM species ratios (e.g., $Br_2ClAA/Cl_3AA < CHBr_2Cl/CHCl_3$), and the magnitude. This disparity increased sharply with the number of bromine substituents, in conformance with the order of methylated solvent extract stability and overall analytical method recovery, supporting the hypothesis that analytical method issues account, in part, for depression of brominated X_3AA concentrations relative to proportionate levels projected from THM species patterns. Results are also consistent with the order of chemical stability for brominated X_3AAs in water (Zhang and Minear 2002). Moreover, because X_3AA decay proceeds through decarboxylation, yielding the THM species analogue (e.g., Br_3AA decays to $CHBr_3$), these losses simultaneously enrich the brominated THMs (see also Heller-Grossman et al. 1993). Because THMs are terminal byproducts, the same can be expected for

brominated species in other classes of trihalogenated DBPs (e.g., haloacetaldehydes, haloketones, and halopicrins).

The last five rows in Table 4.1 summarize the augmented data set for $BrCl_2AA$, Br_2ClAA, Br_3AA, X_3AA, and HAA9, based on application of Equations 4.1-4.3 using the fitting parameters from Table 4.5 to estimate missing data. Together, the three brominated X_3AA species represented, on average, 13% of HAA9 on a molar basis in the augmented database. Because $BrCl_2AA$ was the dominant brominated X_3AA species for ICR treatment plants and it was well estimated by Equation 4.1, the lack of precision for Br_2ClAA and Br_3AA estimates should have a minimal impact on projected X_3AA and HAA9 values.

4.4.2 Regression Model Formulation

Equation 4.4 shows the generalized multiple linear regression model with abbreviations for turbidity (turb), bromide (br), temperature (temp), TOC (toc), UV254 (uv), alkalinity (alk), chlorine consumed (cl2), chlorine contact time (t), chlorine residual (res), and Cl_2 point (precl2). Units are as shown in Tables 4.1 and 4.2.

$$\begin{aligned}\log(\mathrm{DBP}) = \mathrm{intercept} &+ k_{turb}\log(\mathrm{turb}) + k_{br}\log(\mathrm{br}) + k_{temp}\,\mathrm{temp} + k_{alk}\,\mathrm{alk} + k_{toc}\log(\mathrm{toc}) \\ &+ k_{uv}\log(\mathrm{uv}) + k_{cl2}\log(\mathrm{cl2}) + k_{t}\log(\mathrm{t}) + k_{ph}\,\mathrm{pH} + k_{res}\log(\mathrm{res}) + k_{precl2}\,\mathrm{precl2}\end{aligned} \quad (4.4)$$

Results for THM4 are used to illustrate the model development procedure. Further details are provided in the Supplementary Information (Appendix B). Using half the available data for exploratory analysis (N = 357, excluding records with missing data for any candidate variables), stepwise regression indicated that eight of the eleven variables were

significant at the 0.05 level. Adjusted R^2 and Mallow's *Cp* statistics confirmed the choice of eight variables as optimal. The best eight-variable model had an adjusted R^2 of 0.70 for the exploratory data and 0.65 for the validation data, indicating acceptable cross-validation shrinkage of 7.7% (i.e., < 10%). This model was then re-calibrated using the combined data set ($N = 741$) to obtain final parameter estimates and model statistics. The overall model was highly significant (F statistic $p < 0.0001$) and all variables were significant at the 0.001 level and had variance inflation factors below 2.0. The parameterized THM4 model is shown in Equation 4.5.

$$\begin{aligned}\log(\text{THM4}) &= -1.371 + 0.015 * \text{temp} - 0.0005 * \text{alk} + 0.188 * \log(\text{toc}) \\ &+ 0.326 * \log(\text{uv}) + 0.291 * \log(\text{cl2}) + 0.119 * \log(\text{t}) + 0.087 * \text{pH} + 0.167 * \log(\text{res})\end{aligned} \quad (4.5)$$

The final validated model provided a good fit to the data overall ($R^2 = 0.707$), as illustrated by Figure 4.1, showing predicted versus observed values for log(THM4). Residual diagnostic plots are shown in Figure 4.2: the error variance was reasonably homogeneous across the range of predicted values (Figure 4.2a) and the residuals were well approximated by a normal distribution (Figure 4.2b). The Kolmogorov-Smirnov empirical distribution function test statistic p value of 0.055 indicated a normal distribution, although the residuals were slightly negatively skewed (Figure 4.2b inset). Partial regression leverage plots did not suggest non-linear effects for any of the independent variables or the presence of obvious outliers or influential observations warranting examination. Thus, the assumptions required for hypothesis testing using the model were considered valid. Model results are summarized in Tables 4.6 and 4.7, showing parameter estimates and key statistics for final validated

models for DBP species and aggregate DBP variables, respectively. Because $CHBr_3$, Br_2AA, Br_2ClAA, and Br_3AA exhibited consistently low concentrations in the data set (85, 75, 74, and 99% of results were below the MRL, respectively (see Table 4.1)), results were not acceptable for these brominated species and thus they are not shown in Table 4.6. All final validated models were significant at the 0.0001 level. Results are discussed further below.

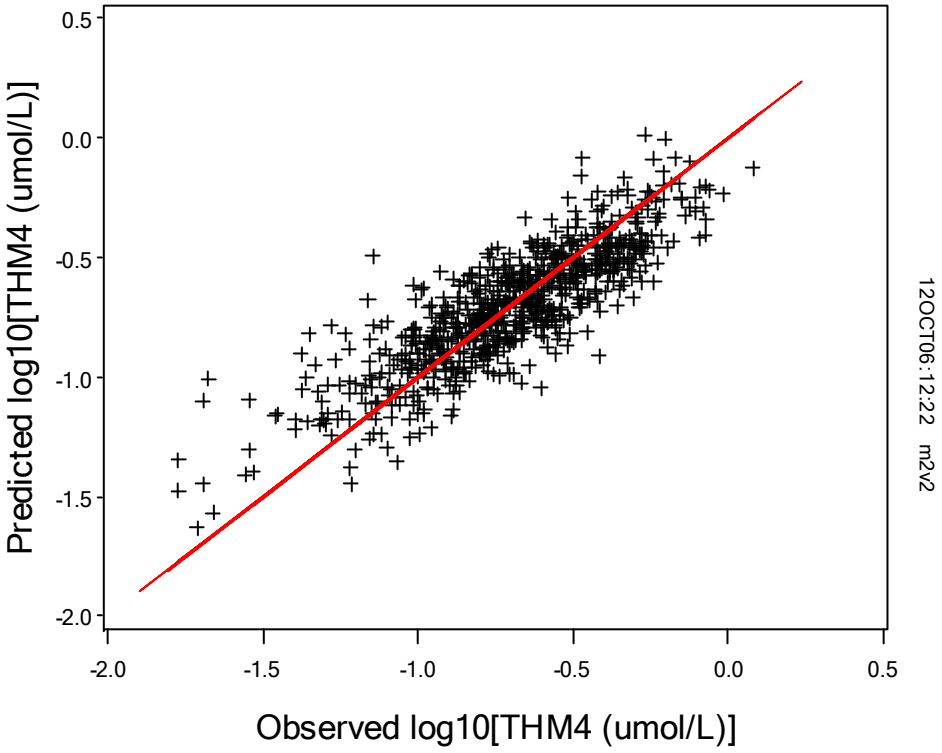


Figure 4.1 Scatterplot of predicted versus observed values for final total trihalomethane model with 1:1 reference line

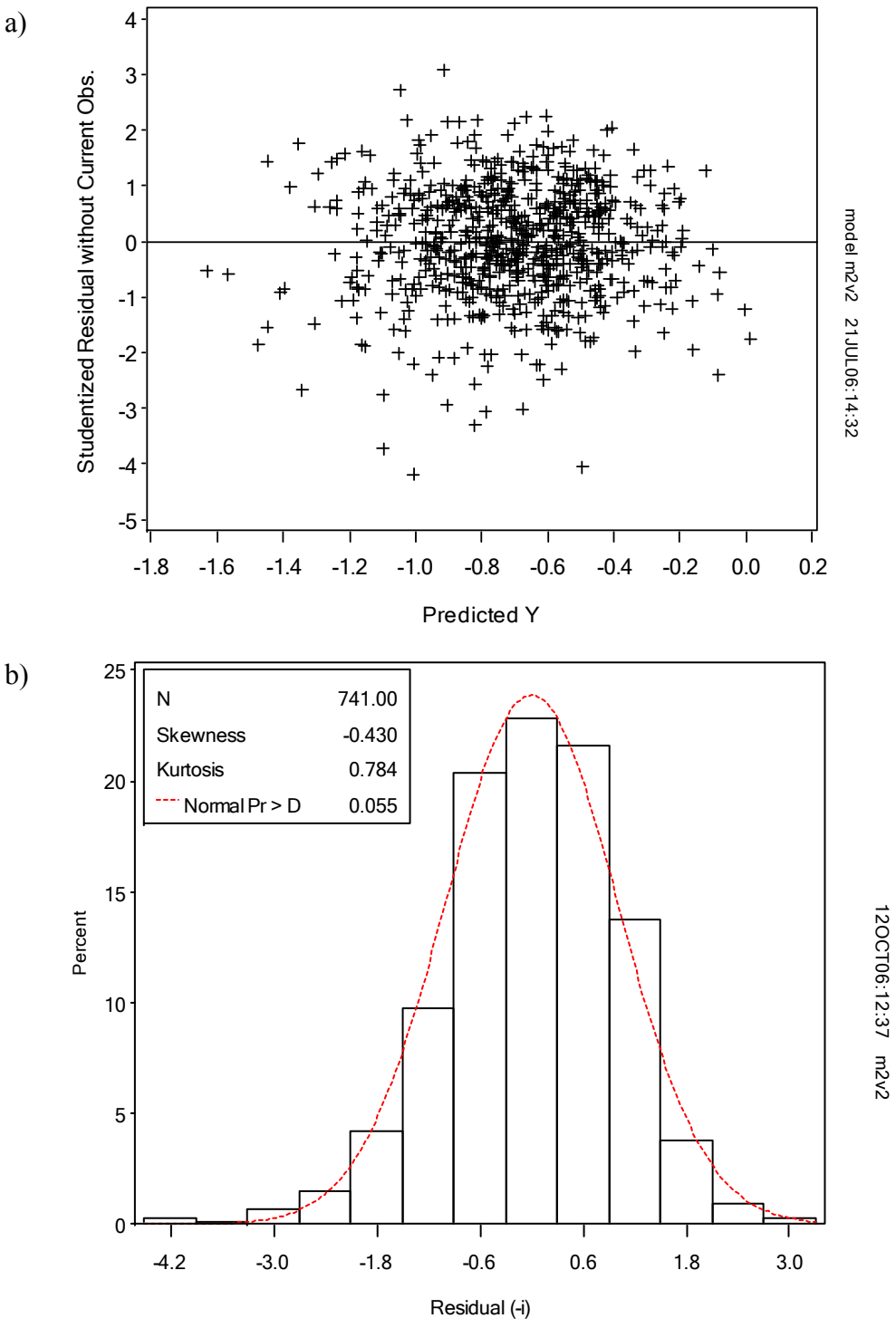


Figure 4.2 Residual diagnostic plots for total trihalomethane model: a) studentized residuals as a function of predicted log[THM4]; b) histogram of studentized residuals with fitted normal curve

Table 4.6 DBP species model results

a) Parameter estimates

Model	Intercept	Turb	Br	Temp	Alk	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual	Cl_2 point
$CHCl_3$	-1.935		-0.2393	0.0170	-0.0012	0.1993	0.4450	0.3824	0.0921	0.1133		
$CHBrCl_2$	-0.984	0.0483	0.4723	0.0088		0.4013		0.2493				
$CHBr_2Cl$	-1.247	0.0501	0.9070	0.0073	0.0009	0.1624	-0.2569		0.1174			
Cl_2AA	-1.284		-0.2779	0.0069	-0.0009	0.2864	0.3189	0.3898				-0.1010
BrClAA	-1.456	0.0472	0.4083	0.0054	0.0007	0.2820		0.1710			0.1890	-0.0621
Cl_3AA	-0.653	-0.0681	-0.2814		-0.0015	0.1645	0.4376	0.5563	0.0673	-0.0775	0.1732	
$BrCl_2AA$	-0.532		0.3224			0.3320		0.1966	0.0526	-0.1247	0.1987	
Cl_3AH	-2.787		-0.2072	0.0225	-0.0008	0.4220		0.3852	0.2051			

b) Model statistics

Model	N [a]	Adj R^2 [b]	p [c]	t statistics for parameter estimates											
				Intercept	Turb	Br	Temp	Alk	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual	Cl_2 point
$CHCl_3$	730	0.6854	9	-15.0		-9.2	13.5	-8.6	4.1	11.2	11.9	5.0	10.1		
$CHBrCl_2$	848	0.6579	6	-19.9	3.8	20.7	7.7		10.3		9.0				
$CHBr_2Cl$	783	0.6721	8	-11.8	3.2	29.1	5.3	5.8	2.8	-5.6		5.5			
Cl_2AA	749	0.5042	8	-13.3		-9.1	4.8	-5.5	5.3	7.1	11.1				-5.3
BrClAA	808	0.5647	9	-21.9	3.2	14.2	4.1	4.9	6.6		5.4			3.9	-3.4
Cl_3AA	697	0.5762	10	-3.9	-3.8	-8.4		-8.4	2.6	8.4	14.5	2.7	-5.3	3.0	
$BrCl_2AA$	737	0.5127	7	-5.0		13.1			7.7		6.6	2.6	-10.5	4.3	
Cl_3AH	758	0.5514	7	-36.8		-6.6	14.7	-4.7	8.4		10.5	9.0			

[a] number of records used to parameterize model
[b] adjusted R-squared
[c] number of model parameters including intercept

Table 4.7 THM4, X_2AA, X_3AA, HAA9, and TOX model results

a) Parameter estimates

Model	Intercept	Turb	Br	Temp	Alk	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual	Cl_2 point
THM4	-1.3708			0.0146	-0.0005	0.1881	0.3258	0.2905	0.1186	0.0871	0.1667	
X_2AA	-1.1353		-0.1531	0.0072		0.2497	0.2488	0.3239			0.1310	-0.0917
X_3AA	0.0776	-0.0455			-0.0009		0.3826	0.4282	0.0450	-0.0885		
HAA9	-0.5038			0.0034	-0.0008	0.2072	0.1970	0.3434				-0.0577
TOX	2.4590				-0.0007	0.3037	0.3673	0.2260			0.2689	

b) Model statistics

Model	N [a]	Adj R^2 [b]	p [c]	t statistics for parameter estimates											
				Intercept	Turb	Br	Temp	Alk	TOC	UV254	Cl_2 consumed	Cl_2 contact time	pH	Cl_2 residual	Cl_2 point
THM4	741	0.7073	9	-14.4			14.7	-4.7	4.9	10.2	11.2	7.7	9.6	4.8	
X_2AA	737	0.4822	8	-13.3		-6.1	5.6		5.1	6.2	10.3			2.9	-5.2
X_3AA	655	0.4819	7	0.65	-3.0			-5.8		10.3	13.3	2.2	-7.0		
HAA9	595	0.4964	7	-7.4			2.8	-5.9	4.4	5.1	11.2				-3.4
TOX	764	0.4963	6	36.8				-5.2	6.3	9.2	7.5			6.2	

[a] number of records used to parameterize model
[b] adjusted R-squared
[c] number of model parameters including intercept

It was not always the case that stepwise regression results and patterns of adjusted R^2 and Mallow's *Cp* statistics pointed to the same number of variables for model inclusion, as observed for THM4. Best judgment was used considering the available information and model selection was repeated if validation failed or any of the selected variables were not significant at the 0.05 level. In all cases, low variance inflation factors indicated that standard errors of parameter estimates were not adversely affected by collinearity and that initial efforts at addressing this through elimination of variables were effective (see Supplemental Information in Appendix B). Model residuals were slightly negatively skewed in all cases, but Kolmogorov-Smirnov statistic *p* values always exceeded 0.01, indicated lack of strong departure from normality. Figure 4.3 illustrates the model fits for $CHBrCl_2$, $BrCl_2AA$, X_2AA, and TOX.

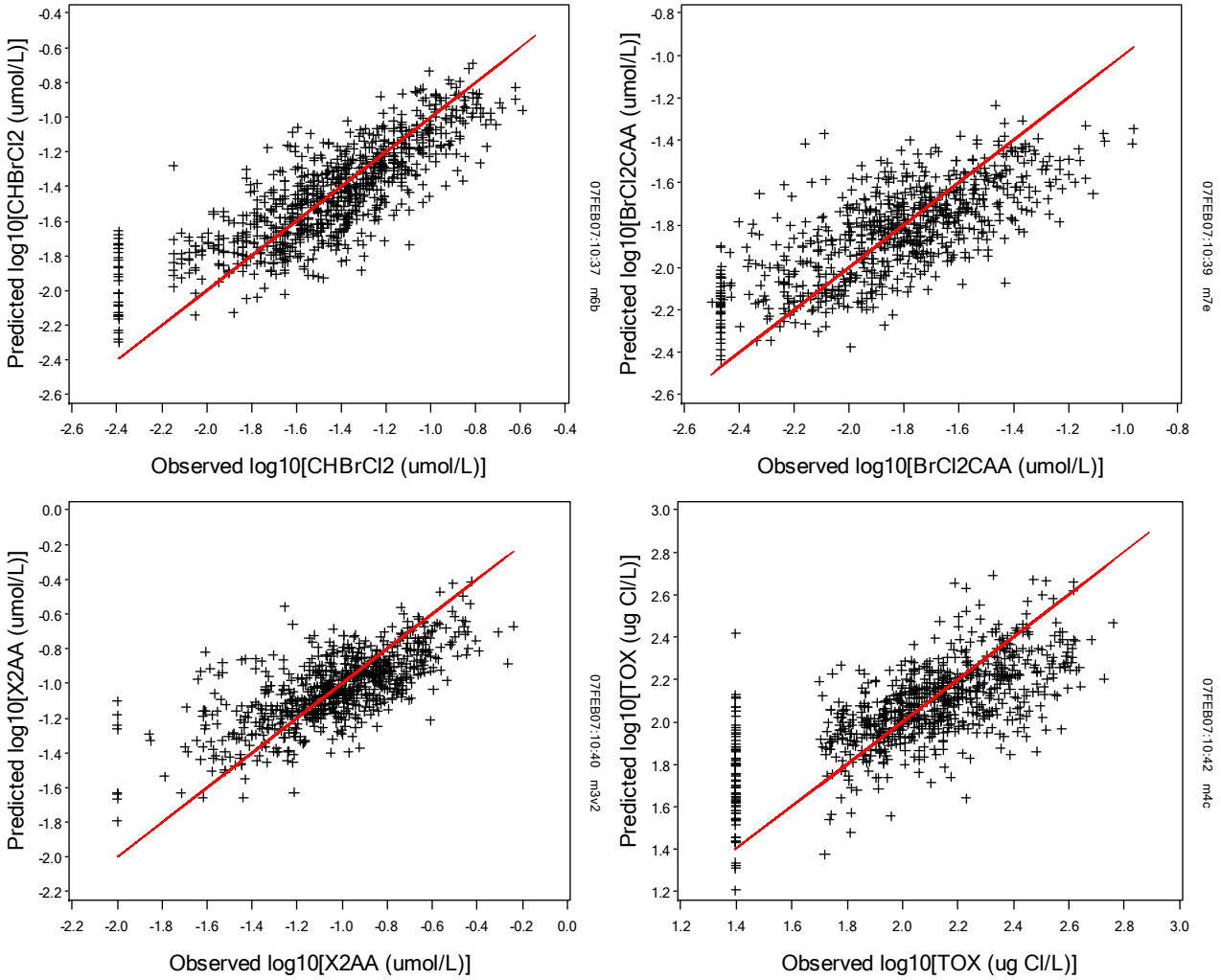


Figure 4.3 Scatterplots of predicted versus observed values for final models with 1:1 reference lines: a) $CHBrCl_2$, b) $BrCl_2AA$, c) X_2AA, and d) TOX

As Figure 4.1 illustrates, the THM4 model had a slight tendency toward overprediction at low concentrations. This was a consistent issue for all the DBP models (see Figure 4.3) and indicated some type of misspecification that was not identifiable from residual diagnostics. Attempts to address this were made through two alternate modeling approaches. Considering that analytical uncertainty for very low concentration data might be responsible for the problem, the model was reformulated after truncating the data set at different minimum THM4 values (note that log THM4(μmol/L) = −1.5 is equivalent to 3.8 μg/L $CHCl_3$) However, the misspecification pattern persisted and appeared to worsen as higher truncation levels were tested. Another potential source of model misspecification was conceptualizing the treatment train as a single input/output process whereas, in reality, it is a series of unit processes with continuously changing water quality conditions and often multiple chlorine application points. Because DBP results were available for filtered water locations, alternative models for THM4 and X_2AA were developed based on conceptualizing the treatment train as sequential raw to filtered water, and filtered to finished water process segments. This approach did not alleviate the misspecification and yielded poorer models with unacceptable residual patterns. However, the analysis showed that the pattern of overprediction at low concentrations was associated only with records for incremental filtered to finished water DBP formation for plants adding chlorine upstream of filtration, suggesting that these secondary stages of DBP formation were more difficult to describe in the linear modeling framework (see Supplemental Information in Appendix B). Absent a clear way to account for this with available data, the original unsegmented model formulation was considered the best approach within the framework of a multiple linear regression model.

4.4.3 DBP Species Model Results

DBP species model results are summarized in Table 4.6. Table 4.6a lists the parameter estimates (i.e., intercept and coefficients for each independent variable) and Table 4.6b shows the numbers of records available to parameterize each model, adjusted R^2 values, numbers of parameters in the model, and *t* statistics (the parameter estimate normalized by its standard error) for each parameter. A parameter's *t* statistic indicates the direction of influence, relative magnitude, and statistical significance of the model variable. Thus, it provides a basis for comparing the influence of variables within and between models.

The DBP species models show several important trends within the THM, X_2AA, and X_3AA classes. Bromide was a significant variable for all DBP species and its influence changed in sign and magnitude with the number of bromine substituents (see Table 4.6b). Bromide had a moderately negative influence on all fully chlorinated species: *t* statistics ranged from −6.6 to −9.2 for $CHCl_3$, Cl_2AA, Cl_3AA, and Cl_3AH. With a single bromine substituent, bromide became a strongly positive predictor and was the most significant variable in each model: $_t$ statistics ranged from 13.1 to 20.7 for $CHBrCl_2$, BrClAA, and $BrCl_2AA$. The *t* statistic for bromide in the $CHBr_2Cl$ model was 29.1 and it accounted for 90% of the variation explained by the model overall.

The role of alkalinity in these DBP species models followed a pattern similar to that seen for bromide, but with generally lower significance levels. Alkalinity had a negative influence on all the fully chlorinated species but became a positive factor with bromine substitution ($CHBr_2Cl$ and BrClAA). This result was somewhat surprising. Previous DBP models have not included alkalinity as an independent variable, although one early study noted a negative correlation between alkalinity and $CHCl_3$ formation (Otson et al. 1981).

Given the inclusion of bromide in the models, the role of alkalinity cannot be attributed to its moderate correlation with bromide (see Table 4.4). However, results are consistent with alkalinity being an indicator of NOM hydrophobicity and reactivity. Archer and Singer (2006) demonstrated that the UV254/TOC ratio, an indicator of NOM hydrophobicity and reactivity toward DBP formation, tended to decrease with increasing alkalinity in ICR raw waters, and related this to immobilization of hydrophobic NOM under the high ionic strength and high hardness conditions of such waters. Thus, high alkalinity waters are enriched in hydrophilic NOM. Because reactive DBP precursor sites are more concentrated in hydrophobic NOM fractions compared to hydrophilic fractions, fewer reactive centers for DBP formation would be available per unit of TOC in such waters. The higher bromide to precursor ratios would lead to greater extents of bromine substitution in high alkalinity waters, for fixed bromide and TOC levels. Observed alkalinity patterns in the DBP species models, which control for both TOC and bromide, are in accord with this concept. In laboratory studies of NOM isolates, Liang and Singer (2003) found hydrophilic NOM fractions to be more reactive towards bromine substitution than hydrophobic fractions, for a fixed bromide to TOC ratio. This increased reactivity may be attributable to lower precursor concentrations in hydrophilic NOM, per unit TOC, and consequently higher bromide to precursor ratios. It may also be related to halogen substitution being more important than oxidation for reaction centers in hydrophilic NOM, compared to those in hydrophobic NOM, because bromine is a stronger reagent than chlorine for halogen substitution, whereas chlorine is a stronger oxidant (Rook et al. 1978; Boyce and Hornig 1983; Johnson and Jensen 1986). The absence of correlation between alkalinity and either UV254 (Table 4.2) or the UV254/TOC ratio (not shown) in the current data set, as observed by Archer and Singer

(2006), may be a consequence of the predominantly low organic precursor levels for ICR chlorine plants considered here (Obolensky, Singer, and Shukairy 2007) or may be related to covariances in the dataset that mask such a relationship. Nevertheless, the interpretation of alkalinity's association with NOM reactivity is consistent with the observed role of UV254 in the DBP species models. The influence of UV254, when present in the models, was always opposite in direction to that for alkalinity (see Table 4.6).

Chlorine consumed and organic precursor measures (TOC and UV254) were significant variables in almost all DBP species models (see Table 4.6b). With respect to numbers of bromine substituents, these terms exhibited trends opposite to those seen for alkalinity and bromide. With *t* statistics ranging from 10.5 to 14.5, chlorine consumed had a strong positive influence on all fully chlorinated species ($CHCl_3$, Cl_2AA, Cl_3AA, and Cl_3AH); chlorine consumed was less important for species with one bromine substituent (*t* statistics ranged from 5.4 to 9.0 for $CHBrCl_2$, BrClAA, and $BrCl_2AA$); with two bromine substituents, chlorine consumed was not a significant variable for $CHBr_2Cl$. TOC exerted a significant influence in all DBP species models and UV254 provided additional information in 4 of the 8 models. Considering these two variables together, the influence of organic precursors decreased overall with increasing numbers of bromine substituents within each DBP class. In particular, UV254 had a strong positive influence on most of the fully chlorinated species ($CHCl_3$, Cl_2AA, and Cl_3AA) but was either absent ($CHBrCl_2$, BrClAA, and $BrCl_2AA$) or had a negative influence ($CHBr_2Cl$) on brominated compounds. Without interaction terms, the models indicate that higher UV254 absorbance at fixed TOC and bromide levels (a decrease in bromide to precursor ratio) is associated with higher

chlorinated DBP concentrations and lower brominated DBP concentrations, consistent with the general hypothesis discussed above.

Temperature was the most significant variable in both $CHCl_3$ (t = 13.5) and Cl_3AH (t = 14.7) models; there appeared to be a pattern of declining significance for temperature with increasing bromine substitution across the THM class. This may reflect the kinetic disadvantage of chlorine relative to bromine in substitution reactions, which is leveled to some extent with increasing temperature (Oliver 1980). Curiously, temperature was not identified as an important variable in either of the X_3AA species models but was moderately significant in both X_2AA species models (t < 5). Cl_3AH was the only DBP species in which chlorine contact time was found to be a highly significant model variable.

pH was identified as an important variable in only three of the eight DBP species models. Notably, pH had a strong positive influence on $CHCl_3$ formation but was not identified as an important predictor for either of the brominated THM species modeled. Conversely, pH exerted a negative influence on both X_3AA species, with a substantially greater effect on $BrCl_2AA$ than on Cl_3AA. pH was not a significant variable in either X_2AA species model. These findings are generally consistent with a large body of research showing enhanced $CHCl_3$ and THM4 formation at high pH, enhanced Cl_3AA and X_3AA formation at low pH, and minimal effect of pH on X_2AA species formation (Oliver and Thurman 1983; Reckhow and Singer 1985; Reckhow, Singer, and Malcom 1990; Stevens et al. 1989; Pourmaghaddas et al. 1993; Liang and Singer 2003). A mechanistic explanation by Reckhow and Singer (1985) for the opposite effects of pH on THMs and X_3AAs entailed their production from a common intermediate wherein base hydrolysis at an alpha carbonyl center favors cleavage to yield the THM analogue. It is difficult to explain the much stronger

observed negative influence of pH on $BrCl_2AA$, compared to Cl_3AA with no corresponding pattern observed for the THM species. $BrCl_2AA$ decomposition has been shown to be insensitive to pH (Weinberg 2000; Zhang and Minear 2002), so base-hydrolysis of formed $BrCl_2AA$ cannot account for this. However, the results are consistent with Cowman and Singer's (1996) findings that brominated HAAs were more enhanced than chlorinated HAAs for pH 6 chlorination of NOM isolates relative to pH 8 chlorination, suggesting that bromine substitution reactions are more favored at a lower pH.

Although turbidity and chlorine residual appeared in several of the DBP species models, there was no clear pattern to their inclusion and significance levels were generally low ($t \leq 4.3$). The categorical variable Cl_2 point, distinguishing whether plants applied their first chlorine dose to raw or settled/filtered water, was a significant predictor for both X_2AA species but not for any other DBP species modeled. The negative coefficient sign indicated that plants practicing pre-chlorination (Cl_2 point = 1) formed lower X_2AA levels than plants that delayed chlorine addition until after clarification processes (Cl_2 point = 0), other factors being fixed. This finding was counter-intuitive because clarification processes are known to reduce NOM reactivity towards DBP formation. Closer examination of treatment practices for the two groups of plants suggested that the analysis was confounded by differences in the number of chlorine doses utilized by plants practicing pre-chlorination, compared to those with delayed chlorine addition. After controlling for this difference, re-formulated models indicated that pre-chlorination had a small positive influence in two of the DBP species models. This is addressed further in a separate paper (see Chapter 5).

4.4.4 DBP Class Sum and TOX Models

Table 4.7 summarizes model results for aggregate DBP measures (THM4, X_2AA, X_3AA, HAA9, and TOX). Chlorine consumed, organic precursor variables (TOC and UV254), and alkalinity were moderately significant in all models, with the exception of alkalinity for the X_2AA model. Interestingly, bromide appeared only in the X_2AA model, where alkalinity was absent. The bromide and alkalinity parameters were consistently negative, reflecting dominance of the aggregate DBP measures by fully chlorinated DBP species in the low bromide data set under study. As discussed above, the influences of bromide and alkalinity on DBP concentrations strongly depended on the extent of bromine substitution in a particular compound and were always negative for fully chlorinated species. Similarly, the high level of significance of chlorine consumed for all five models was commensurate with its greater importance for chlorinated species than for brominated species. UV254, present in all aggregate DBP models, tended to be a more significant predictor than TOC, which was present in 4 of the 5 models. Temperature was highly significant only for THM4, consistent with its strong influence on $CHCl_3$ (see Table 4.6b). Similarly, pH was significant only for THM4 (positive) and X_3AA (negative), reflecting its importance and direction of influence in the $CHCl_3$, Cl_3AA, and Cl_2BrAA models discussed above.

The wide differences observed among individual DBP species models within the THM, X_2AA, and X_3AA classes suggest that aggregate DBP model results will be strongly dependent on the extent of bromine substitution in the particular data set used for model development. This is determined by water quality conditions, which tend to drive treatment conditions. In addition to such intraclass differences among DBP species models, distinctions

between classes or subclasses (interclass differences) would be expected to shape TOX model results. The data set analyzed here encompassed moderate pH, low bromide, and low organic precursor conditions (see Table 4.2) and DBPs were dominated by the fully chlorinated species in each class (see Table 4.1). Thus, models for TOX and DBP class or subclass sums were driven by the factors influencing these fully chlorinated species. Different results might be expected for data sets with higher pH, bromide, or organic precursor levels that shift the composition of TOX and/or DBP class/subclass sums with respect to either intraclass or interclass speciation patterns. For this reason, interpretation of models for total class sum or TOX concentrations, with respect to the influences of water quality and treatment, may not be as generally applicable as corresponding interpretation of model results for individual species. The same caveat would apply to use of aggregate models for predicting future values.

CHAPTER 5: APPLICATIONS OF MODELS TO DESCRIBE THE IMPACTS OF TREATMENT ON DBP FORMATION USING THE ICR DATABASE

5.1 BACKGROUND AND INTRODUCTION

A previous paper described development of multiple linear regression models for disinfection byproducts (DBPs) based on Information Collection Rule (ICR) data, and their use as analytical tools for the study of relationships between DBP formation and water quality and treatment processes (see Chapter 4). This paper presents extended applications of the this modeling work to examine the specific impacts of point of chlorination, alternative disinfectant use, and softening treatment on finished water DBPs. The research was performed as part of a broader study aimed at extracting untapped information related to DBPs from the ICR (Obolensky and Singer, 2005; Obolensky, Singer, and Shukairy 2007). Linear regression models were used as a platform for analysis of these data because of their ability to isolate and characterize effects from numerous factors of influence that vary simultaneously in a data set, and to provide for straightforward interpretation. The empirical models, based on current knowledge of drivers for DBP formation in drinking water, encompass commonly measured variables representing water quality and treatment processes. These included: turbidity, bromide, temperature, and alkalinity (plant influent); total organic carbon (TOC) and ultraviolet absorbance at 254 nm (UV254) (at the first point of chlorine addition in the treatment train); pH and total chlorine residual (plant finished

water); chlorine consumed (through the plant); chlorine contact time in the plant; and a coded variable to describe the location of first chlorine addition in the treatment train (1 = raw water, 0 = settled/filtered water).

Control of halogenated DBP formation in water treatment is accomplished by limiting the contact of chlorine with reactive organic precursors present in natural water. This may be effected by reducing concentrations of chlorine, organic substrates, or both. Coagulation and softening processes can remove substantial fractions of TOC during treatment, lowering concentrations of organic matter that exert chlorine demand and act as substrates in DBP formation reactions. Thus, for a given set of water quality and treatment conditions, moving the point of chlorine addition forward in the treatment train from the head of the plant to a location following clarification can lead to significant reductions in finished water DBP concentrations. Stevens and coworkers (1976) identified point of chlorine addition as the most important factor for trihalomethane (THM) control, and early laboratory and full scale studies showed the associated benefit of delayed chlorine addition (Babcock and Singer 1979; Young and Singer 1979). Moving the point of chlorination was the most common treatment modification implemented in the U.S. to lower total trihalomethane (TTHM) for compliance with the 1979 Interim TTHM Rule rule, according to a 1987 survey (McGuire and Meadow 1988). Since these early studies, research has addressed optimization of coagulation and softening for natural organic matter (NOM) and DBP precursor reduction, since these processes were originally designed and employed for particulate (e.g. turbidity), color, or hardness removal (Randtke 1999). As part of an integrated DBP control strategy, Safe Drinking Water Act regulations now require specified percentages of TOC removal between influent and finished water (linked to source water conditions), though utilities are

not restricted from adding chlorine upstream of settled water in the treatment train since this practice is often driven by competing treatment objectives (U.S. EPA 1998a). Though most ICR plants (67%) applied chlorine to raw water, those that delayed chlorine addition until settled or filtered water locations had significantly higher organic precursor levels (i.e. TOC and UV254), indicating that this practice was implemented for DBP control (Obolensky, Singer, and Shukairy 2007).

Because NOM in source waters is composed of a varied mixture of complex macromolecular structures that preclude specific chemical identification, measurements of bulk water properties (e.g. ultraviolet absorbance, TOC, color) and other indirect approaches (e.g. resin fractionation, molecular size, ^{13}C NMR, model compound studies) have been used to characterize NOM, its role in DBP formation, and its alteration or removal in drinking water treatment processes (Croue et al. 1999). Research has shown that certain types of moieties in NOM are more or less reactive towards DBP formation, and that specific DBP yields (i.e. per unit carbon) correlate with UV254 and activated aromatic content (Reckhow, Singer, and Malcom 1990). Thus UV254 is a more specific indicator of DBP formation potential than TOC. Hydrophobic and hydrophilic NOM components, differentiated by resin fractionation methods, have been shown to differ substantially with respect to DBP formation. Hydrophobic NOM fractions (humic substances, including humic and fulvic acids) generally have greater aromatic content, higher UV254, higher molecular weight, and higher chlorine consumption and DBP formation yields than corresponding hydrophilic NOM fractions, regardless of origin (Reckhow, Singer, and Malcom 1990; Croue et al. 1999). NOM fractions have also been found to differ with respect to relative formation yields of specific DBP classes (Croue et al. 1999; Marhaba and Van 2000; Liang and Singer 2003).

For example, aliphatic (i.e., non-aromatic) fractions have been found to contain more THM precursors than HAA precursors. At the same time, differences in chemical and structural properties render hydrophobic NOM more amenable to removal by coagulation and softening processes than hydrophilic NOM so that, in addition to lowering absolute TOC concentrations, these clarification processes alter TOC composition, preferentially removing UV254 absorbance and DBP precursor moieties over organic carbon generally (Babcock and Singer 1979; Chadik and Amy 1987; Randtke 1999; Liang and Singer 2001; Archer and Singer 2006). Therefore plants that delay chlorine addition until after NOM removal processes are expected to have lower specific DBP yields (per unit carbon) and possibly different relationships between DBP formation and other water quality and treatment factors, compared to plants chlorinating raw water. Liang and Singer (2001, 2003) found that precursors for THMs, dihaloacetic acids (X_2AAs), and trihaloacetic acids (X_3AAs) were removed to different extents by coagulation processes, and related this to differences in hydrophobicity and aromaticity of organic structures serving as substrates for the different classes of DBPs. Reckhow, Singer, and Malcom (1990) found that highly conjugated structures (i.e. aromatic and hydrophobic NOM) were enriched in Cl_3AA precursors, compared to precursors for $CHCl_3$, Cl_2AA, or total organic halogen (TOX) generally, and related this to a proposed reaction mechanism wherein the pathway to Cl_3AA was stabilized by conjugation. Thus, one objective of the current research was to evaluate the specific impact of delayed chlorine addition on DBP and TOX formation at ICR plants while accounting for related effects (e.g. TOC, UV254, chlorine consumed, and chlorine contact time).

A second objective of the current research was to examine the impact of alternative disinfectant use on DBP formation at ICR plants. Application of alternative oxidants/disinfectants by water utilities to replace some or all their use of free chlorine has been an important aspect of DBP control in the U.S. since the 1979 THM regulation (Singer 1994). Use of chloramines to arrrest continued THM formation, especially in distribution system transit, has been widespread. Historically, chlorine dioxide was applied during drinking water treatment primarily for taste and odor control (Hoehn and Gates 1999). After THMs became regulated, application of chlorine dioxide as a preoxidant/predisinfectant in place of chlorine became an important THM control strategy, since it acts primarily an oxidant and produces few halogenated byproducts. A survey of 35 U.S. water utilities utilizing chlorine dioxide indicated that reducing THM formation was the main purpose for its use in 65% of systems (Dietrich et al. 1992). Not surprisingly, the mean source water TOC concentration for these systems was quite high (7.65 mg/L), while EPA's consideration at the time of a 1 mg/L limit on total oxychloride residuals (sum of ClO_2, ClO_2^-, and ClO_3^-) effectively limited chlorine dioxide doses to an average of 1.24 mg/L (Werdehoff and Singer 1987). The low concentrations of organohalogens formed with chlorine dioxide contact are thought to be products of trace free chlorine used in its generation or produced as a reaction product of chlorine dioxide and phenolic moieties in NOM (Werdehoff and Singer 1987; Richardson 1998; Hoehn and Gates 1999).

Ozone has also been used as an alternative oxidant/disinfectant for control of halogenated DBPs and many other treatment objectives (e.g. taste and odor control, iron and manganese oxidation, disinfection, coagulant aid, removal of color and micropollutants), though the potential for bromate formation has presented a challenge for ozonation of high

bromide waters, and ozone reacts to form nonhalogenated DBPs such as formaldehyde. Ozone may also form halogenated DBPs in the presence of bromide. Ozone differs significantly from chlorine dioxide in that, in addition to selective direct oxidation pathways, it can act through highly reactive and nonselective hydroxyl radical reactions. Ozone is most effective for controlling halogenated DBPs when used as replacement for chlorine as a preoxidant/predisinfectant. Neither chlorine dioxide nor ozone leads to appreciable loss of TOC, although both oxidants alter NOM by reacting with DBP and TOX precursor sites (Werdehoff and Singer 1987; Reckhow 1999). Preozonation may improve particle removal in coagulation but may also lower the efficiency of TOC and DBP precursor removal in coagulation, due to oxidative cleavage of NOM into smaller, more highly charged and hydrophilic molecules (Randtke 1999; Reckhow 1999), though lower ozone doses have been demonstrated not to diminish TOC removal (Dowbiggin and Singer 1989). Some research has indicated that ozone consumes Cl_3AA precursors but not Cl_2AA precursors and that ozone may enhance the formation of certain compounds like haloketones and haloacetaldehydes (Reckhow and Singer 1984; Reckhow 1999). In a survey of twelve utilities treating high TOC or high bromide waters, Krasner and coworkers (2006) found that alternative disinfectants generally (i.e. chloramines, chlorine dioxide, and ozone) favored formation of dihalogenated DBPs over their trihalogenated analogues. Krasner et al. (2006) postulated that the preferential formation of dihalogenated acetaldehydes in ozone plants was attributable sequential chlorine substitution reactions with formaldehyde that was a product of prior ozone contact. Depending on treatment conditions, ozonation may lead to increased formation of some brominated DBP species following subsequent chlorination (e.g. Br_2AA and $CHBr_2Cl$), due to increased bromide to chlorine and/or bromide to precursor ratios

resulting from precursor oxidation and reduced chlorine demand, which led to lower chlorine doses (Miltner, Shukairy, and Summers 1992; Reckhow 1999; Westerhoff et al. 2000). Ozone reacts much more slowly with bromide than with organic DBP precursors, but generated hypobromous acid may react to yield bromoform or other fully brominated DBPs even in the absence of downstream free chlorine contact.

Softening treatment is distinguished from conventional treatment by the use of high pH conditions and addition of lime and/or sodium hydroxide to effect the removal of hardness ions from high alkalinity waters. With respect to influences on DBP formation, important characteristics of softening treatment include the characteristically high alkalinity source waters, the high pH of treatment, and the formation of calcium carbonate precipitates during treatment. Since both ozone and chlorine dioxide decompose at high pH, many softening plants rely on chloramines to control THM formation, especially those with high TOC source waters. Increases in THMs and decreases in TOX with elevated pH are well known (Oliver and Thurman 1983; Fleischaker and Randtke 1983; Reckhow and Singer 1985). Studies have shown dramatic reductions in Cl_3AA yields at high pH with little effect on Cl_2AA yields (Miller and Uden 1983; Reckhow and Singer 1985; Stevens et al. 1989; Reckhow, Singer, and Malcom 1990; Liang and Singer 2001), and reduced trichloroacetaldehyde (Cl_3AH) formation at high pH (Miller and Uden 1983). Due to the selective adsorption and precipitation of strongly ultraviolet-absorbing high molecular weight constituents, soft water systems contain greater proportions of high molecular weight dissolved organic matter than hard water systems. These differences have been observed in limnology studies. In an analysis of data for 55 lakes, Stewart and Wetzel (1981) showed that the loss of high molecular weight humic material was related to calcium concentrations, and

that both fluorescense and absorbance declined as dissolved calcium increased. Liming of acidified scandinavian lakes, increasing pH and calcium concentrations, was found to lower proportions of high molecular weight in in lake outlet NOM (Anderson, Alberts, and Takacs 2000). Thus, the effects of softening treatment on DBP formation are expected to stem from source water quality, treatment pH, and NOM alteration and removal in the plant. The third objective of this research was to apply multiple linear regression models to investigate these effects.

5.2 METHODS

Using regression models, the effects of point of chlorine addition, alternative disinfectant use, and softening treatment were examined for TOX and eight DBP species: namely chloroform ($CHCl_3$); bromodichloromethane ($CHBrCl_2$); dibromochloromethane ($CHBr_2Cl$); dichloroacetic acid (Cl_2AA); bromochloroacetic acid (BrClAA); trichloroacetic acid (Cl_3AA); bromodichloroacetic acid ($BrCl_2AA$); and chloral hydrate (trichloroacetaldehyde hydrate, Cl_3AH). Procedures used for ICR data handling and model development are described in Chapter 4 (see also Appendix B). Five sets of finished water DBP models were utilized, each set comprising nine distinct models corresponding to the above-listed dependent variables. Models were based on various subsets of data from ICR plants using only free chlorine for disinfection (chlorine plants), according to requirements of the particular analysis. The first set of models are referred to here as chlorine plant models. These models, described in detail in Chapter 4, were based on data for all chlorine plants with some exceptions. Additional sets of models were based on data for chlorine plants applying the first chlorine dose to raw or unclarified water (raw water models); chlorine

plants applying the first chlorine dose to settled or filtered water (settled water models); chlorine plants using a single chlorine dose (single dose models); and chlorine plants using conventional treatment (conventional plant models, see following).

The influence of point of chlorine addition on DBP formation was evaluated through the role of a binary classification variable in chlorine plant models (Cl_2 point=1 for raw, 0 for settled/filtered) and by comparing results for raw and settled water models. The former approach provides more power for detecting a main effect (i.e. point of chlorination). The latter approach allows for observing how point of chlorination might modify the effects of other factors present in the models, without necessitating use of interaction terms which introduce collinearity and present interpretation challenges. To control for possible confounding effects of multiple chlorine doses, the binary classification variable Cl_2 point was also tested in the single dose models.

Chlorine plant models were used as the baseline for evaluating alternative disinfectants' impacts on DBP formation. For plants utilizing chloramines, ozone, or chlorine dioxide, the chlorine plant models were employed to estimate concentrations of finished water DBPs that would be expected under existing water quality and treatment conditions without alternative disinfectant use. Discrepancies between observed results and model predictions were then attributed to the alternative disinfectant application, under the assumption that the models controlled for other significant influences on DBP formation. This assumption should be valid to the extent that these other influences were represented in the models. The relative percent difference (RPD) between predicted and observed values was used to characterize alternative disinfectant treatment effects. The RPD, computed as [observed-predicted]/predicted, was computed for each data point after back-transformation

to express results in μmol/L units (or μg Cl/L for TOX), since models utilized log-transformed dependent variables. Thus, negative RPDs indicated that, under the specified treatment (i.e. alternative oxidant), DBPs were lower than would be expected in its absence, or equivalently that DBP formation was reduced by use of the treatment. Conversely, positive RPDs indicated that DBPs were higher than expected, or that DBP formation was enhanced by the treatment.

In an approach analogous to that used for the alternative disinfectant analysis, the impact of softening treatment on DBP formation was evaluated using predictions from conventional plant models as the baseline. With input data from softening plants, the conventional plant models were employed to estimate concentrations of finished water DBPs that would be expected under existing water quality and treatment conditions in the absence of softening treatment (i.e. under conventional treatment), with differences between predicted and observed values attributed to the effect of softening. RPD values were used to quantify effects, as described above. A second approach to evaluating effects of softening treatment on DBP formation entailed conducting tests of statistical significance on a binary treatment classification variable (1 for softening, 0 for conventional) using the chlorine plant models (which spanned both conventional and softening treatment plants).

It was important that the design matrix for conventional plant models encompass the data domain of softening plants, in order that water quality and treatment effects (other than softening practice) impacting DBP formation would be adequately accounted for in generating predicted DBP values, and that model boundary conditions would not be violated. Distributions of model independent variables for conventional and softening plants were thus compared within categories of source water type (ground water versus surface water) and

treatment plant disinfectant type (chlorine versus chlorine/ chloramine) to identify a subset of data on which to base conventional plant models that could be appropriately applied to data from softening plants. Among model independent variables, alkalinity, pH, and TOC exhibited the greatest overall differences between conventional and softening plants. As expected, softening plants had a significantly higher alkalinity range than conventional plants, regardless of source water or disinfectant type category. TOC and pH differences between conventional and softening plants were much more pronounced for chlorine/chloramine plants (Cl_2/ClM) than for chlorine plants: a substantial fraction of softening Cl_2/ClM plants had TOC and pH values well outside the range for conventional Cl_2/ClM plants. However, TOC and pH values for softening chlorine plants were within the range of values for conventional chlorine plants. Therefore, conventional plant models were based on data for conventional plants using chlorine only, and the evaluation of softening treatment was restricted to plants using free chlorine disinfection. This also allowed the analysis of softening effects to be isolated from the influence of chloramine use. Although all conventional chlorine plants utilized surface water sources, it was not considered necessary to limit the analysis of softening effects only to plants using surface water (note that about half the softening chlorine plants utilized ground water sources). Although there were some distinctions between data characteristics of softening chlorine plants based on source type (e.g. higher turbidity and chlorine consumed at surface water plants), data from the ground water softening plants were within the ranges for conventional chlorine plants.

5.3 RESULTS AND DISCUSSION

5.3.1 Point of Chlorine Addition

As previously reported, the binary Cl_2 point variable was identified as a significant predictor in X_2AA species models (Cl_2AA and BrClAA), but not in models for THM species ($CHCl_3$, $CHBrCl_2$, or $CHBr_2Cl$), X_3AA species (Cl_3AA or $BrCl_2AA$), Cl_3AH, or TOX (see Chapter 4). The negative coefficient sign for Cl_2 point in the X_2AA species models indicated that plants applying chlorine to raw water (raw water plants) formed lower X_2AA concentrations than plants applying chlorine to settled or filtered water (settled water plants), controlling for other factors represented in the models (i.e. turbidity, bromide, temperature, alkalinity, TOC, UV254, chlorine consumed, and chlorine residual). This was counterintuitive and prompted further investigation. A comparison of results for separate raw and settled water models showed that much better fits were obtained for settled water models (R^2 = 0.6252-0.7443, n = 230-331), compared to raw water models (R^2 = 0.3774-0.6561, n = 475-549), despite the much larger data set available for the latter group. These precision differences were most pronounced for the X_2AA species models (60-70% improvement for settled over raw models, compared to 10-30% improvement for X_3AAs, THMs, Cl_3AH, or TOX). Although reasons for the superiority of settled water models were not immediately clear, these results confirmed the existence of a significant distinction for the X_2AA DBP class with respect to the impact of chlorination practices.

Recognizing that use of multiple chlorination points was not accounted for in the models heretofore, treatment practices were examined to determine how they may have affected the results. The number of applied chlorine doses was found to differ substantially

based on the location of first applied chlorine dose. Most raw water plants used two chlorine doses whereas settled water plants most often used one chlorine dose. Table 5.1 shows that 85% of raw water plants applied additional chlorine downstream of the first applied dose, compared to 46% for settled water plants. Considering the fraction of total dose applied at the first chlorination point (which decreased with the number of addition points used), raw water plants applied 38% of their total dose at downstream points, compared to only 19% for settled water plants (weighted averages from Table 5.1). Thus, organic precursor model input variables (i.e. TOC and UV254), for which values at the first point of chlorine addition were employed, more accurately reflected conditions during chlorine contact for settled water plants than for raw water plants, a plausible explanation for the generally better performance of settled water models. Given the much greater importance of this distinction for X_2AAs, results indicate that initial chlorine dose and corresponding precursor conditions are especially relevant for X_2AA formation.

Table 5.1 Summary of chlorine application practices by location of first dose and total number of chlorine addition points in the treatment train

location of first Cl_2 dose	n doses	n	n plants	% of data	mean fraction of total dose at first Cl_2 location
raw water	1	119	28	15%	100%
	2	506	110	64%	61%
	3	157	44	20%	40%
	4	10	3	1%	32%
settled/filtered water	1	211	50	54%	100%
	2	151	35	39%	62%
	3	26	6	7%	37%

In single dose models, the classification variable Cl_2 point was only found to be weakly significant for Cl_3AA and Cl_3AH, with positive coefficients indicating the formation of higher concentrations at raw water plants, controlling for other factors present in the models. The Cl_2 point variable was not significant for other DBP species or TOX in the single dose models. The general lack of importance of Cl_2 point as an influential predictor could stem from the predominantly low organic precursor levels in the data set (chlorine plant TOC 90^{th} percentile = 3.3 mg/L), which was more pronounced when considering the subset of plants using a single chlorine dose (TOC 90^{th} percentile = 2.9 mg/L). It may also be that other variables in the models adequately accounted for any effects of delayed chlorine addition on DBP formation (e.g. UV254, chlorine consumed, and chlorine contact time).

It was noteworthy that single dose models exhibited the same strong patterns with respect to directions of influence and significance levels for bromide, alkalinity, chlorine consumed, TOC, UV254, and pH as were previously observed in models for all chlorine plants, providing validation of those trends (see Chapter 4). Although based on substantially fewer data (n ~ 200 compared to ~ 700-800), models for plants with a single chlorine dose provided much better fits to the data in all cases except for THM species, where little difference was observed. The most dramatic improvement in fit was observed for Cl_2AA, with an increase in adjusted R^2 from 0.5042 (all chlorine plants) to 0.7119 (plants with a single chlorine dose) (see Figure 5.1).

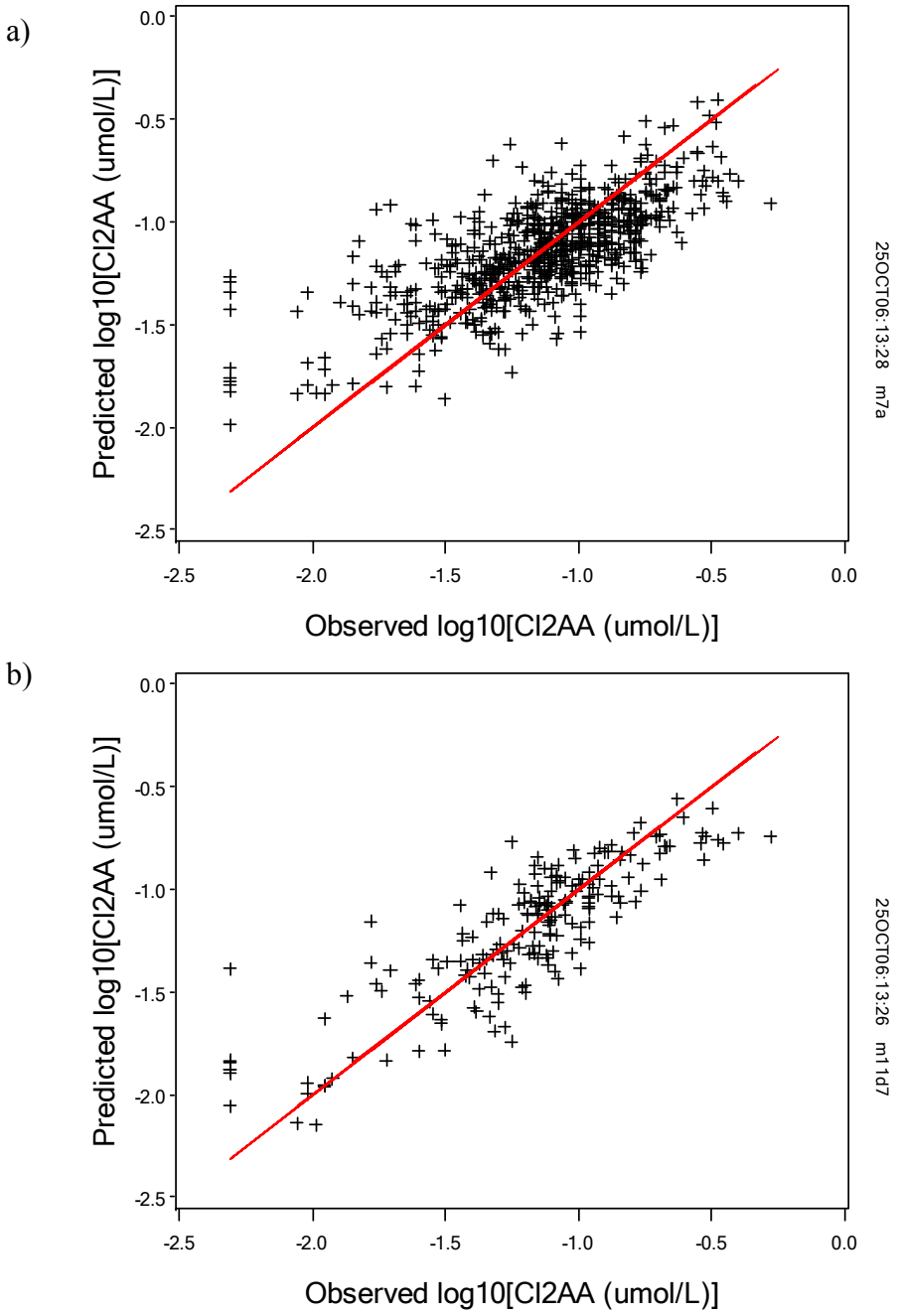


Figure 5.1 Predicted versus observed log10[Cl_2AA (μmol/L)] for models bases on a) all chlorine plant data and b) data for plants with a single chlorine dose, with 1:1 reference line

5.3.2 Alternative Disinfectants

Results for $CHBrCl_2$ are used to illustrate analysis of the impacts of alternative disinfectants on DBP formation. Figure 5.2 shows scatter plots of model predictions versus observed values for logmolar $CHBrCl_2$ concentration, for each disinfectant type (1:1 reference lines superimposed). The chlorine plant $CHBrCl_2$ model included as independent variables influent turbidity, bromide, and temperature, TOC at the first point of chlorine addition, and chlorine consumed through the plant. Points to the left of the reference lines in Figure 5.2 indicate observed concentrations lower than estimated values. Points to the right of the line indicate observed concentrations exceeding estimated values. Figure 5.2a illustrates the model's fit for chlorine plant data, showing that overprediction and underprediction were similar in prevalence and extent. In contrast, Figures 5.2b-5.2d show that observed $CHBrCl_2$ concentrations at plants using chloramines, chlorine dioxide, or ozone were predominantly lower than would be expected under existing water quality and treatment conditions, in the absence of alternative disinfectants. This can be interpreted as indicating that alternative disinfectant use led to a reduction in $CHBrCl_2$ formation. Plot symbols in Figures 5.2b-5.2d differentiate disinfectant subcategories: chloramines used with or without free chlorine; chlorine dioxide used in conjunction with chloramines and/or free chlorine; ozone used in conjunction with chloramines and/or free chlorine. For example, the data in Figure 5.2c shows that $CHBrCl_2$ levels were consistently and significantly lower than expected only when chlorine dioxide was used without free chlorine (i.e. ClO_2/ClM). For chlorine dioxide plants using free chlorine (with or without chloramines), $CHBrCl_2$ levels tended to be close to or slightly lower than expected for chlorine plants.

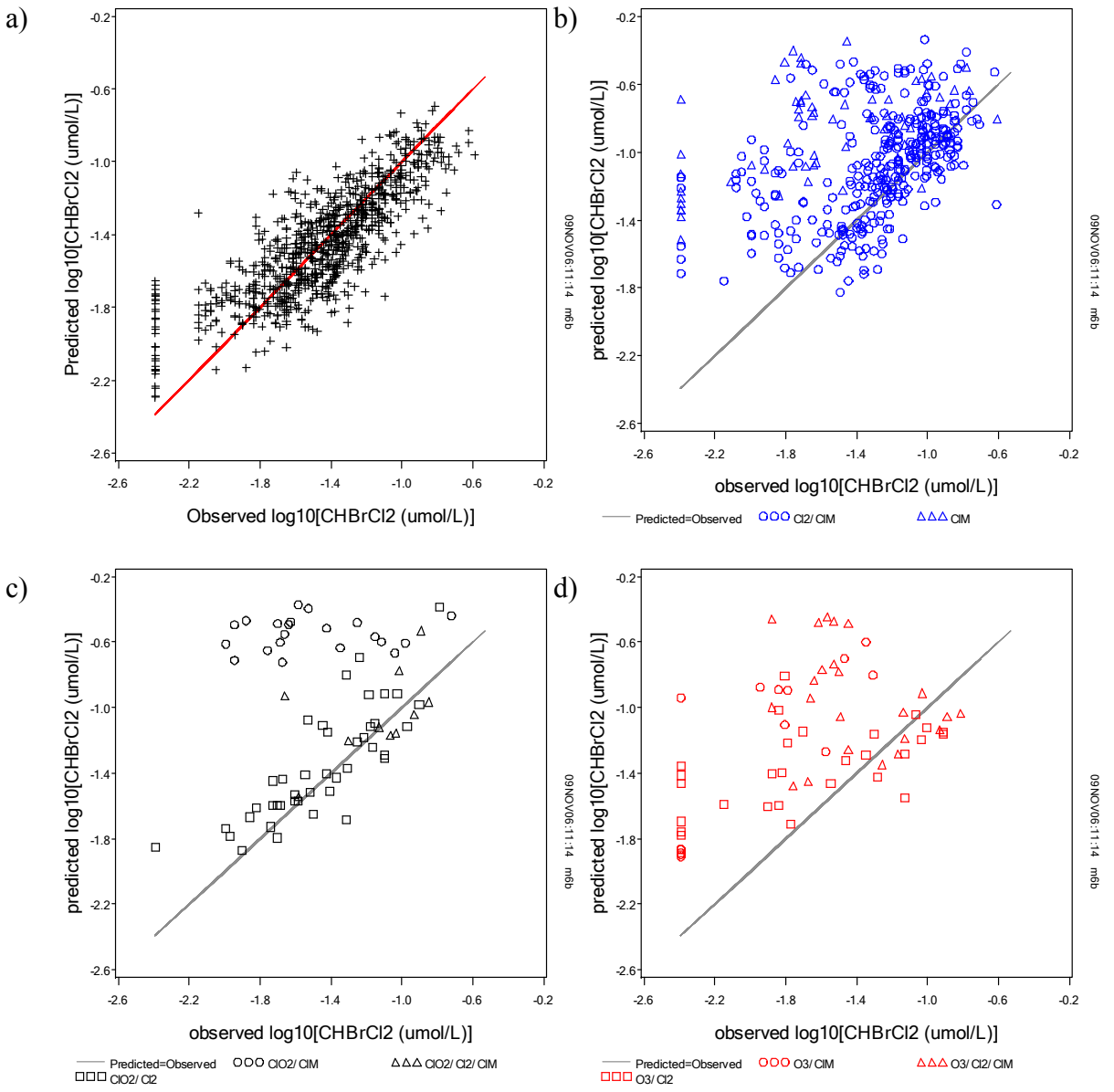


Figure 5.2 Predicted Log10[$CHBrCl_2$ (µmol/L)] from chlorine plant model versus observed values, with 1:1 reference line: results for data from a) chlorine plants, b) chloramine plants, c) chlorine dioxide plants, and d) ozone plants

Histograms of the distributions of RPD values from Figure 5.2 are shown in Figure 5.3, without distinction for disinfectant subcategories. Note that models were developed using log transformed dependent variables (see Figure 5.2), but RPD values were computed from back-transformed results, to allow for more straightforward interpretation. Back-transformation caused the skewness in RPD distributions seen in Figure 5.3. Nevertheless, the shift towards negative RPD values for alternative disinfectants is strongly evident. Results indicate that use of chloramines (Figure 5.3b) and chlorine dioxide (Figure 5.3c) had similar overall impacts on $CHBrCl_2$ formation, lowering finished water concentrations by 29 and 34%, respectively, on average, compared to expected values for use of chlorine alone. Ozone resulted in larger and more consistent $CHBrCl_2$ reduction (Figure 5.3d), with a mean RPD of −45%.

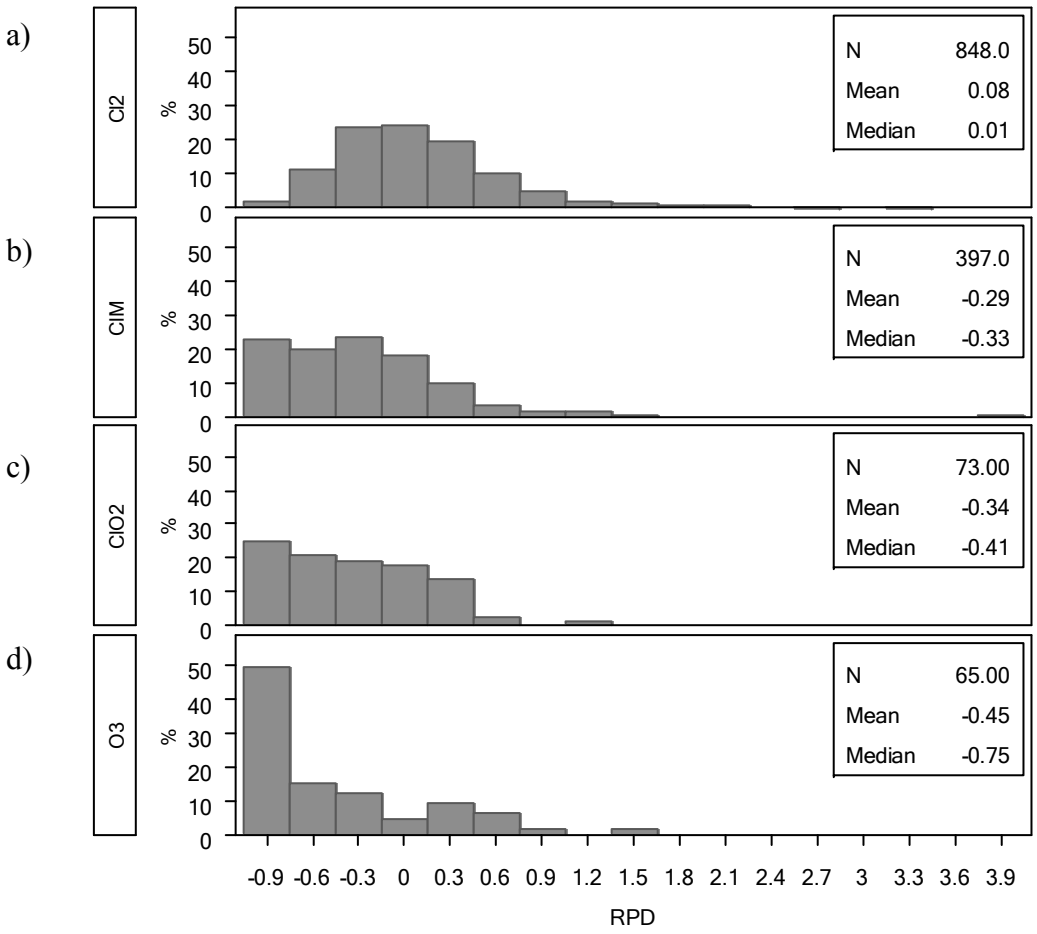


Figure 5.3 Distributions of relative percent difference between observed $CHBrCl_2$ concentrations and values predicted from chlorine plant model: data for a) free chlorine plants, b) chloramine plants, c) chlorine dioxide plants, and d) ozone plants (see Figure 2)

Results for all analytes are summarized in Table 5.2, showing mean RPD values for each disinfectant category and subcategory. Effects for alternative disinfectants should be evaluated in relation to results for free chlorine plants, which comprise the baseline for this analysis. As shown in the first row of Table 5.2, mean RPD values for free chlorine plants ranged from +2.7% to +14.6%, reflecting the slight tendency of models towards underprediction at high values (see Figure 5.2a and Chapter 4). With the exception of TOX and $BrCl_2AA$, the overall impacts of chloramines and chlorine dioxide were comparable for each analyte, based on mean RPDs for the disinfectants (compare second and third rows of Table 5.2). Reductions on the order of 20-30% were observed for THM species and Cl_3AH, with smaller effects for BrClAA, and negligible effects for Cl_2AA. The mean reduction in $BrCl_2AA$ formation was much greater for chlorine dioxide than for chloramines whereas the opposite was true for TOX. For most DBPs, ozone led to substantially greater overall reductions than either chloramines or chlorine dioxide, on average (see fourth row of Table 5.2). In marked contrast to all other analytes and alternative disinfectants, results indicated that Cl_3AH formation was enhanced by use of ozone. It should be noted here that all ICR plants utilizing chlorine dioxide or ozone applied these oxidants in conjunction with free chlorine and/or chloramines (see below).

Table 5.2 Mean relative percent difference between predicted and observed values for DBP species (as μmol/L) and TOX (as μg/L as Cl) by disinfection categories and subcategories

Disinfectant	N	$CHCl_3$	$CHBrCl_2$	$CHBr_2Cl$	Cl_2AA	BrClAA	Cl_3AA	$BrCl_2AA$	Cl_3AH	TOX
Cl_2 only	697-848	+9.7%	+8.1%	+2.7%	+11.9%	+4.3%	+14.6%	+4.7%	+9.5%	+12.7%
any ClM	368-409	-30.6%	-29.3%	-21.4%	+8.7%	-14.8%	-16.5%	-16.9%	-34.5%	-33.4%
any ClO_2	64-75	-25.1%	-34.5%	-32.5%	+5.4%	-17.4%	-22.1%	-37.3%	-32.6%	-5.6%
any O_3	51-65	-45.9%	-45.0%	-14.9%	-26.3%	-49.8%	-61.6%	-51.9%	+18.6%	-28.1%
Cl_2/ClM	297-340	-21.9%	-21.5%	-15.8%	+15.4%	-7.1%	-7.8%	-9.5%	-27.5%	-28.7%
ClM	69-76	-67.9%	-62.7%	-46.4%	-20.3%	-48.3%	-54.1%	-46.0%	-65.1%	-56.5%
ClO_2/Cl_2	36-44	-0.3%	-15.0%	-14.1%	+16.3%	+2.1%	-15.1%	-28.5%	-1.2%	+25.9%
ClO_2/Cl_2/ClM	9-10	-14.4%	-10.8%	+27.3%	+16.3%	+3.5%	+45.8%	+22.1%	+8.4%	-28.4%
ClO_2/ClM	18-23	-74.7%	-84.8%	-79.5%	-19.7%	-68.1%	-67.6%	-83.6%	-92.3%	-55.5%
O_3/Cl_2	18-28	-43.5%	-29.4%	+12.0%	-49.1%	-42.1%	-65.1%	-59.5%	+11.4%	+8.7%
O_3/Cl_2/ClM	21-24	-30.4%	-42.8%	-30.9%	-3.6%	-49.4%	-51.1%	-61.4%	+78.7%	-44.5%
O_3/ClM	12-15	-79.2%	-80.0%	-42.5%	-20.4%	-60.2%	-74.5%	-27.1%	-70.4%	-59.3%

Analysis of results within alternative disinfectant subcategories suffered from small sample size (especially for chlorine dioxide and ozone subcategories), and associated greater uncertainty. Nevertheless it is important to note the substantial variation associated with differences in oxidant/disinfectant application. As shown in Table 5.2 and Figure 5.4, reductions on the order of 50-70% were observed for all DBPs except Cl_2AA when chloramines were employed alone (ClM), compared to corresponding reductions of only 7-30% when free chlorine was applied with chloramines (Cl2/ClM). With conjunctive use of free chlorine, chloramines appeared to have no effect on Cl_2AA formation, whereas chloramines used alone resulted in a moderate 20% reduction in Cl_2AA, on average. With the possible exception of TOX, the benefits of chlorine dioxide appeared to be limited primarily to plants employing chlorine dioxide with no conjunctive free chlorine use (ClO_2/ClM). Use of free chlorine in conjunction with chlorine dioxide appeared to enhance formation of several analytes ($CHBr_2Cl$, Cl_3AA, $BrCl_2AA$), although results were highly variable among the DBP species and must be interpreted with caution due to the sparseness of data (see Table 5.2). As with chloramine and chlorine dioxide, ozone generally led to much greater reductions in formation of the halogenated DBPs studied here when no free chlorine was utilized (O_3/ClM), as compared to use of ozone together with free chlorine (O_3/Cl_2 or O_3/Cl_2/ClM). However, unlike chlorine dioxide, there was little evidence of ozone enhancing formation of these DBPs, with the exception of Cl_3AH. The previously noted overall mean increase in Cl_3AH for ozone treatment was attributable to a small subset of plants in the O_3/Cl_2/ClM category (see Table 5.2 and Figure 5.4). Although most of the seven plants in this category had substantially lower than expected Cl_3AH concentrations, data for two plants with extreme RPD values between +170% and +650% skewed both the subcategory and

overall ozone category means. These two plants applied chlorine following preozonation of high TOC raw waters. Thus, the potential for ozone to enhance Cl_3AH formation may be associated only with a particular set of treatment conditions.

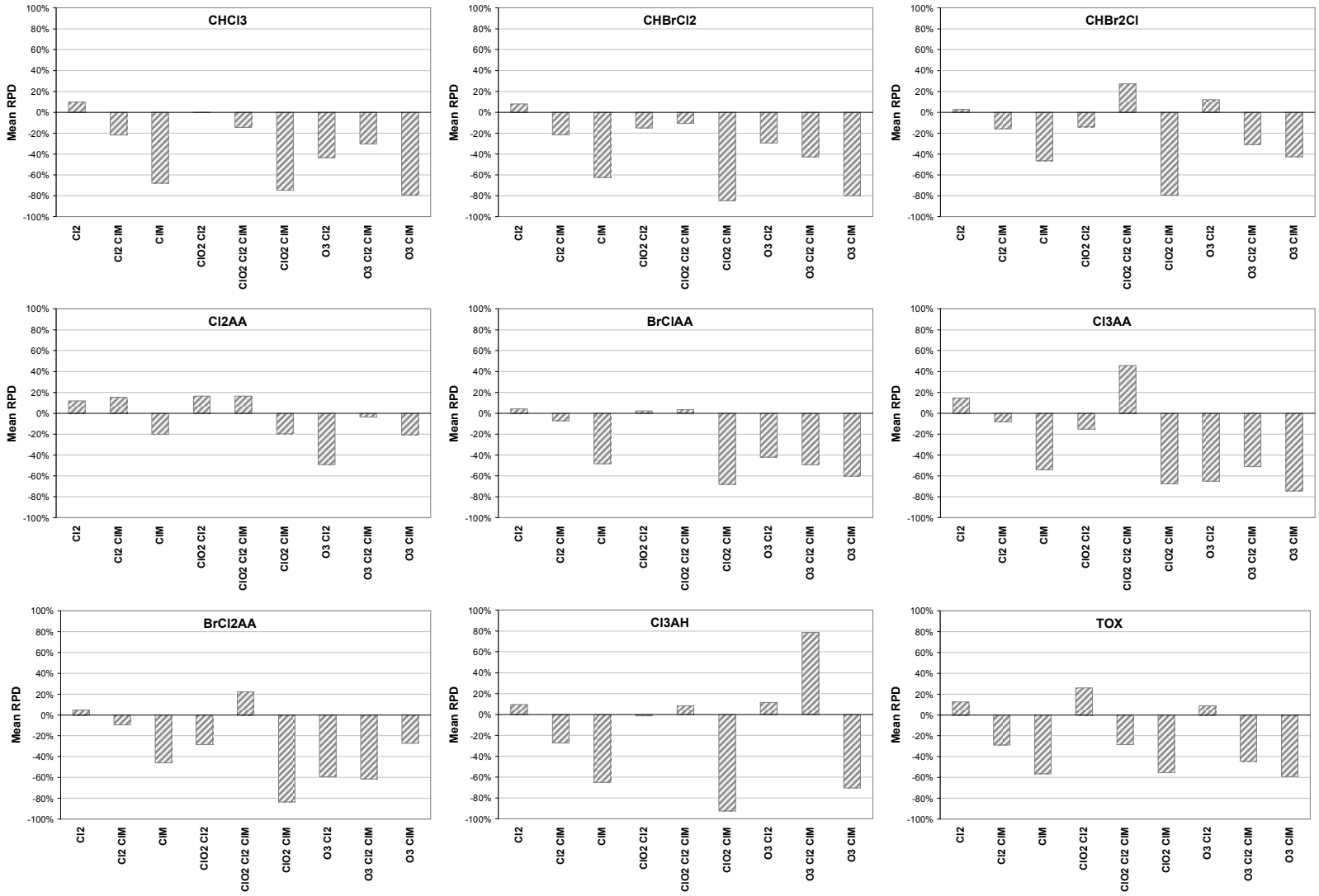


Figure 5.4 Mean relative percent difference between observed DBP concentrations and chlorine plant model predictions, by disinfectant category (see Table 5.2 for *N* values)

The generally greater impact on DBP formation of ozone, compared to chlorine dioxide treatment, can be related to differences in application practices for the two oxidants. For the ICR plants in question, chlorine dioxide was used primarily as a preoxidant at low doses whereas ozone was used in several configurations including pre- and/or post-ozonation, with much higher doses. Statistics listed in Table 5.3 show that, on an equivalent basis, the overall mean ozone dose (0.12 meq/L) was an order of magnitude greater than the mean chlorine dioxide dose (0.013 meq/L). Although ICR data collection preceeded implementation of the national Maximum Contaminant Level for chlorite, chlorine dioxide doses were probably constricted by the understanding that approximately 70% of applied chlorine dioxide is reduced to chlorite (Werdehoff and Singer 1987; Lykins and Griese 1986). Table 5.3 also shows that plants using no free chlorine (i.e. ClO_2/ClM or O_3/ClM) applied higher oxidant doses than those using free chlorine in conjunction with chlorine dioxide or ozone. Thus, higher alternative oxidant doses were generally associated with greater control of DBP formation.

Table 5.3 Summary of ozone and chlorine dioxide doses: mg/L [meq/L]

	n *	n plants	mean	median	min	max
ClO_2/Cl_2	68	16	0.8 [0.011]	0.5 [0.007]	0.1 [0.001]	2.1 [0.031]
ClO_2/Cl_2/ClM	10	4	0.8 [0.012]	1.0 [0.015]	0.2 [0.003]	1.3 [0.019]
ClO_2/ClM	42	9	1.1 [0.017]	1.1 [0.016]	0.5 [0.007]	2.0 [0.030]
Any ClO_2	120	29	0.9 [0.013]	1.0 [0.015]	0.1 [0.001]	2.1 [0.031]
O_3/Cl_2	40	10	1.8 [0.08]	1.5 [0.06]	0.4 [0.02]	5.0 [0.21]
O_3/Cl_2/ClM	26	7	3.1 [0.13]	2.5 [0.11]	0.5 [0.02]	7.0 [0.29]
O_3/ClM	16	3	5.0 [0.21]	3.3 [0.14]	1.2 [0.05]	17.0 [0.71]
Any O3	82	20	2.9 [0.12]	1.7 [0.07]	0.4 [0.02]	17.0 [0.71]

* DBP results not available for all sampling months

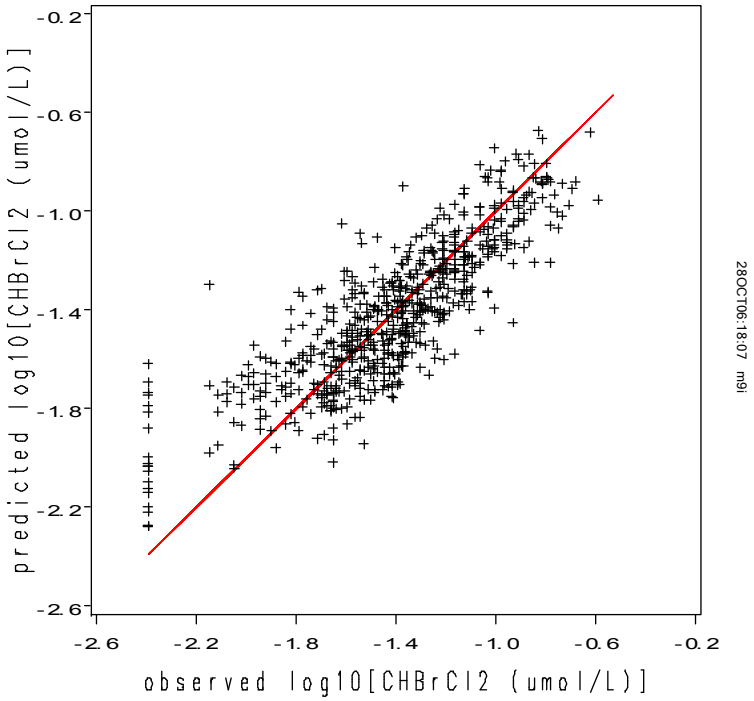
Both ozone and chlorine dioxide treatment are expected to lower specific NOM reactivity towards halogenated DBP formation by oxidizing precursor sites without significantly diminishing TOC concentration. Thus, for a given input TOC value (i.e. TOC at the first point of chlorine addition), lower DBP yields would be expected following exposure to these oxidants, and the magnitude of this impact would increase with oxidant dose. For plants using these oxidants with no free chlorine, substantial fractions of organic chlorine demand would be consumed without significant changes in TOC or formation of chlorinated DBPs. Models should have accounted for reduced chlorine demand through the chlorine consumed variable, athough yields on the basis of TOC (i.e. model slope for TOC) would be expected to decrease, accounting in part for the lower than expected DBP concentrations. Since ozone and chlorine dioxide may both react with source water bromide to generate hypobromous acid, the potential for bromine substitution reactions would remain, even in the absence of free chlorine. This may explain why $CHBr_2Cl$, the only dibrominated compound

studied here (no tribrominated compounds were studied), was the also the only DBP other than Cl_3AH which increased, on average, for any ozone treatment category (see O_3/Cl_2 in Table 5.2 and Figure 5.4).

5.3.3 Softening Treatment

Results of the analysis of softening treatment's impacts on DBP formation are illustrated in Figure 5.5 for $CHBrCl_2$. The fit of the conventional plant $CHBrCl_2$ model is shown in Figure 5.5a. Figure 5.5b shows that, for softening chlorine plants, observed $CHBrCl_2$ concentrations were predominantly lower than values predicted from the conventional plant model, with no apparent distinction with respect to source water type. Results indicate that softening treatment was associated with substantial reduction of $CHBrCl_2$ formation, characterized by a mean RPD of −31%. Results for all analytes are summarized in Figure 5.6 which shows mean RPD values side by side for conventional and softening plants, indicating that softening treatment resulted in moderate to large relative reductions in $CHBrCl_2$, $CHBr_2Cl$, BrClAA, Cl_3AA, $BrCl_2AA$, and Cl_3AH concentrations, but had little effect on $CHCl_3$, Cl_2AA, or TOX. One plausible explanation for softening treatment's impact in generally lowering DBP concentrations could be related to the effect of coprecipitative removal of reactive NOM precursors due to the affinity of higher molecular weight hydrophobic NOM for complexation with calcium precipitates (Randtke 1998). Before further consideration of these findings, some discussion of the models used for this analysis is warranted, especially with regard to the role of pH.

a)



b)

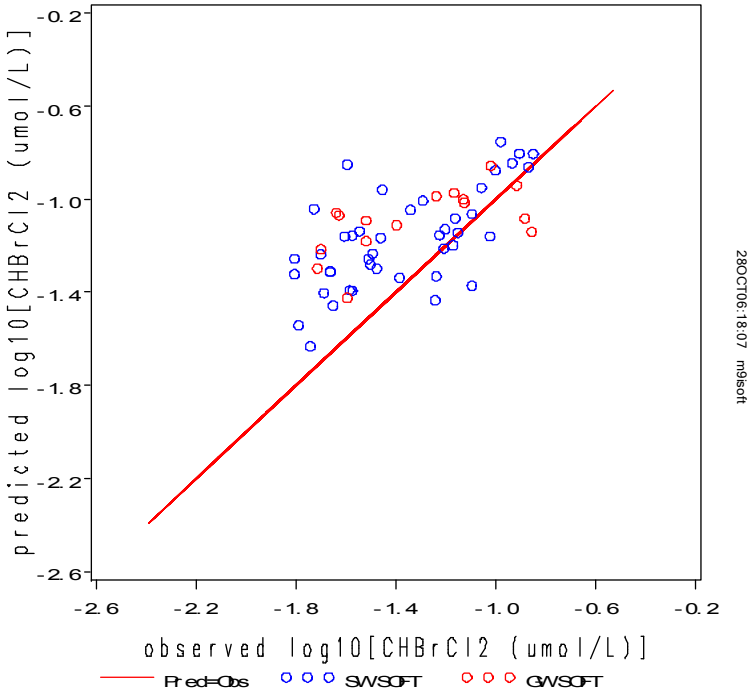


Figure 5.5 Predicted Log10[$CHBrCl_2$ (µmol/L)] values from conventional chlorine plant model versus observed values for a) conventional chlorine plants and b) softening chlorine plants, with 1:1 reference line

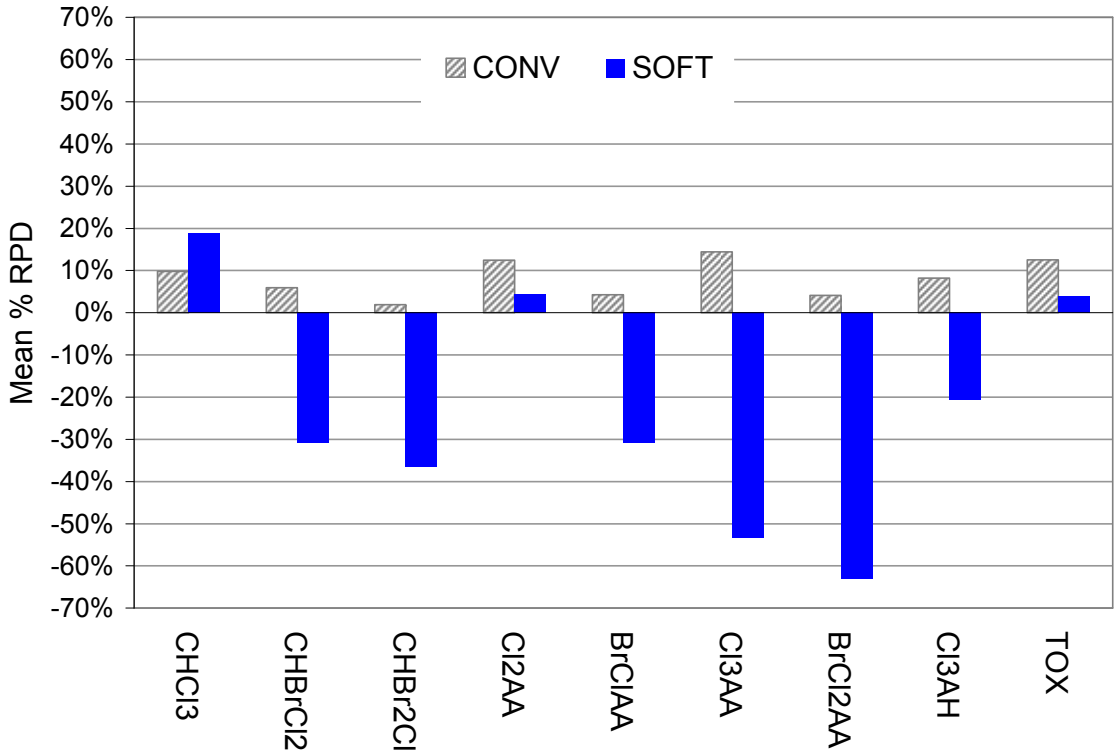


Figure 5.6 Mean relative percent difference between observed DBP concentrations and conventional chlorine plant model predictions for conventional chlorine plants (CONV) and softening chlorine plants (SOFT)

There was some concern that pH effects of softening treatment might not have been adequately captured in models based on conventional plant data because relatively few of those data represented the high pH conditions prevalent in softening treatment. If significant pH effects were not captured in the model, they might account for the apparent effects of softening, illustrated in Figure 5.6. In the chlorine plant models, which were based on data from both conventional and softening plants, pH was identified as a highly significant variable for $CHCl_3$ (positive effect) and $BrCl_2AA$ (negative effect), with a moderate influence on Cl_3AA (negative effect) (see Chapter 4). However, for the conventional plant

models used as the baseline for softening analysis here, pH was identified only as a weakly significant variable in $CHCl_3$, $CHBrCl_2$, and Cl_3AH models (all positive effects), but was not present in either of the X_3AA species models. A comparison of pH sampling distributions for conventional and softening chlorine plants shows that less than 10% of data from conventional plants exceeded pH 9, whereas almost 50% of the data for softening plants were in this high range (see Figure 5.7). This suggests that the conventional plant data may not have provided adequate leverage to detect the influence of pH on X_3AA species formation in corresponding models. An unaccounted for negative influence of high pH on Cl_3AA and $BrCl_2AA$ formation could explain the significant apparent reduction of these species associated with softening treatment (see Figure 5.6). However, the same reasoning cannot not be extended to explain the apparent effects of softening on $CHBrCl_2$, $CHBr_2Cl$, BrClAA, and Cl_3AH, since pH was either accounted for in the conventional plant models for these species or not expected to exert a strong influence.

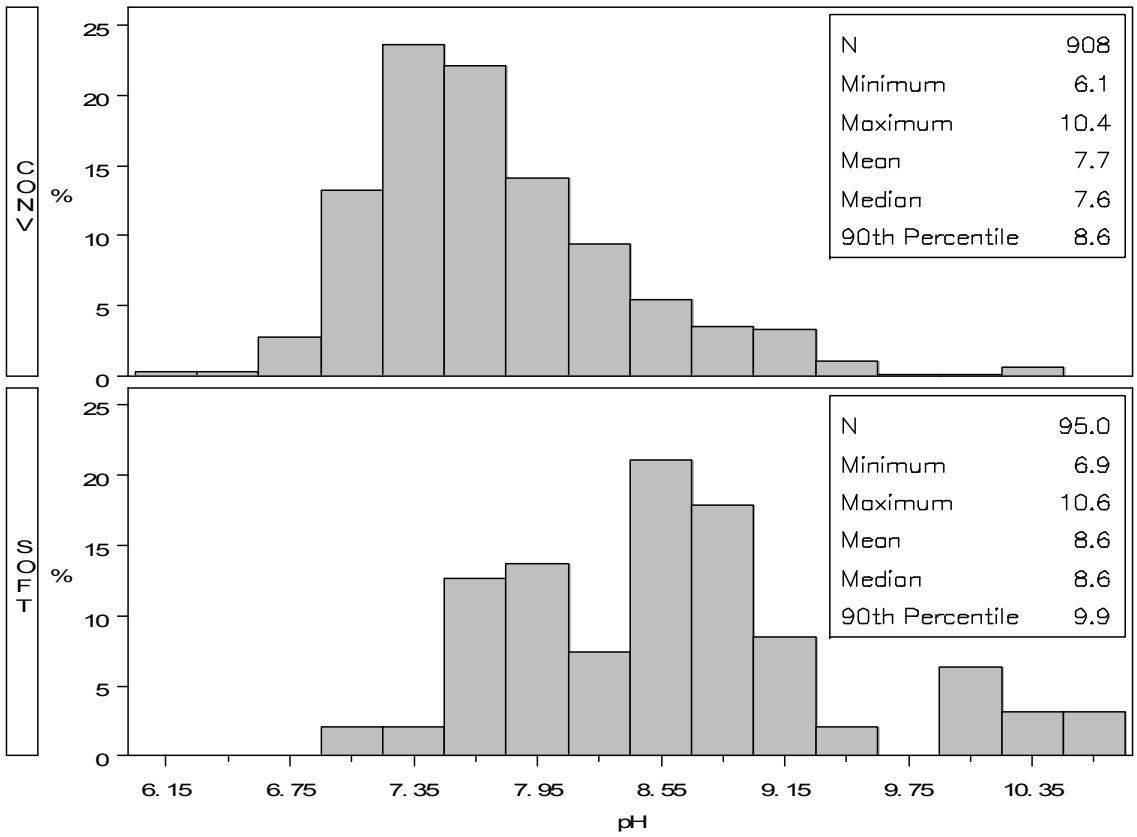


Figure 5.7 Distributions of finished water pH at conventional and softening chlorine plants

A related issue concerning pH effects being sufficiently captured in the conventional plant models was the choice of location type in the treatment train from which to obtain representative pH data, as pH levels change through the treatment train. Since it was impractical to uniquely extract, for each treatment plant and sampling month, the most relevant pH value with respect to chlorine exposure, a decision had to be made about which location yielded the most appropriate value considering the data set overall (e.g. influent, settled water, first point of chlorine addition, etc.). On the basis of an analysis of pH adjustment practices, finished water was judged to best represent pH conditions during

chlorine contact in ICR chlorine plants, overall, so this value was used consistently for model input (see Chapter 4). However, conventional and softening plants differed fundamentally with respect to pH adjustment between the initial chlorine application point and finished water. For softening plants, finished water pH was always either equal to or lower than pH at the first point of chlorine addition: on average the difference was 0.6 pH units, ranging up to 3 pH units. Thus, even with pH present in the model, actual pH conditions during chlorine contact in softening plants would tend to be more alkaline than represented by the finished water value used for model input. On the other hand, for conventional plants, upward and downward pH adjustments between the first point of chlorine addition and finished water were of similar frequency, with typically larger upwards adjustments (mean 0.8, max 4 pH units) than downward adjustments (mean 0.5, max 5 pH units). As a sensitivity test for the choice of pH location in the models, an alternative set of models was developed using pH data from the location of first chlorine dose, instead of finished water. In these models, pH was not identified as a significant variable for any of the DBPs examined, regardless of whether the models were based on data for all chlorine plants or just for conventional chlorine plants. This provided a validation of the choice of finished water location as the preferred alternative for pH model input data.

Given the concern over whether conventional plant models adequately represented important pH effects, especially for the X_3AA species, the impacts of softening on DBP formation were evaluated using a second approach. This involved testing a binary classification variable (1 for softening, 0 for conventional) for statistical significance in the original chlorine plant models, which included pH effects for $CHCl_3$, Cl_3AA, and $BrCl_2AA$. These tests indicated that softening treatment had a highly significant negative influence on

$CHBr_2Cl$, BrClAA, Cl_3AA, and $BrCl_2AA$ formation, a marginally significant negative influence on $CHBrCl_2$, and Cl_3AH formation, and no effect on $CHCl_3$, Cl_2AA, or TOX. These findings were consistent with those described above in conjunction with Figure 5.6, lending credence to the hypothesis that softening treatment has beneficial impacts on DBP reduction through NOM removal, apart from effects strictly related to treatment pH.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

This dissertation presents results of statistical analyses of the Information Collection Rule database addressing DBP occurrence patterns and their relationships with water quality conditions and treatment practices. Review, screening, and descriptive summary of the database were also conducted to support the research goals.

6.1 CONCLUSIONS

A new metric to quantify the extent of specific halogen substitution in any byproduct class was developed and used to examine bromine fractions of total halogen in each of four byproduct classes (THMs, X_2AAs, X_3AAs, and X_2ANs (DHANs)). This metric facilitates direct interclass comparisons of halogen speciation patterns because it is normalized the same way for each byproduct class and ranges from zero to one, regardless of the number of halogen substituents in that class. In each class examined, bromine fraction ranged widely across the data set overall, and within subsets of data for narrow bromide ranges. However, on an individual sample basis, the four bromine fractions were strongly interdependent, indicating a high degree of redundancy in the 14 individual compounds' monitoring data. Bromine fractions in the X_2AA and THM classes were nearly identical whereas bromine fractions in the X_3AA and X_2AN classes were 10% lower and 60% higher, on average, than in the THM class, respectively. The lower extent of bromine substitution in the X_3AA class is

hypothesized to be related to brominated X_3AA species stability and/or analytical methods issues, whereas the higher extent of bromine substitution in X_2AN is hypothesized to be related to faster formation kinetics of the more basic nitrogen-containing organic substrates. The bromine fraction measurements, particularly for the X_3AA class, were significantly impacted by censored data handling because of higher MRLs and lower occurrence levels for brominated compounds in the ICR data. Correlation among the four bromine fractions motivated their treatment as a multivariate response. Application of a test for multivariate outliers successfully exposed inconsistencies associated with data entry or analytical error.

ICR database screening and review resulted in flagging questionable data for exclusion from subsequent analyses and recovery of a large number of missing categorical variable records, affording a more reliable and complete data set for subsequent study. The sparseness of flagged data indicated a high level of ICR data quality for information relevant to this research. A summary of data patterns revealed strong associations between disinfection practices and source water quality conditions: plants with high concentrations of organic precursors preferentially employed chloramines and avoided prechlorination; plants with high bromide levels also tended to employ chloramines although bromide did not impact prechlorination practice. Though plants employing chloramination used significantly higher chlorine doses than plants using only free chlorine, when normalized to TOC this difference largely disappeared. The median chlorine to TOC ratio in the ICR data set was 1.54 [mg Cl_2]/[mg C]. Applied chlorine to ammonia-nitrogen ratios at chloramine plants varied widely but the median value was near the theoretical 1:1 molar ratio for monochloramine formation. Significantly higher bromide to TOC ratios at ground water

plants, compared to surface water plants, resulted from the typically lower TOC and higher bromide levels in ground waters.

Multiple linear regression models for finished water DBPs at chlorine plants indicated significant changes in the direction and magnitude of the influence of bromide, alkalinity, chlorine consumed, and organic precursor concentrations (TOC and UV254) on concentrations of individual DBP species across compound classes. Results suggested that alkalinity serves as an indicator of NOM hydrophobicity and reactivity towards DBP formation. In agreement with previous research, pH was found to have a positive influence on $CHCl_3$ formation, a negative influence on X_3AA species formation, and no significant impact on X_2AA species. The large differential impacts of pH observed across species within the THM and X_3AA classes in these regression models have generally not been noted in previous studies. Modeling results for aggregate DBP class sums and TOX concentrations reflected the dominance of the data set by chlorinated DBPs, suggesting that inferences from models for these types of aggregate variables are less generally applicable than inferences from models for specific individual compounds.

A degree of misspecification in some of the regression models was traced, in part, to effects of multiple chlorine dose applications. This impacted Cl_2AA most strongly, such that model precision increased almost 50% when data were restricted to those for plants using a single chlorine dose. Results suggest that chlorine consumed in subsequent stages of treatment (after initial chlorine dose) are less relevant for X_2AA species formation than for other DBPs examined. After controlling for the number of applied chlorine doses in the treatment train, the location of chlorine addition (raw or settled/filtered water) did not have a strong influence on DBP concentrations, based on models that accounted for TOC, UV254, chlorine consumed, and chlorine contact time.

Use of chloramines, chlorine dioxide, and ozone were each shown to result in substantially lower DBP concentrations relative to levels projected from models developed for plants using free chlorine alone under otherwise identical water quality and treatment conditions. For most DBPs examined, these differences were greatest for ozone, which may be related more to treatment practices than to intrinsic oxidant properties (ozone doses were generally much higher than chlorine dioxide doses). A notable exception was Cl_3AH which was shown to be elevated under certain ozone treatment conditions, namely application of chlorine following preozonation of high TOC waters.

For softening treatment, many of the DBP species studied were found to be much lower than projected using models developed for conventional treatment, other conditions being equal. Significant impacts were observed for brominated THM and X_2AA species and for all X_3AA species. TOX appeared to be unaffected, which may be explained by the relatively small contribution of the affected species to TOX in this dataset. It was hypothesized that the effect of softening treatment on limiting DBP formation is related to improved organic precursor removal.

6.2 RECOMMENDATIONS

Findings of this research point to several recommendations in the area of DBP monitoring, data analysis, and regression modeling. The examination of halogen substitution patterns by graphic inspection of arcsine-transformed bromine fraction results is recommended as a simple screening step for quality assurance when more than one class of DBP monitoring data is collected. The concept of parallel halogen substitution patterns should be further tested as data are obtained for other classes of DBPs such as tri- and di-

haloacetaldehydes and halopicrins, which have not been widely monitored to-date. Censored data handling methods may significantly impact inferences from data analysis when moderate to large fractions of data are below MRL, and study reports should state the methods employed for processing censored data as a matter of course. This has not generally been practiced in DBP research to date, and is a particular concern for studies addressing HAA9 in view of the different MRLs for many of the species. Methodological problems associated with brominated X_3AA species analysis must be considered in interpretation of existing data and it is strongly recommended that recent advances in method development be adopted for future studies so that better information about occurrence and formation of this important class of DBPs can be obtained (Domino et al. 2004).

Whereas most of the variables found to be important predictors of DBP formation in this research are usually obtained in laboratory and field studies, source water alkalinity, though usually available, has not generally been utilized as an indicator of NOM characteristics with respect to DBP formation. It is recommended that future studies routinely consider alkalinity together with TOC and UV254 as indicators of organic precursor characteristics.

Regression modeling has been shown to be a valuable tool for analysis of full-scale water treatment plant data. However, important treatment factors may be difficult to represent explicitly as model variables so that their effects must be controlled for by subsetting data. This may lead to small data sets that lack the power to demonstrate significant trends, or limited data ranges for important variables. The latter concern stems from the fact that inherent relationships exist between water quality and treatment. For example, plants using only free chlorine for disinfection tend to have a narrow range of

source water TOC concentrations, and plants using conventional treatment generally don't operate under the high pH conditions used in softening. Significant effects may become non-discernable without sufficient leverage in the design matrix. No simple solution to these challenges can be recommended here, but awareness of these issues should inform study design. Finally, whether regression models are developed for the purpose of prediction or as tools for data analysis, validation is essential to ensure that the models developed describe inherent relationships among variables and do not merely provide a best fit to the particular data set at hand. Accordingly, appropriate validation procedures must be considered as part of any overall study design.

APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 2

A.1 ADDITIONAL EXAMPLES OF MULTIVARIATE OUTLIER DETECTION TEST RESULTS

Figure A.1 shows transformed bromine fraction data for distribution system samples from 5 treatment plants. Identified multivariate outliers (2) are indicated by star symbols. DBP species data for the plant/sampling-periods with outliers are shown in Tables A.1 and A.2. Both outliers were due to unexpectedly high bromine fraction in the DHAA class. Data review suggests that the below minimum reporting level result for DCAA in outlier 1.1 (location D) is in error, causing an elevated DHAA bromine fraction value for this sample (see Table 1). For the second outlier, the reported DBAA result of 206 μg/L for outlier 1.2 (location A) almost certainly represents a data entry error. A value of 2.6 μg/L would be commensurate with results for other locations from the same plant/sampling-period (see Table A.2).

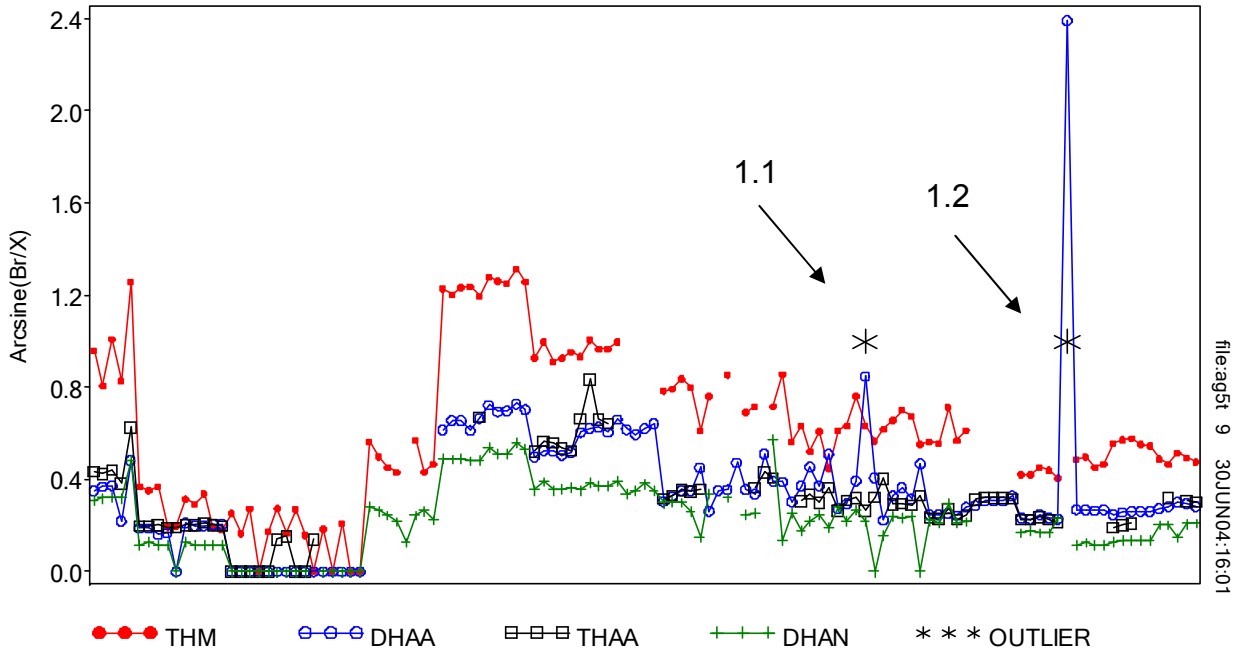


Figure A.1 Arcsine transformed bromine fraction data for 5 plants showing detected outliers; N=121, Br <20-150 µg/L (median= <20 µg/L), TOC 0.4-5.8 mg/L (median= 1.7 mg/L), Br/TOC 0.003-0.171 mg/mg (median= 0.011 mg/mg)

Table A.1 Distribution system sample DBP species concentration data (µg/L) for plant/sampling-period of outlier 1.1 in Figure A.1; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A	31.0	11.0	2.5	0.0	14.0	3.2	0.0	14.0	3.6	0.0	0.0	2.6	1.2	1.0
B	41.0	12.0	2.5	0.0	17.0	3.7	0.0	19.0	4.9	0.0	0.0	2.6	1.0	0.5
C	45.0	15.0	5.1	0.0	17.0	4.9	1.0	18.0	5.2	0.0	0.0	2.3	1.3	0.8
D *	42.0	12.0	2.5	0.0	0.0	17.0	4.4	15.0	3.6	0.0	0.0	2.6	1.0	0.5
E	90.0	10.0	1.9	0.0	62.0	7.1	0.0	36.0	5.4	0.0	0.0	0.7	0.0	0.0

Table A.2 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 1.2 in Figure A.1; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A *	47					16.0	206.0	0.0	14.0	3.2				
B	50	29.0	7.5	1.3	0.0	20.0	3.1	0.0	18.0	3.7			1.8	0.5
C	51	34.0	7.9	1.3	0.0	21.0	3.1	0.0	19.0	3.9			1.8	0.6
D	52	25.0	6.5	1.1	0.0	18.0	3.0	0.0	16.0	3.5			1.7	0.5
E	53	32.0	7.2	1.0	0.0	22.0	3.1	0.0	20.0	3.9			1.8	0.5

Figure A.2 shows transformed bromine fraction data for distribution system samples from a second set of 5 treatment plants. Identified multivariate outliers (1 in this example) are indicated by star symbols. DBP species data for the plant/sampling-period with an outlier are shown in Table A.3. The outlier (location A) was due to unexpectedly high bromine fraction in the DHAA class. Data review suggests that reported values for both DCAA and BCAA (60 μ/L and 30 μ/L, respectively) represent data entry errors involving decimal place shift (i.e. analytical result were probably 6.0 μ/L and 3.0 μ/L, respectively).

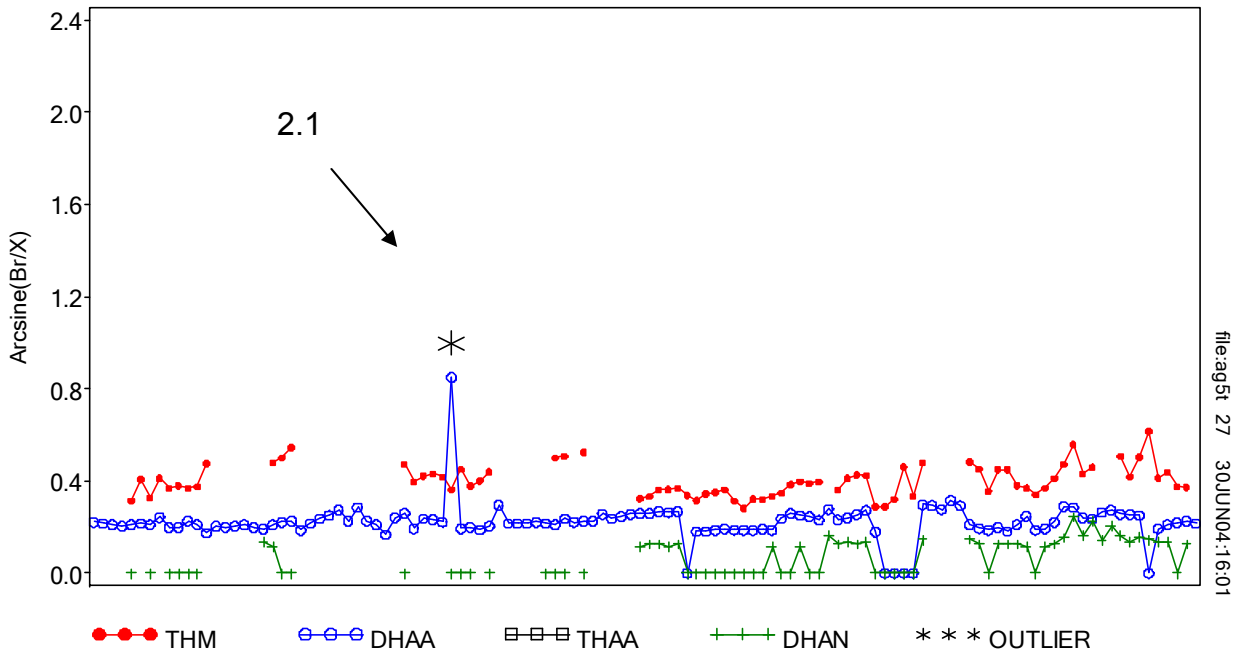


Figure A.2 Arcsine transformed bromine fraction data for 5 plants showing detected outliers; N=118, Br <20-40 µg/L (median= <20 µg/L), TOC 1.3-6.2 mg/L (median= 2.0 mg/L), Br/TOC 0.002-0.011 mg/mg (median= 0.005 mg/mg)

Table A.3 Distribution system sample DBP species concentration data (µg/L) for plant/sampling-period of outlier 2.1 in Figure A.2; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A *	17.0	5.3	0.0	0.0	67.0	30.0	0.0	9.3				1.6	0.0	0.0
B	25.0	6.4	1.1	0.0	9.4	1.6	0.0	8.5				1.2	0.0	0.0
C	23.0	5.8	0.0	0.0	10.0	1.7	0.0	8.0				1.3	0.0	0.0
D	29.0	6.5	0.0	0.0	8.8	1.5	0.0	10.0				0.0	0.0	0.0
E	24.0	6.2	1.0	0.0	12.0	1.8	0.0	9.7				1.4	0.0	0.0

Figure A.3 shows transformed bromine fraction data for distribution system samples from a third set of 5 treatment plants. Identified multivariate outliers (4) are indicated by star symbols. DBP species data for the plant/sampling-periods with outliers are shown in Tables A.4-A.6. The first two outliers, due to unexpectedly high bromine fraction in the DHAA class, are from the same plant/sampling-period (locations A and C, see Table A.4) and appear to involve decimal place data entry errors in BCAA (i.e. analytical results were probably 3.0 μg/L and 3.8 μg/L, not 30 μg/L and 38 μg/L). The third and fourth outliers, due to unexpectedly low THM bromine fraction values of zero, from two sampling periods at the same treatment plant, are both attributable to below minimum reporting level BDCM results that are incommensurate with THM species results for other locations in the same sampling-periods (see Tables A.5-A.6). A third similar anomalous result from this plant was not identified as an outlier because of measurable bromoform (location E, Table A.6). The THM data pattern shown in Table A.6, wherein no dibromochloromethane was reported for samples with measurable bromodichloromethane and bromoform, suggests a general problem with THM data for this plant/sampling-period.

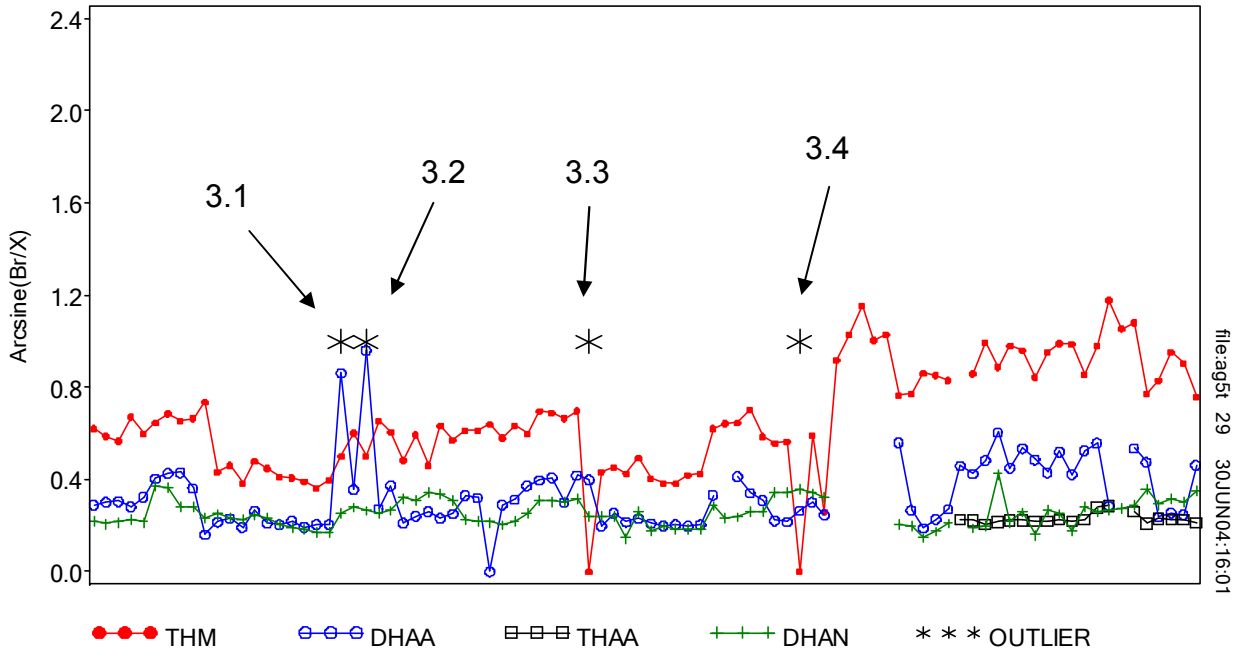


Figure A.3 Arcsine transformed bromine fraction data for 5 plants showing detected outliers; N=90, Br <20-61 µg/L (median= 0.030 µg/L), TOC 2.6-6.7 mg/L (median= 3.2 mg/L), Br/TOC 0.001-0.020 mg/mg (median= 0.020 mg/mg)

Table A.4 Distribution system sample DBP species concentration data (µg/L) for plant/sampling-period of outliers 3.1 and 3.2 in Figure A.3; below minimum reporting level results represented as zero; locations identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A *	60.0	10.2	0.0	0.0	36.0	30.0	0.0	33.0				6.9	2.4	0.0
B	87.6	12.9	1.1	0.0	68.0	3.4	1.3	50.0				7.6	2.9	0.0
C *	65.0	10.1	0.0	0.0	70.0	38.0	0.0	49.0				6.8	2.6	0.0
D	114.0	15.1	1.4	0.0	67.0	3.2	0.0	52.0				7.1	2.4	0.0
E	102.0	13.0	1.2	0.0	72.0	3.9	1.3	51.0				7.2	2.6	0.0

Table A.5 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 3.3 in Figure A.3; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A *	43.0	0.0	0.0	0.0	74.0	2.5	2.7	39.0				4.2	2.1	0.0
B	38.0	7.5	0.0	0.0	68.0	1.7	0.0	34.0				4.1	2.1	0.0
C	42.0	8.2	0.0	0.0	71.0	2.8	0.0	52.0				4.6	2.1	0.0
D	37.0	7.3	0.0	0.0	60.0	2.0	0.0	35.0				4.9	0.8	0.0
E	53.0	9.8	0.0	0.0	72.0	2.3	0.0	40.0				5.4	2.5	0.0

Table A.6 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 3.4 in Figure A.3; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A	42.3	10.1	0.0	1.2	17.8	2.1	0.0	19.9				0.0	2.4	0.7
B	52.9	9.9	0.0	1.5	28.5	2.0	0.0	40.6				0.0	2.5	0.7
C *	59.2	0.0	0.0	0.0	42.9	3.0	0.0	34.7				0.0	2.8	0.7
D	66.9	11.5	0.0	1.3	47.9	3.9	0.0	55.3				0.0	2.4	0.7
E	52.3	0.0	0.0	1.4	42.2	2.6	0.0	42.4				0.0	2.0	0.7

Figure A.4 shows transformed bromine fraction data for distribution system samples from a fourth set of 5 treatment plants. Identified multivariate outliers (3) are indicated by star symbols. DBP species data for the plant/sampling-periods with outliers are shown in Tables A.7-A.9. The first two outliers are from September and December sampling, respectively, at the same ICR plant. The first outlier, due to an unexpectedly low bromine fraction value in the THAA class is attributable to the below minimum reporting level DBCAA result at location E (see Table A.7). The second outlier is attributable to very high THM species concentrations at location D (an “AVG” type ICR location) and probably does

not represent data entry or analytical error (see Table A.8). It is noteworthy that THM concentrations at this location were also high (relative to other locations) in the September sampling round (Table A.7), but not as extreme as in December. The third outlier in Figure A.4, due to an unexpectedly low DHAA bromine fraction value (zero) is attributable to below minimum reporting level results for both brominated DHAA species (BCAA and DBAA) while brominated DBPs occurred at moderate concentrations in the THM and DHAN classes for the same location (see Table A.9).

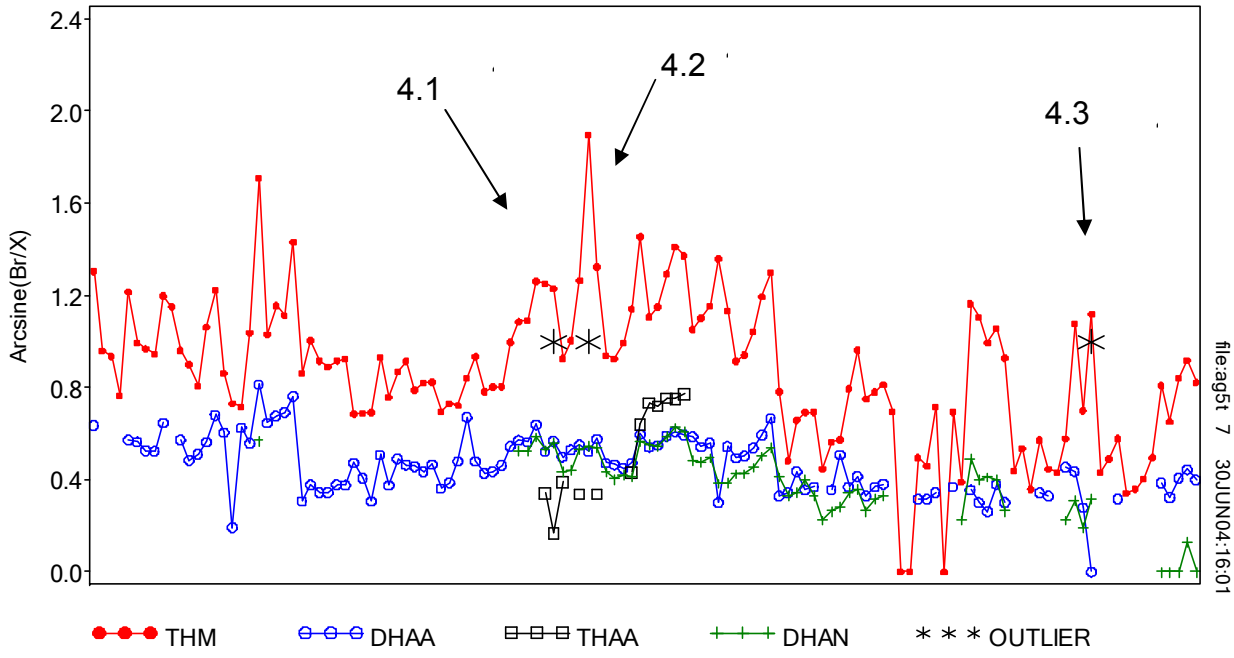


Figure A.4 Arcsine transformed bromine fraction data for 5 plants showing detected outliers; N=128, Br <20-280 µg/L (median= 120 µg/L), TOC <0.7-2.6 mg/L (median= 0.8 mg/L), Br/TOC 0.004-0.743 mg/mg (median= 0.250 mg/mg)

Table A.7 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 4.1 in Figure A.4; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A	0.0	0.0	4.3	11.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	0.6	3.5
B	0.0	0.0	4.5	11.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0	0.6	3.5
C	0.0	1.4	6.5	15.0	0.0	0.0	4.5	0.0	0.0	0.0	0.0	0.0	0.6	4.3
D	3.6	5.4	9.2	15.0	1.8	1.8	4.6	1.7	1.9	2.1	0.0	0.8	1.0	4.3
E *	0.0	2.0	6.8	14.0	0.0	1.0	4.5	0.0	1.3	0.0	0.0	0.0	0.7	4.0

Table A.8 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 4.2 in Figure A.4; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A	0.0	1.0	4.3	7.7	2.0	2.6	3.8	2.4	2.8	2.6	0.0	0.0	0.6	2.4
B	0.0	1.2	5.1	9.0	0.0	0.0	3.1	0.0	0.0	0.0	0.0	0.0	0.6	2.5
C	0.0	2.2	8.6	14.0	0.0	1.5	4.2	0.0	1.1	2.2	0.0	0.0	0.9	3.6
D *	7.0	10.0	15.0	39.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	1.1	3.8
E	0.0	3.3	10.0	15.0	0.0	1.6	4.6	0.0	1.1	2.2	0.0	0.0	0.9	3.7

Table A.9 Distribution system sample DBP species concentration data (μg/L) for plant/sampling-period of outlier 4.3 in Figure A.4; below minimum reporting level results represented as zero; location identified as outlier indicated by asterisk

LOC	CHCL3	BDCM	DBCM	CHBR3	DCAA	BCAA	DBAA	TCAA	BDCAA	DBCAA	TBAA	DCAN	BCAN	DBAN
A	0.0	0.0	0.0	1.6	0.0	0.0	0.0	0.0				0.	0.0	0.0
B	0.0	0.0	1.0	3.2	8.5	4.4	2.4	4.1				0.	0.0	0.5
C	13.0	16.0	13.0	3.0	7.5	3.9	2.3	3.5				2.	1.5	1.2
D	4.4	5.6	5.0	2.1	2.5	1.3	1.1	1.1				0.	0.5	0.5
E *	13.0	17.0	14.0	3.3	2.8	0.0	0.0	0.0				2.	1.6	1.2

A.2 SAS PROGRAM CODE FOR MULTIVARIATE OUTLIER TEST

```
*************************************************************************************************;
* CARONI & PRESCOTT MULTIVARIATE OUTLIER DETECTION METHOD
        MODIFIED FROM PENNELL (2002)
* 4 MACROS DEFINED FOR p=4, 3, 2, 1 CORRESPONDING TO MV DIMENSION
* K= # outliers to check
* ALPHA = experimentwise typeI error
* INPUT = data in
* OUTPUT = data out
* LIBR=library for output file
*************************************************************************************************;

/* p=4 */
%MACRO OUT4(K,ALPHA,INPUT,OUTPUT,LIBR);
        proc iml;
        use &input;
        /* use arcsine transformed variables */
        read all var{thmpctbrt thaapctbrt dhaapctbrt dhanpctbrt} into X; /* X is n x 4 */
        read all var{event_id} into ID; /* ID is n x 1 */
        ALPHA=&ALPHA;
        do j=1 to &K;
                n=nrow(X);
                if j=1 then n1=n; /* n1 is original number of observation */
                XBAR=(X[:,]); /* XBAR is 1 x p, averages across observations */
                ONE=J(n,1,1); /* ONE is n x 1 column vector of 1's */
                XBARN=ONE*XBAR[,1]||ONE*XBAR[,2]||ONE*XBAR[,3]||ONE*XBAR[,4];
                        /* XBARN is n x p */
                A=(X-XBARN)`*(X-XBARN); /* A is p x p */
                do i=1 to N;
                /* N is number of observations at current iteration of the K-loop */
                        Xi=X[i,]; /* i-th row of X */
                        WTEMP=1-((N/(N-1))*(XI-XBAR)*inv(A)*(XI-XBAR)`); /* ith Wilk's lr statistic */
                        if i=1 then W=WTEMP; else W=W//WTEMP;
                                /* W becomes n x 1 col vector of Wilk's */
                        end; /* end i loop*/
                W=ID||X||W; /* W is n x 6 */
                create wilks from W; append from W; close wilks; /* read into data set to sort */
                sort data=wilks by COL6; /* sort so smallest Wilk's statistic is first observation */
                use wilks;
                read all var{COL1 COL2 COL3 COL4 COL5 COL6} into TEMP;
                                /* read data set into n x 6 matrix */
                close wilks;
                STEP=1+(n1-n); /* there will be a total of K steps */
                /* check for tie in smallest Wilk's statistic */
                *TIECHK=TEMP[1:2,1:6];
                *mattrib TIECHK label='2 smallest W'     colname=({ID THM THAA DHAA DHAN W});
                *print TIECHK;
                OUT=TEMP[1,]; /* new 1x6 OUT matrix holds one observation with smallest Wilk's */
                PVAL=probbeta(OUT[1,6],(N-5)/2,2);
                                /* Wilk's statistic p-value from Beta distribution */
                BONALPHA=ALPHA/N;
                CRIT10=betainv(BONALPHA,(N-5)/2,2);
```

```
                        /* crit. value of Beta distr. at bonferoni alpha level */
        OUT=STEP||N||OUT||PVAL||CRIT10||BONALPHA;
                        /* one record for output matrix, 12 variables */
        if j=1 then OUT2=OUT; else OUT2=OUT2//OUT;
                        /* build up output matrix from K rows */
        /* drop observation with smallest (most sign.) Wilk's from data set before next j */
        ID=TEMP[2:N,1]; /* n-1 column vector of IDs*/
        X=TEMP[2:N,2:5]; /* (n-1)x4 */
        end; /* end j loop */
    create &libr..&output from OUT2; append from OUT2; close &libr..&output;
    run; /* run proc iml */
proc sort data=&libr..&output(rename=(COL1=STEP COL2=N COL3=EVENT_ID COL4=THMpctbrt
        COL5=thaapctbrt COL6=dhaapctbrt COL7=dhanpctbrt COL8=W COL9=PVAL
        COL10=CRIT10 COL11=BONALPHA10));
    by descending step;
    run;
data &libr..&output; set &libr..&output;
    format THMpctbrt thaapctbrt dhaapctbrt dhanpctbrt 5.3 pval bonalpha10 9.7 w crit10 8.6;
    retain maxl &output; drop maxl;
    &output=1;
    if _n_=1 then maxl=1;
    if (w>crit10 and maxl) then delete; else maxl=0;
    run;
proc sort data=&libr..&output; by event_id;
    run;
%MEND OUT4;


/* p=3 */
%MACRO OUT3(K, ALPHA, INPUT, OUTPUT, LIBR, VAR1, VAR2, VAR3);
proc iml;
    use &input;
    ** use arcsine transformed variables ;
    read all var{&VAR1 &VAR2 &VAR3} into X; /* X is n x 3 */
    read all var{event_id} into ID; /* ID is n x 1 */
    ALPHA=&ALPHA;
    do j=1 to &K;
        n=nrow(X);
        if j=1 then n1=n; /* n1 is original number of observation */
        XBAR=(X[:,]); /* XBAR is 1 x p, averages across observations */
        ONE=J(n,1,1); /* ONE is n x 1 column vector of 1's */
        XBARN=ONE*XBAR[,1]||ONE*XBAR[,2]||ONE*XBAR[,3]; /* XBARN is n x p */
        A=(X-XBARN)`*(X-XBARN); /* A is p x p */
        do i=1 to N; /* N is number of observations at current iteration of the K-loop */
            Xi=X[i,]; /* i-th row of X */
            WTEMP=1-((N/(N-1))*(XI-XBAR)*inv(A)*(XI-XBAR)`);
                        /* i-th Wilk's likelihood ratio statistic */
            if i=1 then W=WTEMP; else W=W//WTEMP;
                        /* W becomes n x 1 column vector of Wilk's statistics */
            end; /* end i loop*/
        W=ID||X||W; /* W is n x 5 */
        create wilks from W; append from W; close wilks; /* read into data set to sort */
        sort data=wilks by COL5; /* sort so smallest Wilk's statistic is first observation */
        use wilks;
```

```
                read all var{COL1 COL2 COL3 COL4 COL5} into TEMP;
                                        /* read data set back into n x 6 matrix */
                close wilks;
                STEP=1+(n1-n); *print STEP; /* there will be a total of K steps */
                /* check for tie in smallest Wilk's statistic */
                *TIECHK=TEMP[1:2,1:5];
                *mattrib TIECHK label='2 smallest W'     colname={ID &VAR1 &VAR2 &VAR3 W};
                *print TIECHK;
                OUT=TEMP[1,]; /* new 1x6 OUT matrix holds one observation with smallest Wilk's */
                PVAL=probbeta(OUT[1,5],(N-4)/2,1.5);
                                /* Wilk's statistic p-value from Beta distribution */
                BONALPHA=ALPHA/N;
                CRIT10=betainv(BONALPHA,(N-4)/2,1.5);
                                /* critical value of Beta distribution at bonferoni alpha level */
                OUT=STEP||N||OUT||PVAL||CRIT10||BONALPHA;
                                /* one record for output matrix, 12 variables */
                if j=1 then OUT2=OUT; else OUT2=OUT2//OUT;
                                /* build up output matrix from K rows */
/* drop observation with smallest (most significant) Wilk's statistic from data set before next j */
                ID=TEMP[2:N,1]; /* n-1 column vector of IDs*/
                X=TEMP[2:N,2:4]; /* (n-1)x3 */
                end; /* end j loop */
        create &libr..&output from OUT2; append from OUT2; close &libr..&output;
        run; /* run proc iml */

proc sort data=&libr..&output(rename=(COL1=STEP COL2=N COL3=EVENT_ID COL4=&VAR1
                COL5=&VAR2 COL6=&VAR3 COL7=W COL8=PVAL COL9=CRIT10
                COL10=BONALPHA10));
        by descending step;
        run;
data &libr..&output; set &libr..&output;
        format &VAR1 &VAR2 &VAR3 5.3 pval bonalpha10 9.7 w crit10 8.6;
        retain maxl &output; drop maxl;
        &output=1;
        if _n_=1 then maxl=1;
        if w>crit10 and maxl then delete; else maxl=0;
        run;
proc sort data=&libr..&output; by event_id;
        run;
%MEND OUT3;

/* p=2 */
%MACRO OUT2(K, ALPHA, INPUT, OUTPUT, LIBR, VAR1, VAR2);
proc iml;
        use &input;
        ** use arcsine transformed variables ;
        read all var{&VAR1 &VAR2} into X; /* X is n x 2 */
        read all var{event_id} into ID; /* ID is n x 1 */
        ALPHA=&ALPHA;
        do j=1 to &K;
                n=nrow(X);
                if j=1 then n1=n; /* n1 is original number of observation */
                XBAR=(X[:,]); /* XBAR is 1 x p, averages across observations */
                ONE=J(n,1,1); /* ONE is n x 1 column vector of 1's */
```

```
                XBARN=ONE*XBAR[,1]||ONE*XBAR[,2]; /* XBARN is n x p */
                A=(X-XBARN)`*(X-XBARN); /* A is p x p */
                do i=1 to N; /* N is number of observations at current iteration of the K-loop */
                        Xi=X[i,]; /* i-th row of X */
                        WTEMP=1-((N/(N-1))*(XI-XBAR)*inv(A)*(XI-XBAR)`);
                                /* i-th Wilk's likelihood ratio statistic */
                        if i=1 then W=WTEMP; else W=W//WTEMP;
                                /* W becomes n x 1 column vector of Wilk's statistics */
                        end; /* end i loop*/
                W=ID||X||W; /* W is n x 4 */
                create wilks from W; append from W; close wilks; /* read into data set to sort */
                sort data=wilks by COL4; /* sort so smallest Wilk's statistic is first observation */
                use wilks;
                read all var{COL1 COL2 COL3 COL4} into TEMP;
                        /* read data set back into n x 6 matrix */
                close wilks;
                STEP=1+(n1-n); *print STEP; /* there will be a total of K steps */
                /* check for tie in smallest Wilk's statistic */
                *TIECHK=TEMP[1:2,1:4];
                *mattrib TIECHK label='2 smallest W'     colname=({ID &VAR1 &VAR2 W});
                *print TIECHK;
                OUT=TEMP[1,]; /* new 1x6 OUT matrix holds one observation with smallest Wilk's */
                PVAL=probbeta(OUT[1,4],(N-3)/2,1);
                        /* Wilk's statistic p-value from Beta distribution */
                BONALPHA=ALPHA/N;
                CRIT10=betainv(BONALPHA,(N-3)/2,1);
                        /* critical value of Beta distribution at bonferoni alpha level */
                OUT=STEP||N||OUT||PVAL||CRIT10||BONALPHA;
                        /* one record for output matrix, 12 variables */
                if j=1 then OUT2=OUT; else OUT2=OUT2//OUT;
                        /* build up output matrix from K rows */
/* drop observation with smallest (most significant) Wilk's statistic from data set before next j */
                ID=TEMP[2:N,1]; /* n-1 column vector of IDs*/
                X=TEMP[2:N,2:3]; /* (n-1)x4 */
                end; /* end j loop */
        create &libr..&output from OUT2; append from OUT2; close &libr..&output;
        run; /* run proc iml */
proc sort data=&libr..&output(rename=(COL1=STEP COL2=N COL3=EVENT_ID COL4=&VAR1
                COL5=&VAR2 COL6=W COL7=PVAL COL8=CRIT10 COL9=BONALPHA10));
        by descending step;
        run;
data &libr..&output; set &libr..&output;
        format &VAR1 &VAR2 5.3 pval bonalpha10 9.7 w crit10 8.6;
        retain maxl &output; drop maxl;
        &output=1;
        if _n_=1 then maxl=1;
        if w>crit10 and maxl then delete; else maxl=0;
        run;
proc sort data=&libr..&output; by event_id;
        run;
%MEND OUT2;

/* p=1 */
%MACRO OUT1(K, ALPHA, INPUT, OUTPUT, LIBR, VAR1);
```

```
proc iml;
        use &input;
        ** use arcsine transformed variables ;
        read all var{&VAR1} into X; /* X is n x 1 */
        read all var{event_id} into ID; /* ID is n x 1 */
        ALPHA=&ALPHA;
        do j=1 to &K;
                n=nrow(X);
                if j=1 then n1=n; /* n1 is original number of observation */
                XBAR=(X[:,]); /* XBAR is 1 x p, averages across observations */
                ONE=J(n,1,1); /* ONE is n x 1 column vector of 1's */
                XBARN=ONE*XBAR[,1]; /* XBARN is n x p */
                A=(X-XBARN)`*(X-XBARN); /* A is p x p */
                do i=1 to N; /* N is number of observations at current iteration of the K-loop */
                        Xi=X[i,]; /* i-th row of X */
                        WTEMP=1-((N/(N-1))*(XI-XBAR)*inv(A)*(XI-XBAR)`);
                                /* i-th Wilk's likelihood ratio statistic */
                        if i=1 then W=WTEMP; else W=W//WTEMP;
                                /* W becomes n x 1 column vector of Wilk's statistics */
                        end; /* end i loop*/
                W=ID||X||W; /* W is n x 3 */
                create wilks from W; append from W; close wilks; /* read into data set to sort */
                sort data=wilks by COL3; /* sort so smallest Wilk's statistic is first observation */
                use wilks;
                read all var{COL1 COL2 COL3} into TEMP; /* read data set back into n x 6 matrix */
                close wilks;
                STEP=1+(n1-n); *print STEP; /* there will be a total of K steps */
                /* check for tie in smallest Wilk's statistic */
                *TIECHK=TEMP[1:2,1:3];
                *mattrib TIECHK label='2 smallest W'     colname=({ID &VAR1 W});
                *print TIECHK;
                OUT=TEMP[1,]; /* new 1x6 OUT matrix holds one observation with smallest Wilk's */
                PVAL=probbeta(OUT[1,3],(N-2)/2,0.5);
                        /* Wilk's statistic p-value from Beta distribution */
                BONALPHA=ALPHA/N;
                CRIT10=betainv(BONALPHA,(N-2)/2,0.5);
                        /* critical value of Beta distribution at bonferoni alpha level */
                OUT=STEP||N||OUT||PVAL||CRIT10||BONALPHA;
                        /* one record for output matrix, 12 variables */
                if j=1 then OUT2=OUT; else OUT2=OUT2//OUT;
                        /* build up output matrix from K rows */
/* drop observation with smallest (most significant) Wilk's statistic from data set before next j */
                ID=TEMP[2:N,1]; /* n-1 column vector of IDs*/
                X=TEMP[2:N,2:2]; /* (n-1)x4 */
                end; /* end j loop */
        create &libr..&output from OUT2; append from OUT2; close &libr..&output;
        run; /* run proc iml */
proc sort data=&libr..&output(rename=(COL1=STEP COL2=N COL3=EVENT_ID COL4=&VAR1
                COL5=W COL6=PVAL COL7=CRIT10 COL8=BONALPHA10));
        by descending step;
        run;
data &libr..&output; set &libr..&output;
        format &VAR1 5.3 pval bonalpha10 9.7 w crit10 8.6;
        retain maxl &output; drop maxl;
```

```
        &output=1;
        if _n_=1 then maxl=1;
        if w>crit10 and maxl then delete; else maxl=0;
        run;
proc sort data=&libr..&output; by event_id;
        run;
%MEND OUT1;
```

APPENDIX B: SUPPORTING INFORMATION FOR CHAPTER 4

B.1 HALF-MRL METHOD CENSORED DATA REPACEMENT FOR Br_2ClAA AND Br_3AA

Under the "Half-MRL" censored data replacement method used for this research, adjustments were made for the two brominated X_3AA species with high ICR MRLs. The replacement value for censored Br_2ClAA (MRL = 2.0 μg/L) was either 0.5 or 1.0 μg/L and depended on the sample's reported $BrCl_2AA$ concentration. The replacement value for censored Br_3AA (MRL = 4.0 μg/L) was either 0.5, 1.0, 2.0, or 3.0 μg/L and depended on the sample's reported $BrCl_2AA$ and Br_2ClAA concentrations. This approach was expected to provide a more accurate reflection of true occurrence levels for these species than uniform replacement with any single fixed value in the wide range between zero and the MRL.

Censored Br_2ClAA was set to half the MRL (1.0 μg/L) only if measurable $BrCl_2AA$ was reported. Otherwise censored Br_2ClAA was set to 0.5 μg/L. Censored Br_3AA was set to half the MRL (2 μg/L) if measurable Br_2ClAA was reported. If the reported Br_2ClAA concentration exceeded 3 μg/L then censored Br_3AA was set to 3 μg/L. If Br_2ClAA was below MRL but $BrCl_2AA$ was measurable, censored Br_3AA was set to 1.0 μg/L. If Br_2ClAA and $BrCl_2AA$ were both below MRL then Br_3AA was set to 0.5 μg/L.

B.2 VARIABLE CONDITIONING

Adjustments in scale and/or location may be needed to avoid computational instabilities arising from widely disparate ranges or magnitudes among independent variables. Linear dependencies among independent variables also can cause computational problems and result in unstable parameter estimates with inflated variance. Second or higher-order model terms representing interaction or polynomial effects can improve model fit but they introduce collinearity problems and greatly increase the complexity of interpretation. The latter consideration is especially relevant for this research which focused on model interpretation rather than prediction. The benefit of second-order terms was evaluated accordingly. Based on theoretical considerations, temperature and pH were expected to have the greatest potential for interaction effects so interactions terms involving these variables were considered, as well as a quadratic term for temperature to describe a potential nonlinear effect. Interactions with chlorine addition point were included to account for the preferential removal of DBP precursors by coagulation. Altogether, 14 second-order terms were included in model testing. Terms higher than second-order were not considered.

B.3 VALIDATION DATA

A random number from the uniform distribution on the interval (0,1) was generated with the SAS RANUNI function. This was used to assign records to either the exploration or validation subset based on a specified cutoff. Using the computer clock to initialize the seed stream ensured a different record allocation each time the operation was repeated.

B.4 COLLINEARITY ASSESSMENT

A variable with negligible or small variance (i.e. nearly constant value) would tend to be collinear with the intercept term in the design matrix (constant value of 1). Collinearity between the intercept and main independent variables was assessed with principal component analysis using the SAS princomp procedure. Due to its small variance (see Table 4.2), pH exhibited moderate collinearity with the intercept. pH was retained as a candidate variable because the collinearity was not severe and pH is known to significantly influence DBP formation. Collinearity among the 11 main independent variables (see Table 4.2) was assessed with eigenanalysis and decomposition of the variances of the estimates with respect to each eigenvalue, using the SAS reg procedure and *collinoint* option. A collinearity problem occurs when a component associated with a high condition index contributes strongly (variance proportion > 0.5) to the variance of two or more variables (SAS 1999). The condition number (i.e. highest condition index) for the data set was 3.3, indicating that dependency among the main independent variables would not seriously affect parameter estimates. A condition number above 10 has been suggested as evidence of weak dependencies and one above 30 as potentially problematic (Belsley, Kuh, and Welch 1980). Collinearity among independent variables was reassessed after inclusion of second-order terms. As expected, due to inherent dependencies between second-order terms and main variables, there was a substantial amount of collinearity among the 25 variables. After stepwise removal of 5 second-order terms, the condition number was reduced from 68 to 24. The 9 second-order terms remaining for model development included chlorine addition point interactions (with alkalinity, bromide, TOC, chlorine consumed, and chlorine contact time),

temperature interactions (with TOC, chlorine consumed, and chlorine contact time), and the quadratic temperature term.

B.5 DEPENDENT VARIABLE TRANSFORMATION

The Box-Cox series of transforms was applied to the dependent variable according to Equation B.1, where Y_{gm} is the geometric mean of Y. This series maintains the Y units so the sum of squared errors (SSE) can be compared directly in selecting the best π value (Muller and Fetterman 2003). Residual plots and SSE from least-squares regression were compared for π values of -1, 0, 0.5, and 1, corresponding to reciprocal, log, square root, and no transformation. The optimal π was identified as that yielding minimal SSE with acceptable residual patterns.

$$Y' = \frac{\left[\left(Y^{\pi}\right) - 1\right]}{\pi * \left(Y_{gm}^{\pi-1}\right)}, \quad \pi \neq 0$$

$$Y' = Y_{gm} * \log(Y), \quad \pi = 0 \tag{B.1}$$

Box-Cox analysis led to use of log transformation for all dependent variables: SSE were minimized for both log and square root transforms (with negligible difference), and residual patterns were acceptable in both cases. The log transform was chosen for simplicity and for consistency with previous modeling work.

B.6 RESIDUAL DIAGNOSTICS

Model residuals were examined to evaluate conformance with theoretical assumptions of the linear model and to check for evidence of model misspecification. Studentized residuals were used for these diagnostics. This form of residual is normalized by a standard error computed without the current observation. Because the standard error shrinks if an outlier is excluded, outlier points are emphasized by this treatment. Scatter plots of model residuals versus predicted dependent variables were examined for variance homogeneity and patterns indicating non-linearity or other model misspecification. Partial regression leverage plots were also reviewed to check for outliers and signs of non-linear effects with respect to individual independent variables. Residual normality was evaluated by histogram inspection and the Kolmogorov-Smirnov test statistic.

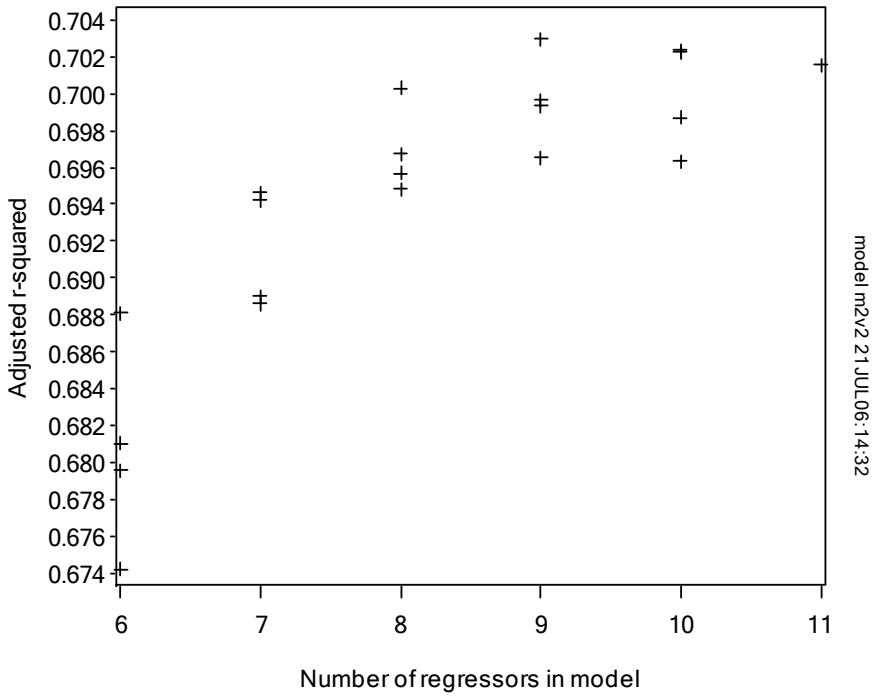
B.7 VARIABLE SELECTION AND MODEL VALIDATION

The SAS reg procedure was used to select model independent variables. The procedure was first run using stepwise selection with significance levels for independent variables to enter and remain in the model set at 0.5 and 0.05, respectively. This imposed a minimal barrier to variables entering the model but only retained them if their significance level stayed below 0.05 (F-statistic, p-value). Stepwise selection results were used to suggest the appropriate number of variables, but the outcome was not taken as the optimal model. The reg procedure was repeated using the R-squared selection method to compute Mallows' Cp and adjusted R^2 statistics for the best 4 models (i.e. highest R^2) of each size in a range bracketing the suggested number of variables. The adjusted R^2 statistic is an alternative to R^2 that accounts for the number of parameters in the model, affording better comparisons

between models of different size. Adjusted R^2 and (Cp-1) were plotted as a function of the number of variables to identify the optimal model size. This was taken as the point where additional variables no longer increased precision meaningfully and/or where (Cp-1) approached the number of regressors. Daniel and Wood (1971) and Muller and Fetterman (2003) discuss the use of Cp to guide model selection. Cp is an estimate of the total variance of prediction (random plus bias) and, thus, minimizing Cp is desirable. Theoretically, a model with minimal bias should have an expected value for Cp equal to the number of parameters in the model (including intercept). Daniel and Wood (1971) note that the correct region is where Cp values for a group of models begin to diverge. The model with the smallest R^2 for the optimal size was chosen. Figure B.1 shows the adjusted R^2 and Cp plots for the THM4 model.

Based on results for THM4 and X_2AA models, it was determined that use of second-order terms was not warranted in the models. Thus, second-order terms were neglected in further model selection.

a)



b)

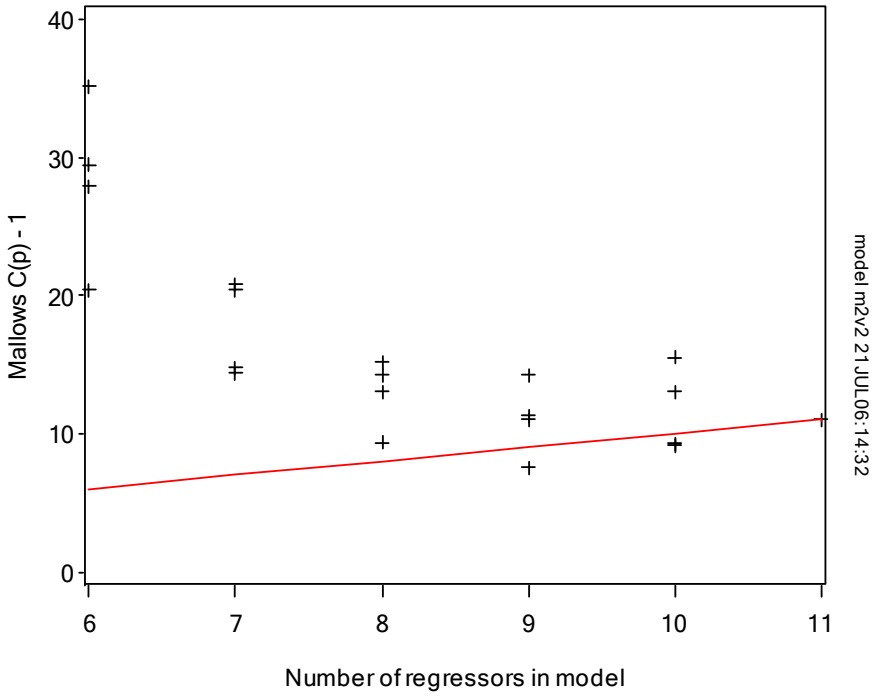


Figure B.1 Plots used to guide variables selection in exploratory THM4 model development showing results for best 4 models in each size: a) adjusted R^2, and b) Mallow's Cp-1 with 1:1 reference line as a function of number of regressors in model

B.8 SEGMENTED MODELING APPROACHES

A potential source of model misspecification was conceptualizing the treatment train as a single input/output process whereas, in reality, it is a series of unit processes with continuously changing water quality conditions and often multiple chlorine application points. Although data were available for most of the important water quality variables through each process in the treatment trains, DBP results were only available for filtered water (if chlorine was added upstream) and finished water. Exploiting these data, alternative models for THM4 and X_2AA were developed based on conceptualizing the treatment train as two segments: raw to filtered water, and filtered to finished water. This was expected to provide better characterization of specific water quality and treatment conditions relevant to incremental DBP formation in each segment. Data were re-processed to obtain appropriate records for each segment as in Table 4.2; the classification variable Cl_2 point was dropped for this analysis. Filtered water DBP results (where available) were associated with water quality and treatment conditions for the raw to filtered water segment of the treatment train. Incremental finished water DBP results (i.e. finished water concentration minus filtered water concentration) were associated with the filtered to finished water segment. Records for the two segments were then modeled both separately and in aggregate.

The THM4 model for combined segment records showed the same misspecification problem as seen in the original model but fit the data more poorly (R^2 decreased from 0.707 to 0.510) and had unacceptable residual patterns. No improvement was found when models were developed separately for the two segment types, although results were noticeably better for the raw to filtered water segment records ($R^2 = 0.540$) than for the filter to finished water segment records ($R^2 = 0.379$): the misspecification persisted in each model. However, when

predictions from the combined segment model were segregated according to treatment configuration, the misspecification pattern was found to be limited to filter to finished water incremental THM4 formation for plants adding chlorine before filtration (Figure B.2b). There was substantial scatter for raw to filtered water segment data (Figure B.2a) but no trend of overprediction at low values and/or underprediction at high values. Finished water THM4 for plants with no chlorine addition before filtration (Figure B.2c) or without any filter process (Figure B2.d) were both quite well represented by the model, though there were relatively few data of these type. Results of the segmented model analysis for X_2AA were similar to those shown for THM4.

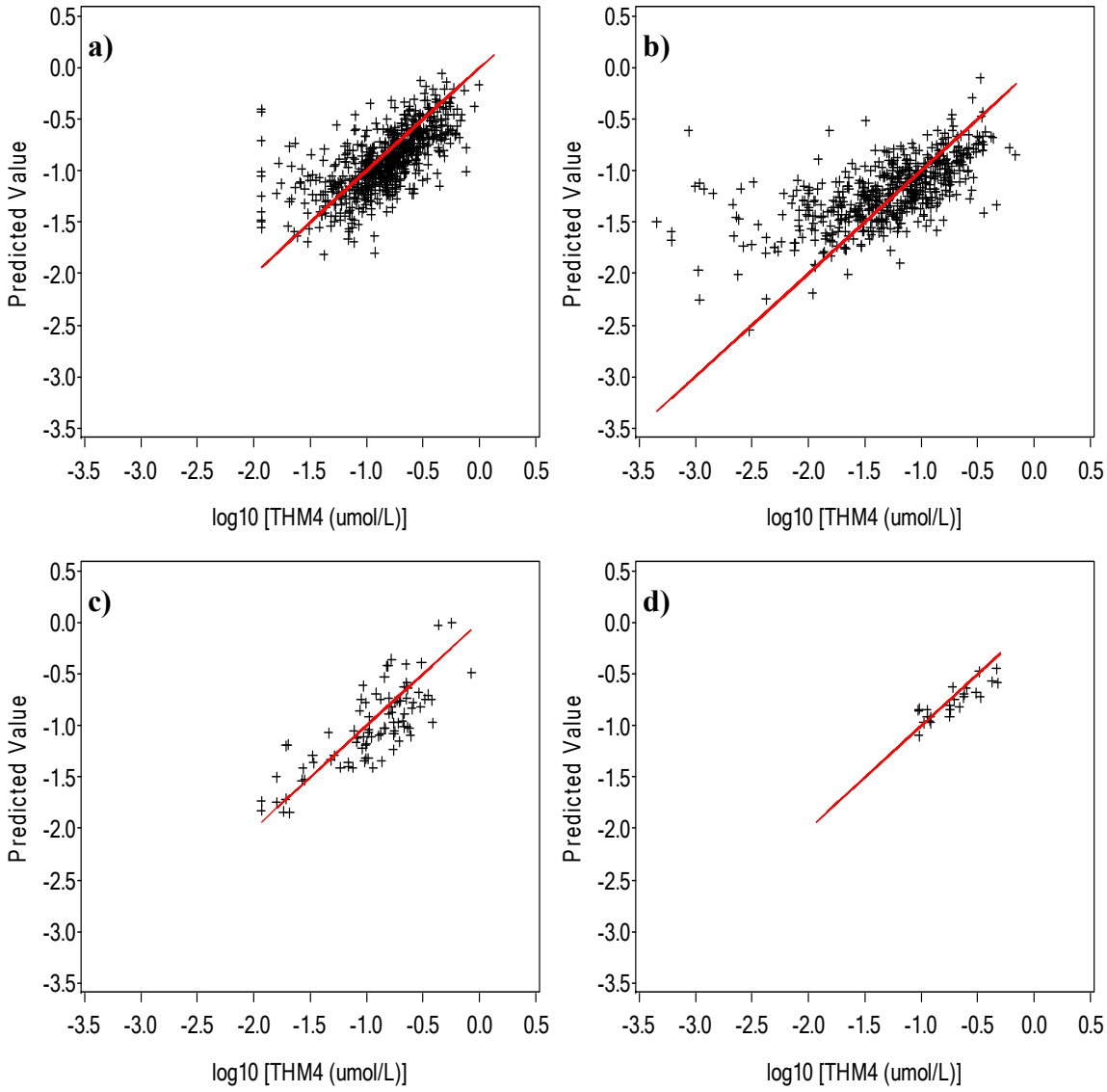


Figure B.2 Predicted versus observed log10[THM4] results for segmented model subset by record type: a) raw to filtered water segments, b) filtered to finished water segments for plants adding chlorine upstream of filter, c) filtered to finished water segments for plants with no chlorine upstream of filter, d) raw to finished water segments for plants with no filter process

Results of the segmented model analysis were provocative and suggested that subsequent stages of DBP formation following initial exposure to chlorine and filtration were difficult to describe in the current framework. However, there was no clear way to account

for this with the available data, given the poor results for the segmented approach. Therefore, the original model formulation was determined to be the best approach within the framework of a multiple linear regression model, given the available data and the primary goal of providing a relatively straightforward platform for examining the effects of water quality and treatment on DBP formation.

REFERENCES

Adin A., Katzhendler J., Alkaslassy D., and Rav-Acha Ch. (1991). Trihalomethane formation in chlorinated drinking water: a kinetic model. Water Research **25**(7), 797-805

Amy G. L., Chadik P. A., and Chowdhury Z. K. (1987). Developing models for predicting trihalomethane formation potential and kinetics. Journal American Water Works Association **79**(7), 89–97.

Amy G. L., Siddiqui M, Ozekin K., Zhu H. W., and Wang C. (1998). Empirically Based Models for Predicting Chlorination and Ozonation By-Products: Trihalomethanes, Haloacetic Acids, Chloral Hydrate, and Bromate. U.S. EPA Office of Ground Water and Drinking Water, EPA-815-R-98-005.

Amy G. L., Tan L., and Marshall K. D. (1991). The effects of ozonation and activated carbon adsorption on trihalomethane speciation. Water Research **25**(2), 191–202.

Amy G. L., Sierka R. A., Bedessem J., Price D., and Tan L. (1992). Molecular size distributions of dissolved organic matter. Journal American Water Works Association **84**(6), 67–75.

Anderson D. O., Alberts J. J., and Takacs M. (2000). Nature of natural organic matter (NOM) in acidified and limed surface waters. Water Research **34**(1), 266–278.

Archer A. and Singer P. C. (2006). An evaluation of the relationship between SUVA and NOM coagulation using the ICR database. Journal American Water Works Association **98**(7), 110–123.

Arguello M. D.; Chriswell C. D.; Fritz J. S.; Kissinger L. D.; Lee K. W.; Richard J. J.; and Svec H. J. (1979). Trihalomethanes in water: a report on the occurrence, seasonal variation in concentrations, and precursors of trihalomethanes. Journal American Water Works Association **71**(9), 504–508.

Arora H.; LeChevallier M. W.; and Dixon K. L. (1997). DBP occurrence survey. Journal American Water Works Association **89**(6), 60–68.

AWWA (2000). Water Quality Division Disinfection Systems Committee report: disinfection at large and medium-size systems. Journal American Water Works Association **92**(5), 32–43.

Babcock D. B. and Singer P. C. (1979). Chlorination and coagulation of humic and fulvic acids. Journal American Water Works Association **71**(3), 149–152.

Baribeau H., Boulos L., Haileselassie H., Crozes G., Singer P. C., Nichols C., Schlesinger S. A., Gullick R., Williams S., Williams R. L., Fountleroy L., Andrews S., Moffat E. (2006). Formation and Decay of Disinfection By-Products in the Distribution System. American Water Works Research Foundation, Denver.

Barrett S. E., Krasner S. W., and Amy G. L., eds. (2000). Natural Organic Matter and Disinfection By-Products: Characterization and Control in Drinking Water. American Chemical Society, Washington D.C.

Bellar T. A., Lichtenberg J. J, and Kroner R. C. (1974). The occurrence of organohalides in chlorinated drinking water. Journal American Water Works Association **71**(3), 703–706.

Belsley D. A., Kuh E., and Welsch R. E. (1980). Regression Diagnostics. John Wiley and Sons, New York.

Bieber T. I. and Trehy M. L. (1983). Dihaloacetonitriles in chlorinated natural waters. In Water Chlorination: Environmental Impact and Health Effects, Vol. 4 No. 1 (eds. R. L. Jolley, W. A. Brungs, J. A. Cotruvo, R. B. Cumming, J. S. Mattice, and V. A. Jacobs), pp. 85–96 Ann Arbor Science Publishers, Ann Arbor.

Bischel Y. and von Gunten U. (1999). Oxidation of iodide and hypoiodous acid in the disinfection of natural waters. Environmental Science and Technology **33**(22), 4040–4045.

Blank V., Shukairy H. M., and Obolensky A. (2002). Unregulated organic DBPs in ICR finished water and distribution systems. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 277–297. American Water Works Association Research Foundation, Denver.

Boyce S. D. and Hornig J. F. (1983). Reaction pathways of trihalomethane formation from the halogenation of dihydroxyaromatic model compounds for humic acid. Environmental Science and Technology **17**(4), 202–211.

Brett R. W. and Calverly R. A. (1979). A one-year survey of trihalomethane concentration changes within a distribution system. Journal American Water Works Association **76**(9):515–520.

Bull R. J., Krasner S. K., Daniel P. A., and Bull R. D. (2001). Health Effects and Occurrence of Disinfection By-Products, American Water Works Association Research Foundation and American Water Works Association, Denver.

Caroni C. and Prescott P. (1992). Sequential application of Wilks's multivariate outlier test. Applied Statistics **41**, 355–364.

Cavanagh J. E., Weinberg H. S., Gold A., Sangaiah R., Marbury D., and Glaze W. H. (1992). Ozonation byproducts: identification of bromohydrins from the ozonation of natural waters with enhanced bromide levels. Environmental Science and Technology **26**(8), 1658–1662.

Chadik P. A. and Amy G. L. (1983). Removing trihalomethane precursors from various natural waters by metal coagulants. Journal American Water Works Association **80**(10), 532–536.

Chadik P. A. and Amy G. L. (1987). Coagulation and adsorbtion of humic substances: an analysis of surrogate parameters for predicting effects on trihalomethane formation potential. Environmental Technology Letters **8**, 261–268.

Chambers J. M., Cleveland W. S., and Tukey P. A. (1983). Graphical Methods for Data Analysis. Duxbury Press, Duxbury, MA.

Chen J. and Regli S. (2002). Disinfection practices and pathogen inactivation in ICR surface water plants. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 371–396. American Water Works Association Research Foundation, Denver.

Choi J. and Valentine R. L. (2002). Formation of N-nitrosodimethylamine (NDMA) from reaction of monochloramine: a new disinfection by-product. Water Research **36**(4), 817–824.

Chowdhury Z. K. and Amy G. L. (1999). Modeling disinfection by-product formation. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 53–64. American Water Works Association, Denver.

Clark R. M. (1998). Chlorine demand and TTHM formation kinetics: a second-order model. Journal of Environmental Engineering **124**(1). 16-24.

Christman R. F. and Oglesby R. T (1971). Microbiological degradation and the formation of humus. In Lignins, Occurrence, Formation, Structure, and Reactions (eds. K. V. Sarkanen and C. H. Ludwig). Wiley-Interscience, New York.

Christman R. F.; Norwood D. L.; Millington D. S.; Johnson, J. D.; and Stevens A. A. (1983). Identity and yields of major halogenated products of aquatic fulvic acid chlorination. Environmental Science and Technology **17**(10): 625-628.

Christman R. F., Norwood D. L., Seo Y., and Frimmel F. H. (1989). Oxidative degradation of humic substances from freshwater environments. In Humic Substances II (eds. M. H. B. Hayes, P. MacCarthy, R. L. Malcom, and R. S. Swift). John Wiley and Sons, New York.

Connell K., Rodgers C. C., Shank-Givens H. L., and Miller K. (2000). Building a better protozoa data set. Journal American Water Works Association **92**(10), 30–43.

Cowman G. A. and Singer P. C. (1996). Effect of bromide ion on haloacetic acid speciation resulting from chlorination and chloramination of aquatic humic substances. Environmental Science and Technology **30**, 16–24.

Craun G. F., Hauchman F. S., and Robinson D. E., eds. (2001). Microbial Pathogens and Disinfection By-products in Drinking Water: Health Effects and Management of Risks. International Life Sciences Institute, Washington D.C.

Croue J. P. and Reckhow D. A. (1989). Destruction of chlorination byproducts with sulfite. Environmental Science and Technology **23**(11), 1412–1419.

Croue J. P., Debroux J. F., Amy G. L., Aiken G. R., and Leenheer J. A. (1999). Natural organic matter: structural characteristics and reactive properties. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 65–93. American Water Works Association, Denver.

Culp G., ed. (1984). Trihalomethane Reduction in Drinking Water: Technologies, Costs, Effectiveness, Monitoring, and Compliance. Noyes Publications, Park Ridge, NJ.

Daniel C. and Wood F. S. (1971). Fitting Equations to Data. Wiley-Interscience, New York.

de Leer E. W. B., Sinninghe Damste J. S., Erkelens C., and de Galan L. (1985). Identification of intermediates leading to chloroform and C-4 diacids in the chlorination of humic acid. Environmental Science and Technology **19**(6), 512-522.

Dietrich A. M, Orr M. P., and Gallagher D. L. (1992). Tastes and odors associated with chlorine dioxide. Journal American Water Works Association **84**(6), 82–88.

DiGiano F. A,. and Bond R. G. (2004). Evaluating GAC performance using the ICR database. Journal American Water Works Association **96**(6), 96–104.

Domino, M. M., Pepich B.V., Munch D. J., and Fair P. S. (2004). Optimizing the determination of haloacetic acids in drinking waters. Journal of Chromatography A **1035**, 9-16.

Dowbiggin and Singer P. C. (1989). Effects of natural organic matter and calcium on ozone-induced particle destabilization. Journal American Water Works Association **81**(6), 77–85.

Draper N. R. and Smith H. (1996). Applied Regression Analysis. Wiley and Sons, New York.

Duke D. T., Siria J. W., Burton B. D., and Amundsen D. W. (1980). Control of trihalomethanes in drinking water. Journal American Water Works Association **72**(8), 470–476.

Edzwald J. K. Becker W.C., Wattier K. L. (1985). Surrogate Parameters for Monitoring Organic Matter and THM Precursors. Journal American Water Works Association **77**(4), 122–132.

El-Shaarawi A. H. (1989). Inferences about the mean from censored water quality data. Water Resources Research **25**(4), 685–690.

Fair P. S., Sorrell R. K., Stultz-Karapondo M., Bohl B. J., Frebis C. P., and Feige M. A. (2002). Quality of ICR monitoring data. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 39–66. American Water Works Association Research Foundation, Denver.

Fleischaker S. J. and Randtke S. J. (1983). Formation of organic chlorine in public water supplies. Journal American Water Works Association **75**(3), 132–138.

Gang D. D., Segar R. L., Clevenger T. E., and Banerji S. K. (2002).Using chlorine demand to predict TTHM and HAA9 Formation. Journal American Water Works Association **94**(10), 76-86.

Gerecke A. C. and Sedlak D. L. (2003). Precursors of N-nitrosodimethylamine in natural waters. Environmental Science and Technology **37**(7), 1331–1336.

Glaze W. H., Weinberg H. S., and Cavanagh J. E. (1993). Evaluating the formation of brominated DBPs during ozonation. Journal American Water Works Association **85**(1), 96–103.

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

Gould J. P., Fitchhorn L. E., Urheim E. (1983). Formation of brominated trihalomethanes: extent and kinetics. In Water Chlorination: Environmental Impact and Health Effects, vol. 4. (eds. R. L. Jolley, W. A. Brungs, J. A. Cotruvo, R. B. Cumming, J. S. Mattice, and V. A. Jacobs), pp. 297–310. Ann Arbor Science Publishers, Ann Arbor.

Haas C. N., Jacangelo J. G. (1993). Development of regression models with below-detection data. Journal Environmental Engineering **119**(2), 214–230.

Haas C. N. (1999). Disinfection. In Water Quality and Treatment: A Handbook of Community Water Supplies (ed. R. D. Letterman), pp. 14.1-14.60. McGraw-Hill, New York.

Hansson R. C., Henderson M. J., Jack P., and Taylor R. D. (1987). Iodoform taste complaints in chloramination. Water Research **21**(10),1265-1271

Harrington G. W., Chowdhury Z. K., and Owen, D. M. (1992). Developing a computer model to simulate DBP formation during water treatment. Journal American Water Works Association **84**(11), 78-87.

Heller-Grossman L., Manka J., Limoni-Relis B., and Rebhun M. (1993). Formation and distribution of haloacetic acids, THM and TOX in chlorination of bromide-rich lake water. Water Research **27**(8), 1323-1331.

Helsel D. R., Gilliom R. J. (1986). Estimation of distributional parameters for censored trace level water quality data 2: verification and applications. Water Resources Research **22**(2), 147–155.

Hoehn R. C. and Gates D. J. (1999). Control of disinfection by-product formation using chlorine dioxide. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 205–221. American Water Works Association, Denver.

Hooper S. M. and Allgeier S. A. (2002). GAC and membrane treatment studies. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 501–538. American Water Works Association Research Foundation, Denver.

Jafvert C. and Valentine R. L. (1992). A reaction scheme for chlorination of ammoniacal water. Environmental Science and Technology **26**(3), 577–586.

Johnson J. D. and Jensen J. N. (1986). THM and TOX formation: routes, rates, and precursors. Journal American Water Works Association **78**(4), 156-162.

Kavanaugh M. C., Trussell A., Croner J., and Trussell R. R. (1980). An empirical kinetic model of trihalomethane formation: applications to meet the proposed THM standard. Journal American Water Works Association **72**(10), 578-582.

Kleinbaum D. G., Kupper L. L., Muller K. E., and Nizam A. (1998). Applied Regression Analysis and Multivariable Methods. Duxbury Press, Duxbury, MA.

Krasner S. W., McGuire M. J., Jacangelo J. G., Patania N. L., Reagan K. M., and Aieta E. M. (1989). The occurrence of disinfection by-products in US drinking water. Journal American Water Works Association **81**(8), 41–53.

Krasner S. W., Reagan K. M., Jacangelo J. G., Patania N. L., Aieta E. M., Gramith K. M. (1990). Relationships between disinfection by-products and water quality parameters: implications for formation and control. Proceedings of the Annual Conference and Exposition. American Water Works Association, Denver.

Krasner S. W., Croue J. P., Buffle J., and Purdue E. M. (1996a). Three approaches for characterizing NOM. Journal American Water Works Association **88**(6), 66–79.

Krasner S. W., Sclimenti M. J., Chinn R., Chowdhury Z. K., Owen D. M. (1996b). 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. (eds. R. A. Minear and G. L. Amy), pp.59–90. CRC Press Inc., Boca Raton.

Krasner S. W., Chowdhury Z. K., Edwards M. A., and Bell K. A. (1997). Use of SUVA in developing TOC removal requirements. In Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.

Krasner S. W., Chinn R., Pastor S. J., Sclimenti M. J., Richardson S. D., Thruston A. D., Weinberg H. W. (2002). Relationships between the different classes of DBPs: formation, speciation, and control. Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.

Krasner S. W., Weinberg H. S.; Richardson S. D.; Pastor S. J., Chinn R., Sclimenti M. J., Onstad G. D., Thruston A. D. Jr. (2006). Occurrence of a new generation of disinfection byproducts. Environmental Science and Technology **40**(23), 7175–7185.

Leenheer J. A., Croue J. P., Benjamin M., Korshin G. V., Hwang C. J., Bruchet A., and Aiken G. R. (2000). Comprehensive isolation of natural organic matter from water for spectral characterizations and reactivity testing. In Natural Organic Matter and Disinfection By-Products: Characterization and Control in Drinking Water (eds. S. E. Barrett, S. W. Krasner, and G. L. Amy), pp. 68-83. American Chemical Society, Washington D. C.

Leenheer J. A. and Croue J. P. (2003). Characterizing aquatic dissolved organic matter. Environmental Science and Technology **37**(1), 19A-26A.

Liang L. and Singer P. C. (2003). Factors influencing the formation and relative distribution of haloacetic acids and trihalomethanes in drinking water. Environmental Science and Technology **37**(13), 2920–2928.

Liu S., Lu J. C., Kolpin D. W., Meeker W. Q. (1997). Analysis of environmental data with censored observations. Environmental Science and Technology **31**(12), 3358–3362.

Lubin J. H., Colt J. S., Camann D., Davis S., Cerhan J. R., Severson R. K., Bernstein L., and Hartge P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. Environmental Health Perspextives **112**(17), 1691–1696.

Lykins B. W. and Griese M. H. (1986).Using chlorine dioxide for trihalomethane control. Journal American Water Works Association **78**(6), 88–93.

Marhaba T. F. and Van D. (2000). The variation of mass and disinfection by-product formation potential of dissolved organic matter fractions along a conventional surface water treatment plant. Journal of Hazardous Materials **A74,** 133–147.

McGuire M. J. and Meadow R. G.. (1988). AWWARF trihalomethane survey. Journal American Water Works Association **80**(1), 61–68.

McGuire M. J. and Hotaling M. L. (2002). Relationships between source water quality and choices of primary and secondary disinfectants. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 299–316. American Water Works Association Research Foundation, Denver.

McGuire M. J., McLain J. L., and Obolensky A., eds. (2002). Information Collection Rule Data Analysis. American Water Works Association Research Foundation, Denver.

McLain J. L., Obolensky A., Shukairy H. M. (2002). Disinfection by-product speciation. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp 141–168. American Water Works Association Research Foundation, Denver.

Messner M. J. and Wolpert R. L. (2002). Cryptosporidium and Giardia occurrence in ICR drinking water sources – statistical analysis of ICR data. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 463–482. American Water Works Association Research Foundation, Denver.

Miller J. W. and Uden P. C. (1983). Characterization of nonvolatile aqueous chlorination products of humic substances. Environmental Science and Technology **17**(3), 150–157.

Miltner R. J., Shukairy H. M., and Summers R. S. (1992). Disinfection by-product formation and control by ozonation and biotreatment. Journal American Water Works Association **84**(11), 53–62.

Minear R. A. and Bird J. C. (1980). Trihalomethanes: impact of bromide ion concentration on yield, species distribution, rate of formation, and influence of other variables. In Water Chlorination: Environmental Impact and Health Effects, vol. 3. (eds. R. L. Jolley, W. A. Brungs, R. B. Cumming, and V. A. Jacobs), pp 151–160. Ann Arbor Science Publishers, Ann Arbor.

Mitch W. A. and Sedlak D. L. (2002). Formation of N-nitrosodimethylamine from dimethylamine during chlorination. Environmental Science and Technology **36**(4), 588–595.

Morris J. C. (1967). Kinetics of reactions between aqueous chlorine and nitrogen compounds. In Principles and Applications of Water Chemistry (eds. S. D. Faust and J. H. Hunter), pp. 23-53. John Wiley and Sons, New York.

Muller K. E. and Fetterman B. A. (2003). Regression and ANOVA. John Wiley and Sons, New York.

Munch D. J, Fair P. S., Hautman D. P., Sorrell R. K., Bohl B., Frebis C. P., Stultz-Karapondo M. L., Domino M. M., and Pepich B. V. (2000). The progression of methods approved by the U.S. EPA for HAA analyses and the near term future for EPA methods development. Workshop presentation, Water Quality Technology Conference, American Water Works Association.

Murphy P. A. and Craun G. F. (1999). Balancing microbial and chemical risks. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 119–138. American Water Works Association, Denver.

Myers A. G. (1990). Evaluating alternative disinfectants for THM control in small systems. Journal American Water Works Association **82**(6), 77–84.

Najm I. N., Patania N., Jacangelo J. G., and Krasner S. K. (1994). Evaluating surrogates for disinfection by-products. Journal American Water Works Association **86**(6), 98–106.

Najm I. N. and Trussell R. R. (2001). NDMA formation in water and wastewater. Journal American Water Works Association **93**(2), 92–99.

Nieminski E. C.; Chaudhuri S.; and Lamoreaux T. (1993). The occurrence of DBPs in utah drinking waters. Journal American Water Works Association **85**(9):98–105.

Norwood D. L., Johnson J. D., Christman R. F., Hass J. R., and Bobenreith M. J. (1980). Reactions of chlorine with selected aromatic models of aquatic humic material. Environmental Science and Technology **14**(2), 187-190.

Norwood D. L. and Christman R. F. (1987). Structural characterization of aquatic humic material, 2, phenolic content and its relationship to chlorination mechanism in an isolated aquatic fulvic acid. Environmental Science and Technology **21**(8), 791-798.

Obolensky A. (1998). Philadelphia Water Department internal memorandum: suspect distribution system data, September 7, 1998.

Obolensky A. and Frey M. (2002). Distribution system DBP results and SDS performance. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp 169–198. American Water Works Association Research Foundation, Denver.

Obolensky A. and Singer P. C. (2005). Halogen substitution patterns among disinfection byproducts in the Information Collection Rule database. Environmental Science and Technology **39**(8), 2719–2730.

Obolensky A., Singer P. C., and Shukairy H. M. (2007). Information Collection Rule data evaluation and analysis to support impacts on disinfection by-product formation. Journal of Environmental Engineering **133**(1), 53–63.

Oliver B. G. (1980). Effect of temperature, pH, and bromide concentration on the trihalomethane reaction of chlorine with aquatic humic material. In Water Chlorination: Environmental Impact and Health Effects, vol. 3. (eds. R. L. Jolley, W. A. Brungs, R. B. Cumming, and V. A. Jacobs), pp 141–149. Ann Arbor Science Publishers, Ann Arbor.

Oliver B. G. and Thurman E. M. (1983). Influence of aquatic humic substance properties on trihalomethane potential. In Water Chlorination Environmental Impacts and Health Effects Vol 4 No. 1 (eds. R. L. Jolley, W. A. Brungs, J. A. Cotruvo, R. B. Cumming, J. S. Mattice, and V. A. Jacobs), pp. 231-241. Ann Arbor Science Publishers, Ann Arbor.

Onstad G. D. (2003). Methods, Formation, and Occurrence of Halogenated Furanones in U.S. Drinking Waters. Doctoral Dissertation, University of North Carolina, Chapel Hill.

Otson R.; Williams D. T.; and Bothwell P. D. (1981). Comparison of trihalomethane levels and other water quality parameters for three treatment plants on the Ottawa river. Environmental Science and Technology **15**(9):1075–1080.

Pennell M. L. (2002). Identification Of Multivariate Outliers In an Environmental Risk Assessment. Master's Thesis, University of North Carolina, Chapel Hill.

Plewa M. J., Wagner E. D., Richardson S. D., Thruston A. D., Woo Y., McKague A. B. (2004). Chemical and biological characterization of newly discovered iodoacid drinking water disinfection byproducts. Environmental Science and Technology **38**(18), 4713–4722.

Pourmoghaddas H., Stevens A. A., Kinman R. N., Dressman R. C., Moore L. A., and Ireland, J. C. (1993). Effect of bromide ion on formation of HAAs during chlorination. Journal American Water Works Association **85**(1), 82–87.

Qualls R. G. and Johnson J. D. (1983). Kinetics of the short-term consumption of chlorine by fulvic acid. Environmental Science and Technology **17**(11), 692-698.

Quimby B. D., Delaney M. F., Uden P. C., and Barnes R. M. (1980). Determination of the aqueous chlorination products of aquatic humic substances by gas chromatography with microwave emission detection. Analytical Chemistry **52**(2), 259–263.

Randtke S. J. (1998). Organic contaminant removal by coagulation and related process combinations. Journal American Water Works Association **80**(5), 40–56.

Randtke S. J. (1999). Disinfection by-product precursor removal by coagulation and precipitative softening. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 237–258. American Water Works Association, Denver.

Reckhow D. A. and Singer P. C. (1984). The removal of organic halide precursors by preozonation and alum coagulation. Journal American Water Works Association **76**(4), 151–157.

Reckhow D. A. and Singer P. C. (1985). Mechanisms of organic halide formation during fulvic acid chlorination and implications with respect to preozonation. In Water Chlorination: Environmental Impact and Health Effects, Vol. 5 (eds. R. L. Jolley, R. J. Bull, W. P. Davis, S. Katz, M. H. Roberts, and V. A. Jacobs), pp. 1229–1257. Lewis Publishers Inc., Chelsea, MI.

Reckhow D. A., Singer P. C., and Malcom R. L. (1990). Chlorination of humic materials: byproduct formation and chemical interpretations. Environmental Science and Technology **24**(11), 1655–1664.

Reckhow D. A. (1999). Control of disinfection by-product formation using ozone. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 179–204. American Water Works Association, Denver.

Richardson S. D. (1998). Drinking water disinfection by-products. In Encyclopedia of Environmental Analysis and Remediation, Vol. 3 (ed. R. A. Meyers). John Wiley and Sons, New York.

Richardson S. D. (2003). Disinfection by-products and other emerging contaminants in drinking water. Trends in Analytical Chemistry **22**(10), 666–684.

Roberson J. A. (2002). An inventory of ICR systems, treatment plants, and source water supplies. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp 67–80. American Water Works Association Research Foundation, Denver.

Roberts M. G., Singer P. C., and Obolensky A. (2002). Comparing total HAA and total THM concentrations using ICR data. Journal American Water Works Association **94**(1), 103–114.

Rook, J. J. (1974). Formation of haloforms during chlorination of natural waters. Journal of Water Treatment Examination **23**(2):234–243.

Rook J. J.; Gras A. A.; van den Heijden J.; and de Wee J. (1978). Bromide oxidation and organic substitution in water treatment. Journal of Environmental Health **A13**(2):91-116.

SAS Institute Inc. (1999). SAS/STAT® User's Guide, Version 8. SAS Institute Inc., Cary, NC.

Scharfenaker M. A. (2001). Reg watch: USEPA offers first glimpse of Stage 2 D/DBPR. Journal American Water Works Association **93**(12), 20–34.

Seidel C. J., McGuire M. J., Summers R. S., and Via S. (2005). Have Utilities Switched to Chloramines? Journal American Water Works Association **97**(10), 87–97.

Shapiro J. (1957). Chemical and biological studies on the yellow organic acids of lake water. Limnology and Oceanography **2**(3), 161-179.

Shukairy H. M., Miltner R.J., Summers R. S. (1994). Bromide's effect on DBP formation, speciation, and control: part 1, ozonation. Journal American Water Works Association **86**(6), 72–87.

Shukairy H. M. and Summers R. S. (1996). DBP speciation and kinetics as affected by ozonation and biotreatment. In Disinfection Byproducts in Water Treatment: The Chemistry of Their Formation and Control (eds. R. A. Minear and G. L. Amy), pp. 311–335. CRC Press Inc., Boca Raton.

Shukairy H. M., Blank V., and McLain J. L. (2002). TOX in finished water and distribution systems: relationships with other DBPs. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 251–275. American Water Works Association Research Foundation, Denver.

Singer P. C., Barry J. J., Palen G. M., and Scrivner A. E. (1982). Trihalomethane Formation in Water Treatment Plants in North Carolina. Water Resources Research Institute of the University of North Carolina, Report No. 179, Raleigh.

Singer P. C. and Chang S. D. (1989). Correlations between trihalomethanes and total organic halides formed during drinking water treatment. Journal American Water Works Association **81**(8), 61–65.

Singer P.C. (1994). Control of disinfection by-products in drinking water. Journal of Environmental Engineering **120**(4)727–744.

Singer P. C.; Obolensky A.; and Greiner A. (1995). DBPs in North Carolina drinking waters. Journal American Water Works Association **87**(10):83–92.

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

Solarik G., Summers R. S., Sohn J., Swanson W. J., Chowdhury Z. K, and Amy G. L. (2000). Extensions and verification of the water treatment plant model for disinfection byproduct formation. In Natural Organic Matter and Disinfection By-Products: Characterization and Control in Drinking Water (eds. S. E. Barrett, S. W. Krasner, and G. L. Amy), pp. 47-66. American Chemical Society, Washington D. C.

Speitel G. E. (1999). Control of disinfection by-product formation using chloramines. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 161–177. American Water Works Association, Denver.

Stevens A. A., Slocum C. J., Seeger D. R., Robeck G. G. (1976). Chlorination of organics in drinking water. Journal American Water Works Association **68**(11), 615–620.

Stevens A. A.; Moore L. A.; and Miltner R. J. (1989). Formation and control of non-trihalomethane by-products. Journal American Water Works Association **81**(8), 54–60.

Stewart A. J. and Wetzel R. G. (1981). Asymmetrical relationships between absorbance, fluorescence, and dissolved organic carbon. Limnology and Oceanography **26**(3), 590–597.

Summers R. S., Benz M. A., Shukairy H. M., and Cummings L. (1993). Effect of separation processes on the formation of brominated THMs. Journal American Water Works Association **85**(1), 88–95.

Suzuki N. and Nakanishi J. (1995). Brominated analogues of MX (3-chloro-4-(dichloromethyl)-5-hydroxy-2-(5H)-furanone) in chlorinated drinking water. Chemosphere **30**(8), 1557–1564.

Swanson, W. J., Chowdhury Z. K., Aburto L, and Summers R. S. (2002). Disinfection practices and pathogen inactivation in ICR surface water plants. In In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 1999–223. American Water Works Association Research Foundation, Denver.

Symons J. M., Stevens A. A., Clark R. M., Geldreich E. E., Love O. T., and DeMarco J. (1981). Treatment Techniques for Controlling Trihalomethanes in Drinking Water. EPA-600/2-81-156. U.S. EPA, Cincinnati.

Symons J. M.; Krasner S. K.; Simms L. A., and Sclimenti M. J. (1993). Measurement of THM and precursor concentrations revisited: the effect of bromide ion. Journal American Water Works Association **85**(1), 51–62.

Symons J. M; Krasner S. K., Sclimenti M. J., Simms L. A., Sorensen H. W., Speitel G. E., and Diehl A. C. (1996). Influence of bromide ion on trihalomethane and haloacetic acid formation. In Disinfection Byproducts in Water Treatment: The Chemistry of Their Formation and Control (eds. R. A. Minear and G. L. Amy), pp. 91–130. CRC Press Inc., Boca Raton.

Symons J. M. (1999). Disinfection by-products: a historical perspective. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 1-26. American Water Works Association, Denver.

Taylor D. J., Kupper L. L., Rappaport S. M., and Lyles R. H. (2001). A mixture model for occupational exposure mean testing with a limit of detection. Biometrics **57**(3), 681–688.

Townend J. (2002). Practical Statistics for Environmental and Biological Scientists. John Wiley and Sons, Chichester.

Trehy M. L. and Bieber T. I. (1981). Detection, identification and quantitative analysis of dihaloacetonitriles in chlorinated natural waters. In Advances in the Identification and Analysis of Organic Pollutants in Water, Vol. 2 (ed.), pp. 941–975. Ann Arbor Science Publishers, Ann Arbor.

Trussell R. R. (1992). Control strategy 1: alternative oxidants and disinfectants. In Proceedings of the Annual Conference and Exposition. American Water Works Association, Denver.

U.S. Environmental Protection Agency (1979). National interim primary drinking water regulations: control of trihalomethanes in drinking water. Federal Register **44**(231), 68624-68707, November 29.

U.S. Environmental Protection Agency (1994a). National primary drinking water regulations; disinfectants and disinfection byproducts; proposed Rule. Federal Register **59**(145), 38668–38829, July 29.

U.S. Environmental Protection Agency (1994b). National primary drinking water regulations: monitoring requirements for public drinking water supplies: Cryptosporidium, Giardia, viruses, disinfection byproducts, water treatment plant data and other information requirements; proposed rule. Federal Register **59**(28), 6332–6444, February 10.

U.S. Environmental Protection Agency (1996a). National primary drinking water regulations: monitoring requirements for public drinking water supplies: Cryptosporidium, Giardia, viruses, disinfection byproducts, water treatment plant data and other information requirements; final rule. Federal Register **61**(94), 24354–24388, May 14.

U.S. Environmental Protection Agency (1996b). DBP/ICR Analytical Methods Manual. EPA 814-B-96-002, U.S. Environmental Protection Agency, Cincinnati.

U.S. Environmental Protection Agency (1998a). National primary drinking water regulations: disinfectants and disinfection byproducts; final rule. Federal Register **63**(241), 69390–69476, December 16.

U.S. Environmental Protection Agency (1998b). National primary drinking water regulations: Interim enhanced surface water treatment rule; final rule. Federal Register **63**(241), 69478–69521, December 16.

U.S. Environmental Protection Agency (2000a). Stage 2 microbial and disinfection byproducts federal advisory committee agreement in principle. Federal Register **65**(251), 83015–83024, December 29.

U.S. Environmental Protection Agency (2000b). ICR Auxiliary 1 Database Version 5.0, Query Tool Version 2.0. EPA 815-C-00-002 (CD-Rom), U.S. Environmental Protection Agency, Cincinnati.

U.S. Environmental Protection Agency (2000c). ICR Treatment Study Database Version 1.0. EPA 815-C-00-003 (CD-Rom), U.S. Environmental Protection Agency, Cincinnati.

U.S. Environmental Protection Agency (2003). National primary drinking water regulations: stage 2 disinfectants and disinfection byproducts rule; national primary and secondary drinking water regulations: approval of analytical methods for chemical contaminants. Federal Register **68**(159), 49548–49681, August 18.

U.S. Environmental Protection Agency (2006a). National primary drinking water regulations: stage 2 disinfectants and disinfection byproducts rule; final rule. Federal Register **71**(2), 388–493, January 4.

U.S. Environmental Protection Agency (2006b). National primary drinking water regulations: long term 2 enhanced surface water treatment rule; final rule. Federal Register **71**(3), 654–786, January 5.

Valentine R. L., Choi J., Chen Z., Barrett S. E., Hwang C., Guo Y., Wehner M., Fitzsimmons S., Andrews S., Werker A. G., Brubacher C., and Kohut, K. (2005). Factors Affecting the Formation of NDMA in Water. American Water Works Association Research Foundation and American Water Works Association, Denver.

Veenstra J. N. and Schnoor J. L. (1980). Seasonal variations in trihalomethane levels in an Iowa river water supply. Journal American Water Works Association **72**(10):583–590.

Wajon J. E. and Morris J. C. (1980). Bromamination chemistry: rates of formation of NH2Br and some N-bromamino acids. In Water Chlorination: Environmental Impact and Health Effects, Vol. 3 (eds. R. L. Jolley, W. A. Brungs, and R. B. Cumming), pp. 171–181. Ann Arbor Science Publishers Inc., Ann Arbor.

Weinberg H. S., Brophy K. S., and Singer P. C. (2000). Aspects of quality control during analysis of HAA9 in drinking water. In Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.

Weinberg H. S., Krasner S. W., Richardson S. D., and Thruston A. D. (2002). The occurrence of disinfection by-products (DBPs) of health concern in drinking water: results of a nationwide DBP occurrence study. EPA/600/R-02/068, U.S. Environmental Protection Agency Office of Research and Development, National Exposure Research Laboratory, Athens, GA.

Werdehoff K. S. and Singer P. C. (1987). Chlorine dioxide effects on THMFP, TOXFP, and the formation of inorganic by-products. Journal American Water Works Association **79**(9), 107–113.

Westerhoff P., Debroux J., Amy G. L., Gatel D., Mary V., and Cavard J. (2000). Applying DBP models to full-scale plants. Journal American Water Works Association **92**(3), 89–102.

Westerhoff P., Chao P., and Mash H. (2004). Reactivity of natural organic matter with aqueous chlorine and bromine. Water Research **38**(6), 1502–1513.

Williams S. L.; Rindfleisch D. F.; and Williams R. L. (1994). Deadend on haloacetic acids. In Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.

Williams S. L.; Williams R. L.; and Gordon A. S. (1996). The impact of bacterial degradation of haloacetic acids (HAA) in the distribution system. In Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.

Wysock B. M., Cummins M. D., Fair P. S., Feige M. A., Sorrell R. K., Walasek J. B., Allgeier S. C., Shukairy H. M., Stultz-Karapondo M., and Bohl B. J. (2002). ICR implementation and data collection challenges. In Information Collection Rule Data Analysis. (eds. M. J. McGuire, J. L. McLain, and A. Obolensky), pp. 1–38. American Water Works Association Research Foundation, Denver.

Young J. S. and Singer P. C. (1979). Chloroform formation in public water supplies: a case study. Journal American Water Works Association **71**(2), 87–95.

Xie Y. and Reckhow D. A. (1996). Hydrolysis and dehalogenation of trihaloacetaldehydes. In Disinfection Byproducts in Water Treatment: The Chemistry of Their Formation and Control (eds. R. A. Minear and G. L. Amy), pp. 283–291. CRC Press Inc., Boca Raton.

Zavaleta J. O., Hauchman F. S., and Cox M. W. (1999). Epidemiology and toxicology of disinfection by-products. In Formation and Control of Disinfection By-Products in Drinking Water (ed. P. C. Singer), pp. 95–117. American Water Works Association, Denver.

Zhang X. and Minear R. A. (2002). Decomposition of trihaloacetic acids and formation of the corresponding trihalomethanes in drinking water. Water Research **36**(14), 3665–3673.

Zheng M., Andrews S. A., and Bolton J. R. (1999). Impacts of medium-pressure UV and UV/H2O2 treatments on disinfection byproduct formation. In Proceedings of the Water Quality Technology Conference. American Water Works Association, Denver.