EXECUTIVE SUMMARY

INTRODUCTION

Most efforts at controlling disinfection by-product (DBP) formation and allied biodegradable compounds have focused on the treatment plant and distribution system. Drinking water providers now have an array of tools for addressing this problem through enhanced removal of precursors, better process sequencing to minimize formation, use of alternative disinfectants to reduce the formation of chlorine-based DBPs, and distribution system management to minimize dead ends.

One area of drinking water system management that has not been widely discussed in the context of DBP control is management of the watershed and raw water quality. One of the reasons this has not been widely **<<<Text seems to be missing. Please provide.>>>**

Aquatic ecologists, limnologists, and geochemists have long studied the generation of dissolved organic carbon (DOC) and its cycling in the freshwater environment. ~~A great deal~~Much has been learned ~~on~~ about the sources, transport, and degradation of bulk organic matter and selected specific organic compounds in the aquatic environment; however, very few studies have been conducted ~~on~~ about the fate and transport of DBP precursors.

DBP PRECURSOR DATA

To help improve our current understanding of the full range of precursor levels in U.S. waters, three ~~different~~ types of data collection efforts were undertaken. The first stage ~~of this data collection~~ involved ~~a mail survey sent~~ mailing a survey to 553 of the country's largest utilities ~~using~~ who use surface water (chapter 3). The second stage called for telephone followups and site visits to selected utilities from the first list of survey recipients (chapter 3). ~~In order to get the most out of this work,~~ To obtain the most data from this work, information was also collected on organic carbon levels, source types, land use characteristics, and utility interest in these issues. Finally, an extensive literature survey was conducted, spanning the period from 1975 to the present (chapter 4).

A substantial number of utilities have collected DBP data beyond what has been mandated by regulatory agencies. About 17% of respondents had collected total organic halide (TOX) data beyond any that were required under the Information Collection Rule (ICR). About 40% had collected data on DBP precursors in their raw water (including trihalomethane [THM], haloacetic acid [HAA], and TOX precursors). Of these, 12 reported that they had collected precursor data prior to 1981. About 60% of the respondents reported that they had collected total organic carbon (TOC) or similar natural organic matter (NOM)-related data at locations in their watersheds. A slightly smaller number of utilities (~40%) reported that they had conducted watershed studies focused on storm events. The vast majority (90%) were willing to share their data. ~~Data from the~~ ICR data revealed that some free chlorinating utilities ~~to have~~had high ratios of known to unknown TOX (>50%) and some had~~have~~ quite low ratios (<15%). There was ~~not an~~no obvious connection between unknown TOX level and~~to~~ geographic location, raw water TOC, or specific ultraviolet absorbance (SUVA).

From these site-specific data, a set of about 50 primary utilities and about 120 additional secondary utilities were identified. Some ~~of these~~ were the focus of this research project;

however, time and resources did not permit exploration of all candidate utilities. ~~A great deal of~~Much unpublished data exist in the ~~hands~~ databases of these utilities. Given their expressed interest in watershed management for precursor control, it is recommended that additional studies be conducted to better develop ~~our~~ anunderstanding of relevant watershed processes. Any future studies ~~on~~ regarding NOM in watersheds should make use of these utility lists and the accompanying database ~~so as~~ to best capture the full range of organic water qualities.

<<<text truncated>>>

Water quality data collected by federal, state, and local agencies are stored in USEPA's STOrage and RETrieval (STORET) system. Data collected by Indian tribes, volunteer groups, academics, and others may also be stored in this system at their initiative. Because of the sheer volume of data, the database is operated as two distinct systems. Historical water quality data collected up to the end of 1998 are contained in the static Legacy Data Center (~~LCD~~LDC). Data collected starting in 1999 are maintained in the STORET system, which is actively populated with new water quality data. Some historical data (pre-1999) have been stored in both the LDC and STORET. ~~However, b~~Because this transfer is neither complete nor well documented, however, both systems must be searched to obtain comprehensive data for a region. The ~~LCD~~ LDC and STORET combined are the largest water quality database for the United States.

In order to build ~~our~~ a project-specific database, TOC and DOC data were downloaded from the ~~LCD~~ LDC and STORET. In these systems, data may only be downloaded by county or hydrologic unit. ~~Due to~~Because of these restrictions and the volume of data contained within the system that needed to be processed, only data available for specific regions of the country were downloaded. These regions were ~~defined~~ identified by ~~as the~~ 8-digit USGS-defined hydrologic unit codes **<<<Figure 9.6 shows both 7- and 8-digit numbers. Are digits missing in the figure?>>>** (HUCs) for the watersheds ~~which~~ that contained ~~our~~ the 60 utility watersheds included in this phase of the study. These 60 utilities were selected in part so that data were collected for at least two utilities, and thus two HUCs, per ecoregion. Data were downloaded from a total of 160 HUCs that ~~which~~ contained our 60 utility watersheds. Thus, data acquisition from USEPA for the continental United States is currently incomplete.**<<<Earlier, the number referenced in text was "about 50" and Table 3.3 shows 52 utilities, not 60. Which is correct?>>>**

As noted ~~above~~previously, surface water quality data collected by USGS have been assessed by other researchers for long-term trends in nutrient and heavy-metals concentrations. Though not as extensive, the USGS data set also contains a substantial amount of information on TOC and dissolved (filtered) organic carbon (DOC). These data are maintained in a separate database maintained by USGS, called the National Water Information System (NWIS), although some of these data are also transferred to STORET. In order to expand ~~our~~ the project database, TOC and DOC data collected by ~~the~~ USGS were downloaded directly from the NWIS. Data were downloaded by state, resulting in complete data acquisition across the continental United States. Coverage, however, is variable ~~due to~~because of state-by-state and year-to-year water quality monitoring schemes.

The data retrieved from the ~~LCD~~LDC, STORET, and the NWIS were combined and reviewed for extent and coverage. The TOC/DOC database was mined to identify stations consisting of TOC or DOC records collected at least quarterly for a minimum of 4 years. Data were divided into historical (1965–2000) and modern (1995–2004) data sets to capture both long-term and modern trends. **<<<These years overlap. Is there an error here? Also, does the**

modernized data set go beyond 2004? (i.e., to the present date?)>>>These criteria resulted in a total of 307 and 188 stations, respectively, in the historical and modern data sets. These combined data are useful for characterizing regional variability, seasonality, and long-term temporal changes in TOC levels across the nation. The methods used for TOC analysis are generally consistent between the two data sets; however, changes in analysis methods over time (not well documented) and detection limits (observable through data analysis) do exist.

As noted previously, a general survey was sent to 400 utilities. **<<<Earlier, the number was 553; which is correct?>>>**Survey results were used to select a set of 60 utilities for subsequent followup personal communications. Raw water data (TOC, DOC, SUVA, UV_{254}, metals, general organics), hydrologic data (reservoir characteristics, streamflow patterns), and GIS boundaries and land use data were requested from these utilities or gathered from other sources. Analyses of these TOC/DOC data are conducted separately ~~as~~ because methodologies, sampling frequencies, and detection limits tend to be more variable. Locations of the 60 utilities included in this phase of the study are shown in relation to USEPA ecoregion level II ~~definitions~~ designations in Figure 9.1. The watershed delineation for each utility is outlined in light blue. In Figure 9.2, the location and name of each utility is identified. In addition,~~ the USGS defined HUC or HUCs containing each utility~~ HUC locations for each utility are also shown, color coded to match the ~~color used to denote~~ ecoregions in Figure 9.1. Recall that data downloads from the ~~LCD~~ LDC and STORET were limited to these ~~HUCs~~HUC locations, ~~while~~ whereas USGS data were obtained for various locations across the United States. **<<<Do edits reflect intended meaning?>>>**The number of USGS stations included, however, was largely constrained by the period-of-record requirements.

For ~~the~~ a majority of ~~our~~ the analyses, TOC and DOC data were lumped together. This ~~is justifiable~~was justified because DOC typically comprises the majority of TOC, typically 90% to 95%. In addition, this combination of TOC and DOC data allowed ~~us to expand~~expansion of the spatial and temporal extent of the ~~our~~ project database.

Variability Within Ecoregions

Ecoregion 8.3 ~~covers~~ encompasses a large ~~extent~~ area of the eastern United States, including many major metropolitan areas. ~~A large number~~ Many of the 60 utilities included in phase II of ~~the~~ this study also lie within ecoregion 8.3. The HUCs within ecoregion 8.3 for which TOC data were obtained through STORET and the NWIS are shown in Figure 9.6. Because of the greater availability of data for this ecoregion within ~~our~~ the project-specific database, ecoregion 8.3 was selected for analyses to evaluate within-ecoregion variability in TOC concentrations.

Ecoregion 8.3 interquartile TOC ranges of each HUC contained within the ecoregion for which data were obtained are presented in Figure 9.7. Both the ~~l~~Legacy and modern data are included. Interquartile ranges for ~~the data when~~data that have been lumped together are provided as dashed lines on the figure for comparison purposes. As previously noted, the width of each box corresponds with the amount of data from which the statistics are drawn. In general, data for the HUCs within the ecoregion are well represented by the overall interquartile range for the region. Some exceptions ~~should be noted, including~~ include HUCs 3020103, 3020201, 3030002, and 1111~~0~~0207, whose 50th and 75th quartiles tend to be higher than for most of the remaining HUCs. HUC 2~~0~~040203 has an interquartile range almost entirely below that for the combined ecoregion data.**<<<Comparing to Figure 9.7, correct to change these HUCs to 11110207 and 2040203?>>>**

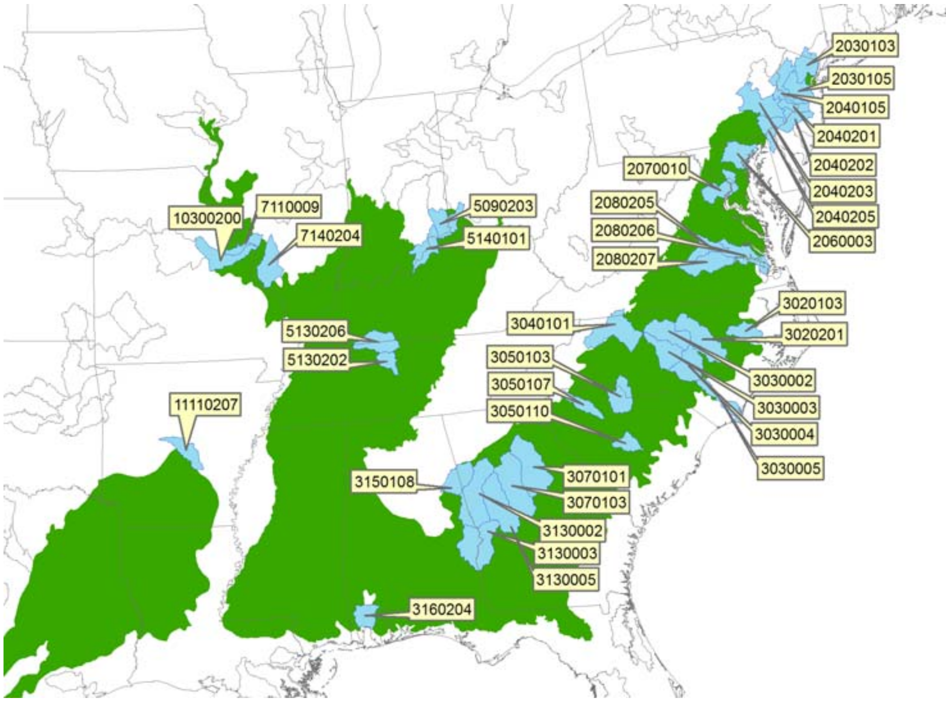

Figure 9.6 Ecoregion 8.3 ~~HUC locations and numbers~~HUCs and locations

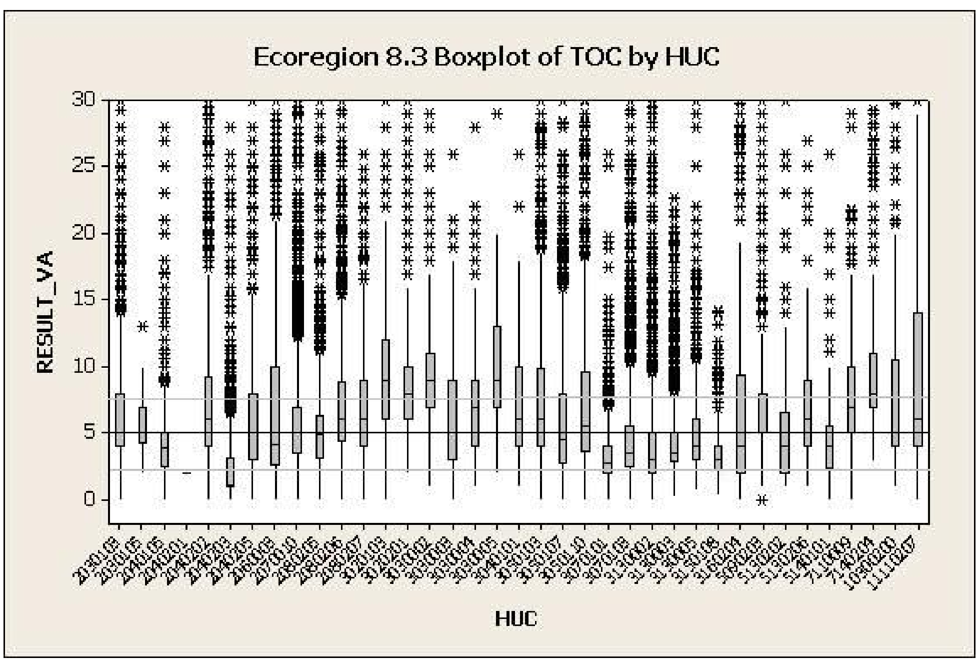

Figure 9.7 Box-plot representation of interquartile range of data for each HUC within ecoregion 8.3

<<<text truncated>>>

The Mann–Kendall technique operates through the analysis of paired sets of data, in this case a unique time value (the independent variable) and the TOC measurement (the dependent variable). The TOC measurements need not be unique, as the Mann–Kendall technique generates a statistical value representative of the monotonic relationship between the independent (time) and dependent (TOC) variables. The governing equation for the Mann–Kendall technique is defined in Equation 9.1, where the ~~S~~summary-test (*S*-test) statistic is a summation of the “discordant pairs” (decreases) and the “concordant pairs” (increases).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sign}\left[\left(X_j - X_i\right) * \left(Y_j - Y_i\right)\right] \qquad (9.1)$$

where S = ~~the~~ summary-test statistic
X_i, X_j = ~~the~~ independent variables (time)**<<<Edits to left of equal sign acceptable?>>>**
Y_i, Y_j = ~~the~~ dependent variables (organic carbon)

The “sign” function within Equation 9.1 is simply a function that converts discordant-pair results to a value of minus one (–1) and concordant-pair results to a value of positive one (+1). This function eliminates false predictions via removal of magnitude of change in trend, outliers, and other variability factors (Millard and Neerchal 2001).

$$\text{sign}(g) = \begin{cases} 1, g > 0 \\ 0, g = 0 \\ -1, g < 0 \end{cases} \qquad (9.2)$$

where g is the operator within the sign function.

The null hypothesis associated with the Mann–Kendall test states that no trend exists, ~~while~~ whereas the alternative hypothesis states that a trend does exist~~s~~. The null hypothesis is true if the *S*-test statistic is near or equal to zero and the alternative hypothesis is rejected. ~~However, t~~The null hypothesis is rejected, however, if the *S*-statistic is significantly different from zero and the alternative hypothesis is accepted. When the null hypothesis is rejected, a resulting negative *S*-test statistic indicates a decreasing trend within the data with time and, conversely, a positive *S*-test statistic indicates an increasing trend with time.

Kendall’s ~~τtau~~, a correlation coefficient, provides a confidence level associated with the ~~S summary~~ *S*-test statistic. Kendall’s ~~τtau~~ is calculated as shown in Equation 9.3.

$$\tau = \frac{S}{n(n-1)/2} \qquad (9.3)$$

where ~~*S* = the summary statistic~~**<<<deleted because it was defined in Eq. 9.1>>>**
n = ~~the~~ total number of measurements within the data set
~~τtau~~ = ~~the~~ correlation coefficient

An absolute value of ~~τtau~~ less than 0.01 indicates a high confidence in the trend postulated by the Mann–Kendall technique, ~~while~~ whereas a τ ~~tau~~ value greater than 0.1 indicates little or no confidence in the predicted trend.

<<<*text truncated*>>>

APPENDIX ~~1~~A
TELEPHONE SURVEY QUESTIONS

Telephone followups were performed by faculty and graduate students or ~~post docs~~postdoctoral scholars. ~~We will~~The intent was to ask for some elaboration on questions asked in the written survey, as well as new topics not ~~really covered~~ included in the ~~earlier~~ survey. This appendix includes the general guidelines and instructions given to university interviewers for conducting the followup telephone surveys.

GENERAL INSTRUCTIONS

You'll be ~~assigned a few utilities to call with~~ given telephone numbers and contact names in order to call utilities and conduct telephone surveys. The numbers and names may or may not be up to date, so you may have to ask whomever answers the telephone ~~if~~ whether there are other ~~people~~ individuals you should talk to.

Please make careful notes of your discussions, and include these in a file ~~on~~ about the utility. You ~~will~~ are to start the file even before you make your first call, using information available at the university and on the Web (see below). It is important to recognize that utility personnel are very busy~~,~~ and ~~that~~ their time is valuable.

Preliminary Work

It is important to do ~~a little~~some preliminary research ~~on~~ about the utilities ~~you're~~ you'll be calling. This will make your call more effective~~,~~ and will minimize~~s~~ the time that a utility person will have to spend on the interview.

1. Look in the compiled utility notebooks. There is ~~some~~ miscellaneous information ~~on~~ about the water treatment plants, utilities, water quality, and watersheds. You'll also find copies of the returned surveys ~~here~~. Look through the notebooks, and make notes or photocopies if there is ~~something~~ documentation you want to have in front of you when you call.

2. Perform a Google search on the water utility. There's no need to spend a lot of time doing this; just try a few phrases to see if there is anything on the Web that would be helpful to ~~us~~the project. Suggestions ~~on~~ for search phrases~~:~~ include

- Drinking water treatment for "city name"
- Potable water for "city name"
- Watershed for "city name"
- Water supply for "city name"