Assessment of Mitigation Strategies for Disinfection By-Product Formation through Bayesian Network Modeling

Zoe J. Y. Zhu, E. McBean and Brett Harper

University of Guelph, Guelph, Ontario

Abstract

Issues of disinfection by-product (DBP) formation at water treatment plants (WTP) in response to chlorination in drinking water treatment systems are common. Controlling the rate of DBP formation is complicated by the presence of zebra mussels, which may inhabit the raw water intake of WTPs. Chlorination to control zebra mussel populations may exacerbate the formation of DBPs. A Bayesian network is developed using the WEBWEAVR-IV Toolkit, utilizing causal relationships between raw water quality parameters in the form of conditional probabilities. Three alternative chlorination scenarios are analyzed, one of which demonstrates the probability of high total trihalomethanes (TTHM) (>0.08 mg/L) can reduce the health risk by 6% to 9%. Given the input parameters of an average WTP, raw water concentrations from the data from 2005 to 2008, the health risk has a 14% chance of being less than ‘de minimus’ (one in a million).

Keywords: disinfection by-products, total trihalomethanes, Bayesian network, dissolved organic carbon (DOC), chlorine dose.

1 Introduction

Disinfection by-products, and specifically trihalomethanes (THMs), are formed during the chlorination reaction of waters containing precursor organics, predominantly dissolved organic carbon (DOC) and inorganic ions. Of the THM subspecies, chloroform (CF) is usually found in the highest concentrations. However, when the bromide ion is present in the source water, three brominated halogenes are produced: bromodichloromethane (BDCM), dibromochloromethane (DBCM), and bromoform (BF). A number of researchers confirm that bromide concentration is a key factor in determining brominated-THM formation and the speciation of THMs (e.g. Jolley 1978; Chang et al. 2001; Hong et al. 2007; Sun et al. 2009; Harper 2012). The lifetime cancer risks through oral ingestion and dermal absorption for BDCM made the highest percentage contribution (59%) to total health risks compared to the other subspecies of THMs. (Lee et al. 2004). Health Canada developed guidelines for THMs and BDCM, as shown in Table 1. There are not enough data to develop guidelines for other THMs species (Wang et al. 2007).

Table 1  Health Canada THM guidelines (Health Canada 2006, 2012).

| Health Canada THM Guidelines |  |
|-----------------------------|--|---|
| THMs | 0.1mg/L |
| BDCM | 0.016mg/L |

The decreasing of BDCM concentrations is important as one of the strategies to reduce the human health risk arising from drinking water. The degree of health risk is also impacted by age and lifestyle.

DBP formation varies with the quality of the source water, as a result of the DOC and inorganic (bromide ion) precursors, as well as chlorine dose, contact time, temperature and pH (Heller-Grossman et al. 1993; Amy et al. 1998). Chlorine is the most common disinfectant used in water treatment. The popularity of chlorine arises due to the combination of low cost, high oxidizing potential, and the chlorine residual which exists within the distribution system protecting against microbial recontamination (Morrow and Minear 1987; Sadiq and Rodriguez 2004; McBean et al. 2008). However, there is merit in reducing the formation of these by-products, if feasible, by altering chlorination dosages. An additional issue which may counter the potential to alter chlorination dosages is the presence of zebra mussels (Dreissena polymorpha), a class of mollusc similar to oysters, clams, and scallops, originating from the Black and Caspian Seas, which have been transported into the Great Lakes water by cargo ships. The mussels grow to 1 in. long and produce 35 000 eggs per season per female (Dermott and Munawar 1993). The proliferation of zebra mussels is causing serious problems by clogging raw water intakes and discharge lines, increasing pipe corrosion and producing massive bio-fouling. Zebra mussels have subsequently become a problem for surface water treatment facilities in parts of the United States and Canada, particularly along the western
shores of Lake Erie. This paper evaluates several chlorination strategies to accomplish lesser formation rates of DBPs while still controlling zebra mussels. In this research, a Bayesian network is created to model the kinetics of the halogenation reactions to mitigate the formation of brominated THMs.

2 Literature Review and Background

2.1 Known Health Effects of DBPs

DBPs in drinking water may lead to potential human health risks and many of the DBPs have been classified as probable or possible carcinogens (Kronberg and Christman 1989; Nikolaou and Lekkas 2001). Table 2 outlines the standards associated with different classes of DBPs and health advisories.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Case number</th>
<th>Status: Regulation</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dibromochloromethane</td>
<td>124-48-1</td>
<td>Final</td>
<td>0.06</td>
</tr>
<tr>
<td>Bromodichloromethane</td>
<td>75-27-4</td>
<td>Final</td>
<td>Zero</td>
</tr>
<tr>
<td>Bromoform</td>
<td>67-66-3</td>
<td>Final</td>
<td>0.08</td>
</tr>
<tr>
<td>Bromate</td>
<td>7789-38-0</td>
<td>Final</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 2 The health advisories related to DBPs (modified from USEPA 2006).

2.2 Research on chlorination and zebra mussel’s control

Verween et al. (2009) have shown that chlorination is an effective tool for eliminating zebra mussels. One strategy to control zebra mussels is continuous treatment, where chlorine is applied consistently to the water at concentrations around 2 mg/L at the intake.

Alternatively, zebra mussel populations may be controlled by either intermittent treatment, or periodic control chlorination (Klerks et al., 1996). Intermittent treatment involves killing the larvae before they settle and change into their more resilient juvenile forms. Sprecher and Getsinger (2000) demonstrate that chlorinating for 30 min every 12 h is effective in controlling veliger (zebra mussel larvae) populations and preventing new zebra mussels from settling; however, this approach is not effective in killing established adult zebra mussels. By chlorinating heavily for two to three weeks in late spring or early summer (in addition to the preceding), the established adults can be killed. Intermittent treatment can be used to kill the young zebra mussels looking to settle and hence represents an important scenario as described below.

It has been demonstrated that chlorination rates can be altered by incorporating knowledge of the changing seasonal tolerances of zebra mussels (Fisher et al. 1994; Costa et al. 2008). An alternative that makes use of this knowledge is periodic treatment, which involves chlorinating on and off throughout the year. This treatment usually consists of three treatment applications spread across the months of April through October. Each treatment gives a dose of chlorine between 0.5 mg/L and 2 mg/L and lasts for 14 d to 28 d. Costa et al. showed that in one experiment, a dose of 0.3 mg/L was applied in 3 sets of 14 d to 21 d periods, resulting in a 95% mortality rate for zebra mussels.

2.3 Bromide Ion

Health Canada has listed the bromide ion as one of the contributing factors in the formation of THMs (Williams et al. 1995). In the presence of chlorine and organic material, as much as 50% of the bromide ion may become incorporated into brominated trihalomethane subspecies (Chang et al. 2001; Harper 2012). Until recently, only very limited bromide data have been available (Kampioti and Stephanou 2002; Uyak and Toroz 2007). However, bromide data over a 4 y period (2005–2008) have recently been released by the Ontario based Drinking Water Surveillance Program (DWSP). This study utilizes this recent bromide data in the development of a Bayesian network.

2.4 Bayesian Networks

This study employs the use of a Bayesian network to predict THMs by utilizing causal relationships between water quality parameters and THMs, as well as causal relationships between the water quality parameters themselves.

Bayesian networks are a type of expert system that represents domain knowledge with a graphical structure that uses nodes to represent variables in a system and arcs between those nodes to represent causal or dependent relationships as well as the strength of uncertain relationships among variables. A Bayesian network quantifies this knowledge structure with probabilistic expressions of the interaction among domain elements, as probabilities are intuitive ways for experts to express
uncertainty. The additivity axiom of probability theory is: If event A and event B are mutually exclusive, then the probability of A or B is the summation of the probability for A and the probability for B. It can be applied to the variable healthrisk. Let A be the event with healthrisk between 1/50 000 and 3/50 000, and B be the event that healthrisk is between 3/50 000 and 7/50 000. Then the chance that healthrisk is between 1/50 000 and 7/50 000 is the chance of events A and B combined. The model of a Bayesian network may be obtained by combining learning and water expert knowledge.

Using a Bayesian model, the treatment plant may control pre-chlorine (PreCl2) and post-chlorine (PostCl2) dosages to minimize THM formation. Since the probabilistic parameters of a Bayesian network can be generated from observed evidence as well as from expert estimates, these networks can easily integrate different types of uncertainty into a single theoretical environment (Xiang 2002).

Bayesian networks rely on Bayes’ theorem to propagate information between nodes. In its simplest manifestation, the theorem is given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}$$

wherein the belief $P(H|E)$ in a hypothesis, given some evidence, depends on the likelihood of observing evidence, E, given both the hypothesis and its negation, $P(E|H)$ and $P(E|\neg H)$, and on the prior probability of the hypothesis $P(H)$ (Pearl 1988). The denominator expresses a normalizing constant that indicates the prior probability of the evidence, $P(E)$. The probabilities that Bayesian techniques rely on can be considered mathematical translations of the linguistic conditional expressions that humans regularly use, making Bayesian methods very effective representations of subjective beliefs.

The Bayesian network’s graph structure accommodates conditional independence between domain elements. It is not necessary to specify the relationship between every variable (and every hypothesis H) in an exhaustive joint probability matrix. Instead, each variable in a Bayesian network requires only a local joint probability distribution that reflects the possible configurations of its immediate parents, if any, in the graph. A node with no parents is said to be marginally independent. The marginal probabilities used to describe the states of such a node may depend on observed data or expert estimates, but are not directly influenced by other nodes in the network. By contrast, conditional probabilities are used to describe the various states of a node with parents, where each state of such a node is described in terms of the various combinations of its parents’ states. These probabilities are conditional on the knowledge of the parents’ states. The Bayesian approach thus limits the number of quantitative parameters needed to describe the probabilistic influence of domain elements on each other; conditional probabilities are described locally, but, through propagation of probabilities from one node to the next, and hence have global influence. The WEBWEAVR-IV Toolkit (Xiang 2006) will be employed to model a Bayesian network that utilizes the concentration levels of raw water parameters to predict levels of TTHM.

3 Study Area and Approach

3.1 WEBWEAVR-IV

The Bayesian network was created using the WEBWEAVR-IV Toolkit, a Java-based research toolkit developed for decision support systems based on graphical models. WEBWEAVR-IV supports the construction of Bayesian networks, inference in standard and dynamic Bayesian networks and decomposable Markov networks, construction and verification of multiply-sectioned Bayesian networks (MSBNs), and inference in multi-agent MSBNs.

3.2 Bayesian Network Developments

Shown in Figure 1 is the Bayesian network developed for TTHM and health risk prediction. Health risk is often measured as a probability that certain disease occurs to an individual in a population. Hence, the probability of health risk is a probability about probability, a second order probability. Although it differs from the common application of probability, namely the first order of probability, it is a standard practice in uncertain reasoning (Fenton and Neil 2012).

![Figure 1](https://example.com/figure1.png)

**Figure 1** Bayesian network to determine TTHM levels and resulting health risk.

The network displays how the various water quality parameters, namely health risk and zebra mussels, rely on the other parameters in the network. For example, dissolved organic carbon (DOC) is dependent on raw water temperature (temp) and zebra mussels (zebra); and TTHM are dependent on bromide and DOC in raw water, and pre- and post-chlorine dosages. We can see that the Bayesian network can explicitly represent the variables and causal relationships between the variables.
3.3 Conditional probabilities

The conditional probabilities for the Bayesian network were developed using data from six water treatment plants (WTPs) in Ontario during 2005 through 2008; five from surface water sources, and one from mixed surface water and groundwater sources. Use of the six WTPs was employed since only for these are all chlorination dosage data and bromide data available. The conditional probability tables (CPTs) are assigned to all child variables. For instance, the conditional probability is calculated for variable DOC:

\[ P(\text{DOC} = 6.3 \text{mg/L} | \text{Zebra} = \text{high, Temp} > 20 \degree \text{C}) = 0.9 \]

The CPT for DOC is shown in Table 3.

Table 3: Conditional probability table for DOC.

<table>
<thead>
<tr>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th>Zebra</th>
<th>Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.08</td>
<td>0.02</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>high</td>
<td>medium</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>medium</td>
<td>high</td>
</tr>
<tr>
<td>0.6</td>
<td>0.2</td>
<td>0.2</td>
<td>medium</td>
<td>medium</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>0.55</td>
<td>0.25</td>
<td>0.2</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
<td>low</td>
<td>medium</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>

The initial probabilities for temperature were developed using the temperature profile in the datasets. Zebra mussels were assigned a range of values which correlated with temperature and zebra mussel spawning rates, as zebra mussel populations tend to increase when the water is >12 \degree C.

Probability levels for other parent nodes of TTHM (bromine, pre- and post-chlorination) in the Bayesian network were determined from data and expert opinions. Given the range of values for the five parent nodes, as well as the ranges of each child node in the Bayesian network, the conditional probabilities for each child node were determined by calculating the probability of the dependent water quality parameter, given the value of the parent water quality parameter.

3.4 The Seasonal Impact on Zebra Mussel Control and Health Risk Classification

To effectively control zebra mussels using chlorination, the seasonal variation of zebra mussel susceptibility to chlorination patterns must be considered. According to experiments by Kilgour and Mackie (1993), Ram et al. (1996), Rajagopal et al. (2003), and Costa et al. (2008), the high susceptibility of zebra mussels to chlorine is observed in the United States to start in June and peak in July and August, after reproduction. This susceptibility occurs due to the mussels’ low body weight, high filtration activity and the high water temperatures. However, given Ontario’s very long, cold winters and short summers, the peaks of susceptibility move to between August and September (Harper 2012). Thus it is in August and September when zebra mussels are most susceptible to chlorine, and so is the optimal period to pre-chlorinate to reduce zebra mussel levels. Given this, consideration of both zebra mussel control and treatment strategies is feasible.

The health risk is quantified using the health advisories related to DBPs (USEPA 2006). For TTHM, the Bayesian network classified by the MCL, if the MCL >0.08 mg/L, is classified as high health risks. For THMs subspecies, the Bayesian network will assess using both MCLs and health risk values.

4 Integrated Analysis

4.1 Comparison of Altered Pre- and Post-Chlorination Levels for TTHM Reduction

Consider three scenarios with varying levels of pre- and post-chlorination in an effort to mimic alternative approaches to chlorination while also incorporating mussel control. The variables being examined are: pre-chlorination, post-chlorination, treated water temperature, raw DOC, and bromine. Using the Bayesian network, values of TTHM and cancer risk are calculated. The three scenarios are as follows:

Scenario 1

Pre-chlorination is set at regular (i.e. the same as current conditions). A level of chlorine between 1.15 mg/L and 2.3 mg/L is applied but post-chlorination was set to the low level (<1.28 mg/L).

Scenario 2

During the months of May, July and September, pre-chlorination is set to 0.5 mg/L and the rest of the year to 0.001 mg/L, and post-chlorination to 3 mg/L. This test follows intermittent treatment (Sprecher and Getsinger 2000), where there are 3 treatments/y using pre-chlorination.

Scenario 3

This scenario is quite different from the preceding scenarios. In this scenario, continuous treatment is used during the months where the water temperature >12 \degree C. During the months of May to November the dose of pre-chlorine is set to 3 mg/L, at 0.001 mg/L for the remainder of the year. During the months of May to November the dose of post-chlorine is set to 0.3 mg/L and 3.3 mg/L for the remainder of the year. This scenario is designed to be the most secure test in terms of microbial disinfection and zebra mussel control. The pre-chlorine dose and post-chlorine dose applied for each scenario is summarized in Table 4.

Table 4: Summary of chlorine dosage for each scenario.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>1-12</td>
<td>5-7, 9</td>
</tr>
<tr>
<td>PreCl</td>
<td>1.15 mg/L-2.3 mg/L</td>
<td>0.5 mg/L-2 mg/L</td>
</tr>
<tr>
<td>Post Cl₂</td>
<td>&lt;1.28 mg/L</td>
<td>3 mg/L</td>
</tr>
</tbody>
</table>
4.2 Bayesian Network Results

The probability results generated by simulation using the Bayesian network are shown in Table 5.

Table 5  Bayesian network results.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>zebra</th>
<th>Temp</th>
<th>bromide</th>
<th>DOC</th>
<th>TTHM</th>
<th>Health risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>High</td>
<td>high</td>
<td>90% high</td>
<td>70%</td>
<td>high</td>
<td>50%</td>
</tr>
<tr>
<td>low</td>
<td>Low</td>
<td>low</td>
<td>4%</td>
<td>16%</td>
<td>3%</td>
<td>14%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario 2</th>
<th>zebra</th>
<th>Temp</th>
<th>bromide</th>
<th>DOC</th>
<th>TTHM</th>
<th>Health risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>High</td>
<td>high</td>
<td>90% high</td>
<td>79%</td>
<td>33%</td>
<td>20%</td>
</tr>
<tr>
<td>low</td>
<td>Low</td>
<td>low</td>
<td>4%</td>
<td>33%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario 3</th>
<th>zebra</th>
<th>Temp</th>
<th>bromide</th>
<th>DOC</th>
<th>TTHM</th>
<th>Health risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>High</td>
<td>high</td>
<td>90% high</td>
<td>76%</td>
<td>high</td>
<td>58%</td>
</tr>
<tr>
<td>low</td>
<td>Low</td>
<td>low</td>
<td>4%</td>
<td>33%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

The results in Table 5 display the predicted probability of each outcome. Three alternative chlorination scenarios are analyzed. Scenario 1 is the best choice compared to the other two scenarios. In the summer months, the probability of a high TTHM level (>0.08 mg/L) can be reduced by 6% to 9%.

In the winter months, when bromide is high, the probability of a high TTHM level (>0.08 mg/L) can be reduced by 17% and the probability of high cancer risk can be reduced by 13%. When bromide is low in winter, the probability of a high TTHM level (>0.08 mg/L) can be reduced by 10%. Of interest is that given the input parameters of average WTP raw water concentrations from the data from 2005 through 2008, health risk has a 14% chance of being less than de minimus (one in a million). The low temperature and low bromide level make the contribution to the low health risk exposure to the DBPs in Ontario.

5 Conclusions

Under current, average WTP conditions, health risk in response to zebra mussel treatment scenarios is found to be above de minimus (10⁻⁶) risk using the Bayesian network developed to predict TTHM levels.

One chlorination scenario (Scenario 1) demonstrates that the probability of a high TTHM concentration level (>0.08 mg/L) can be reduced 6% to 9%.

Given the input parameters of average WTP raw water concentrations from the data from 2005–2008, health risk has a 14% chance of being less than de minimus (one in a million).

The research also indicates that brominated DBPs are negatively correlated with chlorinated DBPs. With elevated levels of bromide level, increased brominated THMs occur and decreased CHCl₃ results, and there is a concomitant increase in health risk.

Acknowledgments

The authors would like to thank the DWSP of Ontario for their contributions in the data collection phase. This research was funded by the Canada Research Chair Program and the Ontario Research Foundation.

References


