Characterization of dissolved organic matter for prediction of trihalomethane formation potential in surface and sub-surface waters

John Awad, John van Leeuwen, Christopher Chow, Mary Drikas, Ronald J. Smernik, David J. Chittleborough, Erick Bestland

Abstract

Dissolved organic matter (DOM) in surface waters used for drinking purposes can vary markedly in character dependent on their sources within catchments. The character of DOM further influences the formation of disinfection by products when precursor DOM present in drinking water reacts with chlorine during disinfection. Here we report the development of models that describe the formation potential of trihalomethanes (THMFP) dependent on the character of DOM in waters from discrete catchments with specific land-use and soil textures. DOM was characterized based on UV absorbance at 254 nm, apparent molecular weight and relative abundances of protein-like and humic-like compounds. DOM character and Br concentration (up to 0.5 mg/L) were used as variables in models (R² > 0.93) of THMFP, which ranged from 19 to 649 μg/L. Chloroform concentration (12 – 594 μg/L) and relative abundance (27 – 99%) were first modeled (R² > 0.85) and from these, the abundances of bromochloromethane and chlorodibromomethane estimated using power and exponential functions, respectively (R² > 0.98). From these, the abundance of bromoform is calculated. The proposed model may be used in risk assessment of catchment factors on formation of trihalomethanes in drinking water, in context of treatment efficiency for removal of organic matter.

Crown Copyright © 2016 Published by Elsevier B.V. All rights reserved.

Keywords:
Catchment runoff
DOM character
THMFP

Abbreviations:
AromDOC, aromatic-organic compounds concentration; Br, bromide ion concentration in raw waters; CHBr3, chloro-bromo-methane formation potential; CHC13Br, bromo-chlorodichloromethane formation potential; CHCl3, chloroform formation potential; DOC, dissolved organic carbon; DOM, dissolved organic matter; FA, percentage abundance of fulvic-like; HA, percentage abundance of humic-like; MW, molecular weight; NOM, natural organics matter; NonAromDOC, non-aromatic-organic compounds concentration; P(1), area under Peak 1; P(2–3), area under peaks 2 & 3; P(4–5), area under peaks 4 & 5; P(6–7), area under peaks 6 & 7; P(8–9), area under peaks 8 & 9; PI, percentage abundance of protein1-like; PII, percentage abundance of protein2-like; R², coefficient of determination; SE, standard deviation; SMP, percentage abundance of soluble microbial protein-like; SUVA, Specific UV absorbance; THMFP, trihalomethane formation potential; UV254, UV absorbance at wavelength 254 nm.

* Corresponding author: Natural & Built Environments Research Centre, University of South Australia, H3-26 Mawson Lakes Campus, Adelaide, South Australia 5095, Australia. Fax: +61 8 830 23386.
E-mail address: John.vanleeuwen@unisa.edu.au (J. van Leeuwen).

http://dx.doi.org/10.1016/j.jhazmat.2016.01.030
0304-3800 (C) 2016 Published by Elsevier B.V. All rights reserved.

Highlights

• Models of THMFP based on the character of precursor organics were developed.
• UV254, HPSEC-UV and F-EEM provided data that could be used for THMFP prediction.
• High removal of humic compounds by enhanced coagulation led to lower THM formation.

Article info

Article history:
Received 19 October 2015
Received in revised form 1 January 2016
Accepted 12 January 2016
Available online 15 January 2016

Keywords:
Catchment runoff
DOM character
THMFP

Journal of Hazardous Materials 308 (2016) 430–439

Contents lists available at ScienceDirect
Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

CrossMark
1. Introduction

Water quality in catchment-fed reservoirs is largely influenced by land management practices within the catchment [1], vegetation types and loading [2–4], topography [5], climate [2,6] and soil types [7–9]. Catchment runoff and stream waters generally have higher concentrations of organics that are more aromatic and humic in nature than of shallow and deep groundwater [10]. This is attributed to either absorption of high molecular weight [11], hydrophobic and aromatic compounds [12] to clay minerals in soils or by microbial modification [13,14] as dissolved organic matter (DOM) moves through the soil profile. Nelson et al. [15] found that differences in dissolved organic carbon (DOC) concentrations within a stream are dependent on the adsorption capacities of the soil types present. Surface horizons high in clay tend to exhibit stronger adsorption characteristics than coarse textured soils [16].

DOM present in catchment water runoff is also influenced by vegetation type, with the highest DOM concentrations having been reported from catchments under forest cover, followed by grassland and then arable soils [17,18]. Naidu et al. [7] reported that DOC in water flows along the A horizon/B horizon boundary and that the concentration under pine was at least twice that under native woodland or pasture. Awad et al. [4] reported that DOC concentrations in water flow from the Australian native vegetation (mainly Eucalyptus) and Pinus radiata were at least twice that of the grassed zero order catchments (ZOCs). The highest concentration of DOC released into subsurface waters and lateral flows occurs after the first major rainfall event in the break-of-season during late autumn/early winter, demonstrating the importance of fresh litter material in generating high amounts of soluble organic matter [19].

Organics present in drinking water sourced from reservoir catchments can pose significant health risks that need to be addressed. These risks include formation of disinfection by-products (DBP), including trihalomethanes (THM), haloacids, iodo-acids, halonitromethanes, nitrosamines, and dichloroacetamide that form when natural organics matter (NOM) reacts with chemical disinfectants [20,21]. Chow et al. [19] reported that significant amounts of DBP precursors in sub-surface waters can be attributed to the leaching of organics from deciduous litter of surrounding vegetation, emphasizing the influence of catchment sub-surface water on reservoir water quality and consequently on DBP formation. The concentration and character of DOM significantly influences raw and treated water quality and DBP formation [9,20,22].

In raw (untreated) waters with high relative abundances of hydrophobic, aromatic and humic compounds, there is generally greater formation of THM following chlorination than in treated waters with compounds of a non-aromatic and hydrophilic character [23,24]. In contrast, Zhao et al. [25] reported that in river water (Pearl River, Guangzhou, China), low molecular weight DOM compounds were the main THM precursors. Besides organic matter, the physical and chemical factors of water also influence the formation of THM. These include bromide (Br) ion concentration, pH and temperature [26].

Factors such as temperature and pH have consistent effects on THM formation [27], whereas effects of DOM concentration and its character are less well understood. The formation of THM is site-specific and varies in response to the character and amount of DOM that has reacted [28]. Water quality parameters such as total or dissolved organic carbon (TOC and DOC, respectively) provide general information on the concentration of precursor material but provide no information on the character of the organics unless other data such as UV–vis absorbance are also included [29]. Hence, incorporating generic parameters such as DOC, TOC or UV absorbance at 254 nm (UV254) into models that describe DBP formation provide relatively limited information on the reactivity of precursor material. A number of studies [30–33] have been reported in which mathematical models were developed to predict THM or THM formation potential (THMFP) based on these generic organic parameters. Few studies have been conducted to develop models to predict THM or THMFP dependent on the character of DOM, such as the specific UV absorbance (SUVA) value (the ratio of absorbance at 254 nm/m to DOC concentration) [34], or relative proportions of fulvic acid [35] and humic-like compounds from fluorescence excitation and emission (F–EEM) spectroscopy [33].

The aim of the study reported here was to develop models for prediction of THMFP and the four compounds, chloroform, bromodichloromethane, chlorodibromomethane and bromoform, based on the character of precursor organics in runoff and subsurface waters sourced from discrete catchments. These were no representative drinking water reservoir with high loading of natural organic matter. In this study, DOM in water samples was quantified by measurement of DOC and UV–vis absorbance and characterized by F–EEM and High Performance Size Exclusion Chromatography (HPSEC) with UV-detection.

2. Methodology

2.1. Sampling and analysis

2.1.1. Water samples for model development

Water samples were collected from six ZOCs with three land management practices (native vegetation, pine plantation, grasslands) with varying surface soil textures of the Myponga reservoir-catchment, South Australia. Water samples were also collected from the Myponga River and Myponga Reservoir. Land-use, predominant vegetation and soil descriptions of ZOCs have been previously reported [4]. Catchments were instrumented with surface runoff collection and sub-surface (~60 cm for most samples) devices, as detailed by Awad et al. [4,36].

Water samples were diluted with high purity Milli-Q water (MQW) to achieve a DOC <20 mg/L. These were treated by coagulation/flocculation at ambient temperature and pH 6 ± 0.1 [37] using aluminium sulfate (alum; Al2(SO4)3·18H2O). A high dose (HD) of alum (>2 times the estimated enhanced dose by mEnCo model [38]) was used to distinguish DOM that could be removed by coagulation from that which is highly recalcitrant to removal.

2.1.2. Water samples for model evaluation

Water samples used for model evaluation were collected from two discrete ZOCs with two land management practices (re-vegetated with Australian native vegetation (8 years) and grasslands), of the Mount Bold reservoir-catchment, South Australia and from a ZOC with pine plantation of the Happy Valley reservoir-catchment, South Australia. Soil samples were collected from each catchment at different topographic positions (upper slope and low slope) and at three soil depths (0–10, 20–30 and 50–60 cm). These were air-dried and passed through a 2 mm sieve. Percentages of clay, silt and sand were determined by mid infrared (MIR) diffuse reflectance spectroscopy and
| Landuse          | Description                          | Depth (cm) | Particle size distribution | Texture                |
|------------------|---------------------------------------|------------|---------------------------|------------------------|
|                  |                                       |            | % Sand | % Silt | % Clay |            |
| Native vegetation| *Eucalyptus leucocory, E. fasciculosa*| 0–10       | 79.3  | 7.4   | 13.3   | Sandy loam over clay |
|                  | Woodland with *E. camaldulensis*      | 20–30      | 79.4  | 4.8   | 15.8   |            |
| Grass            | *Phalaris* and sub clover              | 0–10       | 93.2  | 1.1   | 5.7    | Sand over sandy loam |
|                  | based pasture                         | 20–30      | 97.6  | 1.0   | 1.4    |            |
|                  |                                       | 50–60      | 81.3  | 5.9   | 12.8   |            |
| Forest           | Pine plantation—*P. halepensis* (Aleppo pine) | 0–10       | 87.8  | 2.8   | 9.4    | Sandy loam over clay |
|                  |                                       | 20–30      | 86.6  | 4.3   | 9.1    |            |
|                  |                                       | 50–60      | 51.6  | 6.3   | 42.1   |            |

Table 2

Mean, maximum and minimum values of water quality parameters and THM formation potentials of surface waters, catchment runoff and subsurface waters.

| Parameters<sup>a</sup> | Unit | Surface | Runoff | Subsurface | Evaluation |
|-------------------------|------|---------|--------|------------|------------|
|                         |      | Mean    | Max.   | Min.       | Mean       | Max.       | Min.       | Mean       | Max.       | Min.       |
| n (mg/L)                | cm<sup>–1</sup> | 0.29    | 0.63   | 0.04     | 0.23       | 0.56       | 0.56       | 0.02       | 0.21       | 0.58       | 0.01     |
| DOC (mg/L)              |      | 9.97    | 19.06  | 2.28     | 5.76       | 12.92      | 1.11       | 5.82       | 12.13      | 1.08       |          |
| PI (%)                  |      | 3       | 10     | 1        | 7          | 21         | 1          | 7          | 22         | ND        |          |
| PII (%)                 |      | 17      | 24     | 11       | 17         | 34         | 8          | 16         | 27         | 6         |          |
| SMP (%)                 |      | 9       | 12     | 5        | 9          | 14         | 5          | 9          | 14         | 4         |          |
| FA (%)                  |      | 47      | 58     | 38       | 44         | 57         | 28         | 44         | 58         | 31        |          |
| HA (%)                  |      | 23      | 33     | 13       | 23         | 37         | 8          | 24         | 44         | 7         |          |
| P(1) Area<sup>260</sup> | Area<sup>260</sup> | 13.2    | 128    | ND       | 3.32       | 21.0       | ND         | 5.84       | 81.0       | ND        |          |
| P(2–3) ND              |      | ND      | ND     | ND       | 42.9       | 224        | ND         | 40.5       | 358        | ND        |          |
| P(4–5) ND              |      | 164     | 782    | ND       | 191        | 696        | ND         | 233        | 898        | ND        |          |
| P(6–7) ND              |      | 289     | 670    | 48.0     | 144        | 474        | 7.00       | 171        | 560        | 12.0      |          |
| Br (mg/L)              |      | 0.35    | 0.47   | 0.17     | 0.16       | 0.47       | ND         | 0.08       | 0.36       | ND        |          |
| THMFP (µg/L)           |      | 333     | 649    | 73       | 220        | 528        | 20         | 216        | 623        | 19        |          |
| CHCl<sub>3</sub> (µg/L)|      | 194     | 482    | 28       | 176        | 372        | 18         | 189        | 594        | 12        |          |
| CHCl<sub>3</sub>Br (µg/L)|    | 99      | 162    | 28       | 32         | 150        | 1          | 23         | 104        | 3         |          |
| CHBr<sub>3</sub>Cl (µg/L)|  | 37      | 84     | 8        | 10         | 44         | ND         | 4          | 33         | ND        | 10        |
| CHBr<sub>3</sub> (µg/L)|      | 3       | 9      | ND       | 1          | 9          | ND         | 0          | 3          | ND        | 1         |
| CHCl<sub>3</sub> (%)   |      | 51.0    | 77.8   | 26.7     | 80.3       | 98.6       | 24.5       | 82.0       | 97.5       | 39.7      | 77.8     |
| CHCl<sub>3</sub>Br (%) |      | 32.6    | 40.8   | 19.9     | 14.1       | 38.9       | 1.4        | 14.9       | 37.7       | 2.5       | 17.0     |
| CHBr<sub>3</sub>Cl (%) |      | 14.9    | 39.1   | 2.3      | 4.7        | 31.7       | 0.0        | 2.8        | 20.6       | 0.0       | 4.7      |
| CHBr<sub>3</sub> (%)   |      | 1.5     | 4.5    | 0.0      | 0.7        | 6.5        | 0.0        | 0.2        | 3.2        | 0.0       | 0.5      |

<sup>a</sup> Total samples number used to evaluate the models (surface: 4; runoff: 4 and sub-surface: 8).
<sup>b</sup> Area<sup>260</sup> = area of Log<sub>10</sub> (Dalton) detected by absorbance (cm<sup>–1</sup>) at UV260 nm.
<sup>c</sup> Where: n, number of samples; UV<sub>254</sub>, UV absorbance at wavelength 254 nm; DOC, dissolved organic carbon concentration; PI, percentage abundance of protein-1-like; PII, percentage abundance of protein-2-like; SMP, percentage abundance of soluble microbial protein-like; HA, percentage abundance of humic-like; FA, percentage abundance of fulvic-like; P(1), area under Peak 1; P(2–3), area under peaks 283; P(4–5), area under peaks 465; P(6–7), area under peaks 687; P(8–9), area under peaks 899; Br, bromide ion concentration in raw waters; THMFP, trihalomethane formation potential; CHCl<sub>3</sub>, chloroform formation potential; CHCl<sub>3</sub>Br, bromo-chloromethane formation potential; CHBr<sub>3</sub>Cl, chloro-dibromomethane formation potential; CHBr<sub>3</sub>, bromoform formation potential; ND, not detected.

Texture assigned according to the standard Australian Soil Texture Chart [39]. Landuse, vegetation and soil descriptions for each ZOC are given in Table 1. Water samples were treated by alum as described in Section 2.1.1.

2.1.3. The concentration and character of DOM

For determination of the concentration of DOM measured as DOC and UV absorbance, water samples were passed through 0.45 µm pre-sterilized sterile cellulose membrane filters prior to analyses. DOC concentration was determined using a TOC analyser (Model 900, SIEXS Instruments). UV absorbance at 254 nm was measured using a spectrophotometer (UV-120, MOISTECH Instruments) using a 1 cm quartz cuvette. The average, maximum and minimum values of DOC concentration and UV absorbance (UV<sub>254</sub>) are presented in Table 2 for both model fitted and model evaluation water samples (prior to and after alum treatment).

Fluorescence excitation-emission matrix (F-EEM) spectra were acquired (Model LS 55, PerkinElmer) to characterize DOM in terms of humic-like (HA), fulvic-like (FA), Protein1 (PI), Protein2 (PII) and soluble microbial protein (SMP)-like components [40]. A series of emission spectra (280–600 nm) were obtained with 0.5 nm increments over excitation wavelength (200–500 nm) with 5 nm increments. MQW spectra were subtracted from the measured sample spectra. The method of Chen et al. [40] was employed to calculate the relative abundance of each organic group. Samples for F-EEM were pre-diluted by addition of MQW to minimize the inner-filter effect followed by pre-filtering through 0.45 µm pre-sterilized sterile cellulose membrane filters.

Separation of organic compounds based on apparent molecular weight (MW) distributions was performed using HPSEC (A Waters 2,690 Separation Module and Waters 996 photodiode array detector) with UV detection at 260 nm wavelength, as described by Chow et al. [41]. Column size and type, flow rate and other relevant information for the HPSEC-UV detection method have been previously reported.
Water samples for HPSEC analysis were passed through 0.22 \mu m sterile cellulose membrane filters prior to the analysis. HPSEC chromatograms were analyzed using a peak fitting technique (Peak Fit software, Version 4, Systat Software Inc.) to resolve overlapping peaks as reported by Aslam et al. [42]. Peaks detected by absorbance at UV260 nm had MWs of 58,300 (Peak 1), 8,600 (Peak 2), 4,450 (Peak 3), 2,700 (Peak 4), 1,750 (Peak 5), 1,250 (Peak 6), 830 (Peak 7), 515 (Peak 8) and 390 (Peak 9) Da. The average, maximum and minimum values of DOM character parameters are also presented in Table 2.

2.1.4. THMFP and bromide ion concentration

Total THMFP and formation of constituent THM compounds (four chloro/bromo compounds) were determined using a headspace sampler (PerkinElmer, TurboMatrix 110) and gas chromatograph with electron-capture detection (PerkinElmer Clarus® 500 GC). THM compounds were formed after 20 mg/L chlorine addition under controlled laboratory conditions of 35 °C and pH 7.4, and held for 4 h. Br concentration was measured using Standard Method 4110 [43]. The average, maximum and minimum values of THMFP and its species and Br concentration are presented in Table 2.

2.2. Statistical analysis

Microsoft Excel 2010 was used to compare modelled and measured data (F-test, Student’s T-test and standard error values: SE: Eq. (1)). Where F-test values were more than 0.5, the T-test with equal variance was used. A T-test value >0.5 was interpreted to mean that there were no significant differences between measured and model fitted data whereas a T-test value <0.5 was interpreted as indicating a statistically significant difference.

\[
SE = \sqrt{\frac{\sum (\text{Measured} - \text{Predicted})^2}{\text{No. of samples}}}
\]

(1)

3. Model(s) development

Mathematical models were developed that describe the formation potential of THM in the surface and subsurface waters using non-linear regression analysis tools, IBM SPSS Statistics 21 software. The coefficients of models were determined based on statistical analysis (with highest coefficient of determination: $R^2$).

DOM present in runoff and sub-surface waters can be characterized based on the relative abundances of aromatic and non-aromatic compounds (Eq. (2)). SUVA value is correlated to the weighted average molecular weight and the aromatic compounds. A linear function (Eq. (3)) describes the relationship between SUVA and the relative abundance of aromatics [44].

From Eqs. (2) and (3), the relationship between UV absorbance at 254 nm and the percentage abundances of aromatic (Eq. (4)) and non-aromatic (Eq. (5)) compounds are described.

\[
\text{DOC} = (%\text{Aromatic} + %\text{Nonaromatic}) \times \frac{\text{DOC}}{100} = \text{AromDOC} + \text{NonAromDOC}
\]

(2)

where % Aromatic + % Nonaromatic = 100; AromDOC is the aromatic concentration (%Aromatic x DOC/100), NonAromDOC is the non-aromatic concentration [% (100 – % Aromatic) x DOC/100].

\[
%\text{Aromatic} = a_i \times \text{SUVA} + b_i
\]

(3)

\[
\text{AromDOC} = (a_i \times \text{SUVA} + b_i) \times \frac{\text{DOC}}{100} = a_i \times \text{UV}_{254} + b_i \times \frac{\text{DOC}}{100}
\]

(4)

\[
\text{NonAromDOC} = \text{DOC} - (a_i \times \text{UV}_{254}) - (b_i \times \frac{\text{DOC}}{100}) = (100 - b_i) \times \frac{\text{DOC}}{100} - a_i \times \text{UV}_{254}
\]

(5)

Under the standardized laboratory conditions applied and with the bromide concentration less than 0.025 mg/L, an assumption was made that DOM concentration and character are the main parameters that affect the formation of THM (chlorine was added in excess). The THM formation potential could then be described by Eq. (6):

\[
\text{THMFP} \propto \text{DOC}
\]

\[
\text{THMFP} \propto (\text{AromDOC} + \text{NonAromDOC})
\]

\[
\text{THMFP} = c_m \times (\text{AromDOC}) + d_w \times (\text{NonAromDOC}) + e_y
\]

(6)

\[
\text{THMFP} = a_i \times \text{UV}_{254} + (c_6 \times \text{SUVA} + d_b) \times \frac{\text{DOC}}{100} + e_y
\]

where THMFP is in \mu g/L, UV254 is in cm$^{-1}$ and DOC is in mg/L; $a_i, b_i, c_m, d_w$ are the model coefficients and $e_y$ is the model constant; $i, j, m, w$ and $y$ refer to equation number.

From acquired data ($n=15$) with Br not detected (<0.025 mg/L), Eqs. (7), (9) and (10) were developed to describe relationships between THMFP and DOM concentration (measured as DOC) and character.

\[
\text{THMFP} = g_7 \times \text{UV}_{254} + h_7 \times \frac{\text{DOC}}{100} + e_y, \text{[where } g_7 = a_i \times (c_6 - d_b) \text{ and } h_7 = (c_6 \times b_b) + (100 \times d_b) - (d_b \times b_b)]
\]

(7)

In previous studies, different wavelengths (e.g., 254 nm [44], 272 nm [45], 280 nm [46]) have been used to determine the SUVA value. Hur et al. [47] reported that SUVA determined near to 250 nm is correlated with aromatic organic compounds. Based on this, using the resolved peaks of organics (by HPSEC-UV detection) present in waters before and after alum treatment, THMFP was modeled (Eq. (9)) using above equations (Eqs. (2)–(7)) with the assumption made that UV absorbance at 254 nm is about equal to that at 260 nm (Eq. (8)). The mean difference between UV absorbance at 254 nm and UV absorbance at 260 nm was calculated to be $0.01 \pm 0.01 \text{ cm}^{-1}$ (X. S.D.) SUVA.
was determined at UV absorbance of 254 nm while HPSEC at 260 nm. For Eq. (9), each peak area \([P(1):P(9)]\) determined by HPSEC was multiplied by a coefficient \((a_0, b_1, c_m, d_9, f_s)\) to account for the effect of peak number on THM formation. The area under each peak \([\text{area of Log}_{10} (\text{Dalton})]\) detected by absorbance (cm\(^{-1}\)) at UV260 nm was calculated using Peak Fit software, Systat Software Inc., as detailed in Section 2.1.3.

\[
UV_{254} \cong UV_{260} = \sum_{p=1}^{9} \text{Peak area} \times 260 \text{nm}
\]

(8)

\[
\text{THMFP} = g_9 \times [a_0 \times P(1) + b_9 \times P(2–3) + c_9 \times P(4–5) + d_9 \times P(6–7) + f_9 \times P(8–9)] + h_9 \times \frac{\text{DOC}}{100} + e_9
\]

(9)

where: \([P(1)]\), area under Peak 1; \([P(2–3)]\), total area under peaks 2&3; \([P(4–5)]\), total area under peaks 4&5; \([P(6–7)]\), total area under peaks 6&7; \([P(8–9)]\), total area under peaks 8&9.

According to Chu et al. [48], proteinaceous organic compounds (SMP and aromatic protein) determined by F-EEM, have a significant effect on the prediction of DBP formation potential (i.e., dichloroacetic). In our study, F-EEM spectra of DOM were acquired to identify five different EEM regions (HA, FA, PI, PII and SMP) for each sample and to calculate the relative abundance of each organic group as detailed by Chen et al. [40]. DOM present in runoff and sub-surface waters can be characterized based on the relative abundances (as percentages) of these five organic groups i.e., PI, PII, SMP, FA and HA. Under the standard conditions (as detailed in Section 2.1.4) and with bromide not detected, Eq. (10) describes the relationship between the relative abundance of different organic constituents as defined by F-EEM data and THM formation.

\[
\text{THMFP} \propto \text{DOC} \\
\text{THMFP} \propto (\text{PI}+\text{PII}+\text{SMP}+\text{FA}+\text{HA}) \times \frac{\text{DOC}}{100} + e_{10}
\]

(10)

where PI, PII, SMP, FA and HA is in percentage (%).

The presence of Br leads to higher THMFP [32] but the effect of Br on the reaction between chlorine and precursor DOM compounds to produce THM may depend on the nature of the precursor DOM. In order to account for bromide effects and based on the available data \((n = 65; \text{for Br concentration } <0.5 \text{ mg/L})\), model equations (Eqs. 9, 10 and 11) were modified with Br concentration added as an independent variable. Mathematical relationships were then established between the DOM character, Br ion concentration and THMFP (Eqs. (11)–(13)). Bromide effects were expressed as \(1 + k \cdot \text{Br}\) to avoid a zero value for THMFP when Br was not detected.

\[
\text{THMFP} = g_{17} \times (1 + k_{11} \times \text{Br}) \times UV_{254} + h_7 \times (1 + k_{11} \times \text{Br}) \times \frac{\text{DOC}}{100} + e_7 \times (1 + k_{31} \times \text{Br})
\]

(11)

\[
\text{THMFP} = g_9 \times (1 + k_{12} \times \text{Br}) \times [a_0 \times P(1) + b_9 \times P(2–3) + c_9 \times P(4–5) + d_9 \times P(6–7) + f_9 \times P(8–9)]
\]

\[
+ h_9 \times (1 + k_{12} \times \text{Br}) \times \frac{\text{DOC}}{100} + e_9 \times (1 + k_{32} \times \text{Br})
\]

(12)

\[
\text{THMFP} = (a_{10} \times \text{PI} + b_{10} \times \text{PII} + c_{10} \times \text{SMP} + d_{10} \times \text{FA} + f_{10} \times \text{HA}) \times \frac{\text{DOC}}{100} \times (1 + k_{13} \times \text{Br}) + e_{10}
\]

(13)

For development of models that describe the relative abundances and concentrations of constituent THM compounds, the following steps were carried out. For prediction of CHCl\(_3\) concentration (\(\mu g/L\)) an assumption was made that CHCl\(_3\) concentration is equal to the total THM formation potential for sample data \((n = 15)\) with Br <0.025 mg/L (Eqs. 9, 10 and 11) were used and the coefficients were estimated. The relative abundance of CHCl\(_3\) to THM has an inverse correlation with bromide ion concentration in the raw waters (the existence of Br leads to more bromide species), [32]. From this, mathematical relationships were established between the DOM character and Br ion concentration (<0.5 mg/L) with CHCl\(_3\) concentration (Eqs. (14)–(16)).

\[
\text{CHCl}_3 = \frac{g_{14} \times UV_{254}}{1 + k_{14} \times \text{Br}} \times \frac{h_{14} \times \text{DOC}}{100} \times (1 + k_{14} \times \text{Br}) + \frac{e_{14}}{(1 + k_{34} \times \text{Br})}
\]

(14)

\[
\text{CHCl}_3 = \frac{g_{15}}{(1 + k_{15} \times \text{Br})} \times [a_{15} \times P(1) + b_{15} \times P(2–3) + c_{15} \times P(4–5) + d_{15} \times P(6–7) + f_{15} \times P(8–9)]
\]

\[
+ \frac{h_{15} \times \text{DOC}}{100} \times (1 + k_{15} \times \text{Br}) + \frac{e_{15}}{(1 + k_{35} \times \text{Br})}
\]

(15)

\[
\text{CHCl}_3 = \left[ \frac{a_{16} \times \text{PI}}{1 + k_{16} \times \text{Br}} + \frac{b_{16} \times \text{PII}}{1 + k_{16} \times \text{Br}} + \frac{c_{16} \times \text{SMP}}{1 + k_{16} \times \text{Br}} + \frac{d_{16} \times \text{FA}}{1 + k_{16} \times \text{Br}} + \frac{f_{16} \times \text{HA}}{1 + k_{16} \times \text{Br}} \right] \times \text{DOC} + e_{16}
\]

(16)

According to van Leeuwen et al. [49] and Roccaro et al. [27] the relative abundances of constituent THM compounds to total THM is a function of the ratio of bromide concentration to the chloride consumed. van Leeuwen et al. [49] used an exponential function to describe the relationship between the percentage formation of chloroform and Br/chlorine consumed ratio. However for the THMFP test, a very high chlorine dose (20 mg/L excess free chlorine) was consistently used and the chlorine consumed was assumed to be not an independent variable in these THM formation models.
Using Eq. (6), the relative abundance of CHCl₃ (the percentage ratio of CHCl₃ to THMFP concentration) was modeled (Eq. (17)) for data when Br was not detected. Where Br was detected, the concentration effect was taken into account as described by Eq. (18). Similarly, data obtained from HPSEC-UV detection and F-EEM were used to predict the %CHCl₃ (Eqs. (19) and (20) respectively).

\[
\text{%CHCl}_3 = \frac{\text{CHCl}_3 \times \text{(AromDOC + NonAromDOC)}}{\text{THMFP \times (AromDOC + NonAromDOC)}} \times 100
\]

(17)

The second subscript z refers to coefficients for chloroform formation potential or total THMFP (see Table 3).

\[
\text{%CHCl}_3 = \left[ \frac{a_{17} \times (c_{17,1} - d_{17,1}) \times \text{UV}_254 + [(c_{17,1} \times b_{17}) \times (100 \times d_{17,1})] \times \text{DOC} \times 100}{a_{17} \times (c_{17,2} - d_{17,2}) \times \text{UV}_254 + [(c_{17,2} \times b_{17}) \times (100 \times d_{17,2})] \times \text{DOC} \times 100} + e_{17} \right] \times 100
\]

(18)

\[
\text{%CHCl}_3 = \left[ \frac{a_{17} \times (c_{17,1} - d_{17,1}) \times \text{UV}_254 + [(c_{17,1} \times b_{17}) \times (100 \times d_{17,1})] \times \text{DOC} \times 100}{a_{17} \times (c_{17,2} - d_{17,2}) \times \text{UV}_254 + [(c_{17,2} \times b_{17}) \times (100 \times d_{17,2})] \times \text{DOC} \times 100} \right] \times 100
\]

(19)

\[
\text{%CHCl}_3 = \left[ \frac{a_{10} \times (c_{19,1} - d_{19,1}) \times \text{PI} \times \text{b}_{19,1} \times \text{P} \times (2 - 3) \times [c_{19,3} \times P(4 - 5)] \times [d_{19,3} \times P(6 - 7)] \times [f_{19,3} \times P(8 - 9)] \times [(c_{19,1} \times b_{19}) \times (100 \times d_{19,1})] \times (1 + k_{19,1} \times Br) + e_{19} \times (1 + k_{29,1} \times Br)}{a_{10} \times (c_{19,2} - d_{19,2}) \times \text{PI} \times \text{b}_{19,2} \times \text{P} \times (2 - 3) \times [c_{19,4} \times P(4 - 5)] \times [d_{19,4} \times P(6 - 7)] \times [f_{19,4} \times P(8 - 9)] \times [(c_{19,2} \times b_{19}) \times (100 \times d_{19,2})] \times (1 + k_{19,2} \times Br) + e_{19} \times (1 + k_{29,2} \times Br)} \right] \times 100
\]

(20)

In the model developed using the relative abundance of chloroform to THMFP, the percentages of bromodichloromethane (CHCl₂Br) and chlorodibromomethane (CHBr₂Cl) are estimated. A power function (Eq. (21)) was used for %CHCl₂Br, and an exponential function (Eq. (22)) for %CHBr₂Cl. For acquired data, CHCl₂Br was not detected when the %CHCl₃ was more than 92% of the total THMFP. Based on this, Eq. (22) is used only when %CHCl₃ ≤ 92%. When %CHCl₃ > 92%, %CHCl₂Br was not detected and assumed to be 0. From previous models established, the concentrations of each individual THM compound can be estimated (the bromoform formation potential is calculated from the percentages of the other three compounds and the predicted THMFP).

\[
\text{%CHCl}_2\text{Br} = b_{21} \times \text{%CHCl}_3^{2.1}
\]

(21)

\[
\text{%CHBr}_2\text{Cl} = a_{22} + b_{22} \times \exp ^{-c_{22} \times \text{%CHCl}_3}
\]

(22)
4. Results and discussion

4.1. Coefficient parameters and the statistical analysis results

Coefficients of model Eqs. (6) and (11)–(22) are given in Table 3, and Table S1 (Supplementary information) summarizes the statistical analyses of these models. R^2 values were greater than 0.85 and T-test values were greater than 0.82 (T-test value >0.5, indicating no significant difference between measured and model fitted data). The developed models were evaluated using data obtained from Mt. Bold and Happy Valley Reservoir ZOCs (16 point data set) as detailed in Section 2.1.2, and results of statistical analyses (T-test and SE) of this data are presented in Table S1. T-test values were greater than 0.54 and SE values were lower than 72.3 µg/L respectively.

4.2. Trihalomethane formation potential (THMFP)

The character of DOM based on SUVA, MW distribution of constituents by HPSEC-UV and relative abundances of protein-like and humic/fulvic-compounds identified by F-EEM, with DOC and Br concentrations were related to THMFPs of DOM isolates (Eqs. (11) (12) and (13) respectively). These equations provided fitted THMFP data that closely matched measured data (Table S1 and Fig. 1). The equation with the best fit to measured data (Eq. (11), R^2 = 0.97 and SE = 32.9 µg/L) was obtained when UV absorbance at 254 nm was incorporated. For Eq. (11), the coefficient of UV254 (g_1 = 799) was greater than of DOC/100 (h_2 = 503), indicating that THM was more likely to be produced from UV absorbing organic, hydrophobic compounds, than DOC overall, that includes hydrophilic and non-aromatic compounds, a finding similar to that reported by Kitis et al. [50] and Chen and Westerhoff [32].

For both Eqs. (11) and (12), k_1 (the coefficient of Br), the Br effect on THM produced from UV absorbing organic compounds was negative (−0.34 for Eq. (11) and −1.24 for Eq. (12)) whereas, k_2 (the coefficient of Br), the Br effect on the THM produced from overall organic constituents (DOC) was positive (2.1 for Eq. (11) and 5.3 for Eq. (12)). This indicates that with higher concentrations of Br there is a decrease in THM formation from UV absorbing, hydrophobic compounds and an increase in THM formation from DOC overall. Liang and Singer [23] reported that bromine reacts more with hydrophilic compounds than with hydrophobic compounds. Chowdhury et al. [51] reported that waters without Br mainly form the chlorinated only THM (i.e., CHCl_3) due to reactions between hypochlorous acid and hydrophobic fractions of DOM, while for waters with Br, the hydrophilic fractions of DOM form brominated THMs through reactions with hypobromous acid. As such, with enhanced coagulation and high removal of hydrophobic compounds, there would be greater potential for brominated compounds to be formed.

For Eq. (13), the coefficients of HA-like and FA-like compounds are positive whereas the coefficients of protein compounds varied i.e., PI- and SMP-like compounds are negative and PII-like compounds are positive. In order to examine the effect of all protein-like compounds by EEM (PI-, PII- and SMP-like) on the THM formation, the weighted average protein coefficient [\text{AVE}_P = (a_p \times PI + b_p \times PII + c_p \times SMP)/sum (PI + PII + SMP)], where a_p, b_p and c_p are model coefficient parameters and PI, PII and SMP are the relative abundance of protein compounds] was calculated. This was found to be −25.7 ± 12.3 (X, S.D.), implying that waters with a high percentage of protein-like compounds have lower formation of THM compared to waters with a high percentage of humic compounds. As such, with enhanced coagulation and high removal of humic compounds, there would be a lower potential for THM formation. Summers et al. [33] also reported that THMFP in watershed waters (Cache La Poudre and Colorado–Big Thompson, Colorado, USA) had a positive correlation with humic-like compounds and/or UV values at 254 nm.

As detailed in Section 2.1.2, a set of 16 data points was used to evaluate the models developed. The measured and predicted values from these models using Eqs. (11)–(13) are presented in Supplementary information (Table S2). In this data set, THMFP concentrations varied from 20 µg/L to 511 µg/L and DOC concentrations varied from 1.8 mg/L to 16.9 mg/L. The character of DOM varied highly (Table 2). Even though the variation in data of this set was high, the model predicted THMFP values generally matched the measured values (Table S1, T-test > 0.72 and SE < 68.7 µg/L).

4.3. Constituent THM compounds

Coefficient of determination values for the CHCl_3 models (Eqs. (14)–(16)) and %CHCl_3 (Eqs. (18)–(20)) values are slightly lower than for the THMFP models (Eqs. (11)–(13)), shown in Table S1. DOM character, concentration (as DOC) and Br concentrations enabled good statistical fitting between model and measured CHCl_3 concentrations (R^2 > 0.92 and SE < 46.7 µg/L) and its relative abundances (R^2 > 0.85 and SE < 8.9%). The measured and model fitted values for Eqs. (14)–(16) and for Eqs. (18)–(20) are presented in Figs. 2 and 3, respectively.
The measured and predicted values for the evaluation data set are presented in Table S2. These models provided for prediction of CHCl$_3$ formation potential with the following statistics, for Eqs. (14)–(16): $T$-test > 0.65 and SE < 72.3 mg/L and Eqs. (18)–(20): $T$-test > 0.54 and SE < 9.6%. Although the absolute standard errors are high, they are of waters with DOC up to 20 mg/L (mean: 7.8 mg/L) following chlorination at 20 mg/L and CHCl$_3$ formation to 594 mg/L.

Similar to the total THMFP model, the coefficient of $UV_{254}$ ($\beta_r = 782$) for the CHCl$_3$ model (Eq. (14)), was greater than of DOC/100 ($h_n = 349$). For Eq. (16), coefficients of HA- and FA-like compounds were positive, while the weighted average protein coefficient ($Ave_{PC}$) was negative ($-71.5 \pm 18.4$). In comparison to the THMFP model (Eq. (13)), the $Ave_{PC}$ value for the CHCl$_3$ model (Eq. (16)) was found to be higher than that for the THMFP model ($-25.7 \pm 12.3$), implying that with a higher percentage of protein-like compounds, there is higher formation of brominated THM species. These results are consistent with the findings of Rodrigues et al. [35] who reported that with higher fulvic acid levels, there is lower formation of brominated THM species. In contrast, for groundwater disinfected by ozone, Huang et al. [52] reported that hydrophobic, humic–acid compounds produced a higher bromoform level compared with hydrophilic compounds.

For each sample, the relative abundance of CHCl$_2$Br was similar to or higher than that of CHBr$_2$Cl. The relative abundance of CHCl$_2$Br ranged from 1.4% to 40.8% and the %CHBr$_2$Cl ranged up to 39.1%. In this study, models used to predict the %CHCl$_2$Br and %CHBr$_2$Cl from %CHCl$_3$ (Eq. (21): power function and Eq. (22): exponential function respectively) fitted well the measured data, as shown in Table S1 and Fig. 4. The measured and predicted values for the evaluation data set are presented in Supplementary information (Table S2).

Few studies have reported the development of models designed to predict THM species based on the character of natural organic matter present in water. Rodrigues et al. [35] reported models to predict THM species based on physical and chemical factors (chlorine dose, temperature, pH and Br concentration) and DOM concentration. However, they used synthetic water spiked with fulvic acid to develop these models. Based on DOC, $UV_{254}$ and Br levels in drinking water and wastewater samples, Chen and Westerhoff [32] developed a power function model to predict only the CHCl$_3$ concentration. From our study, different equations are proposed that include parameters based on the character of DOM to predict THMFP species under standardized chlorination conditions (Eqs. (14)–(16), CHCl$_3$ concentration model; Eqs. (18)–(20), %CHCl$_3$ model; Eq. (21), %CHCl$_3$Br model and Eq. (22), %CHBr$_2$Cl model). Using these equations, the concentrations of each individual chloro- and/or bromo-THM compound can be estimated. From this study, it was found that the distribution of THM compounds is a function of DOM character and the Br concentration.

From studies conducted to determine the impacts of catchment characteristics on the quality of waters used for domestic supply, we previously reported the influence of vegetation (native, pine, grasslands) and soil texture (sandy/clayey) on the characters and concentrations of DOMs present in surface and sub-surface runoff waters [4]. The character of residual DOM in these waters after enhanced coagulation by alum [53] and subsequently on the THM formation potential [54] have also been reported. The findings from these studies indicated that both soil texture and vegetation type and loading are key factors that determine DOM levels and their character in source waters used for drinking water supply. From models reported here that relate the character of organics present in surface and sub-surface waters to catchment characteristics, THMFP in waters can be predicted. It is proposed that these models may be used by catchment managers as part of decision support systems for selection of land-uses that lead to high quality water resources for drinking water supply.
5. Conclusions

The character of DOM and presence of Br in surface and subsurface waters derived from catchments with varied land-uses and surface soil textures can be related to THMFP and relative abundances of constituent compounds. Models were developed of THMFP and the constituent compounds, chloroform, bromodichloromethane and chlorodibromomethane based on the character of precursor organics. The bromoform formation potential is then determined from the percent of the other three compounds and predicted THMFP. Various DOM characterization techniques (i.e., UV absorbance at 254 nm, HPSEC-UV with peak fitting analysis and F-EEM), provided data that related to THM formation potential and the THM species. The best fitted THMFP model was obtained when UV absorbance at 254 nm was used to characterize the dissolved organic constituents of the surface and subsurface waters.

THM was more likely to be produced from UV absorbing organic, hydrophobic compounds, than DOC overall, that includes hydrophilic, non-aromatic compounds. Further, waters with a high percentage of protein-like compounds have lower formation potential of THM but have higher relative abundances of brominated THM species compared to waters with a high percentage of humic compounds. As such, with enhanced coagulation and high removal of humic compounds, there would be a lower potential for THM formation but the relative abundances of brominated THM species would be higher.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Australian Research Council under Grant LP110200208, SA Water and the SA Department for Environment, Water and Natural Resources and the Goyder Institute for Water Research. The authors thank Mr Jonathan Cohen for his valuable support in field work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.jhazmat.2016.01.030.

References

[1] B.A. Bryan, J. Kandulu, D.A. Deere, M. White, J. Frizenschaf, N.D. Crossman, Adaptive management for mitigating cryptosporidium risk in source water: a case study in an agricultural catchment in South Australia, J. Environ. Manag. 90 (2009) 3122–3134.
[2] Y. Yang, Z. He, Y. Wang, J. Fan, Z. Liang, P.J. Stoffella, Dissolved organic matter in relation to nutrients (N and P) and heavy metals in surface runoff water as affected by temporal variation and land uses—a case study from Indian River Area, south Florida, USA, Agric. Water Manag. 118 (2013) 38–49.
[3] M.C. Peer, Hydrology: catchment vegetation and runoff, Prog. Phys. Geogr. 33 (2009) 837–844.
[4] J. Awad, J. van Leeuwen, D. Abate, M. Pichler, E. Bestland, D.J. Chittleborough, N. Fleming, J. Cohen, J. Liffner, M. Drikas, The effect of vegetation and soil texture on the nature of organics in runoff from a catchment supplying water for domestic consumption, Sci. Total Environ. 529 (2015) 72–81.
[5] A. Ágren, I. Buffam, M. Jansson, H. Laudon, Importance of seasonality and small streams for the landscape regulation of dissolved organic carbon export, J. Geophys. Res.: Biogeosci. 112 (2007) 2005–2012.
[6] T. Matisson, P. Kortelainen, A. Laubel, D. Evans, M. Pujo-Pay, A. Räike, P. Conan, Export of dissolved organic matter in relation to land use along a European climatic gradient, Sci. Total Environ. 407 (2009) 1967–1976.
[7] R. Naidu, D. Williamson, R. Fitzpatrick, I. Hollingsworth, Effect of landuse on the composition of throughflow water immediately above clayey B horizons in the Warren Catchment, South Australia, Anim. Prod. Sci. 33 (1993) 239–244.
[8] J.A. Aitkenhead-Peterson, J.E. Alexander, T.A. Clair, Dissolved organic carbon and dissolved organic nitrogen export from forested watersheds in Nova Scotia: identifying controlling factors, Glob. Biogeochem. Cycles 19 (2005) GB4016.
[9] K. Nosrat, G. Govers, E. Smolders, Dissolved organic carbon concentrations and fluxes correlate with land use and catchment characteristics in a semi-arid drainage basin of Iran, CATENA 95 (2012) 177–183.
[10] S. Inamdar, N. Finger, S. Singh, M. Mitchell, D. Leiva, H. Bais, D. Scott, P. McHale, Dissolved organic matter (DOM) concentration and quality in a forested mid-Atlantic watershed, USA, Biogeochemistry 108 (2012) 55–76.
[11] M.R. Banaitis, H. Waldrip-Dail, M.S. Diehl, B.C. Holmes, J.F. Hunt, R.P. Lynch, T. Ohno, Investigating sorption-driven dissolved organic matter fractionation by multidimensional fluorescence spectroscopy and PARAFAC, J. Colloid Interface Sci. 364 (2006) 271–276.
[12] D.A.N. Usuru, C.E. Johnson, Sorption of organic carbon fractions by spodosol mineral horizons, Soil Sci. Soc. Am. J. 68 (2004) 253–262.
[13] J. Aitkenhead-Peterson, W. McDowell, J. Neff, E. Stuart, L. Robert, Sources, Production, and Regulation of Allochthonous Dissolved Organic Matter Inputs to Surface Waters, Academic Press, San Diego, 2003.
