ISSN 1001-0742
CN 11-2629/X

JES

JOURNAL OF ENVIRONMENTAL SCIENCES

January 1, 2015 Volume 27
www.jesc.ac.cn

Could wastewater analysis be a useful tool for China?

ELSEVIER
Science Press

Sponsored by
Research Center for Eco-Environmental Sciences
Chinese Academy of Sciences

Journal of Environmental Sciences Volume 27 2015

www.jesc.ac.cn

1 The potential risk assessment for different arsenic species in the aquatic environment
Meng Du, Dongbin Wei, Zhuowei Tan, Aiwu Lin, and Yuguo Du

9 Synthesis of linear low-density polyethylene-*g*-poly (acrylic acid)-co-starch/organo-montmorillonite hydrogel composite as an adsorbent for removal of Pb(II) from aqueous solutions
Maryam Irani, Hanafi Ismail, Zulkifli Ahmad, and Maohong Fan

21 Research and application of kapok fiber as an absorbing material: A mini review
Yian Zheng, Jintao Wang, Yongfeng Zhu, and Aiqin Wang

33 Relationship between types of urban forest and $PM_{2.5}$ capture at three growth stages of leaves
Thithanhthao Nguyen, Xinxiao Yu, Zhenming Zhang, Mengmeng Liu, and Xuhui Liu

42 Bioaugmentation of DDT-contaminated soil by dissemination of the catabolic plasmid pDOD
Chunming Gao, Xiangxiang Jin, Jingbei Ren, Hua Fang, and Yunlong Yu

51 Comparison of different combined treatment processes to address the source water with high concentration of natural organic matter during snowmelt period
Pengfei Lin, Xiaojian Zhang, Jun Wang, Yani Zeng, Shuming Liu, and Chao Chen

59 Chemical and optical properties of aerosols and their interrelationship in winter in the megacity Shanghai of China
Tingting Han, Liping Qiao, Min Zhou, Yu Qu, Jianfei Du, Xingang Liu, Shengrong Lou, Changhong Chen, Hongli Wang, Fang Zhang, Qing Yu, and Qiong Wu

70 Could wastewater analysis be a useful tool for China? – A review
Jianfa Gao, Jake O'Brien, Foon Yin Lai, Alexander L.N. van Nuijs, Jun He, Jochen F. Mueller, Jingsha Xu, and Phong K. Thai

80 Controlling cyanobacterial blooms by managing nutrient ratio and limitation in a large hypereutrophic lake: Lake Taihu, China
Jianrong Ma, Boqiang Qin, Pan Wu, Jian Zhou, Cheng Niu, Jianming Deng, and Hailin Niu

87 Reduction of NO by CO using Pd-CeTb and Pd-CeZr catalysts supported on SiO_2 and La_2O_3-Al_2O_3
Victor Ferrer, Dora Finol, Roger Solano, Alexander Moronta, and Miguel Ramos

97 Development and case study of a science-based software platform to support policy making on air quality
Yun Zhu, Yanwen Lao, Carey Jang, Chen-Jen Lin, Jia Xing, Shuxiao Wang, Joshua S. Fu, Shuang Deng, Junping Xie, and Shicheng Long

108 Modulation of the DNA repair system and ATR-p53 mediated apoptosis is relevant for tributyltin-induced genotoxic effects in human hepatoma G2 cells
Bowen Li, Lingbin Sun, Jiali Cai, Chonggang Wang, Mengmeng Wang, Huiling Qiu, and Zhenghong Zuo

115 Impact of dissolved organic matter on the photolysis of the ionizable antibiotic norfloxacin
Chen Liang, Huimin Zhao, Minjie Deng, Xie Quan, Shuo Chen, and Hua Wang

124 Enhanced bio-decolorization of 1-amino-4-bromoanthraquinone-2-sulfonic acid by *Sphingomonas xenophaga* with nutrient amendment
Hong Lu, Xiaofan Guan, Jing Wang, Jiti Zhou, Haikun Zhang

131 Winter survival of microbial contaminants in soil: An *in situ* verification
Antonio Bucci, Vincenzo Allocca, Gino Naclerio, Giovanni Capobianco, Fabio Divino, Francesco Fiorillo, and Fulvio Celico

139 Assessment of potential dermal and inhalation exposure of workers to the insecticide imidacloprid using whole-body dosimetry in China
Lidong Cao, Bo Chen, Li Zheng, Dongwei Wang, Feng Liu, and Qiliang Huang

CONTENTS

147 Biochemical and microbial soil functioning after application of the insecticide imidacloprid
Mariusz Cycoń and Zofia Piotrowska-Seget

159 Comparison of three-dimensional fluorescence analysis methods for predicting formation of trihalomethanes and haloacetic acids
Nicolás M. Peleato and Robert C. Andrews

168 The migration and transformation of dissolved organic matter during the freezing processes of water
Shuang Xue, Yang Wen, Xiujuan Hui, Lina Zhang, Zhaohong Zhang, Jie Wang, and Ying Zhang

179 Genomic analyses of metal resistance genes in three plant growth promoting bacteria of legume plants in Northwest mine tailings, China
Pin Xie, Xiuli Hao, Martin Herzberg, Yantao Luo, Dietrich H. Nies, and Gehong Wei

188 Effect of environmental factors on the complexation of iron and humic acid
Kai Fang, Dongxing Yuan, Lei Zhang, Lifeng Feng, Yaojin Chen, and Yuzhou Wang

197 Resolving the influence of nitrogen abundances on sediment organic matter in macrophyte-dominated lakes, using fluorescence spectroscopy
Xin Yao, Shengrui Wang, Lixin Jiao, Caihong Yan, and Xiangcan Jin

207 Predicting heavy metals' adsorption edges and adsorption isotherms on MnO_2 with the parameters determined from Langmuir kinetics
Qinghai Hu, Zhongjin Xiao, Xinmei Xiong, Gongming Zhou, and Xiaohong Guan

217 Applying a new method for direct collection, volume quantification and determination of N_2 emission from water
Xinhong Liu, Yan Gao, Honglian Wang, Junyao Guo, and Shaohua Yan

225 Effects of water management on arsenic and cadmium speciation and accumulation in an upland rice cultivar
Pengjie Hu, Younan Ouyang, Longhua Wu, Libo Shen, Yongming Luo, and Peter Christie

232 Acid-assisted hydrothermal synthesis of nanocrystalline TiO_2 from titanate nanotubes: Influence of acids on the photodegradation of gaseous toluene
Kunyang Chen, Lizhong Zhu, and Kun Yang

241 Air-soil exchange of organochlorine pesticides in a sealed chamber
Bing Yang, Baolu Han, Nandong Xue, Lingli Zhou, and Fasheng Li

251 Effects of elevated CO_2 on dynamics of microcystin-producing and non-microcystin-producing strains during *Microcystis* blooms
Li Yu, Fanxiang Kong, Xiaoli Shi, Zhen Yang, Min Zhang, and Yang Yu

259 Sulfide elimination by intermittent nitrate dosing in sewer sediments
Yanchen Liu, Chen Wu, Xiaohong Zhou, David Z. Zhu, and Hanchang Shi

266 Steel slag carbonation in a flow-through reactor system: The role of fluid-flux
Eleanor J. Berryman, Anthony E. Williams-Jones, and Artashes A. Migdisov

276 Amine reclaiming technologies in post-combustion carbon dioxide capture
Tielin Wang, Jon Hovland, and Klaus J. Jens

290 Do vehicular emissions dominate the source of C6-C8 aromatics in the megacity Shanghai of eastern China?
Hongli Wang, Qian Wang, Jianmin Chen, Changhong Chen, Cheng Huang, Liping Qiao, Shengrong Lou, and Jun Lu

298 Insights into metals in individual fine particles from municipal solid waste using synchrotron radiation-based micro-analytical techniques
Yumin Zhu, Hua Zhang, Liming Shao, and Pinjing He

Available online at www.sciencedirect.com

ScienceDirect

www.journals.elsevier.com/journal-of-environmental-sciences

JOURNAL OF ENVIRONMENTAL SCIENCES

www.jesc.ac.cn

Comparison of three-dimensional fluorescence analysis methods for predicting formation of trihalomethanes and haloacetic acids

Nicolás M. Peleato, Robert C. Andrews*

Department of Civil Engineering, University of Toronto, Toronto M5S 1A4, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 February 2014
Revised 17 April 2014
Accepted 25 April 2014
Available online 13 November 2014

Keywords:
Natural organic matter
Fluorescence
Principal component analysis
PARAFAC
Disinfection byproducts
Drinking water

ABSTRACT

This work investigated the application of several fluorescence excitation–emission matrix analysis methods as natural organic matter (NOM) indicators for use in predicting the formation of trihalomethanes (THMs) and haloacetic acids (HAAs). Waters from four different sources (two rivers and two lakes) were subjected to jar testing followed by 24 hr disinfection by-product formation tests using chlorine. NOM was quantified using three common measures: dissolved organic carbon, ultraviolet absorbance at 254 nm, and specific ultraviolet absorbance as well as by principal component analysis, peak picking, and parallel factor analysis of fluorescence spectra. Based on multi-linear modeling of THMs and HAAs, principle component (PC) scores resulted in the lowest mean squared prediction error of cross-folded test sets (THMs: 43.7 $(\mu g/L)^2$, HAAs: 233.3 $(\mu g/L)^2$). Inclusion of principle components representative of protein-like material significantly decreased prediction error for both THMs and HAAs. Parallel factor analysis did not identify a protein-like component and resulted in prediction errors similar to traditional NOM surrogates as well as fluorescence peak picking. These results support the value of fluorescence excitation–emission matrix–principal component analysis as a suitable NOM indicator in predicting the formation of THMs and HAAs for the water sources studied.

Introduction

Chlorine remains as a common disinfectant used by water utilities which, when added to natural waters, forms potentially hazardous organic halides through reactions with natural organic material (NOM) (Richardson and Postigo, 2012). Disinfection byproduct (DBP) control and regulation for utilities using chlorine typically revolve around two organic halide groups, trihalomethanes (THMs) and haloacetic acids (HAAs), which are reported to occur at the highest concentrations (Hua and Reckhow, 2007).

To facilitate DBP control, efforts have been directed towards developing predictive models. A common source of error in such models is estimation of NOM concentration. NOM is a complex mixture of humic and fulvic acids, proteins, carbohydrates, as well as other groups of organic compound classes (Her et al., 2003), all of which have unique reactivity with oxidants to form DBPs (Barrett et al., 2000). Historically, predictive DBP models most commonly utilize NOM estimation parameters including total organic carbon (TOC), dissolved organic carbon (DOC), UV-absorbance (UVA) at 254 nm, and specific UV-absorbance (SUVA) (Sadiq and Rodriguez, 2004; Chowdhurry et al., 2009). However, these parameters provide little or no information on individual NOM fractions. It is postulated that methods which can separately quantify reactive fractions will improve the accuracy of DBP formation models.

* *Corresponding author.*
E–mail address: nicolas.peleato@mail.utoronto.ca

http://dx.doi.org/10.1016/j.jes.2014.04.014
1001-0742/

Analysis using fluorescence excitation–emission matrices (FEEM) has been gaining traction as a promising method for determining information regarding organic matter composition and function in water (Zepp et al., 2004; Bieroza et al., 2010). Several components of NOM fractions exhibit unique fluorescence signatures that allow them to be distinguished in a fluorescence spectrum or FEEM. The excitation–emission position of the peak is representative of compound and/or functional groups; peak intensity is correlated with concentration (Bieroza et al., 2010). A variety of FEEM analysis techniques are described in the literature, with the majority tracking intensity changes at a few selected wavelengths (Murphy et al., 2011), whereas some incorporate changes to peak shape and position (Roccaro et al., 2009). Molecular structure has a noted effect on peak position which may have a degree of commonality among distinct molecules. By peak picking, the nature of overlapping structural properties between potentially relevant fluorophores is neglected (Persson and Wedborg, 2001).

To account for this overlapping nature of organic classes the application of multivariate analysis techniques have been successful (Stedmon et al., 2003; Peiris et al., 2010). Common advanced analysis techniques involve using two-way principal component analysis (PCA) and multi-way parallel factor (PARAFAC) analysis (Bahram et al., 2006; Stedmon et al., 2003). These techniques provide significant dimensionality reduction, while incorporating the entire fluorescence spectrum (Bieroza et al., 2010). Unlike PARAFAC, PCA models have rotational freedom (Stedmon et al., 2003) such that loading values do not necessarily represent real profiles, however will capture a higher degree of variance within the dataset (Bro, 1997). Further to the lack of rotational freedom when using PARAFAC, commonly employed constraints of unimodality and non-negativity allow for much greater interpretability of results in the context of individual organic matter components.

Several studies have reported the application of fluorescence-based measurements to determine NOM reactivity and predict DBP formation with varying degrees of success. Hao et al. (2012) reported strong correlations (R^2: 0.87 to 0.95) between fluorescence intensity for selected humic and fulvic acid peaks as well as THM and HAA formation potential in reclaimed water. Pifer and Fairey (2012) found increased correlation strength between a humic-like fluorophore and chloroform (R^2: 0.84) using PARAFAC, in comparison to SUVA (R^2: 0.51). Hua et al. (2010) identified a slight increase in correlation between two PARAFAC factors and total THM concentrations when compared to SUVA (R^2: 0.58 vs 0.64, 0.54 *vs.* 057) for a large range of THM concentrations (100–600 μg/L).

This work set out to apply PCA to fluorescence spectra for the prediction of THM and HAA formation in four different source waters to compliment recent results using the PARAFAC approach. Unlike with PARAFAC, PCA was used with the objective of reducing dimensionality representation of the FEEM rather than for identifying individual components in the fluorescence spectra. It was hypothesized that an increased variance explained by PCA along with constraints of component orthogonality would ensure variable independence in the DBP model and could improve prediction of DBP formation for varying water types and precursor concentrations. As such, the FEEM–PCA approach is compared to traditional NOM indicators, including DOC, UVA, SUVA, as well as PARAFAC and peak picking.

1. Material and methods

1.1. Source waters

Four distinct natural surface waters in Ontario, Canada, were selected to cover a range of NOM concentrations (2.4–5.9 mg/L DOC) and represent both lakes and river sources. Water quality parameters are provided in Table 1.

1.2. Jar tests

A bench-scale jar test approach was used to simulate conventional water treatment, including coagulation, flocculation, sedimentation, and filtration. All waters were coagulated with aluminum sulfate (alum) (General Chemical, Parsippany, New Jersey, USA). To ensure a range of DOC alum was dosed between 5 and 70 mg/L (5, 10, 20, 30, 40, 50, 60, 70 mg/L alum or 0.45, 0.89, 1.78, 2.67, 3.56, 4.45, 5.34, 6.23 mg/L as Al). Tests were conducted using a PB-700 Standard Jar Tester paddle stirrer with six square, acrylic 2 L square containers (Phipps & Bird, Richmond, Virginia, USA). The test protocol was adapted from the US EPA's Enhanced Coagulation Guidance Manual (US EPA, 1999). Coagulation was simulated through rapid mix (100 r/min) for 1.5 min followed by reducing the mixing speed to 30 r/min for 15 min to provide flocculation. To simulate sedimentation, the water was then allowed to stand for 30 min. Vacuum filtration was applied using 1.2 micron glass microfiber filters (Whatman, Florham Park, NJ, USA) to represent anthracite–sand media filters. The finished water was analyzed for organic material and pH. For DOC analysis the water was also filtered through 0.45 micron membrane filters (Pall Corporation, Port Washington, NY, USA) to ensure that all particulates had been removed (US EPA, 1999).

1.3. Disinfection byproduct formation and analysis

To ensure consistent conditions, the pH of the finished water was adjusted to 7.0 ± 0.1 using sulfuric acid or sodium hydroxide post-filtration, prior to chlorination. Chlorine dosages of 2.5 and 3.5 mg/L were applied to represent those typically used at water treatment plants associated with the source waters. Chlorinated samples were sealed head space free in pre-cleaned chlorine demand free (acid washed, distilled water rinse, soaked in dilute sodium hypochlorite solution for 8 hr) and incubated at 21 ± 1°C, for 24 hr. The free chlorine residual was then measured (0.02 to 2.54 mg/L) and the remaining chlorine quenched using ascorbic acid (50 mg/L) (Westerhoff et al., 2005). Duplicate samples were then sealed head space free and retained for DBP analysis. All waters were tested for THM formation and all except Lake Simcoe were analyzed for formation of nine haloacetic acids (HAA_9).

Trihalomethane analysis was conducted using EPA liquid–liquid extraction Method 551.1 using methyl tert-butyl ether (MTBE) (US EPA, 1995); for HAA_9, EPA Method 552.3 was used (US EPA, 2003). This allowed for quantification of four THM species and nine HAA species listed in the methods. Analyses were conducted using a Hewlett Packard 5890 Series II Plus gas

Table 1 – Source water characteristics.

	Otonabee River	Lake Simcoe	Lake Ontario	Ottawa River
DOC (mg/L)	5.6	4.1	2.4	5.9
UVA at 254 nm (1/cm)	0.128	0.065	0.017	0.200
SUVA (L/(mg·m))	2.29	1.66	0.71	3.39
Alkalinity (mg/L as $CaCO_3$)	101	121	87	35
pH	8.3	8.2	8.3	7.4

DOC: dissolved organic carbon; UVA: ultraviolet absorbance; SUVA: specific ultraviolet absorbance.

chromatograph (Hewlett Packard, Mississauga, ON, Canada) equipped with an electron capture detector and a J&W Science DB-5.625 durabond column (length: 30 m, inner diameter: 0.25 mm, film: 0.25 μm) (Agilent Technologies Canada Inc., Mississauga, ON, Canada). Injections were run in splitless mode, with helium as the carrier gas and an argon/methane (95%/5%) mix as make-up gas.

1.4. DOC, TOC, and UVA measurements

Concentrations of DOC and TOC were determined *via* heated persulfate oxidation using an Aurora 1030 organic carbon analyzer (O.I. Analytical, College Station, TX, USA) following Standard Method 5310 D (APHA, 2005). UVA at 254 nm was determined using a CE 3055 model spectrophotometer (Cecil Instruments, Cambridge, UK) with a quartz cuvette following Standard Method 5910 B (APHA, 2005).

1.5. Fluorescence spectra collection

FEEMs were collected using a Luminescence Spectrometer LS50B (Perkin-Elmer, Waltham, MA, USA). No pre-treatment of the samples was applied, except that all had been previously adjusted to a pH of 7.0 ± 0.1. A common pH among samples ensured that fluorescence characteristics of the acidic functional groups in humic molecules remained constant (Mobed et al., 1996). Possible inner-filtering effects, which cause peak shifts and intensity reduction, were not accounted for since it has been reported that effects are not expected below 25 mg/L TOC (Henderson et al., 2009). Collection of intensity values occurred within excitation–emission ranges of 250–380 (10 nm increment) and 300–600 nm (1 nm increments), respectively. Scan rate was set to 600 nm/min, slit width 10 nm, and photomultiplier tube voltage 775 V. Instrument settings were determined based on ranges used in previous studies (Bieroza et al., 2010), that were shown to increase resolution (Peiris et al., 2009), and in-house testing to optimize FEEM collection. UV-Grade polymethylmetacrylate cuvettes (VWR, Mississauga, ON, Canada) with four optical windows were used which have been shown to be appropriate for the purpose of distinguishing NOM elements using fluorescence (Peiris et al., 2008). Spectra for Milli-Q® water were subtracted from intensity values of sample spectra to reduce background noise effects.

To track any potential instrumental changes, Milli-Q® samples that were collected over the experimental period were compared using the uncorrected matrix correlation method. This method allows comparison of two entire matrices and indicates similarity using a value from 0 to 1 (1 representing a perfect correlation). Using this technique the relative mean square error (RMSE) between the matrices can be estimated (Burdick and Tu, 1989). All matrix correlations had RMSE values below 0.06 indicating a high degree of similarity between Milli-Q® spectra and supporting instrument and hardware stability during the experimental period.

1.6. Fluorescence data analysis

Each sample produced a total of 4214 fluorescence intensity values at unique excitation–emission wavelength pairs. In total, 35 different samples were run in duplicate (70 FEEMs collected in total). Prior to data analysis using PARAFAC or PCA, Rayleigh scattering regions were removed with a 15 nm margin. Furthermore, emissions above second order (emission twice the excitation wavelength) and below first order (emission equal to excitation wavelength) were removed. For PCA, each variable (excitation/emission pair), was mean centered and scaled to unit variance in order to remove bias towards compounds and spectral regions with higher variability. PCA was performed using R V3.0.2 (R Core Team, 2013). PARAFAC analysis was implemented using the N-way toolbox (Andersson and Bro, 2000) in MATLAB 7.12.0 (MathWorks, Natick, MA, USA). Constraints of non-negativity were used in all modes and un-modality for excitation and emission modes. A second model was made without constraints for comparison. Spectra were pre-processed with scaling and centering, as described by Bro (1997). A peak picking method was implemented using scripts written in R V3.0.2. Locations of distinctive peaks were identified from the fluorescence spectra of raw water. The fluorescence intensity at this excitation/emission pair was then used to track peak intensity changes between samples.

2. Results and discussion

2.1. Fluorescence results from jar tests

Fluorescence intensity values for each coordinate pair were plotted to visualize the spectra. Based on the location of the two main intensity peaks (Table 2) and comparison to the literature, peak *a* was attributed to represent fulvic-acid type matter (Ex/Em: 270 nm/430 nm) while peak *b* represented humic-acid type material common to fresh waters (Ex/Em: 340 nm/435 nm) (Murphy et al., 2008; Chen et al., 2003). Peaks of high intensity at Ex/Em: 250–300/500–600 nm and Ex/Em: 300–380/300–380 nm are representative of the second and the first order Rayleigh scattering, respectively, which are related to the concentration of particulates in the sample (Peiris et al., 2010). Other studies have reported a third peak in the region Ex/Em: 250–290/300–350 nm, which is attributed to protein-like material (Chen et al., 2003). This peak was not apparent in the raw fluorescence spectra from this study. Each of the four water sources differed in overall intensity of peaks *a* and *b*; however their presence and location were approximately consistent, when considering peak locations (Table 2, Fig. 1).

PCA was applied to the dataset of 140 spectra with scattering regions removed (all water sources; 70 unique samples in duplicate). The majority of the variance in the dataset was explained by the first two principle components; 87.94% and 7.82% variance explained by PC1 and PC2, respectively (Fig. 2).

Spectral regions represented by each PC were identified using loading values. Loading plots for the first 6 principal components

Table 2 – Locations of peaks *a* and *b* in each source water.

Peak	Excitation/emission of peak (nm/nm)			
	Otonabee River	Lake Simcoe	Lake Ontario	Ottawa River
a	280/430	280/437	280/428	280/437
b	340/434	340/431	340/429	340/443

are shown in Fig. 2 along with the variance explained. PC1 is generally attributed to humic-like material (Peiris et al., 2010); however there were no pronounced peaks for the water sources examined in this study. The broad area of high loading values (in the negative direction) in PC1 indicated that this region varied equally between samples after the data was mean centered and scaled. The most pronounced loading values from PC2 in low excitation/emission regions (Ex/Em: 250–290/300–350 nm) have been identified to be representative of protein-like substances (Chen et al., 2003). PC3 indicated high loading values in a region around Ex/Em: 330/375 nm may be attributed to the presence of polycyclic aromatic hydrocarbons (Murphy et al., 2008). As with PC2, high loadings in low excitation–emission regions for PC4 and PC5 were thought to represent soluble microbial byproducts. PC6 shows a high degree of similarity to PC3, although there were pronounced negative loading values in fulvic-acid like regions. Based on the wide regions in the loading plots, it should be emphasized that each PC did not represent singular compounds. As such, labels of humic-like and protein-like are used to most accurately represent the compound classes identified. Furthermore, the intention of applying PCA was not for identification of singular components, but rather to provide a reduced dimensional mathematical representation of the full fluorescence spectra.

The optimal number of factors to be included in the PARAFAC solution was determined to be 2 through analysis of sum of squared error and core consistency (Andersen and Bro, 2003). A good fit is indicated by a core consistency close to 100% as well as minimal error reduction through addition of another factor. A marked drop in core consistency (82% to 26%) from 2 to 3 factors was found for the constrained model and conformed well to a minimal reduction in sum of squared errors shown in Fig. 3. Results from the unconstrained model demonstrated identical trends.

PARAFAC loading plots were analyzed to determine the two components identified. Excitation and emission loadings are shown in Fig. 4. Despite broad excitation loadings, both factors 1 and 2 resided in the humic-acid like region (factor 1: Ex/Em 360/450 nm; factor 2: Ex/Em 290/385 nm) (Chen et al., 2003).

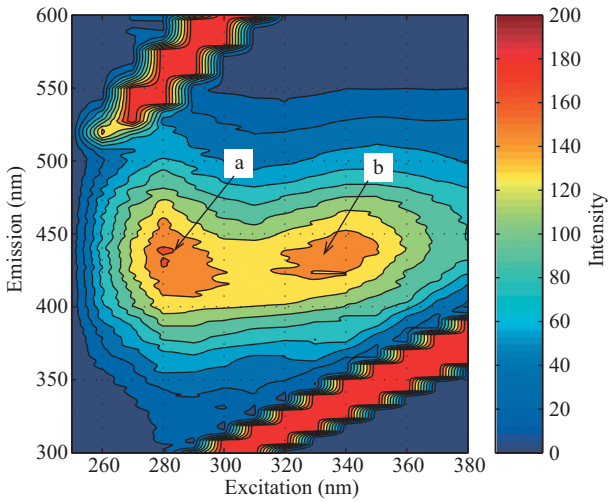

Fig. 1 – Example raw fluorescence spectra of Ottawa River raw water.

2.2. Disinfection by-product formation prediction

All NOM estimation parameters showed a decreasing organic concentration with an increasing alum dose in filtered water; THM and HAA concentrations resulting from chlorination also decreased. Total concentrations of THMs and HAAs varied from 9.2 to 112 μg/L and from 11.7 to 128 μg/L, respectively for all waters and chlorine doses. A reduction in DBPs with increasing coagulant dose was more pronounced for the river sources. Results from other studies have shown similar results of decreasing DBP concentrations with increasing NOM removal following coagulation (Sadiq and Rodriguez, 2004).

For THMs, only trichloromethane (TCM) and bromodichloromethane (BDCM) were observed above the method detection limits (MDL) (2.1 to 3.0 μg/L for each specie). For HAAs, only dichloroacetic acid (DCAA) and trichloroacetic acid (TCAA) were greater than the MDL (0.7–9.6 μg/L for each specie). When considering river source waters at either chlorine dose, BDCM represented 20%–30% of total THMs and 35%–40% of the total for lake sources. TCAA represented 40%–50% of HAA_9 for river sources and 30%–40% for lake sources. Others have presented similar speciation and total THM and HAA concentration ranges for waters with low bromide ion concentrations (<0.01 mg/L) and similar organic content (2–7 mg/L TOC) (Williams et al., 1996; Ates et al., 2007).

2.3. DBP modeling

A multi-linear model was developed to fit the DBP data. Reaction time, chlorination pH, and temperature were controlled to be equal for all waters and samples. Bromide ion concentrations were not explicitly controlled, although the only brominated DBP observed above its MDL was BDCM. Since only organic concentration and chlorine dose were varied between samples, the model was simplified to only include these variables:

$$\text{DBP concentration} = c_1 + c_2 \times \text{chlorine dose} + c_3 \times \text{NOM concentration}$$

where, c_1, c_2 and c_3 are constants determined *via* regression. For NOM measures with more than one variable (*i.e.* PCA), each variable was included individually into the model.

Models with respect to DBPs (total THMs and HAA_9) and each NOM indicator (DOC, UVA, SUVA, PC scores, PARAFAC scores, and peak intensities) were regressed using *R*. A cross-fold validation approach was applied where the full dataset was randomly split into 7 equal sets of 10 samples, where 6 sets were used to train the model while the remaining set was used as test data. Mean squared error (MSE) was calculated based on the difference between predicted and actual DBP concentrations in the test set for each cross validation fold. The average MSE of all folds was used for comparing model performance. To further ensure that the average MSE was representative of model performance, a cross validation was performed 10 times with different randomized training and test sets.

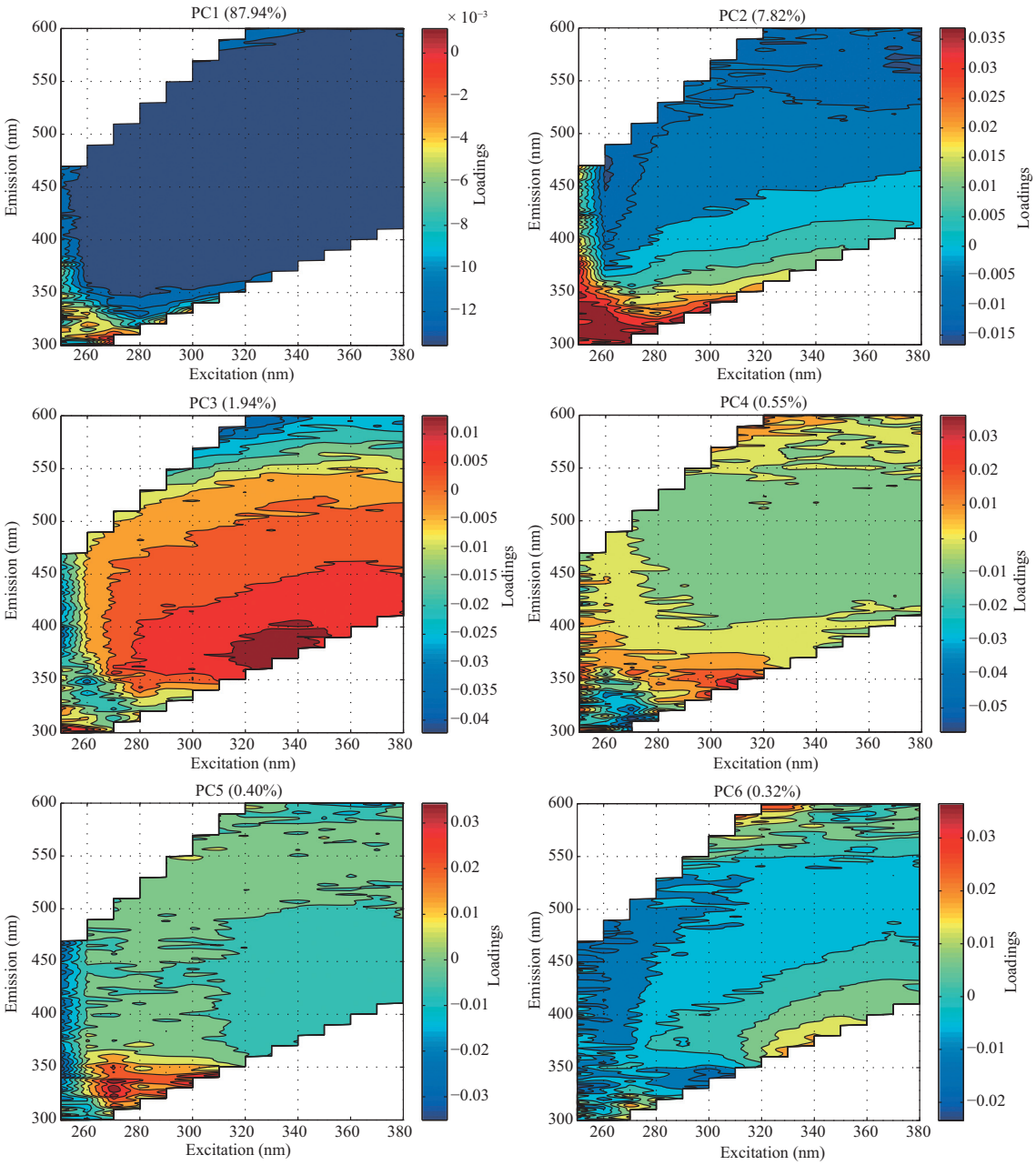

Fig. 2 – Loading plots from principal component analysis along with variance explained by each principal component (PC) for components 1 to 6.

The optimal number of PCs to be used in the model was determined by sequentially including PCs 1 to 14 and observing changes to MSE. A plot of the number of PCs *vs*. MSE for prediction of THMs and HAAs is shown as Fig. 5.

Based on the average MSE of tests sets for prediction of THMs and HAAs, a clear optimum number of PCs was observed. The optimum for THMs occurred at 5 PCs while 2 PCs were ideal for HAAs. Since inclusion of PC3 increased the MSE, modeling with PCs 1, 2, 4, and 5 resulted in a slightly reduced MSE (change of 1.8 $(\mu g/L)^2$). MSE of HAA modeling was reduced significantly with the addition of PC7. However, by combining PCs 1, 2 and 7 error was not reduced when compared to using only the first two PCs. Error did not continue to decline with inclusion of greater number of PCs. It is postulated that past an optimum value, PCs generally represented non-DBP producing fractions and/or modeled noise in the FEEMs, ultimately introducing error into the prediction model. Furthermore, it should be noted that FEEMs can only directly identify fluorescing compounds. Non-fluorescing compounds, which possibly contribute to DBP formation, are therefore not directly identifiable through this method.

PCs 4 to 14 explained less than 1% of the variance in the dataset; however exhibited a significant impact on THM modeling error. Inclusion of PC4, which explained 0.55% of variance, had the most substantial effect. PCs explaining such low amounts of variance are typically excluded from further

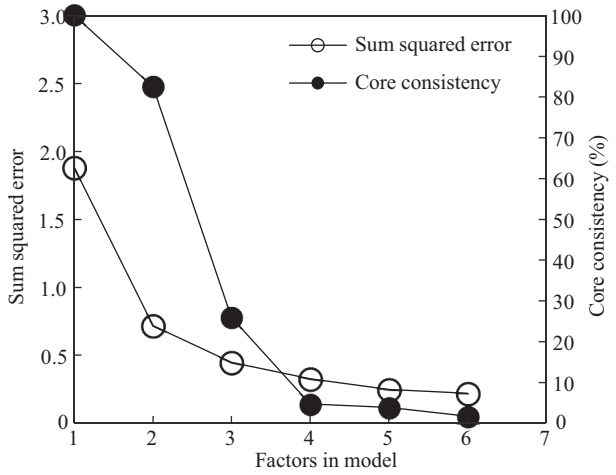

Fig. 3 – Determination of number of factors through error explained and core consistency.

Table 3 – Results from optimized multi-linear models.

Natural organic matter (NOM) surrogate	Mean squared error ((μg/L)^2)	
	Trihalomethanes (THMs)	Haloacetic acids (HAAs)
Dissolved organic carbon	137.5	406.1
Ultraviolet absorbance	97.8	247.0
Specific ultraviolet abosrbance	68.2	480.2
Principal components 1, 2, 4, 5	43.7	–[a]
Principal components 1, 2	–[a]	233.3
Peak *a* and *b*	81.1	242.3
Parallel factors analysis (2 factors; constrained)	96.3	266.8
Parallel factors analysis (2 factors; no constraints)	94.9	264.1

[a] Optimized models for THMs and HAAs required differing numbers of principal components, non-optimized results are not shown.

analysis (Peiris et al., 2010), although appear to be important in this application and waters studied. Loading values for this PC showed positive representation of Ex/Em region 300–320/340–350 nm, which potentially are characteristic of aromatic amino acids (Murphy et al., 2008). The PC also had negative representation at lower excitation/emission wavelengths, also representative of soluble microbial by-products or protein-like substances. Similarly for HAAs, PC2 which had strong representation of protein-like material and reduced the average MSE of test sets to a minimum. This suggests that a portion of THM and HAA formation occurred from protein-like material, which is supported by previous studies (Huang et al., 2009; Henderson et al., 2008).

In comparison to other NOM characterization measures, including DOC, UVA, and SUVA, as well as other fluorescence-based measures, PC scores demonstrated the lowest average MSE for both THMs and HAAs (Table 3). Errors for all THM models were markedly lower than those for HAAs. Example

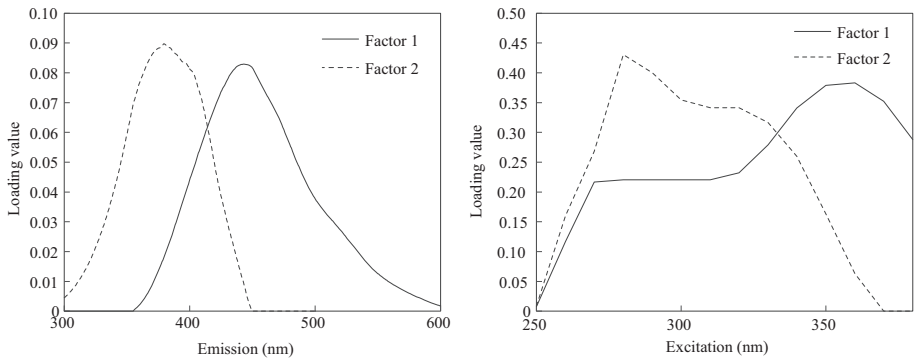

Fig. 4 – Parallel factors analysis (PARAFAC) loading plots for the constrained model.

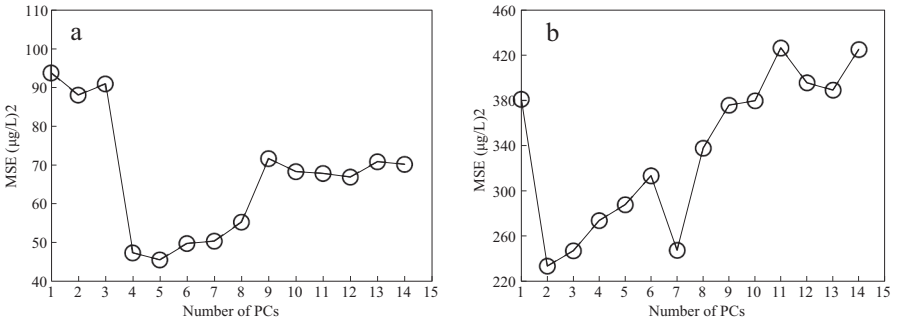

Fig. 5 – Number of principal components (PCs) used in linear model (a) trihalomethanes (THMs) and (b) haloacetic acids (HAAs) *vs.* average mean squared error (MSE) from cross fold validation.

THM models using each organic matter surrogate are shown in Fig. 6. Using PC scores, the line of best fit for actual *vs.* predicted THM values conformed well with the ideal 1:1 slope; all other surrogates showed underestimation at high concentrations and overestimation at low concentrations. Similar results for HAA prediction were also observed (not shown). DOC did not represent NOM reactivity for formation of THMs or HAAs well and, in particular, when considering low organic content samples from Lake Ontario. Fluorescence and UV based measures better accounted for NOM reactivity for DBP formation, likely due to better representation of aromatic structures which have been shown to be correlated with DBP formation (Chen et al., 2003; Barrett et al., 2000).

Error rates from PARAFAC results were similar to SUVA and peak picking for both THMs and HAAs. The constraints of non-negativity and unimodality applied to the PARAFAC model were chosen for better representation of individual fluorophores. In this way, two modeling approaches, one utilizing more pure representation of individual components and the other with abstract mathematical representations (PCA), were compared. While the components identified from PARAFAC were much more interpretable, it resulted

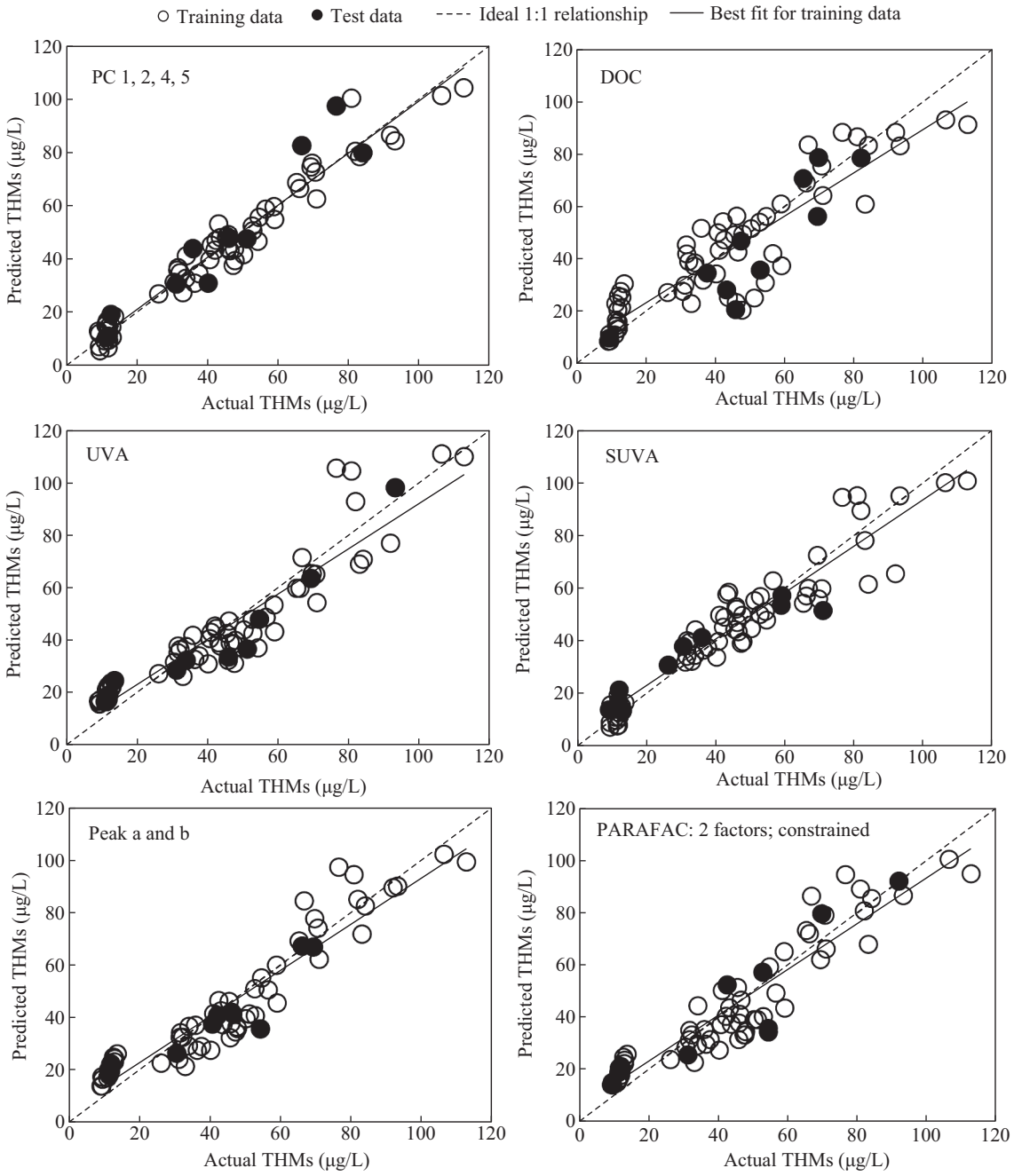

Fig. 6 – Example modeling results (actual trihalomethanes (THMs) *vs.* predicted THMs) using dissolved organic carbon (DOC), principal components (PCs), ultraviolet absorbance (UVA), specific ultraviolet absorbance (SUVA), and parallel factors analysis (PARAFAC).

in increased modeling error when compared to PCA. To identify the impact of applied constraints, a non-constrained PARAFAC model was also fitted, which did not yield a substantial reduction in error.

3. Conclusions

Results from this modeling study demonstrated FEEM–PCA to be a strong indicator of NOM reactivity for DBP formation. Multilinear modeling using PC scores resulted in the lowest prediction error for test sets (THMs: 43.7 $(\mu g/L)^2$, HAAs: 233.3 $(\mu g/L)^2$) when compared to DOC, UVA, SUVA, PARAFAC, and fluorescence peak picking. A pronounced optimum number of PCs were identified which included components representing less than 1% of the variance in the dataset. For both THMs and HAAs, inclusion of protein-like components resulted in reduced prediction error. Modeling results were conducted using pooled data for four unique water sources, thereby in part identifying the ability of NOM surrogates to represent reactivity under different source conditions. Due to variability in NOM components in source waters as well as resulting from treatment, the optimized models of this study are not universally applicable. Further work is needed to identify suitability of the proposed approach to a wider range of source waters. While resulting PCs have vague physical representations of individual NOM components, it is hypothesized that the orthogonality of principle components and limited constraints makes PCA an attractive method for dimensionality reduction of fluorescence spectra when results are subsequently utilized in a statistical correlation model.

Acknowledgments

This work was funded in part by the Canadian Water Network and the Natural Sciences and Engineering Research Council of Canada Chair in Drinking Water Research at the University of Toronto. We would like to thank Dr. Raymond Legge (University of Waterloo) for his constructive comments and suggestions. We would also like to thank the personnel at the Peterborough Utilities Commission, Toronto Water, Barrie surface water treatment plant, and City of Ottawa Britannia water purification plant for providing samples.

REFERENCES

Andersen, C.M., Bro, R., 2003. Practical aspects of PARAFAC modeling of fluorescence excitation–emission data. J. Chemometr. 17 (4), 200–215.

Andersson, C.A., Bro, R., 2000. The *N*-way toolbox for MATLAB. Chemometr. Intell. Lab. 52 (1), 1–4.

APHA (American Public Health Association), AWWA (American Water Works Association), WEF (Water Environment Federation), 2005. Standard Methods for the Examination of Water and Wastewater 21st ed., (Washington D.C.).

Ates, N., Sule Kaplan, S., Sahinkaya, E., Kitis, M., Dilek, F.B., Yetis, U., 2007. Occurrence of disinfection by-products in low DOC surface waters in Turkey. J. Hazard. Mater. 142 (1–2), 526–534.

Bahram, M., Bro, R., Stedmon, C., Afkhami, A., 2006. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. J. Chemometr. 20 (3–4), 99–105.

Barrett, S.E., Krasner, S.W., Amy, G.L., 2000. Natural organic matter and disinfection by-products: characterization and control in drinking water—an overview. In: Barrett, S.E., Krasner, S.W., Amy, G.L. (Eds.), Natural Organic Matter and Disinfection By-products. Oxford University Press, Oxford, pp. 2–14.

Bieroza, M., Baker, A., Bridgeman, J., 2010. Classification and calibration of organic matter fluorescence data with multiway analysis methods and artificial neural networks: an operational tool for improved drinking water treatment. Environmetrics 22 (3), 256–270.

Bro, R., 1997. PARAFAC. Tutorial and applications. Chemometr. Intell. Lab. 38 (2), 149–171.

Burdick, D.S., Tu, X.M., 1989. The wavelength component vectorgram: a tool for resolving two-component fluorescent mixtures. J. Chemometr. 3 (2), 431–441.

Chen, W., Westerhoff, P., Leenheer, J.A., Booksh, K., 2003. Fluorescence excitation–emission matrix regional integration to quantify spectra for dissolved organic matter. Environ. Sci. Technol. 37 (24), 5701–5710.

Chowdhurry, S., Champagne, P., McLellan, P.J., 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: a chronological review. Sci. Total Environ. 407 (14), 4189–4206.

Hao, R., Ren, H., Li, J., Ma, Z., Wan, H., Zheng, X., et al., 2012. Use of three-dimensional excitation and emission matrix fluorescence spectroscopy for predicting the disinfection by-product formation potential of reclaimed water. Water Res. 46 (17), 5765–5776.

Henderson, R.K., Baker, A., Parsons, S.A., Jefferson, B., 2008. Characterisation of algogenic organic matter extracted from cyanobacteria, green algae and diatoms. Water Res. 42 (13), 3435–3445.

Henderson, R.K., Baker, A., Murphy, K.R., Hambly, A., Stuetz, R.M., Khan, S.J., 2009. Fluorescence as a potential monitoring tool for recycled water systems: a review. Water Res. 43 (4), 863–881.

Her, N., Amy, G., McKnight, D., Sohn, J., Yoon, Y., 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. Water Res. 37 (17), 4295–4303.

Hua, G., Reckhow, D.A., 2007. Characterization of disinfection by-product precursors based on hydrophobicity and molecular size. Environ. Sci. Technol. 41 (9), 3309–3315.

Hua, B., Veum, K., Yang, J., Jones, J., Deng, B., 2010. Parallel factor analysis of fluorescence EEM spectra to identify THM precursors in lake waters. Environ. Monit. Assess. 161 (1–4), 71–81.

Huang, J., Graham, N., Templeton, M.R., Zhang, Y., Collins, C., Nieuwenhuijsen, M., 2009. A comparison of the role of two blue-green algae in THM and HAA formation. Water Res. 43 (12), 3009–3018.

Mobed, J.J., Hemmingsen, S.L., Autry, J.I., McGown, L.B., 1996. Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. Environ. Sci. Technol. 30 (10), 3061–3065.

Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. Mar. Chem. 108 (1), 40–58.

Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., et al., 2011. Organic matter fluorescence in municipal water recycling schemes: toward a unified PARAFAC model. Environ. Sci. Technol. 45 (7), 2909–2916.

Peiris, B.R.H., Hallé, C., Haberkamp, J., Legge, R.L., Peldszus, S., Moresoli, C., et al., 2008. Assessing nanofiltration fouling in drinking water treatment using fluorescence fingerprinting and LC-OCD analyses. Water Sci. Technol. 8 (4), 459–465.

Peiris, R.H., Budman, H., Moresoli, C., Legge, R.L., 2009. Acquiring reproducible fluorescence spectra of dissolved organic matter at very low concentrations. Water Sci. Technol. 60 (6), 1385–1392.

Peiris, R.H., Hallé, C., Budman, H., Moresoli, C., Peldszus, S., Huck, P.M., et al., 2010. Identifying fouling events in a membrane-based drinking water treatment process using principal component analysis of fluorescence excitation–emission matrices. Water Res. 44 (1), 185–194.

Persson, T., Wedborg, M., 2001. Multivariate evaluation of the fluorescence of aquatic organic matter. Anal. Chim. Acta. 434 (2), 179–192.

Pifer, A.D., Fairey, J.L., 2012. Improving on $SUVA_{254}$ using fluorescence-PARAFAC analysis and asymmetric flow-field flow fractionation for assessing disinfection byproduct formation and control. Water Res. 46 (9), 2927–2936.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (URL: http://www.R-project.org/).

Richardson, S.D., Postigo, C., 2012. Drinking water disinfection by-products. In: Barceló, D. (Ed.), Emerging Organic Contaminant and Human Health. Springer, Berlin Heidelberg, pp. 93–137.

Roccaro, P., Vagliasinid, F.G.A., Korshin, G.V., 2009. Changes in NOM fluorescence caused by chlorination and their associations with disinfection by-products formation. Environ. Sci. Technol. 43 (3), 724–729.

Sadiq, R., Rodriguez, M.J., 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Sci. Total Environ. 321 (1–3), 21–46.

Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Mar. Chem. 82 (3), 239–254.

US EPA, 1995. Determination of chlorination disinfection byproducts, chlorinated solvents, and halogenated pesticide/herbicides in drinking water by liquid–liquid extraction and gas chromatography with electron-capture detection. Method 551.1 Revision 1.0.

US EPA, 1999. Microbial and disinfection by-product rules simultaneous compliance guidance manual. EPA 815-R-99-015.

US EPA, 2003. Determination of haloacetic acids and dalapon in drinking water by liquid–liquid microextraction, derivatization, and gas chromatography with electron capture detection. EPA 815-B-03-002.

Westerhoff, P., Yoon, Y., Snyder, S., Wert, E., 2005. Fate of endocrine-disruptor, pharmaceutical, and personal care product chemicals during simulated drinking water treatment processes. Environ. Sci. Technol. 39 (17), 6649–6663.

Williams, D.T., Lebel, G.L., Benoit, F.M., 1996. Disinfection by-products in Canadian drinking water. Chemosphere 34 (2), 299–316.

Zepp, R.G., Sheldon, W.M., Moran, M.A., 2004. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. Mar. Chem. 89 (1–4), 15–36.

Editorial Board of Journal of Environmental Sciences

JOURNAL OF ENVIRONMENTAL SCIENCES

环境科学学报(英文版)

www.jesc.ac.cn

Aims and scope

Journal of Environmental Sciences is an international academic journal supervised by Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. The journal publishes original, peer-reviewed innovative research and valuable findings in environmental sciences. The types of articles published are research article, critical review, rapid communications, and special issues.

The scope of the journal embraces the treatment processes for natural groundwater, municipal, agricultural and industrial water and wastewaters; physical and chemical methods for limitation of pollutants emission into the atmospheric environment; chemical and biological and phytoremediation of contaminated soil; fate and transport of pollutants in environments; toxicological effects of terrorist chemical release on the natural environment and human health; development of environmental catalysts and materials.

For subscription to electronic edition

Elsevier is responsible for subscription of the journal. Please subscribe to the journal via http://www.elsevier.com/locate/jes.

For subscription to print edition

China: Please contact the customer service, Science Press, 16 Donghuangchenggen North Street, Beijing 100717, China. Tel: +86-10-64017032; E-mail: journal@mail.sciencep.com, or the local post office throughout China (domestic postcode: 2-580).

Outside China: Please order the journal from the Elsevier Customer Service Department at the Regional Sales Office nearest you.

Submission declaration

Submission of the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere. The publication should be approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out. If the manuscript accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Editorial

Authors should submit manuscript online at http://www.jesc.ac.cn. In case of queries, please contact editorial office, Tel: +86-10-62920553, E-mail: jesc@rcees.ac.cn. Instruction to authors is available at http://www.jesc.ac.cn.

Journal of Environmental Sciences (Established in 1989)
Volume 27 2015

Supervised by	Chinese Academy of Sciences	**Published by**	Science Press, Beijing, China
Sponsored by	Research Center for Eco-Environmental		Elsevier Limited, The Netherlands
	Sciences, Chinese Academy of Sciences	**Distributed by**	
Edited by	Editorial Office of Journal of	Domestic	Science Press, 16 Donghuangchenggen
	Environmental Sciences		North Street, Beijing 100717, China
	P. O. Box 2871, Beijing 100085, China		Local Post Offices through China
	Tel: 86-10-62920553; http://www.jesc.ac.cn	Foreign	Elsevier Limited
	E-mail: jesc@rcees.ac.cn		http://www.elsevier.com/locate/jes
Editor-in-chief	X. Chris Le	**Printed by**	Beijing Beilin Printing House, 100083, China

CN 11-2629/X **Domestic postcode: 2-580** **Domestic price per issue RMB ¥ 110.00**

ISSN 1001-0742