ELSEVIER

Journal of Hazardous Materials B100 (2003) 117–130

Journal of Hazardous Materials

www.elsevier.com/locate/jhazmat

Comparison of spectral fluorescent signatures-based models to characterize DOM in treated water samples

Karim Bengraïne, Taha F. Marhaba*

Department of Civil and Environmental Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

Received 8 January 2003; received in revised form 27 January 2003; accepted 1 March 2003

Abstract

Statistical procedures enable a multivariate analysis of the measurements to identify specific characteristics of the dissolved organic matter (DOM) fractions in raw natural water, including the concentrations. In this work, three already established models were used to predict the concentrations of fractions of DOM from spectral fluorescent signatures (SFSs): a general linear regression (GLR), loadings and scores of a principal components analysis (PCA), and a partial least squares regression (PLS). Details about the method undertaken to prepare the fractions were given. Water samples from surface water treatment plants in New Jersey were used for the testing. In all cases, PLS have shown much better biases and accuracies than GLR and PCA models. Hydrophilic neutral, however, showed poor performances (bias 33%) due to the isolation technique used. Recommendations were provided in order to improve the DOM characterization through SFS, which linked to PLS make a powerful and cost-effective surrogate parameter to characterize DOM.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Dissolved organic matter; Hydrophilic; Hydrophobic; Spectral fluorescence signatures

1. Introduction

Dissolved organic carbon (DOC) is typically used as an aggregate measure of the organic content in water. However, DOC as a conventional parameter does not accurately indicate

Abbreviations: DBPs, disinfection by-products; DOC, dissolved organic carbon; DOM, dissolved organic matter; GLR, general linear regression; mg/L, milligram/liter; NOM, natural organic matter; PCA, principal component analysis; PLS, partial least squares regression; PVWC, Passaic Valley Water Commission; SFS, spectral fluorescent signature; WTP, water treatment plant

* Corresponding author. Tel.: +1-973-642-4599; fax: +1-973-596-5790.
E-mail address: marhaba@adm.njit.edu (T.F. Marhaba).

0304-3894/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0304-3894(03)00071-2

the character of the organic matter in water. Further, it does not distinguish the problematic fraction to disinfection by-products (DBPs) formation. Spectrofluorescence has shown to provide an excellent diagnostic of organic pollution in open sea [1]. Thus, it could represent a promising alternative to DOC. This explains why both DOM fractionation and the SFS technique development are currently valuable in water treatment research. Other advantages of SFS are diverse. First, it does not require the labor-intensive and time-consuming pretreatment of the water sample. Second, it allows online processing and analysis, while minimizing both the operation cost and the time for analysis [2,3]. Third, SFS could help understand the role of each DOM fraction in initiating, promoting, or inhibiting the DBPs formation at a given step of the water treatment, thus answering one of the major concerns of the disinfection and disinfection by-products and the enhanced surface water treatment rules [4].

Practically, SFS is the sum of emission spectra across the entire usable region for an aqueous sample at different excitation wavelengths. In this research, excitation and emission wavelengths are varied together to build extended multivariate datasets. Thus, an excitation–emission matrix (EEM) of fluorescence obtained for each sample can be arranged either in data vector arrays or in data matrices for further analysis [5].

Looking closely to the available literature involving fluorescence shows that:

- the majority of the work dealing with the application of chemometric methods in fluorescence spectroscopy, concern traditional techniques like near-infrared (NIR) spectroscopy [6–10];
- only synthetic samples or natural samples that contain a few chemical species are analyzed [6];
- the measured EEM is described as a pure excitation and emission spectra of the chemical composition in the samples [10];
- the main objective is the prediction of a given analyte concentration by rank annihilation methods [11–13].

Regression techniques were used intensively during the last decade to model and correlate water chemistry parameters. Indeed, DBPs, chlorophyll-a, UV_{254}, and chlorine dosage have been investigated [14–18]. Associated to multivariate methods, fluorescence has proven to be effective with respect to prediction of quality parameters in different type of samples [19]. Near infrared spectroscopy was used successfully to analyze sugar samples with principal components analysis (PCA) and partial least squares regression (PLS) [20]. Further, untreated and complete SFS and PLS can be used to predict DOC in natural water samples, and suggestion has been made on the possibility of applying the SFS–PLS method to isolated DOM fractions—hydrophilics (HPI) and hydrophobics (HPO)—concentrations, thus allowing the DOM characterization [21].

The most commonly used DOM isolation techniques are fractionation on XAD-type macroporous resins [22–24] by gel filtration [25,26], and by ultrafiltration [27,28], while the characterization of the isolated fractions is made possible by pyrolysis GC–MS [29], C-NMR spectra [30].

The resin fractionation procedure developed by Leenheer and Huffman [22] aimed to determine the fraction distribution of DOM in water, therefore was described as analytical. The one developed by Leenheer [23] aimed to prepare organic-concentrated fractions for

subsequent tests without further coping with tremendous volumes of sample, which makes it preparative but also comprehensive since all fractions are obtained rather than humic substances. The concerns that have been raised about the existing fractionation protocols are the following.

- Lack of a universal fractionation procedure for all application in accordance with the variety of research objective and sample matrix [31,32].
- Fractions are more operationally defined than structurally.
- Limited applications since they are not recommended for DOM <5 mg/l [33].

Depending on the type of resins used, the result of a comprehensive fractionation broadly falls into two categories. First, the separation hydrophobics (HPO) versus hydrophilics (HPI) through a system of nonionic resins XAD-8 and XAD-4 [34,35]. Second, a complete separation through ionic and nonionic resins which gives six groups [22,23,32,33] is defined as given further.

Table 1
Conditions of the different models developed to characterize the DOM of treated water sample

Models' calibration and fractionation conditions
SFS–GLR [17]
Calibrated using samples from PVWC, CR and RM WTPS sampled on 16 April 1998.
Analysis of the response surface and the emission characteristics, which involved the area, the slope and the intensities of fluorescence of the major peaks of each fraction, were documented manually.
Fractions were obtained from a modified protocol of the one originated by Leenheer [31] through a combination of XAD-8, AG-MP-50 and Dualite A7 resins.
Model expression
$C = -0.0074 + 0.0003548$ area $+ 3.317$ slope $- 0.00445$ slope $\times$ area $+ 0.10875$ HPOA $+ 0.004$ HPOB $+ 0.16475$ HPON $+ 0.80225$ HPIA $- 0.225$ HPIB
SFS–PCA [3,31]
Calibrated using samples from CR and RM WTPs sampled 16 April and 26 May 1998.
In this linear model, the calibration step was ameliorated using PCA to analyze the surface responses and the emission characteristics. Cumulative variance of the SFSs ranged from 47.0 to 86.2%, which was explained by two factors to allow the determination of excitation and emission frequencies that characterize a unique signature for each fraction [3,36].
The same fractionation protocol and resins, as cited above, was used. Calibration was made of a range of concentration for each fraction. The same post-processing of SFS, composed of slope, area and intensities of fluorescence, was used.
Model expression
$C = [0.06761 + 0.00006 \times IA_{HPOA}] + [0.01659 + 0.00004 \times IA_{HPOB}] + [0.20198 + 0.00016 \times IA_{HPON}] + [0.42156 + 0.00042 \times IA_{HPIA}] + [0.06034 + 0.00003 \times IA_{HPIB}] + [0.11550 + 0.00072 \times IA_{HPIN}]$ where, IA = intensity $\times$ area values from E_x 225–525 emission spectra, (relative intensity units determined by chemical fractionation)
SFS–PLS [21]
Calibrated using samples from CR and RM sampled 9 October 2001.
In this multilinear regression model, no SFS–post-processing was necessary.
Each combination of $E_x - E_m$ was considered as one unique variable. The matrix size was of (13×1951).
Fractionation was performed using a different protocol combining three columns of DAX-8 for HPOs, one AG-MP-50 for HPIA and one Diaion WA-10 providing a DOC recovery of 105%.

- Hydrophilic base (HPIB), amphoteric proteinaceous materials containing amino acids, amino sugars, peptides and proteins.
- Hydrophilic acid (HPIA), organic compound containing the hydroxyl acid group.
- Hydrophilic neutral (HPIN), an organic compound made up of polysaccharides.
- Hydrophobic base (HPOB), the portion of the humic substance retained by DAX-8 resin at normal pH (∼7), which can be eluted by hydrochloride acid.
- Hydrophobic acid (HPOA), a soil fulvic acid.
- Hydrophobic neutral (HPON), a mix of hydrocarbon and carbonyl compounds.

Based on modified versions of Leenheer's fractionation protocol [23], different models have been developed to characterize the DOM and predict the DOC in water samples of low DOM content (<5 mg/l). In this work, three models, the conditions of which are presented in Table 1, were tested. They were designed as given further.

- SFS–GLR, a landscape and contour model [17], based on a post-processing involving a combination of spectral characteristics such as the intensity of fluorescence, the slope and the area under the major peak of fluorescence of a given fraction of DOC in a general linear regression (GLR).
- SFS–PCA, a landscape and contour model [36], in which combination of score and loading factors are used to develop linear models for each fraction of the DOC.
- SFS–PLS, in which each combination of excitation–emission wavelength is considered as a variable, thus correlations were established between two matrices of identical length, one for the SFS (12×1950) and the other for the DOC (12×1) [21].

  Models' performances were compared and commented in relation with the DOM fractionation procedure and recommendations for future research were suggested.

2. Materials and methods

2.1. Samples

Water samples were collected in April during 1998 and 2002 along Passaic Valley Water Commission (PVWC) water treatment plant WTP. Located in Little Falls, NJ, PVWC WTP draws water from the Passaic River and utilizes conventional treatment that includes coagulation–sedimentation, dual media filtration with intermediate and post chlorination. Samples were filtered through a 0.45 μm cellulose filter prior to analysis and fractionation to remove suspended particles, and fractionated into the six components of the DOM, hydrophilics and hydrophobics acid, base and neutral, which were used to build correlations between SFS and DOC. For the laboratory analysis, the samples were taken in 5 l wide-mouthed polyethylene jars and kept cool in dark. All analyses were made in the 14 days after sampling. For the fractionation, samples were collected in 10 l wide-mouthed polyethylene jars and treated as mentioned above.

2.2. Analytical methods

Two protocols of DOM fractionation derived from the one of Leenheer (1981) but using different sets of resins, as seen in Table 1, were carried out. Amberlite resin DAX-8,

a macroporous methylmethacrylate copolymer (Supelco, Bellafonte, PA), AG-MP-50, a strong acid, sulfonated, polystyrene macroporous resin (Biorad, Hercules, CA), Duolite A7, a weak base, phenol–formaldehyde condensation macroporous resin (Supelco, Bellafonte, PA), and Diaion WA 10, a weak anion exchange resins, were all purified by solvent extraction prior to being used in the process. This step is important to eliminate any impurities that may have occurred during the manufacturing process.

As a result, of the fractionation protocol, six fractions of the DOM were isolated based on chemical characteristics. They were termed operationally as hydrophilic and hydrophobic acids (HPIA and HPOA, respectively), base (HPIB and HPOB, respectively) and neutral (HPIN and HPON, respectively), preserved in the applicable eluting hydrochloric acid or sodium hydroxide, and kept refrigerated in quality-assured amber glass bottles (I-Chem from 70 to 250 ml).

Initial DOM fraction concentrations were determined using an O.I. Analytical 700 system (O.I. Corp., College Station, TX) total organic carbon analyzer using the method of sodium persulfate oxidation [37]. Samples of the DOM fractions were volumetrically diluted using Class A glassware to obtain the calibration standards used in this study. The instrument error was controlled within 4% with runs of 5 mg/l standards after every five samples and sample precision of three repeats was controlled within 5%. All fraction samples were appropriately pH adjusted and diluted to reduce the contribution of eluant chemicals to DOC. Milli-Q (Millipore Corp, Bedford, MA) was used for all dilutions, sample preparation, and final glassware washing. All sample glassware was oven dried at a temperature of 500 °C.

The Hitachi F4500 fluorescence spectrophotometer (Tokyo, Japan) equipped with 150 W ozone free Xenon lamp was used for the fluorescence measurements. The samples were recorded in a 1 cm quartz cuvette of 4 ml volume sample size and excited from 225 to 525 nm wavelengths in the backward mode to minimize high-energy molecular damage. The SFS matrix consisted of chemical fluorescence intensity responses across the entire usable fluorescent range between emission (E_m) wavelengths 249–633 nm and excitation (E_x) wavelengths from 225 to 525 nm. Both the emission and excitation slits were 12 nm for SFS–PCA and SFS–GLR models and of 6 nm for the SFS–PLS model. A value of zero has been applied to the data matrix where no emission intensity was observed to avoid case wise deletion of the signature at those frequency intervals.

2.3. Models

The SFS–GLR and SFS–PCA models were built using the peak of maximum intensity of fluorescence of sample of known concentrations for each fraction. A post-processing included the slope, the area, and the intensity for each maximum emission peak of the fractions at different concentration. At each E_m emission spectrum, the starting intensity (i.e. at $E_m = E_x + 12$ nm) and the maximum intensity (i.e. spectrum peak) is determined. The rising slope for each E_x emission spectrum is then calculated as:

$$\text{slope} = \frac{[P - P_i]}{[E_{m(p)} - E_{m(i)}]} \quad (1)$$

P = maximum relative intensity (relative intensity units); P_i = relative intensity at $E_{m(i)}$ (relative intensity units); $E_{m(i)}$ = starting emission wavelength of spectrum = $E_x + 12$ (nm); $E_{m(p)}$ = emission wavelength at maximum relative intensity (nm).

The area under each E_x emission spectrum (area relative intensity units, nm) was determined as the area from $E_m = (E_x + 12)$ to 633 nm. The average slope and the area of four spectra were then calculated for use in statistical post-processing analysis. Details on the SFS post-processing are available in Marhaba and Pu [17]. While the SFS–GLR uses contour factors, such as slope and area; the SFS–PCA introduces the shape factor with the product area by slope.

In the SFS–PLS model, no particular post-processing was necessary. PLS used the matrix variance to decompose the SFSs and calculate a model within the error limits. The working database was made of raw SFSs, which corresponded to 1950 combinations of $\lambda_x - \lambda_m$ for each sample (or SFS). The matrix was then transposed in order to have each sample defined as an object (row) and each of the 1950 wavelength combinations $\lambda_x - \lambda_m$ defined as a variable (column). The strategy adopted is fully described in Marhaba et al. [21]. Six models were obtained, in which hydrophobics and acidic compounds were presenting good robustness and fitness after full cross-validation. Bias were low, between 10^{-4} and 10^{-2} mg/l, while the root mean square error of validation (RMSEV) went from 10^{-3} to 0.13 for HPIN.

3. Results and discussion

The models were used to predict the concentrations of the fractions at five sampling events of PVWC WTP: the intake, the sedimentation basin, the filter influent and effluent, and the finished effluent. As seen in Table 2, and for each column corresponding to a given model, DOM composition is not dramatically different from 1 year to another. Generally, neutral fractions represent 52%, basics 10%, and acids, 38% of the DOM, or hydrophilics 38% and hydrophobics 62% of the DOM. Specifically, in 2002, a slightly different order of composition was noticed with less neutrals and acids but much more basics compounds 40%, or 42% of hydrophilics and 58% of hydrophobics compounds. During the covered period, New Jersey has been submitted to a drought in 1998, a flooding in 1999, and an extended drought period from October 2001 to October 2002.

Overall, the efficiency of PVWC WTP in removing DOM varied from a fraction to another. The average removal of each fraction from the intake to the finished step was found to be 45% for HPIB, HPIN and HPON, 65% for HPOB, and over 80% for HPIA and HPOA. During droughts, less basic and acid, but more neutral compounds were efficiently removed. In fact, the fraction composition might be intrinsically different from a period to another [38]. Therefore, one important axis of research would be, to first investigate the more resistant fraction, and second the disinfection by-product formation potential per fraction in order to further eliminate the one responsible of DBPs.

The predicted DOC (mg/l) values, varied from a model to another, which means that models' performances strongly depended upon the post-processing, and the DOM fractionation technique.

Table 2
Application of the SFS–PCA and SFS–PLS models to characterize and predict the DOC in mg/l through the mass balance concept to PVWC WTP samples in April and between 1998 and 2002

PVWC	HPIB				HPOB				HPIA				HPOA				HPIN				HPON			
Events	PCA	PLS	GLR	M	PCA	PLS	GLR	M	PCA	PLS	GLR	M	PCA	PLS	GLR	M	PCA	PLS	GLR	M	PCA	PLS	GLR	M
1998																								
Intake	0.25	0.18	0.24	0.2	0.25	0.18	0.21	0.18	0.59	0.46	0.55	0.26	1.09	1.12	0.88	1.08	1.13	0.99	1.1	0.96	0.91	0.8	0.93	0.98
Sedimentation basin	0.21	0.14	0.19	0.13	0.21	0.12	0.18	0.1	0.55	0.4	0.54	0.25	1.07	0.79	0.99	1.26	1.05	0.98	1.15	0.57	0.88	0.51	0.81	0.65
Filter influent	0.19	0.13	0.15	0.11	0.21	0.11	0.15	0.14	0.51	0.27	0.51	0.27	1.03	0.74	0.81	1.29	1.1	0.99	0.88	0.27	0.85	0.37	0.79	0.55
Filter effluent	0.17	0.1	0.13	0.11	0.17	0.1	0.15	0.13	0.48	0.27	0.49	0.27	1	0.65	0.75	1.18	1.07	0.99	0.71	0.27	0.78	0.33	0.71	0.48
Finished	0.16	0.1	0.11	0.1	0.15	0.1	0.12	0.11	0.44	0.26	0.47	0.26	0.95	0.55	0.71	0.95	1.04	0.98	0.51	0.26	0.71	0.29	0.66	0.4
1999																								
Intake	0.23	0.21	0.25	0.25	0.21	0.18	0.19	0.18	0.6	0.46	0.57	0.26	1.01	0.99	0.89	1.06	1.13	0.99	1.04	0.92	1.15	0.78	1	0.88
Sedimentation basin	0.21	0.19	0.27	0.1	0.19	0.12	0.17	0.1	0.58	0.31	0.53	0.25	0.91	0.93	1.2	1.26	1.11	0.98	0.98	0.88	0.97	0.44	0.91	0.35
Filter influent	0.19	0.15	0.14	0.11	0.18	0.15	0.15	0.14	0.59	0.28	0.56	0.27	0.88	0.88	1.18	1.29	1.05	0.98	1	0.49	0.91	0.33	0.77	0.5
Filter effluent	0.17	0.12	0.17	0.13	0.17	0.16	0.25	0.2	0.55	0.27	0.55	0.27	0.71	0.69	1.23	1.18	1.05	0.97	0.98	0.37	0.79	0.27	0.66	0.28
Finished	0.16	0.1	0.12	0.11	0.15	0.11	0.12	0.11	0.54	0.26	0.54	0.27	0.68	0.51	1.15	1	1.05	0.95	0.95	0.36	0.75	0.23	0.49	0.31
2000																								
Intake	0.25	0.18	0.23	0.21	0.25	0.19	0.26	0.21	0.57	0.41	0.53	0.33	1.01	1.01	0.99	1	1.03	0.88	0.99	0.91	1.2	0.99	1.15	0.96
Sedimentation basin	0.22	0.17	0.2	0.18	0.23	0.17	0.24	0.17	0.55	0.3	0.49	0.29	0.85	0.77	1.03	0.98	0.99	0.98	0.91	0.69	1.05	0.6	1.01	0.59
Filter influent	0.19	0.11	0.15	0.11	0.21	0.1	0.21	0.14	0.57	0.27	0.53	0.27	0.71	0.71	0.97	0.93	1	0.97	0.89	0.67	0.99	0.55	0.89	0.55
Filter effluent	0.17	0.1	0.16	0.13	0.17	0.1	0.19	0.12	0.55	0.26	0.5	0.27	0.55	0.63	0.95	0.91	0.97	0.96	0.79	0.47	0.93	0.5	0.85	0.52
Finished	0.16	0.11	0.13	0.11	0.16	0.1	0.16	0.1	0.53	0.25	0.49	0.26	0.56	0.53	0.91	0.89	0.96	0.96	0.79	0.46	0.94	0.5	0.77	0.51
2001																								
Intake	0.29	0.26	0.27	0.27	0.28	0.19	0.25	0.18	0.58	0.42	0.5	0.38	0.85	1.12	0.98	1.1	1.13	0.99	1.07	0.98	1.15	0.93	1.05	0.98
Sedimentation basin	0.25	0.21	0.2	0.21	0.2	0.13	0.21	0.15	0.56	0.31	0.47	0.31	0.71	0.99	0.99	1.03	1.07	0.97	1	0.91	0.95	0.44	1	0.4
Filter influent	0.19	0.14	0.17	0.11	0.21	0.12	0.18	0.14	0.59	0.27	0.48	0.27	0.55	0.91	0.88	1.29	0.98	0.97	0.99	0.56	0.91	0.41	0.93	0.39
Filter effluent	0.17	0.12	0.14	0.13	0.17	0.11	0.19	0.2	0.57	0.26	0.45	0.27	0.51	0.71	0.81	1.18	0.91	0.96	0.93	0.36	0.79	0.39	0.9	0.35
Finished	0.16	0.13	0.11	0.1	0.17	0.1	0.18	0.2	0.55	0.25	0.44	0.26	0.49	0.51	0.8	0.85	0.87	0.95	0.8	0.36	0.73	0.28	0.87	0.32
2002																								
Intake	0.31	0.25	0.33	0.28	0.28	0.27	0.28	0.25	0.61	0.46	0.58	0.41	0.97	1.15	1	1.18	1.15	1.03	1.1	1.01	1.23	0.84	1.2	0.9
Sedimentation basin	0.22	0.19	0.25	0.23	0.21	0.19	0.2	0.2	0.58	0.31	0.55	0.33	0.91	1.11	0.98	1.12	1.11	1.01	1.09	1	0.95	0.53	1	0.55
Filter influent	0.19	0.17	0.17	0.18	0.21	0.18	0.2	0.17	0.6	0.28	0.57	0.29	0.89	1.03	0.96	1.05	1.05	0.84	0.94	0.88	0.71	0.51	0.95	0.49
Filter effluent	0.17	0.15	0.17	0.15	0.17	0.15	0.17	0.13	0.57	0.27	0.55	0.27	0.79	1	0.9	1.01	1.05	0.95	0.9	0.8	0.69	0.36	0.85	0.45
Finished	0.16	0.11	0.16	0.11	0.17	0.1	0.14	0.1	0.55	0.22	0.54	0.21	0.77	0.98	0.9	0.99	0.98	0.81	0.83	0.7	0.55	0.38	0.8	0.44

PCA: predicted DOC with model SFS–PCA; PLS: predicted DOC with model SFS–PLS; GLR: predicted DOC with model SFS–GLR; M: measured DOC (mg/l).

The predicted concentrations obtained through simple linear models were exaggerated compared to those resulting from the PLS, which shows that the partial use of SFS do not translate the reality of the DOM composition at different periods, even by incorporating a shape factor such as in SFS–PCA model. The landscape and contour post-processing technique weaken in the choice of the maximum energy of emission–energy of excitation–intensity of fluorescence corresponding to a given fraction. It was obvious at this time that part of the information was missed. This problem was circumvented in SFS–PLS model by using the entire spectra for each fraction of DOM of a given water sample, therefore assimilating each combination E_x–E_m–I as one unique variable.

Chemically, SFS–GLR and SFS–PCA models were obtained using the same fractionation protocol, in which a sequence of XAD-8, AG-MP-50 and Duolite A-7 was used; while SFS–PLS was built from a different sequence of resins: XAD-8, AG-MP-50 and WA-10. One of the major difference between the protocols is the use of same column packed with the appropriate amount of XAD-8 resin to separate the hydrophobic fractions in the case of SFS–GLR and SFS–PCA, while a triple XAD-8 resin columns was set-up for the SFS–PLS model, each column removing only one hydrophobic fraction. HPON was the first fraction to be separated by gravity through XAD-8 resin at pH 7 ± 0.2 at a flow rate <12 bed volumes per hour. After a de-protonation of the second column of resin XAD-8 to pH 10 with NaOH, elution of HPOB was conducted with 0.25 bed volume of 0.1N HCl, followed by 1.5 bed volume of 0.01N HCl at a flow rate less than two bed volume per hour. Finally, elution of HPOA was conducted on a third column of XAD-8.

The set-up of three columns of XAD-8 independently for HPON, HPOB, and HPOA was based on the fact that XAD-8, resin used to separate hydrophobics from hydrophilics, does not discriminate the adsorption between HPON and HPOB. Qualls and Haines [39] noted that the HPON analysis, based on difference of DOM, could be contaminated by some HPOB. Further, it was noted that the adsorbates on XAD-8 under natural pH contain both fractions, HPON and HPOB, which consequently make their measurement as a DOM difference after each adsorption run inappropriate since the mass decrement is the summation of them [40]. All the sample used herein had a DOC below 5 mg/l, therefore, the suggestion to test DOM directly on both HPOA and HPOB fractions to quantify all hydrophobic fractions [22], could not be used without overcoming the inevitable noise from instrument and eluants. Another reason of independently fractionating the hydrophobics is the slight wash out of the precedent fractions noted in HPOA fractions. This wash out might be due to the intermediate polarity showed by some organics [38], probably the neutrals. Overall, using three columns of XAD-8 resins might not be practically convenient, but it guaranteed a clear-cut in terms of DOC among the fractions, while maintaining the backgrounds errors as low as possible.

The fractionation protocols that leaded to SFS–GLR and SFS–PCA models included the Duolite A-7 resin, which was used for HPIA. This resin was replaced with WA-10 in the SFS–PLS model with a service flow and elution rates of 8 and 4 bed volumes/h, respectively.

Duolite A-7, the bleeding [39] of which might explain the higher values obtained for HPIA and HPIN for GLR- and PCA-based models, was replaced by WA-10 resin in the PLS model. WA-10 is a physically and chemically stable weak anionic resin, less hydrophilic than Duolite A-7. As a result, HPIA predictions obtained with the SFS–PLS model showed

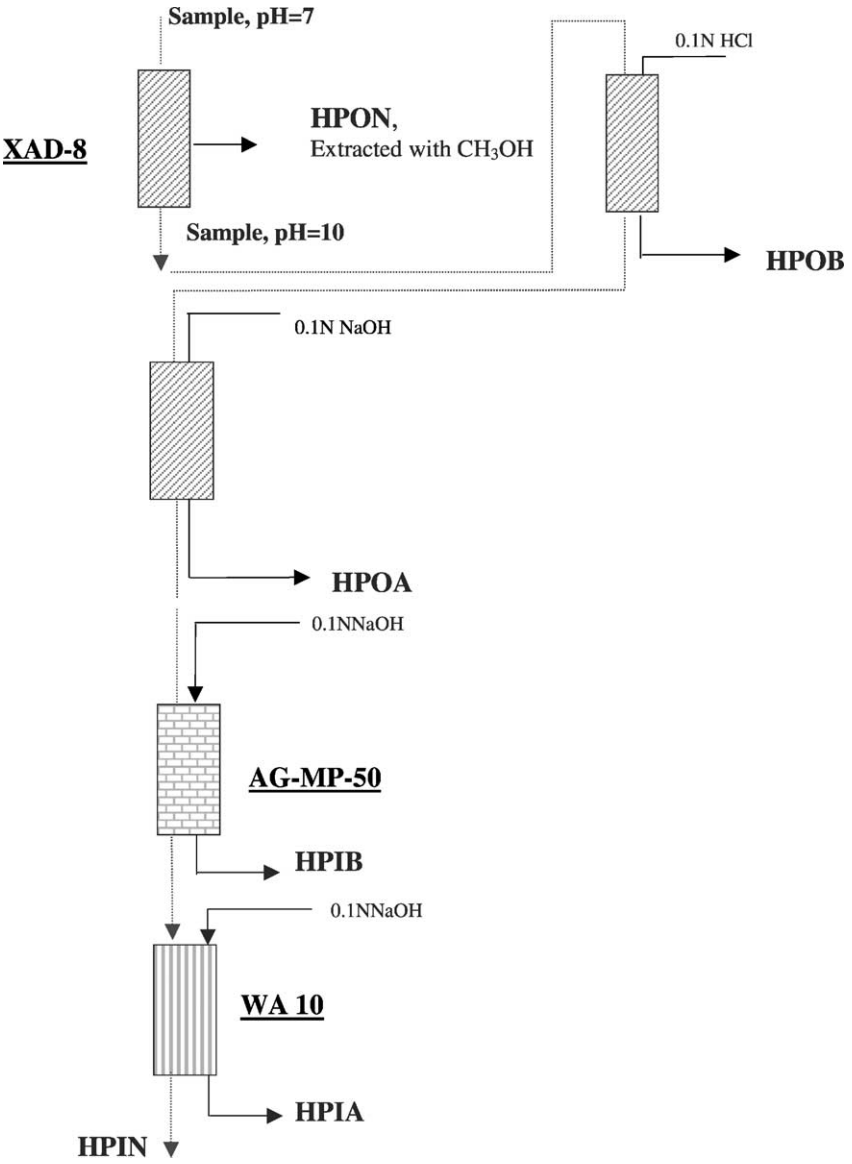

Fig. 1. Fractionation protocol followed to build the model SFS–PLS.

lower biases and better accuracies, yet much higher than the normally expected. Results could be ameliorated if the capacity (meq/g wet) and the amount of WA-10 used, which was determined with the same formula as in Leenheer [23], and multiplied by a safety factor of 1.5 at the least, were more accurate.

The modifications introduced to the DOM fractionation protocol presented in Fig. 1, and at the base of which the SFS–PLS model was calibrated, explain the better prediction

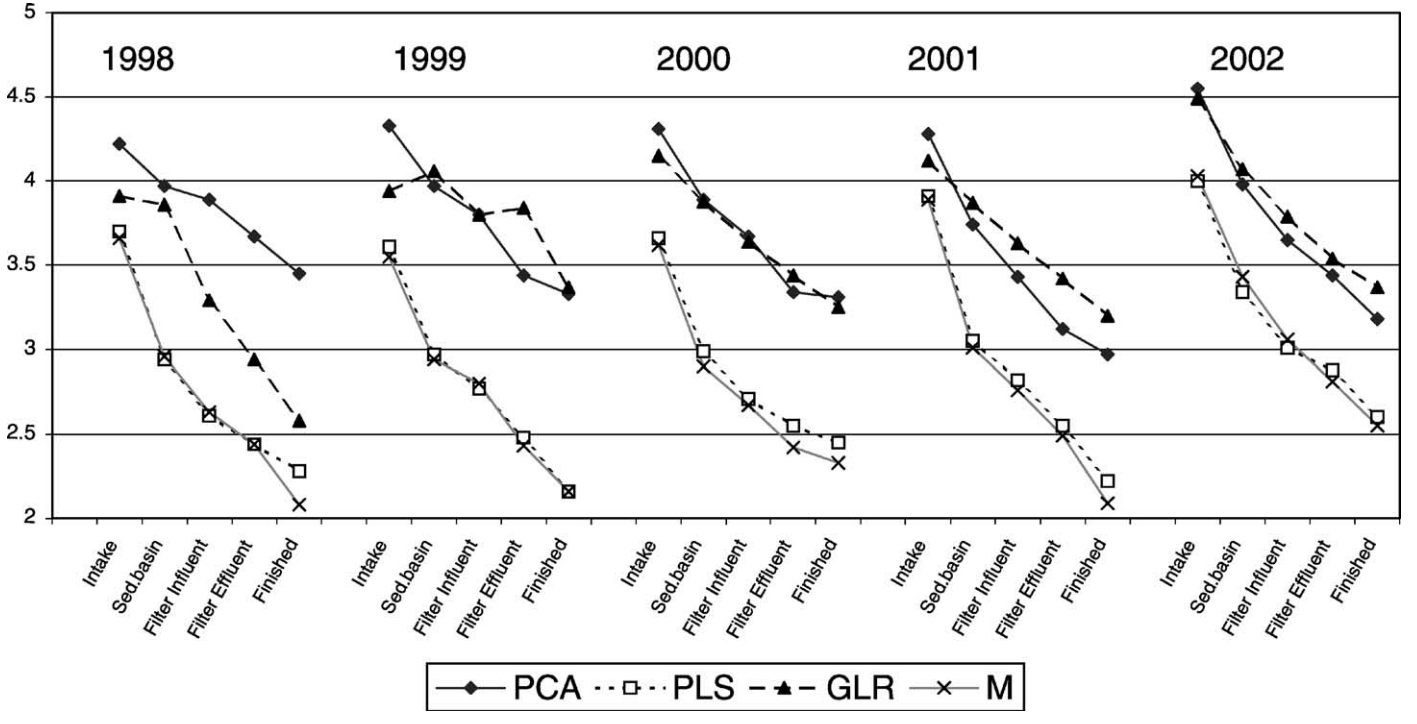

Fig. 2. Comparison of the predicted DOC (mg/l) values obtained through mass balance with the measured DOC values in water samples collected in April between 1998 and 2002 at Passaic Valley Water Commission water treatment plant, NJ.

Table 3
Statistical analysis

	Y			e			S_e			S_Y			Bias			Accuracy		
	GLR	PCA	PLS	GLR	PCA	PLS	GLR	PCA	PLS	GLR	PCA	PLS	GLR	PCA	PLS	GLR	PCA	PLS
HPOA	0.954	0.818	0.84	−0.128	−0.264	−0.242	0.21	0.326	0.324	0.133	0.186	0.209	1.576	−0.323	−0.288	−0.135	1.758	1.549
HPOB	0.19	0.199	0.14	0.036	0.045	−0.014	0.047	0.058	0.034	0.042	0.037	0.043	0.189	0.227	−0.1	1.097	1.588	0.79
HPON	0.878	0.899	0.502	0.327	0.348	−0.049	0.373	0.386	0.09	0.158	0.171	0.212	0.372	0.387	−0.097	2.363	2.263	0.429
HPIA	0.519	0.558	0.311	0.237	0.276	0.029	0.248	0.286	0.07	0.039	0.038	0.074	0.457	0.495	0.093	6.33	7.47	0.936
HPIB	0.018	0.202	0.153	0.018	0.035	−0.136	0.045	0.049	0.041	0.057	0.042	0.046	0.097	0.175	−0.089	0.789	1.174	0.886
HPIN	0.933	1.041	0.961	0.288	0.397	0.327	0.358	0.468	0.422	0.142	0.071	0.049	0.309	0.381	0.33	2.523	6.595	8.589
HPI	1.636	1.801	1.452	0.556	0.72	0.344	0.608	0.775	0.437	0.211	0.122	0.139	0.34	0.4	0.242	2.889	6.336	3.153
HPO	2.021	1.916	1.484	0.234	0.128	−0.304	0.344	0.204	0.395	0.238	0.314	0.417	0.115	0.067	−0.204	1.446	0.65	0.948
All data	0.61	0.62	0.485	0.132	0.141	0.007	0.247	0.301	0.219	0.318	0.31	0.324	0.215	0.228	0.014	0.775	0.97	0.676
DOC	3.658	3.717	2.909	0.79	0.849	0.041	0.86	0.905	0.076	0.423	0.42	0.529	0.216	0.224	0.014	2.03	2.155	0.145
Intake	0.687	0.772	0.630	0.062	0.098	0.005	0.143	0.176	0.074	0.364	0.384	0.367	0.09	0.135	0.008	0.394	0.458	0.202
Sedimentation basin	0.658	0.652	0.509	0.15	0.143	0.0016	0.257	0.28	0.156	0.372	0.355	0.344	0.228	0.22	0.0032	0.68	0.79	0.453
Filter Inf.	0.605	0.614	0.464	0.141	0.15	0	0.286	0.344	0.245	0.325	0.338	0.34	0.233	0.245	0	0.813	1.01	0.725
Filter Eff	0.572	0.567	0.43	0.153	0.147	0.01	0.288	0.36	0.285	0.33	0.328	0.329	0.267	0.26	0.024	0.873	1.095	0.868
Finished	0.525	0.541	0.39	0.152	0.167	0.016	0.245	0.328	0.272	0.322	0.319	0.314	0.289	0.31	0.042	0.759	1.027	0.867
April-98	0.553	0.64	0.455	0.113	0.2	0.015	0.291	0.342	0.316	0.325	0.371	0.35	0.204	0.313	0.033	0.895	0.92	0.903
April-99	0.634	0.629	0.466	0.172	0.167	0.005	0.266	0.356	0.266	0.385	0.362	0.341	0.271	0.266	0.01	0.69	0.984	0.779
April-00	0.612	0.629	0.466	0.147	0.164	0.002	0.195	0.264	0.199	0.345	0.362	0.341	0.24	0.261	0.004	0.565	0.729	0.582
April-01	0.608	0.584	0.485	0.133	0.11	0.01	0.292	0.345	0.219	0.351	0.323	0.354	0.219	0.188	0.021	0.832	1.066	0.619
April-02	0.642	0.627	0.528	0.113	0.097	−0.002	0.21	0.205	0.046	0.354	0.351	0.368	0.175	0.155	−0.003	0.592	0.584	0.125

Y: mean concentration (mg/l); e: mean error; S_e: standard mean error; S_Y: concentration standard error.

obtained. However, concerns still exist, especially for HPIN fraction, for which the predicted values are constant along the WTP train, and HPIA fractions that were still exaggerated.

Based on the mass balance concept, theoretically, the sum of all predicted values must be comparable to the measured original water sample DOC (mg/l) as follow:

$$\sum_{i=1}^{6}[\mathrm{C}]_{\text{predicted fractions}} = \mathrm{DOC}_{\text{predicted}} \tag{2}$$

When comparing the sum of the predicted fractions' (Eq. (2)) DOC values to the measured value of the original pre-fractionated sample, the landscape and contour models were presenting a gap, as shown in Fig. 2. This gap reflected the differences in terms of fractionation and post-processing exploitation of the SFS. The recovery of the DOC was abnormally high for the fractionation protocol at the base of the landscape and contour models but was minimal (105%) for the fractionation protocol used to build the SFS–PLS model. Further, the statistical analysis presented in Table 3 confirmed that SFS–PLS model offer best bias and accuracy in almost all cases: by fraction, bias varied from 8.9% (HPIB) to 0.33% (HPIN) and accuracy from 42.9 (HPON) to 858% (HPIN); by group of fractions: hydrophilics and hydrophobics, bias was below 25% while accuracy was higher for the hydrophilics; by type of model independently of the year or the sampling event, where bias was 1.4% and accuracy 67%; by DOC-mass balance bias in this case was 1.4% while accuracy reached 14.5%; by sampling WTP train locations, maximum bias was 2.4% at the filter effluent sampling point while accuracy was minimal at the delivery point (86.8%); and by year of sampling, where bias was below 3.3%.

The results obtained with the SFS–PLS methodology are still somewhat limited by the fractionation protocol used, especially the portion of the protocol dealing with the separation of the hydrophilics acid and neutral. HPIN fraction had the maximal bias (33%) and accuracy (858%). Therefore, future models must be build on a DOM fractionation sequence that would keep the sequence used for the SFS–PLS model, which included a set of three columns of XAD-8 resins to separate the hydrophobics, but a better knowledge of the adsorptive capacity of the WA-10 weak base anion exchange resin used to separate HPIA and HPIN is needed.

4. Conclusion

The main objective of this study was to compare different models built with isolate enriched fractions of DOM of a surface water WTP, which are responsible for the chlorine demand and disinfection by-products formation potential. SFS was considered herein as a surrogate parameter to characterize low organic matter content, which induces most problems encountered during treatment and distribution of drinking water, leading to water quality depreciation. DOM fraction concentrations have been determined by using two different isolation–fractionation techniques and confirmed by modeling the spectral fluorescent signatures (SFSs) of the pre-fractionated NOM in the water sample. Three modeling techniques were tested for predicting fraction concentrations at different sampling events of a major New Jersey water treatment plant.

From the comparison of the three models: SFS–GLR, SFS–PCA, and SFS–PLS, on a large set of data compiled during 5 years at the same period, SFS–PLS showed the lowest biases and the better accuracies. Therefore, it is suggested that no SFS post-processing is necessary to reliably characterize the DOC content of water samples.

From the comparison of two sequences of resins, the sequence XAD-8, AG-MP-50 and WA-10, showed better performances in fractionating low DOM content, although some ameliorations are still needed in order to ameliorate the percentage of recovery while keeping the background errors as minimum as possible.

Overall, while SFS–PLS was the model found to be more precise and accurate, future improvement of the calibration through the fractionation protocol is still necessary. Therefore, the equations obtained for the model SFS–PLS represent a first step in developing a generally applicable DOM-based model for DBPs formation control in drinking water treatment and distribution systems. Research is being conducted on the effectiveness of the models obtained by combining more data of different WTPs, understanding and quantifying the inhibition and synergy effects between fractions that might be involved in natural waters, especially between neutral fractions.

Acknowledgements

This research has been funded in part by the New Jersey Department of Environmental Protection A-280 safe drinking water fund. The authors would like to thank Dr. R. Lee Lippincott (New Jersey Department of Environmental Protection), Yong Pu and Jaime Arago (New Jersey Institute of Technology) for their significant contributions to the work.

References

[1] Y.V. Orlov, L.G. Persiantsev, S.P. Rebrick, S.M. Babichenko, J. Soc. Photo-Opt. Instrum. Eng. 2503 (1995) 159.
[2] S.M. Babitchenko, J. Lapimma, L. Poryvkina, V. Varlamov, J. Soc. Photo-opt. Instrum. Eng. 2503 (1995) 157.
[3] T.F. Marhaba, D. Van, R.L. Lippincott, Water Res. 34 (14) (2000) 3543.
[4] United States Environmental Protection Agency. National Primary Drinking Water Regulations: Dinsinfection/Disinfection By-Products (D/DBP) Rule, Federal Register 59, 38668, 1989.
[5] J. Saurina, C. Leal, R. Campano, M. Granados, R. Tauler, M.D. Prat, Anal. Chim. Acta 409 (2000) 237.
[6] L. Norgaard, Talanta 42 (1995) 1305.
[7] S.D. Brown, T.B. Blank, S.T. Sum, L.G. Weyer, Anal. Chem. 66 (1994) 315.
[8] W.F. McClure, Anal. Chem. 66 (1994) 43.
[9] D. Bertrand, C.N.G. Scotter, Appl. Spectrosc. 46 (1992) 1420.
[10] Z. Wang, B.C. Pant, C.H. Langford, Anal. Chim. Acta 232 (1990) 43.
[11] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 53 (1981) 92.
[12] D.S. Burdick, X.M. Tu, L.B. McGown, D.W. Millican, J. Chemom. 4 (1990) 15.
[13] O. Berntsson, L.G. Danielson, M.O. Johansson, S. Folestad, Anal. Chim. Acta 419 (2000) 45.
[14] C.M. Morrow, R.A. Minear, Water Res. 21 (1987) 41.
[15] G.W. Harrington, A. Bruchet, D. Rybacki, P. Singer, Water Disinfection and Natural Organic Matter: Characterization and Control, ACS, Washington, DC, 1992.
[16] W.A. Elshorbagy, H.A. Qdais, M.K. Elsheamy, Water Res. 34 (13) (2000) 3431.
[17] T.F. Marhaba, Y. Pu, J. Hazard. Mater. 73 (2000) 221.

[18] M.J. Rodriguez, J.B. Serodes, Water Res. 35 (6) (2001) 1572.
[19] T. Persson, M. Wedborg, Anal. Chim. Acta 434 (2001) 179.
[20] D. Baunsgaard, L. Muck, L. Nørgaard, Appl. Spectrosc. 54 (3) (2000) 438.
[21] T.F. Marhaba, K. Bengraïne, Y. Pu, J. Arago, J. Hazard. Mater. B97 (2003) 83.
[22] J.A. Leenheer, E.W.D. Huffman, US Geol. Survey 79 (4) (1976) 737.
[23] J.A. Leenheer, Environ. Sci. Technol. 15 (1981) 578.
[24] E.M. Thurman, R.L. Malcolm, Environ. Sci. Technol. 15 (1981) 463.
[25] A. Bruchet, Y. Tsutsumi, J.P. Duguet, J. Mallevialle, Adv. Chem. 214 (1987) 381.
[26] P.J. Shaw, H. De Haan, R.I. Jones, Environ. Sci. Technol. 15 (1994) 753.
[27] G.L. Amy, C.J. Kuo, R.A. Sierke, Ozone Sci. Eng. 10 (1) (1988) 39.
[28] T. Kainulainen, T. Tuhkanen, T. Vatitiainien, H. Heinonen-Tanski, P. Kaliokoski, Water Sci. Technol. 30 (9) (1994) 169.
[29] A. Bruchet, C. Rousseau, J. Mallevialle, J. AWWA 82 (9) (1990) 66.
[30] R.L. Malcolm, Environ. Int. 18 (1992) 609.
[31] G.R. Aiken, J.A. Leenheer, Chem. Ecol. 8 (1993) 135.
[32] J.A. Leenheer, J.P. Croue, M. Benjamin, G.V. Korshin, C.J. Hwang, A. Bruchet, G.R. Aiken, in: S.E. Barrett, S.W. Krasner, G.L. Amy (Eds.), Natural Organic Matter and Disinfection By-Products Characterization and Control in Drinking Water, ACS, Washington, DC, 2000, p. 68.
[33] J.A. Leenheer, E.W.D. Huffman, J. Res. US Geol. Survey 4 (1976) 737.
[34] G.R. Aiken, D.M. McKnight, K.A. Thorn, E.M. Thurman, Org. Geochem. 18 (1992) 567.
[35] R.L. Malcolm, P. McCarthy, Environ. Int. 18 (1992) 597.
[36] T.F. Marhaba, R.L. Lippincott, D. Van, Fresenius J. Anal. Chem. 366 (2000) 22.
[37] Standard Methods for the Examination of Water and Wastewater, 19th edn., American Public Health Association, American Water Works Association, Water Pollution Control Federation, USA, 1995.
[38] J.P. Croué, J.F. Debrous, G.L. Amy, G.R. Aiken, J.A. Leenheer, in: D. Philip, C. Singer, (Eds.), Formation and Control of Disinfection By-Products in Drinking Water, AWWA, 1999, p. 65.
[39] R.G. Qualls, B.L. Haines, J. Soil Sci. Soc. Am. 55 (1991) 1112.
[40] B. Gasparovic, V. Vojvodic, B. Cosovic, Anal. Chim. Acta 338 (1997) 179.