

## Removal of NOM-constituents as characterized by LC-OCD and F-EEM during drinking water treatment

S. A. Bagtho, S. K. Sharma, M. Guitard, V. Heim, J.-P. Croué and G. L. Amy

### ABSTRACT

Natural organic matter (NOM) is of concern in drinking water because it causes adverse aesthetic qualities such as taste, odour, and colour; impedes the performance of treatment processes; and decreases the effectiveness of oxidants and disinfectants while contributing to undesirable disinfection by-products. The effective removal of NOM during drinking water treatment requires a good understanding of its character. Because of its heterogeneity, NOM characterization necessitates the use of multiple analytical techniques. In this study, NOM in water samples from two drinking water treatment trains was characterized using liquid chromatography with organic carbon detection (LC-OCD), and fluorescence excitation-emission matrices (F-EEMs) with parallel factor analysis (PARAFAC). These characterization methods indicate that the raw and treated waters are dominated by humic substances. The results show that whereas the coagulation process for both plants may be optimized for the removal of bulk DOC, it is not likewise optimized for the removal of specific NOM fractions. A five component PARAFAC model was developed for the F-EEMs, three of which are humic-like, while two are protein-like. These PARAFAC components and the LC-OCD fractions represented effective tools for the performance evaluation of the two water treatment plants in terms of the removal of NOM fractions.

**Key words** | characterization, drinking water treatment, natural organic matter

**S. A. Bagtho** (corresponding author)

**S. K. Sharma**

**G. L. Amy**

UNESCO-IHE Institute for Water Education,  
Westvest 7, AX 2611 Delft,  
The Netherlands  
E-mail: [s.bagtho@unesco-ihe.org](mailto:s.bagtho@unesco-ihe.org)

**S. A. Bagtho**

**G. L. Amy**

Faculty of Civil Engineering and Geosciences,  
Delft University of Technology,  
GA 2600 Delft,  
The Netherlands

**M. Guitard**

Veolia Eau Banlieue de Paris,  
Direction technique,  
6 esplanade Charles de Gaulle,  
92751 Nanterre Cedex,  
France

**V. Heim**

Syndicat des Eaux d'Ile de France (SEDIF),  
14 rue Saint Benoît, Paris 75006,  
France

**J.-P. Croué**

**G. L. Amy**

King Abdallah University of Science and  
Technology, Thuwal,  
Saudi Arabia

### INTRODUCTION

Natural organic matter (NOM) is a heterogeneous mixture of compounds found abundantly in natural waters. NOM originates from living and dead plants, animals and microorganisms, and from the degradation products of these sources (Chow *et al.* 1999). Because of its heterogeneity and complexity, it is not practical to characterize NOM in terms of all of its constituents; rather, it is commonly characterized into groups of compounds with similar physicochemical properties. NOM significantly affects water treatment processes such as coagulation, oxidation, adsorption, and membrane filtration. It affects drinking water quality by contributing to formation of potentially carcinogenic disinfection by-products (DBPs) (Sharp *et al.* 2004), by

promoting biological regrowth in the water distribution system and by contributing to colour, tastes and odours. The extent to which NOM affects water treatment processes depends on its quantity and physicochemical characteristics.

It is now widely accepted that the efficiency of drinking water treatment is greatly influenced by the amount and character of NOM present in water. Consequently, many water treatment utilities monitor NOM in their source waters in order to optimize treatment processes. Typically, this optimization has been obtained using bulk water quality parameters such as dissolved organic carbon (DOC) and ultraviolet absorbance at a wavelength of 254 nm (UVA<sub>254</sub>). Specific ultraviolet absorbance (SUVA) is another bulk parameter

that has been used as a surrogate for NOM composition and reactivity (Weishaar 2003). However, the use of these bulk parameters has limitations. Many waters may have NOM with similar DOC concentrations or  $UV_{254}$  absorptivities but with different characteristics such as molecular weight and reactivity, resulting in different removal efficiencies during treatment. A better understanding of its quantity as well as character is therefore required to improve the performance of treatment processes and to optimize the removal of NOM.

High performance size exclusion chromatography (HPSEC), which separates NOM mainly according to molecular size/weight, has been widely applied in characterization of NOM in aquatic environments (Frimmel 1998; Her et al. 2003; Croué 2004). A liquid chromatography system coupled with organic carbon and organic nitrogen detectors (LC-OCD-OND) (Huber et al. 2011) may be used to fractionate NOM into five fractions: biopolymers (such as polysaccharides, polypeptides, proteins and amino sugars); humic substances (fulvic and humic acids); building blocks (hydrolysates of humic substances); low molecular weight (LMW) humic substances and acids; and low molecular weight neutrals (such as alcohols, aldehydes, ketones and amino acids).

Fluorescence is another property that is frequently used for NOM characterization. The relatively low expense and high sensitivity of fluorescence measurements, coupled with rapid data acquisition of water samples at low natural concentrations, have made fluorescence spectrophotometry using fluorescence excitation emission matrices (F-EEM) attractive for NOM characterization of water samples. This characterization has typically involved the use of excitation-emission wavelength pairs to identify fluorophores based on the location of peaks on F-EEM contour plots (Coble 1996). These peaks have been used to distinguish between humic-like NOM, with longer emission wavelengths ( $>350$  nm), and protein-like NOM, with shorter emission wavelengths ( $\leq 350$  nm). Other methods include: fluorescence regional integration (FRI) (Chen et al. 2003); multivariate data analysis (e.g. principal component analysis, PCA, and partial least squares regression, PLS) (Persson & Wedborg 2001); and multi-way data analysis using parallel factor analysis (PARAFAC) (Stedmon et al. 2003). PARAFAC has been used to decompose F-EEMs into individual components some of which have been attributed to protein-like or humic-like

NOM (Stedmon et al. 2003; Stedmon & Markager 2005a; Hunt & Ohno 2007; Yamashita et al. 2008).

The primary objective of this study was to characterize NOM in water samples taken across two drinking water treatment plants serving the suburbs of Paris. This was carried out in order to improve our understanding of the character of the NOM and its temporal variation in waters treated by the two plants. A secondary objective was to evaluate the performance of the treatment processes in terms of NOM removal. Samples were collected from the two treatment plants and analyzed using bulk water quality parameters as well as LC-OCD and F-EEM with PARAFAC.

## MATERIALS AND METHODS

### Sampling

Water samples were collected from two drinking water treatment plants of Syndicat des Eaux d'Ile de France (SEDIF) located in the suburbs of Paris (France) between March 2008 and September 2009. The two plants, Choisy-le-Roi (CR) and Neuilly-sur-Marne (NM), comprise conventional treatment using aluminium sulphate ( $Al_2(SO_4)_3$ ) as coagulant, coupled with biofiltration using ozonation followed by biological activated carbon (BAC) filtration. The following water samples were collected once a month from CR: (i) raw water; (ii) preozonated water; (iii) settled water; (iv) sand filtrated water; (v) ozonated water; (vi) BAC filtrated water; and (vii) product water. The following samples were collected once a quarter from NM: (i) raw water; (ii) settled water; (iii) sand filtrated water; (iv) ozonated water; (v) BAC filtrated water; and (vi) product water. On average, seven samples were collected from CR every month and six samples every quarter from NM. The samples were collected in clean glass bottles and immediately filtered through  $0.45 \mu m$  before being transported, within 24 hours, to the laboratory for analysis. The pre-filtered samples were stored at  $5^\circ C$  until required for analysis, normally within one week of sampling. All the samples were analyzed for DOC,  $UVA_{254}$  and F-EEMs. Selected samples were analyzed using LC-OCD. For CR, samples collected during 10 of the 18 months of sampling were analyzed (67 samples in total), while for NM, samples

collected during 3 of the 7 quarters were analyzed (17 samples in total).

### DOC and UVA<sub>254</sub> measurements

DOC concentrations of all pre-filtered samples were determined by the catalytic combustion method using a Shimadzu TOC-V<sub>CPN</sub> organic carbon analyzer. UVA<sub>254</sub> of each sample was measured at room temperature (20 ± 1 °C) and ambient pH using a Shimadzu UV-2501PC UV-VIS scanning spectrophotometer. SUVA was determined by dividing the UVA<sub>254</sub> by the corresponding DOC concentration.

### Characterization with LC-OCD

Size exclusion chromatography of water samples was performed with a LC-OCD system (DOC-LABOR, Germany) at Het Waterlaboratorium, Haarlem, The Netherlands. An accompanying software (Fiffikus) was used to calculate the DOC and SUVA of the NOM fractions.

### Fluorescence excitation emission matrices (F-EEM)

To account for fluorescence quenching resulting from relatively high concentration of DOC in water samples, absorbance corrections have to be applied; however, these corrections are not necessary if the sample UVA<sub>254</sub> is less than 0.05 cm<sup>-1</sup> (Kubista *et al.* 1994) or if the DOC concentration of the sample is diluted to about 1 mg C/L prior to fluorescence measurement. Since UVA<sub>254</sub> was more than 0.05 cm<sup>-1</sup> for nearly all raw water samples from the two water treatment plants, the prefiltered samples were diluted to a DOC concentration of 1 mg C/L using Milli-Q water prior to fluorescence measurements. Fluorescence intensities for all samples were measured, at ambient pH and at room temperature (20 ± 1 °C), using a FluoroMax-3 spectrofluorometer (Horiba Jobin Yvon). EEMs were generated for each sample by scanning over excitation wavelengths between 240 and 450 nm at intervals of 10 nm and emission wavelengths between 290 and 500 nm at intervals of 2 nm. An EEM of Milli-Q water was obtained and this was subtracted from the EEM of each sample in order to remove most of the water

Raman scatter peaks. Since samples were diluted to a DOC concentration of 1 mg/L prior to measurements, each blank subtracted EEM was multiplied by the respective dilution factor and Raman-normalized by dividing by the integrated area under the Raman scatter peak (excitation wavelength of 350 nm) of the corresponding Milli-Q water, and the fluorescence intensities was reported in Raman units (RU).

### PARAFAC modeling of fluorescence EEM

PARAFAC was used to model the dataset of F-EEMs generated for samples from both treatment plants. It uses an alternating least squares algorithm to minimize the sum of squared residuals in a trilinear model, thus allowing the estimation of the true underlying EEM spectra (Harshman & Lundy 1994; Bro 1997). It reduces a dataset of EEMs into a set of trilinear terms and a residual array (Andersen & Bro 2003):

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + \varepsilon_{ijk},$$

$$i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K$$

where  $x_{ijk}$  is the fluorescence intensity of the  $i$ th sample at the  $j$ th emission and  $k$ th excitation wavelength;  $a_{if}$  represents the concentration of the  $f$ th fluorophore in the  $i$ th sample (defined as scores),  $b_{jf}$  and  $c_{kf}$  are estimates of the emission and excitation spectra respectively for the  $f$ th fluorophore (defined as loadings),  $F$  is the number of fluorophores (components) and  $\varepsilon_{ijk}$  is the residual element, representing the unexplained variation in the model (Stedmon *et al.* 2003).

While component scores indicate the relative concentrations of groups of organic fractions represented by the components, excitation and emission loadings indicate their characteristic excitation and emission spectra (Stedmon *et al.* 2003). The maximum fluorescence intensity for each component obtained from the PARAFAC analysis was used to illustrate the quantitative and qualitative differences between samples.

It is generally difficult to decide the most appropriate number of components of a PARAFAC model. There are several tools that may be used to select the appropriate

number of components but only two were used in this study: the split-half analysis (Harshman 1984), and the examination of residual error plots (Stedmon & Bro 2008). For split-half analysis, the dataset of EEMs was randomly split into two halves and a PARAFAC model obtained for each half. The excitation and emission spectral loadings of the two independent halves were then compared to ascertain whether they were similar.

A total of 179 water samples were collected for this study: 137 from CR and 42 from NM water treatment trains, respectively. A dataset comprising F-EEMs for 145 of these samples was used in the PARAFAC analysis. A series of PARAFAC models consisting of three to seven components were generated using DOMfluor toolbox (Stedmon & Bro 2008), which was specifically developed to perform PARAFAC analysis of DOM fluorescence. It contains all the tools used to perform split-half and residual errors diagnostics.

## RESULTS AND DISCUSSION

### Variation of DOC and SUVA

The mean DOC concentrations and SUVA values of the samples collected from CR and NM treatment process trains are shown in Table 1.

Figure 1 shows the variation of DOC and SUVA for raw and product water samples collected from CR and NM treatment plants. In the case of CR, the DOC concentrations ranged between 2.0 and 4.3 mg L<sup>-1</sup> for raw water, and between 1.0 and 1.9 mg L<sup>-1</sup> for product water. The SUVA values varied between 1.7 and 3.7 L mg<sup>-1</sup> m<sup>-1</sup> for raw water, and between 0.2 and 2.0 L mg<sup>-1</sup> m<sup>-1</sup> for product water. For NM, the DOC concentrations ranged between 2.1 and 3.2 mg L<sup>-1</sup> for raw water, and between 1.1 and 1.6 mg L<sup>-1</sup> for product water. The SUVA values varied between 1.7 and 3.1 L mg<sup>-1</sup> m<sup>-1</sup> for raw water, and between 0.7 and 1.3 L mg<sup>-1</sup> m<sup>-1</sup> for product water. For both cases, there was no clear seasonal pattern in the variation of either DOC concentration or SUVA for the raw water. The treated-water DOC concentrations were fairly stable, indicating that two treatment plants were generally effective in maintaining a relatively constant product water DOC

**Table 1** | Mean value and standard deviations of DOC concentrations and SUVA values for water samples from Choisy-le-Roi and Neuilly-sur-Marne process trains

Water sample	Choisy-le-Roi (n = 20)		Neuilly-sur-Marne (n = 7)	
	DOC (mg L <sup>-1</sup> )	SUVA (L mg <sup>-1</sup> m <sup>-1</sup> )	DOC (mg L <sup>-1</sup> )	SUVA (L mg <sup>-1</sup> m <sup>-1</sup> )
Raw water	2.7 ± 0.6	2.5 ± 0.5	2.5 ± 0.4	2.4 ± 0.5
Preozonated water	2.8 ± 0.6	1.9 ± 0.3		
Settled water	2.1 ± 0.3	1.5 ± 0.4	1.9 ± 0.2	1.8 ± 0.3
Sand filtrated water	1.9 ± 0.4	1.6 ± 0.5	1.6 ± 0.3	1.9 ± 0.4
Ozonated water	1.8 ± 0.3	1.2 ± 0.6	1.6 ± 0.2	1.1 ± 0.3
BAC filtrated water	1.5 ± 0.3	1.2 ± 0.5	1.3 ± 0.2	1.1 ± 0.3
Product water	1.6 ± 0.3	1.1 ± 0.4	1.3 ± 0.2	1.0 ± 0.3

concentration. On the contrary the treated-water SUVA values were not as stable, indicating that the plants were not as effective in maintaining a more uniform NOM character in the treated water.

### Treatment efficiencies in terms of DOC removal and SUVA reduction

For both treatment plants, the operational objective for the removal of NOM is to maintain a TOC concentration of  $\leq 2.0 \pm 0.2$  mg L<sup>-1</sup>, a French treatment goal, as long as the raw water TOC concentration is  $\leq 5.0$  mg L<sup>-1</sup>. Since the TOC and DOC concentrations did not differ by more than 5% for any of the samples analyzed, the treatment efficiencies for the two process trains were assessed in terms of DOC removal and SUVA reduction.

Figure 2 shows the mean DOC removed across each treatment process for both CR and NM process trains, respectively. In both plants, DOC is removed mainly by coagulation-filtration, which removed 0.9 mg L<sup>-1</sup> in CR and 0.8 mg L<sup>-1</sup> in NM, and by BAC filtration, which removed a further 0.3 mg L<sup>-1</sup> in CR and 0.2 mg L<sup>-1</sup> in NM. The results show that whereas the mean DOC concentrations of the raw and product waters for CR were slightly higher than for NM, the difference in the DOC removal efficiencies for the two treatment trains were not statistically significant for any of the processes. For both plants, the maximum DOC concentrations of the raw and

treated waters were not more than 4 and 1.9 mg L<sup>-1</sup>, respectively. While the latter satisfies the treatment objective stated earlier, the fact that it was obtained when the raw water DOC concentration was only 3.1 mg L<sup>-1</sup> implies that it is likely that the objective could be compromised if the raw water DOC concentration approached 5 mg L<sup>-1</sup>, the maximum for which the treatment objective was optimized.

### NOM characterization using LC-OCD

LC-OCD was used to obtain DOC concentrations of the five chromatographic fractions (biopolymers, humic substances, building blocks, LMW acids and LMW neutrals) before and

after each treatment step for water samples from CR and NM. Sampling for LC-OCD analyses was carried out in such a manner as to capture periods of low as well as high algal counts and as such, samples were collected once a month in ten of the twenty and in three of the seven months of sampling for CR and NM, respectively. Figures 3 and 4 show the mean DOC concentrations of LC-OCD fractions of the raw and treated waters, the removals across each treatment process, and the percentage DOC contribution of the fractions in the raw and treated waters for the two process trains, respectively. Humic substances were the dominant fraction in all water samples from both plants, contributing on average to 55% of the DOC. Since the source waters for both plants are river water, it would be

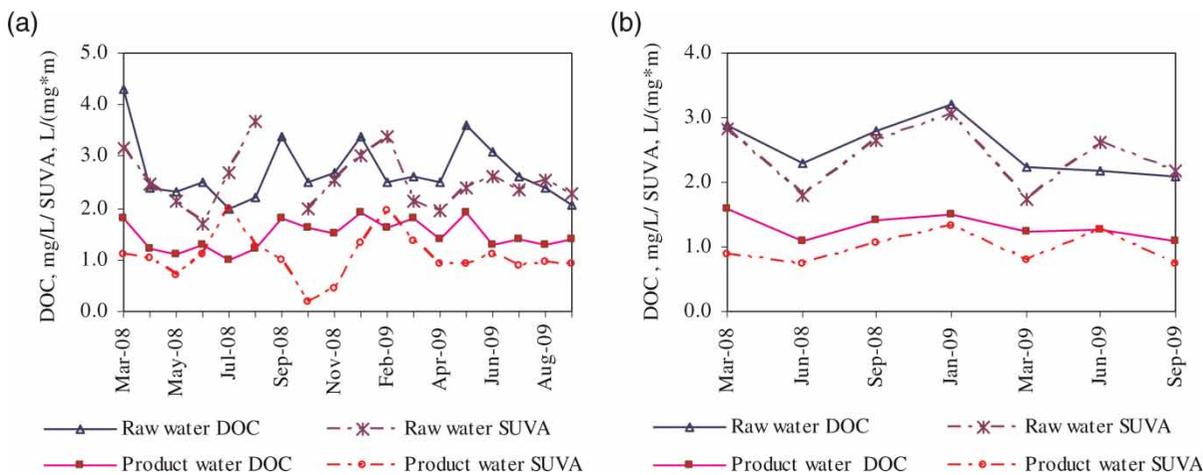


Figure 1 | Variation of DOC concentrations and SUVA values for raw and treated water samples collected from (a) Choisy-le-Roi and (b) Neuilly-sur-Marne treatment plants.

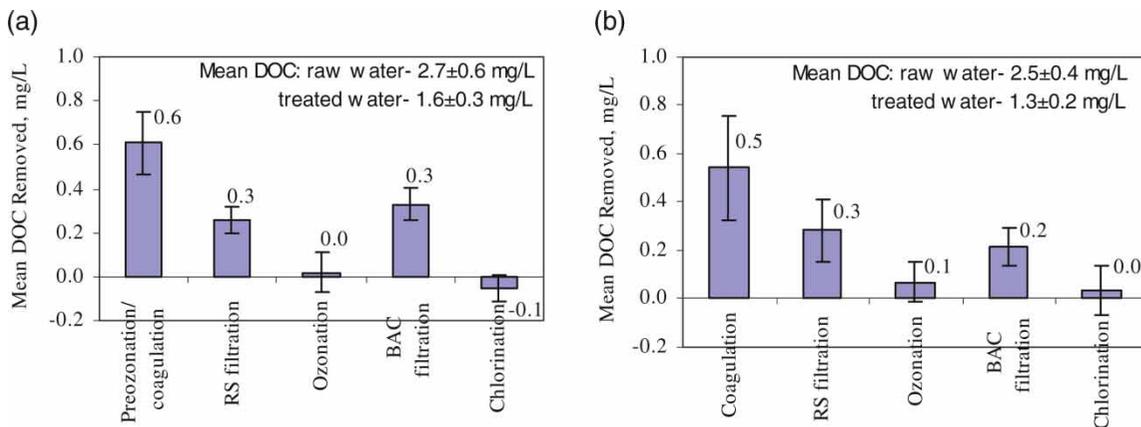
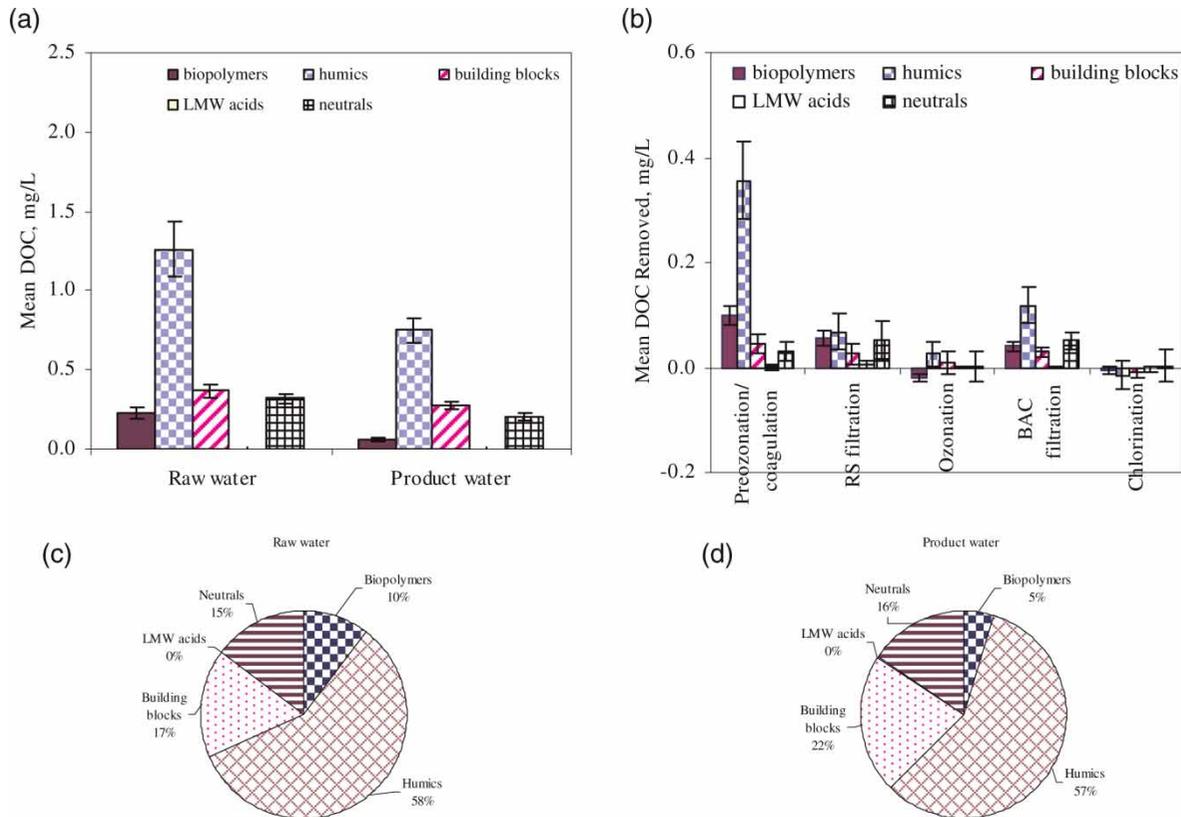


Figure 2 | Mean DOC removed for water samples collected from (a) Choisy-le-Roi and (b) Neuilly-sur-Marne treatment trains. The mean values are based on the complete datasets of samples from Choisy-le-Roi ( $n = 20$ ) and Neuilly-sur-Marne ( $n = 7$ ). The error bars indicate 95% confidence intervals.



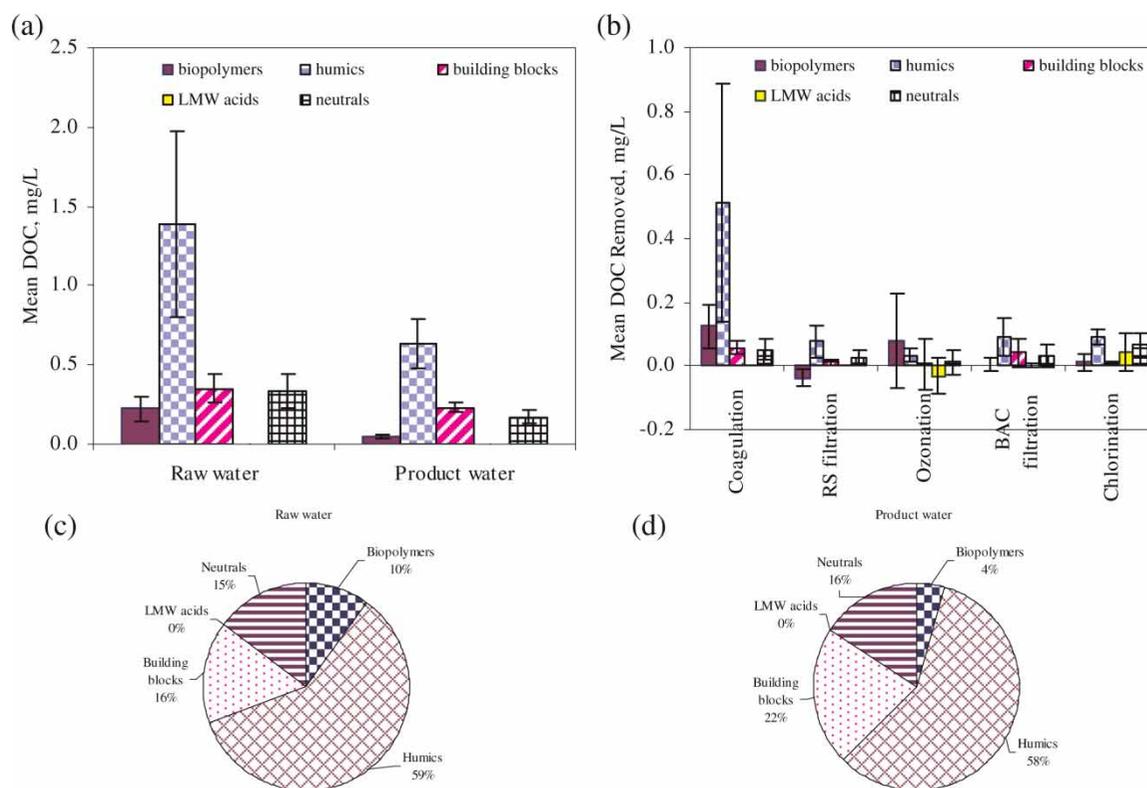
**Figure 3** | (a) Mean DOC concentrations of the five LC-OCD fractions for raw and treated waters, (b) mean DOC removal across each treatment process, and the fractional DOC composition of (c) raw and (d) treated waters for samples collected across Choisy-le-Roi treatment train. The error bars indicate 95% confidence intervals ( $n = 10$ ).

reasonable to expect the NOM composition to be typical of natural waters dominated by terrestrial runoff, in which humic substances (fulvic and humic acids) are 50% of the DOC (Thurman 1985). In both plants, LMW acids were below the detection limit in nearly all samples.

The removal of the LC-OCD fractions occurred mainly by coagulation, followed by BAC filtration. The trend in the change of NOM composition after treatment is similar for both process trains. The large MW fractions were preferentially removed, with the percentage contribution of the biopolymer fraction decreasing by a half, from 10 to ~5% in the raw and treated waters, respectively, while that of humic substances decreasing only slightly, by 1% for both trains. In contrast, there was a relative increase in the LMW fractions, with the building blocks increasing from 17 and 16% in the raw water for CR and NM, respectively, to 22% in the treated water for both plants, and the neutral fractions increasing slightly from 15 to 16% for both process trains. The reduction of the high molecular weight (HMW)

hydrophobic humic substances and biopolymer fractions (which possibly include nitrogenous organic compounds) before chlorine disinfection, which is applied in both process trains, decreases the potential for formation of potentially carcinogenic disinfection by-products such as total trihalomethanes. The relative increase in the LMW fractions, which are generally more biodegradable, could potentially increase bacterial re-growth in the distribution system but the application of chlorination in both plants should minimize this.

To further evaluate the performance of the two process trains in terms of NOM removal, LC-OCD data for a selection of sampling dates were examined in more detail. Since it removed the most DOC and is also a process that is routinely used to optimize DOC removal, the coagulation/flocculation process was used for the evaluation. In order to achieve the treatment plants' objective of maintaining a TOC concentration of  $\leq 2 \text{ mg L}^{-1}$  in product water (there was no statistical difference between TOC and DOC for



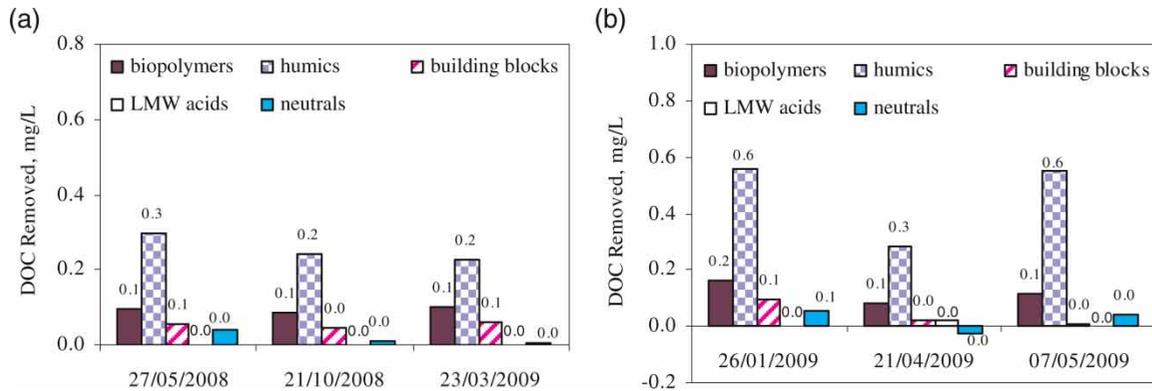
**Figure 4** | (a) Mean DOC concentrations of the five LC-OCD fractions for raw and treated waters, (b) mean DOC removal across each treatment process, and the fractional DOC composition of (c) raw and (d) treated waters for samples collected across Neuilly-sur-Marne treatment train. The error bars indicate 95% confidence intervals ( $n = 3$ ).

both plants), a calculated coagulant dose, which includes the raw water TOC as one of the parameters, is applied in both plants. As this objective was generally met on all the sampling dates for both plants, the performance was evaluated in terms of the removal efficiency of specific NOM fractions as measured by LC-OCD.

For the CR process train, the removal of LC-OCD fractions by coagulation/flocculation was evaluated for two cases in each of which three samples were selected: (1) with similar raw water DOC concentrations and coagulant dosages (Figure 5(a)), and (2) with different raw water DOC concentrations and coagulant dosages (Figure 5(b)). For the first case, the selected samples had DOC concentrations of 2.3–2.6 mg L<sup>-1</sup> and the applied coagulant doses were 13.9–14.4 mg Al/L. The SUVA was ~2.0 L mg<sup>-1</sup> m<sup>-1</sup> and turbidity 2.4–4.8 NTU for all of the three samples. The similarity in DOC concentrations, the SUVA values, and the applied Al<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub> doses was reflected in the removal of the NOM fractions that are amenable to coagulation. The amounts of the HMW biopolymer and humic fractions

as well as the building blocks removed were similar for all of the three samples.

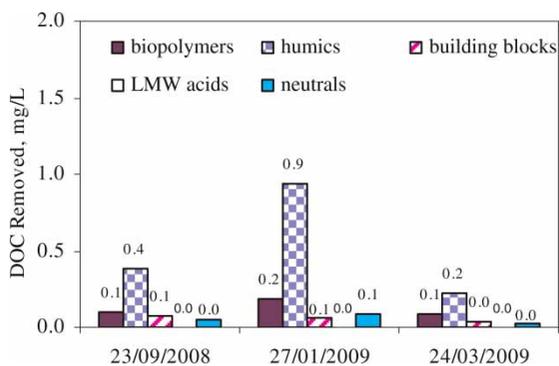
For the second case, comprising samples with different DOC concentrations and Al<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub> doses (Figure 5(b)), the DOC concentrations were ~3.5 mg L for January and May 2009 samples, and 2.5 mg L for April 2009 sample. Al<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub> doses were ~32 mg Al/L for January and May 2009 samples, and 14.4 mg Al/L for April 2009 sample. The SUVA values were 3.0 L mg<sup>-1</sup> m<sup>-1</sup> for January, 2.0 L mg<sup>-1</sup> m<sup>-1</sup> for April, and 2.4 L mg<sup>-1</sup> m<sup>-1</sup> for May. The removal of humic substances (0.6 mg L<sup>-1</sup>) was similar for January and May samples, indicating that the difference in SUVA values did not significantly affect the removal efficiency of this fraction. In contrast, whereas 0.1 mg L<sup>-1</sup> of building blocks was removed for the January sample, hardly any was removed for the May sample, indicating that the lower SUVA for the latter may have made the removal of this fraction more difficult. The removal of humic substances for April sample (0.3 mg L<sup>-1</sup>) was 50% of that for January or May, which is in roughly the same



**Figure 5** | DOC of LC-OCD fractions removed by coagulation process of Choisy-le-Roi treatment plant on three selected dates with (a) similar raw water DOC and coagulant dose, and (b) different raw water DOC and coagulant dose.

ratio as the applied  $\text{Al}_2(\text{SO}_4)_3$  doses. As for May sample, which had a similar SUVA, significantly less building blocks were removed for April as for January sample, which had a higher SUVA.

For NM process train, LC-OCD analyses were performed for three of the seven sets of samples and the DOC of LC-OCD fractions removed by coagulation/flocculation are shown in Figure 6. For these sampling dates, the DOC concentrations and SUVA values were higher for January 2009 than for either September 2008 or March 2009 samples. Furthermore, the raw water turbidity was significantly higher for the January sample. Consequently, the  $\text{Al}_2(\text{SO}_4)_3$  dose for the sample of January (50 mg Al/L) was more than twice as much as for the other two samples (~20 mg Al/L). There was a correspondingly higher removal of the HMW (biopolymers and humics) and neutral fractions for the January 2009 sample than for the other two samples dates. However, whereas

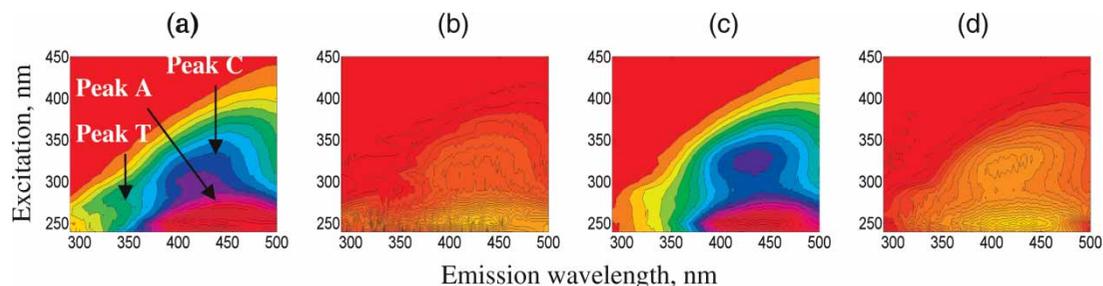


**Figure 6** | DOC of LC-OCD fractions removed by coagulation process on three selected samples with different raw water DOC concentrations and coagulant doses collected from Neuilly-sur-Marne water treatment plant.

the removal of biopolymers (normalized to the applied  $\text{Al}_2(\text{SO}_4)_3$  dose) was the same (0.004 mg C/L per mg Al/L) for the three samples, that of humics was the same (0.02 mg C/L per mg Al/L) for September 2008 and January 2009 but less (0.01 mg C/L per mg Al/L) for March 2009. The lower removal for the latter is consistent with its lower SUVA (1.7 L  $\text{mg}^{-1} \text{m}^{-1}$ ) compared to that for September (2.6 L  $\text{mg}^{-1} \text{m}^{-1}$ ) or January (3.1 L  $\text{mg}^{-1} \text{m}^{-1}$ ).

### Fluorescence EEMs

Three main fluorescence intensity peaks were obtained for all samples from both CR and Neuilly process trains that were analyzed. These previously identified peaks were observed at the following excitation and emission wavelengths: humic-like fluorescence (peak A) at 240–260 nm and 420–470 nm, respectively; fulvic-like fluorescence (peak C) at 300–340 nm and 400–450 nm respectively; and tryptophan-like fluorescence (peak T) at 240–280 nm and 300–360 nm, respectively. Figure 7 shows typical contour plots of F-EEMs for raw and product water samples for CR and NM, respectively. In both cases, the fluorescence of the raw and treated waters was dominated by the humic-like peak A. There was substantial reduction of all of the three fluorescence peaks across the two treatment process trains. The percentage reduction (relative to the raw water) of the three peaks across the treatment processes were similar for both plants: 55% after coagulation/flocculation; 85% after BAC filtration; and 86% after chlorination (final water).



**Figure 7** | Fluorescence EEM contour plots for raw (a and c) and treated (b and d) water samples collected from Choisy-le-Roi (a and b) and Neuilly-sur-Marne (c and d) water treatment plants on 22 July and 27 March 2008, respectively.

### PARAFAC components extracted from fluorescence EEM

A dataset comprising fluorescence EEMs for 145 water samples from both CR and NM were used for PARAFAC analysis. The analysis produced five models with the number of components in each ranging from 3 to 7. These models were subjected to a series of tests in order to determine the one with the most appropriate number of components. Split-half analyses were carried out for all the five models

but only the three, the four and the five component models could be split-half validated. These were split-half validated in the sense that the corresponding components in the split halves had equal excitation and emission loadings as verified by the corresponding Tucker's congruence coefficients being greater than 0.95 (Lorenzo-Seva & Berge 2006). Of the three validated models, only the one with the highest (five) number of components was considered for further analysis.

The five components of the selected model have spectral features similar to those previously extracted from

**Table 2** | Comparison of the spectral characteristics of five components identified in this study with those of similar components identified in previous studies

Component of this study	Excitation/Emission wavelength (nm)	Description and source assignment (References)
1	260(360)/480	Terrestrial humic substances Peak P3: <260(380)/498, (Ref. 3) Component 3: 270(360)/478, (Ref. 4) Component 3: 275(390)/479, (Ref. 7)
2	250(320)/410	Terrestrial/anthropogenic humic substances Component 6: <250(320)/400, (Ref. 5) Component 2: 315/418, (Ref. 2) Component 3: 295/398, (Ref. 6) Component 3: 250(310)/400, (Ref. 9)
3	<250(290)/360	Amino acids, free or protein bound Component 7: 240(300)/338, (Ref. 3) Component 4: <260(305)/378, (Ref. 8) Component 6: 250(290)/356, (Ref. 9)
4	<250(300)/406	Terrestrial humic substances Component 1: <260(305)/428, (Ref. 8) Component 3: 295/398, (Ref. 6) Peak C or M: (Ref. 1)
5	270/306	Amino acids, free or protein bound Component 4: 275/306, (Ref. 6) Component 8: 275/304, (Ref. 5) Peak B: 275/310, (Ref. 1)

Secondary excitation wavelength is given in brackets.

Ref. 1. (Coble 1996), Ref. 2. (Murphy et al. 2006), Ref. 3. (Murphy et al. 2008), Ref. 4. (Stedmon et al. 2003), Ref. 5. (Stedmon & Markager 2005a), Ref. 6. (Stedmon & Markager 2005b), Ref. 7. (Yamashita et al. 2008), Ref. 8. (Yamashita & Jaffe 2008), Ref. 9. (Kowalczyk et al. 2009).

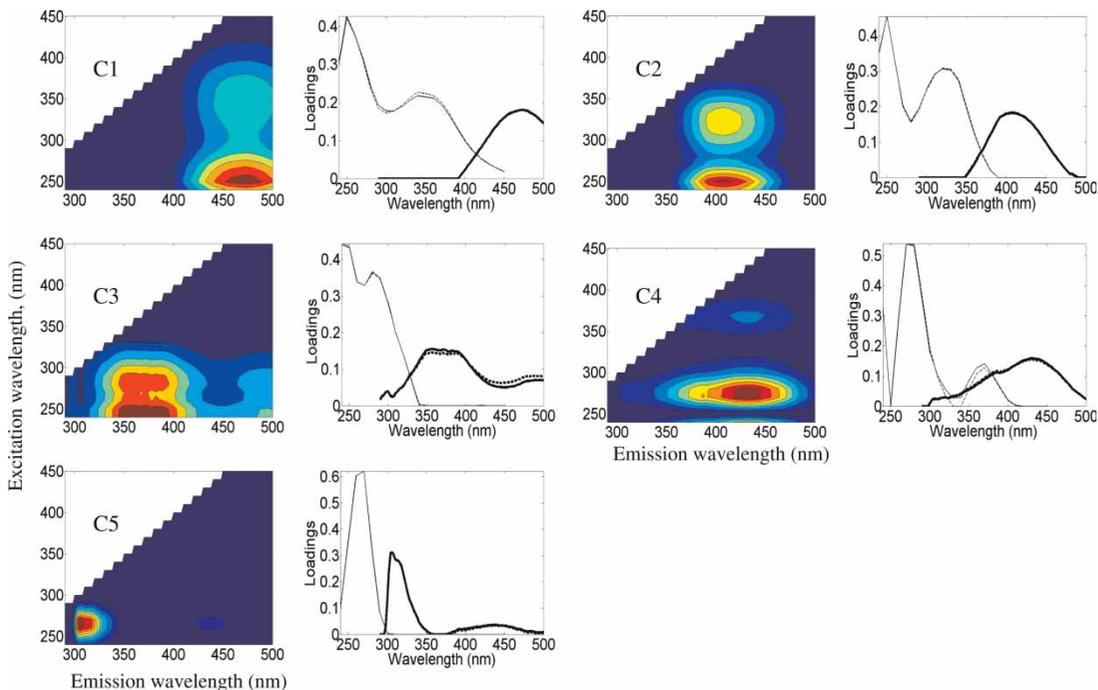
fluorescence EEMs of dissolved organic matter (DOM) (Stedmon *et al.* 2003; Stedmon & Markager 2005a, b; Murphy *et al.* 2006; Murphy *et al.* 2008; Yamashita & Jaffe 2008; Yamashita *et al.* 2008; Borisover *et al.* 2009; Zhang *et al.* 2009). Table 2 shows a comparison of the excitation and emission wavelengths for the fluorescence maxima of the five components identified in this study with those of similar components identified in previous studies. Three humic-like components of terrestrial origin were identified: two dominant ones, component 1 (C1) and component 2 (C2); and a secondary one, component 4 (C4). Two of the components have excitation/emission characteristics similar to those of fluorescent protein-like compounds (Cory & McKnight 2005): component 3 (C3) is spectrally similar to tryptophan-like fluorophore; and component 5 (C5) is spectrally similar to tyrosine-like fluorophore.

The spectral contour plots and excitation and emission spectra of each of the identified components are shown in Figure 8. The spectra show relative fluorescence intensities (loadings), in Raman units, as a function of excitation and emission wavelengths for the complete dataset (solid) and for one of the independent halves used for validation (dotted).

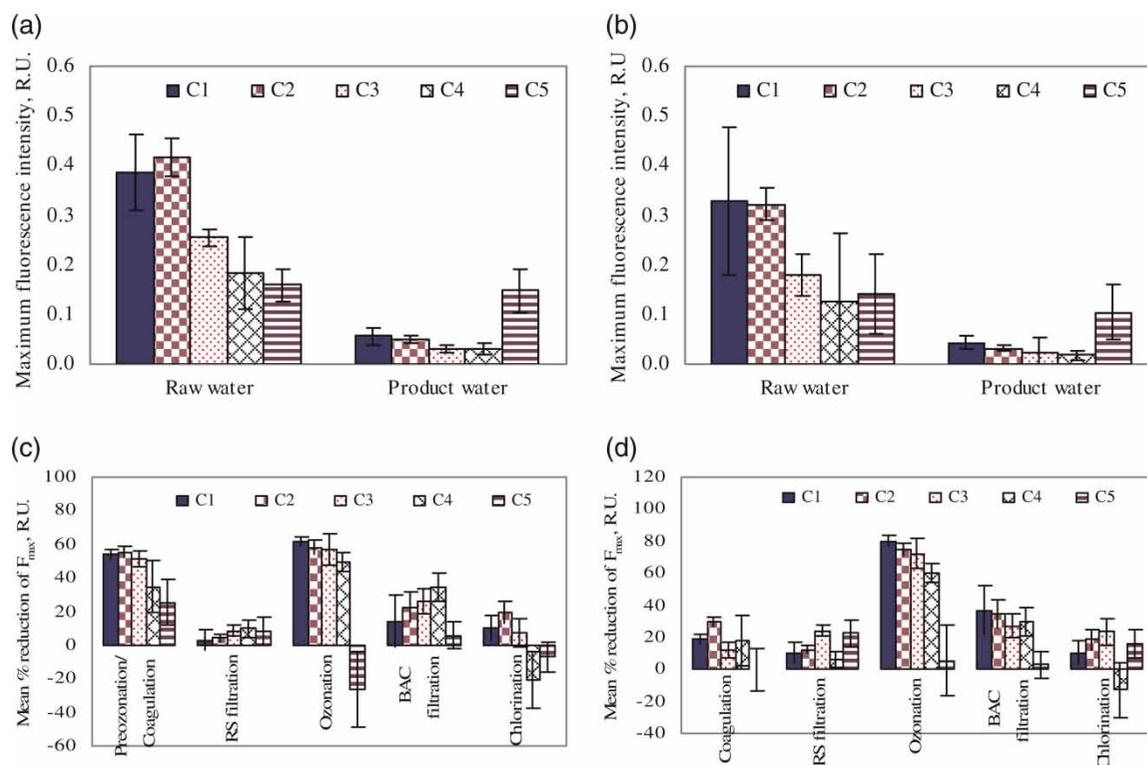
### PARAFAC component scores across treatment

Figures 9(a) and (b) show the average maximum fluorescence intensity ( $F_{\max}$ ) of the five components for raw and product waters for CR and NM treatment trains. These fluorescence intensities give estimates of the relative concentrations of each component. For both plants, raw water samples exhibited higher  $F_{\max}$  for terrestrial humic-like components C1 and C2 than for the other three components. Whereas the results appear to indicate that the raw water samples were dominated by humic-like fluorescent compounds, they are not sufficient to draw conclusions about the relative concentrations of all the seven components without prior knowledge of their respective quantum yields. However, results of LC-OCD also showed quantitatively that for both treatment plants, humic substances comprised on average 50–60% of all samples analyzed. For both plants,  $F_{\max}$  for tyrosine-like component C5 was generally stable across the treatment.

Figures 9(c) and (d) show the mean percentage  $F_{\max}$  reduction (relative to influent  $F_{\max}$  at each process) for



**Figure 8** | Contour plots of the five components identified by PARAFAC using the combined EEMs dataset of Choisy-le-Roi and Neuilly-sur-Marne samples. The line plots on the right show split-half validations of excitation (thin) and emission (thick) loadings between the complete dataset (solid) and one of the independent halves (dotted). The nearly perfect coincidence between the excitation and emission spectra of the complete dataset and of the independent halves shows the robustness of the model.



**Figure 9** | Maximum fluorescence intensities of PARAFAC components of raw and product waters for (a) Choisy-le-Roi and (b) Neuilly-sur-Marne treatment plants, and the mean percentage reductions across each treatment process at (c) Choisy-le-Roi and (d) Neuilly-sur-Marne. The error bars indicate 95% confidence intervals of the mean.

the five components across the two treatment plants. For CR,  $F_{max}$  for humic-like components C1, C2 and C4, and tryptophan-like component C3 were reduced by 30–60% after preozonation/coagulation, by 50–60% after ozonation, and by 10–30% after BAC filtration. The effective removal by coagulation process of HMW organic substances consisting of humic substances and proteins, represented by humic-like and tryptophan-like components respectively, has been observed in other studies (Haberkamp *et al.* 2007; Humbert *et al.* 2007). It should be noted, however, that except for coagulation and filtration processes, the reduction in fluorescence does not always result in DOC reduction. Oxidation processes like ozonation only transform HMW NOM into smaller and less aromatic organic compounds which have lower UV absorptivities and fluorescence. For NM, coagulation reduced  $F_{max}$  for components C1, C2, C3 and C4 by 15–30%, which is substantially less than that for CR. The higher reduction for the latter may be due mainly to the preozonation, which is applied in CR but not in NM; this may also

partly explain why the reduction by ozonation is higher for NM (~60–80%) than for CR. The effect of ozonation is not *intact* removal of a component but rather quenching of its fluorescence.

## CONCLUSIONS

Based on the characterization of NOM in water samples from CR and NM drinking water treatment plants using bulk NOM measurements, F-EEMs and LC-OCD, the following conclusions can be drawn from this study:

- Whereas the treated water DOC concentrations were relatively stable for both treatment plants, indicating the effectiveness of bulk DOC removal, the SUVA values were not as stable, indicating that the NOM character of the treated water is more variable.
- Fluorescence and LC-OCD measurements both showed that the raw water treated at the two water treatment plants is comprised mostly of humic substances.

- For both treatment plants, the HMW fractions, comprising biopolymers and humic substances, were preferentially removed while the relative contribution of the low molecular weight fractions, comprising building blocks and neutrals, increased after treatment.
- LC-OCD results indicate that for both plants, the coagulation process is not optimized for the removal of specific NOM fractions.
- A five component PARAFAC model of F-EEMs for samples from the CR and NM drinking water treatment plants was developed, comprising three humic-like and two protein-like substances (components).
- The fluorescence of samples from both treatment plants was dominated by terrestrial humic-like components, C1 and C2.
- The modelled PARAFAC components and the LC-OCD fractions demonstrated the effectiveness of these tools for the performance evaluation of the two water treatment plants in terms of the removal of NOM fractions.

## ACKNOWLEDGEMENTS

The authors acknowledge the financial support of Syndicat des Eaux d'Ile de France (SEDIF), France.

## REFERENCES

- Andersen, C. M. & Bro, R. 2003 Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* **17** (4), 200–215.
- Borisover, M., Laor, Y., Parparov, A., Bukhanovsky, N. & Lado, M. 2009 Spatial and seasonal patterns of fluorescent organic matter in Lake Kinneret (Sea of Galilee) and its catchment basin. *Water Research*. **43**, 3104–3116.
- Bro, R. 1997 PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **38** (2), 149–171.
- Chen, W., Westerhoff, P., Leenheer, J. A. & Booksh, K. 2003 Fluorescence excitation-emission matrix regional integration to quantify spectra for dissolved organic matter. *Environmental Science and Technology* **37**, 5701–5710.
- Chow, C. W. K., van Leeuwen, J. A., Drikas, M., Fabris, R., Spark, K. M. & Page, D. W. 1999 The impact of the character of natural organic matter in conventional treatment with alum. *Water Science and Technology* **40** (9), 97–104.
- Coble, P. G. 1996 Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry* **51**, 325–346.
- Cory, R. M. & McKnight, D. M. 2005 fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter. *Environmental Science and Technology* **39**, 8142–8149.
- Croué, J.-P. 2004 Isolation of humic and non-humic fractions: structural characterization. *Environmental Monitoring and Assessment* **92**, 193–207.
- Frimmel, F. H. 1998 Characterization of natural organic matter as major constituents in aquatic systems. *Journal of Contaminant Hydrology* **35**, 201–216.
- Haberkamp, J., Ruhl, A. S., Ernst, M. & Jekel, M. 2007 Impact of coagulation and adsorption on DOC fractions of secondary effluent and resulting fouling behaviour in ultrafiltration. *Water Research* **41**, 3794–3802.
- Harshman, R. A. 1984 'How can I know if it's real?' A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling. In: *Research Methods for Multimode Data Analysis* (H. G. Law, C. W. Snyder, J. A. Hattie & R. P. McDonal, eds.) Praeger, New York, pp. 566–591.
- Harshman, R. A. & Lundy, M. E. 1994 Parafac – parallel factor analysis. *Computational Statistics & Data Analysis* **18** (1), 39–72.
- Her, N., Amy, G., McKnight, D., Sohna, 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 Research* **37**, 4295–4303.
- Huber, S. A., Balz, A., Abert, M. & Pronk, W. 2011 Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography-organic carbon detection-organic nitrogen detection (LC-OCD-OND). *Water Research* **45**, 879–885.
- Humbert, H., Gallard, H., Jacquemet, V. & Croue, J.-P. 2007 Combination of coagulation and ion exchange for the reduction of UF fouling properties of a high DOC content surface water. *Water Research* **41**, 3803–3811.
- Hunt, J. F. & Ohno, T. 2007 Characterization of fresh and decomposed dissolved organic matter using excitation-emission matrix fluorescence spectroscopy and multiway analysis. *Journal of Agricultural and Food Chemistry* **55** (6), 2121–2128.
- Kowalczyk, P., Durako, M. J., Young, H., Kahn, A. E., Cooper, W. J. & Gonsior, M. 2009 Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: interannual variability. *Marine Chemistry* **113** (3–4), 182–196.
- Kubista, M., Sjoback, R., Eriksson, S. & Albinsson, B. 1994 Experimental correction for the inner-filter effect in fluorescence-spectra. *Analyst* **119** (5), 417–419.
- Lorenzo-Seva, U. & Ten Berge, J. M. F. 2006 Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology* **2**, 57–64.
- Murphy, K. R., Ruiz, G. M., Dunsmuir, W. T. M. & Waite, T. D. 2006 Optimized parameters for fluorescence-based

- verification of ballast water exchange by ships. *Environmental Science and Technology* **40** (7), 2357–2362.
- 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. *Marine Chemistry* **108** (1–2), 40–58.
- Persson, T. & Wedborg, M. 2001 Multivariate evaluation of the fluorescence of aquatic organic matter. *Analytica Chimica Acta* **434** (2), 179–192.
- Sharp, E. L., Parsons, S. A. & Jefferson, B. 2004 The effects of changing NOM composition and characteristics on coagulation performance, optimisation and control. *Water Science and Technology: Water Supply* **4** (4), 95–102.
- Stedmon, C. A. & Bro, R. 2008 Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology and Oceanography: Methods* **6**, 572–579.
- Stedmon, C. A. & Markager, S. 2005a Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnology and Oceanography* **50** (2), 686–697.
- Stedmon, C. A. & Markager, S. 2005b Tracing the production and degradation of autochthonous fractions of dissolved organic matter by fluorescence analysis. *Limnology and Oceanography* **50** (5), 1415–1426.
- Stedmon, C. A., Markager, S. & Bro, R. 2003 Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry* **82**, 239–254.
- Thurman, E. M. 1985 *Organic Geochemistry of Natural Waters*. Martinus Nijhoff/Dr. W. Junk Publishers, Dordrecht (Netherlands).
- Weishaar, J. L. 2003 Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environmental Science and Technology* **37**, 4702–4708.
- Yamashita, Y. & Jaffe, R. 2008 characterizing the interactions between metals and dissolved organic matter using excitation-emission matrix and parallel factor analysis. *Environmental Science and Technology* **42**, 7374–7379.
- Yamashita, Y., Jaffe, R., Maie, N. & Tanoue, E. 2008 Assessing the dynamics of dissolved organic matter (DOM) in coastal environments by excitation emission matrix fluorescence and parallel factor analysis (EEM-PARAFAC). *Limnology and Oceanography* **53** (5), 1900–1908.
- Zhang, Y., Van Dijk, M. A., Liu, M., Zhu, G. & Qin, B. 2009 The contribution of phytoplankton degradation to chromophoric dissolved organic matter (CDOM) in eutrophic shallow lakes: field and experimental evidence. *Water Research* **43** (18), 4685–4697.

First received 5 October 2010; accepted in revised form 4 August 2011