Available online at www.sciencedirect.com

ScienceDirect

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

Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: From PARAFAC analysis to online monitoring wavelengths

Yulia Shutova [a], Andy Baker [b], John Bridgeman [c], Rita K. Henderson [a,d,*]

[a] UNSW Water Research Centre, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia
[b] Connected Waters Initiative Research Centre, The University of New South Wales, Sydney, NSW 2052, Australia
[c] School of Civil Engineering, University of Birmingham, Birmingham, UK
[d] School of Chemical Engineering, The University of New South Wales, Sydney NSW 2052, Australia

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form
22 January 2014
Accepted 26 January 2014
Available online 7 February 2014

Keywords:
Coagulation
Dissolved organic carbon
Drinking water treatment
Fluorescence excitation emission matrix (FEEM)
SUVA

ABSTRACT

Organic matter (OM) causes many problems in drinking water treatment. It is difficult to monitor OM concentrations and character during treatment processes due to its complexity. Fluorescence spectroscopy is a promising tool for online monitoring. In this study, a unique dataset of fluorescence excitation emission matrixes (EEMs) ($n = 867$) was collected from all treatment stages of five drinking water treatment plants (WTPs) situated in diverse locations from subtropical to temperate climate. The WTPs incorporated various water sources, treatment processes and OM removal efficiencies (DOC removal 0%–68%). Despite these differences, four common fluorescence PARAFAC components were identified for characterisation of OM concentration and treatability. Moreover, fluorescence component ratios showed site-specific statistically significant correlations with OM removal, which contrasted with correlations between specific UV absorbance at 254 nm (SUVA) and OM removal that were not statistically significant. This indicates that use of fluorescence spectroscopy may be a more robust alternative for predicting DOC removal than UV spectroscopy. Based on the identified fluorescence components, four optical locations were selected in order to move towards single wavelength online OM monitoring.

1. Introduction

Over recent decades, researchers have highlighted increasing dissolved organic carbon (DOC) concentrations in fresh water sources in the Northern Hemisphere (Clark et al., 2010; Monteith et al., 2007). These long-term increases, along with seasonal variations and high organic matter (OM) surges during extreme weather events, are presenting challenges during drinking water treatment operation, and particularly process control and optimisation, with respect to ensuring that drinking water quality guidelines are met (Worrall et al., 2002; Matilainen et al., 2011). The OM found in drinking water sources is a heterogeneous mixture of organics of

* *Corresponding author*. Tel.: +61 293855383.
E-mail address: r.henderson@unsw.edu.au (R.K. Henderson).

0043-1354/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.watres.2014.01.053

various chemical compositions, resulting in a range of sizes, charge and hydrophobicity, which may have terrestrial, microbial or anthropogenic origin. The link between OM concentration and character and water treatment process efficiency is significant. For example, OM contributes to coagulant demand where OM fractions of higher aromaticity, molecular weight and hydrophobicity are more easily removed by the coagulation process (Edzwald, 1993; Kim and Yu, 2005; Sharp et al., 2006). OM may also control disinfectant demand (Kitis et al., 2002) and, moreover, react with disinfectant to form potentially harmful disinfection by-products (DBPs) such as trihalomethanes (THMs) among others (Gallard and Von Gunten, 2002). It is therefore important to reduce OM concentration by treatment optimisation such that DBPs are minimized.

Researchers have highlighted the importance of having reproducible methods for the bulk characterisation of OM in regard to drinking water treatment (Matilainen et al., 2011). Conventional OM characterisation techniques comprise total organic carbon (TOC), colour and UV absorbance spectroscopy (Volk et al., 2002). However, there are associated limitations; for example, TOC is a concentration measurement and therefore gives no information on OM character. Similarly, colour only gives an indication of the concentration of humic and fulvic substances but no information on other OM fractions (Uyguner et al., 2007), while specific UV absorbance at 254 nm (SUVA) only gives an indication of the aromaticity of the OM (Weishaar et al., 2003). Of these, SUVA has been more closely correlated with treatability (Edzwald and Kaminski, 2009) whereby if SUVA in the raw water was higher than 4 L (mg m)$^{-1}$, the DOC removal obtained is typically higher than 50%. However, this method has the limitation that SUVA is an average value of dissolved OM aromaticity. For example, two samples with SUVA equal to 4 L (mg m)$^{-1}$ could have dramatically different OM character. One sample may contain a distribution of dissolved OM with SUVA that is normally distributed with a mean of 4 L (mg m)$^{-1}$ and narrow standard deviation, while another sample could have a highly asymmetric distribution of OM, or a wider standard deviation, and contain a substantially greater proportion of DOC with lower SUVA. Overall, a method that gives more information with respect to OM concentration, character and treatability in regards to DOC removal is required.

Recent studies have shown that fluorescence spectroscopy is a useful tool for determination of OM concentration, character and function (Beggs and Summers, 2011; Bieroza et al., 2009), particularly since the development of technology that enables the rapid acquisition of 3D fluorescence excitation-emission matrices (EEMs). Traditional classification of EEM data defined specific regions, or peaks, where DOM was found to fluoresce: Peak C (λex/λem = 350/420–480 nm) and Peak A (λex/λem = 260/380–460 nm) regions as 'humic-like' and that of Peak T (Peak T λex/λem = 275/340 nm) and B (λex/λem = 275/310 nm) as 'tryptophan-like' and 'tyrosine-like', respectively, and more generally 'protein-like' (Coble, 1996). More recently, Peak M (λex/λem = 312/380-420 nm) was identified in marine water samples and has subsequently been linked to microbial activity (Jørgensen et al., 2011). Peak T has also been associated with microbially-derived OM (Elliott et al., 2006; Henderson et al., 2008). Peak C has been linked to terrestrially-derived or reprocessed OM fractions (Baker et al., 2008). Discrete optical properties including Peak C (λex/λem = 300–370/400–500 nm) emission wavelength, the ratio of Peak T (λex/λem = 275/340 nm) to Peak C and the ratio of Peak C to absorbance at 340 nm were shown to correlate with functional assays including benzo[a]pyrene binding, alumina adsorption, hydrophilicity and buffering capacity (Baker et al., 2008). Furthermore, OM aromaticity can be estimated using a fluorescence index (McKnight et al., 2001). However, current studies tend to include analysis of EEMs using multivariate techniques that can better reveal underlying fluorescing regions, significantly advancing understanding of the fluorescence signal captured in fluorescence EEM of OM.

Multi-way analytical techniques, in particular parallel factor analysis (PARAFAC), have been shown to be useful for fluorescence EEM analysis in marine, fresh water, ground water environments, as well as in wastewater, recycled and drinking water systems (Osburn and Stedmon, 2011; Stedmon et al., 2011; Staehr et al., 2012; Esparza-Soto et al., 2011; Murphy et al., 2011; Baghoth et al., 2011). In PARAFAC modelling, fluorescence EEMs are mathematically split into a set of independent components. The number of identified components, excitation-emission positioning and intensity of components have been linked to changes of concentration and character of OM within datasets (Baghoth et al., 2011; Ishii and Boyer, 2012; Sanchez et al., 2013). For example, humic-like components were linked to different sources of OM (e.g. terrestrial, microbial) and were shown to have dissimilar treatability during water and waste treatment (Ishii and Boyer, 2012; Baghoth et al., 2011). However, PARAFAC analysis is a complex procedure, which means that despite the relative ease of data acquisition, the complex post-processing limits the wider application of fluorescence EEM spectroscopy for OM monitoring. Fortunately, recent studies suggest that there may be a small number of components that are found frequently in drinking water sources that can be linked to treatability (Baghoth et al., 2011; Sanchez et al., 2013). Furthermore, in wastewater systems, a universal PARAFAC model was determined for several different plants that indicated variability could be monitored using only a few excitation-emission wavelength pairs (Murphy et al., 2011). This suggests that, using PARAFAC, key fluorescence wavelengths that can be used to monitor OM characterisation and concentration through drinking water treatment plants universally may be identified. Since the number of studies including PARAFAC analysis of drinking water treatment plants is rare, there is a pressing need to study a wide range treatment plants with differing OM removal efficiencies.

The aim of this paper was therefore to investigate similarities and differences in OM fluorescence in samples obtained from a variety of drinking water sources that differ with respect to catchment conditions, geographical region, treatment processes and OM removal efficiency. Fluorescence data were analysed using PARAFAC to identify key optical locations that were common to a number of treatment plants and investigate the link between these wavelengths and OM character and concentration changes in order to assess their potential application for online monitoring. Overall, a set of wavelength pairs for monitoring using single wavelength probes were identified and the information from these

compared against more conventional DOC and SUVA monitoring with respect to OM removal.

2. Method

2.1. Sampling protocol

Five water treatment plants (WTPs) located in Queensland (QLD), New South Wales (NSW) and Victoria (VIC), Australia, were sampled over a year between October 2011 and September 2012 on a monthly base (Table S1): Capalaba WTP (QLD); South Maclean WTP (QLD); Grahamstown WTP (NSW); Gresford WTP (NSW); and, Yarra Glen WTP (VIC). Sites were selected in order to maximise potential fluorescence variability due to location, catchment type and treatment process applied. For example, WTPs were located in both subtropical and temperate climate zones and included both unprotected and protected reservoirs and river systems (Table S1).

Water samples were collected at each water treatment stage (Table S1). Treatment stages included: powdered activated carbon (PAC), aeration, coagulation, sedimentation, membrane and sand filtration, and disinfection processes. In Capalaba WTP, PAC was used during algal blooms, which have caused the treatment plant to be taken off-line in the past. At South Maclean WTP, aeration is used to remove dissolved gases, metals and volatile compounds and to increase the dissolved oxygen level of the water. Both Capalaba and South Maclean WTPs had chlorine dosed prior to the sand filters in order to catalyse soluble manganese oxidation. Capalaba, South Maclean and Grahamstown WTPs utilised coagulation/sedimentation processes. Yarra Glen WTP used coagulation coupled with continuous microfiltration. Gresford WTP included just microfiltration and disinfection processes.

All samples were collected in triplicate by plant operators in pre-labelled, sterilised, polypropylene (PP) 50 mL tubes which had previously been shown to have minimal fluorescent leachate (Hambly et al., 2010). Samples were kept cold and dark during overnight delivery to The University of New South Wales where they were analysed within 72 h. All samples were filtered through 0.45 μm sterilised syringe filters prior to analysis. Samples were stored at 4 °C in the dark to minimise potential sample changes. All water samples had average pH of 7.3 ± 0.6.

2.2. Analytical techniques

Each of the water samples were analysed using fluorescence spectroscopy, UV absorbance spectroscopy and dissolved organic carbon (DOC) analysis as follows:

DOC concentrations were determined using a Shimadzu TOC_{CSH} total organic carbon analyser. The nonpurgeable organic carbon (NPOC) determination method was employed. Samples were acidified to pH 2–3 using 2 M HCl and then sparged with nitrogen to remove all inorganic carbon. The resultant NPOC was calculated as a mean of three measurements from the TOC analyser.

UV absorption data was obtained using a Varian Cary 50 Bio UV/Visible spectrometer. The data was collected in triplicate using a 1 cm path length quartz cuvette (Starna, Australia) for an absorption range of 200–600 nm at an increment of 1 nm and a scan speed of 600 nm min^{-1}. UV absorption data was used to assess optical density and thus determine the need for application of either dilution or inner filtration correction factors for samples as appropriate (Ohno, 2002). UV absorption at 254 nm corrected to a path length of 1 m was used to calculate specific UV absorption, SUVA (i.e. ratio of UV_{254}:DOC).

Fluorescence EEMs were obtained using a 1 cm path length quartz cuvette (Starna, Australia) and a Varian Cary Eclipse Spectrophotometer. Fluorescence intensities were measured in triplicate at excitation wavelengths of 200–400 nm in 5 nm increments and emission wavelengths of 280–500 nm in 2 nm increments. The photomultiplier tube (PMT) voltage was set at 800 V and excitation and emission slit widths of 5 nm were utilised. Raman scans of MilliQ water in a sealed cell (Varian, Australia) were obtained at an excitation wavelength of 348 nm over the emission range of 380–410 nm, for the calculation of the area of the Raman peak. The area of the Raman peak was used to normalise the fluorescence intensity of all spectra, which are expressed in Raman units (RU) (Lawaetz and Stedmon, 2009; Murphy et al., 2010).

2.3. Data processing

Spectral correction of the fluorescence EEMs was undertaken to minimise instrumental and sample-related biases, potentially including wavelength-dependent variability in the transmission efficiency of monochromators, fluctuations in spectrometer light intensity and sample inner filter effects (Coble, 1996; Cory et al., 2010; Murphy et al., 2010). Specifically, raw data were spectrally corrected using previously determined instrument specific excitation and emission factors. Rayleigh–Tyndell and Raman scatter lines were removed to limit interference during quantitative analysis and to improve fluorescence EEM display (Zepp et al., 2004). Sample-specific matrices of correction factors for inner filter effects calculated from UV absorbance scans were applied to the corrected data (Parker and Barnes, 1957), which was then normalised to RU. Data manipulations were performed using Matlab 2009b software (Mathworks, US) and the FDOMcorr toolbox for Matlab (Murphy et al., 2010). Spectral correction was undertaken for all samples replicates, and then the average matrix of particular sample was calculated. The average matrices were used for subsequent PARAFAC modelling.

PARAFAC models were generated using DOMFluor Toolbox for Matlab (Stedmon and Bro, 2008). Fluorescence intensity below 240 nm excitation was excluded from all EEMs. The data where the first order Rayleigh line and Raman line dominated the signal were replaced with missing values. Non-negativity constraints were applied for all models. Outlier samples were excluded from the dataset, as identified using outlier tests and leverage plots. When the number of components had been chosen, split half analysis and validation were applied. All models were validated with split-half analysis. Analyses using random initialisation of the model were performed to ensure that the model derived was the least squares result and not a local minimum.

It was not possible to develop a validated universal PARAFAC model in this study, instead seven PARAFAC models

Table 1 – PARAFAC component locations and description.

Component label	Approximate location	Traditional classification	Description	PARAFAC model
Component 1	Ex: <290 (330–400) nm	Peak C + Peak A	Humic-like	A, B, C, D, F, G, H
C1	Em: 400–500 nm		Terrestrial delivered OM	
Component 2	Ex: <260 (280–340) nm	Peak C + Peak A	Humic-like	A, B, C, D, F, G, H
C2	Em: 340–460 nm		Terrestrial delivered Reprocessed OM	
Component 3	Ex: <290 nm	Peak A	Humic-like	A, B, C, D, H
C3	Em: 340–500 nm		Terrestrial delivered	
Component 4	Ex: <250 (260–300) nm	Peak T + Peak B	Protein-like	A, C, D, F, G, H
C4	Em: 320–360 (280–300) nm		Microbial delivered	
Component 5	Ex: <260 (300–400) nm	Peak C + Peak A	Humic-like	C, G,
C5	Em: 400–500 nm		Terrestrial delivered	
Component 6	Ex: <300 nm	Peak B	Protein-like	G
C6	Em: 280–380 nm		Microbial delivered	
Component 7	Ex: <260 (280–340) nm	Peak M	Humic-like	H
C7	Em: 340–420 nm		Terrestrial delivered Reprocessed OM	

were developed: Model A, the raw water model, which included all source water EEMs from all sites at all time points; Models B, C, D, F and G were site-specific models; and, Model H included EEMs of water samples collected after coagulation/separation (membrane filtration or sedimentation) processes.

Tucker congruence coefficients were calculated for each excitation and emission spectrum from each WTP model (B-G) and compared to the raw water model (Model A) to determine quantitatively whether the spectra of components could be considered identical for each model.

Spearman's correlation coefficients between component score and DOC concentration and UV_{254} absorbance were calculated to investigate the possibility of OM concentration monitoring during treatment using these techniques. This was also undertaken for DOC removal by coagulation-sedimentation or coagulation-membrane filtration processes, component score removal, UV_{254} absorbance removal, source water SUVA and source water component ratios to examine OM changes during water treatment.

Selection of the specific regions for online monitoring was undertaken by choosing fluorescence excitation/emission wavelengths pairs at: (1) the maximum variation between two components excitation/emission loadings, components C1 and C2; (2) the maximum overlap between three components, C1, C2, and C3; and (3) the maximum of excitation/emission loadings of the component C4.

3. Results

3.1. PARAFAC Modelling

Within the seven developed PARAFAC models, a total of seven components were identified, named C1–C7. Components C1 and C2 appeared to be common to all models, while components C3 and C4 appeared to be present in six out of seven and five out of seven models, respectively (Table 1, Fig. 1 and S1). C1, C2, C3 and C4 may be considered almost identical between the models, as most of the correlations showed Tucker congruence coefficients higher than 0.95 (Table S2) (Lorenzo-Seva and ten Berge, 2006). In addition to Tucker congruence coefficients, we assessed how similar the PARAFAC components were between PARAFAC models by examining if the r^2 and slope of linear correlations between component scores were close to 1.0 (Table S3) (Murphy et al., 2011). A strong linear correlation was observed between components scores in independent models and in Model A (the raw water model) and Model H (the coagulated water model) in most of the cases (Table S3). Exceptions were the correlation of C2 in Model D and correlation of C4 in Models C and G with $r^2 < 0.55$. Some of the components were found to be interchangeable (e.g. C1 in Models A, B, C, D, F, and H; C2 in Model C and A). However, it was not the case for the majority of components scores due to variations in the slopes. Therefore, C1, C2, C3 and C4 spectra were considered comparable between the models but component scores were not interchangeable between the models.

3.1.1. *Raw water Model A*

A four component (C1–C4) PARAFAC was developed for all raw water samples (Model A) (Table 1, Fig. 1). According to traditional classification, C1 consisted of a combination of Peak A and Peak C, where terrestrial and non-processed OM would dominate (Ishii and Boyer, 2012). C2 was a composite of Peak A and Peak C, linked to terrestrial and processed OM. C3 was related to Peak A and terrestrially-delivered OM. C4 was a combination of Peak T and Peak B fluorescence, linked to microbial protein-like OM. All four components are commonly reported in surface and treated water (Bridgeman et al., 2011; Ishii and Boyer, 2012; Baghoth et al., 2011).

In all source water, C1 and C2 had higher fluorescence intensities than C3 and C4 (Table S4); for example, in samples from Capalaba WTP, C1 was seven times higher than C4. Furthermore, rivers and reservoirs exhibited contrasting OM properties. Specifically, Allyn River, Paterson River and Logan River (South Maclean) had higher variability of C1 and C2 (e.g. in Logan river water C1 was 0.85 ± 0.38 RU) in contrast to reservoir water, which tended to be very stable in all components (e.g. in Grahamstown raw water C1 was 0.79 ± 0.09 RU). The exception was Capalaba WTP where the source water had variable fluorescence intensities of 2.7 ± 0.2 RU and 1.8 ± 0.1 RU for C1 and C2, respectively, at least three times higher than observed at other reservoirs. Yarra Glen WTP source water had the lowest intensities in all four components (e.g. in feed

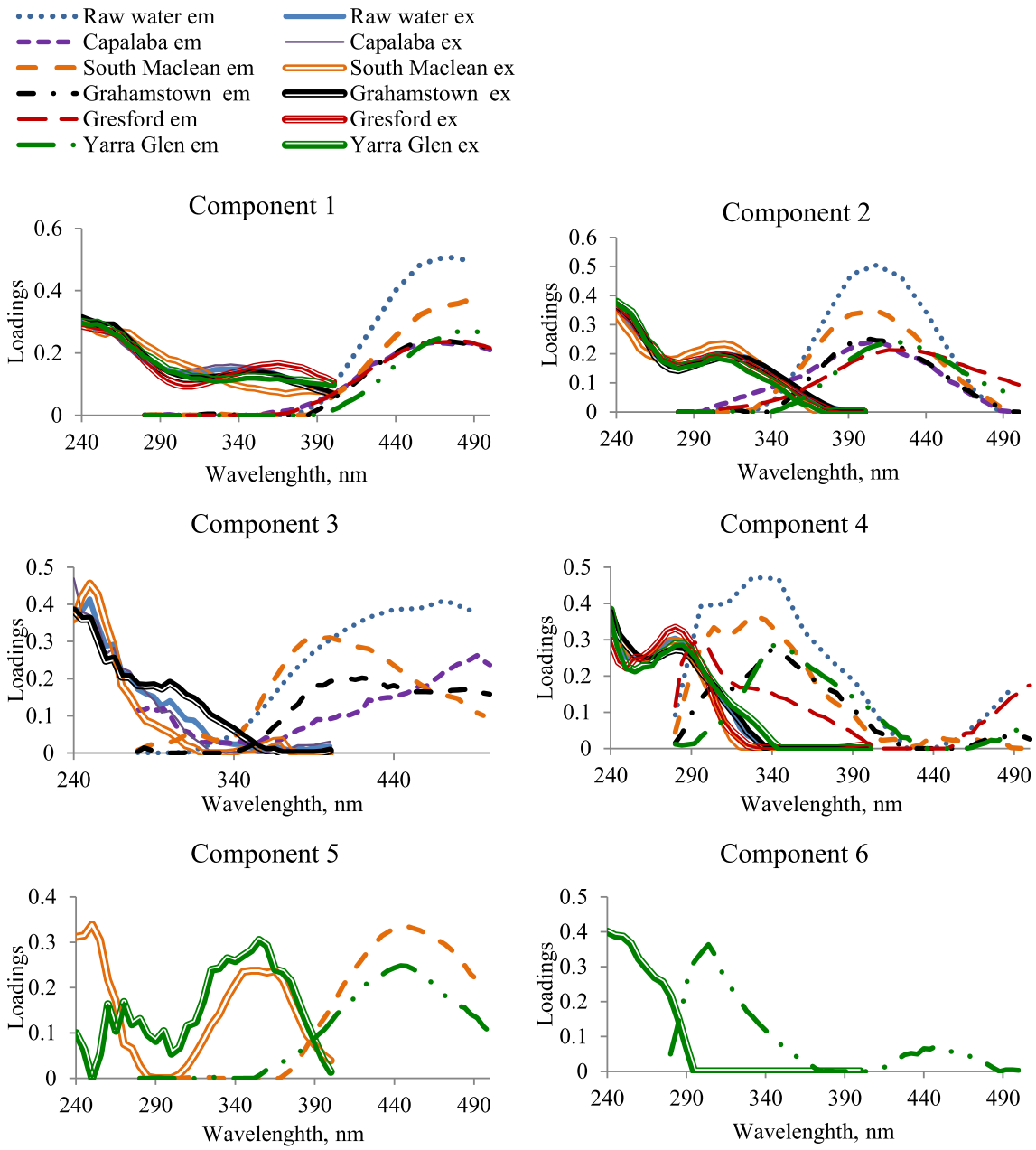

Fig. 1 – Comparison of PARAFAC components in samples from five water treatment plants combined in the Raw water model of all source water samples (Model A), and independent models of Capalaba WTP, South Maclean WTP, Grahamstown WTP, Gresford WTP and Yarra Glen WTP water samples (Models B–G). em = emission spectra, ex = excitation spectra.

water C1 was 0.49 ± 0.09 RU), and it was the only site that contained almost no C3 (Table S4). No significant seasonal variations of OM were observed during the sampling period in all sites (Table S4), potentially due to unusually high rainfall and colder summer temperatures in comparison to the means of records from the years 1967–2012 (NSW), 1953–2012 (QLD) and 1994–2012 (VIC) (Australia's official weather forecasts).

3.1.2. *WTP specific Models B – G*

PARAFAC modelling identified three to five fluorescence components present in each plant (Table 1; Fig. 1). A total of six independent components were identified. C1 and C2 were found at all sites, comparing well with those observed in Model A. C3 was not found in Gresford and Yarra Glen WTPs. C4 was identified in all models, except that of Capalaba WTP, where concentrations of OM were the highest (Table S4); hence, we hypothesise that this absence could be due to dominance of fluorophores contained in components C1 – C3. C4 and C6 had the lowest fluorescence intensity (e.g., in Model G 0.08 ± 0.04 and 0.08 ± 0.02, respectively) in comparison to other peaks. In the Models C–F, PARAFAC had difficulty distinguishing between two components that are in similar positions to Peak T and Peak B; therefore, these models returned hybridised spectra of C4. However, for Yarra Glen WTP (Model G), where concentrations of OM were the lowest in the dataset (Table S4), C4 was split into two components (C4 and C6,

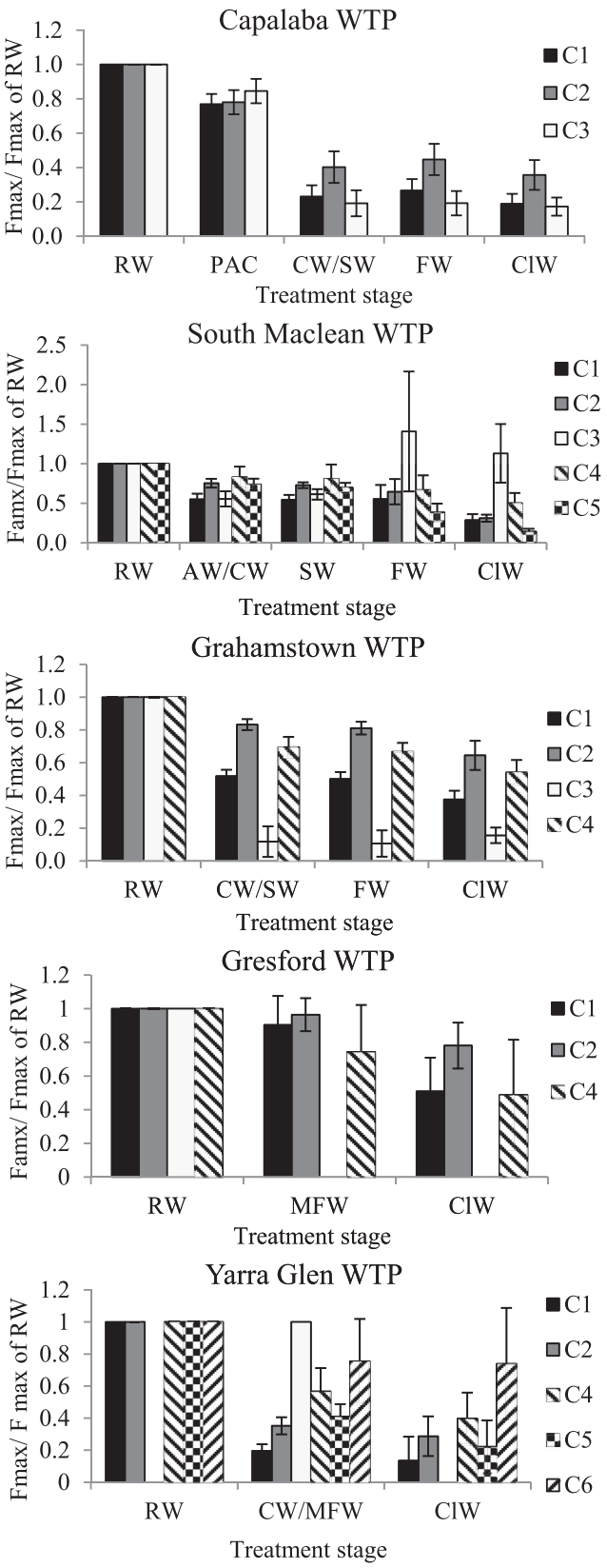

Fig. 2 – Normalised removal of components through water treatment processes in Capalaba, South Maclean, Grahamstown, Gresford and Yarra Glen water treatment plants (WTPs). Error bars are the standard deviation of samples analysed in triplicate during the 12 month sampling period. RW = raw water, PAC = powdered activated carbon treated water, AW = aerated water, CW = coagulated water, SW = settled water, FW = sand filtered water, MFW = membrane filtered water, ClW = chlorinated water. C1–C6 are the PARAFAC components.

Fig. 1), which were linked to Peak T (tryptophan-like) and Peak B (tyrosine-like) respectively (Table 1). Component C5, linked to terrestrially-delivered OM, was identified in South Maclean and Yarra Glen WTP models (Model C and G, respectively), but was not found in the raw water model (Model A).

3.1.3. *Coagulated water Model H*

A five component model was developed using EEMs of water samples obtained after clarification but prior to sand filtration and chlorination treatment stages (Model H). Components C1, C2, C3, and C4 were identified in Model H and compared well with those observed in Models A-G (Figure S1, Tables S2 and S3). Component C7 was unique to Model H and was associated with the Peak M fluorescence area (Table 1; Coble, 1996). C2 excitation loadings were similar in both Model A and H but emission loadings were lower in coagulated water and the maximum of the peak shifted from 408 nm to 430 nm (Figure S1). Component C7, that was unique to Model H, had similar excitation loadings to C2, but a much lower emission loadings peak at 370 nm (Figure S1).

3.2. *OM removal*

3.2.1. *DOC removal*

The coagulation-sedimentation and coagulation-membrane filtration processes removed the majority of OM; for example, the highest DOC removal was achieved in Yarra Glen (54 ± 10%) and Capalaba (51 ± 11%) WTPs, followed by Grahamstown (37 ± 3%) and South Maclean (31 ± 13%) WTPs. Sand filtration in Grahamstown and Capalaba WTPs as well as membrane filtration in Gresford WTP did not affect OM concentrations as determined by DOC, fluorescence and UV_{254} absorbance analysis. In South Maclean WTP, all components decreased after the aeration/coagulation stage and there was no significant difference in DOC, component intensities or UV_{254} absorbance after sedimentation (Table S5).

3.2.2. *PARAFAC component removal*

Similarly to DOC removal, the majority of fluorescent OM was removed during coagulation-sedimentation and coagulation-membrane filtration processes; for example, in Capalaba WTP samples, the highest OM removal of 80 ± 8% was achieved for C3 while the lowest removal was observed for C2 at 55 ± 10%. Similarly to Capalaba WTP, in Grahamstown WTP, C3 and C2 had the highest (88 ± 9%) and lowest (17 ± 3%) removal efficiencies, respectively. In Yarra Glen WTP, maximum and minimum fluorescent OM removal was achieved for C1 (80 ± 4%) and C4 (43 ± 15%) respectively. It was also noted that components C1 and C3 were removed preferentially at Capalaba (C1 – 74 ± 7%, C3 – 74 ± 7%), South Maclean (C1 – 46 ± 6%, C3 – 39 ± 7%), Grahamstown (C1 – 48 ± 4%, C3 – 88 ± 9%) and Yarra Glen WTPs (C1 – 80 ± 4%).

Table 2 – Summary of Spearman's rank correlation coefficients between PARAFAC component scores (in Models B, C, D, F and G), UV_{254} absorbance and dissolved organic carbon (DOC) concentrations (* correlations are significant, $p < 0.05$, p-value calculated for two tailed test, chlorinated water samples were excluded from the dataset).

Site	DOC concentration				
	Capalaba	South Maclean	Grahamstown	Gresford	Yarra Glen
Component 1 score	0.89*	0.63*	0.65*	0.78*	0.61*
Component 2 score	0.87*	0.61*	0.49*	0.86*	0.55*
Component 3 score	0.74*	0.61*	0.70*		
Component 4 score		0.62*	0.67*	0.84*	0.40*
Component 5 score		0.48*			0.48*
Component 6 score					0.46*
UV_{254} absorbance	0.86*	0.72*	0.60*	0.87*	0.48*

C2 and C4 removal was particularly poor in South Maclean (C2 – 27 ± 4%, C4 – 20 ± 16%) and Grahamstown WTPs (C2 – 17 ± 3%, C4 – 31 ± 6%). This indicates that terrestrially-delivered OM (C1 and C3) had higher treatability than processed or microbially-delivered OM (C2 and C4). In South Maclean WTP, located in a subtropical climate, there was a significant increase of fluorescence intensities for C3 at the sand filtration stage (Fig. 2, Table S5), which was supported by an increase of DOC concentrations from 3.4 ± 0.8 mg L^{-1} to 4.1 ± 0.8 mg L^{-1} over the 12 month period (Table S5).

3.2.3. *Site-specific correlations*

Based on an individual model's data (Models B, C, D, F, and G), correlations between component scores, UV_{254} absorbance and DOC concentrations were site specific (Table 2). To illustrate, for Capalaba WTP, C1 had the highest correlation with DOC concentration (r_s = 0.89) in comparison to other sites, followed by C2 (r_s = 0.87) and UV_{254} absorbance (r_s = 0.86). Results obtained for Grahamstown, Gresford and South Maclean WTP samples were in a medium range of DOC (Table S4). In these sites, the relationship between fluorescence components, UV_{254} absorbance and DOC concentrations was weaker than for Capalaba WTP samples (Table 2). The strongest correlations of DOC concentration was observed with C3 score (Grahamstown WTP, r_s = 0.71), C2 score (Gresford WTP, r_s = 0.86) and UV_{254} absorbance (South Maclean WTP, r_s = 0.73). DOC concentrations were the lowest in Yarra Glen WTP (Table S4) and relationships between DOC concentration, fluorescence components and UV_{254} absorbance had the lowest Spearman's rank correlation coefficient in comparison to other sites (Table 2).

Components C1 and C2 removal consistently had strong correlations with DOC removal in site-specific datasets in contrast to correlations with UV_{254} removal where no significant correlations were observed with the exception of Capalaba WTP (Table 3). No significant correlations were found between DOC removal and other parameters in Grahamstown WTP (Table 3), potentially due to very low DOC removal variations 37 ± 3% during the sampling period.

It was observed that the C1:C2 ratio had a strong positive correlation with DOC removal in Capalaba WTP (r_s = 0.69) and that the C1:C4 ratio had a strong positive correlation with DOC removal in South Maclean (r_s = 0.88) and Yarra Glen (r_s = 0.70) WTPs (Table 3). Interestingly, in contrast to component ratios, there was no significant correlation observed between raw water SUVA and DOC removal in site-specific data-sets (Table 3).

3.3. *Relating OM character and treatability*

The relationship between raw water character and treatability was further investigated to determine whether fluorescence could be used to predict DOC removal by coagulation processes, for example, in a similar way in which SUVA is

Table 3 – Summary of Spearman's rank correlation coefficients between PARAFAC component scores removal, UV_{254} absorbance removal and dissolved organic carbon (DOC) removal (*correlations are significant, $p < 0.05$, p-value calculated for two tailed test, chlorinated water samples were excluded).

Site	DOC removal				
	Capalaba	South Maclean	Grahamstown	Yarra Glen	All sites combined
	Model B	Model C	Model D	Model G	Models A–G
Component 1 removal	0.74*	0.54	0.42	0.95*	0.81*
Component 2 removal	0.69*	0.75*	0.51	0.92*	0.73*
Component 3 removal	0.70*	0.29	0		0.38*
Component 4 removal		0.14	0.06	−0.17	0.41*
Component 5 removal		0.54		0.77*	
Component 6 removal				−0.14	
UV_{254} absorbance removal	0.72*	0.22	−0.25	0.50	0.69*
SUVA raw water	0.26	0.58	0.15	−0.18	0.50*
Component1/component 2	0.69*	0.14	0.26	−0.25*	0.64*
Component1/component 3	0.26	−0.36	0.07		0.42*
Component1/component 4		0.88*	−0.01	0.70*	0.60*

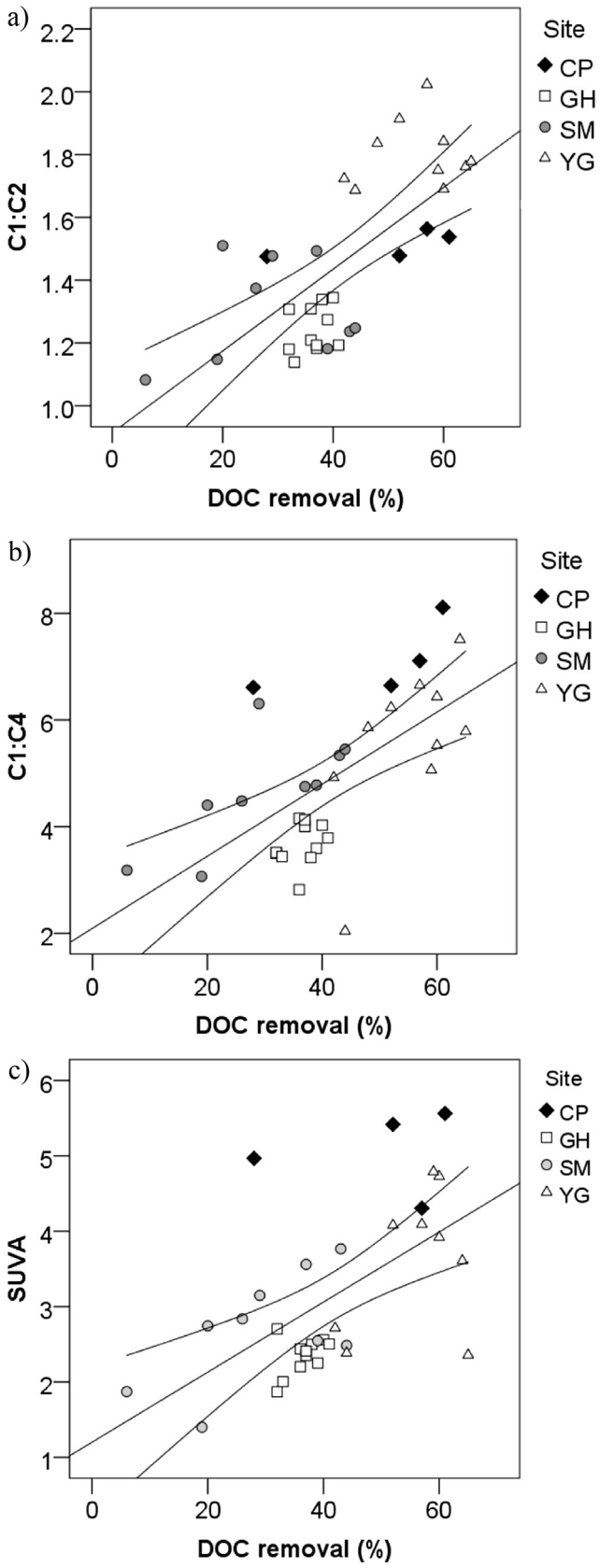

Fig. 3 – Correlation between variations of PARAFAC component score ratios and dissolved organic carbon (DOC) removal when using coagulation/sedimentation and coagulation/membrane filtration processes, for: a) C1:C2 = Component 1 to Component 2 ratio; b) C1:C4 = Component 1 to Component 4 ratio; and, c) SUVA = specific ultraviolet absorbance at 254 nm. CP = Capalaba WTP, GH = Grahamstown WTP, SM = South Maclean WTP, YG = Yarra Glen WTP. Lines represent mean confidence intervals.

currently utilised (Weishaar et al., 2003). To this end, the following parameters were plotted against DOC removal: C1:C2 ratio, C1:C4 ratio and SUVA (Fig. 3). Note that Gresford WTP was excluded from this particular analysis as coagulation was not performed and therefore results were not comparable with those from other treatment plants. Component C3 removal and C1:C3 ratio had the weakest correlation with DOC removal (Table 3) and, therefore, C3 is not further discussed.

Similarly to site-specific relationships (Table 4), significant positive correlations between the C1:C2 ratio ($r_s = 0.64$), C1:C4 ratio ($r_s = 0.60$), SUVA ($r_s = 0.50$) and DOC removal were observed in all sites where coagulation treatment was applied (Table 3, Fig. 3). For example, if the C1:C2 ratio was higher than 1.4, the DOC removal was demonstrated to be higher than 40%. Furthermore, there was no statistically significant difference between SUVA or fluorescence component ratio correlations with DOC removal when combining all sites ($p > 0.05$) (Table 3).

Multiple regression analysis was applied in order to investigate if a combination of the C1:C2, C1:C4 and SUVA or just C1:C2 and C1:C4 ratios can be used to predict DOC removal (Equations (1) and (2)). Results showed that there was a significant relationship between DOC removal and the combination of C1:C2, C1:C4 and SUVA ($p = 0.000017$), and also with just C1:C2 and C1:C4 ratios ($p = 0.000003$). However, exclusion of SUVA from Equation (1) had no demonstrable effect (Equation (2)) and r^2 remained the same. Correlation between SUVA and DOC removal alone was significant ($p = 0.00053$) but had a lower r^2 of 0.35.

$$\text{DOC removal} = 26.053\,(\text{C1:C2}) + 3.137\,(\text{C1:C4}) + 0.072\,(\text{SUVA}) - 11.912;\ (r^2 = 0.56,\ p = 0.000017) \quad (1)$$

$$\text{DOC removal} = 26.161\,(\text{C1:C2}) + 3.171\,(\text{C1:C4}) - 12.003;\ (r^2 = 0.56,\ p = 0.000003) \quad (2)$$

Site-specific relationships between OM character and DOC removal can also be observed in Fig. 3. The data was observed to form two major clusters. The cluster at higher DOC removal values included Capalaba and Yarra Glen WTP samples, where OM had a higher proportion of more aromatic, hydrophobic, charged material and therefore higher DOC removal in comparison to the other cluster. This latter cluster comprised samples from South Maclean and Grahamstown WTPs with a higher proportion of less aromatic and more hydrophilic OM in comparison to the first group and therefore consistently less than 50% DOC removal was observed in these sites (Fig. 3). These relationships were conducted under the assumption that treatment processes were optimised; if this was not the case then it is possible that this correlation can be improved.

Table 4 – Potential points (P) for online monitoring.

Point label	Corresponding component	Excitation wavelength (nm)	Emission wavelength (nm)
P1	Component 1	380	488
P2	Component 2	310	392
P3	Components 1,2,3	240	440
P4	Component 4	280	328

3.4. *Selection of wavelengths for online monitoring*

Components C1, C2, C3 and C4 were found to be common for all PARAFAC models and useful for tracking OM changes during water treatment. Component C3 has not been shown to be useful for OM removal and treatability monitoring; however, it may by employed to monitor DOC concentrations in some cases (for example at Capalaba, South Maclean and Grahamstown WTPs, Table 2). Based on these results, four points for online monitoring were chosen (Table 4). On plotting component scores against fluorescence intensity, a linear correlation was observed for points P1, P2 and P4 with r_s varying from 0.90 to 0.99 in both raw water and coagulated water models (Figure S2, Table S6, S7). There was a weaker correlation between C3 and point P3 intensity with r_s of 0.77 in raw water and 0.84 in coagulated water as P3 is related to the number of components (Table S6, S7). A linear correlation was observed for point P3 and sum of C1, C2 and C3 scores with r_s varying from 0.99 to 0.99 in both raw water and coagulated water models.

4. Discussion

4.1. *OM treatability*

It was observed that terrestrially delivered OM (C1 and C3) had higher treatability than processed or microbially-delivered OM (C2 and C4). A similar finding was presented by Baghoth et al. (2011) where preferential reduction of humic-like components relative to protein-like components was observed during coagulation. The fact that it is possible to identify fluorescent OM fraction that have different treatability on coagulation means they may be used for assessing source water character in terms of prediction of OM removal. Thus, terrestrially- and microbially-delivered OM fractions play an important role in OM matrix, especially at lower concentrations of DOC and in a cooler climate, where fluorescence spectra have a better correlation with OM concentration and changes than UV_{254} absorbance (Tables 2 and 3). This is in agreement with previous research, where UV absorbance was reported as a less selective technique than fluorescence spectroscopy (Bieroza et al., 2009). In warmer climates and at higher concentrations of DOC, such as those experienced at Capalaba WTP, terrestrially-delivered fluorescence spectra were shown to correlate well with OM concentration and changes through the WTP.

It was demonstrated that with an increase of the C1:C2 ratio in raw water, there was an increase in DOC removal by coagulation processes. Components C1 and C2 were linked to terrestrially-derived and terrestrially-derived reprocessed OM, respectively (Table 1). An increase in the C1:C2 ratio therefore suggests the presence of more aromatic, hydrophobic, charged material versus that of lower charged and more hydrophilic content. The strong correlation observed is therefore explained as such material would be preferentially targeted by coagulation treatment (Kim and Yu, 2005; Sharp et al., 2006). Similarly to previous research (Bieroza et al., 2009), the ratio between terrestrial and microbial fractions of OM, determined by C1:C4 ratio, was also found to be useful to assess OM treatability. While the weakest relationship was between raw water SUVA and DOC removal, it was still possible to observe the general relationship frequently observed whereby if SUVA in the raw water was higher than four, the DOC removal obtained was typically higher than 50% (Edzwald, 1993). The fact that a stronger correlation was found for fluorescence across the sites indicates that use of fluorescence C1:C2 or C1:C4 ratios may be a more robust alternative to predicting DOC removal than SUVA.

In addition, increased DOC and C3 concentrations observed post-sand filtration suggest that the filter beds may host microbial communities that contribute to OM concentration in the warmer regions, especially during the summer period. This was not apparent on UV_{254} analysis. Bridgeman et al. (2014) also determined, using carbon isotopes analysis, that a new source of organic carbon may be released during water treatment, for example, from biofilms or from the abrasion of filter media, and fluorescence therefore has a potential to be used to monitor OM production.

There also was no difference between correlations observed between fluorescence components, UV_{254} absorbance and DOC concentration within the site-specific datasets ($p > 0.05$) (Table 2). Similar results were reported by Baghoth et al. (2011), where significant correlations were obtained between all PARAFAC components, UV_{254} and DOC concentrations. We suggest that this indicates a strong relationship between the fluorescent, absorbent fractions of OM and the total DOC pool. The greatest OM concentrations, component intensities and variability were observed in unprotected water sources located in subtropical climate zone (Capalaba and South Maclean WTPs) in comparison to other sites with protected catchments located in a temperate climate zone.

4.2. *From PARAFAC modelling to on-line monitoring*

The presence of non-interchangeable components, and the fact that composition of components varied site by site in raw water and in coagulated water models, can explain the impossibility of developing a universal model. This result agreed with another study (Murphy et al., 2011) where a diverse sample set of 1479 EEMs from six recycled WTPs was analysed. As Murphy et al. (2011) reported, differences between individual and global models highlight the problem of interpreting PARAFAC results of diverse sample sets, where PARAFAC global models may generate systematically biased estimates of fluorescence intensities and fluorescence ratios. Hence, significantly different fluorescence intensities and ratios of components may be predicted using individual models. However, for a smaller data-set with less diverse samples, it may be possible to have a universal PARAFAC model for all samples (e.g. Sanchez et al., 2013). The adequacy of universal or individual models should be judged accurately in each case.

Overall, there is a need to extend the use of PARAFAC analysis to identify key regions of the OM fluorescence spectra associated with OM treatability during the water treatment process. Hence, in this study, despite the differences in catchments (climate zone, protected/unprotected), types of water sources (rivers, reservoirs) and treatment technologies (different coagulants and floc separation techniques), it was possible to identify four common fluorescence PARAFAC components (C1, C2, C3 and C4) to be useful for characterisation of OM. Furthermore, key pairs of excitation/emission wavelength that have a potential for on-line monitoring were identified. Fluorescence intensities subtracted from EEMs at selected excitation/emission wavelength correlated well with PARAFAC components scores. This indicates that fluorescence intensities at points P1–P4 are representative for C1–C4 and, similarly to component ratios (Fig. 3), we propose that P1:P2 and P1:P4 ratios can be used in drinking water treatment to assess OM treatability (Figure S3).

5. Conclusion

Based on OM characterisation using fluorescence EEMs, UV_{254} absorbance, SUVA, DOC analysis, as well as PARAFAC analysis of fluorescence data, the following conclusions have been drawn:

- Despite the differences in catchments, types of water sources, treatment technologies, concentrations of OM and differences in individual PARAFAC models, it was possible to identify a number of common fluorescence parameters (C1, C2, C3, C4) that are useful for characterising OM concentrations and changes.
- PARAFAC component ratios C1:C2 and C1:C4 in source water samples had statistically significant correlations with OM treatability by coagulation-sedimentation and coagulation-membrane filtration processes.
- C1, C2, C3 and C4 fluorescence components and their ratios had statistically significant correlations with DOC removal in contrast to UV_{254} and SUVA, indicating that use of fluorescence spectroscopy may be a more robust alternative to OM monitoring.
- Currently, there is no online monitoring equipment available that combines EEM-PARAFAC; therefore, a set of wavelength pairs was identified for use as online OM monitoring points: P1 (λex/λem = 380/488 nm), P2 (λex/λem = 310/392 nm), P3 (λex/λem = 240/440 nm) and P4 (λex/λem = 280/328 nm).
- With fluorescence probes on the market that measure fluorescence in the selected points, it will be possible to use fluorescence probes to measure specific excitation and emission spectra in order to determine OM character, concentration and treatability.

Acknowledgements

This research was supported under Australian Research Council's Linkage Projects funding scheme (project number LP100200259) which included support from Seqwater, Hunter Water, Melbourne Water and Water Research Australia Ltd (WaterRA). In addition, the Authors would like to thank Water Research Australia for the PhD top-up scholarship provided for this project. Finally, the Authors thank Dr Kate Murphy for her helpful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2014.01.053

REFERENCES

Australia's official weather forecasts & weather radar – Bureau of Meteorology [WWW Document], n.d. URL http://www.bom.gov.au/ (accessed 8.7.13).

Baghoth, S.A., Sharma, S.K., Amy, G.L., 2011. Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC. Water Res. 45, 797–809.

Baker, A., Tipping, E., Thacker, S.A., Gondar, D., 2008. Relating dissolved organic matter fluorescence and functional properties. Chemosphere 73, 1765–1772.

Beggs, K.M.H., Summers, R.S., 2011. Character and chlorine reactivity of dissolved organic matter from a Mountain Pine Beetle impacted Watershed. Environ. Sci. Technol. 45, 5717–5724.

Bieroza, M., Baker, A., Bridgeman, J., 2009. Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. Sci. Total Environ. 407, 1765–1774.

Bridgeman, J., Bieroza, M., Baker, A., 2011. The application of fluorescence spectroscopy to organic matter characterisation in drinking water treatment. Rev. Environ. Sci. Biotechnol. 10, 277–290.

Bridgeman, J., Gulliver, P., Roe, J., Baker, A., 2014. Carbon isotopic characterisation of dissolved organic matter during water treatment. Water Res. 48, 119–125.

Clark, J.M., Bottrell, S.H., Evans, C.D., Monteith, D.T., Bartlett, R., Rose, R., Newton, R.J., Chapman, P.J., 2010. The importance of the relationship between scale and process in understanding long-term DOC dynamics. Sci. Total Environ. 408, 2768–2775.

Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation–emission matrix spectroscopy. Mar. Chem. 51, 325–346.

Cory, R.M., Miller, M.P., McKnight, D.M., Guerard, J.J., Miller, P.L., 2010. Effect of instrument-specific response on the analysis of fulvic acid fluorescence spectra. Limnol. Oceanogr. Methods 8, 67–78.

Edzwald, J.K., 1993. Coagulation in drinking water treatment: particles, organics and coagulants. Water Sci. Technol. 27, 21–35.

Edzwald, J.K., Kaminski, G.S., 2009. A practical method for water plants to select coagulant dosing. J. N. Engl. Water Works Assoc. 123, 15–31.

Elliott, S., Lead, J.R., Baker, A., 2006. Characterisation of the fluorescence from freshwater, planktonic bacteria. Water Res. 40, 2075–2083.

Esparza-Soto, M., Núñez-Hernández, S., Fall, C., 2011. Spectrometric characterization of effluent organic matter of a sequencing batch reactor operated at three sludge retention times. Water Res. 45, 6555–6563.

Gallard, H., Von Gunten, U., 2002. Chlorination of natural organic matter: kinetics of chlorination and of THM formation. Water Res. 36, 65–74.
Hambly, A., Henderson, R.K., Baker, A., Stuetz, R.M., Khan, S.J., 2010. Probabilistic analysis of fluorescence signals for monitoring dual reticulation water recycling schemes. Water Sci. Technol. 62, 2059.
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, 3435–3445.
Ishii, S.K.L., Boyer, T.H., 2012. Behavior of reoccurring PARAFAC components in fluorescent dissolved organic matter in natural and engineered systems: a critical review. Environ. Sci. Technol. 46, 2006–2017.
Jørgensen, L., Stedmon, C.A., Kragh, T., Markager, S., Middelboe, M., Søndergaard, M., 2011. Global trends in the fluorescence characteristics and distribution of marine dissolved organic matter. Mar. Chem. 126, 139–148.
Kim, H.-C., Yu, M.-J., 2005. Characterization of natural organic matter in conventional water treatment processes for selection of treatment processes focused on DBPs control. Water Res. 39, 4779–4789.
Kitis, M., Karanfil, T., Wigton, A., Kilduff, J.E., 2002. Probing reactivity of dissolved organic matter for disinfection by-product formation using XAD-8 resin adsorption and ultrafiltration fractionation. Water Res. 36, 3834–3848.
Lawaetz, A.J., Stedmon, C.A., 2009. Fluorescence intensity calibration using the Raman scatter peak of water. Appl. Spectrosc. 63, 936–940.
Lorenzo-Seva, U., ten Berge, J.M.F., 2006. Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology 2, 57–64.
Matilainen, A., Gjessing, E.T., Lahtinen, T., Hed, L., Bhatnagar, A., Sillanpää, M., 2011. An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. Chemosphere 83, 1431–1442.
McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnol. Oceanogr. 46, 38–48.
Monteith, D.T., Stoddard, J.L., Evans, C.D., de Wit, H.A., Forsius, M., Høgåsen, T., Wilander, A., Skjelkvåle, B.L., Jeffries, D.S., Vuorenmaa, J., Keller, B., Kopácek, J., Vesely, J., 2007. Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. Nature 450, 537–540.
Murphy, K.R., Butler, K.D., Spencer, R.G.M., Stedmon, C.A., Boehme, J.R., Aiken, G.R., 2010. Measurement of dissolved organic matter fluorescence in aquatic environments: an interlaboratory comparison. Environ. Sci. Technol. 44, 9405–9412.
Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., Khan, S.J., 2011. Organic matter fluorescence in municipal water recycling schemes: toward a Unified PARAFAC model. Environ. Sci. Technol. 45, 2909–2916.
Ohno, T., 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. Environ. Sci. Technol. 36, 742–746.
Osburn, C.L., Stedmon, C.A., 2011. Linking the chemical and optical properties of dissolved organic matter in the Baltic-North Sea transition zone to differentiate three allochthonous inputs. Mar. Chem. 126, 281–294.
Parker, C.A., Barnes, W.J., 1957. Some experiments with spectrofluorimeters and filter fluorimeters. Analyst 82, 606–618.
Sanchez, N.P., Skeriotis, A.T., Miller, C.M., 2013. Assessment of dissolved organic matter fluorescence PARAFAC components before and after coagulation–filtration in a full scale water treatment plant. Water Res. 47, 1679–1690.
Sharp, E.L., Parsons, S.A., Jefferson, B., 2006. Seasonal variations in natural organic matter and its impact on coagulation in water treatment. Sci. Total Environ. 363, 183–194.
Staehr, P.A., Baastrup-Spohr, L., Sand-Jensen, K., Stedmon, C., 2012. Lake metabolism scales with lake morphometry and catchment conditions. Aquat. Sci. 74, 155–169.
Stedmon, C.A., Bro, R., 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. Limnol. Oceanogr. Met. 6, 572–579.
Stedmon, C.A., Seredyńska-Sobecka, B., Boe-Hansen, R., Le Tallec, N., Waul, C.K., Arvin, E., 2011. A potential approach for monitoring drinking water quality from groundwater systems using organic matter fluorescence as an early warning for contamination events. Water Res. 45, 6030–6038.
Uyguner, C.S., Suphandag, S.A., Kerc, A., Bekbolet, M., 2007. Evaluation of adsorption and coagulation characteristics of humic acids preceded by alternative advanced oxidation techniques. Desalination 210, 183–193.
Volk, C., Wood, L., Johnson, B., Robinson, J., Hai, W.Z., Kaplan, L., 2002. Monitoring dissolved organic carbon in surface and drinking waters. J. Environ. Monit. 4, 43–47.
Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. Environ. Sci. Technol. 37, 4702–4708.
Worrall, F., Burt, T.P., Jaeban, R.Y., Warburton, J., Shedden, R., 2002. Release of dissolved organic carbon from upland peat. Hydrol. Process 16, 3487–3504.
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.